

TRƯỜNG ĐẠI HỌC Y HÀ NỘI
BỘ MÔN DỊCH TỄ HỌC

THỰC HÀNH DỊCH TỄ HỌC

TRƯỜNG ĐẠI HỌC Y HÀ NỘI
THU VIỆN

**THU VIỆN
HUYT**
NHÀ XUẤT BẢN Y HỌC
Hà nội - 1995
TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

Chủ biên:

PGS, PTS. DƯƠNG ĐÌNH THIÊN

Tham gia biên soạn:

PTS, THẠC SĨ. LÊ VŨ ANH

BS. TRẦN VĂN CHÍ

PSG, PTS. PHẠM HUY DŨNG

THẠC SĨ. NGUYỄN TRẦN HIỂN

BS. HOÀNG THỊ HOA



THƯ VIỆN
HUBT

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

LỜI NÓI ĐẦU

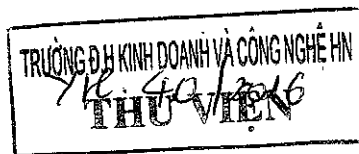
Để phục vụ nhu cầu học tập của sinh viên các cấp học của Trường Đại học Y Hà Nội, chúng tôi biên soạn cuốn "Thực hành Dịch tễ học" nhằm minh họa cho phần lý thuyết về phương pháp Dịch tễ học hiện đại và nâng cao khả năng thực hành thực tế của anh chị em. Nó cũng có thể giúp ích ít nhiều cho cán bộ làm công tác dịch tễ và các cán bộ nghiên cứu y học về phương pháp học chuẩn mực.

Cuốn sách chắc chắn còn chứa đựng những thiếu sót nhất định. Các tác giả chân thành cảm ơn và tiếp thu những chỉ giáo của đồng nghiệp xa gần.

Hà nội, ngày 2 tháng 1 năm 1995

Thay mặt Tập thể tác giả

DƯƠNG ĐÌNH THIÊN





**THƯ VIỆN
HUBT**

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

X TỶ LỆ MẮC

I. MỤC TIÊU

1. Định nghĩa và phân biệt được tỷ lệ mới mắc và tỷ lệ hiện mắc.
2. Nêu mối tương quan và ứng dụng của hai loại tỷ lệ mắc trên đây.
3. Liệt kê những yếu tố có thể gây nên sự khác nhau trong quá trình tính toán tỷ lệ mới mắc và tỷ lệ hiện mắc.

II. NỘI DUNG

Một trong những mục tiêu cơ bản của dịch tễ học nhằm nghiên cứu về sự xuất hiện bệnh hay một trạng thái sinh lý hoặc một hiện tượng sức khỏe nào đó trong xã hội. Việc đo lường sự xuất hiện bệnh tật là nhiệm vụ trung tâm, sống còn của dịch tễ học.

Có thể đo lường sự xuất hiện bệnh theo nhiều cách khác nhau:

1. Số tuyệt đối và số tương đối liên quan tới kích thước quần thể.

Dịch tễ học khác với những môn y học lâm sàng ở hai điểm quan trọng: Thứ nhất, các nhà dịch tễ học quan tâm nghiên cứu một nhóm người chứ không phải từng cá thể riêng biệt. Thứ hai, các nhà dịch tễ học nghiên cứu cả những người được coi là khỏe mạnh lẫn những người ốm đau, họ cố gắng tìm kiếm sự khác biệt cơ bản giữa những người này. Các nhà dịch tễ học cố gắng đo lường sự xuất hiện bệnh tật trong những quần thể khác nhau. Để có thể so sánh sự xuất hiện bệnh ở các quần thể có kích thước khác nhau thì điều cần thiết là tính toán thành tỷ lệ. Một tỷ lệ bao gồm hai thành phần sau:

$$\frac{\text{Tử số}}{\text{Mẫu số}} = \frac{\text{Số sự kiện xảy ra}}{\text{Dân số có nguy cơ}}$$

Bằng cách tính toán một tỷ lệ như vậy, nhằm đo lường tình trạng xuất hiện của bệnh tật, đã tạo lập nên một con số độc lập so với kích thước của quần thể. Tuy nhiên trong một vài trường hợp, tùy theo mục đích, số tuyệt đối về những trường hợp bệnh vẫn có thể có ích. Song hầu hết việc mô tả sự xuất hiện bệnh hay một trạng thái, sinh lý cần phải tính đến độ lớn của quần thể.



THƯ VIỆN
HUBT

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

2. Tỷ lệ mới mắc và tỷ lệ hiện mắc.

Tỷ lệ mới mắc và tỷ lệ hiện mắc là hai hình thái đo lường chủ yếu về sự xuất hiện bệnh trong dịch tễ học.

a) tỷ lệ hiện mắc. (*Prevalence*)

Tỷ lệ hiện mắc thường được ký hiệu là P. Người ta còn gọi tỷ lệ này bằng một thuật ngữ khác là tỷ số hiện mắc. Tỷ lệ hiện mắc được định nghĩa như sau:

$$P = \frac{\text{Tổng số các trường hợp bệnh/thời điểm/quần thể}}{\text{Tổng dân số của quần thể đó trong thời điểm đó.}}$$

Tỷ lệ hiện mắc đo lường số người mắc bệnh trong một quần thể tại thời điểm nhất định. Tuy nhiên tỷ lệ hiện mắc phụ thuộc vào hai yếu tố đó là tỷ lệ mới mắc và thời gian kéo dài trung bình của bệnh. Ở đây nếu tỷ lệ hiện mắc một bệnh bất kỳ nào đó có sự thay đổi thì tỷ lệ này đã phản ánh sự thay đổi của tỷ lệ mới mắc hoặc thời gian kéo dài trung bình của bệnh hoặc cả hai. Tỷ lệ hiện mắc giống như tỷ số dân số mắc bệnh trong một quần thể nhất định tại một thời điểm. Như vậy tỷ lệ này có giá trị đi từ 0 đến 1.

b) tỷ lệ mới mắc tích lũy. (*Cumulative Incidence*)

Tỷ lệ mới mắc tích lũy còn có tên gọi là tỷ số mới mắc, ký hiệu là CI. Tỷ lệ mới mắc tích lũy được định nghĩa như sau:

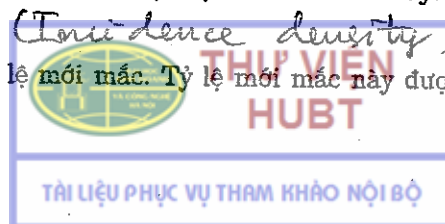
$$CI = \frac{\text{Số cá thể mới mắc bệnh trong thời gian nghiên cứu.}}{\text{Tổng số dân của quần thể nghiên cứu tại thời điểm bắt đầu nghiên cứu.}}$$

Như vậy, tỷ lệ mới mắc tích lũy chính là tỷ số của các cá thể không có bệnh tại thời điểm ban đầu nghiên cứu, nhưng đã phát triển bệnh trong giai đoạn nghiên cứu. Tỷ lệ này đo lường những trường hợp bệnh mới xuất hiện trong một quần thể trong một thời gian nhất định. Hay nói cách khác, tỷ lệ mới mắc tích lũy được xem như nguy cơ trung bình của cá thể trong quần thể có thể mắc bệnh trong một thời gian nhất định. Ở đây, công thức tính tỷ lệ này cho thấy tử số là một phần của mẫu số, tỷ lệ này có giá trị đi từ 0 đến 1.

Trong một số nghiên cứu những nhóm dân cư khác nhau trong quần thể nghiên cứu đã dẫn đến sự khác biệt về nguy cơ mắc bệnh. Sự khác biệt của những nhóm dân cư trong quần thể nghiên cứu ở đây chủ yếu là thời gian theo dõi được đối với từng cá thể. Xuất phát từ thực tế rằng trong nghiên cứu tương lai những cá thể tham gia nghiên cứu khác nhau theo thời gian, có những cá thể tham gia từ lúc bắt đầu cho đến khi kết thúc. Nhưng có khá nhiều cá thể rút ra khỏi nghiên cứu theo các thời điểm khác nhau có thể vì không muốn tham gia nghiên cứu, hoặc đã di cư trong thời gian nghiên cứu. Trong tình trạng này thì tỷ lệ mới mắc tích lũy không thể tính được trực tiếp từ những số liệu như vậy (ở đây, độ dài thời gian quan sát đối với từng cá thể ảnh hưởng trực tiếp đến tỷ lệ mới mắc tích lũy) mà phải dùng tới khái niệm mật độ mới mắc sau đây.

c) Mật độ mới mắc (Ký hiệu là ID).

Mật độ mới mắc cũng là một loại tỷ lệ mới mắc. Tỷ lệ mới mắc này được định nghĩa



như sau:

$$ID = \frac{\text{Tổng số trường hợp bệnh mới mắc xảy ra trong quần thể trong một khoảng thời gian nhất định.}}{\text{Tổng số đơn vị độ dài thời gian có nguy cơ theo dõi được đối với từng cá thể trong quần thể đó}}$$

Mẫu số là tổng số độ dài thời gian có nguy cơ mắc bệnh. Đơn vị này thường được đo lường theo năm và biểu thị là "Năm-Người" hay "Thời gian-Người" hoặc "Thời gian có nguy cơ". Đối với từng cá thể trong quần thể nghiên cứu, thời gian có nguy cơ là thời gian mà trong đó từng cá thể tham gia nghiên cứu trong quần thể còn giữ được tình trạng không mắc bệnh.

Bằng cách chia số trường hợp bệnh mới mắc cho thời gian có nguy cơ thì độ dài thời gian của từng cá thể tham gia nghiên cứu cũng như tổng số độ dài thời gian quan sát đã được tính đến.

Tỷ lệ mới mắc đo lường tỷ lệ phát triển bệnh trong một thời gian nhất định. Tỷ lệ mới mắc phản ánh tốc độ xuất hiện bệnh. Một sự thay đổi về tỷ lệ mới mắc thì có nghĩa là đã có sự thay đổi về cân bằng của các yếu tố căn nguyên. Tỷ lệ mới mắc là một tỷ lệ cơ bản rất quan trọng đối với các nhà nghiên cứu tìm căn nguyên gây bệnh.

Trên thực tế, người ta không thể tính thời gian có nguy cơ cho từng cá thể, mà là ước tính thời gian có nguy cơ tương đối. Chúng ta có thể tính bằng cách cộng dân số đầu lúc bắt đầu thời gian nghiên cứu và dân số cuối thời gian nghiên cứu, sau đó chia cho 2 và nhân với độ dài thời gian nghiên cứu. Như vậy kích thước dân số ở thời điểm giữa của thời gian nghiên cứu có thể được sử dụng theo cách tính tương đối như trên.

Mật độ mới mắc không phải là tỷ số như 2 tỷ lệ trên vì tử số là số trường hợp mới mắc bệnh, còn mẫu số là số đơn vị thời gian-người. Trị số của mật độ mới mắc không bao giờ nhỏ hơn 0.

3. Mối quan hệ giữa các tỷ lệ trên.

Tỷ lệ hiện mắc phụ thuộc vào tỷ lệ mới mắc và thời gian kéo dài trung bình của bệnh. Tỷ lệ hiện mắc là sản phẩm của tỷ lệ mới mắc và thời gian kéo dài của bệnh. Trong những tình trạng ổn định như các bệnh mạn tính chẳng hạn, thì mối quan hệ giữa tỷ lệ mới mắc và tỷ lệ hiện mắc có thể được biểu thị như sau:

$$P = I \times D$$

Trong đó P = Tỷ lệ hiện mắc

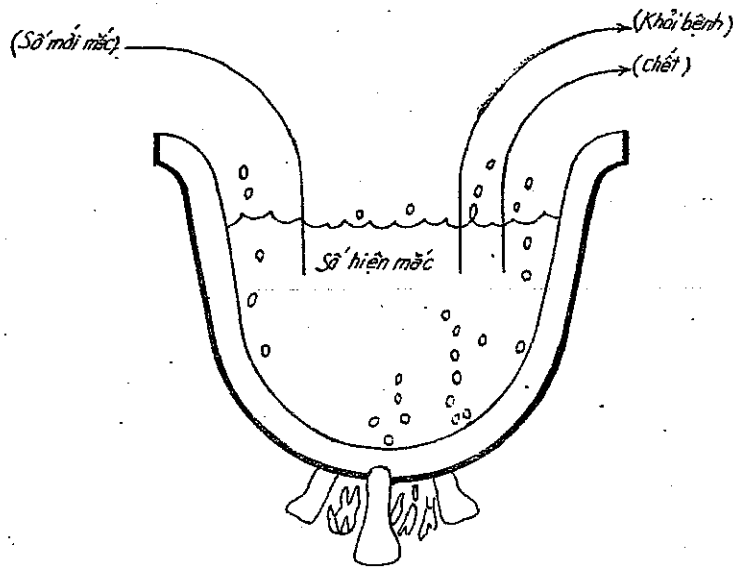
I = Tỷ lệ mới mắc

D = Thời gian kéo dài trung bình của bệnh.

Mối liên quan giữa các tỷ lệ có thể thấy rõ trong hình vẽ sau đây:

Các nhà lập kế hoạch cho những hoạt động y tế thường dựa trên tỷ lệ hiện mắc vì tỷ lệ này đo lường nhu cầu như nhu cầu giường bệnh cũng như những trang thiết bị y tế và nhu cầu về nhân lực.

Tỷ lệ hiện mắc được xác định bằng các cuộc điều tra ngang, trái lại tỷ lệ mới mắc là một loại tỷ lệ rất cơ bản và hữu ích nhưng khó đo lường. Tỷ lệ mới mắc thường được xác



Hình 1 - Mối liên quan giữa số trường hợp mới mắc và hiện mắc.

định trong các nghiên cứu theo dõi về tương lai bằng cách theo dõi một quần thể nhất định mà khởi điểm nghiên cứu từng cá thể tham gia đã được xác định không mắc bệnh mà ta định nghiên cứu, sau đó người ta tiến hành theo dõi từng cá thể trong một thời gian nhất định để xác định những trường hợp bệnh mới xuất hiện và đánh giá tốc độ phát triển bệnh. Tỷ lệ mới mắc được so sánh giữa các nhóm quần thể có phơi nhiễm với yếu tố nguy cơ ở các mức độ khác nhau nhằm đo lường ảnh hưởng của các yếu tố nguy cơ tới sự xuất hiện bệnh. Bằng cách này, người ta có thể ước tính nguy cơ tương đối (xem thêm chương Nghiên cứu tương lai).

III. BÀI TẬP

1. Mỗi một đường kẻ đậm ngang dưới đây biểu thị cho một trường hợp bị bệnh tăng bạch cầu đơn nhân nhiễm khuẩn. Đường nối giữa hai điểm biểu thị thời kỳ bị bệnh của từng trường hợp mang các ký hiệu khác nhau (Giả sử rằng quần thể nghiên cứu cố định là 20 người). Dựa vào hình vẽ sau đây hãy tính các tỷ lệ sau:

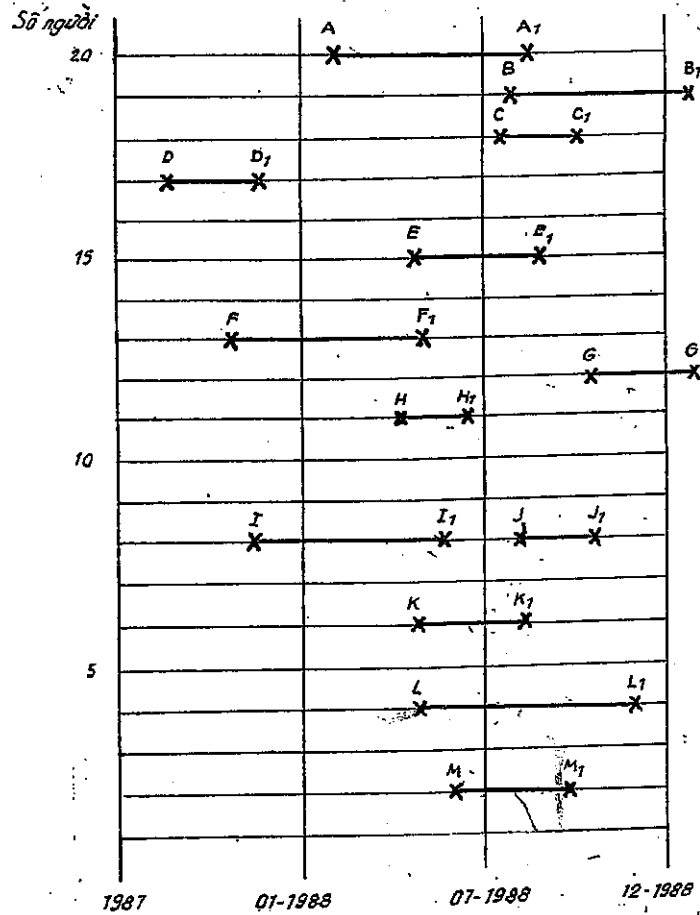
- Tính tỷ lệ mới mắc trong năm 1988.
- Tính tỷ lệ hiện mắc trong năm 1988.
- Giả sử có một cuộc điều tra tiến hành vào tháng 7/1988. Hãy tính tỷ lệ hiện mắc bệnh tại thời điểm này.

2. Từ số liệu ở bảng 1 dưới đây, hãy tính thời gian kéo dài trung bình của từng bệnh mạn tính trong nhiều năm liền.



$$P = I \cdot D$$

$$D = \frac{P}{I}$$



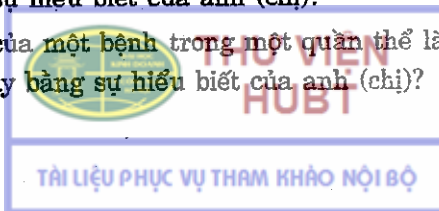
Bảng 1- TỶ LỆ HIỆN MẮC VÀ TỶ LỆ MỚI MẮC CỦA 5 BỆNH MẠN TÍNH TẠI MỘT QUẦN THỂ X. TÍNH TRÊN 100.000 DÂN.

Bệnh	Tỷ lệ mắc tính trên 100.000 dân	
	Tỷ lệ hiện mắc (P)	Tỷ lệ mới mắc (I)
Động kinh	376	30,8
Viêm đa khớp	55	5,0
Parkinson	157	20,0
Bệnh về dây TK vận động	7	1,7
Bệnh khối u TKTU	69	17,3

3. Nghiên cứu mối tương quan giữa chế độ ăn và bệnh tật, ví dụ người ta biểu thị tình hình mắc bệnh bằng con số tuyệt đối (như tổng số trường hợp mắc bệnh) hoặc biểu thị bằng con số tương đối (con số có liên quan tới độ lớn của quần thể nghiên cứu). Cách biểu thị nào là thích hợp, hãy giải thích.

4. a) Tỷ lệ hiện mắc của một bệnh trong một quần thể là 0,02. Hãy giải thích ý nghĩa của con số này bằng sự hiểu biết của anh (chị)?

b) Tỷ lệ mới mắc của một bệnh trong một quần thể là 5 trên 10000 năm/người. Nếu ý nghĩa của con số này bằng sự hiểu biết của anh (chị)?



5. Giả sử một quần thể có phân bố tuổi ổn định, có thể có cách lý giải như thế nào khi tỷ lệ hiện mắc của bệnh gia tăng, mặc dù tỷ lệ mới mắc vẫn giữ mức độ hằng định.

6. Trong thời gian 4 năm tại một phòng thí nghiệm y học lớn đã có 532 người bị tai nạn lao động. Con số người làm việc tại đây lúc ban đầu của thời kỳ này là 520 người và tính đến thời điểm cuối cùng là 680 người. Người ta có thể tính được tỷ lệ nào trong nghiên cứu này, hãy tính cụ thể.

7. Để có thể phát hiện và điều trị sớm ung thư cổ tử cung, người ta tiến hành khám sản khoa hàng loạt và thường xuyên cho phụ nữ trong độ tuổi từ 30 đến 59. Sau đó tiếp tục theo dõi và phát hiện những phụ nữ mới bị bệnh này trong một khoảng thời gian nghiên cứu nhất định. Nếu tính từ thời điểm bắt đầu cho đến khi kết thúc người ta tính được khoảng thời gian nguy cơ là 338294 năm-người và xác định được 123 ca bệnh mới xuất hiện trong suốt thời gian nghiên cứu. Có thể tính được tỷ lệ nào trong nghiên cứu này, hãy tính cụ thể.

8. Trong một nghiên cứu sàng lọc tiến hành trên 5000 phụ nữ. Người ta đã tìm thấy có 25 người mắc bệnh ung thư vú. Năm năm sau đó, người ta đã phát hiện thêm 10 trường hợp bị bệnh. Có thể tính được những tỷ lệ nào trong nghiên cứu này, hãy tính cụ thể?

9. Có 270 trường hợp bị bệnh loét tá tràng ở nam giới trong thời gian 5 năm nghiên cứu theo dõi tại một thành phố. Tổng số những người dân ông sinh sống trong thành phố này là 18500 người, thời điểm lúc bắt đầu nghiên cứu và 21500 người vào thời điểm cuối của cuộc điều tra. Người ta có thể tính được tỷ lệ nào, hãy tính cụ thể?

10. Theo trung tâm nghiên cứu ung thư của Thụy Điển, tính từ năm 1971, 1972 và 1973 có số trường hợp bị ung thư tụy lần lượt từng năm là 97, 121 và 112 trường hợp nam giới bị bệnh này trong đó tuổi từ 70-74. Tại thời điểm đầu năm 1971 có 309949 người dân ông trong nhóm tuổi này và đến cuối năm 1973 có 332400 dân ông cùng nhóm tuổi. Người ta có thể tính được những tỷ lệ nào dựa trên các dữ kiện trên. Hãy tính cụ thể.

11. Trong một nghiên cứu sàng lọc trên 1329 nam giới tuổi từ 40 đến 59 tuổi, người ta đã tiến hành khám kiểm tra mức độ cholesterol huyết thanh và huyết áp tâm trương cho những đối tượng này. Sau đó tiến hành theo dõi những đối tượng trên trong vòng 6 năm nhằm phát hiện những trường hợp bị bệnh nhồi máu cơ tim. Biết rằng tại thời điểm bắt đầu nghiên cứu tất cả các đối tượng đều không bị bệnh này. Kết quả nghiên cứu được trình bày trong bảng dưới đây

Bảng 2- MỨC ĐỘ HUYẾT ÁP TÂM TRƯƠNG (mmHg).

Mức cholesterol huyết thanh (mg/100ml)	<147		147-166		>167	
	Tổng số	T.hợp bệnh	Tổng số	T.hợp bệnh	Tổng số	T.hợp bệnh
<220	431	10	93	3	49	7
220-259	347	19	74	6	49	6
>260	185	19	57	11	44	11
	963	48	224	20	142	24



PAI.DV
Z

Dựa vào bảng số liệu trên người ta có thể tính được loại tỷ lệ nào, hãy tính các tỷ lệ đối với từng trường hợp cụ thể sau:

a) Nhóm dân ông có lượng cholesterol huyết thanh dưới 220 mg/100ml và huyết áp tâm trương dưới 147 mmHg.

b) Nhóm dân ông có lượng cholesterol huyết tương > 260 mg/100ml và huyết áp tâm trương dưới 147mmHg

c) Nhóm dân ông có mức cholesterol huyết tương < 220 mg/100ml và huyết áp tâm trương > 167mmHg.

d) Nhóm dân ông có mức cholesterol huyết tương > 260 mg/100ml và huyết áp tâm trương > 167mmHg.

TRẢ LỜI BÀI TẬP

1. a) 10/20. Hay 50% (50%)

b) $11/20 = 50\%$ (10 trường hợp mới mắc + 1 trường hợp mắc từ trước).

c) $5/20 = 25\%$

2. Thời gian kéo dài trung bình của bệnh ký hiệu là D.

D = Bệnh động kinh = 12,2

D = Viêm đa khớp = 1,1

D = Parkinson = 7,8

D = Bệnh về dây thần kinh vận động = 4,1

D = Bệnh khối u hệ thống thần kinh trung ương = 4,0

3. Con số tương đối. Khi so sánh tình trạng bệnh tật, người ta cần phải đo lường làm sao cho các chỉ số biểu thị tình trạng bệnh độc lập với độ lớn của quần thể.

4. a) 2% dân số bị bệnh tại một thời gian cụ thể nào đó.

b) 5 trường hợp bị bệnh xảy ra trên 10000 năm-người có nguy cơ mắc bệnh.

5. thời gian kéo dài trung bình của bệnh giảm.

6. Số trường hợp mắc bệnh = 532. ✓

Dân số trung bình = $\frac{520+680}{2} = 600$ ✓

Thời gian quan sát là 4 năm. ✓

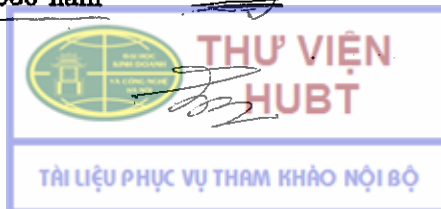
Tổng số thời gian nguy cơ là = $600 \times 4 = 2400$ năm-người. ✓

Tỷ lệ mới mắc (ID) = $\frac{532}{2400} = 0,22$ năm ✓

7. Tỷ lệ mới mắc bệnh (ID) *Mật độ mới mắc.*

ID = $\frac{123}{338294} = 0,00036$ năm

8. Tỷ lệ hiện mắc P



$$P = \frac{25}{5000} = 0,005$$

Tỷ lệ mới mắc tích lũy CI

$$CI = \frac{10}{1000-25} = 0,002 \text{ trong 5 năm nghiên cứu.}$$

9. Số trường hợp mắc bệnh là 270.

$$\text{Dân số trung bình} = \frac{18500+21500}{2} = 20000$$

Thời gian quan sát là 5 năm.

Tổng số thời gian nguy cơ = $20000 \times 5 = 100000$ năm-người.

$$\text{Tỷ lệ mới mắc I} = \frac{270}{100.000} = 0,0027 \text{ năm}$$

TỶ LỆ TỬ VONG

I. MỤC TIÊU

- Viết ra được định nghĩa của các tỷ lệ sau:
 - Tỷ lệ tử vong thô.
 - Tỷ lệ tử vong riêng phần (theo tuổi, giới, dân tộc và theo nguyên nhân).
 - Tỷ lệ chết/mắc (Case - Fatality Rates).
- Phân tích, sử dụng đúng được các tỷ lệ trên.
- Nêu những lý do cần phải chuẩn hóa tỷ lệ.

II. NỘI DUNG

- Tỷ lệ tử vong thô (CDR). (*Crude Death Rate*)

$$\text{CDR} = \frac{\text{Tổng số trường hợp chết vì mọi nguyên nhân/quần thể/thời gian}}{\text{Dân số trung bình/quần thể/thời gian đó}}$$

Nếu tính tỷ lệ tử vong thô trong vòng 1 năm thì tử số bao gồm tổng số tử vong trong năm, còn mẫu số chính là dân số giữa năm đó (30-6 hàng năm).



THƯ VIỆN
HUBT

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

Tỷ lệ tử vong thô rất dễ tính, nó biểu thị tỷ lệ tử vong quan sát được trong quần thể nghiên cứu và nó là điểm tính toán khởi đầu cho việc tính toán các tỷ lệ chuẩn hóa.

Tỷ lệ tử vong thô đo lường tỷ số tử vong của một quần thể hàng năm. Tỷ lệ tử vong này bị ảnh hưởng bởi 2 yếu tố sau:

- Thứ nhất là xác suất chết (yếu tố này được đo lường một cách chính xác bằng tỷ lệ chết riêng phần theo tuổi, sẽ giải thích ở phần sau).

- Thứ hai là đặc trưng theo tuổi hay phân bố theo tuổi của quần thể đang xem xét tỷ lệ tử vong thô.

Do ảnh hưởng của 2 yếu tố trên, người ta thường xem xét tỷ lệ tử vong riêng phần theo tuổi và cấu trúc dân số của quần thể trước khi tiến hành so sánh tỷ lệ tử vong thô giữa các nhóm quần thể với nhau.

2. Tử vong và tuổi.

Tuổi là một trong những đặc trưng quan trọng nhất ảnh hưởng tới phân bố bệnh. Trước khi có thể tiến hành so sánh phân bố tình hình bệnh tật giữa 2 quần thể, người ta cần phải tính đến sự khác biệt về cấu trúc tuổi giữa 2 quần thể này.

Ví dụ: Tại Maine năm 1973 tỷ lệ tử vong thô là 10,6 trên 1000 dân và tại Alaska cùng năm tỷ lệ này là 4,6 trên 1000 dân. Dựa trên 2 tỷ lệ tử vong thô ở 2 khu vực trên cho thấy tỷ lệ tử vong ở Maine cao hơn hẳn Alaska, nhưng xem xét sự phân bố tuổi tại 2 quần thể trên cho thấy tỷ lệ tử vong thô ở Maine cao do quần thể ở đây có cấu trúc dân số già và tại Alaska có cấu trúc dân số trẻ một cách điển hình.

3. Tỷ lệ tử vong riêng phần (hoặc tỷ lệ tử vong đặc hiệu) theo tuổi.

Yếu tố tuổi ảnh hưởng sâu sắc lên tỷ lệ tử vong thô do vậy điều cần thiết là tính tỷ lệ tử vong trong từng nhóm tuổi để so sánh. Bảng 1 sau đây trình bày tỷ lệ tử vong thô là tỷ lệ tử vong riêng phần theo tuổi tại thành phố Baltimore cho thấy: Người da trắng tại Baltimore có tỷ lệ tử vong thô cao hơn (15,2) người da màu (9,8). Nhưng nếu xem xét tỷ lệ tử vong trong từng nhóm tuổi ở cả 2 nhóm người lại cho thấy hình ảnh ngược lại là người da màu có tỷ lệ tử vong riêng phần trong từng nhóm tuổi đều cao hơn người da trắng. Vậy dường như có sự mâu thuẫn gì ở đây giữa tỷ lệ tử vong thô và tỷ lệ tử vong riêng phần theo tuổi? Câu trả lời ở đây là sự phân bố tuổi khác nhau tại thành phố Baltimore khác biệt giữa người da trắng và người da màu. Những người da trắng đa phần là nhiều tuổi và những người da màu chủ yếu là ở lứa tuổi trẻ.

Bảng 1- TỶ LỆ TỬ VONG TÍNH TRÊN 1000 DÂN THEO TUỔI VÀ CHỦNG TỘC TẠI THÀNH PHỐ BALTIMORE.

	Tất cả các nhóm tuổi	1 tuổi	1-4	5-17	18-44	45-65	>65
Da trắng	15,2	13,5	0,6	0,4	1,5	10,7	19,7
Da đen	9,8	22,6	1,0	0,5	3,6	18,8	61,1

4. Tỷ lệ tử vong chuẩn hóa theo tuổi.

Khi một quần thể có sự khác biệt về phân bố tuổi, nếu chúng ta muốn so sánh thì

cần phải tính tỷ lệ tử vong chuẩn hóa theo tuổi.

Để có thể hiểu tỷ lệ tử vong chuẩn hóa theo tuổi là gì và cách tính như thế nào? Trước hết cần phải biết rằng một tỷ lệ tử vong thô chính là tổng quát của các phức hợp sau đây:

Tỷ số dân trong nhóm tuổi x tỷ lệ tử vong riêng phần của nhóm tuổi đó.

Tỷ lệ tử vong thô được chuẩn hóa bằng cách đưa sự khác biệt về dân số trong từng nhóm tuổi tương ứng với tỷ lệ dân số trong các nhóm tuổi trong quần thể chuẩn. Quần thể chuẩn thường là quần thể được điều tra trong các cuộc điều tra dân số.

Trình tự cách chuẩn hóa sẽ được minh họa ở ví dụ sau đây:

Bảng 2- TỶ LỆ TỬ VONG Ở 2 CỘNG ĐỒNG

Cộng đồng	Tỷ lệ tử vong riêng phần theo tuổi/1000	
	Trẻ	Già
A	4 ^a	16 ^c
B	5 ^b	20 ^d

$$A. \frac{1}{2} + 16 \cdot \frac{1}{2} = 10$$

$$B. \frac{2}{3} + 20 \cdot \frac{1}{3} = 10$$

Cộng đồng A có dân số bao gồm một ^{nửa} người trẻ và một nửa người già. Cộng đồng B có 2/3 người trẻ và 1/3 người già sống trong cộng đồng. Tỷ lệ tử vong riêng phần trong từng cộng đồng được đưa ra ở bảng 2.

Từ các thông tin trên, chúng ta có thể tính tỷ lệ tử vong thô cho 2 cộng đồng như sau:

- Tỷ lệ tử vong thô ở cộng đồng A là:

$$(1/2)(4) + (1/2)(16) = 10 \text{ trên } 1000.$$

- Tỷ lệ tử vong thô ở cộng đồng B là:

$$(2/3)(5) + (1/3)(20) = 10 \text{ trên } 1000.$$

Lưu ý rằng mặc dầu xác suất chết ở cộng đồng A là thấp hơn ở các nhóm tuổi, nhưng tỷ lệ tử vong thô lại tương tự như cộng đồng B vì rằng cộng đồng A có dân số già hơn cộng đồng B.

Bây giờ chúng ta hãy tính tỷ lệ chết chuẩn hóa cho từng cộng đồng sử dụng quần thể chuẩn. Quần thể chuẩn được chọn ở đây bao gồm 1/3 là người trẻ và 2/3 người già.

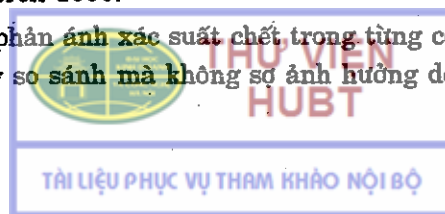
- Tỷ lệ tử vong chuẩn hóa ở cộng đồng A là:

$$(1/3)(4) + (2/3)(16) = 12 \text{ trên } 1000.$$

- Tỷ lệ tử vong chuẩn hóa ở cộng đồng B là:

$$(1/3)(5) + (2/3)(20) = 15 \text{ trên } 1000.$$

Tỷ lệ tử vong chuẩn hóa theo tuổi phản ánh xác suất chết trong từng cộng đồng và ở đây chúng ta có thể sử dụng tỷ lệ này so sánh mà không sợ ảnh hưởng do sự khác biệt



về phân bố tuổi giữa 2 cộng đồng này.

5. Tỷ lệ tử vong đặc hiệu theo giới.

Nói chung đa số các nước trên thế giới tỷ lệ tử vong ở nam giới thường cao hơn so với nữ giới. Ví dụ ở Mỹ năm 1973 tỷ lệ chết chuẩn hóa theo tuổi là 9,1 trên 1000 nam giới, trong khi đó tỷ lệ này ở nữ giới chỉ chiếm khoảng một nửa: 5,1. Đã có khá nhiều cách giải thích khác nhau cho tình trạng này như nam giới thường mắc các bệnh mà có nguy cơ tử vong cao như bệnh tim mạch, ung thư, tai nạn hoặc các bệnh về đường hô hấp.

6. Tỷ lệ chết đặc hiệu theo chủng tộc.

Ví dụ vừa nêu trên đã đề cập cho thấy người da màu có tỷ lệ tử vong cao hơn so với người da trắng. Việc xem xét tỷ lệ tử vong đặc hiệu theo các chủng tộc khác nhau và nguyên nhân tử vong là rất cần thiết.

7. Tỷ lệ tử vong xét theo nguyên nhân.

Tỷ lệ tử vong do một nguyên nhân hay một tình trạng bệnh lý nào đó cũng có thể tính cho toàn bộ quần thể hoặc cho bất cứ một nhóm tuổi hay chủng tộc nào hoặc các nhóm tuổi xét theo sự khác biệt về giới.

Tỷ lệ tử vong theo nguyên nhân được tính như sau:

$$\frac{\text{Tổng số chết do một bệnh trong quần thể nhất định trong một khoảng thời gian nào đó}}{\text{Tổng số dân số trong quần thể đó ở thời gian giữa khoảng thời gian đó}} = x \text{ 100000}$$

Tỷ lệ này thường được tính trên 100000 dân trong một năm.

8. Tỷ lệ chết/mắc.

Tỷ lệ chết/mắc được tính bằng:

$$\frac{\text{Tổng số trường hợp chết vì một bệnh nhất định trong một thời gian nhất định}}{\text{Tổng số trường hợp mắc bệnh đó trong thời gian đó}} = x \text{ 100}$$

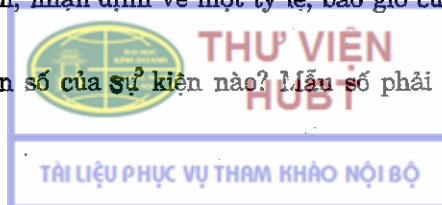
Tỷ lệ này biểu thị xác suất chết của một bệnh trong số những trường hợp đã được chẩn đoán là có mắc bệnh, hay nó biểu thị cho mức độ nguy hiểm của bệnh. Tỷ lệ chết/mắc được sử dụng điển hình trong các bệnh cấp tính, tỷ lệ này ít có ý nghĩa đối với bệnh mạn tính bởi vì trong trường hợp bệnh mạn tính thời kỳ ủ bệnh cũng như thời kỳ mắc bệnh kéo dài và có nhiều thay đổi.

Tỷ lệ chết/mắc của cùng một bệnh thường cũng rất khác nhau ở các khu vực khác nhau.

9. Phân tích, nhận định và sử dụng các tỷ lệ.

Trước khi phân tích, nhận định về một tỷ lệ, bao giờ cũng cần làm rõ những điểm sau đây:

- Tử số phải là con số của sự kiện nào? Mẫu số phải là con số cá thể của quần thể



nào? Liệu những con số đó có được thu thập chính xác không?

- Khoảng thời gian diễn biến của sự kiện (khoảng thời gian thu thập số liệu) nêu là khoảng thời gian nào? Có hợp lý không?

- Tỷ lệ đó liệu có đem so sánh được không?

- Tỷ lệ đó sẽ được sử dụng cho mục tiêu nào? Có phục vụ được cho mục tiêu đó không?

Nếu không phân tích nhận định kỹ lưỡng trước khi sử dụng, có thể sẽ mắc phải những sai lầm phổ biến sau đây:

+ Sai lầm do sử dụng mẫu số không đúng:

Đã có không ít các cuộc điều tra quan tâm đến nguy cơ mắc một bệnh nào đó theo các nhóm tuổi khác nhau, và làm một bảng tổng hợp thí dụ như sau:

Nhóm tuổi	Số mắc	Tỷ lệ phần trăm
Dưới 1 tuổi	8	11.8
1 - 2 tuổi	34	50.0
3 - 4 tuổi	16	23.5
5 - 9 tuổi	7	10.3
10 - 14 tuổi	3	4.4
Cộng	68	100.0

Và từ bảng tổng hợp này, nếu đi đến nhận định rằng nhóm 1-2 tuổi có nguy cơ mắc cao nhất là hoàn toàn không chính xác, vì tỷ lệ phần trăm ở đây chẳng qua chỉ biểu thị tần số mắc ở các nhóm tuổi khác nhau, do đó nó không nói gì được nguy cơ mắc của bất kỳ nhóm tuổi nào. (Trường hợp này phải tính các tỷ lệ mắc riêng phần của từng nhóm tuổi, bằng cách chia số mắc của từng nhóm tuổi cho tổng số cá thể của nhóm tuổi đó trong quần thể, mới so sánh được, và mới nói lên nguy cơ mắc của nhóm tuổi đó).

+ Sai lầm do sử dụng tử số không đúng:

Tính tử số không đúng dễ làm quá trội hoặc quá non nguy cơ mắc đối với một bệnh. Nếu dễ dàng chấp nhận tử số một cách hời hợt, không phân tích xem tử số có thật đúng là con số nói lên sự kiện cần điều tra nghiên cứu không, thì dễ dẫn đến kết luận sai lầm như trong thí dụ sau đây:

Thí dụ người ta đã tính tỷ lệ chết ung thư ở Philipin trong một năm, riêng ở vùng nông thôn thấy $14,2 \cdot 10^{-5}$, nếu tính cả ở thành thị thì con số này lên tới $40,1 \cdot 10^{-5}$. Có nên cho rằng tỷ lệ chết ung thư ở nông thôn là thấp không? Trước khi nhận định về vấn đề này, người ta phải phân tích đến những yếu tố có thể làm sai lầm số chết thu thập được, như: ở thành phố có điều kiện chẩn đoán tốt hơn, dân thành phố quan tâm hiểu biết về bệnh tốt hơn, đi khám chữa nhiều hơn, dễ dàng hơn; số chết lại không chia theo nơi ở cũ mà chỉ ghi nhận vào số chết ở thành phố. Thêm vào đó họ lại còn được kể là có hộ khẩu chính thức sau 6 tháng tạm trú.

+ Sai lầm về thời gian đã thu thập số liệu:

Muốn so sánh được, các tỷ lệ phải được nêu trong cùng những thời gian có hoàn cảnh kinh tế xã hội, có tiến bộ khoa học kỹ thuật... như nhau. Đặc biệt lưu ý đối với các bệnh

10 - 14	47,37	38,09	+8,28
15 - 19	71,43	60,87	+10,56
20 - 29	83,33	76,92	+6,33
30 - 39	60,00	44,44	+15,56
40 trở lên	69,23	41,18	+28,05

Ta thấy rõ ràng ở tất cả các nhóm tuổi có tỷ lệ mắc ở nam đều cao hơn ở nữ (trái hẳn với nhận định trên dùng tỷ lệ mắc chung nữ cao hơn nam là không chính xác). Điều này xảy ra, chính là do chọn mẫu sai, trong đó đáng kể là sự khác biệt về cấu trúc tuổi của 2 quần thể nam và nữ làm cho tỷ lệ mắc chung không thể so sánh được (mà đem so sánh sẽ mắc sai lầm). Trường hợp này phải đem chuẩn hóa theo tuổi trước khi so sánh, đơn giản hơn cả là so sánh bằng tỷ lệ mắc riêng phần theo tuổi như làm trên đây (còn nhiều kỹ thuật chuẩn hóa khác, xin xem sách đã dẫn).

Chúng ta cũng cần phân tích kỹ lưỡng khi sử dụng một số chỉ số y tế thường dùng sau đây:

+ Về dân số học của quần thể

- Dân số toàn bộ: thường được ước tính cho một vài năm sau dựa trên tỷ lệ tăng trưởng dân số.

- Tỷ lệ tăng trưởng = (tỷ lệ sinh + nhập cư) - (tỷ lệ chết - di dân)

- Dân số phân bố theo địa dư, hành chính

- Dân số phân bố theo tuổi và giới.

Có nhiều cách ước tính dân số. Sau đây là phương pháp đơn giản nhất, với giả định rằng dân số tăng trưởng đều hàng năm. Giả sử ta có: Ở cuộc điều tra dân số ngày 01.01.1979 có 800.000 dân vào 01.01.1989 có 1.000.000 dân

Số dân tăng trung bình mỗi năm sẽ là: $\frac{1.000.000 - 800.000}{10} = 20.000$ dân

Tính theo phương pháp gia tăng số học:

Dân số vào 01.01.1994 = 1.000.000 + (20.000 x 5) = 1.100.000 dân

Tính theo phương pháp gia tăng hình học (chi tiết hơn):

Trước hết tính tỷ lệ gia tăng hàng năm: $\frac{1.000.000 - 800.000}{800.000} = 2,5\%$

Bây giờ ta quy ước dân số vào 01.01.1989 là 1, thì trong 5 năm tới, với tỷ lệ gia tăng hàng năm là 2,5%, dựa vào bảng tính sẵn của ngân hàng (lãi mẹ đẻ lãi con) con số 1 này sẽ tương ứng với 1,13140. Do đó dân số vào 01.01.1994 sẽ là: 1.000.000 x 1,13140 = 1.131.400 dân

+ Về các tỷ lệ chết: Thường dùng nhất là:

- Tỷ lệ chết thô

- Tỷ lệ chết đối với từng bệnh, từng nguyên nhân, nhất định

- Tỷ lệ chết riêng phần theo giới

- Tỷ lệ chết riêng phần theo tuổi

Nhóm tuổi	Nam	Nữ	Nam cao hơn nữ
	0,00	0,00	0,00
Dưới 5 tuổi	36,36	31,58	+4,78
Tỷ lệ mắc riêng phần theo tuổi (%)			

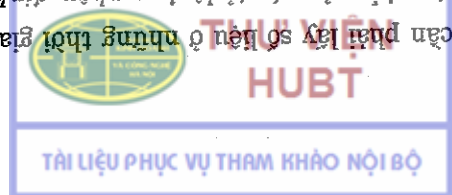
Từ bảng chưa chuẩn hóa theo tuổi này, ta thấy tỷ lệ mắc ở nữ cao hơn ở nam. Tỷ lệ mắc riêng phần theo tuổi của từng giới:

Nhóm tuổi	Nam		Nữ	
	Số khám	Số mắc	Số khám	Số mắc
Dưới 5 tuổi	35	0	11	0
5 - 9	22	8	19	6
10 - 14	19	9	21	8
15 - 19	14	10	23	14
20 - 29	12	10	26	20
30 - 39	10	6	18	8
40 trở lên	13	9	17	7
Cộng	125	52	135	63
Tỷ lệ mắc chung	41,6%		46,7%	

Ta hãy xem thì dù sau đây về việc so sánh nguy cơ mắc giữa nam và nữ:
 + Sai lầm khi đem so sánh những tỷ lệ chưa chuẩn hóa:
 Đây là sai lầm đã gặp phổ biến. Nhiều nghiên cứu đã đem nhân định so sánh ngay những tỷ lệ chưa chuẩn hóa nên đã dẫn đến những kết luận sai lầm, đôi khi rất nghiêm trọng, vì làm đảo ngược cả thực tế. Vì nhất, chúng phải làm chuẩn hóa theo tuổi trước khi đem so sánh những tỷ lệ mắc thông thường.

+ Sai lầm do chọn mẫu không đúng:
 Một thí dụ kinh điển cho ta thấy vấn đề này: một nhà khoa đã nêu lên sự khác biệt về thời gian đau để giải đoạn 2 ở 2 nhóm sản phụ có và không được dùng thuốc giảm đau khi đẻ. Những người không dùng thuốc được tuyên truyền để không đau, và tiếp tục được cũng có cả khi đau đẻ. Sau đó tuyển chọn trong số này, những người tình nguyện sẽ không dùng thuốc làm nhóm đối chứng. Sai lầm chính của nghiên cứu này là chỉ những sản phụ có kinh nghiệm để không đau mới tình nguyện làm nhóm chứng, ngược lại những sản phụ lần trước đẻ có đau hoặc những sản phụ đẻ con số đầu ở nhóm chủ yếu.

+ Sai lầm về ý nghĩa của các tỷ lệ đem nhân định so sánh:
 Thí dụ muốn đánh giá hiệu quả tiêm chủng bạch hầu thì không thể căn cứ vào tỷ lệ chết/mắc (tỷ lệ này để nhân định hiệu quả điều trị, thí dụ điều trị bằng SAD chẳng hạn) mà phải căn cứ vào tỷ lệ mới mắc bạch hầu trước và sau khi dùng vacxin.



theo mùa rõ rệt, cần phải lấy số liệu ở những thời gian tương ứng.

- Tỷ lệ chết trên mắc

Các tỷ lệ chết thô, tỷ lệ chết đối với từng bệnh, tỷ lệ chết trên mắc khi muốn đem so sánh đều phải chuẩn hóa trước đã.

Còn các tỷ lệ chết riêng phần theo tuổi và theo giới thì có thể đem so sánh ngay các nhóm tuổi (không kể giới) hoặc các giới khác nhau (mà không kể tuổi) ở các quần thể khác nhau được.

+ Về các tỷ lệ mắc

Có nhiều (xem bài tỷ lệ mắc) nhưng ngay đối với một bệnh, một nguyên nhân nhất định, người ta có thể tính ra nhiều thể loại tỷ lệ riêng phần khác nhau, tùy theo mục đích sử dụng. Thí dụ người ta có thể tính tỷ lệ mắc riêng phần theo:

- Nguyên nhân - tuổi
- Nguyên nhân - giới
- Nguyên nhân - tuổi - giới
- Nguyên nhân - dự phòng (hoặc điều trị)
- Nguyên nhân - tuổi - giới - dự phòng (hoặc điều trị).

Tuy nhiên, hàng năm, bao giờ cũng cần làm một bảng tổng kết về những nguyên nhân mắc hàng đầu, quy thành phần trăm để nhìn tổng thể về tình hình ở địa phương. Trong số đó, cần làm những nguyên nhân hàng đầu đối với: tỷ lệ mắc chung, tỷ lệ mắc trẻ con, và tỷ lệ mắc ở các bà mẹ.

+ Về các chỉ số kinh tế - xã hội

Trong số các chỉ dẫn chung về kinh tế xã hội, thì chỉ dẫn phân loại các hộ giàu nghèo rất khó xác định. Theo một số tác giả, có thể căn cứ vào thu nhập và chi phí cho ăn uống:

Chi phí đến 80% thu thập cho ăn uống: gia đình nghèo

Chi phí đến 50% thu nhập cho ăn uống: không nghèo không giàu

Chi phí dưới 20% thu nhập cho ăn uống: gia đình giàu

+ Về các dịch vụ y tế

Ở khu vực bệnh viện, cần tính toán đến các chỉ số liên quan giữa dân số - số giường bệnh (kể cả khu vực tư nhân), các chỉ số về các hoạt động của bệnh viện, các tỷ lệ chết và các tỷ lệ mắc ở các bệnh viện.

Đối với các dịch vụ y tế công cộng khác, cần xây dựng các chỉ số về các cộng đồng có trạm y tế, về các trạm y tế có đủ đội ngũ cán bộ quy định, có triển khai (như khám thai, đỡ đẻ, kế hoạch hóa gia đình, tiêm chủng mở rộng, khám chữa bệnh, cung ứng thuốc...).

Dù là chỉ số nào, khi muốn đem so sánh với nhau đều cần xem xét kỹ để tránh những sai lầm có thể mắc phải đối với một tỷ lệ, như đã nêu ở trên.

III. BÀI TẬP

1. Tại một quần thể trong năm 1982 người ta đã thống kê được có 1886 trường hợp chết trong năm đó. Để tính được tỷ lệ tử vong thô, người ta cần phải thu thập thêm các

Thông tin gì?

2. Bảng 3 sau đây trình bày tỷ lệ tử vong thô và tỷ lệ tử vong chuẩn hóa theo tuổi tại thành phố New York và toàn bộ nước Mỹ trong vòng 40 năm. Dựa vào bảng số liệu này có thể giải thích được điều gì ở đây?

Bảng 3- TỶ LỆ TỬ VONG THÔ VÀ TỶ LỆ TỬ VONG CHUẨN HÓA THEO TUỔI DO MỌI NGUYÊN NHÂN TÍNH TRÊN 1000 DÂN

Năm	Thành phố New York		Nước Mỹ	
	Tỷ lệ chuẩn hóa theo tuổi	Tỷ lệ tử vong thô	Tỷ lệ chuẩn hóa theo tuổi	Tỷ lệ tử vong thô
1940	11,3	10,2	10,8	10,8
1950	8,9	10,0	8,4	9,6
1960	8,1	11,1	7,6	9,5
1970	7,7	11,2	7,1	9,5
1980	6,6	10,8	5,9	9,9

3. Năm 1970 tỷ lệ tử vong thô tại Guyana (một nước đang phát triển tại Nam Phi) là 6,8 trên 1000 dân và tỷ lệ này tại Mỹ là 9,8 trên 1000 dân.

Tỷ lệ tử vong thô tại Guyana thấp hơn so với Mỹ có thể được giải thích bằng sự thật là Mỹ có tổng số dân lớn hơn không? Hãy đưa ra các cách giải thích của anh hay chị?

4. Tỷ lệ tử vong thô và tỷ lệ tử vong chuẩn hóa theo tuổi (tính trên 100000 dân) do bị bệnh tim và bệnh xơ cứng động mạch tại Chile và Mỹ năm 1967 trình bày ở bảng sau đây:

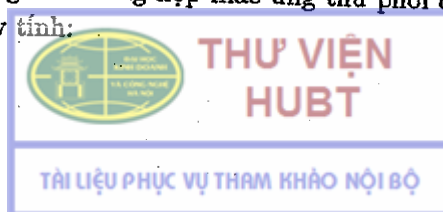
Bảng 4- TỶ LỆ TỬ VONG THÔ VÀ TỶ LỆ TỬ VONG CHUẨN HÓA THEO TUỔI DO BỊ BỆNH TIM MẠCH.

Nước	Tỷ lệ tử vong thô	Tỷ lệ tử vong chuẩn hóa
Chile	67,4	58,2
Mỹ	316,3	131,4
Tỷ suất: Mỹ/Chile	4,7	2,3

a) Để so sánh tỷ lệ tử vong do bệnh tim mạch ở 2 nước thì sử dụng tỷ lệ nào thích hợp hơn? Tại sao?

b) Tại sao tỷ suất của tỷ lệ tử vong thô và tỷ lệ tử vong chuẩn hóa giữa 2 nước lại khác nhau?

5. Một thành phố bao gồm có 100000 dân (trong đó có 45000 nam và 55000 nữ) và có 1000 người chết trong năm. Có 50 trường hợp bị mắc bệnh ung thư phổi (trong đó 40 là nam, 10 nữ) trong năm đó. Trong tổng số trường hợp mắc ung thư phổi cùng năm có 45 trường hợp chết (36 nam, 9 nữ). Hãy tính:



- a) Tỷ lệ tử vong thô.
- b) Tỷ lệ tử vong đặc hiệu theo tuổi. (*giới*)
- c) Tỷ lệ tử vong đặc hiệu do bệnh ung thư phổi.
- d) Tỷ lệ chết/mắc do bệnh ung thư phổi.

Hãy chọn một câu trả lời đúng nhất cho các bài tập từ bài 6 đến bài 7 sau đây.

6. Tỷ lệ chết của trẻ em dưới 1 tuổi được đo bằng số trẻ chết:

- a) Từ 24 giờ đến 1 năm tuổi trên 10000 trẻ đẻ sống.
- b) Dưới 6 tháng tuổi trên 1000 trẻ đẻ sống.
- c) Dưới 1 năm tuổi trên 1000 trẻ đẻ sống.
- d) Dưới 1 năm tuổi trên 1000 cuộc đẻ.
- e) Dưới 1 năm tuổi trên 10000 cuộc đẻ.

7. Trong một cộng đồng bao gồm 100000 người, có 1000 trường hợp mắc 1 bệnh, trong đó 200 trường hợp chết vì bệnh đó trong 1 năm. Tỷ lệ chết vì bệnh này là:

- a) 0.2%.
- b) 1%.
- c) 2%.
- d) 10%.
- e) 20%.

I LOVE YOU

I MISS YOU

TRẢ LỜI BÀI TẬP

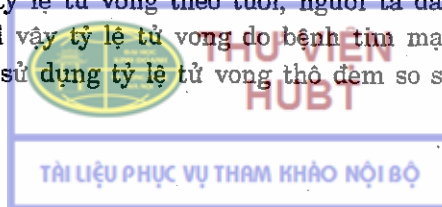
1. Để tính được tỷ lệ tử vong thô của quần thể này trong năm 1982 người ta cần phải biết được dân số giữa 1982 của quần thể đó.

2. Tỷ lệ tử vong thô không phải là tỷ lệ đo lường xu thế tử vong tốt nhất bởi vì có nhiều yếu tố ảnh hưởng tới tỷ lệ này, nhất là yếu tố tuổi. Trong vòng 40 năm (1940-1980) cấu trúc theo tuổi của thành phố này đã thay đổi với sự gia tăng số người nằm trong nhóm tuổi già hơn. Để so sánh, tốt nhất nên sử dụng tỷ lệ tử vong chuẩn hóa theo tuổi.

3. Trên thực tế nước Mỹ có dân số lớn hơn Guyana nhưng không thể lấy đó để giải thích cho sự khác biệt này. Bởi vì, tỷ lệ tử vong đã đề cập đến con số chết trên 1000 dân ở cả 2 nước. Tỷ lệ tử vong thô ở các nước đang phát triển thấp mặc dù tỷ lệ tử vong riêng phần theo tuổi cao, nhất là tỷ lệ tử vong ở trẻ em dưới 5 tuổi rất cao bởi vì số người già chiếm một tỷ lệ nhỏ đã làm cho tỷ lệ tử vong thô thấp.

4. a) Tỷ lệ tử vong chuẩn hóa theo tuổi là thích hợp nhất. Tỷ lệ tử vong thô không chỉ phản ánh tỷ lệ chết đặc hiệu theo tuổi mà còn cả phân bố tuổi của quần thể đó. Vì rằng Chile cũng như Guyana (các nước đang phát triển) có quần thể dân trẻ hơn so với quần thể dân nước Mỹ, do đó muốn đem so sánh nguy cơ chết vì bệnh tim mạch 2 nước, người ta sử dụng tỷ lệ tử vong chuẩn hóa theo tuổi.

b) Khi chuẩn hóa tỷ lệ tử vong theo tuổi, người ta đã loại bỏ được sự khác biệt về phân bố tuổi, chính vì vậy tỷ lệ tử vong do bệnh tim mạch (đã chuẩn hóa) tại Mỹ chỉ bằng 2,3 lần còn nếu sử dụng tỷ lệ tử vong thô đem so sánh, con số khác biệt này lên đến 4,7 lần.



RR C_{Te} I_{De}

C_{To} I_{De}

- 5. a) $1000/100000 \times 1000 = 10$ trên 1000
- b) $600/4500 \times 1000 = 13,3$ trên 1000 đàn ông.
 $400/5500 \times 1000 = 7,3$ trên 1000 đàn bà.
- c) $45/100000 \times 1000 = 0,45$ trên 1000.
- d) $45/50 \times 100 = 90\%$.
- 6. c)
- 7. e)

ĐO LƯỜNG NGUY CƠ

I. MỤC TIÊU

- 1. Định nghĩa nguy cơ tương đối, nguy cơ tuyệt đối và nguy cơ quy thuộc.
- 2. Giải thích ý nghĩa áp dụng của từng loại nguy cơ nêu trên.

II. NỘI DUNG

1. Nguy cơ tương đối và nguy cơ quy thuộc.

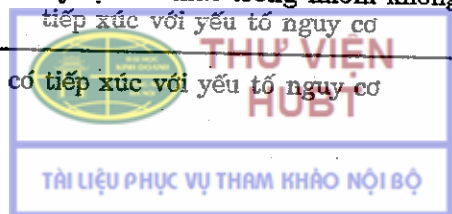
Nguy cơ tương đối và nguy cơ quy thuộc là 2 số đo của sự kết hợp giữa tiếp xúc với một yếu tố đặc biệt nào đó và nguy cơ mắc bệnh.

* Nguy cơ tương đối =
$$\frac{\text{Tỷ lệ mới mắc trong nhóm tiếp xúc với yếu tố nguy cơ}}{\text{Tỷ lệ mới mắc trong nhóm không tiếp xúc với yếu tố nguy cơ}}$$

* Nguy cơ tuyệt đối =
$$\text{Tỷ lệ mới mắc trong nhóm tiếp xúc} - \text{Tỷ lệ mới mắc trong nhóm không tiếp xúc.}$$
 $I_e - I_o$

* Đôi khi người ta còn biểu thị nguy cơ quy thuộc dưới dạng phân số phòng bệnh (trong trường hợp yếu tố nguy cơ ở đây là yếu tố phòng bệnh) trong nhóm tiếp xúc như sau:

$$\frac{\text{Tỷ lệ mới mắc trong nhóm có tiếp xúc với yếu tố nguy cơ}}{\text{Tỷ lệ mới mắc trong nhóm không tiếp xúc với yếu tố nguy cơ}}$$



RR: Relative Risk

Nguy cơ tuyệt đối tương ứng với tỷ lệ mới mắc, có nghĩa là tỷ lệ xuất hiện một bệnh hoặc một tình trạng sinh lý nào đó. Nguy cơ tuyệt đối là một tỷ lệ cơ bản để từ đó người ta tính được nguy cơ tương đối và nguy cơ quy thuộc. Các nhà lâm sàng sử dụng nguy cơ tương đối để biểu thị nguy cơ của một nhóm người có tiếp xúc với một yếu tố (như hút thuốc lá, cao huyết áp, hoặc nhóm nam giới) so với nguy cơ của một nhóm đối chứng tương tự nhưng không tiếp xúc với chính yếu tố đó (như không hút thuốc lá, không bị cao huyết áp, hoặc nhóm phụ nữ).

Giả sử yếu tố xem xét ở đây là hút thuốc lá thì việc tính toán nguy cơ tuyệt đối cho các nhà lâm sàng biết được nguy cơ mắc bệnh cao ở những người hút thuốc lá là bao nhiêu so với nhóm những người không hút thuốc lá. Những người hút thuốc lá có thể là nhóm có nguy cơ bị bệnh nào đó cao hơn (xét theo thói quen hút thuốc lá), và người ta có thể sử dụng thử nghiệm sàng lọc nhằm phát hiện những trường hợp mắc bệnh mà chưa hề có biểu hiện lâm sàng. Những yếu tố mà có liên quan đến tình trạng mắc một bệnh nào đó xét theo nguy cơ thì được người ta gọi là yếu tố nguy cơ đối với bệnh đó.

Nguy cơ tương đối chỉ ra lợi ích mà người bệnh có thể có được, nếu yếu tố nguy cơ bị loại trừ. Điều đó có nghĩa là khi yếu tố nguy cơ giảm xuống hoặc thay đổi (như cai thuốc lá) người ta có thể đo lường nguy cơ tương đối để đoán trước. Nguy cơ tương đối không đo lường xác suất mà một ai đó có yếu tố nguy cơ sẽ phát triển bệnh. Ví dụ nếu người ta tính được nguy cơ tương đối (RR) liên quan với sự có mặt của một yếu tố là 10, thì chỉ có ý nghĩa là xác suất mắc bệnh nào đó ở những người có tiếp xúc với yếu tố nguy cơ cao hơn gấp 10 lần so với những người ở nhóm không tiếp xúc với yếu tố nguy cơ. Tuy nhiên trong trường hợp các bệnh hiếm thì những người có tiếp xúc với yếu tố nguy cơ chỉ có cơ hội bị bệnh rất thấp. Điều này rất rõ ràng trong các nghiên cứu về bệnh hiếm. Ví dụ như nghiên cứu về những phụ nữ sử dụng viên thuốc tránh thai dài hạn có nguy cơ tương đối bị bệnh ung thư gan cao. Tuy nhiên tỷ lệ mới mắc bệnh này rất thấp do vậy nguy cơ tương đối ở nhóm những người tiếp xúc có cao cũng ít có ý nghĩa trong việc so sánh lợi ích có thể thu được. Điều này đặc biệt quan trọng cần phải chú ý trong các thiết kế nghiên cứu hồi cứu. Trong các thiết kế nghiên cứu này, người ta không thể có được tỷ lệ mới mắc trong nhóm tiếp xúc cũng như nhóm không tiếp xúc, do đó không thể áp dụng công thức tính nguy cơ tương đối (RR) này được mà phải ước tính gián tiếp thông qua tỷ suất chênh lệch và tình trạng mắc bệnh (OR).

Nguy cơ tương đối biểu thị độ mạnh của sự kết hợp giữa một yếu tố nguy cơ và bệnh tật. Ở đây, nếu nguy cơ tương đối mà cao, người ta thường hướng tới xem xét vấn đề căn nguyên của bệnh. Nguy cơ tương đối rất có ích trong việc nghiên cứu tìm kiếm căn nguyên gây bệnh.

Nguy cơ quy thuộc đo lường con số nguy cơ tuyệt đối (tỷ lệ mới mắc) mà người ta có thể quy cho do một yếu tố nguy cơ cụ thể nào đó (ví dụ như hút thuốc lá). Nguy cơ quy thuộc được tính toán bằng cách lấy tỷ lệ mới mắc bệnh ở nhóm tiếp xúc (như nhóm những người hút thuốc lá) trừ đi tỷ lệ mới mắc bệnh ở nhóm không tiếp xúc với yếu tố nguy cơ (nhóm không hút thuốc lá). Con số dôi ra đó chính là nguy cơ quy thuộc gây nên do hút thuốc lá. Nói chung, nguy cơ quy thuộc chỉ ra phần khác biệt về tỷ lệ mắc bệnh giữa nhóm tiếp xúc và nhóm không tiếp xúc với yếu tố nguy cơ nghiên cứu gây ra. Nguy cơ

quy thuộc là một số đo rất cần thiết để đo lường tác động của một yếu tố nguy cơ (hoặc một yếu tố có tính chất phòng bệnh) lên sức khỏe của cộng đồng, đặc biệt là nguy cơ quy thuộc quần thể rất có ý nghĩa đối với cán bộ lập kế hoạch sức khỏe cho cộng đồng. Bởi vì nó đo lường lợi ích có thể có được nếu yếu tố nguy cơ làm giảm đi trong quần thể dân cư.

2. Ý nghĩa lâm sàng.

Mặc dù một yếu tố nguy cơ có thể có số đo nguy cơ tương đối RR cao xét theo một bệnh hoặc một trạng thái sinh lý nào đó, nếu yếu tố đó là hiếm gặp trong quần thể thì tác động trên quần thể sẽ là rất nhỏ. Ví dụ những bệnh nhân với tình trạng bị bệnh polip mang tính chất gia đình có nguy cơ tương đối RR bị ung thư đại tràng cao gấp 20 lần so với những bệnh nhân mà không bị bệnh polip mang tính chất gia đình. Nhưng tỷ lệ mới mắc ung thư đại tràng xảy ra do bị bệnh polip mang tính chất gia đình lại rất nhỏ trên quần thể làm cho chỉ số này ít có ý nghĩa. Do đó, để có ý nghĩa điều cần thiết là yếu tố nguy cơ cần có cả nguy cơ tương đối cao và yếu tố này cũng phải là khá phổ biến gặp trong quần thể để nó đủ ảnh hưởng lên tỷ lệ mới mắc bệnh trong quần thể.

III. BÀI TẬP

1. Từ kết quả ở bảng 1 sau đây, hãy tính:

- nguy cơ tương đối giữa nhóm những người hút thuốc lá và nhóm không hút thuốc lá.
- Nguy cơ quy thuộc cho những người hút thuốc lá.

Bảng 1- TỶ LỆ CHẾT DO BỆNH UNG THƯ PHỔI Ở NHỮNG NGƯỜI TUỔI TỪ 35 TRÒ LÊN.

	Tỷ lệ chết do ung thư phổi trên 1000 người tuổi > 35, hàng năm
Những người không hút thuốc lá	0,07
Những người hút thuốc lá	0,96

2. Hãy giải thích kết quả thu được dựa vào số liệu của bài tập trên theo cách hiểu của anh chị.

3. Việc sử dụng, tính toán nguy cơ tương đối có giúp ích gì cho các đối tượng sau đây:

- Các nhà lâm sàng.
- Các nhà nghiên cứu.

4. Việc sử dụng, tính toán nguy cơ quy thuộc có giúp ích gì cho các đối tượng sau đây:

- Những người chịu trách nhiệm các chương trình mang tính chất phòng bệnh.
- Những người chịu trách nhiệm lập kế hoạch y tế ở tuyến trung ương.

5. a) Nguy cơ tuyệt đối đo lường cái gì?

b) Nó được sử dụng khi nào.



Attributable Risk người hút thuốc là có nguy cơ 10 P (10%)
 hơn 7 K^o C^o là 3,1 lần.

6. Nam giới ở lứa tuổi 35 mà nghiện hút thuốc lá nặng có nguy cơ tương đối bị ung thư phổi là 14. Hãy tính xác suất cho 1 người nam giới 35 tuổi hút thuốc lá nặng mắc ung thư phổi trong năm.

7. Một nghiên cứu được tiến hành trên quần thể ở 2 khu vực nông thôn và thành thị nhằm đo lường sự xuất hiện bệnh ung thư đường hô hấp. Hãy tính nguy cơ tương đối (RR) và nguy cơ quy thuộc (AR) so sánh 2 khu vực dựa trên 3 loại tỷ lệ đo lường sự xuất hiện bệnh sau đây:

	Thành thị	Nông thôn
Tỷ lệ mới mắc/100000 năm-người.	60	15
Tỷ lệ mới mắc tích lũy (CI) trong thời gian 5 năm tính trên 100000 dân.	315	80
Tỷ lệ hiện mắc tính trên 100000 dân (P).	275	70

8. Người ta đã tiến hành thử nghiệm vaccin phòng bệnh cúm trên một nhóm đối tượng là sinh viên y khoa tình nguyện. Trong tổng số 95 người nhận vaccin giả có 8 người bị mắc bệnh cúm. Ngoài ra người ta cũng đã thông báo có 27% những người nhận vaccin và 24% những người nhận vaccin giả bị mắc triệu chứng khó chịu trong người. Hãy tính nguy cơ tương đối cho những tình huống sau đây:

- Mắc bệnh cúm.
- Mắc các triệu chứng khó chịu trong người sau khi nhận vaccin.
- Hãy giải thích kết quả tính toán được bằng sự hiểu biết của anh hay chị?

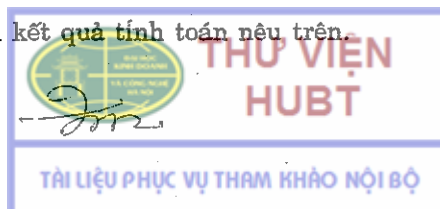
9. Trong một nghiên cứu về tỷ lệ mắc và tỷ lệ tử vong do nhồi máu cơ tim được tiến hành ở nhóm đàn ông có gia đình và nhóm đàn ông sống độc thân. Kết quả thu được như sau.

Tỷ lệ bị nhồi máu cơ tim ở đàn ông 40-64 tuổi (tỷ lệ chuẩn hóa theo tuổi).

	Tỷ lệ mới mắc 100000 năm-người	Tỷ lệ tử vong 100000 năm-người
Nhóm có gia đình	1371	498
Nhóm sống độc thân	1228	683

Hãy tính nguy cơ tương đối trong các tình huống sau:

- Mắc nhồi máu cơ tim của nhóm đàn ông có gia đình so với nhóm sống độc thân.
- Tử vong do nhồi máu cơ tim của nhóm đàn ông có gia đình so với nhóm sống độc thân.
- Hãy giải thích kết quả tính toán nêu trên.



ĐP ĐM

TRẢ LỜI BÀI TẬP

1. a) $0,96/0,07 = 13,7$

b) $0,96 - 0,07 = 0,89$

2. a) Ở những người tuổi từ 35 trở lên hút thuốc lá có nguy cơ chết vì bệnh ung thư phổi cao gấp 13 lần so với những người cùng độ tuổi mà không hút thuốc lá.

b) Tỷ lệ chết do ung thư phổi trong nhóm hút thuốc lá là 0,96. Con số 0,96 đôi ra là quy cho vấn đề hút thuốc lá gây nên, hay nói cách khác nguy cơ quy thuộc do hút thuốc lá là 0,89.

3. a) Nguy cơ tương đối giúp cho các nhà lâm sàng biết được mức độ nguy cơ gia tăng mà bệnh nhân có tiếp xúc với một yếu tố nào đó (như hút thuốc lá, cao huyết áp, lượng Cholesterol máu cao) có được so với một bệnh nhân không tiếp xúc với yếu tố đó. Nguy cơ tương đối có khả năng giúp các nhà lâm sàng xác định những bệnh nhân có nguy cơ cao mắc một bệnh nào đó. Nhưng nó không cung cấp cho các nhà lâm sàng nguy cơ tuyệt đối mắc bệnh đó.

b) Nguy cơ tương đối đo lường độ mạnh của sự kết hợp giữa yếu tố nguy cơ và bệnh. Nếu nguy cơ tương đối mà cao, thường gợi ý cho các nhà nghiên cứu hướng tới vấn đề căn nguyên hoặc nguyên nhân gây bệnh.

4.a) Nguy cơ quy thuộc đo lường tác động ảnh hưởng của một yếu tố nguy cơ nào đó được loại trừ lên tỷ lệ mới mắc bệnh. Ví dụ các chương trình phòng bệnh chẳng hạn, người ta thường sử dụng chỉ số này để chứng minh về cơ bản tác động phòng bệnh của chương trình.

b) Việc xác định nguy cơ quy thuộc đối với từng loại tiếp xúc khác nhau hỗ trợ cho vấn đề lập kế hoạch dịch vụ y tế.

5. a). Nguy cơ tuyệt đối đo lường con số mới mắc bệnh hay nói cách khác nó đo lường tốc độ xuất hiện bệnh.

b) Nó được sử dụng trong các tình huống mang tính chất dự đoán.

6. Nếu chỉ đưa ra một mình con số nguy cơ tương đối thì không thể tính được.

7. Kết quả tóm tắt như sau.

Loại tỷ lệ	Nguy cơ tương đối (RR)	Nguy cơ quy thuộc (AR).
Tỷ lệ mới mắc	$60/15=4$	$60-15=45 \times 10^{-5}$ trong 1 năm
Tỷ lệ mới mắc tích lũy	$315/80=3,9$	$315-80=235 \times 10^{-5}$ trong 5 năm
Tỷ lệ hiện mắc	$275/70=3,9$	$275-70=205 \times 10^{-5}$

8. a) Tỷ lệ mới mắc tích lũy (CIe) ở nhóm nhận vaccin là:

$$3/95 = 316 \times 10^{-4}$$

Tỷ lệ mới mắc tích lũy (CIo) ở nhóm không nhận vaccin là:

(hay nhóm nhận vaccin giả).

$$8/48 = 1667 \times 10^{-4}$$



$$\text{Nguy cơ tương đối RR} = \frac{CI_e}{CI_o} = 0,19$$

$$\text{b) Nguy cơ tương đối RR} = \frac{0,27}{0,24} = 1,1$$

c) Tình trạng xuất hiện mắc bệnh cúm trong nhóm nhận vaccin thật chỉ là 20% so với nhóm nhận vaccin giả (hay nói cách khác giảm được 80%). Sự xuất hiện tình trạng khó chịu trong người sau khi nhận vaccin ở nhóm nhận vaccin thật cao hơn 10%.

$$9. \text{ a) Nguy cơ tương đối RR} = \frac{1371}{1228} = 1,1$$

$$\text{b) Nguy cơ tương đối RR} = \frac{498}{683} = 0,7$$

c) Tỷ lệ tử vong do nhồi máu cơ tim ở những người đàn ông có gia đình thấp hơn so với những người đàn ông sống độc thân. Hoặc có thể giải thích là nhồi máu cơ tim nhẹ hơn (hoặc được chăm sóc tốt hơn) ở những người đàn ông có gia đình so với những người đàn ông sống độc thân.

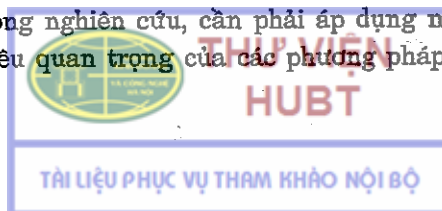
BIẾN SINH HỌC

I. MỤC TIÊU

1. Nhận thức được sự biến thiên và ý nghĩa của nó trong các số liệu điều tra nghiên cứu y sinh học.
2. Nhận định khái quát được phân phối tần số, tần suất các chỉ số tập trung và phân tán qua các bảng phân phối tần số và biểu đồ phân phối tần số.
3. Lập được bảng phân phối tần số và vẽ được biểu đồ phân phối tần số các loại dữ kiện khác nhau.

II. NỘI DUNG

Các số đo trong nghiên cứu y sinh học luôn luôn thay đổi theo từng cá thể, từng người quan sát, từng lần đo... Các số đo có thể luôn luôn thay đổi này gọi là các biến sinh học, mà dưới góc nhìn thống kê học gọi là biến số trong nghiên cứu. Để làm cho các biến sinh học được tin cậy trong nghiên cứu, cần phải áp dụng một số thuật toán thống kê. Một trong những mục tiêu quan trọng của các phương pháp thống kê là: đảm bảo rằng các



nhà nghiên cứu khác nhau đều rút ra những kết luận như nhau trước các dữ kiện như nhau. Vì trong y sinh học, sự biến thiên của các số liệu nghiên cứu là không thể tránh khỏi. Và từ khái niệm biến thiên này cần phải xác nhận ý nghĩa thống kê các kết quả thu được.

Trong việc phân tích các dữ kiện nghiên cứu, như vậy, thì sự biến thiên và ý nghĩa thống kê là 2 điểm có tầm quan trọng đặc biệt.

Cũng như các phương pháp khoa học khác, các phương pháp thống kê phải tuân theo những điều kiện vận dụng chính xác. Đa đa số các thuật toán thống kê chỉ có giá trị khi các dữ kiện nghiên cứu là độc lập với nhau, và phân phối của chúng phù hợp với một quy luật lý thuyết nhất định (phần lớn là với luật phân phối chuẩn).

Chúng ta sẽ lần lượt nghiên cứu:

- Phân phối tần số
- Độ tập trung của dãy thống kê
- Độ phân tán của dãy thống kê
- Ý nghĩa thống kê
- Tổng quát hóa kết quả nghiên cứu mẫu ra quần thể toàn bộ.

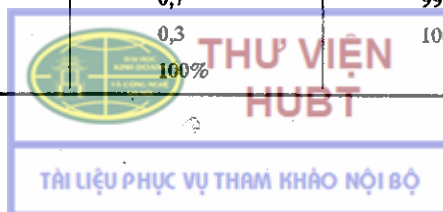
Dù là đối với các dữ kiện định tính hay định lượng, thì công việc đầu tiên, sau khi thu thập kết quả nghiên cứu, là lập một bảng phân phối tần số, và vẽ một biểu đồ phân phối tần số của bảng đó, để có thể sơ bộ nhận định xem kết quả nghiên cứu của chúng ta được phân phối theo quy luật phân phối như thế nào, biến thiên ra sao.

Bảng phân phối tần số.

Xem một vài thí dụ sau đây:

Bảng 1- BẢNG PHÂN PHỐI TẦN SỐ HỘ GIA ĐÌNH THEO SỐ NGƯỜI TRONG HỘ
(biến rời rạc, không gộp lớp)

Số người trong hộ	Tần số (hộ)	Tần suất	Cộng dồn tần suất
1	14	4,7	4,7
2	21	7,0	11,7
3	52	17,4	29,1
4	58	19,4	48,5
5	48	16,1	64,6
6	37	12,4	77,0
7	31	10,4	87,4
8	16	5,3	92,7
9	10	3,3	96,0
10	7	2,3	98,3
11	2	0,7	99,0
12	2	0,7	99,7
13	1	0,3	100,0
Cộng số hộ	299	100%	



Bảng 2- BẢNG PHÂN PHỐI TẦN SỐ TRỌNG LƯỢNG DDT PHUN Ở 1022 NHÀ
(biến liên tục, có gộp lớp đều nhau)

Trọng lượng DDT (g)	Tần số (nhà)	Tần suất	Tần suất dồn
460-469	5	0,5	0,5
470-479	22	2,1	2,6
480-489	24	2,3	4,9
490-499	13	1,3	6,2
500-509	15	1,5	7,7
510-519	20	1,9	9,6
520-529	46	4,5	14,1
530-539	44	4,3	18,4
540-549	73	7,1	25,5
550-559	47	4,6	30,1
560-569	49	4,8	34,9
570-579	91	8,9	43,8
580-589	104	10,2	54,0
590-599	88	8,6	62,6
600-609	77	7,5	70,1
610-619	57	5,6	75,7
620-629	73	7,1	82,8
630-639	101	9,9	92,7
640-649	44	4,3	97,0
650-659	15	1,5	98,5
660-669	3	0,3	98,8
670-679	3	0,3	99,1
680-689	2	0,2	99,3
690-699	2	0,2	99,5
700-709	4	0,5	100,0
Cộng	1022	100,0	100,0

Bảng 3- BẢNG PHÂN PHỐI TẦN SỐ CHIỀU DÀI CÁNH CỦA 93 MUỖI A.CULICIFACIES
(biến liên tục, có gộp lớp đều nhau)

Chiều dài cánh (mm)	Tần số	Tần suất	Tần suất dồn
2,05-2,14	1	1,1	1,1
2,15-2,24	1	1,1	2,2
2,25-2,34	9	9,7	11,9
2,35-2,44	14	15,0	26,3
2,45-2,54	14	15,0	41,9
2,55-2,64	20	21,5	63,4
2,65-2,74	8	8,6	72,0
2,75-2,84	18	19,4	91,4



2,85-2,94	4	4,3	95,7
2,95-3,04	2	2,1	97,8
3,05-3,14	1	1,1	98,9
3,15-3,24	0	0	0
3,25-3,34	1	1,1	100,0
Cộng	93	100,0	

Bảng 4- BẢNG PHÂN PHỐI TẦN SỐ CỦA 1041 TRƯỜNG HỢP MẮC LAO PHỐI THEO TUỔI
(Biến rời rạc, gộp lớp không đều nhau)

Nhóm tuổi	Số mắc	Tần suất%	Tần suất dồn
Dưới 5 tuổi	80	7,7	7,7
5 - 14	161	15,5	23,2
15 - 44	600	57,6	80,8
45 trở lên	200	19,2	100,0
Cộng	1041	100,00	

Từ các bảng phân phối tần số, có thể rút ra một số nhận định về độ tập trung và độ phân tán sau:

- Một: gia đình 4 người (bảng 1) 2,55-2,64 mm (bảng 3)
- Trung vị: 2,65-2,74 mm (bảng 3) gia đình 7 người (bảng 1)
- Cả 2 dãy thống kê trong bảng 1 và bảng 3 thì một và trung vị đều không ở cùng một vị trí, nhưng ở bảng 3 một và trung vị gần sát nhau: phân phối của dãy này gần như cân đối, còn ở bảng 1 thì một và trung vị xa nhau nhiều nên ít cân đối hơn, có thể coi là một phân phối lệch.

(Sau này thì tính số trung bình, ta sẽ xem cả 3 chỉ số một, trung vị, số trung bình để nhận định về phân phối của dãy số).

- Cột tần số cũng có thể cho ta thấy sự phân bố đặc trưng nghiên cứu theo phân bố đối xứng hay lệch (nhìn vào cột tần suất thì dễ thấy rõ hơn) = 58 hộ (tần số) = 19,4% (tần suất) là lớn nhất tương ứng với 4 người/hộ ở bảng 1: phân phối lệch trái.

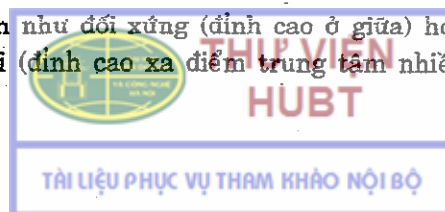
- Tần suất dồn còn cho ta thấy ở một mức nào đó của đặc trưng nghiên cứu có tỷ lệ phần trăm là bao nhiêu: 48,5% số hộ (gần nửa) là có từ 4 người trở xuống (bảng 1).

2. Đồ thị phân phối tần số.

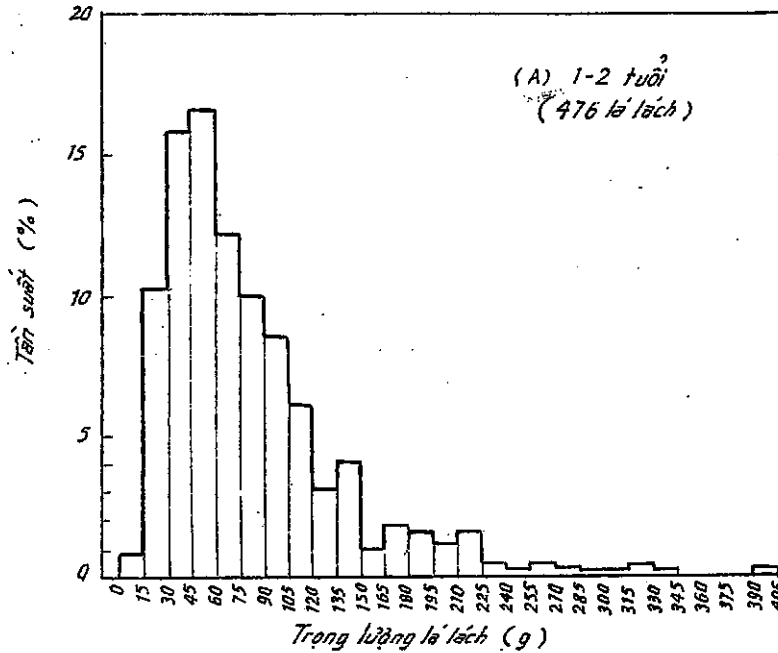
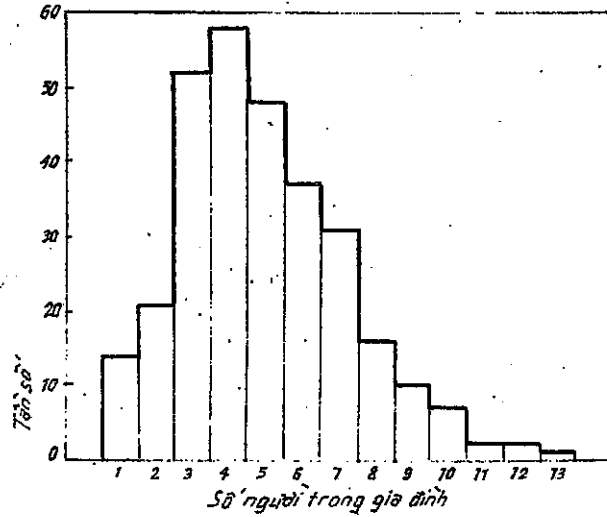
Từ các đồ thị cột và/hoặc đa giác tần số, ta nhận định được:

- Kiểu phân phối tần số:** Ta sẽ xét về tính chất đối xứng của phân phối tần số. Nói chung, đường biểu diễn tần số có 1 đỉnh ứng với tần số lớn nhất (chính là một của dãy số) và giảm dần ở 2 bên đến các tần số nhỏ.

Đường biểu diễn này có thể gần như đối xứng (đỉnh cao ở giữa) hoặc rất không đối xứng, hoặc lệch trái hoặc lệch phải (đỉnh cao xa điểm trung tâm nhiều hay ít, và lệch sang một bên).



Hình 1- Phân bố tần số hộ gia đình theo số người trong hộ (đường biểu diễn gần đối xứng).



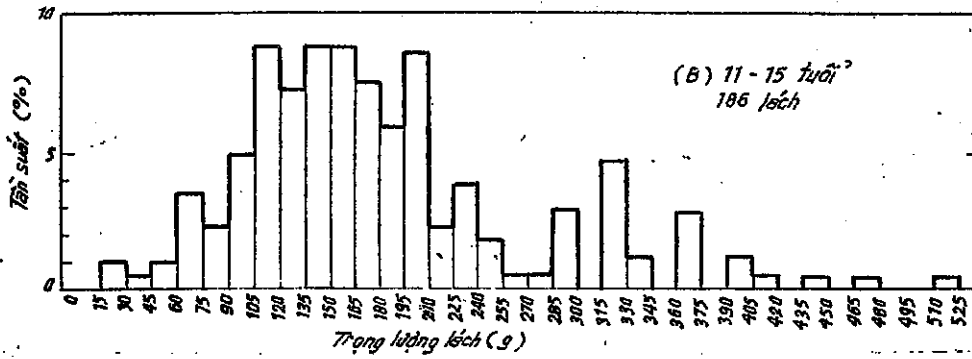
Hình 2- Phân bố tần suất lá lách theo trọng lượng của 476 trẻ 1-2 tuổi. Đường biểu diễn không đối xứng (lệch trái)
Trọng lượng lá lách (g)

Khi đường cong đối xứng (hoặc gần như đối xứng) đối với một tần số đủ lớn, thì có thể coi kết quả nghiên cứu bắt đầu xấp xỉ chuẩn, hoặc chắc chắn xấp xỉ chuẩn (tùy thuộc vào cỡ mẫu, kích thước quần thể, tỷ lệ tiêu thức nghiên cứu trong quần thể, độ phân tán của tiêu thức nghiên cứu...) và được xử lý. Kết quả theo quy luật phân phối chuẩn với trung bình m và độ lệch chuẩn s :

trong khoảng $m \pm 1s$, sẽ chứa đựng 68% số quan sát

trong khoảng $m \pm 2s$ sẽ chứa đựng 95% số quan sát

trong khoảng $m \pm 3s$ sẽ chứa đựng 99% số quan sát



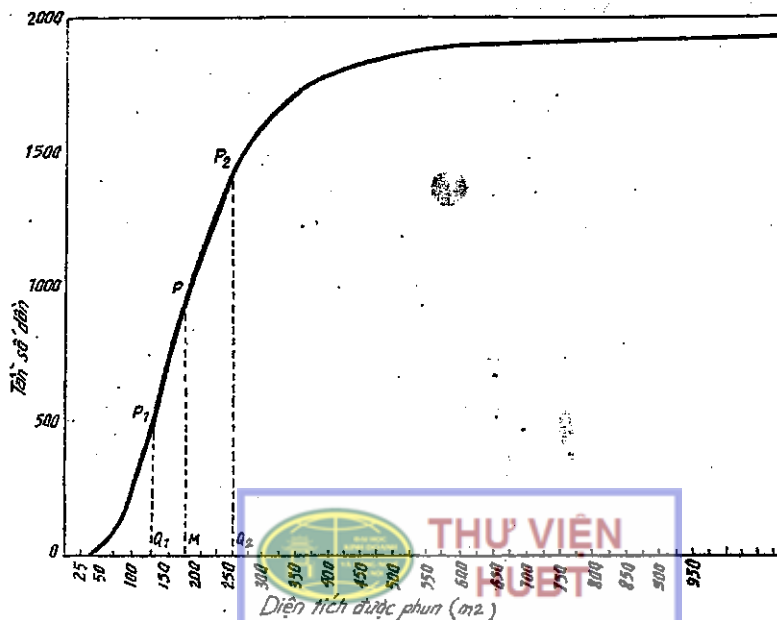
Hình 3- Phân bố tần suất lách theo trọng lượng (tính bằng g) của 186 trẻ 11-15 tuổi. (Đồ thị phân tán)

Bảng 4- PHÂN BỐ TẦN SỐ CỦA 1925 NHÀ THEO DIỆN TÍCH ĐƯỢC PHUN

Diện tích phun (m ²)	Tần số (nhà)	Tần số đơn
26-50	4	4
51-75	40	44
76-100	91	135
101-125	193	328
126-150	206	534
151-175	264	798
176-200	209	1007
201-225	183	1190
226-250	137	1327
251-275	146	1473
276-300	107	1580
301-325	68	1648
326-350	66	1714
351-375	38	1752
376-400	27	1779
401-425	26	1805
426-450	26	1831
451-475	20	1851
476-500	11	1862
501-525	10	1872
526-550	11	1883
551-575	9	1892
576-600	6	1898
601-650	5	1903
626-650	4	1907
651-675	3	1910

676-700	2	1912
701-725	2	1914
726-750	2	1916
751-775	3	1919
776-800	2	1921
801-825	0	1921
826-850	0	1921
851-875	1	1922
876-900	0	1922
901-925	1	1923
926-950	0	1923
951-975	0	1923
976-1000	1	1924
1001-1025	0	1924
1026-1050	0	1924
1051-1075	0	1924
1076-1100	1	1925
Cộng	1925	

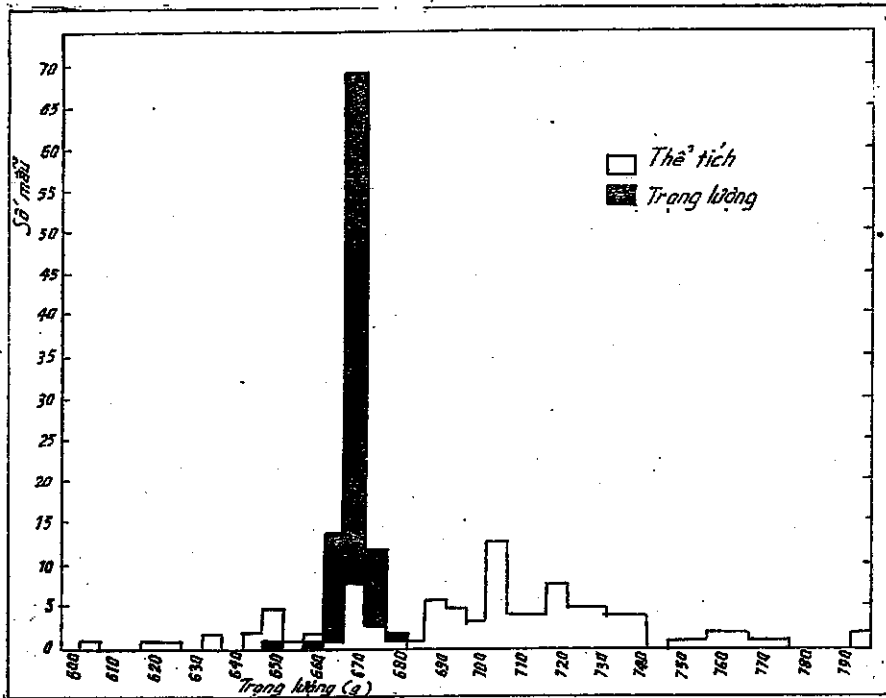
b) *Trung vị*: tứ phân vị và bách phân vị: Khi ta lập bảng tần số dồn (bảng 4) và vẽ đồ thị tần số dồn (hình 4) rồi trên trục tung ta xác định điểm tương ứng với một nửa tần số tổng cộng, từ điểm đó kẻ một đường thẳng ngang song song với trục hoành, đường thẳng này sẽ cắt đường cong ở một điểm P, rồi từ P ta lại hạ một đường thẳng đứng xuống, nó sẽ cắt trục hoành ở điểm M. Trị số của điểm M trên trục hoành rất quan trọng nó chính là trung vị của dãy số, nó có một nửa tần số có trị số cao hơn và một nửa thấp hơn trị số đó. Nếu ta chia trục tung ra làm 100%, thì đồ thị tần số dồn sẽ dùng để tính độ tứ phân vị (25% và 75%) hoặc độ bách phân vị (2,5% và 97,5%), còn trung vị ứng với điểm 50% (cách làm như trên).



Hình 4 - Phân bố tần số dồn của 1925 nhà theo diện tích phun. Đồ thị tần số dồn. Diện tích được phun (m^2)

c) Sự phân tán của dãy thống kê:

Trên một đồ thị phân phối, sự phân bố của các cột liền, hoặc cột chữ nhật tập trung lại một chỗ hoặc phân tán rộng trên trục hoành độ (biểu thị số đo của đặc trưng nghiên cứu) rất có ích cho việc nhận định độ phân tán của dãy thống kê: Nếu chúng sát nhau và gần như quần tụ xung quanh trung bình ở giữa trục hoành làm cho đồ thị có một đỉnh rõ rệt thì phân bố là đối xứng và ít phân tán, ngược lại chúng cách xa nhau, mà các trị số cao lại xa trung tâm trên trục hoành thì dãy số bị phân tán nhiều.



Hình 5- So sánh lượng DDT đã phun theo thể tích và theo trọng lượng.

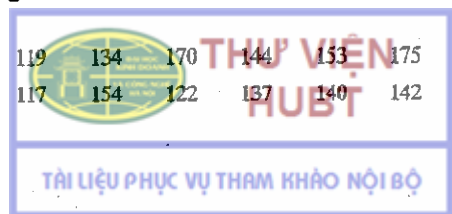
Các hiện tượng y sinh học, như đã đề cập đến, khi được nghiên cứu, thế nào cũng có một độ phân tán nhất định. (Chúng ta sẽ tính độ phân tán trong nội bộ mỗi dãy thống kê và việc so sánh độ phân tán nhiều dãy thống kê này là việc rất quan trọng trong phân tích thống kê). Nếu chỉ căn cứ vào đồ thị phân phối, như ở hình sau đây (hình 5) cũng thấy được rằng, trước hết ống đựng DDT là không đều nhau, thứ hai là, trọng lượng trung bình các mẫu thu được không tương ứng tí nào với trọng lượng ấn định phải phun. Ngoài ra, các mẫu gặp nhiều nhất là 705g và 720g, trong khi một của dãy phải là 670g.

III. BÀI TẬP

Bài tập 1.

Câu hỏi: Hãy sắp xếp chuỗi thống kê sau từ nhỏ đến lớn:

126	113	126	139	171	119	134	170	144	153	175
166	180	139	126	149	117	154	122	137	140	142



137	152	149	129	146	175	168	104	144	134	145
118	131	122	153	149	169	132	147	150	152	140
145	161	153	177	146	152	112	140	145	152	151
155	112	162	188	156	160	170	165	156	157	161
132	162	155	170	160	172	158	155	162	189	105

Trả lời

104, 105

(112)2, 113, 117, 118, 119

(122)2, (126)3, 129

131, (132)2, (134)2, (137)2, (139)2,

(140)3, 142, (144)2, (145)2, 146, 147, 148, (149)3,

150, 151, (152)4, (153)3, 154, (155)3, (156)2, 157, 158,

(160)2, (161)2, (162)2, 165, (168)2, 169,

(170)3, 171, 172, 173, 175, 177,

180, 182, 188, 189,

Bài tập 2

Người ta đã tìm mức acid uric trong huyết thanh của 267 người cho máu, kết quả ghi trong bảng phân phối tần số sau:

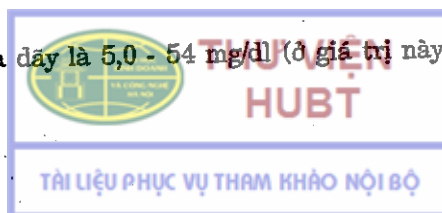
Acid uric (mg/dl)	Tần số	Tần suất	Tần suất dồn
3,0 - 3,4	2	0,8	0,8
3,5 - 3,9	15	5,6	6,4
4,0 - 4,4	33	12,4	18,7
4,5 - 4,9	40	15,0	33,7
5,0 - 5,4	54	20,2	53,9
5,5 - 5,9	47	17,6	71,5
6,0 - 6,4	38	14,2	85,8
6,5 - 6,9	16	6,0	91,8
7,0 - 7,4	15	5,6	97,4
7,5 - 7,9	3	1,1	98,5
8,0 - 8,4	1	0,4	98,9
8,5 - 8,9	3	1,1	100,0

Câu hỏi:

1. Giá trị một của dãy thống kê này là bao nhiêu?
2. Xác định trung vị của dãy thống kê?
3. Phân bố của dãy thuộc loại nào?
4. Có bao nhiêu người có mức acid uric dưới 6 mg/dl?

Trả lời:

- 1- Giá trị một của dãy là 5,0 - 5,4 mg/dl (ở giá trị này có tần suất cao nhất 20,2%).



2- Trung vị: vì dãy có 12 lớp nên trung vị sẽ là trung bình của 2 giá trị nằm giữa lớp: 5,7 - 6,2

3- Phân bố: Mốt và trung vị nằm sát gần trùng nhau phân bố trên gần như đối xứng.

4- Có 71,5% số người: $\frac{267 \times 71,5}{100} = 191$ người.

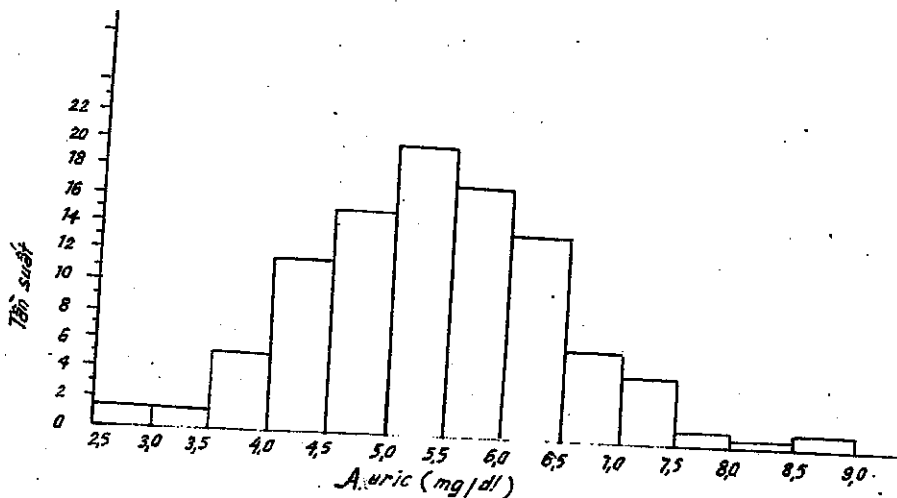
Bài tập 3.

1- Hãy vẽ đồ thị cột cho bảng dữ kiện của bài tập 2.

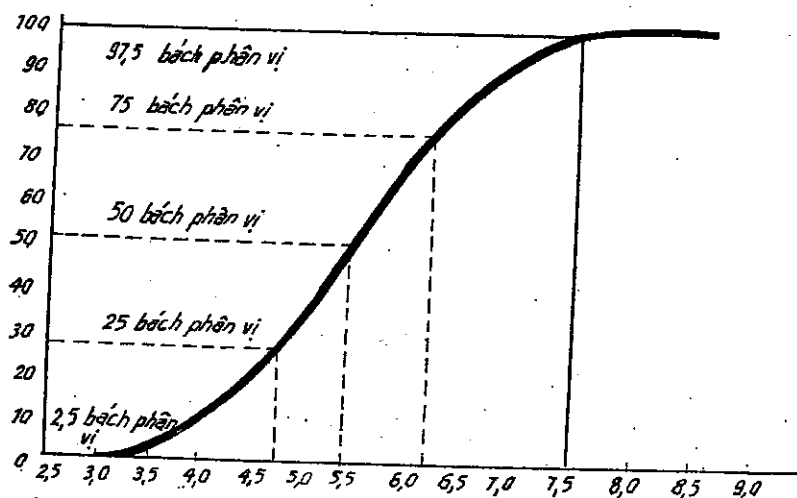
2- Hãy nhận định về kiểu phân phối tần số này.

3- Hãy vẽ đồ thị tần số dồn của bảng đó và nhận định giá trị của trung vị, các tứ phân vị, các bách phân vị.

Trả lời:



2. Biểu đồ cột này cho hình thái của một phân bố gần như đối xứng.



3. Đồ thị tần số dồn, cho ta thấy, khoảng:

- Trung vị: 5,4 mg/dl

- TPV3 = BPV 75: 6,2 mg/dl TPV1 = BPV 25 = 4,7 mg/dl

- BPV 97,5: 7,5 mg/dl; BPV 2,5: 3,5 mg/dl.



ĐỘ TẬP TRUNG VÀ ĐỘ PHÂN TÁN CỦA SỐ LIỆU NGHIÊN CỨU

I. MỤC TIÊU

- 1- Tính được các chỉ số của độ tập trung và độ phân tán đối với các loại biến số nghiên cứu khác nhau.
- 2- Ứng dụng và phiên giải được các chỉ số đó.

II. NỘI DUNG

Độ tập trung của dãy số thống kê được biểu thị chủ yếu bằng các chỉ số trung tâm sau:

- Trung bình số học
- Trung vị
- Mốt
- Trung bình nhân

Độ phân tán của dãy số thống kê được biểu thị chủ yếu bằng các chỉ số phân tán sau:


- Khoảng phân tán
- Độ lệch chuẩn.

A. CÁC DỮ KIẾN ĐỊNH LƯỢNG KHÔNG GỘP LỚP

Các dữ kiện định lượng có thể không gộp lớp hoặc cần gộp lớp trong các bảng phân phối tần số. Người ta cho rằng tần số các dữ kiện không quá 50 thì có thể không gộp lớp, nhưng dãy thống kê vượt quá 50 tần số thì nên gộp lớp.

1. Tính các chỉ số trung tâm.

1.1. Số trung bình (trung bình số học) \bar{x} :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum x}{n}$$


Giả sử có dãy 5 số liệu về diện tích nhà đã được phun DDT:

$$\begin{array}{r}
 227 \ 253 \ 185 \ 101 \ 184 \\
 196 \ 217 \ 122 \ 148 \ 233 \\
 336 \ 114 \ 109 \ 135 \ 198 \\
 227 + 253 + \dots + 198
 \end{array}$$

$$\text{thì } \bar{x} = \frac{\quad}{15} = \frac{2758}{15} = 183,9\text{m}^2$$

1.2. Trung vị

Là giá trị giữa của dãy số, như vậy sẽ có một nửa số đo nhỏ hơn và nửa kia lớn hơn giá trị giữa này. Nếu tần số là một số lẻ, trung vị là giá trị nằm ở vị trí chính giữa dãy số, nếu là số chẵn thì trung vị được quy ước là số trung bình của hai giá trị nằm giữa dãy số.

Muốn tìm trung vị phải xếp dãy số tăng dần (hoặc giảm dần). Ở thí dụ trên, có tần số là 15:

101, 109, 114, 122, 135, 148, 184, 185, 196, 198, 217, 227, 233, 253, 336.

Thì trung vị là giá trị thứ tám: 185m²

1.3. Trung bình nhân

Tính bằng hai cách:

TBN = $\sqrt[n]{x_1 \times x_2 \times \dots \times x_n}$ hoặc đơn giản hơn người ta dùng log

$$(\text{TBN}) = \frac{\log x_1 + \log x_2 + \dots + \log x_n}{n} = \frac{\sum \log x}{n} \quad \text{rồi lấy:}$$

TBN = đối log của log (TBN)

Cần chú ý: Nếu trong dãy số không có trị số nào quá bất thường, thì số trung bình là biểu thị điển hình của độ tập trung. Nếu có, thì người ta dùng trung vị hoặc trung bình nhân. Khi dãy số có quá nhiều trị số, thì điểm tập trung một cũng là một biểu thị điển hình tốt, vì cũng giống như trung vị, nó tương đối không bị ảnh hưởng bởi các trị số bất thường và kiểu phân phối tần số.

Riêng việc dùng TBN phải thận trọng khi có một trị số là số không (zéro) (TBN sẽ là số zéro) hoặc dãy số có trị số âm (không thể tính TBN được).

Trong thí dụ trên:

$$\text{TBN} = \sqrt[15]{227 \times 196 \times \dots \times 198}$$

Đơn giản hơn ta tính:

$$\begin{aligned}
 \log (\text{TBN}) &= \frac{\log 227 + \log 196 + \dots + \log 198}{15} \\
 &= \frac{33,5959}{15} = 2,2397
 \end{aligned}$$

rồi lấy đối log của 2,2397 là 173,7.

173,7 chính là trung bình nhân của dãy số.



THƯ VIỆN
HUBT

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

2. Tính các chỉ số phân tán.

Một đặc trưng quan trọng khác của một dãy thống kê là sự phân tán của các số hạng xung quanh số trung bình của dãy. Nếu các số hạng trong dãy không khác nhau nhiều thì số trung bình, hoặc trung vị... cũng đủ mô tả các số liệu thống kê đó. Nếu khác nhau nhiều, phải tính:

- *Khoảng phân tán*: Là hiệu của số đo lớn nhất với số đo nhỏ nhất. Trong thí dụ trên ta có khoảng phân tán:

$$451m^2 - 101m^2 = 350m^2$$

Khoảng phân tán ít dùng để đo độ phân tán vì nó không kể đến sự phân tán của các số hạng khác, và khi số hạng càng tăng thì sẽ có xác suất gộp các giá trị rất lớn hoặc rất nhỏ càng tăng với tần số này.

- *Độ lệch chuẩn*.

Muốn đánh giá độ phân tán hợp lý, phải chú ý đến khoảng cách của từng số hạng với số trung bình số học của dãy thống kê.

Trong thí dụ trên, ta có lần lượt 15 khoảng cách so với trung bình sau: (số thứ nhất = 227-183,9=43,1...)

$$43,1 \quad -69,9 \quad -82,9 \quad -48,9 \quad -61,9$$

$$12,1 \quad -74,9 \quad -33,1 \quad -1,1 \quad 49,1$$

$$152,1 \quad 69,1 \quad -35,9 \quad 0,1 \quad 14,1$$

Nếu chú ý đến dấu đại số, thì tổng 15 số này sẽ bằng số 0. (trong thí dụ này, tổng = -0,5 vì ta làm tròn số trung bình). Nếu chỉ lấy giá trị tuyệt đối của chúng, rồi lấy trung bình, thì ta được độ lệch trung bình (trong thí dụ này, độ lệch trung bình là 49,9).

Muốn tính độ lệch chuẩn, trước hết ta tính phương sai V_x .

Dãy là một mẫu nhỏ (dưới 50 trị số quan sát), ta dùng công thức tính V_x (nhưng thường viết gọn là V) như sau:

$$V = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1} = \frac{\sum (x - \bar{x})^2}{n - 1}$$

Nhưng người ta thường dùng công thức đơn giản nhất để tính phương sai V sau đây:

$$V = \frac{1}{n - 1} \left[(x_1^2 + x_2^2 + \dots + x_n^2) - \frac{(x_1 + x_2 + \dots + x_n)^2}{n} \right]$$

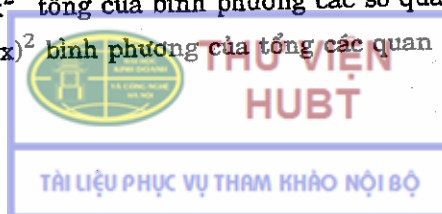
$$V = \frac{1}{n - 1} \left[\sum x^2 - \frac{(\sum x)^2}{n} \right]$$

Trong đó: V là phương sai

n : số quan sát (cỡ mẫu)

$\sum x^2$ tổng của bình phương các số quan sát.

$(\sum x)^2$ bình phương của tổng các quan sát.



rồi tính độ lệch chuẩn s:

$$s = \sqrt{V} = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

Trong thí dụ trên: $V = 4.180,64m^2$

$$\text{và } s = 64,66 m^2$$

Nếu người ta tính theo công thức đơn giản vừa nêu thì

$$V = 4.178,55$$

$$\text{và } s = 64,65 m^2$$

3. Tính sai số chuẩn ($SE_{\bar{x}}$).

Trong thí dụ trên, ta tính diện tích trung bình ở 15 nhà thì ta được $\bar{x}_1 = 183,9m^2$. Bây giờ ta lại lấy một mẫu 15 nhà khác của địa phương đó ta lại được $\bar{x}_2 \neq \bar{x}_1$. Cứ làm như thế ta sẽ được các số trung bình khác nhau: $\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_T$.

(với T là tổng số mẫu cùng cỡ có thể rút ra được từ quần thể).

Người ta đã thấy rằng, ở mỗi mẫu cỡ n ta sẽ được một trung bình và phương sai của các số đo cá thể trong mẫu nghiên cứu là V, thì ta sẽ có phương sai của các số trung bình của tất cả các mẫu này là V/n và độ lệch chuẩn của các số trung bình (chính là sai số chuẩn $SE_{\bar{x}}$, nhưng thường viết gọn là SE).

$$SE = \sqrt{\frac{V}{n}} = \frac{s}{\sqrt{n}}$$

Trong thí dụ trên:

$$SE = \sqrt{\frac{4.178,55}{16}} = \sqrt{278,57} = 16,69$$

B- CÁC DỮ KIẾN ĐỊNH LƯỢNG CÓ GỘP LỚP

(mẫu lớn, trên 50 trị số).

1. Gộp lớp các số liệu:

Không có quy ước chung cho ý niệm về gộp lớp. Song, để cho dễ thu xếp, người ta gợi ý nên căn cứ vào khoảng phân tán, với số lớp không nên quá 30, thường từ 3 đến 20 lớp và điều quan trọng hơn cả là nên căn cứ vào giá trị y học của khoảng cách lớp (đặc biệt là ở ranh giới của tình trạng bình thường và bệnh lý). Khoảng cách lớp quá lớn, sẽ thất thoát mất nhiều thông tin, ngược lại, khoảng cách hẹp sẽ đưa số lớp lên quá nhiều một cách không cần thiết.

Dù gộp lớp như thế nào, ta cũng cần hiểu rõ ý nghĩa và tôn trọng quy ước của các giới hạn trên và dưới của lớp.

Nếu lớp được xếp bằng các số nguyên:

Với các số đo định lượng chung:



Thí dụ: 10-14kg có nghĩa là từ 9,6 đến 14,5

15-19 kg có nghĩa là từ 14,6 đến 19,5

Như vậy: Khi gộp lớp với các số nguyên thì cần hiểu là:

- Giới hạn dưới thực phải thấp hơn giới hạn dưới biểu thị bằng số nguyên đó, và
- Giới hạn trên thực phải cao hơn giới hạn trên biểu thị bằng số nguyên đó đến sát với giới hạn dưới biểu thị bằng số nguyên của lớp tiếp theo. Trong thí dụ trên là ta chỉ lấy đến một số lẻ.

Riêng với số tuổi, lại theo một quy ước khác:

Thí dụ: 10-14 tuổi được hiểu là từ đúng 10 tuổi đến dưới 15 tuổi, có nghĩa là:

- Giới hạn dưới là tuổi chính xác, còn
- Giới hạn trên thực phải kéo dài đến sát ngay giới hạn dưới biểu thị bằng số nguyên của lớp kế tiếp.

Việc hiểu rõ các giới hạn này rất quan trọng trong việc tính các độ tứ phân vị, độ bách phân vị, trị số giữa lớp để tính trung bình và độ lệch chuẩn của dãy thống kê.

Ta sẽ phân biệt các biến rời rạc và các biến liên tục

1.1 Tính các chỉ số trung tâm của dãy số là biến rời rạc gộp lớp

- Số trung bình số học \bar{x} :

$$\bar{x} = \frac{\sum xf}{\sum f}$$

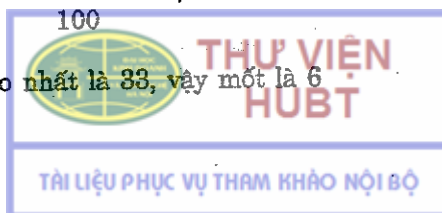
Thí dụ: Bảng phân phối tần số lông ức trái của ấu trùng A.Stephensi:

Bảng 5:

Số lông x	Tần số f	fx	Tần số dồn
3	1	3	1
4	18	72	19
5	30	150	49
6	33	198	82
7	15	105	97
8	2	16	99
9	1	9	100
Cộng	100	533	

Ở thí dụ này: $\bar{x} = \frac{533}{100} = 5,33$

- Mốt: Tần số cao nhất là 33, vậy mốt là 6



- *Trung vị:* Vì tần số là 100 (số chẵn) nên trung vị sẽ là trung bình của hai giá trị thứ 50 và thứ 51. Cả hai giá trị này đều nằm trong lớp 6, vậy trung vị là 6.

- *Trung bình nhân:* Phải lập một bảng mới cho dễ tính và đỡ làm lẫn như sau:

Bảng 6

Số lòng	log x	Tần số f	f.logx
3	0,4771	1	0,4771
4	0,6021	18	10,8378
5	0,6990	30	20,9700
6	0,7782	33	25,6806
7	0,8451	15	12,6765
8	0,9031	2	1,8062
9	0,9542	1	0,9542
Cộng		100	73,4024

$$\text{Tính trung bình các log} = \frac{\sum f \log x}{\sum f} = \frac{73,4024}{100} = 0,7340$$

TBN = đối log của 0,7340 = 5,42

1.2 Tính các chỉ số phân tán của các biến rời rạc gộp lớp.

Cách tính giống như đối với các biến rời rạc khác, chỉ thay $\sum x^2$ bằng $\sum fx^2$ và $(\sum x)^2$ bằng $(\sum fx)^2$. Để cho dễ tính, ta lập một bảng mới (dùng số liệu bảng trong thí dụ trên).

Bảng 6b

Số lòng x	Tần số f	x^2	fx	fx^2
3	1	9	3	9
4	18	16	72	288
5	30	25	150	750
6	33	36	198	1188
7	15	49	105	735
8	2	64	16	128
9	1	81	9	81
$\sum x$ 42	$\sum f = n$ 100	$\sum x^2$ 280	$\sum fx$ 553	$\sum fx^2$ 3179

Bây giờ ta tính:

Dùng công thức đơn giản nhất



THƯ VIỆN
HUBT

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

$$\begin{aligned}
 V &= \frac{1}{n-1} \left[\sum fx^2 - \frac{(\sum fx)^2}{n} \right] \\
 &= \frac{1}{99} \left[3.179 - \frac{553 \times 553}{100} \right] \\
 &= 1,21
 \end{aligned}$$

do đó $s = \sqrt{1,21} = 1,1$

Ta tính tiếp SE:

$$SE = \frac{s}{\sqrt{n}} = \frac{1,1}{10} = 0,11$$

2. Tính các chỉ số trung tâm của các biến liên tục gộp lớp.

2.1. Tính trung bình.

Muốn tính trung bình của một dãy biến liên tục, trước hết phải xây dựng một bảng phân phối tần số:

- Lớp gộp phải hợp lý, tránh để mất những thông tin quan trọng, nhất là đối với những mốc đã hình thành trong y học (mức bình thường, mức bệnh lý, mức dị dạng...) hoặc những tiêu chuẩn đã ấn định (số calo trung bình, số lượng DDT cho $1m^2$...).

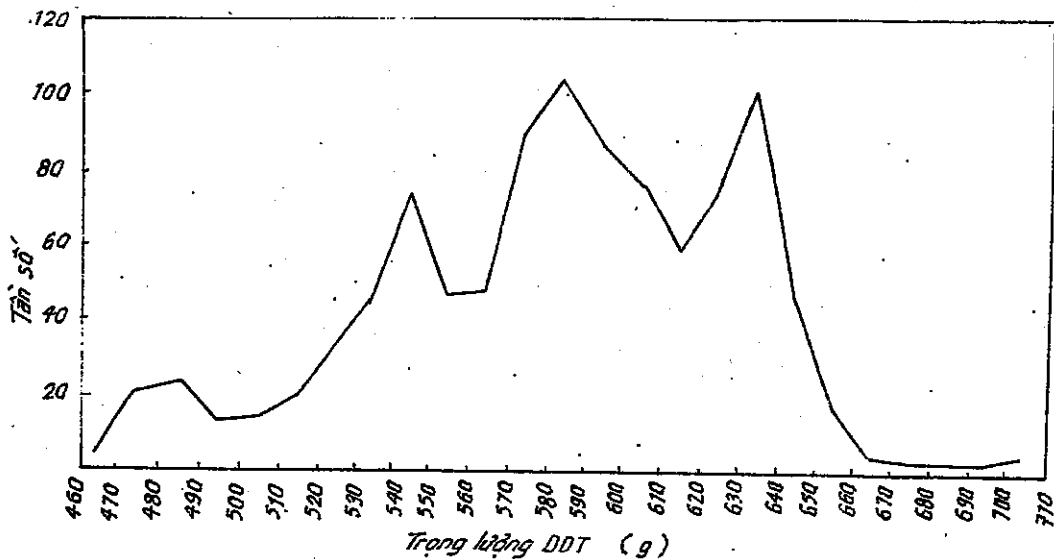
- Số lớp không quá nhiều: người ta cho rằng số lớp nên từ 15 đến 30 là hợp lý nhất.

Thí dụ ta có bảng phân phối trọng lượng DDT đã phun ở các gia đình sau đây:

Bảng 7- BẢNG PHÂN PHỐI TẦN SỐ TRONG LƯỢNG DDT ĐÃ PHUN Ở 1022 GIA ĐÌNH.

Gộp lớp trọng lượng	Trị số giữa lớp x	Khoảng cách lớp đơn vị phụ u	tần số f	uf
g	g			
460-469	465	-11	5	-55
470-479	475	-10	22	-220
480-489	485	-9	24	-216
490-499	495	-8	13	-104
500-509	505	-7	15	-105
510-519	515	-6	20	-120
520-529	525	-5	46	-230
530-539	535	-4	44	-176
540-549	545	-3	73	-219
550-559	555	-2	47	-94
560-569	565	-1	49	-49
570-579	575	0	91	0
580-589	585	1	104	104
590-599	595	2	88	176
600-609	605	3	77	231

610-619	615	4	57	228
620-629	625	5	73	365
630-639	635	6	101	606
640-649	645	7	44	308
650-659	655	8	15	120
660-669	665	9	3	27
670-679	675	10	3	30
680-689	685	11	2	22
690-699	695	12	2	24
700-709	705	13	4	52
Cộng			1.022	705



Hình 6- Đồ thị tần số trọng lượng DDT đã phun

Bảng này được xây dựng, để sử dụng một đơn vị phụ trong việc tính trung bình của dãy số.

Để tính trung bình, ta hãy xem xét một vài lớp xung quanh giá trị nằm ở giữa dãy số để làm gốc. Thí dụ ta lấy giá trị lớp 570-579 với trị số giữa lớp là 575, rồi tính độ lệch của mỗi lớp so với giá trị lớp gốc đó, với đơn vị tính bằng số lớp (mỗi đơn vị lớp tương ứng với 10g - lớp được gộp bằng những khoảng cách đến 10g) và có giữ dấu đại số của u và f. Cuối bảng ta tổng cột 4 để được $\sum uf$ và tổng cột 5 để được $\sum uf$.

Rồi tính trung bình $\frac{\sum uf}{\sum f} = \frac{705}{1022} = 0,69$ khoảng cách lớp.

Từ kết quả trung bình tính bằng khoảng cách lớp, ta tính ra trung bình dãy số bằng số:



Trên đây ta vừa chọn mốc 575g làm gốc, nay ta tính được trung bình tính bằng khoảng cách lớp là +0,69 (tương ứng với $0,69 \times 10g = 6,9g$) nghĩa là trung bình số học của dãy sẽ là:

$$575g + 6,9g = 581,9g$$

- *Mốt.*

Muốn tìm mốt của một dãy biến liên tục có gộp lớp, vẫn căn cứ vào tần số cao nhất. Ở đây, ứng với tần số cao nhất là lớp có trị số giữa lớp 585g.

Nhưng để xem xét rõ hơn, người ta thường vẽ đồ thị đa giác tần số, và dọc trên trục hoành giá trị ở chân đường chiếu của đỉnh đa giác tần số. Nếu đa giác tần số có nhiều đỉnh (như ở hình 6 của bảng 7) ta thấy 3 đỉnh ứng với các giá trị là 545g, 585g và 635g. Trường hợp này người ta thường tìm nguyên nhân của sự phân phối đặc biệt đó, và đôi khi có thể thấy được, như ở đây chính là do sự cân đong không chính xác đã gây ra.

- *Trung vị.*

Ở đây ta cấu tạo 25 lớp. Lớp giữa là lớp thứ 13 ứng với lớp có tần số lớn nhất, nghĩa là trung vị trùng với mốt ở giá trị 585g. Đối với dãy biến liên tục, để cho dễ, người ta vẽ đồ thị tần số dồn, hình sau đây, rồi kẻ một đường ngang từ điểm ứng với một nửa tần số, đường ngang này sẽ gặp đường biểu diễn ở một điểm, từ đó hạ đường vuông góc xuống trục hoành, chân đường vuông góc này xác định trung vị: khoảng 586g.

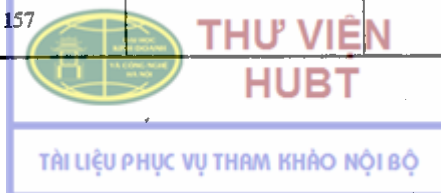
- *Trung bình nhân (TBN).*

Đối với các biến liên tục có gộp lớp, muốn tính TBN, trước hết lấy logarit của các trị số giữa lớp, rồi nhân với tần số của lớp. Lấy tổng của các tích này chia cho tần số tổng cộng ta sẽ được số trung bình của tổng các tích này. Sau đó lấy đối log của số trung bình này ta sẽ được TBN của dãy số.

Ta làm một thí dụ sau đây:

Bảng 8

Số P.vjvax trên 100 bình trường x	Tần số f	Trị số giữa	log x	f log x
1	24	1	0	0
2	26	2	0,3010	7,8260
3-4	27	3,5	0,5441	14,6907
5-8	21	6,5	0,8129	17,0709
9-16	16	12,5	1,0969	17,5504
17-32	12	24,5	1,3892	16,6704
33-60	7	48,5	1,6857	11,7999
65-128	10	96,5	1,9845	19,8450
129-256	5	192,5	2,2844	11,4220
257-512	4	384,5	2,5849	10,3396
513-1024	4	768,5	2,8856	11,5424
1025-2048	1	1536,5	3,1865	3,1965
Cộng	157			141,9438



Ở bảng ta đã có $\sum f \log x = 141,9438$
 $\sum f = 157$

Trung bình tổng các $\log x$ này là:

$$\frac{\sum f \log x}{\sum f} = \frac{141,9438}{157} = 0,9041$$

Vậy TBN của dãy số = đối log 0,9041 = 8,02

2.2. tính các chỉ số phân tán của dãy thống kê biến liên tục gộp lớp.

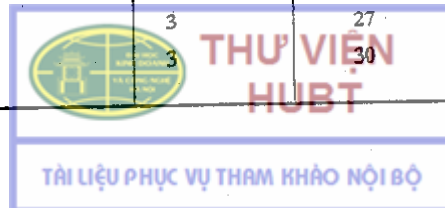
- Tính phương sai và độ lệch chuẩn.

Muốn tính phương sai của các biến liên tục gộp lớp, mỗi lớp gộp sẽ được tính căn cứ vào trị số giữa của lớp đó. Và để đơn giản các phép tính, người ta lấy khoảng cách lớp làm đơn vị phụ để tính, sau đó lại trả lại đơn vị gốc như trên đây đã làm.

Ta trở lại thí dụ ở mục 1B1 bảng 9 sau đây:

Bảng 9

Gộp lớp trọng lượng	Trị số giữa lớp x	Khoảng cách lớp đơn vị phụ u	Tần số f	uf	u ² f
g	g				
460-469	465	-11	5	-55	605
470-479	475	-10	22	-220	2200
480-489	485	-9	24	-216	1944
490-499	495	-8	13	-104	832
500-509	505	-7	15	-105	735
510-519	515	-6	20	-120	720
520-529	525	-5	46	-230	1150
530-539	535	-4	44	-176	704
540-549	545	-3	73	-219	657
550-559	555	-2	47	-94	138
560-569	565	-1	49	-49	49
570-579	575	0	91	0	0
580-589	585	1	104	104	104
590-599	595	2	88	176	352
600-609	605	3	77	231	693
610-619	615	4	57	228	912
620-629	625	5	73	365	1825
630-639	635	6	101	606	3636
640-649	645	7	44	308	2156
650-659	655	8	15	120	960
660-669	665	9	3	27	243
670-679	675	10	3	30	300



680-689	685	11	2	22	242
690-699	695	12	2	24	288
700-709	705	13	4	52	676
Cộng			1022	705	22.171

Ở bảng ta đã có: $\sum f = 1022$ (cột 4)

và: $\sum uf = 705$ (cột 5)

- Bây giờ ta tính $\sum u^2f$ bằng cách đem uf của từng hàng (cột 5) nhân với u ở hàng tương ứng (cột 3), rồi cộng lại sẽ được u^2f (cột 6)

$$\sum u^2f = 22,171$$

- Tính số hiệu chỉnh:

$$\frac{(\sum uf)^2}{\sum f} = \frac{705^2}{1022} = 486,3$$

- Tính các bình phương độ lệch

$$\sum uf^2 - \frac{(\sum uf)^2}{\sum f} = 22.171 - 486,3 = 21.684,7$$

- Tính phương sai: $\frac{21.684,7}{1022} = 21,22$ đơn vị phụ

- Tính độ lệch chuẩn $\sqrt{21,22} = 4,61$ đơn vị phụ

4,61 đơn vị phụ quy về đơn vị gốc là $10 \times 4,61 = 46,1g$ (vì đơn vị phụ ở đây được lấy là khoảng cách lớp, nên mỗi đơn vị phụ tương ứng với một khoảng cách lớp = 10g)

Vậy $s = 46,1g$

Sau khi đã tính độ lệch chuẩn quy về đơn vị gốc, nếu cần ta sẽ tính phương sai V theo đơn vị gốc:

$$V = s^2 = 10^2 \times 21,22 = 2122g \text{ hoặc } = 46,1^2 = 2122g$$

Chú ý là đơn vị phụ được chọn như thế nào cũng không ảnh hưởng gì đến kết quả tính V và s .

3. Tính sai số chuẩn SE.

Cũng như các phân phối tần số khác, phương sai của các số trung bình các mẫu được tính bằng cách đem phương sai của các số hạng của phân phối chia cho cỡ mẫu

$V_{\bar{x}} = \frac{V_x}{n}$ rồi tính tiếp sai số chuẩn $SE_{\bar{x}}$.

$$SE = \sqrt{V_{\bar{x}}} = \sqrt{\frac{V_x}{n}} = \frac{s}{\sqrt{n}} = \frac{46,1}{\sqrt{1022}} = \frac{46,1}{32} = 1,44g$$

C. CÁC DỮ KIẾN ĐỊNH TÍNH

Đối với các dữ kiện định tính thì việc nhận định về độ tập trung và độ phân tán cũng tuân theo những nguyên tắc của các dữ kiện định lượng vừa trình bày trên đây, nhưng đơn giản hơn nhiều vì chúng là một phân phối nhị thức.

1. Bảng phân phối tần số.

Có thể trình bày thành bảng phân phối tần số một chiều hay hai chiều (khi có mô tả với một yếu tố liên quan nào đó).

a) *Bảng phân phối tần số một chiều.*

Số lần tính	Con trai	Con gái	Tỷ suất nam nữ
1000	510	490	51/49

b) *Bảng phân phối hai chiều:*

Giới/bệnh	Cỡ mẫu	Không mắc	Cộng
Nam	31	59	100
Nữ	37	63	100
Cộng	69	132	200

2. Tính chất đặc biệt của phân phối nhị thức.

Phân phối nhị thức có thể đối xứng hoặc không đối xứng.

a) *Phân phối nhị thức không đối xứng:*

Trong một phân phối nhị thức ta phải chú ý đến hai con số:

Tổng số đơn vị thống kê n trong phân phối (cỡ mẫu n trong nghiên cứu mẫu) và tỷ lệ p có đặc trưng nghiên cứu (hoặc tỷ lệ q không có đặc trưng nghiên cứu).

Khi n nhỏ mà lại có p khác q thì phân phối nhị thức không đối xứng.

b) *Phân phối nhị thức sẽ đối xứng khi:*

- $p = q$ dù n bé đi chăng nữa, còn khi

- p khác q thì cần phải có n lớn.

Khi phân phối nhị thức đối xứng, thì phân phối này được coi như xấp xỉ phân phối chuẩn.

3. Tính độ tập trung và độ phân tán của phân phối nhị thức.

Bắt đầu xấp xỉ chuẩn hoặc chắc chắn xấp xỉ chuẩn.

a) *Số trung bình và độ lệch chuẩn của số trung bình:*

"Số trung bình" của tiêu thức nghiên cứu: np , và phương sai bằng npq và độ lệch chuẩn

bằng \sqrt{npq} . Nhưng số trung bình và độ lệch chuẩn này của nó ít được sử dụng trong những nghiên cứu thường gặp trong dịch tễ học.

b) Tần suất trung bình (hoặc tỷ lệ).

Trong hầu hết các nghiên cứu thường gặp trong dịch tễ học là các nghiên cứu về tỷ lệ và ta phải tính tỷ lệ tiêu thức nghiên cứu, cùng với độ lệch chuẩn của tỷ lệ đó.

- Tỷ lệ tiêu thức nghiên cứu: $p = \frac{m}{n}$ (với n là cỡ mẫu, và m là số có mang đặc trưng cần nghiên cứu).

4. Tính sai số chuẩn SE_p .

- Sai số chuẩn của các tỷ lệ các mẫu trong quần thể được tính từ: $V_{\bar{p}} = \frac{V_p}{n}$

$$SE_p = \sqrt{V_{\bar{p}}} = \sqrt{\frac{V_p}{n}}$$

Ở thí dụ này (bảng phân phối một chiều):

- Tỷ lệ sinh con trai

$$p = \frac{510}{1000} = 0,51$$

- Khoảng dao động của p (khoảng tin cậy) ở ngưỡng xác suất 0,05:

$$0,51 \pm 0,032 = \text{từ } 0,478 \text{ đến } 0,542$$

Trên đây ta vừa đề cập rằng phân phối nhị thức chỉ được xử lý kết quả theo các chỉ tiêu kể trên với điều kiện là phân phối nhị thức phải chắc chắn xấp xỉ chuẩn, hoặc ít nhất cũng phải bắt đầu xấp xỉ chuẩn, nếu không các kết quả sẽ không được chấp nhận.

Nói khác đi là ở các nghiên cứu tỷ lệ, phải có mẫu lớn. Ở đây khái niệm về mẫu lớn khác hẳn với mẫu lớn khi nghiên cứu về số trung bình (mẫu được coi là lớn khi $n \geq 30$), mẫu lớn ở các nghiên cứu tỷ lệ không chỉ phụ thuộc vào n , mà phụ thuộc vào cả p và q nữa.

Nói một cách chung, đơn giản nhất thì điều kiện của một mẫu lớn trong nghiên cứu các tỷ lệ là:

n phải lớn hơn 100

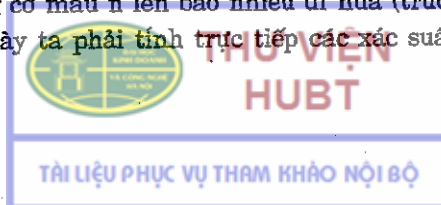
và p (hoặc q) phải lớn hơn 0,10

Hoặc $np = 5$ và $nq = 5$ thì phân phối nhị thức là bắt đầu xấp xỉ chuẩn, và

$np = 10$ và $nq = 10$ thì phân phối nhị thức là chắc chắn xấp xỉ chuẩn.

Cho nên, một cách thực hành ứng dụng, muốn nghiên cứu tỷ lệ, để tính các chỉ số với các khoảng tin cậy của chúng theo luật chuẩn nên áp dụng np hoặc $nq \geq 5$ đến 10 với p (hoặc q) lớn hơn 0,10.

Chú thích: Đối với các mẫu bé, độ chính xác bao giờ cũng rất thấp, và đối với các tỷ lệ nhỏ ($p < 0,1$ hoặc $q < 0,1$) phân phối nhị thức không thể coi được là một phân phối chuẩn, dù ta có tăng cỡ mẫu n lên bao nhiêu đi nữa (trường hợp này nó sẽ gần với phân phối Poisson). Lúc này ta phải tính trực tiếp các xác suất, dài dòng và phức tạp, không trình bày ở đây.



D. Ý NGHĨA VÀ ỨNG DỤNG CỦA ĐỘ LỆCH CHUẨN VÀ SAI SỐ CHUẨN

Người ta sử dụng độ lệch chuẩn để xác định khoảng dao động của các số hạng của dãy thống kê xung quanh số trung bình (hoặc tỷ lệ) của chúng. Còn sai số chuẩn dùng để hình thành khoảng tin cậy của kết quả mẫu số với kết quả quần thể.

1. Độ lệch chuẩn và khoảng dao động.

Độ lệch chuẩn dùng để xác định một khoảng rộng nhất định sao cho đối với một số phần trăm nhất định của trị số quan sát được trên từng đơn vị thống kê, rơi vào trong khoảng đó.

Thí dụ: Khoảng 95% được tính bằng: kết quả nghiên cứu ± 2 SD, có nghĩa là có 95% các số đo quan sát được trên từng đơn vị thống kê sẽ nằm trong khoảng đi từ kết quả nghiên cứu + 2SD đến kết quả nghiên cứu - 2SD.

Như vậy độ lệch chuẩn nói lên độ phân tán của các số đo quan sát: độ lệch chuẩn càng lớn chứng tỏ các số đo quan sát càng phân tán, và ngược lại.

2. Sai số chuẩn và khoảng tin cậy.

Các kết quả từ nghiên cứu mẫu (trung bình, tỷ lệ, tỷ số phần trăm) là những số ước lượng có thể được đối với kết quả ở quần thể. Những ước lượng này có giá trị như thế nào? Ta phải sử dụng khoảng tin cậy:

Khoảng tin cậy là một khoảng mà người ta tin tưởng rằng kết quả thực của quần thể sẽ rơi vào trong khoảng đó ở một mức tin cậy nhất định.

Khi cỡ mẫu là hợp lý, thì khoảng tin cậy CI được tính như sau:

$$CI = \text{Kết quả nghiên cứu mẫu} \times k (SE)$$

Trong đó k là số giới hạn tin cậy chuẩn (k = 2 với mức tin cậy 95%, k = 2,6 với mức tin cậy 99%).

SE là sai số chuẩn của ước lượng kết quả mẫu như đã trình bày:

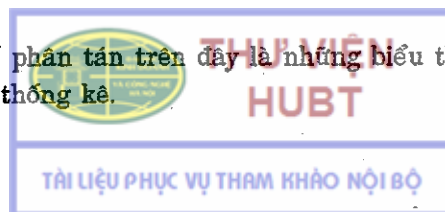
Nói rằng cỡ mẫu là hợp lý là nói đến cỡ mẫu tối thiểu vừa đủ, được tính theo các công thức sát hợp tùy theo loại mẫu, và đặc biệt là tùy theo các ước lượng thống kê mà ta ấn định sử dụng.

Trường hợp không làm được, có thể coi các hướng dẫn sau đây đối với cỡ mẫu là chấp nhận được:

1. Khi ước lượng một số trung bình, cỡ mẫu tối thiểu là 30 trở lên.
2. Khi ước lượng các tỷ lệ nằm trong khoảng 10% đến 90% thì cỡ mẫu phải là 100 trở lên.
3. Trường hợp tỷ lệ với lượng dưới 10% hoặc trên 90% thì cỡ mẫu bắt buộc phải trên 100, tỷ lệ ước lượng càng thấp thì cỡ mẫu càng phải lớn. Đối với các tỷ lệ thấp quá, dưới 1% cần phải hỏi ý kiến một nhà thống kê có kinh nghiệm.

E. HỆ SỐ BIẾN THIÊN (HSBT)

Các chỉ số tập trung và các chỉ số phân tán trên đây là những biểu thị bằng số của tính chất quan trọng của một dãy số thống kê.



Nhưng, độ phân tán của một dãy thống kê chỉ nói lên sự phân tán ở trong dãy đó. Còn khi muốn nhìn xem hai dãy thống kê phân tán khác nhau như thế nào, thì không thể căn cứ vào độ lệch chuẩn của từng dãy một, vì người ta nhận thấy rằng đối với cùng một số đo, thì trong hai dãy thống kê, dãy nào có chỉ số tập trung lớn (số trung bình lớn) sẽ có phương sai lớn.

Cho nên, để so sánh độ phân tán của hai dãy thống kê, người ta xây dựng khái niệm về hệ số biến thiên bằng cách tính độ lệch chuẩn dưới dạng phần trăm đối với số trung bình của dãy đó.

$$\begin{aligned} \text{HSBT} &= \frac{\text{độ lệch chuẩn}}{\text{trung bình}} \times 100 \\ &= \frac{s}{x} \times 100 \end{aligned}$$

rồi mới đem so sánh hệ số biến thiên của hai dãy với nhau: dãy nào có HSBT lớn hơn sẽ là dãy bị phân tán nhiều hơn.

Cần chú ý là các đơn vị của các số đo không ảnh hưởng gì đến hệ số biến thiên cả?

Thí dụ: Khi ta đo chiều cao của người lớn và của trẻ con với cùng một đơn vị đo chiều dài, thì phương sai của dãy thống kê chiều cao người lớn sẽ lớn hơn của dãy thống kê chiều cao trẻ con.

Cũng như khi ta đo chiều dài và chiều rộng của cánh muỗi Anopheles, có kết quả sau:

	Trung bình (mm)	Độ lệch chuẩn (mm)
Chiều dài cánh	2,59	0,208
Chiều rộng cánh	0,622	0,0764

Ta hãy tính HSBT của chiều dài cánh:

$$\frac{0,208}{2,59} \times 100 = 0,083 \times 100 = 8\%$$

Và HSBT của chiều rộng cánh:

$$\frac{0,0764}{0,622} \times 100 = 0,123 \times 100 = 12\%$$

Như vậy độ phân tán của dãy số đo chiều rộng cánh muỗi lớn hơn độ phân tán của dãy số đo chiều dài cánh muỗi với tỷ số: 3/2.

G. PHÂN TÍCH PHƯƠNG SAI

Trên đây chúng ta đã xét về mô tả một dãy thống kê, và so sánh độ phân tán của hai dãy thống kê. Nhưng chúng ta nhiều khi nghiên cứu đến ba hoặc nhiều hơn nữa các dãy thống kê, và lại muốn xem xét xem các dãy thống kê đó là đều được lấy ra từ một quần thể đồng nhất hay không, nếu các dãy thống kê này có những khác biệt nhau có ý nghĩa. Trường hợp này phải dùng kỹ thuật phân tích phương sai, dựa trên nguyên tắc là: nếu

nhiều chuỗi thống kê được nghiên cứu từ cùng một quần thể đồng nhất, thì độ phân tán liên chuỗi sẽ giống như độ phân tán trong mỗi chuỗi.

Thí dụ: người ta muốn phun DDT với định lượng $1,5g/m^2$. Có 5 cán bộ thực hiện phun. Người ta muốn kiểm tra xem họ đã phun như thế nào, bằng cách lấy ngẫu nhiên mỗi người 9 điểm, được kết quả như sau:

Bảng 10- LƯỢNG DDT TÍNH BẰNG g/m^2 /từng người

	1	2	3	4	5
	1,82	1,36	0,90	1,20	1,82
	1,37	1,18	1,42	2,22	2,08
	1,67	1,26	1,34	1,54	1,35
	1,64	1,00	1,04	2,62	1,22
	1,18	1,72	1,60	0,74	1,03
	1,83	0,86	1,70	1,66	1,11
	1,51	0,91	1,05	1,36	2,42
	2,00	0,63	1,73	1,62	2,19
	0,77	1,39	2,10	1,91	2,63
Cộng	13,79	10,34	12,88	14,87	15,85
T.bình	1,53	1,15	1,43	1,65	1,76

Tổng cộng cả 5 người: 67,73

Trung bình cả 45 điểm: 1,505

Vấn đề đặt ra là sự khác biệt trong kết quả lượng DDT/ m^2 phun của 5 người có phải là do ngẫu nhiên hay đó là sự khác biệt có ý nghĩa?

Ta sẽ tính toán để có được bảng sau:

Bảng 10b

Loại khác biệt	Số bậc tự do	Tổng các bình phương	Bình phương trung bình	F
Giữa các người phun (liên chuỗi)	4	1,9825	0,4956	2,31
Giữa các điểm phun của một người (nội chuỗi)	40	8,5994	0,2150	
Cộng	44	10,5819		

Các con số trong bảng 10b được tính như sau:

a) Số bậc tự do:

+ Tổng cộng: Tổng số các số đo -1: $45 - 1 = 44$

+ Giữa các người phun cần so sánh -1: $5-1=4$

+ Đối với mỗi người phun: hiệu 2 số trên: $44-4=40$

b) Tổng số bình phương.

+ Tổng cộng: $\sum x^2 - \frac{(\sum x)^2}{N}$. Ở thí dụ này là 10,5819 (trong đó x là các số đo và N là số tổng cộng các số đo)

+ Giữa các người phun:

$$\frac{\sum X^2}{n} - \frac{(\sum x)^2}{N}$$

trong đó: X là tổng các số đo của mỗi người phun

và n là số lần phun của người phun đó

ở đây:

$$\frac{13,79^2 + 10,34^2 + 12,88^2 + 14,87^2 + 15,85^2}{9} - \frac{67,73^2}{45} = 1,9825$$

+ Đối với mỗi người: Hiệu của hai số vừa tính được:

$$10,5819 - 1,9825 = 8,5994$$

c) Các bình phương trung bình.

Dem tổng các bình phương chia cho số bậc tự do:

+ Giữa các người phun: $1,9825 : 4 = 0,4956$

+ Đối với mỗi người: $8,5994 : 40 = 0,2150$

d) Tỷ số phương sai F.

Dem bình phương trung bình giữa các người phun cho bình phương trung bình của mỗi người phun:

$$0,4956 : 0,2150 = 2,31$$

Sự biến thiên toàn bộ (tổng các bình phương) ở đây là 10,5819 được chia làm hai phần:

1) Tổng các bình phương tương ứng với sự biến thiên riêng của mỗi người phun (biến thiên nội lớp) ở đây là 8,5994 và

2) Tổng các bình phương biểu thị cả các sự khác biệt giữa các người phun và cả phương sai đôi ra trong các kết quả của mỗi người, ở đây là 1,9825.

Lý thuyết thống kê đã chứng minh rằng, nếu không có khác biệt thực sự giữa các kết quả thu được của các người phun khác nhau, thì bình phương trung bình giữa các người phun phải bằng bình phương trung bình của mỗi người phun, và tỷ số của chúng phải xấp xỉ 1

Fisher và Yates đã tính sẵn các giá trị chuẩn của F đối với các mức ý nghĩa khác nhau theo số bài tự do.

Ở thí dụ này $F = 2,31$, nhỏ hơn giá trị F trong bảng với tất cả bậc tự do 4 và 40 ở mức ý nghĩa 5%, do đó ta không thể loại bỏ giả thuyết H_0 (trung bình các kết quả phun của 5 người là khác biệt không có ý nghĩa thống kê).

III. BÀI TẬP

Bài tập 1

Tính giá trị giữa của các dãy số sau (lấy đến một số lẻ thập phân)

Tuổi	(Trả lời)	Kg	(Trả lời)
15-19	17,5	15-19	17
20-24	22,5	20-24	22
25-29	27,5	25-29	17
30-34	32,5	30-34	32
35-39	37,5	35-39	37
40-44	42,5	40-44	42
45-49	-47,5	45-49	47

Bài tập 2

Tính các độ bách phân vị của bảng gộp lớp sau:

Có thể vẽ đồ thị tần suất dồn và lấy các độ bách phân vị, nhưng tính ra sẽ chính xác hơn. Thí dụ ta có cân nặng kni sinh của 10.000 trẻ sơ sinh được gộp lớp như sau:

Cân nặng (g)	Tần số (g)	Cân hiệu là (g)	Tần số dồn
500-800	5	450-849	5
900-1200	40	850-1249	45
1300-1600	102	1250-1649	147
1700-2000	141	1650-2049	268
2100-2400	295	2050-2449	583
2500-2800	929	2450-2849	1512
2900-3200	2321	2850-3249	3833
3300-3600	2826	3250-3649	6659
3700-4000	2125	3650-4049	8784
4100-4400	798	4050-4449	9582
4500-4800	321	4450-4849	9903
4900-5200	76	4850-5249	9979
5300-5600	18	5250-5649	9997
5700-6000	2	5650-6049	9999
6100-6400	1	6050-6449	10.000

Công thức để tính độ BPV

$$P = \left(\frac{p-a}{b-a} \times c \right) + L_b$$

Trong đó: P: Giá trị của độ bách phân vị muốn tính được trong dãy số thống kê.



p: Độ bách phân vị muốn tính x tổng cộng tần số

a: Tần số dồn ngay trước p

b: Tần số dồn ngay sau a

c: Khoảng cách lớp

L_b: Giới hạn dưới của lớp tương ứng với tần số dồn b.

Hãy tính các độ bách phân vị 2,5 và 97,5; 25,50 và 75%

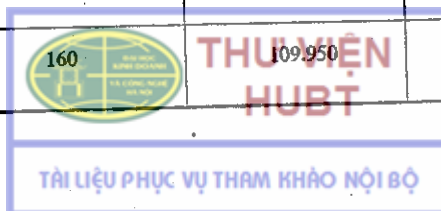
Trả lời: Các độ bách phân vị

thứ	2,5	là	1942g
"	25	"	3020g
"	50	"	3415g
"	75	"	3808g
"	97,5	"	4660g

Bài tập 3

Có bảng phân phối số liệu nghiên cứu sau:

Tiểu thức x	Tần số f	Fx %	Tần suất	Tần suất dồn
605	1	605	0,6	0,6
615	1	615	0,6	1,2
625	1	625	0,6	1,8
635	2	1270	1,2	3,0
645	2	1290	1,2	4,2
650	5	3250	3,1	7,3
655	5	3275	3,1	10,4
660	8	5280	5,0	15,4
665	9	5985	5,6	21,0
670	11	7370	6,9	27,9
675	13	8775	8,1	36,0
680	14	9520	8,8	44,8
685	18	12.330	11,2	56,0
690	13	8970	8,1	64,1
695	10	6950	6,2	70,3
700	9	6.300	5,6	75,9
705	8	5640	5,0	80,9
710	7	4970	4,4	85,3
720	6	4320	3,7	89,0
730	6	4380	3,7	92,7
740	5	3700	3,1	95,8
750	4	3000	2,9	98,7
760	1	760	0,6	99,3
770	1	770	0,6	100,0
Cộng	160	109.950		100,0



Câu hỏi

1. Hãy xác định trung vị, mốt, nhận định kiểu phân phối
2. Hãy tính trung bình, độ lệch chuẩn, sai số chuẩn
3. Vẽ biểu đồ cột, vẽ biểu đồ tần suất dồn và từ đó xác định các bách phân vị 97,5 và 2,5 và tứ phân vị (1,2,3)

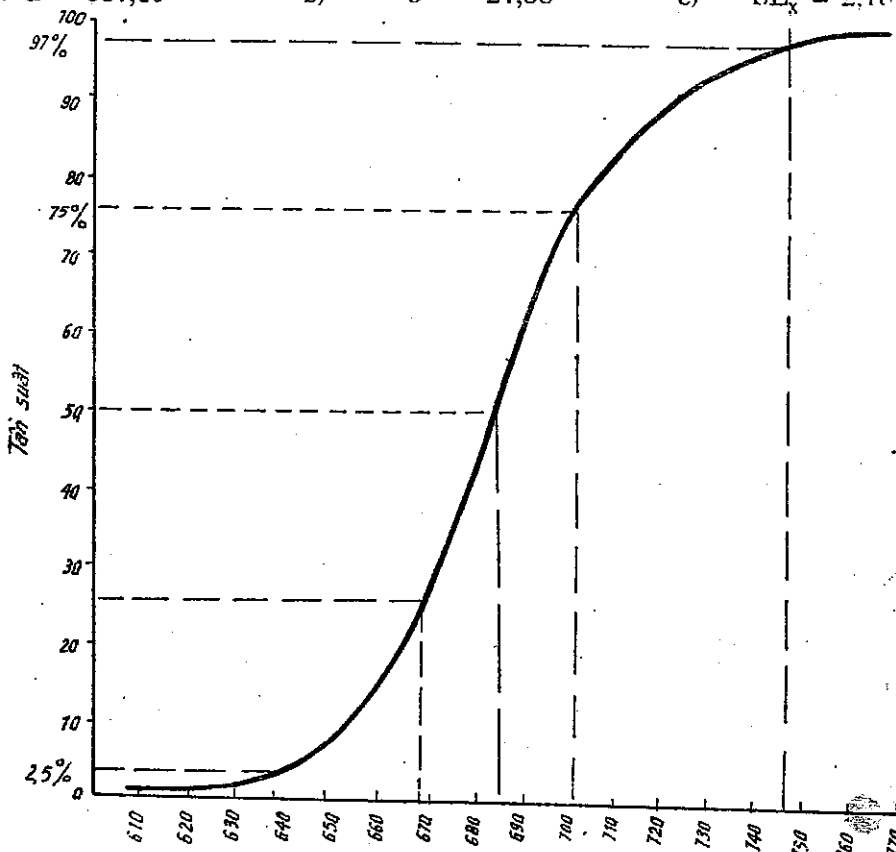
Trả lời:

1- a) Vì tần số là một số chẵn 100 nên trung vị sẽ là trung bình của tần số thứ 50 và 51. Cả hai tần số này đều nằm trong lớp tần suất dồn 56,0. Vậy trung vị là 685

b) Mốt: tần suất cao nhất đây là 11,2: vậy mốt là 685

c) Ta chưa tính trung bình, nhưng nhìn chung vào bảng phân phối, lại có trung vị và mốt trùng nhau: phân phối có nhiều khả năng là phân phối đối xứng.

2. a) $\bar{x} = 687,19$ b) $\sigma = 27,58$ c) $SE_{\bar{x}} = 2,18$



3. b) RPV 1 = 667

TPV 2 = 685

TPV 3 = 710

c) BPV 97,5 : 747

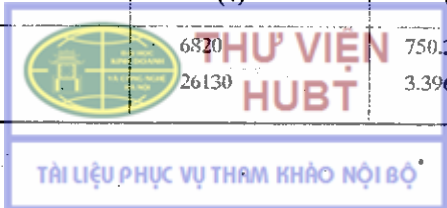
BPV 50 : 685

BPV 2,5 : 635

Bài tập 4

Người ta đã thành lập một bảng gồm những cột sau trong nghiên cứu về huyết áp tâm thu của 448 đàn ông 45-54 tuổi *

HATT(mmHg)	Tần số f	Trị số giữa m	Σx tính bằng m.f	Σx^2 tính bằng m.m.f
(1)	(2)	(3)	(4)	(5)
100-119	62	110	6820	750.200
120-139	201	130	26130	3.396.900



140-159	127	150	19050	2.857.500
160-179	40	170	6800	1.156.000
180-199	14	190	2660	505.400
200-219	3	210	630	132.300
220-239	1	230	230	52.900
Cộng	448		62.320 = $\sum x$	8.851.200 = $\sum x^2$

* Nguồn số liệu: Edinburgh - Fife Heart study

Câu hỏi:

1. Tính huyết áp tâm thu trung bình của họ.
2. Tính độ lệch chuẩn của trung bình đó.
3. Có thể nói gì về HATT của quần thể đàn ông 45-54 tuổi nếu mẫu 448 người này là mẫu đại diện?

Trả lời

1. $m = 139,1$ mmHg $V_x = 19.607$

2. $s = \sqrt{V_x} = 140$ mmHg

3. $SE = 6,62$ mmHg

Nếu mẫu là đại diện thì trung bình huyết áp tâm thu của quần thể đàn ông 45-54 tuổi ở nơi đó, với mức tin cậy 95% là:

$139 \pm 12,98$ mmHg

Bài tập 5

Người ta đã đo độ sâu bằng mét của 10 cái giếng có kết quả sau:

1,4 0,8 3,4 2,7 1,9
3,9 4,2 1,8 2,4 2,1

Câu hỏi

Hãy tính các chỉ số trung tâm và chỉ số phân tán của dãy thống kê này.

Trả lời:

- Độ sâu trung bình của giếng:

$$m = \frac{\sum x}{n} = \frac{24,6}{10} = 2,46 \text{ mét}$$

- Độ lệch chuẩn:

Trước hết, tính phương sai:

$$V = \frac{1}{9} \left[71,2 - \frac{(24,6 \times 24,6)}{10} \right] = 1,2$$

từ đó: $s = \sqrt{1,2} = 1,1$ m.



Ý NGHĨA THỐNG KÊ SAI SỐ CHUẨN - TỰ ĐỐI CHIẾU CẤP

I. MỤC TIÊU

1. Học thuộc được các bước cần tiến hành sau khi sơ bộ xử lý số liệu nghiên cứu.
2. Làm được những trắc nghiệm thống kê phù hợp
3. Hiểu được ý nghĩa của sai số chuẩn (phân biệt được với độ lệch chuẩn) và ứng dụng của nó.
4. Biết sử dụng kỹ thuật tự đối chiếu cấp đúng tình huống nghiên cứu.

II. NỘI DUNG

Chúng ta thường không thể nghiên cứu về một đặc trưng nào đó trên quần thể toàn bộ, mà chỉ nghiên cứu đặc trưng đó trên một mẫu rút từ quần thể đó ra một cách ngẫu nhiên. Nhưng không phải nghiên cứu trên một mẫu để chỉ nói về mẫu đó, mà mong muốn rằng các kết quả nghiên cứu trên mẫu đó có thể suy ra kết quả như khi chúng ta nghiên cứu trên toàn bộ quần thể.

Như vậy, một câu hỏi sẽ được đặt ra là: giả sử mẫu ta lấy ra nghiên cứu là đại diện cho quần thể toàn bộ, thì bằng cách nào để có thể coi các kết quả nghiên cứu mẫu là có giá trị phản ánh kết quả nghiên cứu quần thể toàn bộ?

Cần phải lưu ý đến 2 vấn đề: trước hết là mẫu phải đại diện, thứ hai là tiến hành các thuật toán thống kê để tổng quát hóa cho giá trị quần thể toàn bộ.

* Mẫu phải đại diện quần thể:

Mẫu đại diện quần thể được nói một cách rất đơn giản là về số lượng và về chất lượng, nghĩa là cả cỡ mẫu và chọn mẫu. Cỡ mẫu đã được tính theo những công thức phù hợp với từng loại hình mẫu nghiên cứu (đơn, chùm, tầng...) với những quy định nhất định của chúng (các điều kiện áp dụng để kéo về phân phối xấp xỉ chuẩn). Lý thuyết thống kê đã chứng minh rằng ảnh hưởng của sai số ngẫu nhiên càng giảm thiểu khi cỡ mẫu càng lớn lên. Còn chọn mẫu là bước cực kỳ quan trọng đối với tính đại diện của mẫu: nhà nghiên cứu phải định liệu xem cấu trúc của mẫu có đại diện hay không.

Thí dụ: ta muốn mô tả chỉ số lách trong một vùng lưu hành sốt rét, ở học sinh cấp 2 trong vùng, với cỡ mẫu là 100. Nhưng nếu chỉ lấy mẫu nghiên cứu từ các học sinh có đến

SE: Standard Error.

trường học trong 1 ngày nào đó, mà lại bỏ qua các học sinh nghỉ học (có thể do bị sốt rét mà phải nghỉ học) thì rõ ràng kết quả nghiên cứu mẫu đó không phải là ước lượng của kết quả quần thể học sinh cấp 2 ở vùng này được.

* Khả năng tổng quát hóa kết quả mẫu ra quần thể.

Ngoài việc xác định xem cấu trúc mẫu như thế nào để mẫu là đại diện quần thể (quần thể định danh, quần thể tham khảo), thì kỹ thuật ngẫu nhiên là bắt buộc phải tôn trọng nghiêm ngặt, nếu bất kỳ ở bước nào của nghiên cứu mà tính ngẫu nhiên không được áp dụng đúng thì kết quả nghiên cứu mẫu cũng không thể tổng quát hóa cho quần thể được, vì đã mắc phải những sai số nghiêm trọng. Thí dụ trong điều tra đánh giá tỷ lệ tiêm chủng mở rộng, theo mẫu chùm thì từ việc chọn chùm, chọn ngõ, chọn nhà ngẫu nhiên đến chọn trẻ ngẫu nhiên... đều rất quan trọng không thể sai phạm tính ngẫu nhiên ở bất kỳ một bước nào. Ngay cả việc thay thế một trẻ được chọn ngẫu nhiên (vì cháu đó vắng mặt chẳng hạn) bằng 1 trẻ khác vẫn được chọn ngẫu nhiên sau đó, cũng làm cho kết quả đánh giá là không tin cậy được.

Một điểm quan trọng khác để tổng quát hóa kết quả mẫu cho quần thể là tính sai số chuẩn (SE).

* Sai số chuẩn là một ước lượng thống kê quan trọng. Nó được sử dụng trong việc ước lượng kết quả ở quần thể toàn bộ từ kết quả của mẫu nghiên cứu.

Vì kết quả của mẫu nghiên cứu luôn luôn khó có thể và hầu như không thể chấp hoàn toàn với kết quả của quần thể đã lấy mẫu, mà mục tiêu của nghiên cứu mẫu lại muốn là khái quát ra cho quần thể toàn bộ, nên trong việc này bao giờ cũng phải sử dụng đến sai số chuẩn.

Sai số chuẩn là một số đo xác định một khoảng để cho các ước lượng kết quả mẫu sẽ có thể dao động trong đó đối với các mẫu khác nhau được rút ra từ cùng một quần thể với các cỡ mẫu như nhau. Nói khác đi sai số chuẩn biểu thị sự biến thiên của kết quả nghiên cứu mẫu, nếu nghiên cứu được làm lại nhiều lần.

A. SAI SỐ CHUẨN CỦA SỐ TRUNG BÌNH TRONG CÁC NGHIÊN CỨU MÔ TẢ

Nếu trong một quần thể, ta nghiên cứu 1 mẫu, ta sẽ có được một số trung bình mẫu \bar{x}_1 . Nếu ta nghiên cứu lần lượt nhiều mẫu, ta sẽ có được các số trung bình mẫu $\bar{x}_2, \bar{x}_3, \dots$. Khi ta nghiên cứu được với cùng điều kiện nghiên cứu tất cả T mẫu cùng cỡ: $T = C_n^N$ thì ta sẽ thu được các số trung bình: $\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_T$

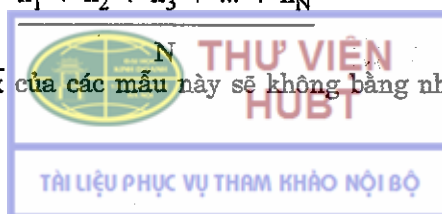
Bây giờ ta tính trung bình \bar{X} của tất cả các số trung bình mẫu \bar{x} kể trên, ta sẽ được một số trung bình khác: trung bình quần thể \bar{X} (mà ta thường gọi là trung bình lý thuyết)

$$\bar{X} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \dots + \bar{x}_T}{T}$$

giống như khi ta làm nghiên cứu trên quần thể toàn bộ gồm N cá thể:

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_N}{N}$$

Các số trung bình \bar{x} của các mẫu này sẽ không bằng nhau mà chúng cũng dao động



xung quanh số trung bình X của chúng với một độ phân tán nhất định: độ lệch chuẩn của tất cả các số trung bình của các mẫu có thể lấy ra được từ một quần thể. Độ lệch chuẩn của tất cả các số trung bình này được gọi dưới một tên riêng là sai số chuẩn (SE).

Chúng ta cũng đã đề cập đến, việc nghiên cứu quần thể toàn bộ thường là không thể, và cũng không cần: ta chỉ làm nghiên cứu 1 mẫu theo đúng nghĩa của nó, rồi tổng quát ra quần thể, nhờ vào sai số chuẩn SE.

Người ta đã chứng minh rằng, nếu trong một mẫu nghiên cứu, phương sai của các số đo cá thể là V , và số trung bình được tính từ n cá thể trong mẫu, thì phương sai của các trung bình mẫu sẽ bằng V/n , và độ lệch chuẩn của các số trung bình mẫu (độ lệch chuẩn của các số trung bình mẫu = sai số chuẩn) sẽ là:

$$SE_x = \sqrt{\frac{V}{n}} = \frac{s}{\sqrt{n}}$$

Và bây giờ từ số trung bình mẫu nghiên cứu \bar{x} , ta có thể xây dựng ước lượng ra trung bình quần thể \bar{X} :

$$\bar{X} = \bar{x} \pm k \left(\frac{s}{n} \right)$$

Tất nhiên, mẫu nghiên cứu phải là đại diện, cả về số lượng và về chất lượng (không có bất kỳ một loại sai số nào, với sai số ngẫu nhiên là tối thiểu).

B. SAI SỐ CHUẨN CỦA TỶ LỆ TRONG CÁC NGHIÊN CỨU MÔ TẢ

Cũng như đối với trung bình, sai số chuẩn của một tỷ lệ cũng có ý nghĩa tương tự (để xây dựng nên khoảng tin cậy).

Ta nhắc lại khoảng tin cậy 95% được xây dựng như sau:

$$CI = \text{ước lượng} \pm 1,96SE \text{ (của ước lượng).}$$

Đây là công thức chung áp dụng cho bất kỳ ước lượng của tham số nào có phân bố chuẩn. Nên đối với các biến nhị thức, có cỡ mẫu n , và ta đếm được m đơn vị có tiêu thức nghiên cứu, nghĩa là tỷ lệ tiêu thức mẫu là p , thì khoảng tin cậy 95% của tiêu thức nghiên cứu là:

$$CI = m \pm 1,96 \sqrt{npq}$$

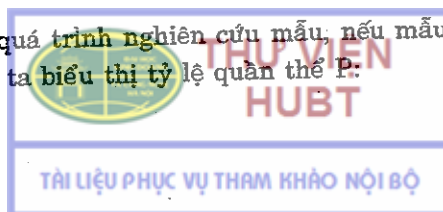
$$= m \pm 1,96 \sqrt{n \left(\frac{m}{n} \right) \left(1 - \frac{m}{n} \right)}$$

Dem chia các giới hạn đó cho n , sẽ được khoảng tin cậy của tỷ lệ.

Nghĩa là sai số chuẩn của tỷ lệ được tính bằng:

$$SE_p = \sqrt{\frac{pq}{n}}$$

Và sau khi đã xem xét toàn bộ quá trình nghiên cứu mẫu, nếu mẫu thực sự đại diện cho quần thể từ đó đã lấy mẫu ra, ta biểu thị tỷ lệ quần thể P :



$$P = p \pm kSE_p$$

$$= p \pm k \sqrt{pq/n}$$

Trong các kết quả này, k là giới hạn tin cậy ứng với các ngưỡng xác suất hay mức tin cậy mong muốn:

$$k = 1,96 \text{ (#2) ở mức tin cậy 95\%}$$

$$k = 2,58 \text{ (# 2,6) ở mức tin cậy 99\%}$$

$$k = 1,65 \text{ ở mức tin cậy 90\%}$$

$$k = 1,28 \text{ (# 1,3) ở mức tin cậy 80\%}$$

$$k = 1 \text{ ở mức tin cậy 68,27\%}$$

$$k = 0,68 \text{ ở mức tin cậy 50\%}$$

Và như vậy, trong các kết luận phiên giải kết quả, kèm theo các kết quả bằng số, bao giờ cũng phải nêu kết quả đó ở ngưỡng xác suất là bao nhiêu (hoặc ở mức tin cậy là bao nhiêu).

C. ƯỚC LƯỢNG SỐ LƯỢNG TUYỆT ĐỐI TOÀN BỘ QUẦN THỂ CÁC TIÊU THỨC NGHIÊN CỨU TỬ NGHIÊN CỨU MÔ TẢ

Trong các nghiên cứu mô tả theo phương pháp dịch tễ học, nhất là trong việc tìm hiểu về các nhu cầu của một cộng đồng, thì cần thiết phải biết đến số lượng tuyệt đối toàn bộ tiêu thức nghiên cứu ở cộng đồng đó. Chẳng hạn như khi muốn lập kế hoạch để trợ giúp cho các dịch vụ y tế thì rất cần biết tổng số dân với các nhóm tuổi có số lượng toàn bộ là bao nhiêu. Vấn đề này thường không đòi hỏi đến một độ chính xác tuyệt đối, nhưng lại cần phản ánh đúng tình trạng thực ở địa phương đó. Thí dụ ta muốn làm kế hoạch cho dịch vụ đỡ đẻ cho 10.000 phụ nữ tuổi 15 - 49 bằng cách đào tạo một số nữ hộ sinh nhất định cho họ, và sẽ không phải điều chỉnh nếu số phụ nữ đó là 10.500, thậm chí 11.000. Nhưng nếu số phụ nữ đó lên tới 20.000 thì sẽ thiếu nữ hộ sinh trầm trọng và chất lượng dịch vụ đỡ đẻ sẽ giảm đi nhiều.

Có nhiều cách ước lượng số toàn bộ, nhưng có 2 phương pháp đặc biệt quan trọng trong công tác điều tra, bằng cách sử dụng phân số lấy mẫu hoặc sử dụng tỷ xuất mẫu.

1. Sử dụng phân số lấy mẫu.

$$\text{Phân số lấy mẫu} = \frac{\text{số đơn vị được chọn vào mẫu}}{\text{tổng số đơn vị của quần thể}}$$

$$\text{Nhu ở mẫu 5\% thì phân số lấy mẫu là} = \frac{5}{100} = \frac{1}{20}$$

Và ước lượng số lượng tuyệt đối toàn bộ (SLTB) sẽ được tính bằng

$$\text{SLTB} = \frac{\text{tần số gặp trong mẫu}}{\text{phân số lấy mẫu}}$$

Thí dụ ở 1 thị xã có 5983 hộ, bây giờ ta nghiên cứu một mẫu hệ thống với khoảng



cách mẫu là 20, ta sẽ có mẫu là 299 hộ (phân số lấy mẫu $299/5983 = 1/20$ số hộ) và ta chỉ điều tra được ở 294 hộ với kết quả như sau:

Nhóm tuổi	Kết quả mẫu		Ước lượng số lượng toàn bộ	
	Nữ	Nam	Nữ	Nam
Dưới 5 tuổi	34	39	692	794
5 - 14	67	70	1365	1426
15 - 19	49	51	998	1039
20 - 24	58	59	1181	1202
25 - 29	44	46	896	937
30 - 34	41	40	835	815
35 - 39	39	35	794	713
40 - 49	67	64	1365	1303
50 - 59	63	57	1283	1161
60 trở lên	226	148	4603	3014
Cộng	688	609	14.012	12.404

Thí dụ đã làm trên đây cho thấy: phân số lấy mẫu đơn vị thống kê $(14.012 + 12.404)/(688 + 609) = 20,37$ hoặc $(688 + 609)/(14.012 + 12.404) = 0,0491 \neq 0,05$ là hơi nhỏ hơn phân số lấy mẫu đã định.

Bây giờ ta lấy phân số $\frac{1}{20,37} = 0,0491$ để tính số toàn bộ cho từng nhóm tuổi:

Nhóm dưới 5 tuổi: $34/0,0491 = 692,45 \approx 692$ nữ. Các nhóm tuổi khác cũng làm tương tự ở cột 4 và 5 bảng trên.

Sau đó nên kiểm tra lại bằng tổng số:

nữ $\frac{688}{0,0491} = 14.012$ và nam $\frac{609}{0,0491} = 12403$ (coi như hoàn toàn phù hợp với tổng ước lượng từng lứa tuổi ở bảng).

2. Sử dụng các tỷ xuất mẫu.

Trong thí dụ trên, nếu ta muốn ước lượng số lượng xe đạp chẳng hạn, giả sử trong 294 hộ điều tra thấy có 213 xe đạp thì tỷ xuất mẫu $= 213/294 = 0,724$ xe đạp mỗi hộ.

Bây giờ ta có thể ước lượng số xe đạp của thị xã đó:

$$0,724 \times 5983 = 4.332 \text{ xe đạp cho toàn thị xã.}$$

Phương pháp ước lượng trên đây, còn có thể áp dụng để ước lượng một vài số liệu khác. Như khi ta muốn ước lượng số người có tuổi 60 trở lên khỏe mạnh có nhu cầu sử dụng nhưng lại không mua được thuốc chẳng hạn. Cách ước lượng như sau:

Vẫn lấy thí dụ trên, trong cuộc điều tra 294 hộ, ta đã gặp 148 cụ ông, trong đó có 26 cụ rất cần một loại thuốc nào đó nhưng lại không mua ở đâu được, và trong số 226 cụ bà thì số đó là 53.

Bây giờ ta tính tỷ suất nhu cầu/cung ứng thuốc riêng cho các cụ ông và các cụ bà:

- Cụ ông: tỷ suất mua được/cần mua là $26 : 148 = 0,176$

- Cụ bà: tỷ suất mua được/cần mua là $53 : 226 = 0,2345$

Và tính ra số lượng toàn bộ cộng đồng:

- Cho các cụ ông: $0,176 \times 3014 = 530$

- Cho các cụ bà: $0,2345 \times 4603 = 1079$

Như vậy số lượng toàn bộ các cụ ông và các cụ bà 60 tuổi trở lên cần nhưng không mua được loại thuốc đó là:

$$530 + 1079 = 1609$$

Con số này sẽ là ước lượng số liều thuốc cần cung ứng thêm trong cộng đồng đó cho các cụ già 60 tuổi trở lên.

Chú ý là số lượng toàn bộ cần cung ứng thêm kể trên không phải là con số chính xác tuyệt đối, vì con số toàn bộ về số các cụ trên 60 tuổi cũng là một con số ước lượng, nhưng sự không chính xác đó không phải là điều cung ứng thuốc loại bất kỳ một dịch vụ nào sát hợp, không thừa không thiếu quá nhiều.

D. ĐỐI VỚI CÁC NGHIÊN CỨU PHÂN TÍCH

Trong các nghiên cứu phân tích bao giờ chúng ta cũng mô tả trên 2 nhóm cá thể rồi đem so sánh với nhau. Để đảm bảo cho các kết quả nghiên cứu là so sánh được trước hết phải làm một trắc nghiệm ý nghĩa thống kê. Chúng ta không nhắc lại ở đây ý nghĩa của một trắc nghiệm ý nghĩa thống kê (đã được trình bày ở sách) mà sẽ vận dụng nó trong các kết quả nghiên cứu y học, với quan niệm rằng trắc nghiệm thống kê là bắt buộc phải làm trước tất cả các công việc xử lý số liệu, nhận định và suy lý thống kê, suy lý nguyên nhân khác.

1. Ý nghĩa thống kê.

Ý nghĩa thống kê được dùng rất chặt chẽ và bắt buộc phải sử dụng trong các nghiên cứu phân tích, mà thực chất thống kê trong các nghiên cứu phân tích là một sự so sánh (so sánh 2 số trung bình đối với các nghiên cứu định lượng và so sánh 2 tỷ lệ đối với các nghiên cứu định tính). Trước khi đem so sánh, người ta phải xem các kết quả nghiên cứu có so sánh được với nhau không: phải có ý nghĩa thống kê mới đủ điều kiện để đem so sánh. Như vậy, sau khi tập hợp được các kết quả của một nghiên cứu phân tích, bao giờ người ta cũng phải tìm cách kiểm định lại giả thuyết về những yếu tố có thể chịu trách nhiệm về sự khác biệt tìm thấy cần đem so sánh. Vì, ngoài yếu tố được nghiên cứu, sự khác biệt đó còn có thể gây nên yếu tố quan trọng khác mà ta không ngờ tới, hoặc do một số yếu tố ít ảnh hưởng tập hợp lại mà ta không thể chú ý đến một cách đầy đủ: đó chính là những tác động của ngẫu nhiên, sự biến thiên ngẫu nhiên hay sai số chọn mẫu. Chúng ta cũng công nhận rằng, ngay cả khi các nhóm nghiên cứu giống nhau về mọi phương diện thì vẫn có thể có những biến thiên nhất định trong các kết quả nghiên cứu, chỉ là do ngẫu nhiên. Nhưng xác suất để cho ngẫu nhiên gây ra sự quan sát khác biệt đó

là bao nhiêu, thì chúng ta cần xác định và có thể xác định được. Một câu hỏi lúc này sẽ được đặt ra như sau: Nếu ta lặp lại nghiên cứu đó rất nhiều lần trong những điều kiện nghiên cứu giống hệt lần đó, thì có ít nhất là bao nhiêu lần thu được sự khác biệt quan sát lớn bằng sự khác biệt quan sát dưới ở lần nghiên cứu thứ nhất đó? Chỉ có lý thuyết toán xác suất mới có câu trả lời này. Nếu tính toán lý thuyết nêu lên rằng sự khác biệt quan sát thu được đó chỉ có thể xảy ra bởi ngẫu nhiên ở 1 lần trong 100 lần, thì người ta có quyền kết luận rằng, trong nghiên cứu đã tiến hành, tính ngẫu nhiên không phải là nguyên nhân duy nhất gây ra sự khác biệt đó. Thực ra, không thể có cách nào biết chắc rằng kết quả quan sát được không phải là do ngẫu nhiên dù chỉ là đối với 1 trường hợp hiếm hoi đó. Nhưng tỷ lệ 1/99, thậm chí 5/95 là rất nhỏ, nên hợp lý nhất là quy sự khác biệt đó cho các nguyên nhân khác hơn là do ngẫu nhiên (và ta loại bỏ giả thuyết H_0).

Và, ngay cả khi đã ấn định mức ý nghĩa của sự khác biệt quan sát được, thì cũng không vội cho rằng yếu tố nghiên cứu là nguyên nhân duy nhất gây ra sự khác biệt đó, nếu không biết được căn cứ tất cả mọi yếu tố ảnh hưởng đến kết quả nghiên cứu. Thật khó có thể hình dung được tất cả mọi yếu tố như vậy, nhưng nếu biết được, càng nhiều càng tốt, để có thể, dưới phép nghi ngờ rằng mẫu đã được cấu trúc không đúng, và để tránh các sai số đó, thì phương pháp duy nhất là thực hiện lấy mẫu một cách ngẫu nhiên theo đúng kỹ thuật lấy ngẫu nhiên quy định. Ngoài ra cũng cần chắc chắn đã loại bỏ được các yếu tố gây nhiễu đầy đủ rồi.

Ý nghĩa thống kê với giá trị của xác suất P:

— Khái niệm ý nghĩa thống kê bao giờ cũng gắn liền với khái niệm xác suất P, nói khác đi khi nêu "có ý nghĩa thống kê hay không có ý nghĩa thống kê bao giờ cũng phải nêu kèm theo là ở một mức ngưỡng xác suất nào đó". Vì mục tiêu của việc tính ý nghĩa thống kê là để hoặc không phủ nhận (không có ý nghĩa thống kê) hoặc phủ nhận (có ý nghĩa thống kê) giả thuyết H_0 . Còn giá trị P là mức ý nghĩa thấp nhất mà người ta có thể phủ nhận giả thuyết H_0 hoặc không phủ nhận nó, nói khác đi giá trị P nói lên khả năng lớn nhất, để quy cho hiện tượng may rủi chịu trách nhiệm về sự khác biệt trong kết quả, nghĩa là sự khác biệt nếu có ấy chỉ là do ngẫu nhiên tối đa ở ngưỡng xác suất đó. Ví dụ khi ta nêu $P < 0,05$ thì có nghĩa rằng chỉ có nhiều nhất là 5% khả năng của sự khác biệt nghiên cứu thấy là do ngẫu nhiên thôi, và như vậy ta phủ nhận giả thuyết H_0 , và sự khác biệt là có ý nghĩa thống kê ở giá trị P này.

Trong việc xử lý các kết quả nghiên cứu về ý nghĩa thống kê có rất nhiều trắc nghiệm có thể làm, chúng ta cần vận dụng đúng cho từng nghiên cứu.

Chúng ta đã biết trong các nghiên cứu phân tích theo phương pháp dịch tễ học, để so sánh 2 nhóm. Chúng ta sẽ xây dựng bảng phân phối tần số 2 chiều (còn gọi là bảng tiếp liên 2×2) gồm 2 hàng và 2 cột hình thành 4 ô, mỗi ô có chứa một trị số, được quy ước là a, b, c, d, với tổng của 4 trị số đó là N.

Điều cần rất phải chú ý là:

1. Các trị số ghi ở trong các ô đó, phải là các số thực = số đếm được trong nghiên cứu (chứ tuyệt đối không được dùng số phần trăm, tỷ lệ loại chỉ số nào đó).

2. Các trị số đó đều là các giá trị độc lập: mẫu được lấy ra ngẫu nhiên, các nhóm chủ

Error

cứu và đối chứng phải được lấy ngẫu nhiên độc lập từ quần thể (chứ không được áp dụng với các tình huống khác nhau trong cùng một nhóm: thí dụ tình trạng sức khỏe, bệnh tật trước và sau một đợt điều trị của cùng một nhóm người bệnh. Trường hợp này phải áp dụng kỹ thuật tự đối chiếu theo phương pháp ghép cặp số liệu: paired t test)

3. Cũng như các so sánh khác: ta đặt giả thuyết H_0 là giả thuyết không có sự khác biệt (trong việc so sánh) và giả thuyết H_1 (gọi là "giả thuyết cơ" hay "giả thuyết đối lập" hoặc "giả thuyết lựa chọn") là giả thuyết ta muốn mong chứng minh là nó có sự khác biệt giữa 2 đại lượng muốn đem so sánh. Và ở một mức ý nghĩa nhất định nào đó, thì kết quả sẽ là "không loại bỏ được giả thuyết H_0 " đồng nghĩa với sự khác biệt không có ý nghĩa thống kê ở ngưỡng xác suất đó, hoặc sẽ là "loại bỏ giả thuyết H_0 " đồng nghĩa với sự khác biệt là có ý nghĩa thống kê ở ngưỡng xác suất đã ổn định.

~~X~~ Sai lầm loại I và sai lầm loại II (α, β)

Trong việc tìm ý nghĩa thống kê đối với các kết quả của một nghiên cứu phân tích so sánh, dù có loại bỏ giả thuyết H_0 , hay không loại bỏ H_0 , thì chúng ta đều có khả năng mắc 2 loại sai lầm: sai lầm loại I (gọi là sai lầm alpha) và sai lầm loại II (sai lầm beta)

+ Sai lầm α là sai lầm mắc phải khi ta loại bỏ giả thuyết H_0 , trong khi trong thực tế thì H_0 là đúng, sai lầm này biểu thị xác suất các kết quả quan sát là có khác biệt nhưng sự khác biệt đó do các yếu tố may rủi gây nên mà ta loại bỏ nó.

+ Sai lầm β , ngược lại là sai lầm mắc phải khi ta không loại bỏ H_0 trong khi trong thực tế H_0 nó không đúng, nói khác đi, sai lầm biểu thị xác suất của sự thất bại trong quá trình tìm lấy sự khác biệt có ý nghĩa trong khi trên thực tế sự khác biệt có ý nghĩa này là có thật.

+ Hai loại sai lầm có quan hệ chặt chẽ với nhau. Và các số bù của nó đến 1 có những ý nghĩa nhất định: $1 - \alpha$ gọi là mức tin cậy của kết hợp, và $1 - \beta$ gọi là lực mẫu. Chúng ta không thể làm nhỏ cùng một lúc 2 loại sai lầm này.

Sau khi đã xem xét đảm bảo những điểm kể trên là đầy đủ và đúng, ta tính các giá trị của các trắc nghiệm ý nghĩa thống kê. Ta có thể dùng:

Cho các biến định lượng.

Trong việc so sánh các biến định lượng (gộp lớp hay không gộp lớp) để tìm ý nghĩa thống kê, người ta dùng trắc nghiệm t:

- Trắc nghiệm t (test Student)

$$t = \frac{\bar{x}_a - \bar{x}_b}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}} \quad (\text{khi } n_a \neq n_b)$$

nếu $n_a = n_b$ thì:

$$t = \frac{\bar{x}_a - \bar{x}_b}{\sqrt{\frac{s_a^2 + s_b^2}{n}}}$$

trong đó

\bar{x}_a là trung bình của chuỗi A



s_a^2 là phương sai của chuỗi A

n_a là cỡ mẫu của chuỗi A

\bar{x}_b là trung bình của chuỗi B

s_b^2 là phương sai của chuỗi B

n_b là cỡ mẫu của chuỗi B

Điều kiện: mỗi chuỗi phải có ít nhất 30 đơn vị thống kê đối với mẫu ngẫu nhiên đơn trong một quần thể vô hạn. Đối với các loại hình mẫu khác, việc tính trung bình \bar{x} và phương sai s^2 cũng như cỡ mẫu n phải tuân theo các loại hình của mẫu đó. Thắc nghiệm này cũng được áp dụng khi ta rút 2 nhóm trong cùng 1 quần thể để nghiên cứu.

Ta sẽ đem giá trị t tính được so với giá trị t tính sẵn trong bảng t , nếu giá trị t tính được lớn hơn t trong bảng với số bậc tự do tương ứng, thì được kết luận là "sự khác biệt là có ý nghĩa thống kê ở ngưỡng xác suất đã chọn" và ta loại bỏ giả thuyết H_0 .

~~X~~ Thắc nghiệm t kể trên còn có thể tính qua sai số chuẩn.

Nếu trong nghiên cứu có tính sẵn SEP thì có thể sử dụng giá trị của SEP này để tính t , rồi đối chiếu với giá trị t trong bảng như trên. Cách tính:

$$t = \frac{\bar{x}_a - \bar{x}_b}{SE_p} \quad (\text{khi } n_a = n_b)$$

$$t = \frac{\bar{x}_a - \bar{x}_b}{\sqrt{\frac{n_q + n_b}{(n_a + n_b - 2)n_a n_b} [(n_a - 1) V_a + (n - 1) V]}}$$

(khi $n_a \neq n_b$)

Chú ý: số bậc tự do ở đây được tính là $(n_a + n_b) - 2$

Cho các biến định tính.

Có thể dùng trắc nghiệm t cho các tỷ lệ, hoặc trắc nghiệm sử dụng sai số chuẩn của khác biệt, nhưng thường dùng trắc nghiệm χ^2 . Tất nhiên các điều kiện nghiên cứu đã được xem xét đúng như trên đây sẽ đề cập. Và từ kết quả ghi trong bảng tiếp liên 2×2 , ta tính:

Trắc nghiệm χ^2 .

Có nhiều công thức để tính χ^2 tùy thuộc vào kết quả nghiên cứu được, bằng cách nhìn vào giá trị từng ô của bảng 2×2 kể trên:

Nếu không có một trị số nào nhỏ dưới 50:

$$\chi^2 = \frac{(ab - bc)^2 (a + b + c + d)}{(a + b)(c + d)(a + c)(b + d)}$$

Nếu có một trong 4 trị số nhỏ hơn 50:

$$\chi^2 = \frac{(|ad - bc| - \frac{a + b + c + d}{4})^2 (a + b + c + d)}{(a + b)(c + d)(a + c)(b + d)}$$

Nếu không có một trị số nào nhỏ hơn 3:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Chúng ta đem χ^2 tính được với giá trị χ^2 trong bảng tính sẵn nếu χ^2 tính được lớn hơn χ^2 bảng thì kết quả nghiên cứu là "khác biệt có ý nghĩa thống kê ở ngưỡng xác suất đã ấn định" với số bậc tự do bằng 1, và ta loại bỏ giả thuyết H_0 .

Sai số chuẩn của sự khác biệt SE_{Δ} .

Tính:

$$SE_{\Delta} = \sqrt{\frac{p_a q_a}{n_a} + \frac{p_b q_b}{n_b}}$$

Rồi lấy $\Delta = p_2 - p_1$ chia cho SE sẽ được 1 giá trị t' . Đem t' so với biến chuẩn tắc t : nếu $t' > t$ ở ngưỡng xác suất nào đó, thì ta sẽ có thể kết luận khác biệt là có ý nghĩa thống kê ở ngưỡng xác suất ấn định, và ta loại bỏ giả thuyết H_0 (như ở ngưỡng xác suất 0,05 thì $t = 1,96$, nếu $t' > 1,96$ thì là khác biệt có ý nghĩa thống kê ở ngưỡng xác suất 0,05).

Sử dụng trắc nghiệm t .

Tính:

$$t = \frac{|p_a - p_b|}{\sqrt{\frac{pq}{n_a} + \frac{pq}{n_b}}}$$

trong đó: p_a và p_b là tỷ lệ tìm thấy ở nhóm A và B

$$p = \frac{p_a + p_b}{2}$$

n_a và n_b là cỡ mẫu của nhóm A và B

Nếu $n_a = n_b = n$ thì:

$$t = \frac{|p_a - p_b|}{\sqrt{\frac{2pq}{n}}}$$

Và ta cũng nhận định và biện luận giống như đối với test t với số trung bình đã trình bày trên đây.

2. Suy lý căn nguyên.

Sau khi đã xác định kết quả nghiên cứu phân tích bằng trắc nghiệm ý nghĩa thống kê, thì dù "sự khác biệt có ý nghĩa thống kê ở ngưỡng xác suất nhất định nào đó" thì cũng chỉ đưa chúng ta đến nhận định là "Kết hợp đó có ý nghĩa thống kê" mà thôi, chứ chưa thể nhận định rằng đó là một mối kết hợp nhân quả (nghĩa là không được vội vàng kết luận rằng chính cái nguy cơ chúng ta nghiên cứu đã gây ra bệnh nghiên cứu đó) mà phải qua 5 bước suy lý căn nguyên. Vấn đề này đã được trình bày rõ ở sách giáo khoa (sách

đã dẫn) và ở bài kết hợp và cân nguyên trong cuốn thực hành này.

E. TRẮC NGHIỆM T CHO TỪNG CẶP TỰ ĐỐI CHIẾU

Tất cả những trắc nghiệm trình bày trên đây là để sử dụng cho việc so sánh các kết quả của 2 nhóm được rút ra độc lập trong cùng 1 quần thể, hoặc từ 2 quần thể khác nhau.

Còn trong trường hợp ta chỉ làm nghiên cứu trên 1 nhóm có thể được rút ra từ 1 quần thể nào đó, rồi ra tiến hành một can thiệp nhất định, và để đánh giá hiệu quả của can thiệp này, ta thường so sánh về một tiêu thức nào đó TRƯỚC và SAU khi áp dụng can thiệp của chúng ta, nghĩa là trên mỗi cá thể dự nghiên cứu, chúng ta có một cặp số liệu trước và sau cuộc nghiên cứu.

Trong những nghiên cứu như vậy, về cỡ mẫu và chọn mẫu, chúng ta cũng phải tuân thủ như tất cả các nghiên cứu theo mẫu khác. Thí dụ chúng ta định đánh giá hiệu quả của 1 chế độ ăn đặc biệt, hoặc 1 chế độ tập luyện đặc biệt nào đó để cho trẻ em tăng cân hoặc để cho phụ nữ giảm béo, thì tiêu thức nghiên cứu, với mẫu ngẫu nhiên đơn chẳng hạn. Chúng ta, sẽ cân trước khi áp dụng biện pháp can thiệp: sẽ được kết quả trước khi nghiên cứu

$$\bar{x}_1 \quad \text{và} \quad \sigma_1^2$$

và cân sau khi áp dụng biện pháp can thiệp: sẽ được kết quả sau khi nghiên cứu:

$$\bar{x}_2 \quad \text{và} \quad \sigma_2^2$$

Ta sẽ so sánh 2 số trung bình quan sát này bằng test t, giống như đối với các so sánh khác:

$$t = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{\frac{\sigma_2^2 + \sigma_1^2}{n}}} \quad (\text{khi } n_1 = n_2 = n)$$

Cũng đem so sánh t tính được so với t tra bảng và sẽ kết luận "sự khác biệt là có hoặc không có ý nghĩa ở 1 ngưỡng xác suất ấn định nào đó" và ta sẽ loại bỏ hoặc chấp nhận giả thuyết H_0 .

Chúng ta cũng còn có thể so sánh, vẫn dùng trắc nghiệm t, nhưng theo một cách khác: vì là số liệu từng cặp ở cùng 1 cá thể, ta sẽ lấy khác biệt d của từng cá thể đó, có giữ nguyên dấu đại số (nếu tăng cân thì giá trị d mang dấu +, ngược lại là mang dấu -) rồi ta tính tổng đại số của d này, ta sẽ được: $\sum d$.

Và từ $\sum d$ này ta tính $\bar{x}_d = \frac{\sum d}{n}$ và $\sigma_d^2 = \frac{1}{n-1} [\sum d^2 - \frac{(\sum d)^2}{n}]$. Đồng thời ta cũng làm một cột d^2 cho từng cá thể, rồi cộng lại ta sẽ được $\sum d^2$ để tính σ_d^2 ở trên.

Giả thuyết H_0 lúc này sẽ là "không có sự tăng cân nặng" thì giá trị lý thuyết sẽ bằng số 0, và công thức áp dụng sẽ là:

$$t = \frac{\bar{x}_d - 0}{\frac{\sigma_d}{\sqrt{n}}}$$



nghĩa là ta sẽ áp dụng công thức:

$$t = \frac{\bar{x}_d}{\left(\frac{\sigma_d}{\sqrt{n}}\right)}$$

trong đó: \bar{x}_d là trung bình của các khác biệt từng cặp đại số số liệu
 σ_d là độ lệch chuẩn của các khác biệt này
 n là số cá thể quan sát (2 lần trước và sau)

Cũng đem so sánh t tính được với t tra bảng và đi đến kết luận theo quy ước.

Bài tập 1.

Nhân dịp có 4 trường hợp chết vì ung thư (K) gan ở nhà máy nhựa polyvinyl chloride, người ta đã làm 1 nghiên cứu về phân tích tỷ lệ chết chung từ 1947 đến 1974 trong số công nhân nhà máy nhựa polyvinyl chloride, có kết quả như sau:

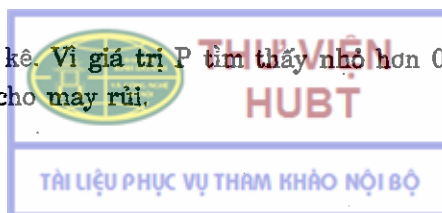
Nguyên nhân chết	Tần số quan sát	Tần số lý thuyết	QS/LT
Mọi nguyên nhân	161	161,0	1,0
Ung thư các loại	41	27,9	1,5
<i>Trong đó:</i>			
tiêu hóa	13	8,3	1,6
gan mật	8	0,7	11,0
phổi	13	7,9	1,6
não	5	1,2	4,2
hạch huyết và tạo huyết	5	3,4	1,5
các loại K khác	5	7,1	0,7
Mạch máu não	8	9,5	0,8
Tim mạch	66	68,6	1,0
Ngoại lai	22	24,3	0,9
Các nguyên nhân khác	24	30,5	0,8

Câu hỏi.

- Số chết vì K có ý nghĩa thống kê không? Tại sao?
- χ^2 đối với bệnh mạch máu não tính được là 0,04 (lbsd) điều đó có nghĩa như thế nào?
- Có sự khác biệt giữa chết vì ung thư và chết chung không? Nếu có thì có thể nhận định sự khác biệt đó thế nào?
- Những khó khăn chủ yếu nào sẽ gặp phải trong việc phân tích tỷ lệ chết trong việc muốn khám phá ra nguyên nhân tiềm ẩn gây ung thư của nhựa polyvinyl chloride?

Trả lời:

- Có ý nghĩa thống kê. Vì giá trị P tìm thấy nhỏ hơn 0,01, điều đó có nghĩa là xác suất là rất nhỏ để quy cho may rủi.



2. $\chi^2 = 0,04$ là rất nhỏ, giá trị P sẽ nhỏ hơn 0,05 thường dùng trong ý nghĩa thống kê.

3. Ta hãy lập 1 bảng phụ cho dễ nhìn nhận:

Nguyên nhân chết	Tần số quan sát	Tần số lý thuyết
Ung thư	41	27,9
Các nguyên nhân khác	120	133,1

rồi tính χ^2 của bảng này:

$$\chi^2_{\text{btd}} = \frac{(41 - 27,9)^2}{27,9} + \frac{(120 - 133,1)^2}{133,1} = 7,44$$

χ^2 tính được lớn hơn χ^2 trong bảng ở P = 0,05 với 1 bậc tự do nên chết vì ung thư là khác biệt có ý nghĩa so với chết chung ở ngưỡng xác suất 0,05.

4. Ở đây có 2 vấn đề nổi bật:

a) Phân tích tỷ lệ chết không đề cập đến nguy cơ tuyệt đối của cái chết toàn quần thể. Nên có thể rằng tỷ lệ chết trong số công nhân V.C thấp hơn ở quần thể nam da trắng nước Mỹ, ngay cả khi có sự không cân xứng với số chết ung thư ở nhóm này.

b) Tỷ suất cao giữa tần số chết quan sát và tần số chết lý thuyết là có thể do hoặc có 1 nguyên nhân chết nào đó hoặc có 1 sự thiếu hụt một nguyên nhân khác.

Bài tập 2.

Bài tập tự đối chiếu cặp.

Pearson và Leer đã so sánh chiều cao của thế hệ bố (1078 người) và thế hệ con (1078 người). Kết quả như sau:

Thế hệ bố: $\bar{x}_1 = 67,71$ inch $\sigma_1^2 = 7,40$

Thế hệ con: $\bar{x}_2 = 68,67$ inch $\sigma_2^2 = 7,60$

Câu hỏi:

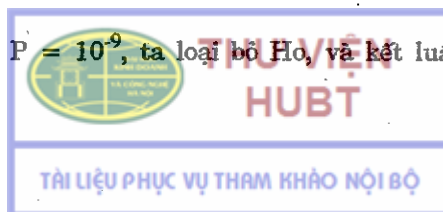
Thế hệ con có thực sự tăng chiều cao so với thế hệ bố?

Trả lời:

Áp dụng:
$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{n}}}$$

Tính ra:
$$t = \frac{67,71 - 68,67}{\sqrt{\frac{7,4 + 7,6}{1078}}} = \frac{0,96}{0,12} = 8$$

So với t bảng thấy $t = 6,1$ ứng với P = 10^{-9} , ta loại bỏ H_0 , và kết luận thế hệ con thực sự cao hơn thế hệ bố.



XÁC SUẤT

I. MỤC TIÊU

Làm được và vận dụng được các phép cộng xác suất và nhân xác suất.

II. NỘI DUNG

Xác suất là gì?

Xác suất Pr cho 1 sự kiện A xảy ra là biểu hiện số lượng của sự xuất hiện sự kiện đó.

Xác suất được định nghĩa là tần suất tương đối được tính theo công thức:

$$\Pr(A) = \frac{\text{Số lần xuất hiện sự kiện A}}{\text{Số lần có khả năng xuất hiện sự kiện A}}$$

*55' lần xuất hiện
nguy cơ.*

Thí dụ:

Trong một vụ dịch ngộ độc thức ăn, 99 trường hợp đã mắc bệnh trên 158 dự tiệc

Số lần xuất hiện sự kiện A = 99

Số lần có khả năng xuất hiện sự kiện A = số lần có nguy cơ xuất hiện sự kiện A = 158

$$\Pr(\text{mắc bệnh}) = \frac{99}{158} = 0.63 \text{ hoặc } 63\%$$

Xác suất có thể được biểu thị bằng một phân số thập phân hoặc bằng tỷ lệ %. Dưới dạng phân số thập phân, xác suất thay đổi trong phạm vi 0 - 1.

Như vậy:

$$\Pr(\text{không xảy ra A}) = 1 - \Pr(\text{xảy ra A})$$

Xác suất có điều kiện.

Xác suất mắc bệnh sẽ thay đổi nếu bệnh xảy ra gắn với một loại thức ăn nào đó. Trường hợp này được gọi là xác suất có điều kiện. Xác suất có điều kiện cho sự kiện A xảy ra gắn với sự kiện B được định nghĩa như sau:

$$\Pr(A/B) = \frac{\text{Số lần A và B cùng xảy ra}}{\text{Số lần B xảy ra}}$$

Thí dụ:

Trong bữa tiệc có 133 người ăn thịt gà trong đó 97 người ăn thịt gà bị ngộ độc

Số lần đồng thời xảy ra A và B = 97

Số lần xảy ra B = 133

$$\Pr(\text{bị bệnh, và có ăn thịt gà}) = \frac{97}{133} = 0,73 \text{ hay } 73\%$$

Hai sự kiện A và B độc lập với nhau có nghĩa là khi A xuất hiện không ảnh hưởng đến sự xuất hiện của B, khi đó.

$$\Pr(AB) = \Pr(A)$$

Xác suất của sự kiện phức hợp.

Sự kiện được coi là phức hợp khi chúng xảy ra đồng thời (A và B) hay khi chúng xảy ra thay thế (A hoặc B)

$\Pr(A \text{ và } B)$ xác suất khi A và B cùng xuất hiện

Khi A và B không thể cùng xuất hiện thì chúng triệt tiêu nhau, và $\Pr(A \text{ và } B) = 0$

$\Pr(A \text{ hoặc } B) =$ xác suất khi có hoặc A, hoặc B, hoặc cả A và B cùng xuất hiện.

Có phép tính giúp ta xử lý dễ dàng hơn các sự kiện phức hợp, đó là:

- Phép nhân xác suất
- Phép cộng xác suất

Phép nhân xác suất.

Phép nhân có thể được phát biểu như sau:

$$\Pr(A \text{ và } B) = \Pr(A/B) \cdot \Pr(B)$$

Trong trường hợp A và B độc lập với nhau, ta có:

$$\Pr(A \text{ và } B) = \Pr(A) \cdot \Pr(B)$$

Thí dụ:

10% bệnh nhân khi uống thuốc có triệu chứng phụ. Có 2 bệnh nhân uống thuốc. Hỏi xác suất khiến cho cả 2 bệnh nhân uống thuốc cùng có phản ứng phụ là bao nhiêu?

Ở đây, ta coi rằng 2 sự kiện xảy ra độc lập với nhau. Điều này có nghĩa là việc xảy ra phản ứng phụ ở bệnh nhân này không liên quan đến việc xảy ra phản ứng phụ ở bệnh nhân khác, vậy:

$$\Pr(\text{cả 2 người uống thuốc đều có phản ứng phụ}) = 0,1 \times 0,1 = 0,01 \text{ hoặc } 1\%$$

Phép cộng xác suất.

Phép cộng được phát biểu như sau:

$$\Pr(A \text{ hoặc } B) = \Pr(A) + \Pr(B) - \Pr(A \text{ và } B)$$

Khi A và B triệt tiêu lẫn nhau, ta có:

$$\Pr(A \text{ hoặc } B) = \Pr(A) + \Pr(B)$$

Thí dụ:

Xác suất khi có ít ra 1 trong 2 bệnh nhân uống thuốc có phản ứng phụ.

Trong trường hợp này, A và B không triệt tiêu nhau, vì vậy ta có $\Pr(\text{ít nhất 1 trong 2 bệnh nhân uống thuốc có phản ứng phụ}) = 0,1 + (0,1 - 0,01) = 0,19 \text{ hoặc } 19\%$.



THƯ VIỆN
HUBT

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

III. BÀI TẬP

Bài tập 1.

Nồng độ Axit Uric trong huyết thanh của 267 người cho máu được sắp xếp theo 12 nhóm nồng độ 3,2-3,4, 3,5-3,9, 4,0-4,4, 4,5-4,9, 5,0-5,4, 5,5-5,9, 6,0-6,4, 6,5-6,9, 7,0-7,4, 7,5-7,9, 8,0-8,4 8,5-8,9; tỷ lệ % và tỷ lệ cộng dồn như trong bảng sau:

Nồng độ axit uric mg/dl	Số người trong từng nhóm nồng độ	Tỷ lệ % so với tổng số 267 người	Cộng dồn tỷ lệ %
3,0-3,4	2	0,8	0,8
3,5-3,9	15	5,6	6,4
4,0-4,4	33	12,4	18,7
4,5-4,9	40	15,0	33,7
5,0-5,4	54	20,2	53,9
5,5-5,9	47	17,6	71,5
6,0-6,4	38	14,2	85,8
6,5-6,9	16	6,0	91,8
7,0-7,4	15	5,6	97,4
7,5-7,9	3	1,1	98,5
8,0-8,4	1	0,4	98,9
8,5-8,9	3	1,1	100

31,5
18,2
52,8

Câu hỏi.

- Một người có nồng độ Urê huyết thanh 4,0-5,9mg/dl có xác suất mạnh khỏe là bao nhiêu %?
- Một người có nồng độ Urê huyết thanh <4,0mg/dl có xác suất mạnh khỏe là bao nhiêu %?
- Một người có nồng độ Urê huyết thanh <4,0mg/dl hoặc >5,9mg/dl có xác suất mạnh khỏe là bao nhiêu?

Trả lời.

- 0,652 hoặc 65,2%
- 0,064 hoặc 6,4%
- 0,349 hoặc 34,9%

Bài tập 2.

Nghiên cứu triệu chứng bệnh lý phụ tạng (đau ngực, nấc, điện tâm đồ bất thường) ở 306 người mới phát hiện có tăng huyết áp (phân loại theo mức nặng nhẹ) cho các số liệu theo bảng sau đây:



		Phân loại huyết áp nặng nhẹ		Tổng số
		Nhẹ và trung bình	Nặng	
Tiền sử đau ngực	Có	18	7	25
	Không	243	38	281
	Tổng số	261	45	306
Tiền sử nấc	Có	4	1	5
	Không	257	41	301
	Tổng số	261	45	306
Điện tâm đồ bất thường	Có	56	22	78
	Không	205	23	228
	Tổng số	261	45	306

Câu hỏi.

- (1) Xác suất một bệnh nhân mới mắc huyết áp cao đến bệnh viện có tiền sử đau ngực là bao nhiêu?
- (2) Cho rằng bệnh nhân mắc huyết áp nặng, xác suất có tiền sử đau ngực là bao nhiêu?
- (3) xác suất một bệnh nhân mới mắc huyết áp cao đến bệnh viện có điện tâm đồ bình thường là bao nhiêu?
- (4) Giả thiết một bệnh nhân huyết áp cao đến khám thấy điện tâm đồ bình thường, hỏi khả năng bệnh nhân có huyết áp nặng là bao nhiêu?
- (5) Xác suất bệnh nhân huyết áp cao đến khám đồng thời có tiền sử đau ngực và điện tâm đồ bình thường là bao nhiêu?
- (6) Xác suất bệnh nhân huyết áp cao đến khám đồng thời có tiền sử đau ngực và điện tâm đồ không bình thường là bao nhiêu?
- (7) 2 bệnh nhân cao huyết áp cùng đến bệnh viện một ngày, hỏi cả 2 cùng có điện tâm đồ bất thường là bao nhiêu?

Trả lời.

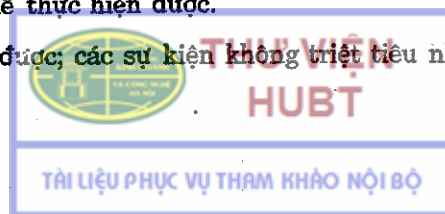
- (1) $25/306=0,082$ hoặc 8,2%
- (2) $7/45=0,156$ hoặc 15,6%
- (3) $228/306=0,745$ hoặc 74,5%
- (4) $23/228=0,101$ hoặc 10,1%
- (5) Người ta không xác định được điều này vì chúng không độc lập với nhau, và do đó:

$Pr(\text{đau ngực và điện tâm đồ bình thường})$ khác

$Pr(\text{đau ngực}) \times Pr(\text{điện tâm đồ bình thường})$

Nếu tính được $Pr(\text{đau ngực điện tâm đồ bình thường})$, ta có thể tính được xác suất cho câu trả lời, tuy nhiên điều này không thể thực hiện được.

- (6) Điều này cũng không thể trả lời được; các sự kiện không triệt tiêu nhau, ta phải



Specificity - Sensitivity

sử dụng.

$\Pr(\text{đau ngực hoặc điện tâm đồ bất thường}) = \Pr(\text{đau ngực}) + \Pr(\text{điện tâm đồ bất thường}) - \Pr(\text{đau ngực và điện tâm đồ bất thường})$; ta không thể có được vế sau cùng của công thức, vì vậy cũng không có đáp số cho câu hỏi này.

(7) Trong trường hợp này, các sự kiện độc lập với nhau, vì vậy có thể trả lời như sau:

$$\frac{78}{306} \times \frac{78}{306} = 0,065 \text{ hoặc } 6,5\%$$

KỸ THUẬT SÀNG TUYỂN PHÁT HIỆN HÀNG LOẠT

I. MỤC TIÊU

1. Phân biệt được nghiệm pháp sàng tuyển với các phương pháp chẩn đoán lâm sàng tại bệnh viện.
2. Nêu được các tiêu chuẩn của những bệnh cần sàng tuyển và những bệnh không thể sàng tuyển.
3. Tính được độ nhạy, độ đặc hiệu, giá trị tiên đoán dương tính của một nghiệm pháp sàng tuyển và ứng dụng của chúng trong những tình huống đúng.

II. NỘI DUNG

Định nghĩa.

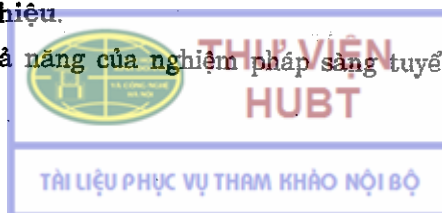
Sàng tuyển là tiến hành một nghiệm pháp phát hiện hàng loạt trong cộng đồng nhằm tìm ra những người có xác suất mắc bệnh cao trong những người có vẻ bình thường, để tiến tới chẩn đoán xem họ đã mắc bệnh chưa, và nếu như chẩn đoán thấy họ đã mắc bệnh, thì chữa bệnh cho họ.

Nói chung, nghiệm pháp sàng tuyển chỉ được thực hiện khi có những tình huống sau đây:

1. Bệnh có tỷ lệ mắc và tỷ lệ tử vong cao.
2. Có kỹ thuật phát hiện bệnh ở giai đoạn rất sớm và được cộng đồng chấp nhận.
3. Khi phát hiện được bệnh thì có biện pháp phòng và chữa hiệu quả.

Độ nhạy và độ đặc hiệu.

Người ta đánh giá khả năng của nghiệm pháp sàng tuyển bằng cách so sánh 2 kết



quả, một của nghiệm pháp sàng tuyển này và một của phương pháp chẩn đoán chính xác đang được chấp nhận nhất (nghiệm pháp chẩn đoán chuẩn).

Có 2 chỉ số được sử dụng để so sánh. Đó là độ nhạy và độ đặc hiệu.

"Độ nhạy" là khả năng của nghiệm pháp sàng tuyển cho kết quả chẩn đoán dương tính khi bệnh nhân thực sự có bệnh; "độ nhạy" được định nghĩa là:

$$\frac{\text{Số người mắc bệnh mà có kết quả nghiệm pháp sàng tuyển dương tính}}{\text{Số người mắc bệnh được sàng tuyển}} \times 100$$

"Độ đặc hiệu" là khả năng của nghiệm pháp sàng lọc cho kết quả âm tính khi người được sàng lọc chắc chắn không có bệnh; "độ đặc hiệu" được định nghĩa là:

$$\frac{\text{Số người không mắc bệnh mà có kết quả nghiệm pháp sàng tuyển âm tính}}{\text{Số người không mắc bệnh được sàng tuyển}} \times 100$$

Tính "độ nhạy" và "độ đặc hiệu".

Dựa trên sơ đồ ma trận của nghiệm pháp sàng tuyển:

Kết quả nghiệm pháp sàng tuyển	Chẩn đoán bằng xét nghiệm chuẩn		Tổng Số
	Có bệnh	Không bệnh	
Dương tính	a	b	a+b
Âm tính	c	d	c+d
Tổng số	a+c	b+d	a+b+c+d

$$\text{Độ nhạy} = \frac{a}{a+c} \quad (\text{Sensitivity})$$

$$\text{Độ đặc hiệu} = \frac{d}{b+d} \quad (\text{Specificity})$$

— positive
— negative

Kết quả dương tính giả và âm tính giả.

b là số dương tính giả

c là số âm tính giả

Nghiệm pháp càng nhạy bao nhiêu thì số c càng nhỏ bấy nhiêu vì theo công thức trên "độ nhạy" bằng $a/(a+c)$, c càng nhỏ đi bao nhiêu thì "độ nhạy" càng tiến tới 100%

Nghiệm pháp càng đặc hiệu bao nhiêu thì số b càng nhỏ bấy nhiêu vì "độ đặc hiệu" bằng $d/(d+b)$, b càng nhỏ đi bao nhiêu thì "độ đặc hiệu" càng tiến tới 100%

Giá trị tiên đoán dương tính của nghiệm pháp.

Tỷ suất $a/(a+b)$ được gọi là giá trị tiên đoán dương tính của nghiệm pháp

Giá trị tiên đoán dương tính của nghiệm pháp tăng khi:

- "Độ nhạy" và "độ đặc hiệu" tăng.



$$Se: 50\% \quad Sp = 50\% \quad \frac{a}{a+b} = 50\%$$

- Tỷ lệ bệnh tồn đọng tại một thời điểm tăng và ngược lại, như trong 3 thí dụ sau đây

Bảng 1

Nghiệm pháp sàng tuyển	Nghiệm pháp chuẩn		Tổng số
	Có bệnh	Không bệnh	
Dương tính	50 <i>a</i>	50 <i>b</i>	100
Âm tính	50 <i>c</i>	50 <i>d</i>	100
Tổng số	100	100	200

Bảng 2

Nghiệm pháp sàng tuyển	Nghiệm pháp chuẩn		Tổng số
	Có bệnh	Không bệnh	
Dương tính	60	40	100
Âm tính	60	40	100
Tổng số	120	80	200

Bảng 3

Nghiệm pháp sàng tuyển	Nghiệm pháp chuẩn		Tổng số
	Có bệnh	Không bệnh	
Dương tính	40	60	100
Âm tính	40	60	100
Tổng số	80	120	200

Trong bảng 1, ta thấy

- Tỷ lệ bệnh tồn đọng là 50%
- Độ nhạy và độ đặc hiệu cùng bằng 50%
- Giá trị tiên đoán dương tính của nghiệm pháp tăng lên $60/100=60\%$

Trong bảng 3, ta thấy

- Tỷ lệ bệnh tồn đọng giảm xuống 40%
- Độ nhạy và độ đặc hiệu vẫn giữ nguyên 50%
- Giá trị tiên đoán dương tính của nghiệm pháp giảm xuống $40/100=40\%$

Mức độ chấp nhận.

Một nghiệm pháp sàng tuyển được chấp nhận, dựa trên:

- Mức độ thực tiễn của nghiệm pháp
- Đáp ứng phát hiện nhóm nguy cơ.



Bảng
a
b
c
d

III. BÀI TẬP

Sử dụng nghiệm pháp sàng tuyển trong 10000 người để phát hiện bệnh nhân đái tháo đường; những người có nồng độ đường máu $\geq 180\text{mg/dl}$ được coi là dương tính.

Kết quả như sau:

Kết quả nghiệm pháp sàng tuyển	Xét nghiệm chuẩn		Tổng số
	Có bệnh	Không bệnh	
Dương tính	34 66	20 88	54 164
Âm tính	116 84	9830 9752	9946
Tổng số	150 150	9850 9850	10000

Câu hỏi.

- Tính "độ nhạy", "độ đặc hiệu" và "giá trị tiên đoán dương tính" của bảng số liệu trên? Khi người ta quy định lại nồng độ đường máu 130mg/dl đã coi là dương tính thì:
 - Có 164 người dương tính thay vì 54 người
 - 98 trong số 164 người dương tính trong số 9850 người không có bệnh đái đường (theo nghiệm pháp chuẩn)
- Tính "độ nhạy", "độ đặc hiệu" và "giá trị tiên đoán dương tính" của bảng số liệu mới này?
- Khi ta giảm nồng độ đường máu coi là dương tính từ 180 xuống 130mg/dl , ảnh hưởng nào đã xảy ra trên kết quả dương tính giả, kết quả âm tính giả và kết quả giá trị tiên đoán dương tính?
- Hiệu quả này ảnh hưởng đến "độ nhạy" và "độ đặc hiệu" như thế nào?
- Nếu nghiệm pháp chỉ được coi là dương tính khi nồng độ đường máu cao trên 180mg/dl , khi đó:
 - Hiệu quả này tác động trên "độ nhạy" và "độ đặc hiệu" như thế nào?
 - Hiệu quả này tác động lên số âm tính giả và dương tính giả như thế nào?
- Trong bảng số liệu đầu tiên, giả thiết tỷ lệ bệnh đái đường tồn đọng tăng từ 1,5% ($150/10000$) lên 2%, và giả thiết độ nhạy và độ đặc hiệu vẫn giữ nguyên, hãy tính giá trị tiên đoán dương tính của nghiệm pháp?
- Giả thiết khả năng chẩn đoán sàng tuyển là 1000 người/tuần, giả thiết bệnh có tỷ lệ tồn đọng là 2%, giả thiết nghiệm pháp có độ nhạy là 95% và độ đặc hiệu là 90%
 - Hỏi bao nhiêu người sàng tuyển có nghiệm pháp dương tính mỗi tuần?
 - Trong số họ, bao nhiêu người dương tính thật và bao nhiêu người dương tính giả?
 - Giá trị tiên đoán dương tính của nghiệm pháp là bao nhiêu?
- Hãy cho thí dụ một hai bệnh mà nghiệm pháp sàng tuyển phát hiện bệnh sớm có ý nghĩa thuận lợi cho:
 - Cá nhân
 - Cộng đồng



Trả lời.

1- Độ nhạy = $34/150$ hoặc 22,6%

Độ đặc hiệu = $9830/9850$ hoặc 99,7%

Giá trị tiên đoán dương tính của nghiệm pháp = $34/54$ hoặc 63%

2- Bảng số liệu được thay đổi như sau:

Sàng lọc	Chẩn đoán chuẩn		Tổng số
	Có bệnh	Không bệnh	
Dương tính	66	98	164
Âm tính	84	9752	9836
Tổng số	150	9850	10000

Độ nhạy = $66/150$ hoặc 44%

Độ đặc hiệu = $9752/9850$ hoặc 99%

Giá trị tiên đoán dương tính của nghiệm pháp = $66/164$ hoặc 40%.

3- Coi là dương tính khi giảm nồng độ đường máu sẽ làm tăng số dương tính giả và làm giảm số âm tính giả, làm giảm giá trị tiên đoán dương tính của nghiệm pháp

4- Độ nhạy sẽ tăng và độ đặc hiệu sẽ giảm.

5- a. Coi là dương tính khi tăng nồng độ đường máu sẽ làm độ nhạy giảm và làm độ đặc hiệu tăng

b. Âm tính giả tăng và dương tính giả giảm

6- Bảng số liệu sẽ trở thành như sau

Nghiệm pháp sàng tuyển	Chẩn đoán bằng biện pháp chuẩn		Tổng số
	Có bệnh	Không bệnh	
Dương tính	45	20	65
Âm tính	155	9780	9935
Tổng số	200	9800	10000

Giá trị tiên đoán dương tính của nghiệm pháp = $a/(a+b) = 45/65$ hoặc 69%. Ghi nhận rằng khi tăng tỷ lệ bệnh tồn đọng từ 1,5%-2% thì giá trị tiên đoán dương tính của nghiệm pháp cũng sẽ tăng.

7- Bảng số liệu cho 1 tuần sẽ trở thành:

Kết quả sàng tuyển	Chẩn đoán bằng biện pháp chuẩn		Tổng số
	Có bệnh	Không bệnh	
Dương tính	19	98	117
Âm tính	1	882	883
Tổng số	20	980	1000

Trong số 1000 người được sàng tuyển có 20 người bệnh; tỷ lệ bệnh tổn động là 2%, độ nhạy là 95% có nghĩa là $a/(a+c)=95/100$; $a+c=20$, vậy $a=19$ và $c=1$. Độ đặc hiệu là 90% có nghĩa là $d/(b+d)=90/100$; $b+d=980$, vậy $b=98$, $d=882$.

- a. Nghiên cứu bảng trên cho thấy 117 người dương tính hàng tuần
 - b. Trong số họ, 19 người dương tính thật và 98 người dương tính giả
 - c. Giá trị tiên đoán dương tính của nghiệm pháp là $19/117$ hoặc 16,2%
- 8-a. 2 thí dụ về nghiệm pháp sàng tuyển có ích cho cá nhân
- 1) Sàng phát hiện ung thư cổ tử cung bằng xét nghiệm tế bào học âm đạo
 - 2) Sàng phát hiện ung thư vú bằng chụp vú đồ có ích cho phụ nữ trên 50 tuổi
- b. 2 thí dụ về nghiệm pháp sàng tuyển có ích cho cộng đồng
- 1) Sàng tuyển phát hiện nhiễm liên cầu khuẩn phòng thấp khớp
 - 2) Nghiệm pháp trong da phát hiện bệnh lao.

KẾT HỢP VÀ CĂN NGUYÊN

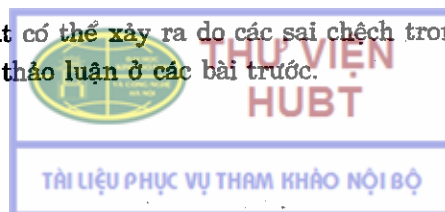
I. MỤC TIÊU

1. Minh họa một ví dụ về khái niệm đa nguyên nhân của bệnh.
2. Định nghĩa những loại kết hợp sau:
 - a. Giả tạo
 - b. Không căn nguyên
 - c. Căn nguyên.
3. Phân biệt giữa kết hợp và căn nguyên và liệt kê những tiêu chuẩn hỗ trợ cho một suy luận về căn nguyên.

II. NỘI DUNG

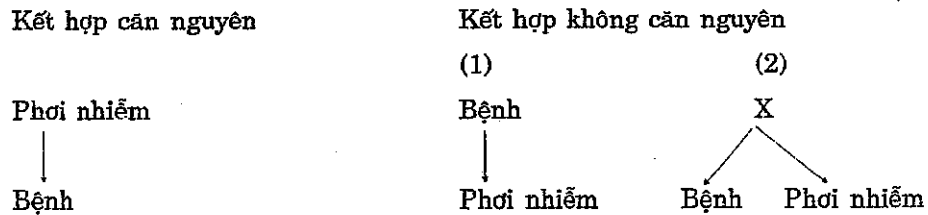
Các nghiên cứu dịch tễ học thường đưa ra sự kết hợp giữa một bệnh và một phơi nhiễm. Đây mới chỉ là bước đầu tiên trong nghiên cứu dịch tễ học. Chúng ta phải phân tích ý nghĩa của sự kết hợp này. Một sự kết hợp có thể là giả tạo, không phải là kết hợp căn nguyên hay là kết hợp căn nguyên.

Một kết hợp giả tạo hay không thật có thể xảy ra do các sai lệch trong nghiên cứu. Nguồn gốc của các sai lệch đã được thảo luận ở các bài trước.



Một kết hợp không căn nguyên xảy ra theo hai cách:

- Bệnh có thể gây ra phơi nhiễm với yếu tố nguy cơ (Hơn là phơi nhiễm gây ra bệnh).
- Cả bệnh và phơi nhiễm kết hợp với một yếu tố thứ ba (yếu tố gây nhiễm) mà ta đã biết hay chưa biết (xem hình 1). Trong trường hợp này, khi đo lường phơi nhiễm, tình cờ chúng ta đo lường yếu tố. Sau đây là một ví dụ về một kết hợp không căn nguyên. Người ta đã tìm thấy có sự kết hợp có ý nghĩa thống kê giữa tử vong do bệnh mạch vành tim và thói quen uống cà phê. Kết quả được trình bày ở bảng 1.



Hình 1 - Kết hợp căn nguyên và không căn nguyên.

Bảng 1

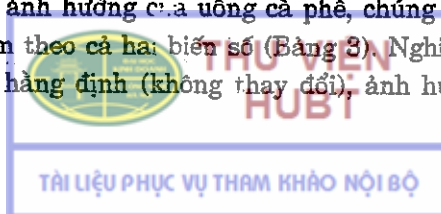
Uống cà phê (số tách mỗi ngày)	Tỷ lệ tử vong do bệnh mạch vành tim ở nam giới 55-64 tuổi (số chết trên 1000 người trong năm)
0	6
1-5	8
≥6	12

Tuy nhiên, người ta cũng biết rằng những người uống và phê thường có hút thuốc lá và kết hợp mạnh giữa hút thuốc lá và tỷ lệ tử vong do bệnh mạch vành tim (xem bảng 2).

Bảng 2

Hút thuốc lá (số bao mỗi ngày)	Tỷ lệ tử vong do bệnh mạch vành tim ở nam giới 55-64 tuổi (số chết trên 1000 người trong năm)
0	4
1-2	10
≥3	15

Do đó, để tách biệt ảnh hưởng của uống cà phê, chúng ta xếp loại chéo tỷ lệ tử vong do bệnh mạch vành tim theo cả hai biến số (Bảng 3). Nghiên cứu bảng 3, chúng ta thấy rằng khi hút thuốc lá hằng định (không thay đổi), ảnh hưởng của uống cà phê đến tử



vong do bệnh mạch vành tim biến mất. Do đó, sự kết hợp giữa uống cà phê và tử vong do bệnh mạch vành tim là sự kết hợp không cân nguyên bị tác động trung gian bởi yếu tố gây nhiễm là hút thuốc lá (hình 1). Điều này có nghĩa là nếu uống cà phê độc lập với hút thuốc lá thì tỷ lệ tử vong do bệnh mạch vành tim là không thay đổi.

Bảng 3

Tỷ lệ tử vong do bệnh mạch vành tim.				
Uống cà phê (số tách mỗi ngày)	Hút thuốc lá (số bao mỗi ngày)			Tổng số
	0	1-2	≥ 3	
0	4	9	15	6
1-5	6	10	13	8
≥ 6	5	9	16	12
Tổng số	4	10	15	

Hút thuốc lá
(Yếu tố gây nhiễm)

Uống cà phê Tử vong do bệnh mạch
vành tim

Hình 2 - Sự kết hợp không cân nguyên giữa uống cà phê và tử vong do bệnh mạch vành tim.

1. Đa nguyên nhân.

Khi một kết quả bị ảnh hưởng bởi nhiều biến số, để kiểm tra ảnh hưởng của từng biến số, cần thiết phải điều chỉnh ảnh hưởng của các biến số khác. Tỷ lệ chuẩn hóa theo tuổi là một ví dụ không chế ảnh hưởng của một biến số là xem xét các kết quả xảy ra ở các mức độ khác nhau của biến số đó, trong khi giữ các biến số khác ở mức độ hằng định. Một kỹ thuật phức tạp hơn là phân tích hồi quy nhiều biến số để xác định ảnh hưởng tương đối của từng biến số một đối với một kết quả. Trong y học có nhiều ví dụ về tính đa nguyên của bệnh. Ví dụ, tỷ lệ chết mẹ bị ảnh hưởng rõ rệt bởi cả tuổi và số con sinh ra. Cần thiết phải tiến hành nghiên cứu phụ nữ với số con nhất định nào đó, ví dụ, 1, 2-3, 4-5 và ≥ 6 . Và ở từng mã nhóm con này, người ta xem xét mối liên quan giữa tuổi và tỷ lệ tử vong mẹ. Kết quả cho thấy là ảnh hưởng của tuổi đều như nhau ở các nhóm mẹ có số con khác nhau, và ảnh hưởng này là rất rõ rệt. Nếu phụ nữ được phân thành các nhóm tuổi khác nhau <20, 20-29, 30-39 và ≥ 40 và ở mỗi nhóm tuổi, họ lại được phân loại theo số con, ta có thể thấy rằng số con tăng lên thì tỷ lệ chết mẹ cũng tăng lên chút ít nhưng không nhiều như ảnh hưởng của tuổi mẹ.

2. Kết hợp cân nguyên.

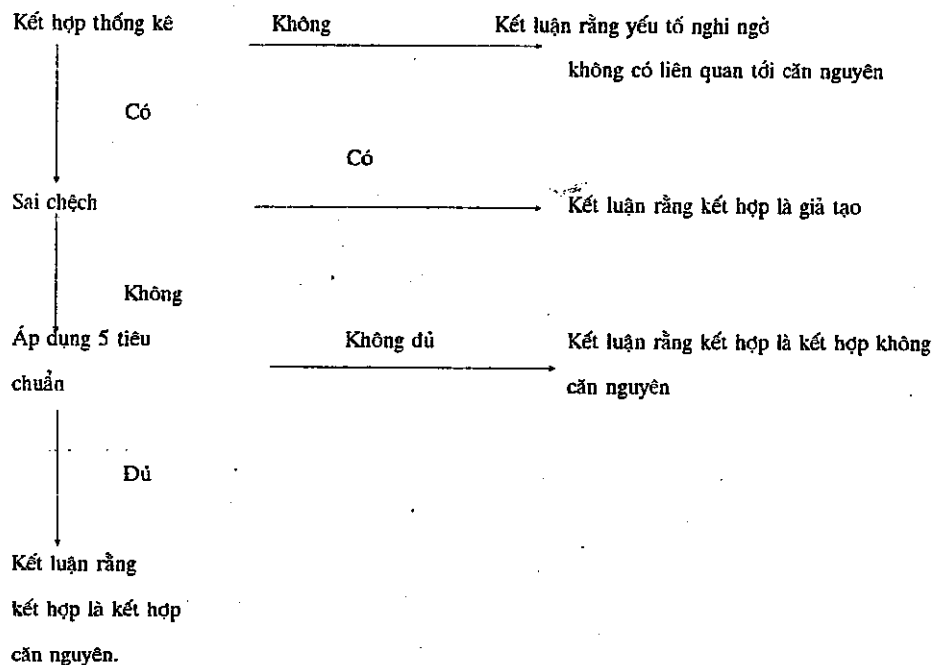
Y học thường quan tâm đến phòng bệnh hay giảm tỷ lệ mắc bệnh. Người ta tiến hành



các nghiên cứu về căn nguyên nhằm tìm ra nguyên nhân gây bệnh để phòng bệnh. Một yếu tố được coi là có tính căn nguyên khi nó góp phần vào sự phát triển bệnh, và nếu loại trừ yếu tố đó đi thì tần số mắc bệnh sẽ giảm đi. Khái niệm về tính căn nguyên trong dịch tễ học khác với trong luật pháp hay triết học. Trong phòng bệnh, xác định một phơi nhiễm là đủ mà không cần thiết phải xác định nguyên nhân cuối cùng của bệnh. Ví dụ, hút thuốc lá được coi là yếu tố nguy cơ làm tăng tỷ lệ hay ung thư phổi và các ung thư khác cũng như các bệnh đường hô hấp và bệnh tim. Việc xác định chính xác thành phơi nhiễm nào trong thuốc lá là thủ phạm chính để tiến hành các biện pháp phòng bệnh là không cần thiết.

3. Xác định nguyên nhân.

Phương pháp thống kê một mình nó chưa đủ bằng chứng cho một kết hợp nhân quả đối với một sự kết hợp. Việc giải thích một kết hợp phải được tiến hành một cách có hệ thống (Hình 2)



Hình 2 - Phiên giải kết quả của một nghiên cứu dịch tễ học.

Ủy ban cố vấn của hội y tế công cộng quân y Mỹ đã đưa ra 5 tiêu chuẩn dùng để xác định mối quan hệ nhân quả. Sử dụng 5 tiêu chuẩn này được coi như là một kiểm định về nguyên nhân gây bệnh. 5 tiêu chuẩn đó là.

- Tính hằng định của sự kết hợp
- Độ mạnh của sự kết hợp
- Độ đặc hiệu của sự kết hợp
- Mối quan hệ về thời gian của sự kết hợp
- Tính chặt chẽ của sự kết hợp



5 tiêu chuẩn trên được định nghĩa tóm tắt như sau:

- Tính hằng định có nghĩa rằng các nghiên cứu khác nhau đều cho một kết quả giống nhau về sự kết hợp, mặc dù các nghiên cứu có thiết kế khác nhau và được tiến hành trên các quần thể khác nhau đôi khi ở những nước khác nhau.

- Độ mạnh của sự kết hợp có liên quan với giá trị của nguy cơ tương đối mà ta tính được. Nguy cơ tương đối càng lớn, khả năng kết hợp đó là kết hợp căn nguyên càng lớn. Hơn nữa, nếu có bằng chứng của quan hệ liều lượng - đáp ứng, khả năng một phơi nhiễm là nguyên nhân gây bệnh càng tăng lên. Quan hệ liều lượng - đáp ứng được coi như là mức độ (ví dụ số liều thuốc lá hút hàng ngày) hay thời gian phơi nhiễm (ví dụ như khoảng thời gian uống thuốc tránh thai).

- Độ đặc hiệu đo lường mức độ mà một phơi nhiễm đặc biệt gây nên một bệnh nào đó. Nếu đáp ứng sinh học đối với phơi nhiễm là thay đổi, ít khi có khả năng phơi nhiễm đó là một nguyên nhân.

- Mối quan hệ thời gian có nghĩa là phơi nhiễm với yếu tố nguy cơ phải xảy ra trước khi mắc bệnh.

- Tính chặt chẽ có nghĩa là sự hợp lý về sinh học được chứng minh ở động vật.

5 tiêu chuẩn nêu trên là chỉ dẫn tốt nhất trong việc xác định một yếu tố căn nguyên gây bệnh trong các trường hợp mà ta không tiến hành được các thực nghiệm có kiểm soát. Tuy nhiên nó không thể thay thế được các thực nghiệm có kiểm soát.

III. BÀI TẬP

A. CÂU HỎI

1- Bảng 4, 5, 6 trình bày kết quả của một nghiên cứu về các yếu tố có liên quan đến đáp ứng đối với chương trình sàng lọc ung thư cổ tử cung.

Bảng 4 - ĐÁP ỨNG VỚI CHƯƠNG TRÌNH THEO TẦNG LỚP XÃ HỘI

Tầng lớp xã hội	Tỷ lệ phần trăm dân chúng đáp ứng
Tầng lớp xã hội cao	75
Tầng lớp xã hội thấp	46
Tất cả các tầng lớp xã hội	53

Bảng 5 - ĐÁP ỨNG ĐỐI VỚI CHƯƠNG TRÌNH THEO TÌNH TRẠNG HÔN NHÂN.

Tình trạng hôn nhân	Tỷ lệ phần trăm dân chúng đáp ứng
Có gia đình	82
Sống độc thân	68
Góa hay li dị	43
Tổng cộng các tình trạng hôn nhân	53



Bảng 6 - PHẦN TRĂM SỐ NGƯỜI Ở TỪNG NHÓM NHỎ ĐÁP ỨNG VỚI CHƯƠNG TRÌNH (CHIA THEO TẦNG LỚP XÃ HỘI VÀ TÌNH TRẠNG GIA ĐÌNH)

Tầng lớp xã hội	Có gia đình	Độc thân	Góa vợ/hay li dị	Các tình trạng hôn nhân
Tầng lớp xã hội cao	83	67	43	75
Tầng lớp xã hội thấp	81	69	43	46
Tất cả các tầng lớp xã hội	82	68	43	53

Từ các số liệu trên, suy luận nào sau đây là đúng.

- Phụ nữ có gia đình đáp ứng tốt hơn
- Không thể rút ra kết luận gì từ những số liệu này vì không biết được tình trạng hôn nhân của những người ở tầng lớp xã hội cao và thấp.
- Những phụ nữ có gia đình có tỷ lệ đáp ứng cao hơn những người sống độc thân hay góa bởi vì nhiều người trong số họ thuộc tầng lớp xã hội cao.
- Không thể rút ra kết luận gì từ các bảng số liệu này vì không biết được đây là nghiên cứu thuần tập hay nghiên cứu bệnh chứng.

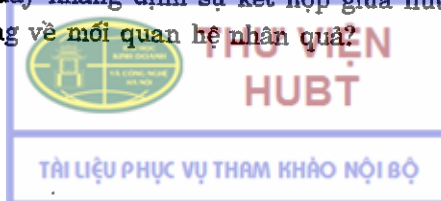
2- Bảng 7 trình bày số liệu thu được từ một nghiên cứu ngang về bệnh béo phì và huyết áp.

Bảng 7 - MỐI LIÊN QUAN GIỮA BÉO PHÌ VÀ HUYẾT ÁP
(số liệu được trình bày trong bảng là số người)

	Huyết áp thấp	Huyết áp trung bình	Huyết áp cao	Tổng cộng
Béo phì	50	50	100	200
Cân nặng bình thường	170	30	100	300
Gầy	380	20	100	500
Tổng cộng	600	100	300	1.000

Từ bảng số liệu này, những kết luận nào sau đây là đúng:

- $50/100 = 50\%$ số người có huyết áp trung bình là béo phì
 - $100/1.000 = 10\%$ số người có huyết áp cao là béo phì
 - $50/200 = 25\%$ số người béo phì có huyết áp thấp
 - $100/500 = 20\%$ số người có huyết áp cao là gầy.
- 3- Người ta đã nghiên cứu kỹ lưỡng sự kết hợp giữa hút thuốc lá và ung thư phổi. Những kết luận nào sau đây khẳng định sự kết hợp giữa hút thuốc lá và ung thư phổi và đưa ra được bằng chứng về mối quan hệ nhân quả?



a. Nguy cơ ung thư phổi tăng lên khi số thuốc lá hút hàng ngày tăng lên và hoặc khi khoảng thời gian hút thuốc lá dài hơn.

b. Những người bỏ hút thuốc lá có tỷ lệ ung thư phổi ở mức trung gian so với những người không hút thuốc lá và hiện đang hút thuốc lá.

c. Các nghiên cứu thực nghiệm trên động vật cho thấy tỷ lệ có vết loét tiền ung thư tăng lên sau khi hít khói thuốc lá vào phổi.

d. Các nghiên cứu tương lai cho kết quả tương tự như các nghiên cứu hồi cứu về sự có mặt và chiều hướng của sự kết hợp.

4- Các nghiên cứu hồi cứu cho thấy ở những người sống sót có tỷ lệ sang chấn về tinh thần trong năm trước khi bị cơn nhồi máu cơ tim so với những người ở nhóm chứng. Liệu có thể kết luận được rằng sang chấn tinh thần gây ra nhồi máu cơ tim?

5- Các nghiên cứu ngang cho thấy rằng những người sống ở khu vực A có tỷ lệ bệnh đường hô hấp cao hơn so với những người sống ở khu vực B. Liệu có thể kết luận được rằng sống ở khu vực A là nguyên nhân gây nên bệnh đường hô hấp?

6- Một nghiên cứu về thai chết lưu và dị dạng bẩm sinh cho thấy rằng các bà mẹ sinh con như thế sử dụng Steroid trong thời kỳ mang thai nhiều hơn các bà mẹ sinh con bình thường ở nhóm chứng. Từ kết quả này, liệu chúng ta có thể kết luận rằng sử dụng steroid trong thời gian mang thai gây nên thai chết lưu và dị dạng bẩm sinh?

B- TRẢ LỜI

1-a. Phụ nữ có gia đình có đáp ứng tốt hơn không liên quan với tầng lớp xã hội.

Bảng 4 cho thấy rằng hình như đáp ứng có liên quan với tầng lớp xã hội. Bảng 5 cho thấy rằng hình như đáp ứng có liên quan với tình trạng hôn nhân. Tuy nhiên, bảng 6 cho thấy rằng nếu tầng lớp xã hội là hằng định, ảnh hưởng của tình trạng hôn nhân vẫn tồn tại, nhưng nếu tình trạng hôn nhân là hằng định, ảnh hưởng của tầng lớp xã hội biến mất.

2. a và c đúng.

3. a, b, c và d tất cả đều đúng.

4. Không. Không thể kết luận được rằng sang chấn tinh thần gây ra nhồi máu cơ tim hay nhồi máu cơ tim gây ra sang chấn tinh thần hay một yếu tố không biết nào đó có thể gây nên cả sang chấn tinh thần và nhồi máu cơ tim. Cần phải tiến hành thêm nghiên cứu tương lai để kiểm tra giả thuyết này.

5. Không nhất thiết. Những người bị bệnh đường hô hấp có thể di chuyển đến vùng A vì lý do khí hậu.

6. Không nhất thiết là như vậy. Điều khó khăn trong nghiên cứu này là steroid có thể được điều trị trong những trường hợp có tiền sử chảy máu hoặc các bệnh xảy ra trong thời kỳ mang thai hay cho những phụ nữ có tiền sử sản khoa bất thường. Sự kết hợp có thể là kết hợp không cần nguyên.



Attack Rate

ĐIỀU TRA XỬ LÝ DỊCH

I. MỤC TIÊU

1. Giải thích phân bố bệnh trong quần thể dân chúng theo thời gian, địa điểm và con người.
2. Mô tả thành phần của một tỷ lệ. Giải thích mối quan hệ giữa tử số và mẫu số.
3. Định nghĩa và tính một tỷ lệ tấn công trong một vụ dịch cấp tính thông thường.

II. NỘI DUNG

Dịch tễ học là một môn khoa học nghiên cứu về tình trạng phân bố bệnh tật, cùng các yếu tố quyết định sự phân bố đó.

Trước một dịch vụ, bao giờ người ta cũng cố gắng tìm kiếm, phát hiện xem ai mắc bệnh và tại sao họ lại mắc bệnh? Ví dụ: Chúng ta phát hiện xem liệu bệnh tật thường xảy ra ở nam giới hay nữ giới, người già hay người trẻ, người giàu hay người nghèo. Có phải họ bị bệnh do mang đặc điểm di truyền, do tiếp xúc nghề nghiệp hay do một thói quen như hút thuốc lá chẳng hạn.

- Để xác định một nghiên cứu có phải thuộc loại nghiên cứu dịch tễ học hay không, người ta thường tìm nhóm đối chứng trong nghiên cứu đó. Để tiến hành so sánh, cần biểu diễn kết quả dưới dạng một tỷ lệ. Một tỷ lệ được tính như sau:

$$\frac{\text{Tử số}}{\text{Mẫu số}} = \frac{\text{Số sự kiện xảy ra}}{\text{Dân số có tiếp xúc với yếu tố nguy cơ}}$$

Các tỷ lệ thường được biểu thị thành số sự kiện tính trên 1000 dân hoặc các bội số thập phân khác.


- Tử số bao gồm những người có xảy ra sự kiện nào đó (ốm hoặc tử vong) trong quần thể tiếp xúc với yếu tố nguy cơ.

1. Tỷ lệ tấn công

Tỷ lệ tấn công đo lường tỷ lệ những người có tiếp xúc với một yếu tố nguy cơ cụ thể nào đó, phát triển thành bệnh.

Tỷ lệ tấn công
(TLTC)

(Attack Rate)
(Attack Rate)

	Số người bị mắc bệnh
	Tổng số người có tiếp xúc với yếu tố nguy cơ của bệnh đó
TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ	

87

Trong một vụ dịch ngộ độc thức ăn, tỷ lệ tấn công được tính cho tất cả các món ăn dùng trong bữa đó. Người ta tính tỷ lệ tấn công ở nhóm những người tiếp xúc (tức là có ăn thức ăn nghi ngờ) và quan trọng hơn nữa là tính cho cả những người không tiếp xúc (tức là không ăn thức ăn đã nêu), số liệu được minh họa trong bảng 1.

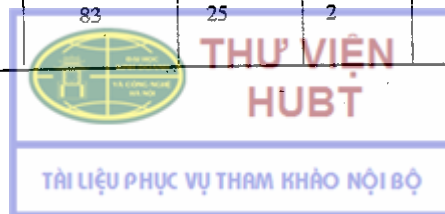
Bảng 1- TỶ LỆ TẤN CÔNG XÉT THEO TIỀN SỬ DÙNG THỨC ĂN TRONG MỘT VỤ DỊCH NGỘ ĐỘC THỨC ĂN.

Món ăn	Những người ăn món ăn đó			Những người không ăn món ăn đó			Chênh lệch giữa các tỷ lệ
	số người	số ốm	TLTC (%)	số người	số ốm	TLTC (%)	
Thịt gà	133	97	73	23	2	8	+65 ✓
Nước sốt	121	88	73	37	11	30	+43
Xúp khoai tây	127	92	72	31	7	22	+50 ✓
Đậu Hà Lan	105	77	73	53	22	41	+32
Bánh mì	66	50	76	92	49	53	+23
Mỡ hóa học	66	50	76	92	49	53	+23
Xà lách	4	1	25	154	98	64	-39
Bánh ngọt	11	1	9	147	98	67	-58
Tráng miệng	36	22	61	122	77	63	-2
Cà phê	98	59	60	60	40	67	-7
Sữa	18	12	67	140	87	62	+5

Bằng cách đánh giá các tỷ lệ tấn công ở những người ăn từng món ăn cụ thể, chúng ta không thể xác định được phương tiện lan truyền bệnh. Tuy nhiên bằng việc so sánh các tỷ lệ tấn công giữa những người ăn và những người không ăn món ăn đó, sẽ cho ta những thông tin có giá trị. Cột cuối cùng bên phải cho thấy sự khác biệt này. Để phân biệt giữa món thịt gà và món xúp khoai tây, hai món này có tỷ lệ chênh lệch lớn, nhất thiết phải dấu chéo như minh họa ở bảng 2. Nghiên cứu bảng này cho thấy món gà tây là món ăn nghi ngờ mang mầm bệnh.

Bảng 2- TỶ LỆ TẤN CÔNG KẾT HỢP CẢ HAI MÓN ĂN NGHI NGỜ

	Có ăn món thịt gà			Không ăn món thịt gà		
	số người	số ốm	TLTC (%)	số người	số ốm	TLTC (%)
Có ăn xúp khoai tây	127	92	72	0	0	0
Không ăn xúp khoai tây	6	5	83	25	2	8



2. Điều tra một vụ dịch.

Một vụ dịch xảy ra khi số trường hợp mắc bệnh gia tăng hơn so với cùng thời kỳ những năm trước đó, ở tại quần thể đó một cách có ý nghĩa thống kê. Bệnh xảy ra ở một cá thể có thể xem như là kết quả tương tác giữa ba yếu tố: tác nhân, vật chủ và môi trường. Trong điều tra một vụ dịch xảy ra ở một cộng đồng nào đó, người ta thường nhìn nhận dưới ba góc độ: Thời gian, địa điểm và con người để mô tả.

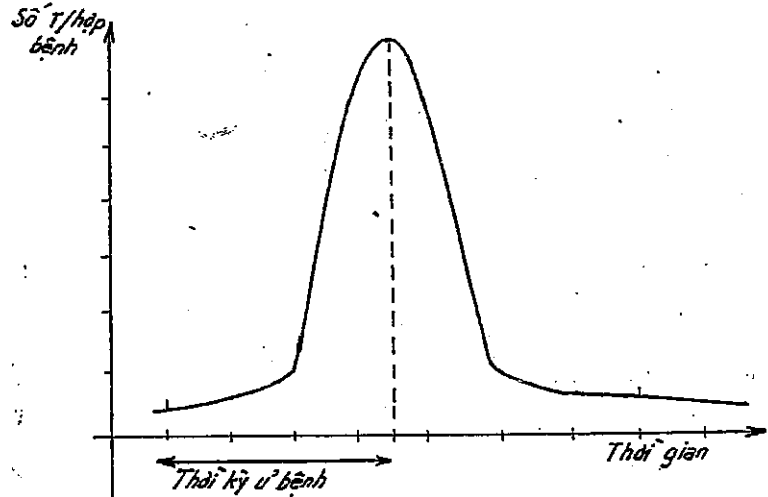
Các trường hợp bệnh có thể được phân chia theo 3 nhóm sau:

1. Thời gian, bao gồm cả ngày khởi bệnh
2. Địa điểm, gồm cả nơi ở và nơi làm việc
3. Các đặc trưng về con người, bao gồm tuổi, giới tính và nghề nghiệp.

Xem xét kết quả sau khi đã phân loại như trên, có thể giúp chúng ta nhận biết được những đặc điểm phổ biến của những người ốm nhưng lại hiếm thấy ở những người khỏe mạnh.

3. Đường biểu diễn dịch.

Đường biểu diễn dịch phác họa hình ảnh một vụ dịch, mà nhìn vào đó chúng ta có thể rút ra một số nhận định. Trong một vụ dịch ngộ độc thức ăn, chẳng hạn như ví dụ trên, thời gian từ khi tiếp xúc với chung (như ăn uống) và đỉnh cao xuất hiện nhiều trường hợp bệnh là xấp xỉ với thời kỳ ủ bệnh. Không có các trường hợp bệnh thứ phát, đường cong thuộc loại chỉ có một đỉnh (xem hình)



Hình 1- Đường biểu diễn vụ dịch ngộ độc thức ăn.

Tuy nhiên, với một bệnh truyền nhiễm điển hình, chẳng hạn như sởi, lan truyền từ người sang người, hình ảnh lan truyền minh họa bằng đường cong vụ dịch không được rõ ràng. Ở những cộng đồng khép kín, chẳng hạn như trường học nội trú, trại lính hoặc tàu thủy, đường cong vụ dịch có thể vạch ra những đỉnh sóng lan truyền kế tiếp nhau, mỗi đỉnh biểu thị một đợt các trường hợp mới, các đỉnh cách đỉnh trước bởi một thời kỳ ủ bệnh. Dịch ngừng lại khi nguồn truyền nhiễm không còn nữa.

4. Phân tích vụ dịch

Số liệu sau khi đã được thu thập, cần phân tích như sau:

- a. Về đường biểu diễn dịch. Có thể ước tính được thời gian ủ bệnh nếu biết thời gian tiếp xúc.
- b. Tính các tỷ lệ tấn công theo tuổi, giới tính và các nhóm nghề nghiệp khác nhau.
- c. Vẽ bản đồ phân bố những trường hợp mắc bệnh theo địa dư hoặc nơi làm việc.

Sau khi phân tích theo các bước như trên, có thể xác định được phương tiện lan truyền bệnh nghi ngờ. Sau đó tính các tỷ lệ tấn công cho nhóm tiếp xúc và nhóm không tiếp xúc với phương tiện này. Tuy nhiên, đôi khi cũng không phát hiện được phương tiện lan truyền bệnh. Chẳng hạn, khi điều tra một vụ dịch ngộ độc thức ăn xảy ra tại một khách sạn đông khách, thường không thể phát hiện được tổng số người có tiếp xúc với một loại thức ăn đặc biệt. Trong những trường hợp như vậy, chúng ta có thể so sánh tỷ lệ tiếp xúc với phương tiện lan truyền bệnh nghi ngờ giữa nhóm mắc bệnh và nhóm những người khỏe mạnh, như trong ví dụ minh họa sau đây.

Tại một khách sạn xảy ra một vụ dịch nhiễm độc thức ăn ở 30 người. Trong đó có 24 người chiếm 80%, có ăn món sò huyết. Nếu chỉ căn cứ vào những tỷ lệ này thì chưa đủ kết luận món sò huyết là phương tiện truyền bệnh. Nhất thiết phải điều tra một mẫu những người có ăn bữa tối đó mà không mắc bệnh. Có 3 người, chiếm 10% không mắc bệnh nhưng có ăn món sò huyết. Đây là một bằng chứng chắc chắn để quy kết cho món sò huyết. Cần ghi nhớ là phép so sánh giữa 2 tỷ lệ nói trên là giữa nhóm người khỏe mạnh và nhóm người mắc bệnh.

III. BÀI TẬP

VỤ DỊCH VÀNG DA XẢY RA TẠI MỘT VÙNG NÔNG THÔN

1. Giới thiệu.

Thứ sáu, ngày 17 tháng 5 năm 1968, đơn vị chịu trách nhiệm về viêm gan thuộc trung tâm kiểm soát bệnh tật (CDC) Atlanta, Georgia nhận được bức điện yêu cầu hỗ trợ điều tra một vụ dịch viêm gan truyền nhiễm. Bức điện cho biết, từ 30/4 đến 16/5/1968, có 32 ca viêm gan truyền nhiễm được báo cáo cho phòng y tế thuộc phường N.T quận L bang Michigan (Mỹ).

Câu hỏi 1: Nếu chỉ biết từ 30/4 đến 16/5/1968 có 32 ca mắc vàng da được thông báo tới phòng y tế huyện thì có thể kết luận rằng ở đây đang có một vụ dịch vàng da xuất hiện hay không? Tại sao?

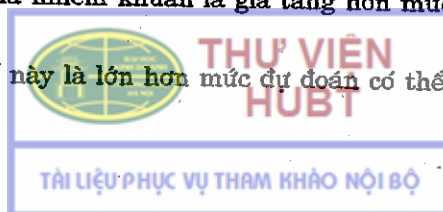
Quận L. nằm ở khu vực phía Bắc của bang Michigan. Diện tích 576km², trong đó có 2025 người sống ở phường N.T. Cộng đồng này mang tính thôn quê điển hình, chia thành hai bộ phận: nông nghiệp và lâm nghiệp.

Công tác điều tra dịch được tiến hành tập trung vào phường NT. Ngoài bộ phận này, chỉ còn 2 quần thể khác có dân số đáng kể nằm trong quận L là làng có 435 người và làng B - dân số 308 người. Khu vực còn lại được chia thành 14 xã, dân chúng hầu hết thuộc tầng lớp dưới. Về mùa hè thường có nhiều dân các nơi đến du lịch tại đây.

Trong năm trước năm 1967, đến tháng 4/1968, có 7 trường hợp viêm gan truyền nhiễm được báo cáo tới phòng y tế quận. Trong đó 4 ca xảy ra ở cùng một gia đình vào thời điểm tháng 4/1967, 3 ca còn lại xảy ra không xác định được mối quan hệ giữa chúng.

Câu hỏi 2: a. Có phải 32 ca bệnh vàng da nhiễm khuẩn là gia tăng hơn mức dự đoán thông thường không?

b. Làm thế nào để xác định được con số này là lớn hơn mức dự đoán có thể có được?



2. Điều tra vụ dịch.

Đến 19/5 có 39 ca viêm gan được báo cáo. Tới 25/5, con số này tăng lên 61, và tới 1/6, hai ca cuối cùng được thông báo đưa tổng số trường hợp lên đến con số 63.

Bước đầu tiên trong cuộc điều tra bao gồm việc phỏng vấn trực tiếp từng bệnh nhân. Tất cả các cuộc phỏng vấn đều cùng do hai điều tra viên thực hiện tại nhà bệnh nhân. Bệnh nhân được hỏi về ngày khởi bệnh, triệu chứng lâm sàng, tiền sử tiếp xúc với bệnh nhân viêm gan trước đó, tiền sử di chuyển ra ngoài phạm vi của công đồng đang ở, và tiền sử tiêm truyền các sản phẩm của máu. Ngoài ra còn thu thập thông tin về tình trạng bệnh và tiền sử tiêm gammaglobin của tất cả các thành viên khác trong gia đình. Các bệnh nhân và gia đình của họ được hỏi về nguồn nước sử dụng, nơi mua sữa, thức ăn và những dịp tụ tập ở các nơi công cộng. Đồng thời các nhà điều tra còn lấy các mẫu nước ở vòi nước máy của từng gia đình để xét nghiệm về mặt vi khuẩn học.

Câu hỏi 3: Dựa trên các thông tin kể trên, lý tưởng nhất cần lấy thêm thông tin gì?

3. Các đặc điểm của vụ dịch.

Theo thời gian:

Tổng số có 63 ca viêm gan truyền nhiễm trong thời gian từ 1/4 đến 8/5/1968. Trong số đó, 61 trường hợp có ngày khởi bệnh nằm trong khoảng từ 28/4 đến 26/5 (Hình 2, các trường hợp bệnh theo thời gian tạo thành đường cong vụ dịch).

Câu hỏi 4: Có thể rút ra điều gì nếu xét thời gian tiếp xúc ban đầu với người nhiễm khuẩn?

Theo con người:

Bảng 3 minh họa các tỷ lệ tấn công tính theo từng độ tuổi và giới đối với những ca bệnh có ngày khởi bệnh từ 28/4 đến 26/5. Tính chung tỷ lệ tấn công ở nam giới cao gần gấp đôi ở nữ: Nam 8,1/1000 so với Nữ là 4,5/1000.

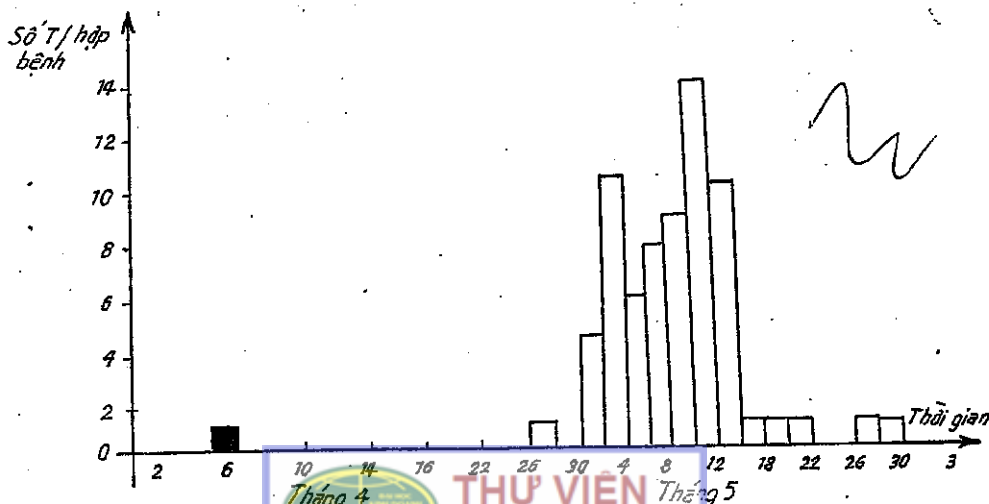
43 trường hợp (70%) xảy ra ở trẻ em ở độ tuổi đi học. Số còn lại xảy ra ở những thanh niên đã tốt nghiệp phổ thông. Trong số 43 ca bệnh là học sinh có 36 trường hợp học ở trường phổ thông N.T.

Các trường khác lác đác có các trường hợp bệnh. Người ta còn tìm thấy trường hợp bệnh ở trường thuộc bang khác nhưng lại là người sống ở quận L.

Câu hỏi 5: a. Có bao nhiêu ca bệnh xảy ra ở trẻ em dưới 5 tuổi?

b. Khoảng tuổi nào có tỷ lệ tấn công cao nhất?

c. Ở thời điểm này của cuộc điều tra có thể lập giả thuyết gì khi xem xét sự phân bố các trường hợp bệnh theo thời gian và đặc điểm con người?



Hình 2. Số trường hợp viêm gan truyền nhiễm từ tháng 4 đến tháng 5/1968, tại quận L.

Bảng 3- TỶ LỆ TẤN CÔNG XÉT THEO TUỔI VÀ GIỚI CỦA CÁC CA BỆNH VIÊM GAN, QUẬN L, TỪ NGÀY 28/4 ĐẾN 26/5

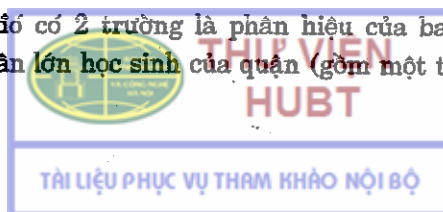
Nhóm tuổi	Dân số	Số mắc			Tỷ lệ tấn công/1000		
		Nam	Nữ	Tổng số	Nam	Nữ	Tổng số
0-4	1740	0	0	0	0,0	0,0	0,0
5-9	1000	2	2	4	3,7	4,5	4,0
10-14	989	12	6	18	22,2	13,4	18,2
15-19	868	16	7	23	35,9	16,6	26,5
20-24	494	1	3	4	4,2	11,8	8,1
25-29	455	0	1	1	0,0	4,6	2,2
30-34	435	3	0	3	14,1	0,0	6,9
35-39	545	1	2	3	4,0	6,7	5,5
40-44	540	2	0	2	7,4	0,0	3,7
45-49	588	1	0	1	3,4	0,0	1,7
50-54	526	2	0	2	7,6	0,0	3,8
55+	1500	0	0	0	0,0	0,0	0,0
Tổng cộng	9680	40	21	61	8,1	4,5	6,3

Bảng 4- TỶ LỆ TẤN CÔNG CỦA BỆNH VIÊM GAN TRUYỀN NHIỄM THEO THỨ TỰ LỚP CỦA TRƯỜNG PHỔ THÔNG CƠ SỞ QUẬN L. 28/4-26/5.

Lớp	Trường N.T			Trường S.L		
	Số học sinh	Số ốm	Tỷ lệ tấn công %	Số học sinh	Số ốm	Tỷ lệ tấn công %
Mẫu giáo	126	2	1,6			
1	128	0	0,0	37	1	2,7
2	121	0	0,0	41	1	2,4
3	107	0	0,0	37	0	0,0
4	106	2	1,9	26	0	0,0
5	120	1	0,8	30	0	0,0
6	111	1	0,9	32	0	0,0
7	110	3	2,7	26	0	0,0
8	120	6	5,0	21	0	0,0
9	143	7	4,9	0		
10	128	11	8,6			
12	93	2	2,2			

Theo địa điểm:

Quận L có 4 trường học, trong đó có 2 trường là phần hiệu của bang P bên cạnh. Trường công cộng N phục vụ cho phần lớn học sinh của quận (gồm một tòa nhà cao tầng



khép kín nằm ở phía nam N, phục vụ cho 1525 học sinh- Từ mẫu giáo đến lớp 12). 70% số học sinh trường này đi học bằng xe ô tô của trường. Còn có một số trường cơ đốc giáo với 240 học sinh gồm từ lớp 1 đến lớp 8 (trường S.L). Học sinh trường này dùng chung xe ô tô với học sinh của trường N.T.

Bảng 4 trình bày tỷ lệ tấn công ở cả 2 trường N.T và S.L. Tỷ lệ tấn công là tương đối thấp cho các lớp từ lớp 6 trở xuống, nhưng từ lớp 7 bắt đầu có sự gia tăng tỷ lệ tấn công đối với trường công cộng N.T. Tỷ lệ tấn công cao nhất xảy ra ở lớp 10, rồi xuống thấp khá rõ rệt ở lớp 11 và 12. Có sự khác biệt rõ rệt về tỷ lệ tấn công giữa lớp 7 và lớp 8 của trường công cộng N.T và trường cơ đốc giáo S.L

Câu hỏi 6: Có thể rút ra kết luận gì về sự phân bố của bệnh xét theo địa điểm dựa vào các thông tin kể trên?

Phân tích xét theo yếu tố tuổi:

Bảng 3 cho thấy sự phân bố tần số bệnh viêm gan theo tuổi và bảng 4 cho thấy sự phân bố bệnh theo trường và lớp học.

Câu hỏi 7: Tại sao việc tính tỷ lệ tấn công theo tuổi lại là quan trọng? Một lần nữa cần ghi nhớ là đại đa số các ca bệnh là xảy ra ở nhóm tuổi 10-19, điều này được chứng minh bằng các tỷ lệ tấn công.

Câu hỏi 8: Bằng cách sử dụng những thông tin tính theo các đặc trưng: con người và địa điểm, người ta có thể rút ra kết luận gì vào lúc này?

4. Nguồn gốc của vụ dịch.

Qua xem xét sự phân bố các ca bệnh theo thời gian, địa điểm và con người, các nhà điều tra đã thu thập được những thông tin quan trọng có liên quan tới quần thể nhóm học sinh. Những học sinh từ lớp mẫu giáo đến lớp 6 bắt buộc phải ăn trưa tại trường. Các cháu có thể ăn thức ăn mang từ ở nhà đến hoặc tại căng tin của trường. Tuy nhiên học sinh từ lớp 7 đến lớp 12 của trường công cộng N.T có thể được phép không ăn trưa tại trường. Vì trường nằm trên một phố chính của N.T nên nhiều học sinh hàng ngày thường ăn trưa ngoài cửa hiệu. Còn trường S.L lại không cho phép bất kỳ một học sinh nào từ lớp 1 đến lớp 8 ăn trưa ở ngoài trường. Tất cả các học sinh của trường bắt buộc phải ăn trưa ở căng tin nhà trường hoặc mang thức ăn từ nhà đến ăn tại trường.

Khẳng định rằng các trường hợp mắc bệnh đều có một nguồn tiếp xúc chung, các nhà điều tra bắt đầu xem xét phương tiện lan truyền bệnh. Nguồn lây nhiễm quan trọng đầu tiên được điều tra là các phương tiện liên quan đến thức ăn và nước. Bảng 5 cho thấy tiền sử tiếp xúc của 41 bệnh nhân viêm gan, tuổi từ 10 đến 19 ở quận L trong tháng 5/1968.

Bảng 5- TIỀN SỬ TIẾP XÚC CỦA 41 CA BỆNH VIÊM GAN, TUỔI TỪ 10 ĐẾN 19, QUẬN L., 5/1968

Thức ăn hoặc nước	có	không	không nhớ	% có tiếp xúc
Khách sạn A	15	25	1	36,6
Khách sạn B	17	23	1	41,5
Hiệu ăn Nữ Hoàng N.T	28	12	1	68,3

Hiệu ăn Nữ Hoàng S.	8	32	1	19,5
Hiệu bánh N.T	37	3	1	90,2
Nhà máy nước N.T	36	5	0	87,8

Câu hỏi 9: a. Từ thông tin cung cấp ở bảng 5, có thể đưa ra giả thuyết gì về nguồn gốc của vụ dịch?

b. Liệu chỉ dựa vào một số liệu này thôi có đủ để chỉ ra nguồn gốc của vụ dịch không?

5. Tiền sử ăn uống của những người khỏe mạnh.

Cho đến lúc này các nhà điều tra mới chỉ tập trung vào các đối tượng bị bệnh trong quần thể. Người ta đã tính tỷ lệ tấn công theo từng loại thức ăn khác nhau. Nhưng để chỉ ra được loại thức ăn nghi ngờ, bây giờ cần phải xem xét nơi ăn uống của nhóm đối tượng khỏe mạnh. Bằng các so sánh tỷ lệ tiếp xúc đối với các nguồn ăn uống giữa những người khỏe mạnh và những người ốm, người ta có khả năng chỉ ra chính xác hơn nữa nguồn gốc thực sự của vụ dịch.

Câu hỏi 10: a. Hãy nghiên cứu bảng 6 dưới đây. Bằng cách so sánh 6 nguồn tiếp xúc đã đưa ra trong bảng, nguồn tiếp xúc nào cho thấy sự khác biệt lớn nhất giữa người khỏe và người ốm?

b. Hãy giải thích tỷ lệ tiếp xúc cao với nước sử dụng ở cả 2 nhóm người khỏe và nhóm người ốm?

Bảng 6 - SO SÁNH TIỀN SỬ TIẾP XÚC CỦA 41 BỆNH NHÂN VIÊM GAN Ở NHÓM TUỔI 10-19 VỚI TIỀN SỬ TIẾP XÚC CỦA 56 ĐỐI TƯỢNG KHỎE MẠNH CÙNG ĐỘ TUỔI QUA ĐIỀU TRA TẠI NHÀ, 6/1968.

Thức ăn và nước ở	Nhóm bệnh nhân viêm gan			
	có	không	không nhớ	% có tiếp xúc
Khách sạn A	15	25	1	36,6
Khách sạn B	17	23	1	41,5
Hiệu ăn Nữ Hoàng N.T	28	12	1	68,3
Hiệu ăn Nữ Hoàng S.	8	32	1	19,5
Hiệu bánh N.T	37	3	1	90,2
Nhà máy nước khu vực N.T	30	5	0	87,8
Thức ăn và nước ở	Nhóm đối tượng khỏe mạnh			
	có	không	không nhớ	% có tiếp xúc
Khách sạn A	22	31	3	39,3
Khách sạn B	15	39	2	26,8
Hiệu ăn Nữ Hoàng N.T	39	17	0	69,6
Hiệu ăn Nữ Hoàng S.	6	50	0	10,7
Hiệu bánh N.T	26	29	1	46,4
Nhà máy nước khu vực N.T	51	4	1	91,1

MỤC LỤC

	<i>Trang</i>
Lời nói đầu	3
X Tỷ lệ mắc (HT Hoa)	5
X Tỷ lệ tử vong (HT Hoa, DD Thiện)	12
X Đo lường nguy cơ (HT Hoa)	22
Biến sinh học (DD Thiện)	27
X Độ tập trung và độ phân tán (DD Thiện)	37
Ý nghĩa thống kê-Sai số chuẩn-Tự đối chiếu cặp (DD Thiện)	58
X Xác suất (PH Dũng)	71
X Kỹ thuật sàng tuyển phát hiện hàng loạt (PH Dũng)	75
Kết hợp và căn nguyên (NT Hiến)	80
Điều tra xử lý dịch (HT Hoa)	87
X Nghiên cứu bệnh chứng (HT Hoa, DD Thiện)	99
X Nghiên cứu thuần tập (HT Hoa, DD Thiện)	108
X Thử nghiệm lâm sàng ngẫu nhiên (NT Hiến, DD Thiện)	117
Điều tra nghiên cứu sức khỏe cộng đồng - Nghiên cứu ngang (DD Thiện)	126
Lập kế hoạch và tổ chức một cuộc điều tra cộng đồng (NT Hiến)	133
Sử dụng thông tin sẵn có (NT Hiến)	149
Chọn mẫu (DD Thiện)	159
Cỡ mẫu (DD Thiện)	164
Mẫu ngẫu nhiên đơn (DD Thiện)	173
Mẫu tầng (DD Thiện)	177
Mẫu chùm (DD Thiện)	181
Thiết kế bộ câu hỏi, kỹ thuật phỏng vấn và điền bộ câu hỏi (LV Anh)	188
Thu thập và trình bày kết quả nghiên cứu (TV Chí)	201
Bài tập tự đánh giá (NT Hiến, HT Hoa)	214



THƯ VIỆN
HUBT

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ



**THƯ VIỆN
HUBT**

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ



**THƯ VIỆN
HUBT**

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

Hiệu ăn Nữ Hoàng S.	8	32	1	19,5
Hiệu bánh N.T	37	3	1	90,2
Nhà máy nước N.T	36	5	0	87,8

Câu hỏi 9: a. Từ thông tin cung cấp ở bảng 5, có thể đưa ra giả thuyết gì về nguồn gốc của vụ dịch?

b. Liệu chỉ dựa vào một số liệu này thôi có đủ để chỉ ra nguồn gốc của vụ dịch không?

5. Tiền sử ăn uống của những người khỏe mạnh.

Cho đến lúc này các nhà điều tra mới chỉ tập trung vào các đối tượng bị bệnh trong quần thể. Người ta đã tính tỷ lệ tấn công theo từng loại thức ăn khác nhau. Nhưng để chỉ ra được loại thức ăn nghi ngờ, bây giờ cần phải xem xét nơi ăn uống của nhóm đối tượng khỏe mạnh. Bằng các so sánh tỷ lệ tiếp xúc đối với các nguồn ăn uống giữa những người khỏe mạnh và những người ốm, người ta có khả năng chỉ ra chính xác hơn nữa nguồn gốc thực sự của vụ dịch.

Câu hỏi 10: a. Hãy nghiên cứu bảng 6 dưới đây. Bằng cách so sánh 6 nguồn tiếp xúc đã đưa ra trong bảng, nguồn tiếp xúc nào cho thấy sự khác biệt lớn nhất giữa người khỏe và người ốm?

b. Hãy giải thích tỷ lệ tiếp xúc cao với nước sử dụng ở cả 2 nhóm người khỏe và nhóm người ốm?

Bảng 6 - SO SÁNH TIỀN SỬ TIẾP XÚC CỦA 41 BỆNH NHÂN VIÊM GAN Ở NHÓM TUỔI 10-19 VỚI TIỀN SỬ TIẾP XÚC CỦA 56 ĐỐI TƯỢNG KHỎE MẠNH CÙNG ĐỘ TUỔI QUA ĐIỀU TRA TẠI NHÀ, 6/1968.

Thức ăn và nước ở	Nhóm bệnh nhân viêm gan			
	có	không	không nhớ	% có tiếp xúc
Khách sạn A	15	25	1	36,6
Khách sạn B	17	23	1	41,5
Hiệu ăn Nữ Hoàng N.T	28	12	1	68,3
Hiệu ăn Nữ Hoàng S.	8	32	1	19,5
Hiệu bánh N.T	37	3	1	90,2
Nhà máy nước khu vực N.T	30	5	0	87,8
Thức ăn và nước ở	Nhóm đối tượng khỏe mạnh			
	có	không	không nhớ	% có tiếp xúc
Khách sạn A	22	31	3	39,3
Khách sạn B	15	39	2	26,8
Hiệu ăn Nữ Hoàng N.T	39	17	0	69,6
Hiệu ăn Nữ Hoàng S.	6	50	0	10,7
Hiệu bánh N.T	26	29	1	46,4
Nhà máy nước khu vực N.T	51	4	1	91,1

Xem xét nguồn sữa:

Tất cả các loại sữa có bán ở quận L. là lấy từ các cửa hàng bò sữa đặt ở ngoài quận L. Về mặt thương mại, thì không có loại sữa nào mà sản xuất ra chỉ là để phục vụ riêng cho quận.

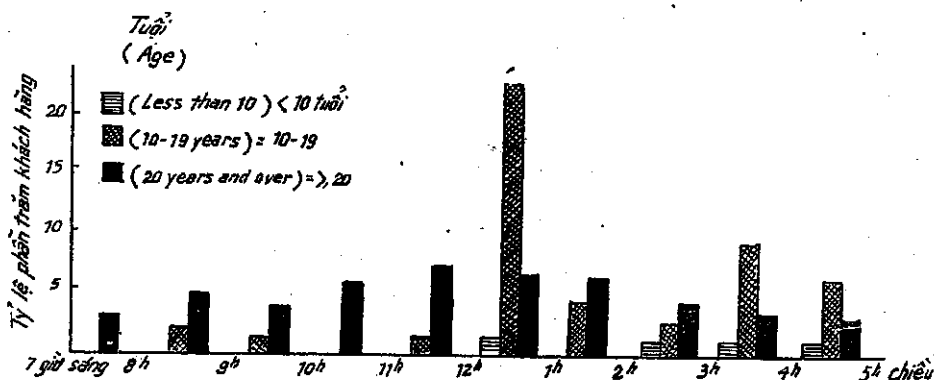
Câu hỏi 11: Có thể rút ra kết luận gì về sữa được nghi ngờ là nguồn truyền nhiễm?

Xem xét nguồn thức ăn:

Cùng tiếp xúc với một loại thức ăn có thể giải thích được các đặc điểm của loại dịch kiểu này. Các loại thức ăn duy nhất có thể được chế biến và tiêu thụ trong khu vực của quận là những thức ăn phục vụ cho khách sạn, rau sống bán ở cửa hàng bán thức ăn làm sẵn, kem ở cửa hàng ăn uống và các loại bánh mỳ.

Tuyệt đại đa số người dân sống tại quận dùng các loại thức ăn đó. Hầu hết các ca bệnh viêm gan sống tại quận L. đều có tiền sử ăn bánh mỳ của cửa hàng bánh N.T. Tuy nhiên chỉ căn cứ vào thông tin này thôi thì chưa thể xác định được liệu cửa hàng bánh mỳ có thể là nguồn gốc của vụ dịch này hay đơn thuần chỉ là nơi tụ tập dân chúng.

Câu hỏi 12: Hãy nghiên cứu hình 3 dưới đây. Những số liệu này giúp cho cuộc điều tra như thế nào?



Hình 3. Tỷ lệ phần trăm khách hàng bị nhiễm độc tại nhà hàng bánh N.T theo giờ và tuổi.

Xem xét tiền sử của các trường hợp bệnh:

Để nghiên cứu vấn đề kỹ càng hơn, các nhà điều tra cố gắng đi khai thác tìm kiếm những trường hợp bệnh nhân mà họ lại sống ngoài khu vực quận L. Trong bất cứ một cuộc điều tra nào có dạng biểu diễn kiểu như thế này thì việc khai thác chi tiết tiền sử của các ca bệnh thường đưa ra những bằng chứng trợ giúp khá lý thú.

Bệnh nhân 1: Đối tượng là nữ giáo viên 45 tuổi sống cách N.T 60 dặm. Cô ta chỉ ghé qua N.T vào dịp 20/3; 5/4 và 14/4 trên đường đi thăm người cha bị ốm, ông ta sống ở bang cạnh đó. Cô ta chỉ dừng lại ở N.T trong hai chuyến đi đầu. Ngày 20/3, cô uống một tách cà phê tại khách sạn N.T. Ngày 5/4 cô ăn vài chiếc bánh ngọt và uống một tách cà phê. Ngày 5/5, cô mắc bệnh viêm gan.

Bệnh nhân 2: Là một người nội trợ 35 tuổi, sống ở quận bên cạnh.

Bệnh nhân 3: Cháu gái 8 tuổi, con của bệnh nhân 2.

Bệnh nhân 4: Là người nội trợ 49 tuổi, sống ở khu vực bên cạnh, đồng thời là chị dâu của bệnh nhân 2 và là dì của bệnh nhân 3. Cả hai gia đình của bệnh nhân 2 và bệnh nhân 3 đều có nhà riêng ở quận L. (nhưng cách khu vực N.T từ 10-15 dặm). Họ về nhà riêng vào các ngày 8, 9, 10 tháng 4 để sửa sang nhà chuẩn bị cho mùa hè tới. Bệnh nhân 3 và bệnh nhân 4 chắc chắn không hề đi vào khu vực N.T vào ngày 9/4, bệnh nhân 4

trông nom hai đứa trẻ của gia đình bệnh nhân 2, để bệnh nhân 2 đi vào N.T mua sắm. Trong thời gian này, bệnh nhân 2 có mua bột mì của cửa hàng bánh mì N.T mang về nhà chuẩn bị cho bữa ăn trưa. Cả 3 bệnh nhân này đều ăn cùng 1 loại bánh mì tự họ làm. Bệnh nhân 2 và 3 bắt đầu ốm vào ngày 7/5 và bệnh nhân 4 ốm vào ngày 11/5. Ngoài ra không có ai khác trong 2 gia đình này ăn loại bột mì trên và không ai bị ốm trong thời gian họ ở quận L.

Bệnh nhân 5: Là một phụ nữ trẻ 35 tuổi, mẹ của 6 đứa con, đã mua bánh mì và bánh tráng kem tại hiệu bánh N.T vào ngày 6/4. Chị mang bánh về nhà, và cùng gia đình đã ăn loại bánh này.

Bệnh nhân 6 và 7: Là hai con gái lớn của bệnh nhân 5.

Bệnh nhân 8 và 9: Là hai cậu con trai út của bệnh nhân 5. Chúng trở về nhà muộn hơn và ăn số bánh còn lại, trong đó chỉ còn 2 chiếc bánh tráng kem.

Bệnh nhân 10: Là con gái 5 tuổi của bệnh nhân 5. Bệnh nhân 10 và người con sinh đôi 5 tuổi về nhà sau cùng, cô bé đói và giành lấy chiếc bánh để ăn. Cô ta bị ốm, còn người con sinh đôi thì không bị. Người cha đi làm từ sáng đến tối không ăn trưa ở nhà, không ăn bánh do vợ mua về. Ông ta không bị ốm.

Câu hỏi 13: Các số liệu trên có phù hợp với giả định rằng hiệu bánh N.T là nơi gây dịch không?

Sự xuất hiện bệnh viêm gan truyền nhiễm sau một tháng kể từ khi có tiếp xúc với hiệu bánh N.T được minh họa ở trường hợp bệnh nhân 1. Bệnh nhân 3 và 4 cho thấy chỉ có mua bột mì của cửa hàng thôi cũng bị viêm gan truyền nhiễm. Bởi vì những bệnh nhân này không hề sử dụng nước của nhà máy nước N.T hoặc ăn ở bất kỳ khách sạn hoặc nhà hàng nào của N.T. Tất cả 4 bệnh nhân này đều không có tiền sử tiếp xúc với những người đã bị viêm gan hoặc vàng da trước đó. Nhưng ai có tiền sử nhiễm khuẩn hoặc truyền máu trong vòng 6 tháng trước khi bệnh khởi phát và không ai trong thời gian gần đây có tiền sử ăn sò hến cả.

Câu hỏi 14: Bước tiếp theo trong cuộc điều tra của bạn sẽ là gì?

Một trong số những bệnh nhân của quận L. là một người giúp việc cho ông chủ hiệu bánh mì N.T. Đó là một người đàn ông da trắng 34 tuổi, không được khỏe mạnh lắm xét về cả mặt thể chất lẫn tinh thần. Ngày 6/4/1968, người này đến gặp thầy thuốc và phàn nàn rằng "cảm thấy lạnh và nôn". Hai ngày sau đó vợ của anh ta cũng đến khám đúng người thầy thuốc đó và cũng kêu rằng "buồn nôn và nhức đầu". Anh ta vẫn tiếp tục làm việc cho đến ngày 11/4 mới nghỉ việc vì lúc này anh ta được chẩn đoán xác định là mắc viêm gan truyền nhiễm. Những người cùng làm việc với anh tại cửa hàng nói rằng anh ta đi nước tiểu xầm màu ít nhất là 4 ngày trước khi nghỉ việc. Cho đến ngày 23/4 anh ta mới trở lại làm việc. Điều cần lưu ý là: Người giúp việc cho ông chủ hiệu bánh là ca bệnh đầu tiên (xem hình 2 trường hợp đầu tiên biểu thị màu đen).

Câu hỏi 15: a. Đường cong vụ dịch (hình 2) có chỉ ra thời kỳ ủ bệnh của bệnh viêm gan truyền nhiễm không?

b. Đường cong này có phù hợp với đặc điểm của một vụ dịch có chung một nguồn lây nhiễm không?



Câu hỏi 16: Biết rằng virus gây viêm gan truyền nhiễm bị diệt bởi sức nóng, vậy phải điều tra gì thêm để xác nhận nguồn lây truyền virus?

6. Điều tra hiệu bánh N.T.

Hiệu bánh N.T đã phục vụ cho khu vực này trong suốt 34 năm qua. Cửa hàng sản xuất các loại bánh mì, các loại bột, bánh ngọt, bánh kem. Bên cạnh những thứ bán ở quầy hàng, hiệu bánh còn cung cấp các loại kẹo, mứt và một phần bánh mì cho các khách sạn thuộc khu vực N.T và một số cửa hàng đồ khô ở quận L.

Trên thực tế, người giúp việc cho ông chủ hiệu bánh tham gia vào tất cả các công đoạn của quá trình làm bánh. Đặc biệt anh ta chịu trách nhiệm công đoạn làm bột kem và tráng bánh mứt hoa quả. Qua quan sát, các nhà điều tra phát hiện ra rằng: kem được hòa lẫn bột bằng tay và công đoạn tráng bánh mứt hoa quả cũng được thực hiện bằng tay. Vì rằng sau giai đoạn này không được hấp lại, do đó chắc chắn những công đoạn này là khởi điểm bị nhiễm khuẩn.

Cả bột làm kem và bột hoa quả để tráng lên bánh được giữ trong vài ngày và các lò làm bánh cũ được sử dụng cũng như các lò mới, không có sự phân biệt các sản phẩm của hiệu bánh không chỉ bán ra trong ngày hôm đó mà còn có thể bán trong những ngày sau đó hoặc được để đông lạnh để rồi bán trong vòng một đến hai tuần sau. Như vậy là thức ăn đã bị nhiễm virus được bán trong vòng vài ngày hoặc vài tuần. Vào giữa giai đoạn của vụ dịch, lúc này người ta đã khẳng định rằng hiệu bánh đúng là nguồn gốc gây dịch. Các nhà điều tra tiến hành lấy máu của từng người làm việc trong hiệu bánh để xác định xem liệu còn có ca viêm gan nào vào thời gian này không?

Người ta kiểm tra nồng độ men SGPT trong từng mẫu máu và kết quả cho thấy tất cả các mẫu máu đều bình thường. Bởi vì đường cong vụ dịch cũng cho thấy vụ dịch kết thúc vào thời điểm này (3/7). Cũng bởi vì không thấy có mẫu máu nào có nồng độ men SGPT gia tăng nên người ta cho phép hiệu bánh được mở cửa lại.

Câu hỏi 17: Bạn có đồng ý với quyết định này không? Hãy đưa ra quan điểm của bạn nhận định về quyết định này.

Câu hỏi 18: Vì rằng không có người công nhân nào làm việc tại hiệu bánh bị bệnh, vậy thì tại sao người ta lại phải tiến hành xét nghiệm men SGPT?

7. Biện pháp khống chế dịch.

Ngay lập tức tất cả mọi người dân sống trong khu vực N.T được tiêm gamma globulin và sau ngày 3/7/1968, tổng cộng có từ 7000 đến 8000 liều được phát ra.

Câu hỏi 19:

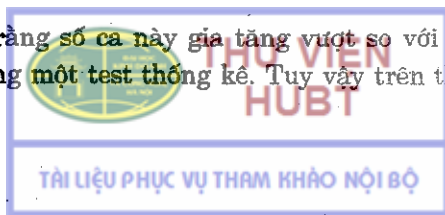
- Tại sao lại phải tiêm gamma globulin mặc dù vụ dịch đã kết thúc?
- Làm thế nào để có thể đánh giá hiệu quả của biện pháp phòng chống này?

Trả lời bài tập.

1. Người ta không thể xác định được liệu 32 ca mắc vàng da có thể lập nên một vụ dịch hay không, trừ khi người ta biết được rằng có bao nhiêu ca dự đoán cho khu vực đó trong cùng thời gian đó.

2. a. Phải

b. Để khẳng định rằng số ca này gia tăng vượt so với số ca dự đoán có thể có được, người ta có thể áp dụng một test thống kê. Tuy vậy trên thực tế thông thường số ca mắc



hàng nam tăng gấp 4 lần trong một thời gian ngắn thì có thể tiến hành mở một cuộc điều tra dịch.

3. Thông tin cần thiết phải có thêm là những số liệu ở dạng giống như thế cho nhóm đối tượng khỏe mạnh.

4. Chúng ta có thể giả định rằng những trường hợp viêm gan này tất cả đều tiếp xúc trong tuần đầu của tháng 4, vào khoảng 30 ngày trước đỉnh điểm của vụ dịch.

5. a. Không có ca bệnh nào xảy ra ở trẻ dưới 5 tuổi.

b. Khoảng tuổi từ 10-19 có tỷ lệ tấn công cao nhất.

c. Tỷ lệ tấn công cao nhất xảy ra ở nam giới. Các ca bệnh này xảy ra một khoảng thời gian nhất định. Dường như dịch xảy ra là do nguồn lây truyền chung nào đó, mà trẻ em ở tuổi chưa đi học thì không có tiếp xúc với yếu tố nguy cơ bị bệnh trong vụ dịch này.

6. Trừ 8 trường hợp còn lại, còn tất cả đều xảy ra ở trường công cộng. Điều này gợi ý rằng nguồn nhiễm bệnh của hầu hết các trường hợp có thể liên quan với sự thường trú ở tại trường.

7. Tuổi và nơi ở có thể đưa ra những đầu mối để tìm ra cơ chế lan truyền và nguồn lây nhiễm dẫn tới vụ dịch. Tuổi còn biểu thị tình trạng miễn dịch với những tiếp xúc trước đó.

8. Dịch xảy ra chủ yếu tại trường công cộng N.T và tập trung vào các học sinh cấp II và cấp III.

9. a. Từ bảng 5 chúng ta không thể rút ra được kết luận gì, vì không có nhóm đối chứng bao gồm những người khỏe mạnh và không có tiền sử tiếp xúc. Nhưng chúng ta có thể giả định rằng: Một trong những nơi được coi là nguồn gốc của vụ dịch sẽ có tỷ lệ tấn công cao. Chẳng hạn hiệu an Nữ hoàng N.T, hiệu bánh N.T và nhà máy nước N.T. Có thể tất cả mọi người dù ốm hay khỏe đều sử dụng nước và bánh mì, do đó giải thích tại sao các tỷ lệ tấn công lại cao. Do vậy phải điều tra thêm, chi tiết về 3 nguồn này.

b. Không.

10. a. Các nhà điều tra đã sử dụng các thành viên khỏe mạnh trong các hộ gia đình làm nguồn đối chứng và thấy không có sự khác biệt trong việc sử dụng nguồn nước và tiếp xúc ăn uống tại hiệu an Nữ hoàng. Nhưng họ đã tìm thấy sự khác biệt rõ rệt giữa những người ốm và tiếp xúc ăn uống tại hiệu bánh N.T

b. Về tiền sử sử dụng nước, người ta thấy cả 2 nhóm đều như nhau vì rằng tất cả các thành viên trong từng hộ gia đình đều sử dụng nước tương tự như nhau. Trừ khu vực ngoại thành, người ta có sử dụng nước giếng.

11. Sữa không phải là nguồn lây nhiễm ở đây, bởi vì: Người ta không hề phát hiện ra trường hợp mắc bệnh nào xảy ra ở trẻ em là đối tượng sử dụng sữa nhiều nhất. Hơn nữa là không hề có vụ dịch nào tương tự xảy ra ở các quận bên cạnh, nơi mà sữa cũng được bán và sử dụng đều từ một nguồn sản xuất.

12. Hình vẽ số 3 cho thấy rằng tỷ lệ mắc bệnh cao ở nhóm trẻ từ 10-19 tuổi và thời gian mắc cao nhất vào lúc ngay sau ăn trưa và sau buổi tối trường.

13. Có, vì rằng hầu hết các ca bệnh, nhất là những trường hợp mắc bệnh không phải là dân thuộc khu vực này, cũng đều có tiếp xúc với sản phẩm sản xuất ra của hiệu bánh.

14. Bước tiếp theo cần phải làm đó là cần phải cố gắng xác định nguồn truyền nhiễm

trong cửa hàng bánh.

15. a. Từ hình vẽ đường cong của vụ dịch, có thể suy luận ra thời kỳ ủ bệnh khoảng từ 24 đến 27 ngày.

b. Có.

16. Chắc chắn rằng nếu người giúp việc thực hiện khâu làm bánh thì sau đó bánh không được hấp lại.

17. Cũng như tất cả những sự kiện đã xảy ra theo thời gian của cuộc điều tra thì vụ dịch cũng đã đến giai đoạn kết thúc. Không hề có dấu hiệu về sự tồn tại của bệnh trong các công nhân của hiệu bánh. Tuy nhiên hiệu bánh cũng cần phải thấy rõ rằng giai đoạn làm bánh bằng tay sau đó không hấp lại là giai đoạn nguy hiểm dễ gây cho bánh nhiễm vi khuẩn gây bệnh để hạn chế tối thiểu hậu quả có thể xảy ra.

18. Hàm lượng men SGPT có thể cho biết mức độ chức năng gan bị phá hủy nên có thể chỉ ra trường hợp viêm gan chưa có biểu hiện lâm sàng.

19. a. Người ta sử dụng globulin để phòng trường hợp mắc bệnh viêm gan thứ phát sau đó. Có một số trường hợp bệnh xảy ra sau ngày 15/5 vì việc sử dụng gamma globulin bắt đầu sớm nhất vào đầu tháng 7. Vụ dịch có thể kéo dài một vài tháng sau nhưng chỉ là những trường hợp lác đác thông qua những tiếp xúc trong gia đình.

b. So sánh những cá thể có tiếp xúc với nguy cơ bị bệnh viêm gan mà sử dụng gamma globulin và không sử dụng gamma globulin với nhau nhằm để xác định những trường hợp bị bệnh viêm gan có biểu hiện lâm sàng và chưa có biểu hiện lâm sàng.

NGHIÊN CỨU BỆNH CHỨNG

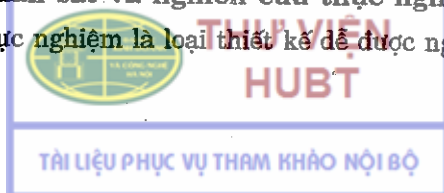
I. MỤC TIÊU

1. Phân biệt nghiên cứu bệnh chứng với nghiên cứu thuần tập (đặc biệt là phân biệt với nghiên cứu thuần tập hồi cứu).
2. Mô tả phương pháp nghiên cứu bệnh chứng.
3. Ghi nhận được những ưu điểm và đánh giá được vai trò của các loại sai số trong nghiên cứu bệnh chứng.

II. NỘI DUNG

1. Nghiên cứu quan sát và nghiên cứu thực nghiệm.

Các nghiên cứu thực nghiệm là loại thiết kế dễ được người ta nhận biết nhất, vì trong



những nghiên cứu này, nhà nghiên cứu đã kiểm soát một yếu tố tác động nào đó, và khi thay đổi yếu tố này, người ta có thể nhận ra sự thay đổi khác nhau hậu quả thu được. Về cơ bản, nghiên cứu này thường tiến hành trên động vật.

Thí dụ: Nghiên cứu về chế độ ăn, người ta có thể kiểm soát chế độ ăn uống và đo lường tỷ lệ sinh sản phát triển. Tuy nhiên trong các nghiên cứu trên người, những vấn đề đạo đức đặt ra đã giới hạn khả năng áp dụng cách đề cập nghiên cứu thực nghiệm. Với cách đề cập nghiên cứu quan sát người ta không hề có một cố gắng tác động nào, mà các kết quả khác nhau được quan sát dưới những điều kiện tự nhiên đối với các tiếp xúc khác nhau. Những đối tượng có một đặc điểm nào đó, chẳng hạn nhóm máu A, hoặc có tiếp xúc với một yếu tố nào đó, chẳng hạn dùng thuốc tránh thai, được người ta xem xét một bệnh nào đó và so sánh với nhóm không có đặc điểm hoặc không tiếp xúc. Khó khăn hay gặp trong loại nghiên cứu này là các nhóm nghiên cứu bằng các cách hỏi khác nhau có thể cho kết quả khác nhau, điều này có thể dẫn đến sự sai lầm trong việc so sánh giữa hai nhóm.

Để nhận biết một nghiên cứu thuộc loại quan sát hay thực nghiệm, hãy tìm vai trò can thiệp của người nghiên cứu trong nghiên cứu đó. Nếu có can thiệp, nó được xếp vào loại nghiên cứu thực nghiệm. Nếu không, mà đại đa số các nghiên cứu trên người là thuộc loại này, nó được xếp vào loại nghiên cứu quan sát.

2. Trình tự điều tra tìm căn nguyên của bệnh.

Trước hết, nhà lâm sàng tiến hành một quan sát hướng về vấn đề tìm căn nguyên, dựa trên kinh nghiệm của họ. Ví dụ: năm 1941, Gregg, một nhà nhân khoa người Úc, báo cáo một hội chứng mới về mù bẩm sinh, và liên hệ nó với tình trạng mắc rubella của người mẹ trong khi mang thai. Sau chiến tranh thế giới thứ hai, các nhà ngoại khoa lồng ngực lưu ý đến hiện tượng: nhiều bệnh nhân mắc ung thư phổi là những người nghiện thuốc lá nặng. Những quan sát lâm sàng như vậy đóng vai trò như những gợi ý nghiên cứu, và dựa trên đó người ta tiến hành lập ra các giả thiết. Sau đó tiến hành kiểm định các giả thiết này hoặc bằng các nghiên cứu hồi cứu, và nếu kết quả của các nghiên cứu này phù hợp, thì tiếp tục kiểm định bằng cách sử dụng nghiên cứu tương lai. Qua đó xác định được các yếu tố nguy cơ và sau đó người ta thực hiện các thử nghiệm can thiệp để đánh giá xem liệu sự thay đổi của những yếu tố nguy cơ có kéo theo việc thay đổi số trường hợp mắc bệnh hay không.

3. Thiết kế nghiên cứu bệnh chứng.

Đại đa số các nghiên cứu dịch tễ học phân tích phải hoàn thành bảng 2x2.

Bảng 2x2:

		Bệnh	
		mắc bệnh	không mắc (chưa)
Tiếp xúc	có tiếp xúc	a	b
	không tiếp xúc	c	d

"Tiếp xúc" được sử dụng ở đây mang một ý nghĩa tổng quát, bao gồm cả sự có mặt

của một đặc tính chẳng hạn như cao huyết áp, hoặc bất kỳ yếu tố bệnh nguyên nào khác.

"Bệnh" ở đây mang ý nghĩa rộng hơn, bao hàm cả các kết quả xảy ra như khỏi bệnh hoặc sống sót.

Sử dụng cách đề cập bệnh chứng, các nhà nghiên cứu bắt đầu chọn a+c các trường hợp bệnh. Rồi chọn nhóm so sánh gồm b+d các trường hợp không mắc bệnh.

Sau khi đã chọn hai nhóm mắc bệnh và không mắc bệnh, người ta xem xét tình trạng tiếp xúc và không tiếp xúc với yếu tố nguy cơ đối với từng cá thể một. Sau đó phân bố các đối tượng nghiên cứu vào hàng tiếp xúc, hay không tiếp xúc, lúc này ta đã hoàn thiện được bảng 2x2 ở trên.

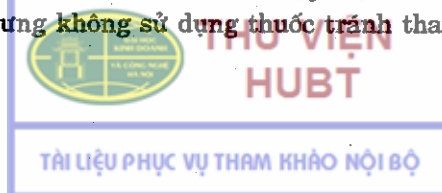
Bước tiếp theo thực hiện phân tích kết quả bằng cách so sánh các tỷ lệ tiếp xúc giữa các nhóm bệnh và đối chứng.

$$\begin{aligned} \text{Tỷ lệ tiếp xúc trong nhóm bệnh} &= \frac{a}{a+c} = Se \\ \text{Tỷ lệ tiếp xúc trong nhóm} & \\ \text{đối chứng không mắc bệnh} &= \frac{b}{b+d} = Sp \end{aligned}$$

Ưu điểm của thiết kế nghiên cứu bệnh chứng là nó có thể sử dụng để nghiên cứu bệnh hiếm. Bệnh nhân mắc những bệnh này được người ta thu thập từ trước lấy từ các bệnh viện và so sánh với nhóm đối chứng là những người không mắc bệnh. Ngược lại, thiết kế nghiên cứu tương lai, trong điều kiện ban đầu là các đối tượng được chọn đều không mắc bệnh rồi theo dõi các đối tượng này theo thời gian để quan sát sự phát triển của bệnh. Tuy nhiên, đối với một bệnh hiếm, số mắc sẽ là rất khó. Hơn nữa, nghiên cứu bệnh chứng đưa ra kết quả trong một thời gian tương đối ngắn. Vấn đề này là rất quan trọng, nếu như hậu quả xem xét lại là một bệnh đang được cho là do tiếp xúc với thuốc gây nên. Nhưng cái giá phải trả cho hai điểm thuận lợi này là xác suất sai chệnh trong nghiên cứu bệnh chứng lớn hơn so với cách đề cập nghiên cứu theo dõi trong tương lai.

4. Sai số hệ thống.

Sai số hệ thống gây ra ước lượng cao hơn hoặc thấp hơn về độ mạnh của sự kết hợp. Độ tin cậy của bất kỳ nghiên cứu nào phụ thuộc vào độ chính xác theo dõi các đối tượng được phân bố vào 4 nhóm a, b, c, d (bảng 2x2). Có thể xảy ra sự phân bố nhầm bởi vì việc chẩn đoán lẫn bỏ sót. Nếu một bệnh thực thể có chẩn đoán rõ ràng, chẳng hạn như ung thư phổi, tiêu chuẩn chẩn đoán được xác định thống nhất. Đại đa số các trường hợp bệnh đều phải có đủ các tiêu chuẩn chẩn đoán y học, và thầy thuốc lâm sàng ít có sự chọn lọc riêng trong việc săn sóc điều trị bệnh nhân, thì sự phân bố nhầm giảm xuống tới thiểu. Nhưng nếu nghiên cứu hiện tượng nghẽn mạch chẳng hạn, thì ngược lại. Bệnh này khó chẩn đoán, có thể biểu hiện như là một biến chứng của bệnh nội khoa hoặc ngoại khoa khác, và tiêu chuẩn chẩn đoán chưa được thống nhất. Tương tự như vậy việc xác định có tiếp xúc hay không có tiếp xúc với yếu tố nguy cơ cũng có thể ảnh hưởng tới việc phân loại. Nếu thầy thuốc gặp một người phụ nữ trẻ, có sử dụng thuốc tránh thai, hiện đang có điểm đau ở chân, thì sẽ có thể cho nhập viện dễ dàng hơn so với bệnh nhân cùng mắc bệnh như vậy nhưng không sử dụng thuốc tránh thai. Nhóm đối chứng đại diện bởi



các đối tượng trong cột b+d. Điều cần thiết là phải xác định không có mắc bệnh ở nhóm này, trong trường hợp ung thư phổi thì vấn đề chẩn đoán sẽ trở nên đơn giản hơn nhưng sẽ khó chẩn đoán trong trường hợp bệnh là viêm tắc mạch. Số đối tượng bệnh được chọn nghiên cứu có thể là không đại diện cho tình trạng bệnh trong quần thể. Những trường hợp được nhập viện có thể đã loại ra các trường hợp bệnh nhẹ hoặc tử vong trước khi được nhập viện. Bệnh mạch vành là một thí dụ.

Mô hình thiết kế bảng 2x2 đối với nghiên cứu hồi cứu cho thấy là việc xác định có tiếp xúc hoặc không tiếp xúc, đôi khi ở cả hai nhóm mắc bệnh và không mắc bệnh có thể loại trừ lẫn nhau. Khó mà định nghĩa và đo đạc sự tiếp xúc, ví dụ những người nhóm máu A, hoặc khó khai thác để cho đối tượng nhớ lại chính xác, ví dụ trong việc dùng thuốc. Hơn nữa tiếp xúc có thể không xảy ra liên tục, ví dụ dùng thuốc tránh thai. Do đó hàng a+b có thể là cao hơn hoặc thấp hơn so với con số thực trong mẫu.

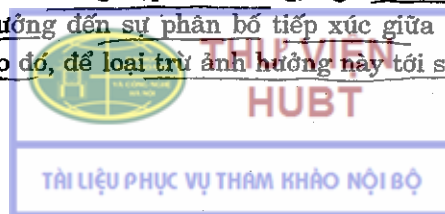
Nếu nguồn thông tin cố định và có sẵn, chẳng hạn như nhóm máu, thì độ sai số hệ thống là nhỏ tới thiểu. Còn thông thường hơn, thông tin cần thiết thường không có sẵn và chỉ thu được qua câu hỏi hoặc điều tra phỏng vấn. Việc gợi nhớ lại những sự kiện đã qua, xảy ra trong quá khứ có thể không còn nhớ được chính xác, hoặc là các thông tin bổ sung từ nguồn không chính thức khác cũng bị sai lệch.

Việc nhớ lại có chọn lọc các thông tin cũng có thể xảy ra đối với những đối tượng bị mắc bệnh. Vì biết mình mắc bệnh, và cũng có thể đã biết tình trạng mắc bệnh này có liên quan với một yếu tố nguy cơ, do đó có thể dẫn tới tình huống người bệnh trả lời nhớ lại việc tiếp xúc với yếu tố nguy cơ quá mức hoặc ngược lại. Người điều tra biết đối tượng nào là mắc bệnh và đối tượng nào là đối chứng, có thể điều tra nhiều hơn đối với những trường hợp bệnh để tìm sự tồn tại của mối quan hệ giữa yếu tố nguy cơ và mắc bệnh. Để hạn chế tới thiểu sai lệch từ nguồn này, lý tưởng nhất là người điều tra không hề biết trước đối tượng được phỏng vấn thuộc nhóm nào, nhưng điều này lại khó mà thực hiện được trong nghiên cứu hồi cứu.

Độ sai lệch có thể xảy ra trong việc chọn nhóm đối chứng, đặc biệt nếu nhóm bệnh được chọn từ các bệnh nhân nằm viện. Điều cần bản là nhóm đối chứng phải càng giống nhóm bệnh càng tốt, tuy nhiên cũng phải tương tự với quần thể dân nói chung trong việc tiếp xúc, nếu như kết quả nghiên cứu mong muốn suy rộng cho cả quần thể. Các đối tượng trong nhóm chứng mà được chọn từ bệnh viện có thể chứa một tỷ lệ không đại diện xét theo một đặc điểm nhất định, ví dụ cao huyết áp và hút thuốc lá. Nếu chọn nhóm chứng trong bệnh viện mà mắc bệnh hệ thống nghiêm trọng chẳng hạn như bệnh ung thư, thì sự có mặt của bệnh này có thể ảnh hưởng tới việc xem xét tiếp xúc với yếu tố nguy cơ nghi ngờ là nguyên nhân của bệnh.

5. Ghép cặp (Matching)

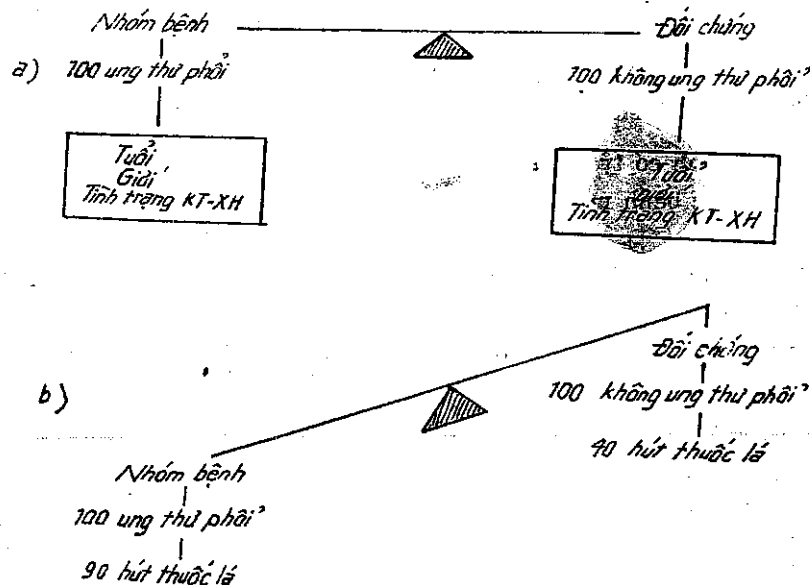
So sánh giữa các trường hợp bệnh và đối chứng có thể phát hiện ra một sự khác biệt về tỷ lệ tiếp xúc, và sự phát triển bệnh, có thể được quy cho sự khác biệt này, với điều kiện là các yếu tố khác là tương tự giữa 2 nhóm nghiên cứu. Để đạt được tính có thể so sánh giữa các trường hợp bệnh và các trường hợp đối chứng, người ta thường ghép cặp các đặc điểm mà cho rằng có ảnh hưởng đến sự phân bố tiếp xúc giữa 2 nhóm. Thông thường tuổi là một yếu tố như vậy. Do đó, để loại trừ ảnh hưởng này tới sự so sánh người



ta ghép cặp các trường hợp bệnh và trường hợp chứng về yếu tố tuổi. Tình trạng kinh tế xã hội, thông thường ảnh hưởng tới tỷ lệ tiếp xúc, bởi vì nó ảnh hưởng đến các yếu tố môi trường và kiểu cách sống, và do đó chúng cũng được ghép cặp giữa nhóm bệnh và nhóm đối chứng. Cũng cần phải nhấn mạnh là, một khi ghép cặp một yếu tố nào đó thì người ta không thể phát hiện vai trò căn nguyên của nó bởi vì lúc này nhóm bệnh và đối chứng là tương tự nhau xét về đặc điểm đó.

6. Kiểm định giả thiết.

Chúng ta hãy xét một giả thiết là hút thuốc lá liên quan đến ung thư phổi. Lấy 100 bệnh nhân ung thư phổi, và rồi chọn 100 đối chứng từ quần thể không mắc ung thư phổi. Hai nhóm được ghép cặp với nhau theo tuổi, giới tính và tình trạng kinh tế xã hội. Kết quả được minh họa trong hình vẽ dưới đây (hình 1). Bây giờ việc tìm hiểu thói quen hút thuốc của hai nhóm, hình 1 miêu tả sự không cân xứng trong kết quả thu được. Bạn đã thấy một sự kết hợp giữa hút thuốc lá và ung thư phổi.



Hình 1- Kiểm định một giả thiết bằng cách ghép cặp nhóm bệnh và nhóm chứng (a) và rồi điều tra về yếu tố căn nguyên nghi ngờ (b).

7. Phân tích kết quả.

Bắt đầu nghiên cứu bằng 2x2 ta có bảng sau đây:

	Nhóm bệnh	Đối chứng
Hút thuốc lá	?	?
Không hút thuốc lá	?	?
Tổng cộng	100	100

phổi mãn

Trong lúc nghiên cứu, chúng ta xác định ngược trở về quá khứ số người hút thuốc và không hút thuốc ở cả hai nhóm bệnh và đối chứng. Do vậy, chúng ta đi tìm a, b, c và d

odd ratios (OR)

cho bảng 2x2, giả sử ta có số liệu tìm được như sau:

	Nhóm bệnh	Đối chứng
Hút thuốc lá	90 ^a	40 ^b
Không hút thuốc lá	10 ^c	60 ^d
Tổng cộng	100	100

Bây giờ chúng ta so sánh tỷ lệ tiếp xúc ở nhóm bệnh và nhóm đối chứng:

Tỷ lệ tiếp xúc:

$$\text{Nhóm bệnh} = \frac{a}{a+c} = \frac{90}{100} \quad \text{hay } 90\%$$

$$\text{Đối chứng} = \frac{b}{b+d} = \frac{40}{100} \quad \text{hay } 40\%$$

8. Ước lượng nguy cơ tương đối.

Do bản chất hồi cứu của nghiên cứu này chúng ta không thể rút ra tỷ lệ mới mắc bệnh ở nhóm hút thuốc hoặc không hút thuốc. Tuy nhiên, có thể ước lượng nguy cơ tương đối với hai điều kiện dưới đây:

1. Bệnh có tỷ lệ mới mắc thấp hơn ở quần thể dân chúng. Điều này là đúng với đại đa số các bệnh mạn tính.
2. Nhóm chứng đại diện cho quần thể dân chúng xét về tần số xuất hiện đặc tính nghiên cứu.

Tỷ suất chênh lệch (odd ratios = OR) là con số ước lượng nguy cơ tương đối trong các thiết kế nghiên cứu:

$$\text{Tỷ suất chênh lệch OR} = \frac{ad}{bc} = \frac{a/c}{b/d} = \frac{a/b}{c/d}$$

Trong ví dụ này tỷ suất chênh lệch (OR) là:

$$\text{OR} = \frac{90 \times 60}{10 \times 40} = 13,5$$

III. CÁCH TIẾN HÀNH MỘT NGHIÊN CỨU BỆNH CHỨNG

A. NHẮC LẠI MỘT VÀI KHÁI NIỆM CẦN THIẾT

1. Nghiên cứu bệnh chứng là một nghiên cứu phân tích dịch tễ học, nó sử dụng để kiểm định một giả thuyết, nên bao giờ cũng phải có nhóm đối chứng (để so sánh với nhóm chủ cứu).
2. Nghiên cứu bệnh chứng thuộc loại nghiên cứu hồi cứu, nghĩa là nó được tiến hành sau khi bệnh đã xảy ra xong đâu đấy (nhóm chủ cứu).

3. như vậy điểm xuất phát của nghiên cứu bệnh chứng là căn cứ vào bệnh, nghĩa là khi xây dựng nghiên cứu, phải căn cứ vào sự kiện là có bệnh (để chọn ngẫu nhiên vào nhóm chủ cứu) và không có bệnh (để chọn ngẫu nhiên vào nhóm đối chứng). Và công việc sau đó mới là khai thác lại trong tiền sử sức khỏe của họ (mọi cá thể ở cả 2 nhóm) xem tình trạng phơi nhiễm với yếu tố nguy cơ cần nghiên cứu của bệnh như thế nào.

4. Nghiên cứu bệnh chứng thuộc loại hình nghiên cứu quan sát, nghĩa là người nghiên cứu không có can thiệp gì vào cả quá trình của bệnh lẫn quá trình phơi nhiễm với yếu tố nguy cơ.

5. Vì là một nghiên cứu hồi cứu, đòi hỏi người được vào nghiên cứu phải nhớ lại về việc họ có bị phơi nhiễm với yếu tố nguy cơ không, họ đã bị phơi nhiễm như thế nào (cả về định tính và định lượng) nên sai số nhớ lại hoặc nghiên cứu này có xác suất lớn, nhất là đối với những bệnh có thời kỳ phơi nhiễm dài mới có thể mắc (phần lớn là những bệnh mạn tính). Các thầy thuốc làm nghiên cứu bệnh chứng phải rất chú ý đến loại sai số này. Còn các loại sai số khác, cũng có thể mắc phải thì như trong các thiết kế nghiên cứu khác.

6. Nghiên cứu bệnh chứng có thể tiến hành trên các sổ sách, nhất là trên các hồ sơ bệnh án bệnh lịch... Trường hợp này cần chú ý là các loại sổ sách hồ sơ đó có tin cậy được không, có được ghi chép đầy đủ không (vì thường những sổ sách hồ sơ đó không nhằm phục vụ cho nghiên cứu lúc này). Nếu chúng đầy đủ, thì có ưu điểm là rất khách quan với nghiên cứu (giảm được sai số chẩn đoán khá nhiều) nhưng nếu không đầy đủ thì không sử dụng được.

7. Việc chọn các cá thể vào nhóm đối chứng có tầm quan trọng lớn đến kết quả nghiên cứu, và rất khó chọn được ra đủ một nhóm đối chứng sát hợp, cần bỏ ra nhiều suy nghĩ để cân nhắc đầy đủ.

8. Cũng cần nêu là, nếu nghiên cứu bệnh chứng được tiến hành trong quần thể bằng phỏng vấn, thì người thực hiện cuộc phỏng vấn phải được tập huấn đầy đủ, và trung thực với trả lời của người được hỏi, nếu không sai số do người phỏng vấn gây nên cũng sẽ lớn làm ảnh hưởng không nhỏ đến kết quả nghiên cứu.

B. MỘT BÀI TẬP VỀ NGHIÊN CỨU BỆNH CHỨNG

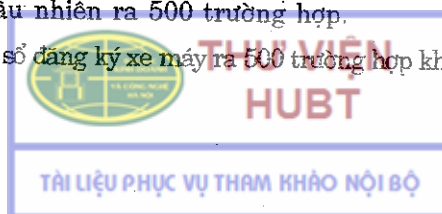
Trong năm 1993, người dân Hà Nội có nhận xét rằng tai nạn xe máy có tăng lên, nhưng vẫn là nhiều ở những người điều khiển phương tiện còn quá trẻ: dưới 18 tuổi. Một nhà dịch tễ học muốn kiểm định giả thuyết này bằng một nghiên cứu bệnh chứng. Được biết rằng có tới 20% số người điều khiển xe máy trẻ đã gây những tai nạn giao thông lớn nhỏ. Và nghiên cứu được định ở mức tin cậy 95%, với $OR \geq 2$, và độ chính xác tương đối là 0,25p.

- Tính cỡ mẫu: 500 cho mỗi nhóm.

- Chọn cá thể vào nghiên cứu:

+ Xìn toàn bộ danh sách tai nạn xe máy trong năm 1993 vào các bệnh viện ở Hà Nội, lập danh mục, rút ngẫu nhiên ra 500 trường hợp.

+ Rút ngẫu nhiên từ số đăng ký xe máy ra 500 trường hợp không gây tai nạn gì trong năm 1993.



- Tính kết quả:

Giả sử ta có bảng 2x2 sau đây:

	Có gây tai nạn	Không gây tai nạn	Tỷ lệ gây tai nạn
Dưới 18 tuổi	150	350	150/500
18 tuổi trở lên	50	450	50/500
Tỷ lệ lái xe trẻ	150/200	350/800	1.000

Tìm ý nghĩa thống kê bằng trắc nghiệm $\chi^2=61,25$

Tra bảng χ^2 : Nghiên cứu "có ý nghĩa thống kê" ở mức $p < 0,05$ đã định. Ta loại bỏ H_0 và chấp nhận H_1 : tai nạn giao thông có kết hợp thống kê với tuổi của người điều khiển xe máy ở ngưỡng xác suất 0,05.

Tỷ lệ tai nạn ở lái xe trẻ: 0,30

Tỷ lệ tai nạn ở lái xe 18 tuổi trở lên: 0,10

Và có $OR=3,86$

Các vụ tai nạn giao thông do người lái xe máy trẻ dưới 18 tuổi có nguy cơ tương đối ước lượng xấp xỉ gấp 4 lần do người lái xe máy từ 18 tuổi trở lên.

IV. BÀI TẬP

Trong các nghiên cứu bệnh chứng:

1- a. Nhóm mắc bệnh (case) là 100 phụ nữ mắc bệnh ung thư vú. Cần phải chọn nhóm đối chứng (control) là nhóm gì?

b. Giả thiết của bạn có thể là: mang thai lần đầu tiên khi tuổi còn trẻ thì có khả năng ngăn ngừa ung thư vú. Có nên ghép cặp về yếu tố tình trạng kinh tế xã hội ở nhóm mắc bệnh và nhóm đối chứng hay không? Hãy cho biết lý do.

c. Liệu có khả năng kiểm định giả thiết rằng tỷ lệ ung thư vú ở phụ nữ độc thân cao hơn phụ nữ có chồng, bằng cách sử dụng 100 trường hợp bệnh và 100 trường hợp đối chứng không? /

e. Giả sử là 90% nhóm chứng là đã lập gia đình. Hãy ước lượng nguy cơ tương đối mắc ung thư vú cho phụ nữ độc thân.

2- Giả thiết của bạn là những người đàn ông làm nghề lái xe mà nghiện rượu có tỷ lệ bị tai nạn xe hơi cao. Hãy thiết kế một nghiên cứu hồi cứu để kiểm định giả thiết này, sử dụng những gợi ý sau đây:

a. Chẩn đoán một trường hợp mắc bệnh khó hay không khó?

Nên chọn những trường hợp bệnh từ những nguồn nào?

b. Nêu tên của quần thể phù hợp cho nghiên cứu, từ đó ta chọn nhóm đối chứng.

c. Hãy liệt kê các đặc trưng cần ghép cặp với nhóm đối chứng.

case - control.

- d. Đặc trưng nào bạn phải xác định cho từng đối tượng nghiên cứu.
- e. Những khó khăn gì có thể được tính đến trong việc xác định sai đặc trưng nghiên cứu.

3- Một nghiên cứu bệnh chứng cho kết quả như sau:

	Tiếp xúc	Không tiếp xúc
Nhóm bệnh	200	200
Nhóm chứng	50	250

OR = $\frac{200 \cdot 250}{200 \cdot 50}$

Hãy tính tỷ suất chênh lệch OR. (Odds Ratios)

4- Một nghiên cứu hồi cứu về sự kết hợp giữa hút thuốc lá và bệnh nhồi máu cơ tim cho kết quả như sau:

	Không hút thuốc lá (không tiếp xúc)	Hút thuốc lá (tiếp xúc) số bao/ngày		
		1/2	1	2
Nhóm bệnh	31	9	39	18
Nhóm chứng	2706	710	1825	605

Hãy tính nguy cơ tương đối ước lượng hay tỷ suất chênh lệch OR bị nhồi máu cơ tim giữa nhóm không hút thuốc lá và nhóm những người hút thuốc lá ở các mức hút thuốc lá khác nhau sau đây:

- a. 1/2 bao 1 ngày
- b. 1 bao trong 1 ngày.
- c. 2 bao trong 1 ngày.

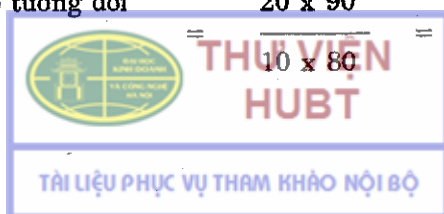
$\frac{9 \times 2706}{31 \times 710}$

Trả lời bài tập

- 1- a. Nhóm chứng là nhóm không bị bệnh ung thư vú. ✓
- b. Có, vì tình trạng kinh tế xã hội ảnh hưởng đến tuổi có thai lần đầu, do vậy nếu bạn chứng minh một sự kết hợp thì có thể sự kết hợp này là do tình trạng kinh tế xã hội. Trừ khi lúc ban đầu nghiên cứu, đặc trưng này đã được ghép cặp giữa hai nhóm.
- c. Được. Vì chúng ta xác định được tỷ lệ phụ nữ độc thân ở từng nhóm. ✓
- d. Không, chúng ta cần tiến hành so sánh tần số lập gia đình ở nhóm đối chứng.

	Nhóm bệnh (Case)	Đối chứng (Control)
<i>single</i> Độc thân	20	10
<i>married</i> Có gia đình	80	90
<i>total</i> Tổng cộng	100	100

Ước lượng nguy cơ tương đối
tỷ suất chênh lệch (OR) = $\frac{20 \times 90}{10 \times 80} = 2,25$



2- a. Không có mấy khó khăn trong việc chẩn đoán chết do tai nạn xe cộ. Có thể tìm nhóm chết do tai nạn từ sổ khám bệnh ở các phòng khám hoặc từ hồ sơ lưu trữ của cảnh sát và văn phòng thống kê tai nạn do các phương tiện tốc độ cao.

b. Nhóm chứng có thể lấy từ danh sách những lái xe khác ở trong cùng địa phương, họ chưa bao giờ gây tai nạn. Chú ý, nhóm chứng phải là lái xe, vì tình trạng nghiện rượu có thể khác biệt ở nhóm lái xe đã mất bằng lái xe.

c. Tuổi, giới, tình trạng kinh tế xã hội (hoặc bất kỳ đặc điểm nào mà bạn nghĩ rằng ảnh hưởng tới việc bị tai nạn xe cộ có thể dùng để đem ghép cặp, chỉ trừ đặc điểm không uống rượu).

d. Mức độ dùng rượu.

e. Uống rượu đối với những lái xe đã thôi nghề, có thể khó xác định, và nhóm chứng có thể đưa lại thông tin sai lạc về điểm này.

3- Tỷ suất chênh lệch OR.

$$\text{OR} = \frac{200 \times 250}{200 \times 50} = 5$$

4- a.

$$\text{OR} = \frac{9 \times 2706}{3 \times 710} = 1,1$$

b.

$$\text{OR} = \frac{39 \times 2706}{31 \times 1825} = 1,9$$

c.

$$\text{OR} = \frac{18 \times 2706}{31 \times 605} = 2,6$$

NGHIÊN CỨU THUẦN TẬP

I. MỤC TIÊU

1. Mô tả được một nghiên cứu thuần tập tương lai.
2. Phân biệt nghiên cứu thuần tập với các nghiên cứu ngang.
3. So sánh ý nghĩa, tác dụng, sử dụng nghiên cứu thuần tập tương lai với nghiên cứu bệnh chứng.



THƯ VIỆN
HUBT

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

II. NỘI DUNG

Người ta đã dùng bốn thuật ngữ khác nhau để chỉ các nghiên cứu trong tương lai là nghiên cứu tỷ lệ mới mắc (Incidence study), nghiên cứu thuần tập (Cohort study), nghiên cứu về tương lai (Prospective study), và nghiên cứu dọc (Longitudinal study).

Sử dụng tên gọi nghiên cứu thuần tập đề cập đến tính chất nhóm nghiên cứu. Nghiên cứu tỷ lệ mới mắc ngụ ý đến vấn đề thực tế là số trường hợp mới mắc có thể xác định được trong nghiên cứu này. Nghiên cứu trong tương lai chỉ ra vấn đề thực tiễn là nhóm nghiên cứu được theo dõi theo thời gian trong tương lai và ngược lại với nghiên cứu hồi cứu là quay lại thời gian đã xảy ra trong quá khứ. Tên gọi nghiên cứu dọc đưa ra vấn đề các đối tượng nghiên cứu, một khi đã được xác định, từng cá thể sẽ được theo dõi trong suốt quá trình nghiên cứu sau này.

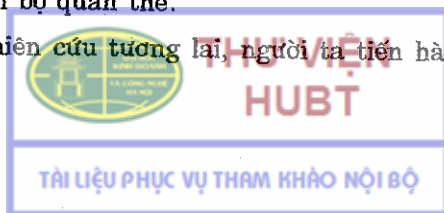
Trong nghiên cứu trong tương lai (Prospective) thời gian lúc bắt đầu nghiên cứu đòi hỏi tất cả các đối tượng tham gia nghiên cứu đều không bị mắc bệnh mà ta đang quan tâm. Trong khi đó, nghiên cứu hồi cứu, tất cả các sự kiện quan tâm (bệnh và tiếp xúc) đã xảy ra khi nghiên cứu bắt đầu.

Như ta thấy ở hình 1 dưới đây, trước hết nhà nghiên cứu tuyển chọn một nhóm người tình nguyện vào thăm khám để khẳng định rằng họ không mắc bệnh ta đang quan tâm ở thời điểm bắt đầu nghiên cứu. Rồi họ tiến hành thu thập một loạt các mục thông tin (các biến số), chẳng hạn số liệu về dân số học, tình trạng kinh tế xã hội, nghề nghiệp và tiền sử bệnh tật. Các biến này rõ ràng phải bao gồm cả những biến thực sự có liên quan tới nguyên nhân của bệnh. Thiết kế tương lai nhờ có đặc tính hết sức tự nhiên của nó, tránh được hai nguyên nhân chính gây sai lệch ở trong cách đề cập hồi cứu, đó là việc chọn đối chứng và sai lệch trong đánh giá sự tiếp xúc. Trong trường hợp này nhóm chứng gồm những người không tiếp xúc với yếu tố nghi ngờ. Do tiếp xúc nhất thiết phải xảy ra trước khi bị bệnh, nên giảm được sai lệch trong việc đánh giá tiếp xúc: cả đối tượng nghiên cứu và nhà nghiên cứu đều không biết trước ai sẽ là người mắc bệnh và ai sẽ không bị mắc bệnh.

Một trong những giới hạn chính của cách đề cập nghiên cứu tương lai nảy sinh từ vấn đề thực tế là các đối tượng nghiên cứu bắt buộc tất cả đều phải tình nguyện, và họ phải quay trở lại đều đặn để được kiểm tra xem có phát triển bệnh hay không và để xác định những thay đổi trong các yếu tố tiếp xúc đang được theo dõi. Những người làm hợp đồng thường hay được chọn bởi vì việc theo dõi họ trong một vài năm là tương đối dễ dàng, có thể sử dụng công nhân đường sắt hoặc các nhân viên làm công ăn lương của chính phủ và những người không thuộc tầng lớp thất nghiệp hoặc tầng lớp làm những nghề nghiệp không ổn định như bán hàng, khuôn vác.

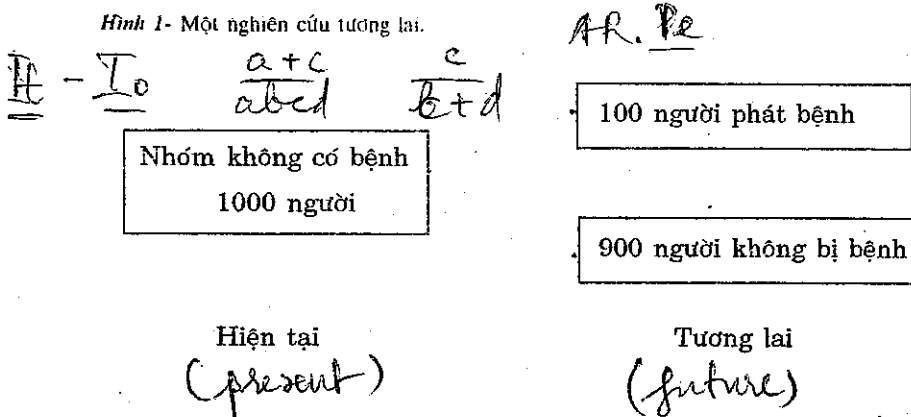
Những người thuộc các nhóm sau này có thể được tiếp xúc với nhiều loại kịch chấn (stress) khác nhau, và sự thanh lọc nghề nghiệp có thể gây ra việc tự chọn lọc trong số những người làm nghề đó. Và kết quả của nghiên cứu trên một nhóm người làm nghề không ổn định này thì giới hạn việc áp dụng cho những nhóm người làm nghề khác, không tổng quát hóa cho toàn bộ quần thể.

Như vậy, trong nghiên cứu tương lai, người ta tiến hành quan sát một nhóm người



khỏe mạnh và rồi phân loại họ xét theo các đặc trưng mà nhà nghiên cứu quan tâm. Nếu bệnh mạch vành (tỷ lệ mới mắc) được nghiên cứu, thì trước hết người ta kiểm tra các thành viên trong nhóm để loại trừ những người có bệnh mạch vành trước đó. Rồi các đối tượng nghiên cứu được theo dõi theo một cách chặt chẽ để nhận biết ai bị mắc bệnh trong quá trình nghiên cứu (Trong ví dụ này là phát triển bệnh mạch vành). Bằng cách này người ta sẽ có được tỷ lệ mới mắc bệnh cả ở những người có đặc trưng hoặc đặc điểm đang xét và cả ở những người không có đặc trưng hoặc đặc điểm này. Xét một ví dụ khác, chúng ta có thể xác định được tỷ lệ mắc bệnh ung thư phổi xảy ra trong một thời kỳ 10 năm ở những người được phân loại là có hút thuốc lá ngay từ lúc bắt đầu thời kỳ nghiên cứu và đem so sánh chung với tỷ lệ như vậy ở những người không hút thuốc lá.

Hình 1- Một nghiên cứu tương lai.



1. Những đối tượng không theo dõi được trong quá trình nghiên cứu theo đối tượng lai.

Trong đại đa số các nghiên cứu tương lai, nhất thiết các đối tượng tham gia nghiên cứu từng thời kỳ một phải được thăm khám hoặc trả lời các câu hỏi giám sát. Có một số nhỏ các đối tượng trong nhóm sẽ ra khỏi nghiên cứu vì nghiên cứu kéo dài trong nhiều năm. Nhìn chung người ta giả sử rằng các đối tượng ra khỏi nghiên cứu có tỷ lệ mắc bệnh xảy ra như nhau và tương tự với những người tiếp tục tham gia nghiên cứu. Điều giả định này có thể là sai. Những người phát triển bệnh có thể thôi không quay trở lại để điều trị, hoặc họ có thể đến gặp nhà nghiên cứu ít hơn so với những người vẫn duy trì ở trạng thái không mắc bệnh, dẫn đến việc nhận định tỷ lệ mắc thấp trong trường hợp trước và nhận định cao trong trường hợp sau. Phải khống chế sự thiếu hụt này ở mức độ tối thiểu chấp nhận được, để cho kết luận của nghiên cứu giữ được giá trị. Chính hạn chế này đã ảnh hưởng đến nghiên cứu theo dõi các nhóm đối tượng trong quần thể dân nói chung.

2. Phân tích kết quả.

Hãy xét một tập hợp 2000 người trong đó 800 hút thuốc lá và 1200 không hút thuốc. Tại thời điểm khởi đầu của cuộc nghiên cứu không ai mắc ung thư phổi.

	Ung thư phổi	Không ung thư phổi	Tổng cộng
Hút thuốc lá	a	b	800
Không hút thuốc	c	d	1200
Tổng cộng	a+c	b+d	2000



Cross sectional / Longitudinal study

Toàn bộ 2000 người này được theo dõi trong suốt 20 năm, 100 trường hợp mới xuất hiện ung thư phổi, trong đó có 90 trường hợp có hút thuốc lá và 10 trường hợp không hút thuốc lá. Bây giờ ta hoàn thành bảng 2x2 như sau:

Tiếp theo, chúng ta so sánh tỷ lệ mới mắc ung thư phổi ở những người hút thuốc lá và không hút thuốc lá:

+ Tỷ lệ mới mắc ở nhóm hút thuốc lá:

$$\frac{a}{a+b} = \frac{90}{800} = 112,5/1000$$

+ Tỷ lệ mới mắc ở nhóm không hút thuốc lá:

$$\frac{c}{c+d} = \frac{10}{1200} = 8,3/1000$$

với $p < 0,001$

(Incidence) Bệnh
phần nhiên

3. Tính nguy cơ tương đối.

Trong nghiên cứu tương lai, chúng ta có khả năng xác định tỷ lệ mới mắc một cách trực tiếp trong những người có tiếp xúc và không tiếp xúc. Do vậy, chúng ta tính nguy cơ tương đối là tỷ suất giữa hai tỷ lệ mới mắc.

$$\text{Nguy cơ tương đối (RR)} = \frac{a/a+b}{c/c+d} \quad (\text{Relative Risk})$$

Trong ví dụ này nguy cơ tương đối như sau:

$$\text{Nguy cơ tương đối (RR)} = \frac{90/800}{10/1200} = 13,5$$

4. So sánh với các nghiên cứu bệnh chứng.

Hạn chế và thuận lợi của các nghiên cứu bệnh chứng và nghiên cứu thuần tập tương lai:

Bệnh chứng

- + Thời gian nghiên cứu ngắn
- + Tương đối ít tốn kém
- + Thích hợp đối với các bệnh hiếm
- + Ít gặp phải các vấn đề về tính nhân đạo
- + Chọn nhóm chứng có thể gặp phải sai số hệ thống.
- Không cần đến đối tượng nghiên cứu là người tình nguyện.

Thuần tập tương lai

- Thời gian nghiên cứu dài
- Chi phí nghiên cứu cao
- Chỉ dùng cho các bệnh tương đối phổ biến
- Vấn đề nhân đạo có thể phải xem xét đến và ảnh hưởng đến thiết kế nghiên cứu.
- + Nhóm chứng ít tạo ra sai số hệ thống.
- + Cần đến người tình nguyện. Do vậy kết quả có thể không tổng quát hóa được.
- + Không có sai số nhớ lại.

✓ - Có sai số nhớ lại



- Số đối tượng nghiên cứu nhỏ hơn
- Không xảy ra hiện tượng bỏ dở cuộc nghiên cứu
- Không thể xác định được tỷ lệ mới mắc
- Nguy cơ tương đối có giá trị gần đúng
- + Cần số đối tượng nghiên cứu lớn.
- + Xảy ra hiện tượng đối tượng nghiên cứu có thể bỏ cuộc giữa chừng.
- + Xác định được tỷ lệ mới mắc
- + Nguy cơ tương đối có giá trị chính xác.

Nghiên cứu hồi cứu và nghiên cứu tương lai về thực chất không phải là các phương pháp hoàn thiện tuyệt đối. Các quan sát lâm sàng hoặc điều tra số hiện mắc, có thể đưa ra gợi ý ban đầu về một sự kết hợp giữa tiếp xúc với một yếu tố và hậu quả phát triển một bệnh. Giả thiết này sau đó được kiểm định bằng một nghiên cứu hồi cứu. Nếu sự kết hợp này được xác nhận, thì người ta tiến hành một nghiên cứu tương lai.

6. Nghiên cứu ngang.

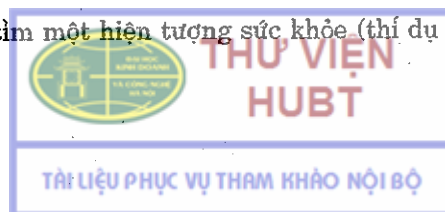
Các nghiên cứu này xác định tỷ lệ hiện mắc, không xác định tỷ lệ mới mắc. Tác dụng của chúng chủ yếu là trong việc xác định sự lưu hành của bệnh hơn là trong việc điều tra căn nguyên. Ví dụ, để đánh giá tần số mắc cao huyết áp ở thành phố Baltimore, người ta tiến hành đo huyết áp của một tập hợp dân tuổi trưởng thành sống trong thành phố đó trong năm 1973. Cuộc điều tra thu thập các mục thông tin khác như tuổi, giới, chủng tộc và nghề nghiệp. Do vậy, chúng ta có thể xác định tỷ lệ mắc cao huyết áp là bao nhiêu xét theo các thông số tiến hành bằng đo đạc. Nghiên cứu kiểu này là một ví dụ về nghiên cứu sự phân bố của một bệnh trong quần thể, hơn là về căn nguyên của bệnh. Tuy vậy, mô hình phân bố cũng có thể gợi ý cho chúng ta về một giả thuyết căn nguyên để rồi kiểm định giả thuyết này bằng nghiên cứu hồi cứu hoặc bằng nghiên cứu tương lai.

III. CÁCH TIẾN HÀNH MỘT NGHIÊN CỨU THUẦN TẬP TƯƠNG LAI

A. TÌM VẤN ĐỀ NGHIÊN CỨU

Tìm một vấn đề sức khỏe để tiến hành nghiên cứu thuần tập tương lai không khó. Trên thế giới những bệnh mạn tính phát triển nhiều, người ta thường tập trung vào đó, và phần lớn là những nghiên cứu giải thích. Ở ta, hiện tại vẫn còn nhiều sự kiện mắc và hoặc chết của một số hiện tượng sức khỏe diễn biến ngắn, phổ biến mà chưa xác định hết được nguy cơ chính xác của chúng. Chúng ta sẽ suy nghĩ về những vấn đề này với những nghiên cứu ứng dụng thực tế, vì nghiên cứu thuần tập tương lai cho những kết quả tin cậy, chính xác. Thí dụ riêng trong sự nghiệp bảo vệ sức khỏe bà mẹ và trẻ em, tỷ lệ chết mẹ do thai sản là bao nhiêu? Có những nguy cơ gì để có thể tác động đề phòng? Tỷ lệ chết trẻ con dưới 1 tuổi? dưới 3 tuổi? dưới 5 tuổi? Nguy cơ của chúng là những gì? Làm thế nào để đề phòng?...

- Xác định vấn đề nghiên cứu: tìm một hiện tượng sức khỏe (thí dụ bệnh tật) phổ biến



với nguy cơ phổ biến của nó (mà có thể tác động cải thiện đối với nguy cơ đó). Cả bệnh tật và yếu tố nguy cơ đều có thể định lượng được rõ ràng.

- Tính cỡ mẫu cần nghiên cứu:

$$n = Z_{1-\alpha/2}^2 \times \frac{\left(\frac{Q_1}{P_1} + \frac{Q_2}{P_2}\right)}{[\log_e(1 - \varepsilon)]^2}$$

(Xin xem thêm ở sách đã dẫn).

- Chọn các cá thể vào nhóm có phơi nhiễm với yếu tố nguy cơ cần nghiên cứu.
- Chọn nhóm đối chứng hợp lý với nhóm chủ cứu trên đây.
- Thăm khám cả hai nhóm này để xác định họ hiện đang không có bệnh tật định nghiên cứu.
- Tiến hành điều tra khảo sát, mô tả... theo quy trình nghiên cứu, rồi tập hợp, xử lý kết quả nghiên cứu như thường lệ (so sánh 2 nhóm).

B. MỘT BÀI TẬP VỀ MẪU NGHIÊN CỨU THUẦN TẬP TƯƠNG LAI

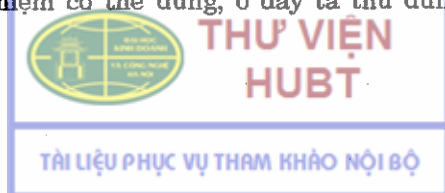
Một nhà dịch tễ học muốn nghiên cứu tỷ lệ chết trẻ con dưới 1 tuổi với những nguy cơ nhất định. Ở bài tập này, chỉ đề cập đến nguy cơ tuổi mẹ sinh con quá trẻ (trước 20 tuổi). Giả sử rằng tỷ lệ chết trẻ con chung là 0,05, và muốn nghiên cứu có mức tin cậy 95% với nguy cơ tương đối khai thác được có độ chính xác tương đối là 0,5 của tỷ lệ và yêu cầu nguy cơ tương đối này phải bằng 2 trở lên.

- Tính cỡ mẫu: 225 con của sản phụ sinh con trước tuổi 20.
- Chọn cá thể vào nghiên cứu: chọn 225 phụ nữ có thai trước tuổi 19, và 225 phụ nữ có thai sau tuổi 20 (cho dễ xác định) họ đều có mang thai lần đầu, thăm khám cho họ để ít nhất xác định họ không có các dị tật về khung chậu, không có các bệnh nguy hiểm, giáo dục cho họ tự chăm sóc thai và khám thai đúng quy định, và đến đẻ tại trạm y tế... rồi tiếp tục theo dõi định kỳ cho họ.
- Theo dõi tất cả 450 trẻ đó cho đến 1 tuổi, ghi lại số chết do mọi nguyên nhân.
- Giả sử có kết quả sau đây:

	Con chết	Con sống	Tỷ lệ
Mẹ dưới 20 tuổi sinh con	34	191	34/225
Mẹ trên 20 tuổi sinh con	11	214	11/225
Cộng	45	405	45/450

- Trước hết ta tìm ý nghĩa thống kê của kết quả này:

Có nhiều trắc nghiệm có thể dùng, ở đây ta thử dùng trắc nghiệm sai số chuẩn của khác biệt:



$$SE_{\Delta} = 2,85$$

$$\frac{\Delta}{SE_{\Delta}} = 3,59$$

Kết quả có nghĩa thống kê với $p < 0,05$ đã định, và như vậy ta loại bỏ giả thuyết H_0 , mà chấp nhận giả thuyết H_1 (nghĩa là ta chấp nhận giả thuyết cho rằng tỷ lệ chết trẻ con dưới 1 tuổi của các bà mẹ có tuổi lúc sinh con dưới 20 tuổi là khác biệt có ý nghĩa thống kê với tỷ lệ chết trẻ con dưới 1 tuổi của các bà mẹ có tuổi lúc sinh các đứa trẻ đã trên 20 tuổi).

- Tính tỷ lệ chết trẻ con dưới 1 tuổi, của:

Mẹ dưới 20 tuổi đã sinh con: $34/225 = p_1 = 15,56\%$

Mẹ trên 20 tuổi mới sinh con: $11/225 = p_2 = 5,34\%$

- Tính sai số chuẩn các tỷ lệ này, của:

Mẹ dưới 20 tuổi đã sinh con:

$$SE_{p1} = \sqrt{\frac{(15,56) \times (84,44)}{(225-1)}} = 2,42\%$$

Mẹ trên 20 tuổi mới sinh con:

$$SE_{p2} = \sqrt{\frac{(5,34) \times (94,66)}{(225-1)}} = 1,50\%$$

Nếu quần thể là lớn, và mẫu nghiên cứu được rút ra một cách ngẫu nhiên hoàn hảo với cỡ mẫu đầy đủ, thì ở ngưỡng xác suất 5%:

- Tỷ lệ chết trẻ con dưới 1 tuổi của các bà mẹ dưới 20 tuổi đã sinh con :

$$15,56 \pm 2,42 = \text{từ } 17,98\% \text{ đến } 13,14\%$$

- Tỷ lệ chết trẻ con dưới 1 tuổi của các bà mẹ trên 20 tuổi mới sinh con:

$$5,34 \pm 1,50 = \text{từ } 6,84\% \text{ đến } 3,84\%$$

Chú ý:

Ở đây, ta tiến hành nghiên cứu thuần tập tương lai, theo dõi được chủ động và chặt chẽ các kết quả thu được, nên coi như kết quả không chứa sai số.

- Tính

$$RR = \frac{p_1}{p_2} = \frac{15,56}{5,34} = 3$$

- Tính $AR = p_1 - p_2 = 15,56 - 5,34 = 10,22\%$

Nghĩa là: Các bà mẹ trẻ dưới 20 tuổi đã sinh con có nguy cơ chết con trội lên 10,22% và cao gấp 3 lần so với ở các bà mẹ trên 20 tuổi mới sinh con.

IV. BÀI TẬP

1. Một nghiên cứu ngang trong năm 1976 phát hiện ra rằng: tỷ lệ dùng thuốc tránh



thai phân bố theo lứa tuổi như sau:

Tuổi	Tỷ lệ dùng TTT
15-19	15%
20-24	25%
25-29	22%
30-34	15%
35-39	7%
40-44	3%

Điều suy ra từ những số liệu trên là phụ nữ càng nhiều tuổi càng hạn chế dùng thuốc tránh thai. Điều suy luận này là:

- Đúng
- Không đúng, vì cần tính tỷ lệ trên quần thể để hỗ trợ cho quan sát này.
- Không đúng, vì nghiên cứu không sử dụng nhóm đối chứng hoặc nhóm so sánh.
- Không đúng vì có thể ảnh hưởng của chính nhóm nghiên cứu.
- Không đúng vì cần đến tần số mới dùng thuốc để khẳng định điều này, trong khi ở đây lại là tần số hiện đang dùng thuốc tránh thai.

Đối với các câu hỏi 2, 3, 4, 5 và 6 dưới đây, hãy nhận định các nghiên cứu được mô tả theo một trong các kiểu nghiên cứu đưa ra dưới đây:

- Nghiên cứu quan sát không kiểm soát.
- Nghiên cứu ngang hoặc nghiên cứu sự lưu hành bệnh.
- Nghiên cứu thực nghiệm.
- Nghiên cứu tương lai.
- Nghiên cứu hồi cứu .

2. Người ta tìm thấy 4 con chuột trong 1 chuồng ở khu nuôi súc vật thí nghiệm của trường đại học bị chết. Trong chuồng bên cạnh, một con bị co giật rồi chết, 2 con ốm nhưng đã qua khỏi và 1 con còn đang ốm. Nhà thú y tuyên bố rằng hiện xảy ra một vụ dịch ở chuột.

3. 1500 nam giới làm việc cho một hãng hàng không được người ta khám sức khỏe lần đầu tiên vào năm 1951 và xếp loại theo tiêu chuẩn chẩn đoán bệnh mạch vành. Cứ 3 năm một lần người ta lại tiến hành khám lại để phát hiện những trường hợp mới mắc bệnh này. Hàng năm người ta tính tỷ lệ tấn công cho từng nhóm tuổi khác nhau.

4. Một mẫu nghiên cứu ngẫu nhiên bao gồm những người nam giới ở nhà không đi làm việc chọn từ kết quả điều tra dân số ở 4 vùng. Người ta tiến hành khám bệnh cho các đối tượng này để phát hiện mắc bệnh động mạch vành. Tất cả những người có bệnh đều được loại ra khỏi cuộc nghiên cứu. Những người còn lại được phân bố vào nhóm tập luyện và nhóm đối chứng không tập luyện. Những người tham gia vào nhóm tập luyện theo một chương trình có hệ thống trong vòng 2 năm. Sau đó cứ nửa năm 1 lần người



ta lại tiến hành khám sức khỏe để phát hiện ra những trường hợp bệnh mạch vành ở cả 2 nhóm.

5. Dựa vào danh bạ điện thoại của thành phố, cứ 10 người, người ta lại chọn ra 1 người để gửi bảng câu hỏi điều tra cho người ấy. Các câu hỏi đề cập đến tuổi, giới, thói quen hút thuốc, các triệu chứng đường hô hấp nếu có trong một tuần lễ trước đó. Trên 90% các đối tượng đã chấp thuận trả lời và gửi kết quả. Qua đó người ta xác định các tỷ lệ hiện có triệu chứng đường hô hấp trên.

- Ung thư âm đạo.

7. Ung thư biểu mô đường âm đạo trước đây rất hiếm gặp. Tuy nhiên từ 1966-1969 tại bệnh viện Vincen Memorial Boston có đến 8 trường hợp mắc bệnh đến điều trị tại đó. Sự xuất hiện loại ung thư này ở những bệnh nhân sinh ra trong khoảng thời gian 1946-1951 tại nước Anh là một sự kiện bất thường khiến người ta phải tiến hành một cuộc điều tra. Hướng chú ý của các nhà điều tra tập trung trực tiếp vào tiền sử dùng thuốc oestrogens của mẹ trong thời gian mang thai dẫn đến hậu quả bệnh nhân mắc ung thư đường âm đạo.

Kiểu nghiên cứu dịch tễ học nào là thích hợp cho trường hợp này? Tại sao?

Trả lời bài tập.

1. d. Điều suy diễn có thể đúng, nhưng không thể rút ra được từ bảng số liệu đó, vì rằng có thể ảnh hưởng của chính nhóm nghiên cứu. Nhất thiết cần phải theo dõi một mẫu nghiên cứu theo thời gian và ghi nhận việc sử dụng thuốc tránh thai của từng cá nhân theo thời gian.

2. a. Quan sát không kiểm soát.

3. d. Nghiên cứu tương lai.

4. c. Nghiên cứu thực nghiệm.

5. e. Nghiên cứu hồi cứu hoặc nghiên cứu bệnh chứng.

6. b. Nghiên cứu cắt ngang hay nghiên cứu tỷ lệ hiện mắc.

7. Tiến hành nghiên cứu bệnh chứng là thích hợp, vì rằng:

a. Đây là bệnh hiếm.

b. Tính chính xác của chẩn đoán cao.

c. Không thể nhầm lẫn các ca bệnh được.

d. Tiền sử tiếp xúc đối với yếu tố nguy cơ nghi ngờ là căn nguyên gây bệnh có thể thu thập được thông qua các hồ sơ hay hỏi tiền sử.

e. Người ta có thể tính được nguy cơ tương đối.

THỬ NGHIỆM LÂM SÀNG NGẪU NHIÊN

I. MỤC TIÊU

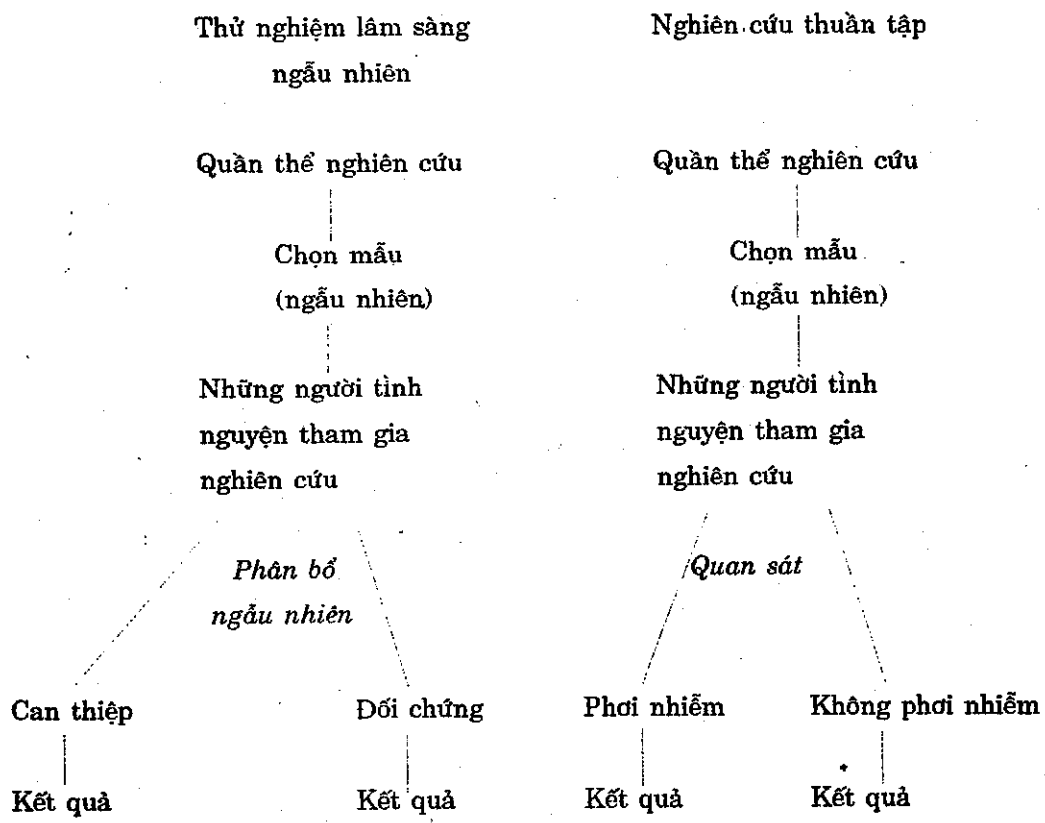
Mô tả, thiết kế đúng, và thực hiện được một thử nghiệm lâm sàng ngẫu nhiên.

II. NỘI DUNG

Nghiên cứu thực nghiệm thường là một phương pháp tốt nhất trong nghiên cứu khoa học nhằm kiểm định một giả thuyết. Trong vật lý học, nghiên cứu thực nghiệm thường hay được sử dụng. Trong sinh học, người ta cũng thường tiến hành nghiên cứu thực nghiệm trên động vật. Nghiên cứu thực nghiệm trên người thường ít được sử dụng.

Mục đích của các nghiên cứu bệnh chứng và thuần tập là nhằm xác định căn nguyên của bệnh. Các nghiên cứu can thiệp thực nghiệm được trình bày ở bài này thường có một mục đích khác là nhằm xác định xem trong các phương pháp điều trị, phương pháp nào là tốt nhất (xem hình 1). Các thông tin thu được từ nghiên cứu thực nghiệm thường rất có ích trong việc xác định mối liên hệ căn nguyên được hình thành từ các nghiên cứu quan sát. Ví dụ, các chương trình chống hút thuốc lá ở phụ nữ có thai rất có hiệu quả trong việc phòng trẻ em sinh ra thiếu cân. Kết quả này sẽ hỗ trợ mạnh mẽ cho một kết luận rằng hút thuốc lá trong thời kỳ mang thai là nguyên nhân dẫn đến trẻ sơ sinh thiếu cân.

Giống như trong các nghiên cứu thuần tập quan sát, các thử nghiệm lâm sàng là một loại nghiên cứu tương lai. Các đối tượng nghiên cứu được theo dõi và giám sát sự xuất hiện bệnh trong quá trình nghiên cứu. Sự khác nhau căn bản giữa nghiên cứu thuần tập và thử nghiệm lâm sàng là trong thử nghiệm lâm sàng, người nghiên cứu chủ động tác động hay can thiệp vào một nhóm (nhóm can thiệp) và nhóm khác không nhận một biện pháp can thiệp nào (nhóm chứng). Trong khi đó, trong nghiên cứu quan sát, người nghiên cứu chỉ quan sát và không hề có một tác động hay một can thiệp nào. Phân bố ngẫu nhiên bệnh nhân vào các nhóm điều trị khác nhau là đặc trưng cơ bản của thử nghiệm lâm sàng.



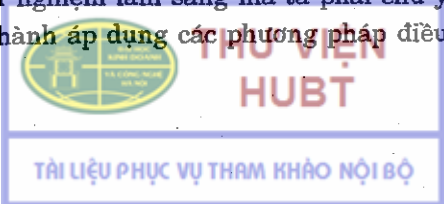
Hình 1- Sự khác nhau giữa thử nghiệm lâm sàng ngẫu nhiên và nghiên cứu thuần tập tương lai.

1. Sự đồng ý tham gia nghiên cứu.

Trong thử nghiệm lâm sàng, chỉ có một nhóm nhận một phương pháp điều trị nào đó, còn nhóm đối chứng thì không nhận can thiệp gì cả. Do đó, các đối tượng tham gia nghiên cứu phải là những người tình nguyện. Nhưng người được đề nghị tham gia nghiên cứu thường là những người có nguy cơ cao xuất hiện một bệnh nào đó mà ta cần nghiên cứu (ví dụ như nhồi máu cơ tim, mất khả năng sinh đẻ hoặc dị tật bẩm sinh). Bệnh nhân phải được thông báo về chế độ nghiên cứu và điền vào mẫu đồng ý tham gia nghiên cứu. Nếu bệnh nhân không đồng ý tham gia nghiên cứu, họ có thể thay đổi lối sống do hiểu biết về yếu tố nguy cơ và dẫn đến làm hạ thấp nguy cơ mắc bệnh. Mặc dù vậy, thử nghiệm lâm sàng ngẫu nhiên vẫn mang lại một kết quả tốt mà không phương pháp nghiên cứu nào có thể đạt được.

2. Phân bố ngẫu nhiên đối tượng tham gia nghiên cứu.

Việc phân bố ngẫu nhiên đối tượng tham gia nghiên cứu vào nhóm can thiệp và nhóm đối chứng sẽ mang lại kết quả có độ tin cậy lớn nhất và các nhóm nghiên cứu có thể so sánh được với nhau. (Điều đó có nghĩa là vào thời điểm bắt đầu nghiên cứu, cả hai nhóm đều giống nhau về tất cả các yếu tố đã biết và chưa biết, trừ yếu tố mà ta nghiên cứu). Đây là điểm cốt yếu khi thiết kế thử nghiệm lâm sàng mà ta phải chú ý đến. Nếu chúng ta có hai nhóm bệnh nhân, và tiến hành áp dụng các phương pháp điều trị khác nhau ở



mỗi nhóm, sự khác biệt về kết quả ở các nhóm sẽ do phương pháp điều trị khác nhau với một điều kiện là các yếu tố khác đều giống nhau và chỉ có một yếu tố khác nhau là phương pháp điều trị. Điều này chỉ có thể đạt được khi các đối tượng nghiên cứu được chọn ngẫu nhiên. Chúng ta cần phải phân biệt được sự khác nhau giữa lựa chọn ngẫu nhiên các đối tượng tham gia vào nghiên cứu (có thể giống nhau giữa nghiên cứu quan sát và nghiên cứu thực nghiệm) và phân bố ngẫu nhiên các đối tượng nghiên cứu đã được chọn vào nhóm can thiệp và nhóm đối chứng (chỉ xảy ra ở nghiên cứu thực nghiệm).

3. Loại trừ các sai lệch.

Cả đối tượng nghiên cứu và người nghiên cứu đều có thể ảnh hưởng đến kết quả của thử nghiệm. Để giảm các sai lệch loại này, người ta thường sử dụng kỹ thuật "làm mù". Trong kỹ thuật "làm mù kép", cả người nghiên cứu và đối tượng nghiên cứu đều không biết bệnh nhân được phân bố vào nhóm nào. Kỹ thuật này đặc biệt phải được tiến hành trong những nghiên cứu mà kết quả thường mang tính chủ quan như "khá hơn", "không thay đổi" hay "xấu đi". Nếu kết quả nghiên cứu có tính khách quan, như chết, cơn đột quỵ, không cần thiết phải áp dụng kỹ thuật "làm mù".

4. Ảnh hưởng giả tạo của chất trợ (Placebo effect).

Placebo effect

Người ta đã biết rõ một hiện tượng trong y học là khi bệnh nhân được điều trị một loại chất trợ (ví dụ như viên đường) họ thường cảm thấy có những tiến triển tốt hơn về mặt lâm sàng so với bệnh nhân không được điều trị bằng chất trợ ấy. Hiện tượng này thường được gọi là "ảnh hưởng giả tạo của chất trợ" và phải được xem xét trong khi thiết kế các thử nghiệm lâm sàng nếu ta muốn quy kết hiệu quả trong nghiên cứu can thiệp là do chính phương pháp điều trị can thiệp chứ không phải là do ảnh hưởng của sự điều trị. Do đó, người ta thường sử dụng một chất trợ để điều trị bệnh nhân mà bệnh nhân ở nhóm chứng không thể nhận biết được, có nghĩa là chất trợ này giống hoàn toàn với thuốc ở nhóm được điều trị can thiệp về hình dạng, màu sắc. Chỉ có như vậy, sự khác nhau về kết quả điều trị giữa các nhóm sẽ là do sự khác nhau về phương pháp điều trị.

5. Phân tích kết quả.

Khi phân tích kết quả của thử nghiệm lâm sàng, người ta thường đưa ra nhiều dự đoán. Đầu tiên người ta tiến hành so sánh giữa các nhóm can thiệp và nhóm chứng xem sự giống nhau về các yếu tố như tuổi, giới, chủng tộc và đặc biệt là các yếu tố nguy cơ đã biết có nằm trong nghiên cứu hay không (kết quả của sự phân bố ngẫu nhiên). Sau đó, người ta tiến hành phân tích ảnh hưởng của phương pháp can thiệp (ví dụ, chương trình chống hút thuốc lá sẽ làm giảm tỷ lệ người hút thuốc lá ở nhóm can thiệp). Cuối cùng, bước phân tích chủ yếu của thử nghiệm lâm sàng là kiểm định xem liệu giả thuyết đưa ra rằng yếu tố nguy cơ ảnh hưởng đến sức khỏe có đúng không (ví dụ, hút thuốc lá làm tăng nguy cơ ung thư đường hô hấp dưới).

Về nguyên lý, bước phân tích chủ yếu của thử nghiệm lâm sàng không khác nhau nhiều lắm so với phân tích kết quả của các nghiên cứu thuần tập quan sát. Ví dụ, trong thử nghiệm lâm sàng, không nhất thiết là phương pháp can thiệp được thực hiện thành công ở mọi cá thể ở nhóm can thiệp (ví dụ, tất cả mọi người đều ngừng hút thuốc lá) và hoàn toàn không xảy ra ở nhóm chứng (ví dụ, không ai ngừng hút thuốc lá). Kết quả của thử nghiệm lâm sàng là phát hiện được sự khác nhau về hiệu quả can thiệp ở hai nhóm. Ngoài

ra, bước phân tích kiểm định giả thuyết phải dựa trên nguyên tắc là mọi đối tượng nghiên cứu một khi đã được phân bố ngẫu nhiên vào các nhóm họ phải được đưa vào phân tích. Nói một cách khác là nếu ta loại bỏ một cách có chọn lọc bệnh nhân khi phân tích kết quả sẽ làm mất đi ý nghĩa của việc phân bố ngẫu nhiên bệnh nhân vào các nhóm nghiên cứu.

III. BÀI TẬP

Người ta đã tiến hành một thử nghiệm lâm sàng ngẫu nhiên để đánh giá các phương pháp điều trị bệnh đái đường.

Thử nghiệm này bao gồm các bước như: xây dựng quy trình thu thập các số liệu có thể so sánh được với nhau, phân bố ngẫu nhiên bệnh nhân vào các nhóm điều trị bằng các phương pháp khác nhau bao gồm cả nhóm so sánh được điều trị bằng chất trợ, đánh giá bằng phương pháp "làm mù kép" các loại thuốc uống, theo dõi bệnh nhân trong một thời gian dài, thu thập, biên soạn và giám sát các số liệu nghiên cứu.

Người ta lựa chọn những bệnh nhân đái đường không phụ thuộc insulin và không đòi hỏi phải điều trị insulin. Loại đái đường này chiếm chủ yếu (trên 90% tổng số bệnh nhân đái đường).

Thử nghiệm này có hai mục tiêu chính:

1. Đánh giá hiệu quả của các phương pháp điều trị làm hạ đường huyết trong việc dự phòng các biến chứng về mạch máu.
2. Nghiên cứu lịch sử tự nhiên của bệnh về mạch máu.

Người ta đã áp dụng những chế độ điều trị sau trong nghiên cứu này:

1. Điều trị bằng insulin với những liều lượng khác nhau để duy trì mức độ đường huyết ở mức bình thường.
2. Điều trị bằng insulin với liều lượng chuẩn: Insulin chậm U-80, 10-16 đơn vị mỗi ngày tiêm theo diện tích cơ thể.
3. Điều trị bằng Tolbutamide 1,5g/ngày theo đường uống.
4. Điều trị bằng lactose trợ (placebo) dưới hình dạng và liều lượng giống như các thuốc điều trị hạ đường huyết.

Tất cả các nhóm trên đều có một chế độ ăn kiêng như nhau.

Kết quả:

Các nhà nghiên cứu đã thu được những kết quả bất ngờ sau:

1. Insulin không có hiệu quả hơn so với chỉ áp dụng chế độ ăn kiêng suốt đời ở những bệnh nhân đái đường.
2. Ăn kiêng cộng với điều trị bằng tolbutamide không có hiệu quả hơn so với chỉ áp dụng chế độ ăn kiêng suốt đời ở những bệnh nhân đái đường.
3. Có sự tăng có ý nghĩa thống kê tỷ lệ chết do bệnh mạch máu ở nhóm được điều trị bằng tolbutamide so với nhóm được điều trị bằng placebo.

Câu hỏi:

1. Sự lựa chọn nhóm bệnh nhân với thể đái đường này ảnh hưởng như thế nào đến thể này áp dụng những kết luận thu được từ thử nghiệm lâm sàng này.

2. Khi nào kỹ thuật "làm mù kép" là cần thiết nhất và ít cần thiết nhất? Nó có nghĩa là như thế nào?

3. Hiện nay rất khó tiến hành được một thử nghiệm lâm sàng như thế này. Tại sao?

4. Nếu không có nhóm điều trị bằng placebo, ta sẽ không thu được những kết luận gì?

5. Tại sao phân bố ngẫu nhiên bệnh nhân vào các nhóm điều trị là cần thiết và nó đạt được mục đích gì?

Trả lời:

1. Những kết quả đạt được từ nghiên cứu này, cũng như những nghiên cứu khác, chỉ áp dụng được cho những quần thể tương tự như quần thể nghiên cứu. Tuy nhiên, trong thử nghiệm này, không có vấn đề gì bởi vì những bệnh nhân nghiên cứu đại diện cho phần lớn những bệnh nhân đái đường phong phụ thuộc insulin.

2. Kỹ thuật "làm mù kép" là cần thiết để loại trừ các sai lệch chủ quan gây ra do cả người nghiên cứu và đối tượng nghiên cứu. Kỹ thuật này là cần thiết nhất trong các nghiên cứu mà kết quả mang tính chủ quan như "khá hơn", "không thay đổi" hay "xấu hơn". Nó ít cần thiết nhất cho các nghiên cứu với những kết quả khách quan. Kỹ thuật này sẽ không cho cả người nghiên cứu và đối tượng nghiên cứu biết về chế độ điều trị trên placebo.

3. Tất cả những bệnh nhân tham gia nghiên cứu đã đồng ý tham gia nghiên cứu sau khi đã được thông báo về nghiên cứu. Họ đã được theo dõi chặt chẽ mọi ảnh hưởng hay triệu chứng. Một chuyên gia giỏi về bệnh đái đường đã tham gia vào thiết kế thử nghiệm này và cho rằng nhóm placebo sẽ có kết quả tương tự như các nhóm điều trị bằng thuốc hạ đường huyết. Và cuối cùng, điều này đã được chứng minh là đúng.

4. Người ta có thể thu được những kết luận quan trọng nhất nếu không tiến hành điều trị placebo. Đó là:

a. Insulin không tốt hơn so với placebo.

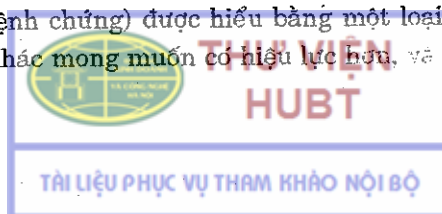
b. Tolbutamide có liên quan với tăng nguy cơ tử vong và bệnh mạch máu (có ý nghĩa thống kê) so sánh với placebo.

5. Phân bố bệnh nhân ngẫu nhiên vào các nhóm điều trị là cần thiết vì chỉ khi đó, sự khác biệt về kết quả điều trị mới có thể qui cho là do tiến hành các thử nghiệm điều trị khác nhau. Nếu không tiến hành phân bố ngẫu nhiên, sự khác nhau về kết quả điều trị có thể là do các đặc trưng khác của các nhóm nghiên cứu. Phân bố ngẫu nhiên dễ mang lại một sự phân bố như nhau các yếu tố nguy cơ (ví dụ tuổi, giới, mức độ nặng của bệnh) ở các nhóm. Đây là một khái niệm cơ bản của thử nghiệm lâm sàng ngẫu nhiên.

IV. CÁCH TIẾN HÀNH NGHIÊN CỨU MỘT THỬ NGHIỆM LÂM SÀNG NGẪU NHIÊN

A. NHẮC LẠI MỘT VÀI KHÁI NIỆM CẦN THIẾT

1. Thử nghiệm lâm sàng thuộc loại nghiên cứu can thiệp. Nói như vậy có nghĩa là: khái niệm "yếu tố nguy cơ" trong các nghiên cứu phân tích dịch tễ (như nghiên cứu thuần tập hoặc nghiên cứu bệnh chứng) được hiểu bằng một loại thuốc điều trị khác, hoặc một phương pháp điều trị khác mong muốn có hiệu lực hơn, và loại thuốc điều trị này, phương



pháp điều trị này... do chính người thầy thuốc điều trị chủ động "áp đặt" vào, can thiệp vào quá trình diễn biến của bệnh (làm cho bệnh mau khỏi, hoặc bệnh sẽ khỏi triệt để, hoặc ít di chứng để lại...).

2. thử nghiệm lâm sàng là một nghiên cứu tương lai, trong đó chính người thầy thuốc điều trị phải theo dõi, giám sát xác nhận sự xuất hiện của hiệu quả điều trị mong đợi trong tương lai.

3. Vì thử nghiệm lâm sàng là một trong những nghiên cứu phân tích dịch tễ, nên bao giờ cũng phải thiết lập một nhóm đối chứng, với tỷ lệ về số lượng hoặc 1/1, hoặc 2/1 tùy thuộc nhóm đối chứng được chọn như thế nào, nghĩa là đảm bảo sự tương đồng đến đâu so với nhóm chủ cứu.

4. Vì cũng là một nghiên cứu ngẫu nhiên, nên yếu tố ngẫu nhiên cũng vẫn là yếu tố chi phối làm giảm các sai số giống như ở các nghiên cứu ngẫu nhiên khác. Song ở đây quá trình chọn các đối tượng bằng ngẫu nhiên, lại biểu thị dưới dạng phân bố việc có nhận và không nhận biện pháp can thiệp cho họ, để cuối cùng ta có được những ai sẽ vào nhóm chủ cứu (có nhận yếu tố can thiệp) và những ai sẽ vào nhóm đối chứng (không nhận yếu tố can thiệp cần nghiên cứu, mà nhận yếu tố can thiệp cũ hoặc placebo để đối chứng).

5. Kỹ thuật "làm mù" phải được vận dụng triệt để. Nên tiến hành kỹ thuật "làm mù kép". Có thể được thì nên tiến hành "làm mù ba" (cả bệnh nhân, cả thầy thuốc làm điều trị và cả người xử lý kết quả đều không biết) trong đó quan trọng hơn cả là những người bệnh được can thiệp dứt khoát không được biết mình đang "bị làm nghiên cứu", trừ những nghiên cứu phải sử dụng người tình nguyện. Ngay cả trường hợp này thì họ lại càng không thể được biết mình được hay bị ở nhóm nghiên cứu nào, nó ảnh hưởng rất lớn đến hiệu lực nghiên cứu (gây những sai chệch không lường được) và lại còn là vấn đề đạo đức y học nữa.

Kỹ thuật làm mù	Người bị làm mù trong quá trình nghiên cứu		
	bệnh nhân	người nghiên cứu	người xử lý số liệu
mù đơn	có	không	không
mù kép	có	có	không
mù ba	có	có	có

B. MỘT BÀI TẬP VỀ MẪU NGHIÊN CỨU THỬ NGHIỆM LÂM SÀNG

Một thầy thuốc muốn thử xem tác dụng nhanh của một loại thuốc sốt rét mới so với loại thuốc cũ hiện đang dùng.

Tiêu chuẩn đạt ra là hết ký sinh trùng vô tính trong vòng 36 giờ sau khi dùng thuốc.

1. Trước hết ta hãy tính cỡ mẫu.

Để tính được cỡ mẫu tối thiểu vừa đủ, cần phải:

a) *Ấn định mức ý nghĩa* cho nghiên cứu, cũng đồng thời là *ấn định mức tin cậy* mà



ta sẽ dùng để kiểm định sự khác biệt tỷ lệ khỏi của 2 loại thuốc trên 2 nhóm người dùng 2 loại thuốc đó. Thường dùng mức ý nghĩa 0,1 hoặc 0,05 hoặc 0,01.

b) *Ấn định một lực mẫu nhất định.* Lực mẫu là mong muốn của chúng ta để phát hiện ra được sự khác biệt (có ý nghĩa thống kê) của 2 loại thuốc này, nói khác đi lực mẫu là xác suất có được kết quả nghiên cứu có ý nghĩa. Lực mẫu phụ thuộc vào:

+ Sự khác biệt của các tỷ lệ khỏi bệnh thực sự là bao nhiêu. Sự khác biệt này càng lớn, thì lực mẫu càng lớn.

+ Cỡ mẫu là bao nhiêu. Cỡ mẫu càng lớn, thì lực mẫu càng lớn.

Những lực mẫu thường dùng là 80%, 90% và 95%

c) *Ấn định một sự khác biệt có giá trị về lâm sàng giữa các tỷ lệ của 2 nhóm người sử dụng 2 loại thuốc đó.* Sự khác biệt này là bao nhiêu để có được giá trị về lâm sàng thì phải do các thầy thuốc lâm sàng tự ấn định ra dựa trên ý nghĩa thực tế lâm sàng. Thí dụ sự khác biệt là 5% sẽ rất có giá trị nếu tỷ lệ khỏi đang từ 5% lên 10%, nhưng sẽ ít có giá trị hơn nhiều nếu tỷ lệ khỏi đang từ 90% lên 95%. Tuy nhiên, đây chỉ là thí dụ về các con số, các thầy thuốc lâm sàng cần căn cứ vào nhiều khía cạnh khác, về ý nghĩa xã hội của vấn đề, ý nghĩa y học và y tế của vấn đề, tính phổ biến và nghiêm trọng của bệnh nghiên cứu...

Để có được sự khác biệt này, tất nhiên chúng ta đã có tỷ lệ khỏi đối với loại thuốc điều trị đang dùng, công việc còn lại là ước lượng tỷ lệ khỏi đối với loại thuốc mới nữa mà thôi.

d) *Công thức tính cỡ mẫu.*

Có nhiều.

Đơn giản nhất là:

$$n = \frac{Z_{1-\alpha/2} p.q.F}{D^2}$$

trong đó: p là trung bình của 2 tỷ lệ ở 2 nhóm:

$$p = \frac{p_1 + p_2}{2}$$

q là số bù của p đến 1:

$$q = 1 - p$$

là bình phương của tổng ($Z_{\alpha/2}$ và Z_{β})

$$F = (Z_{\alpha/2} + Z_{\beta})^2 \text{ với :}$$

$Z_{\alpha/2}$ là giá trị tới hạn của phân bố chuẩn ứng với mức ý nghĩa 2 phía của sai lầm α đã ấn định

$$\alpha = 0,05 \text{ thì } Z_{\alpha/2} = 1,96$$

$$\alpha = 0,01$$

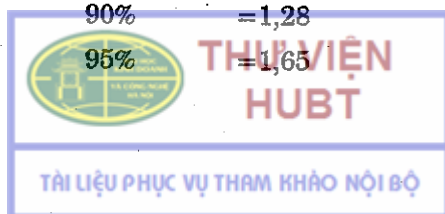
$$\alpha = 0,1$$

Z_{β} là giá trị tới hạn 1 phía tương ứng với lực mẫu $1 - \beta$ đã ấn định

$$1 - \beta = 80\% \text{ thì } Z_{\beta} = 0,86$$

$$90\% = 1,28$$

$$95\% = 1,65$$



Để giúp người nghiên cứu, người ta đã tính sẵn các giá trị của F ở các mức ý nghĩa và lực mẫu thường dùng:

Mức ý nghĩa \ Lực mẫu	50%	80%	90%	95%
0,1	2,7	6,2	8,6	10,8
0,05	3,8	7,9	10,5	13,0
0,01	6,6	11,7	14,9	17,8

Còn D là khác biệt tỷ lệ ở 2 nhóm (nó chính là hiệu số giữa 2 tỷ lệ này, khi chúng được biểu thị bằng cùng đơn vị)

Thí dụ nếu chúng ta ấn định:

Mức ý nghĩa $\alpha = 0,05 \rightarrow Z_{\alpha/2} = 2$ (thay cho 1,96)

Lực mẫu $1 - \beta = 0,90$ thì F sẽ là 10,5 (hàng 0,05 cột 90% trong bảng F trên đây).

Nếu D = 0,10 (Tỷ lệ khỏi của thuốc cũ $p_1 = 70%$, của tỷ lệ thuốc mới $p_2 = 80%$, thì

$$p = \frac{0,7+0,8}{2} = 0,75$$

thì sẽ có:

$$n = \frac{2 \times 0,75 \times 0,25 \times 10,5}{(0,1)^2} = 394 \approx 400$$

như vậy cần phải lấy mỗi nhóm 400 bệnh nhân vào mẫu nghiên cứu: 400 người điều trị thuốc cũ và 400 người điều trị thuốc mới.

2. Chỉ định số bệnh nhân vào 2 nhóm.

Sau khi loại trừ những bệnh nhân có những bệnh kết hợp hoặc mạn tính xét có thể ảnh hưởng đến quá trình điều trị, 800 bệnh nhân đủ tiêu chuẩn được lấy vào nghiên cứu. Lập danh mục của 800 bệnh nhân này, tên xếp theo vần chữ cái rồi chọn ngẫu nhiên họ vào 2 nhóm, bằng các kỹ thuật ngẫu nhiên. Thí dụ nếu người thầy thuốc không có ý định chủ quan nào vào việc can thiệp bằng thuốc cũ hoặc thuốc mới thì có thể chỉ định số chẵn vào nhóm nọ và số lẻ vào nhóm kia... Việc chỉ định ngẫu nhiên này có tầm quan trọng quyết định đến kết quả nghiên cứu.

3. Thực hiện nghiên cứu.

Quá trình này nên áp dụng kỹ thuật "làm mù kép" các bệnh nhân được chăm sóc điều trị như nhau về hình thức, để không ai được biết mình được hoặc bị điều trị thế này thế khác. Mặt khác những người thực hiện điều trị cũng không được biết có sự khác nhau trong việc điều trị này (các gói thuốc, ống thuốc, viên thuốc phải giống nhau, được đánh số theo mã ghi trong danh mục nghiên cứu cho các bệnh nhân tương ứng). Họ cũng được xét nghiệm KST sốt rét vào giờ trước giờ thứ 36 kể từ khi bắt đầu dùng thuốc không sớm hơn, không muộn hơn. Người ta cho rằng trong các thử nghiệm lâm sàng ngẫu nhiên, quá trình thực hiện nghiên cứu là bước khó thực hiện nghiêm ngặt (nhà nghiên cứu phải

giám sát rất chặt chẽ) nhưng cũng phải lại là bước quyết định đến tính chính xác của kết quả nghiên cứu.

4. Tính kết quả nghiên cứu.

Giả sử ta thu được kết quả sau, trong 1 bảng 2x2:

36 giờ sau Thuốc dùng điều trị	hết KST	còn KST	Cộng
Thuốc mới	350	50	400
Thuốc cũ	180	120	400

1. Trước hết ta tìm "ý nghĩa thống kê" của kết quả nghiên cứu. Đây là một nghiên cứu tỷ lệ nên ta sẽ dùng một trắc nghiệm ý nghĩa về sự khác biệt giữa 2 tỷ lệ:

$$\chi^2_{1b+a} = 35,5$$

Kết quả nghiên cứu có ý nghĩa thống kê ở ngưỡng $p < 0,05$ như đã dự kiến, ta loại bỏ H_0 (2 tỷ lệ là không khác biệt) và chấp nhận H_1 sau: Tỷ lệ hết KST vô tính ở nhóm được điều trị bằng thuốc mới có khác biệt tỷ lệ đó ở nhóm điều trị bằng thuốc cũ. Sự khác biệt này có ý nghĩa thống kê ở ngưỡng $p < 0,05$.

2. Ta tính tỷ lệ ở 2 nhóm:

- Nhóm dùng thuốc mới: $350/400 = 0,87 \pm 0,0057$

- Nhóm dùng thuốc cũ: $280/400 = 0,70 \pm 0,0105$

3. Tính RR: $0,87/0,70 = 1,25$

E. KẾT LUẬN

Trong nghiên cứu, việc chỉ định vào 2 nhóm chủ cứu và đối chứng được thực hiện ngẫu nhiên trong số bệnh nhân được tính theo cỡ mẫu phù hợp với các điều kiện ấn định. Khi tiến hành nghiên cứu, kỹ thuật "làm mù kép" được giám sát triệt để, nên có thể cho rằng kết quả nghiên cứu không chứa sai số quan trọng. Bây giờ, vận dụng 6 tiêu chuẩn về tính nguyên nhân, người ta thấy chúng đều được thỏa mãn trong nghiên cứu, nên có thể kết luận rằng thuốc mới có tỷ lệ sạch KST sót rét trong vòng 36 giờ cao hơn so với tỷ lệ đó nếu dùng thuốc cũ: 0,87 so với 0,70.

Sau đây, là trích bảng để tiện việc so sánh. (bảng Fisher và Yates') ở một bậc tự do:

Xác suất P	Giá trị X^2
0,001	10,827
0,01	6,635
0,02	5,412
0,05	3,841
0,10	2,706
0,20	1,642



ĐIỀU TRA NGHIÊN CỨU SỨC KHỎE CỘNG ĐỒNG - NGHIÊN CỨU NGANG

(Cross sectional Study)

I. MỤC TIÊU

- Xác nhận vai trò quan trọng của nghiên cứu ngang trong dịch tễ học.
- Nhận biết những lợi điểm của mẫu chùm ngẫu nhiên.
- Phân biệt được các đặc tính của thiết kế lấy mẫu chùm.

II. NỘI DUNG

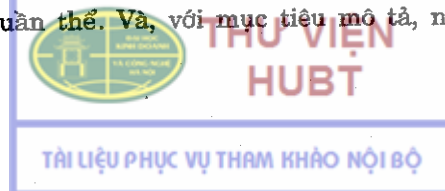
Trước hết cần phân biệt các nghiên cứu quan sát với các nghiên cứu can thiệp: ở nghiên cứu quan sát thì ta cần phải *chọn các cá thể* vào nghiên cứu, còn ở nghiên cứu can thiệp thì ta lại *chỉ định loại can thiệp* cho các cá thể (như 2 phương pháp điều trị cho các bệnh nhân chẳng hạn). Đôi khi, có những nghiên cứu kết hợp cả 2: thí dụ trong nghiên cứu điều trị sốt rét trong quần thể, thì can thiệp điều trị (dùng thuốc hoặc placebo) được *chỉ định* cho mọi trẻ con uống hàng tuần, và một mẫu 20% số trẻ đó được *chọn* ra để giám sát bằng quan sát hàng tháng.

Nghiên cứu ngang thuộc loại nghiên cứu quan sát. Cần chú ý đến một số điểm quan trọng như sau:

1. Nghiên cứu ngang gắn liền với tỷ lệ *hiện mắc*, nghĩa là nghiên cứu ngang chỉ được sử dụng để mô tả? (có rất ít kết quả nghiên cứu ngang được đưa sang dịch tễ học phân tích, khi những đặc trưng nguy cơ nghiên cứu là bẩm sinh, không thay đổi suốt một đời người, thí dụ màu mắt, nhóm máu).

2. Nghiên cứu ngang cũng gắn liền với khoảng thời gian diễn biến của hiện tượng sức khỏe mà ta nghiên cứu trong quần thể: chỉ nhìn nhận hiện tượng sức khỏe đó (thí dụ tình hình bệnh tật) trong một *thời điểm* nhất định: nên người ta hình dung kết quả ngang là một bức ảnh chụp nhanh về tình hình của hiện tượng sức khỏe mà ta nghiên cứu. Tuy nhiên, cụm từ "thời điểm" ở đây được nhìn nhận một cách tương đối, tương ứng với một "thời gian ngắn". Phần lớn các nghiên cứu về tình hình các bệnh truyền nhiễm cấp tính, theo kinh nghiệm, người ta thường ấn định thời gian ngắn này là 2 tuần.

3. Nghiên cứu ngang là công cụ chủ yếu hiện nay, được sử dụng rất rộng rãi để mô tả một hiện tượng sức khỏe trong quần thể. Và, với mục tiêu mô tả, nó cũng được sử



dụng để đánh giá: đánh giá một hiện tượng sức khỏe, đánh giá tiến độ hoặc kết quả của một chương trình y tế.

4. Với ý nghĩa mô tả, trong dịch tễ học hiện đại, việc mô tả một hiện tượng sức khỏe nào đó bao giờ cũng kèm theo việc mô tả yếu tố nguy cơ của nó (nếu không mô tả yếu tố nguy cơ kèm theo sẽ coi là thiếu sót trong việc mô tả, và nghiên cứu không phục vụ lợi ích thực tế).

Với ý nghĩa đánh giá, việc mô tả yếu tố nguy cơ cũng rất cần thiết, nó giải đáp nguyên nhân của tình trạng được đánh giá, và chỉ ngay ra cho cách khắc phục tốt hơn nữa tình trạng đó.

5. Nghiên cứu ngang có thể được sử dụng để mô tả, đánh giá cả các biến định lượng lẫn định tính, các biến liên tục và biến rời rạc. Nhưng đều tuân thủ những bước đi như nhau và với những nguyên lý như nhau.

6. Nghiên cứu ngang mô tả một tình hình y tế, một hiện trạng sức khỏe... Với mục tiêu là hình thành được một giả thuyết mang tính nguyên nhân về vấn đề nghiên cứu (khác hẳn với nghiên cứu tương lai: nghiên cứu tương lai nhằm kiểm định một giả thuyết nhân quả, và bao giờ cũng phải sử dụng nhóm đối chứng trong nghiên cứu của mình).

7. Nghiên cứu ngang mô tả hiện trạng nên kết quả được dùng trong việc quản lý, điều hành, đánh giá tình hình thực thi một chương trình, một chủ trương y tế, tình hình lưu hành của một bệnh, diễn biến của một hiện tượng sức khỏe (còn muốn điều tra, nghiên cứu về nguy cơ hoặc nguyên nhân của một bệnh chẳng hạn, phải sử dụng các thiết kế nghiên cứu dịch tễ học phân tích).

8. Nghiên cứu ngang có thể sử dụng với mẫu ngẫu nhiên đơn, mẫu tầng hoặc mẫu chùm hoặc các mẫu đó kết hợp nhau. Nhưng người ta thường sử dụng mẫu chùm để nghiên cứu ngang trên những quần thể lớn.

9. Nghiên cứu ngang tiến hành trong quần thể, thực chất là những cuộc điều tra cộng đồng, được mô tả chi tiết trong cuốn thực hành này. Đây là một phát triển quan trọng của dịch tễ học.

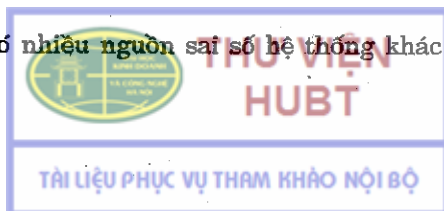
10. Thông thường thì các nghiên cứu ngang đều tiến hành trên một mẫu duy nhất trong quần thể (để tìm ra một tỷ lệ hoặc một số trung bình của quần thể đó). Còn trường hợp đặc biệt như khi muốn ước lượng sự khác biệt các tỷ lệ (hoặc số trung bình) ở 2 quần thể khác nhau chẳng hạn thì công thức để tính cỡ mẫu có khác với cỡ mẫu tính cho trường hợp làm một mẫu duy nhất trong quần thể.

11. Là một nghiên cứu theo mẫu, kết quả nghiên cứu ngang sẽ không "chập" được với kết quả khi nghiên cứu quần thể toàn bộ, vì bao giờ cũng có sai số do nghiên cứu mẫu, sai số này gọi là sai số mẫu, nó bao gồm 2 loại chính:

Sai số hệ thống và sai số ngẫu nhiên.

Chúng ta có thể làm giảm đến tối thiểu sai số hệ thống bằng cách thiết kế thật cẩn thận chi tiết các thủ tục lấy mẫu và tuân thủ triệt để những thủ tục đó khi thực thi. Và cũng có thể làm giảm sai số ngẫu nhiên đến mức chấp nhận được bằng cách chọn một cỡ mẫu hợp lý.

Ngoài ra còn có thể có nhiều nguồn sai số hệ thống khác nữa (sai số không đáp ứng



do một vài cá thể không đủ tiêu chuẩn nghiên cứu lại được chọn vào mẫu, sai số do người trả lời...). Cũng còn có những sai số không thuộc việc lấy mẫu không đề cập ở đây, nhưng cũng cần rất chú ý trong toàn bộ quá trình nghiên cứu.

12. Một vài định nghĩa.

- Đơn vị lấy mẫu: có thể là một người, một hộ gia đình, một trường học...
- Quần thể định danh (hay quần thể tham khảo): là quần thể ta quan tâm đến tiêu thức cần nghiên cứu.
- Quần thể nghiên cứu (hay quần thể điều tra): là tập hợp của các đơn vị lấy mẫu, từ đó ta lấy mẫu ra. Trên thực tế, thì quần thể nghiên cứu hẹp hơn nhiều so với quần thể định danh.
- Thiết kế mẫu: sơ đồ để chọn các đơn vị lấy mẫu vào mẫu từ quần thể nghiên cứu.
- Khung mẫu: danh mục các đơn vị để chọn mẫu.

13. Mẫu ngẫu nhiên đơn có thể được dùng nhưng ít dùng trong điều tra cộng đồng.

Mẫu ngẫu nhiên đơn là hình ảnh cơ bản nhất của các mẫu ngẫu nhiên.

- Phải lập một khung mẫu là một bản danh mục quần thể nghiên cứu, có tên và mã số (số thứ tự) từ 1 đến N.
- Từ đó rút ngẫu nhiên ra số đơn vị cần thiết của cỡ mẫu bằng nhiều kỹ thuật ngẫu nhiên, trong đó bảng số ngẫu nhiên thường được dùng phổ biến.

Trước khi tiến hành một kỹ thuật ngẫu nhiên nào đó, bao giờ cũng cần phải quy ước trước một số điểm như hướng ngẫu nhiên, phía ngẫu nhiên... và cần phải chú ý đến số chữ số, số ký tự của số ngẫu nhiên.

Mẫu ngẫu nhiên đơn còn được thực hiện bằng cách lấy mẫu hệ thống.

14. Mẫu 2 bậc: là loại mẫu được dùng phổ biến nhất để điều tra cộng đồng.

Trên thực tế, nhiều tình huống không thể thực hiện được mẫu ngẫu nhiên đơn, hoặc mẫu hệ thống như trường hợp quần thể quá lớn (quần thể vô hạn), không có khung mẫu (hoặc khung mẫu không tin cậy được) hoặc quần thể quá phân tán trên những lãnh thổ rộng lớn.

Một trong những mẫu 2 bậc là mẫu chùm, lấy như sau:

1. Quần thể thường tập hợp tự nhiên thành các chùm, như 1 làng, 1 quận huyện, một cao ốc... Ta sẽ lập danh mục của các đơn vị bậc 1 đó.
2. Chọn ngẫu nhiên các đơn vị bậc 1 đó ra.
3. Trong mỗi đơn vị bậc 1 vừa được chọn, ta lập danh mục các đơn vị bậc 2 (hộ gia đình hoặc cá thể) rồi ta chọn các đơn vị bậc 2 này vào mẫu cho đủ cỡ mẫu.

Mẫu 2 bậc có nhiều lợi điểm:

- Chi phí: không phải qua quá nhiều chùm.
- Dễ làm: chỉ cần lập danh mục các chùm đã được chọn, và có thể làm khi chúng ta đến chùm đó.

Tuy nhiên, chúng ta phải trả một giá nhất định vì cùng một cỡ mẫu thì mẫu chùm cho những ước lượng ít chính xác hơn mẫu ngẫu nhiên đơn nhiều, cho nên phải làm với

cỡ mẫu lớn hơn. Thí dụ: điều tra số hiện tượng dương tính huyết thanh với HIV ở thành phố Kinshara, Zaire, cỡ mẫu lớn (vì tỷ lệ dương tính thấp) lại phải trải rộng khắp thành phố, nên không làm mẫu ngẫu nhiên đơn được. Người ta đã chia thành phố ra 266 khu vực, chọn ra 100 khu vực, và trong mỗi khu vực chọn ngẫu nhiên 100 người lớn để phỏng vấn và lấy máu xét nghiệm.

15. Mẫu tầng ngẫu nhiên cũng thường được dùng trong điều tra cộng đồng.

Trong các cuộc điều tra nghiên cứu, đôi khi ta muốn có những kết quả riêng của từng khu vực khác nhau, và muốn chắc chắn là mỗi khu vực có riêng cỡ mẫu tương xứng của nó. Trường hợp này, người ta coi mỗi khu vực là 1 tầng, và ở mỗi tầng sẽ chọn ra một mẫu ngẫu nhiên riêng biệt (hoặc ngẫu nhiên đơn hoặc 2 bậc).

Tầng không nhất thiết phải là các khu vực theo địa dư, mà có thể là nông thôn-thành thị, nam-nữ, trẻ-già... miễn là ta thấy nếu làm toàn bộ thì ước lượng sẽ không chính xác, vì các ước lượng ở các tầng là khác nhau.

Thí dụ người ta đã tiến hành một cuộc điều tra về tỷ lệ chết trẻ con ở Gambia, mà những cuộc điều tra cơ bản trước đó cho thấy rằng công tác chăm sóc sức khỏe ban đầu ở các làng là rất khác nhau, nên người ta đã điều tra tỷ lệ chết trẻ con riêng ở tầng có và ở tầng chưa có chăm sóc sức khỏe ban đầu.

III. CÁCH TIẾN HÀNH MỘT NGHIÊN CỨU NGANG ĐỂ ĐIỀU TRA NGHIÊN CỨU SỨC KHỎE CỘNG ĐỒNG

1. Chọn vấn đề nghiên cứu.

Là những vấn đề ưu tiên trong bộ phận công tác của mình, được tính toán phù hợp với các nguồn lực có thể có, đảm bảo nghiên cứu có kết quả, và kết quả là có hiệu quả. (Nghiên cứu ngang thuộc loại nghiên cứu ứng dụng thực tế). Hiệu quả của một chương trình, một dịch vụ, một tình huống, một hoạt động y tế... được hiểu theo nghĩa là cải thiện được một hiện trạng y tế, hoặc làm giảm thiểu một hiện tượng sức khỏe nặng nề. Như vậy phải căn cứ vào mục tiêu của những vấn đề ưu tiên mà đặt vấn đề nghiên cứu. Ở đây, chúng ta cũng cần phân biệt những khái niệm về hiệu lực lâm sàng, hiệu lực dịch tễ, hiệu lực y tế... để đặt vấn đề nghiên cứu rõ ràng.

2. Xác định tiêu thức nghiên cứu: Quần thể và mẫu nghiên cứu.

Dưới đây, sẽ đề cập đến các nghiên cứu ngang nhằm tìm kiếm kết quả dưới dạng một tỷ lệ cho dễ hiểu. Không nhắc lại định nghĩa về một tỷ lệ, ở đây chỉ nói về khía cạnh bản thân nó là một tỷ số dưới dạng một phân số mà tử số là số đếm của đặc trưng nghiên cứu và mẫu số là số đếm của đơn vị nghiên cứu (trong quần thể trên đó đặc trưng nghiên cứu được khảo sát).

a) Đặc trưng hay tiêu thức nghiên cứu.

Thường là một bệnh, một kết quả của một dịch vụ hoặc một can thiệp nào đó. Dù là tiêu thức nghiên cứu nào, thì nó cũng phải được định nghĩa rõ ràng, dễ hiểu, không hiểu sai, và đo đếm được chính xác.



Thí dụ: trong bảng tiêu chuẩn chẩn đoán hội chứng sốc nhiễm độc ở phụ nữ dùng kháng hành kinh siêu thấm Reily có triệu chứng sốt, người ta đã định nghĩa sốt ở đây là người có thân nhiệt từ 38°C trở lên, hay trong chương trình TCMR được coi là có tiêm BCG có kết quả khi có sẹo BCG.

b) Quần thể.

Có nhiều khái niệm về quần thể trên đó xảy ra diễn biến tiêu thức nghiên cứu. Sử dụng loại quần thể nào phụ thuộc vào mục tiêu nghiên cứu. Quần thể cũng phải được định nghĩa rõ ràng. Thí dụ: trẻ 12 - 23 tháng, trong đó bao gồm những trẻ bắt đầu ngày kỷ niệm được sinh nhật lần thứ hai, không phân biệt gái hay trai, kinh tế, văn hóa, xã hội...

c) Mẫu và quần thể.

Thường quần thể là rất lớn khó có thể thu thập thông tin từ tất cả các cá thể trong quần thể đó. Nên chúng ta thường chọn ra một mẫu và thu thập thông tin từ các cá thể trong mẫu này, với mong muốn là mẫu là đại diện cho quần thể.

Nhưng ngay cả khi có thể điều tra quần thể toàn bộ, thì nghiên cứu mẫu vẫn tốt hơn, vì nghiên cứu mẫu:

- Tiết kiệm thời gian, nhân lực và chi phí.
- Chất lượng thông tin tập trung được thời gian và các nguồn lực khác cho các dữ kiện tin cậy trên mỗi cá thể.
- Mẫu chùm ít chính xác hơn mẫu ngẫu nhiên đơn, nhưng vẫn đáp ứng được các lợi điểm này, khi số đơn vị thống kê trong quần thể điều tra là rất lớn.

3. Thiết kế mẫu nghiên cứu.

a) Cơ mẫu: Trong việc tính cỡ mẫu (theo công thức hoặc theo bảng tính sẵn ở sách dịch tễ nào cũng có) bao giờ chúng ta cũng cần phải có một số ý niệm cụ thể về vấn đề nghiên cứu, bao gồm bắt buộc 3 quyết định sau:

- Ước lượng tỷ lệ đặc trưng nghiên cứu nếu nhằm nghiên cứu một tỷ lệ (còn khi nhằm nghiên cứu một số trung bình thì thay vào đó, cần có độ lệch chuẩn của số trung bình đó). Các ý niệm này được xây dựng từ những nghiên cứu trước đó nếu nó còn có giá trị cho lần nghiên cứu này, nếu không, người ta phải làm thăm dò trước với một phần nhỏ của mẫu.

- Ấn định một sai số chọn mẫu chấp nhận cho nghiên cứu, nói khác đi, phải ấn định một mức ý nghĩa nhất định cần thiết và chấp nhận được cho kết quả của nghiên cứu, đây chính là sai lầm alpha cần ấn định (như trong phần lớn các nghiên cứu dịch tễ thông thường, người ta thường lấy mức 5%). Việc ấn định mức ý nghĩa như vậy cũng đồng nghĩa với việc ấn định mức tin cậy của nghiên cứu $1 - \alpha$.

- Xác định một độ chính xác mong muốn (hoặc độ chính xác tuyệt đối, hoặc độ chính xác tương đối). Xác định một độ chính xác nào đó là tương đối khó, nhất là khi dùng độ chính xác tuyệt đối. Cần phải hiểu kỹ về độ chính xác dùng trong công thức, và đặc biệt là cần nắm vững nghiên cứu yêu cầu chính xác đến mức nào mà ấn định ra nó, nếu không cân nhắc đầy đủ thì hoặc sẽ cỡ mẫu không đủ tin cậy vì quá nhỏ, hoặc cỡ mẫu sẽ quá lớn gây những chi phí tốn kém không đáng có.



Có 3 dữ kiện này rồi, chúng ta sẽ thay vào các công thức cho loại mẫu quyết định dùng, sẽ được cỡ mẫu tối thiểu cho nghiên cứu.

b) *Chọn mẫu*: Sau khi đã quyết định quần thể nào sẽ được tiến hành nghiên cứu thì đề cập tốt hơn cả là lấy mẫu ngẫu nhiên. Các cá thể được chọn ngẫu nhiên vào mẫu, sẽ tránh được mọi sai số chọn vào mẫu, nó cũng đảm bảo rằng mọi cá thể đều có một xác suất đã biết được chọn như nhau, điều đó giúp ta áp dụng các nguyên lý lý thuyết thống kê vào kết quả nghiên cứu (thí dụ tính khoảng tin cậy chẳng hạn).

Nói đến chọn mẫu là nói đến những ai sẽ được chọn vào mẫu nghiên cứu (đáp ứng đầy đủ các tiêu chuẩn đã xây dựng) và những người đó sẽ được chọn vào mẫu như thế nào trên nguyên tắc ngẫu nhiên (quy định các bước chọn ngẫu nhiên, và thể thức chọn ngẫu nhiên cho từng bước).

Đây chính là bước rất quan trọng, mà ta thường gọi là thiết kế mẫu, thường được viết dưới dạng một sơ đồ để chọn các đơn vị mẫu vào quần thể nghiên cứu (hay quần thể điều tra) từ quần thể định danh (hay quần thể tham khảo).

c) *Xây dựng quy trình nghiên cứu*.

Nghiên cứu ngang tiến hành trên quần thể là một cuộc điều tra cộng đồng và quy trình nghiên cứu là một bộ câu hỏi. Cần tôn trọng triệt để các quy định trong việc xây dựng bộ câu hỏi sao cho thông tin thu thập được chính xác tối đa. Ngoài ra nó phải có tính thực thi. Để đảm bảo tính thực thi, phải tổ chức thử thách bộ câu hỏi tại quần thể sẽ tiến hành điều tra ở một quy mô nhỏ, do chính người thiết kế bộ câu hỏi và hoặc những điều tra viên sẽ thực hiện cuộc điều tra tiến hành, sau đó, sẽ chỉnh lý cho phù hợp, kể cả về ngôn từ, và bổ sung hoặc giảm nhẹ bộ câu hỏi để đưa ra bộ câu hỏi chính thức làm đại trà.

4. Tuyển chọn và huấn luyện điều tra viên.

Để cho có kết quả khách quan tin cậy, các kỹ thuật "làm mù" được người ta khuyên nên sử dụng tối đa có thể được. Điều này liên quan đến việc tuyển chọn điều tra viên là rất quan trọng: có trình độ để thực hiện tốt những yêu cầu của quy trình nghiên cứu hoặc bộ câu hỏi đặt ra, nhưng cũng cần đảm bảo hỏi và ghi chép lại một cách trung thực các câu hỏi và trả lời của từng đơn vị thống kê (không tự ý hỏi bằng thuật ngữ riêng của mình, không hỏi gợi ý, không hiểu sai hoặc điều chỉnh các trả lời theo ý hiểu biết của riêng mình...).

Không nên dễ dãi thỏa mãn trong việc huấn luyện điều tra viên. Phải huấn luyện sao cho kết quả điều tra của nhiều điều tra viên khác nhau như là của chỉ một điều tra viên tốt đã tiến hành. Muốn được như vậy, nhà nghiên cứu phải huấn luyện thật tỉ mỉ, kỹ lưỡng, từng chi tiết nhỏ một như lời hỏi, cách hỏi, thái độ hỏi, cách cân đo, cách ghi, cách đánh dấu các kết quả trả lời... trong quá trình huấn luyện, phương pháp "đóng vai" là không thể thiếu. Và sau này trong quá trình điều tra thật, họ phải được giám sát chặt chẽ và thay thế nếu cần.

5. Lập kế hoạch và tổ chức triển khai thực hiện điều tra cộng đồng.

Như trên đã nói, các cuộc điều tra trong quần thể thường sử dụng mẫu chùm. Ở đây cũng lấy thí dụ khi sử dụng mẫu chùm trong nghiên cứu. Công tác chuẩn bị khá dài, không trình bày ở đây.

- Chọn ngẫu nhiên chùm.

Nếu nghiên cứu cần chính xác hơn và có đủ dữ kiện, sẽ tính số chùm cần thiết, bằng



không, có thể sử dụng mẫu 30 chòm.

Cần cần nhắc đầy đủ xem chòm trong nghiên cứu cụ thể đó là tập hợp cá thể nào, quy mô nào, tùy thuộc mẫu chòm đích thực hay cải tiến.

Nên tìm mọi điều kiện để có thể chọn chòm vào nghiên cứu tại Ban điều tra nghiên cứu điều tra, tránh những tùy tiện khi về đến địa phương cụ thể.

- Chọn ngẫu nhiên đơn vị thống kê.

Các điều tra viên được huấn luyện làm theo một cách làm thống nhất.

- Tất cả mọi công việc của điều tra viên, từ bước chọn đơn vị thống kê này cho đến mọi công việc sau đó đều phải được đội ngũ giám sát viên giám sát, uốn nắn, bổ sung... nhất là trong những ngày đầu triển khai công việc. Những ưu điểm và thiếu sót của điều tra viên phải được thông báo kịp thời đến đội ngũ điều tra viên để phát huy hoặc để cùng tránh cho thống nhất toàn cuộc điều tra.

6. Tập hợp, xử lý số liệu.

Nên tập hợp vào các bảng tổng hợp cho từng chòm, rồi bảng tổng hợp toàn bộ nghiên cứu, sau đó mới xử lý.

Nếu xử lý trên máy tính, thì đơn giản hơn, nhưng kinh nghiệm đã cho thấy là việc vào số liệu tỉ mỉ, chính xác là cực kỳ quan trọng, khâu này rất dễ nhầm lẫn, và việc phải kiểm tra lại là rất khó khăn, tốn kém, mất nhiều thì giờ. Cũng cần nêu thêm là kết quả đã được xử lý theo hệ thống tin nào.

Trên đây là những điểm khái quát về một cuộc điều tra cộng đồng bằng một nghiên cứu ngang. Tất cả những bài tiếp sau đây sẽ mô tả từng bước chủ yếu của điều tra cộng đồng.

IV. BÀI TẬP MẪU CHÒM

Bài tập 1.

Có một khu vực 10 làng với tình hình dân số như sau:

Làng	Dân số	Cộng dồn dân số
1	1000	1000
2	400	1400
3	200	1600
4	300	1900
5	1200	3100
6	1000	4100
7	1600	5700
8	200	5900
9	350	6250
10	450	6700

Muốn lấy 1 mẫu gồm 3 làng theo kỹ thuật mẫu có xác suất tỷ lệ (pps). Hãy chỉ ra 3 làng vào mẫu.



THƯ VIỆN
HUBT

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

Trả lời:

- Tìm khoảng cách mẫu làng theo số dân: $6700 : 3 = 2233$ ✓
- Chọn 1 số ngẫu nhiên từ 1 đến 2233 (con số ngẫu nhiên có 4 chữ số). Giả sử số ngẫu nhiên đó là 1814.
- Vì 1900 bắt đầu lớn hơn 1814, nên làng số 4 là làng đầu tiên được chọn vào mẫu.
- Cộng 1814 với 2233 được 4047. Làng có số dân cộng dồn 4100 bắt đầu lớn hơn 4047: làng số 6 là làng thứ hai được chọn vào mẫu.
- Cộng tiếp 4047 với 2233 được 6280: làng số 10 là làng thứ ba được chọn vào mẫu.

Bài tập 2.

Muốn tiến hành 1 mẫu chùm, nhưng lại không biết số dân của từng chùm, thì làm thế nào? Có ưu nhược điểm gì?

Trả lời:

- Chọn làng theo mẫu ngẫu nhiên đơn.
- Chọn ra một tỷ số cố định hộ gia đình cho mỗi làng (số hộ điều tra ở từng làng sẽ khác nhau)
- Sẽ rất khó khăn trong việc tiến hành đối với điều tra viên
- Tính chính xác sẽ giảm đi rất nhiều đối với bất kỳ ước lượng nào.

Bài tập 3.

Tính chính xác các ước lượng trong một cuộc điều tra phụ thuộc vào những yếu tố nào?

Trả lời:

- Cỡ mẫu (càng lớn càng chính xác hơn)
- Số chùm và cỡ chùm (cùng 1 cỡ mẫu thì khi số chùm càng nhiều với cỡ chùm càng nhỏ sẽ chính xác hơn khi số chùm càng ít với cỡ chùm càng lớn).
- Tính biến thiên về tiêu thức nghiên cứu giữa các chùm với nhau (tiêu thức nghiên cứu ở chùm nọ khác với ở chùm kia càng nhiều thì tính chính xác của ước lượng càng giảm).

LẬP KẾ HOẠCH VÀ TỔ CHỨC MỘT CUỘC ĐIỀU TRA CỘNG ĐỒNG

I. MỤC TIÊU

1. Nêu được 10 bước lập kế hoạch và 6 bước tổ chức một cuộc điều tra cộng đồng.
2. Ra được những quyết định cần thiết cho cuộc điều tra.
3. Viết được báo cáo về một cuộc điều tra.



II. NỘI DUNG

A. LẬP KẾ HOẠCH CHO MỘT CUỘC ĐIỀU TRA.

1. Bước 1: Phát hiện các vấn đề và nhu cầu của cộng đồng.

Các nguyên lý chung:

Trước hết người nghiên cứu phải xác định rõ mục tiêu và phạm vi của nghiên cứu bằng cách: -

- Xác định những nhu cầu và vấn đề cần điều tra
- Mô tả mục tiêu của cuộc điều tra, những nội dung nào liên quan với nhu cầu của cộng đồng phải được điều tra.

Nhiều cuộc điều tra mắc phải một nhược điểm là tiến hành thu thập quá nhiều thông tin về quá nhiều vấn đề trong một cuộc điều tra. Bằng mọi cách phải tránh lặp lại nhược điểm này.

Một cuộc điều tra về sức khỏe chỉ có giá trị khi các thông tin thu thập được sẽ được sử dụng để đáp ứng một nhu cầu nào đó của một cộng đồng. Các thông tin phải đề cập đến khía cạnh tài chính, công việc phải làm thêm của các cán bộ y tế và thời gian cần thiết để tiến hành điều tra.

Cuộc điều tra phải trả lời các câu hỏi sau:

- Bệnh nhân đến trạm y tế-bệnh viện vì những bệnh gì?
- Những thông tin hiện có liệu có đại diện hay phản ánh đúng tình hình hay không? Nếu không, tại sao?
- Có sự giải thích hợp lý nào về tình trạng bệnh tật hay các nhu cầu của cộng đồng hay không?
- Những thông tin thu thập được có thể giúp để giải quyết các vấn đề hay nhu cầu của cộng đồng hay không?
- Những bệnh nhân đến khám và điều trị có đại diện cho tình hình bệnh tật và sức khỏe của cộng đồng hay không?
- Cộng đồng có được cung cấp những thông tin cần thiết hay không?
- Những thông tin về dân số học và xã hội học có đủ hay không? Còn thiếu những thông tin nào?
- Những nhu cầu ưu tiên của cộng đồng là gì?

2. Bước 2: Quyết định những thông tin cần thu thập.

a) Những thông tin nào là cần thiết giúp cho việc giải quyết những vấn đề sức khỏe và nhu cầu của cộng đồng?

b) Những thông tin nào là cần thiết để đề xuất ra những giải pháp hay để phân bổ các nguồn lực phục vụ sức khỏe hay nhu cầu của cộng đồng.

Trong giai đoạn này cần phải liệt kê những chủ đề quan trọng nhất mà ta cần nghiên cứu. Mỗi chủ đề sẽ có những nhóm nhỏ thông tin như:



THƯ VIỆN
HUBT

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

- Chế độ ăn: những thực phẩm có chứa protein, ngũ cốc, sự thay đổi chế độ ăn theo mùa.

- Nhà ở:

+ Loại nhà, vật liệu, kiến trúc.

+ Số người trong gia đình hiện đang sống ở đó.

+ Số phòng, bếp và nhà vệ sinh.

Một điều quan trọng nhất là phải liệt kê một cách đơn giản những loại thông tin chính trong giai đoạn này và xem xét lại nó nhiều lần trong suốt quá trình lập kế hoạch. Điều này sẽ giúp người lập kế hoạch nhớ thực hiện tất cả các bước và những điểm đã được vạch ra để bảo đảm sự thành công của cuộc điều tra.

Người thiết kế điều tra cũng phải trả lời các câu hỏi sau:

- Cuộc điều tra sẽ tiến hành phỏng vấn bao nhiêu câu hỏi
- Các câu hỏi sẽ được soạn thảo chi tiết đến đâu để giải quyết những vấn đề của cộng đồng?

- Những câu hỏi nào là quan trọng nhất?

3. Bước 3: Tìm hiểu xem những thông tin cần điều tra đã có sẵn hay không.

Một điều quan trọng là phải tìm hiểu xem những thông tin nào ta cần điều tra hiện đã có ở trung tâm y tế, ở bộ y tế, hay ở các trường đại học. Bước này bao gồm những công việc sau:

- Thảo luận về các nhu cầu của cộng đồng và những thông tin cần thiết với các nhà quản lý và nhân viên y tế ở tuyến y tế cơ sở và ở tuyến cao hơn.

- Điều tra xem những thông tin đó có trong sách báo hay các báo cáo đã được công bố về các vấn đề tương tự ở các cộng đồng khác, trong khu vực hay toàn quốc hay không.

- Kiểm tra những thông tin hiện đã có cùng với các nhân viên y tế, nhân viên ở các bộ ngành khác, và ở các trường cao đẳng và đại học.

- Điều tra xem những thông tin cần thiết có liên quan với các yếu tố đặc biệt và điều kiện sống của nhân dân địa phương hay không.

- Nếu người điều tra nhận thấy rằng thông tin mà họ thu được từ kinh nghiệm và sách vở cũng có thể áp dụng được với tình hình của địa phương đó, thì họ có thể đi đến một trong hai kết luận sau:

+ Không cần phải tiến hành cuộc điều tra, và những thông tin hiện có đủ để giải quyết nhu cầu sức khỏe của cộng đồng.

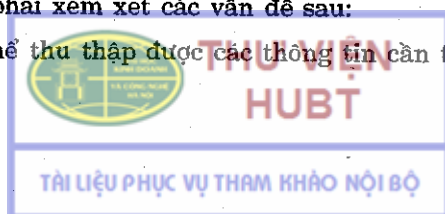
Hoặc:

+ Chỉ cần tiến hành một cuộc điều tra nhỏ để thu thập những thông tin mà hiện chưa có. Kinh nghiệm cho thấy rằng thường vẫn phải tiến hành điều tra vì những thông tin hiện có là chưa đủ và không thích hợp.

4. Bước 4: Xem xét khả năng thực thi thành công cuộc điều tra.

Người điều tra cần phải xem xét các vấn đề sau:

- Cuộc điều tra có thể thu thập được các thông tin cần thiết hay không?



- Những thông tin liệu vẫn có ích sau khi được thu thập và báo cáo hay không?
- Có những nguồn lực cần thiết cho cuộc điều tra hay không?

Việc lập kế hoạch cho một cuộc điều tra nhỏ thường là hấp dẫn và dễ tiến hành. Tuy nhiên, việc lập kế hoạch có thể bỏ qua một số vấn đề quan trọng hoặc những yếu tố làm ảnh hưởng ngược lại đến kết quả. Điều này có thể khắc phục được nếu việc lập kế hoạch được tiến hành một cách có hệ thống, từng bước một. Người lập kế hoạch phải vạch rõ các bước khác nhau của cuộc điều tra và bảo đảm từng bước được thực hiện theo đúng kế hoạch đã vạch ra.

Các vấn đề có liên quan đến cuộc điều tra như nhân lực, thời gian và kinh phí cũng phải được xem xét. Một cuộc điều tra thường đòi hỏi sự tham gia của nhiều người trong một thời gian dài. Đối với các cuộc điều tra nhỏ, các nhân viên y tế ở trạm y tế hay bệnh viện có thể tự tiến hành điều tra được nhưng nó sẽ làm tăng thêm gánh nặng công việc hàng ngày của họ. Cuộc điều tra cũng cần có kinh phí cần thiết để mua giấy in bộ câu hỏi, và những vấn đề hậu cần khác như tàu xe đi lại, ăn ở của điều tra viên.

Hầu hết các cuộc điều tra, thậm chí cả những cuộc điều tra nhỏ, cần phải được tiến hành trong nhiều tháng để chuẩn bị, và tiến hành điều tra trên thực địa.

Sau khi hoàn thành cuộc điều tra, các kết quả phải được phân tích và viết trong báo cáo cùng với kết luận. Một số ít cuộc điều tra có thể được hoàn thành trong một thời gian ngắn hơn 6 tháng, nhưng đa số các cuộc điều tra kéo dài một năm hay hơn.

Trong giai đoạn này, những đòi hỏi cần thiết cho cuộc điều tra không cần phải viết quá chi tiết mà chỉ là những nét đại cương để bảo đảm thực hiện các quyết định đã đưa ra và không quên một chi tiết nào.

Trong giai đoạn này, người lập kế hoạch cũng phải trả lời các câu hỏi sau:

- Có cần thiết phải thu thập các mẫu bệnh phẩm để xét nghiệm hay không?
- Cuộc điều tra sẽ tiến hành bao nhiêu lâu?
- Những thông tin thu thập được liệu vẫn có ích lợi khi hoàn thành cuộc điều tra và viết báo cáo hay không?
- Ai sẽ tiến hành cuộc điều tra?

5. Bước 5: Những quyết định đầu tiên trong việc lập kế hoạch.

a) Liệt kê các câu hỏi chính mà cuộc điều tra phải trả lời và ai trong cộng đồng sẽ phải trả lời những câu hỏi này.

b) Vạch ra phương pháp (kế hoạch lấy mẫu) cần thiết để lựa chọn quần thể và nơi điều tra.

c) Quyết định thời gian cần thiết để hoàn thành cuộc điều tra phỏng vấn ở thực địa.

Để tiến hành thành công cuộc điều tra, cần phải thỏa mãn những điều kiện sau:

- Trong cộng đồng phải có đủ một số người có kiến thức và kinh nghiệm để trả lời những câu hỏi của cuộc điều tra. Ví dụ, không nên hỏi những câu hỏi về mặt kỹ thuật đối với nông dân, hay hỏi những người thành phố về công việc đồng áng vì họ không có kiến thức và kinh nghiệm về những vấn đề này.



THƯ VIỆN
HUBT

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

- Việc xác định và đến thăm hỏi, phỏng vấn những người trong cộng đồng có hiểu biết và kinh nghiệm liên quan đến vấn đề cần điều tra phải dễ dàng thực hiện - Ví dụ, nếu cuộc điều tra muốn biết các gia đình trông nom những trẻ em và người lớn bị tàn tật như thế nào, và các gia đình đã gặp những khó khăn gì. Cần phải đưa ra phương pháp xác định và tìm kiếm những gia đình đó - Chương "kỹ thuật điều tra chọn mẫu" sẽ cung cấp những chỉ dẫn và nguyên tắc lựa chọn một nhóm người đại diện cho cộng đồng đó.

- Những người được phỏng vấn phải sẵn sàng cộng tác và trả lời các câu hỏi điều tra. Điều này rất quan trọng vì đôi lúc họ dè dặt không muốn thảo luận về các vấn đề của họ với những người xa lạ.

- Ở tuyến y tế cơ sở, nhân viên y tế thường có thời gian và sẵn sàng tham gia vào việc tổ chức và giám sát cuộc điều tra, từ bước đầu tiên lập kế hoạch đến bước viết báo cáo. Những câu hỏi cần phải được trả lời trong bước điều tra này là:

- Ai sẽ là đối tượng phù hợp nhất để phỏng vấn.
- Ai sẽ là người phỏng vấn tốt nhất?
- Những thông tin nào phải được thu thập bằng quan sát trực tiếp chứ không phải bằng phỏng vấn?
- Những thông tin nào cần phải có sự hỗ trợ của thăm khám và xét nghiệm?
- Những thông tin nào cần phải thu thập qua thảo luận với những nhà chuyên gia hay những người lãnh đạo?
- Những nguồn lực nào của trạm y tế có thể phục vụ cho cuộc điều tra? Cần phải sắp xếp công việc như thế nào để các nhân viên y tế có thể tham gia vào cuộc điều tra?
- Những yếu tố nghiên cứu có thay đổi trong thời gian điều tra không? Làm thế nào để khắc phục điều đó?
- Mọi ngày hay mỗi tuần, người điều tra viên có thể tiến hành được bao nhiêu cuộc phỏng vấn?

Trước khi chuyển sang giai đoạn tiếp theo, hãy kiểm tra và xem lại bảng liệt kê những thông tin cần điều tra ở bước 2.

6. Bước 6: Chọn đối tượng phỏng vấn.

Những quyết định về kế hoạch chọn mẫu.

- a) Đối tượng phỏng vấn được lựa chọn như thế nào?
- b) Số đối tượng phỏng vấn là bao nhiêu?
- c) Cuộc điều tra được tổ chức như thế nào để các điều tra viên có thể tiếp cận được đối tượng phỏng vấn?

Bước này cần phải được lập kế hoạch một cách cụ thể.

Phương pháp chọn đối tượng phỏng vấn không đúng sẽ dẫn đến kết quả sai lạc và không có giá trị. Mục đích của cuộc điều tra là cung cấp thông tin về những vấn đề hiện tại của cộng đồng. Mục tiêu này chỉ đạt được khi tất cả mọi người trong cộng đồng có cơ hội như nhau được phỏng vấn. Những người giàu và nghèo, ốm và khỏe mạnh, trẻ và già phải được chọn vào mẫu điều tra theo một tỷ lệ giống như cấu trúc của cộng đồng.



Những điều tra viên không được phép tùy tiện chọn người để phỏng vấn. Trước khi tiến hành cuộc điều tra họ phải được phổ biến về kế hoạch điều tra chọn mẫu và ai sẽ được chọn là đối tượng phỏng vấn. Chương sau sẽ trình bày các kỹ thuật chọn mẫu. Cách đơn giản nhất để chọn ngẫu nhiên đối tượng điều tra là viết mỗi tên lên một mảnh giấy, và cho vào một cái hộp và lắc nó lên. Sau đó một người nào đó sẽ rút từng mẫu giấy có tên ra cho đến khi đạt đủ cỡ mẫu cần điều tra thì thôi. Những tên người được bốc thăm sẽ là những đối tượng điều tra và những người còn lại thì sẽ không tham gia vào điều tra.

Một số câu hỏi của bước này là:

- Cuộc điều tra sẽ tiến hành ở bao nhiêu làng?
- Những làng điều tra được lựa chọn như thế nào?
- Cần phải làm những bước gì để xác định đối tượng điều tra?
- Các cuộc phỏng vấn sẽ được nối tiếp nhau như thế nào trong thời gian điều tra?
- Cỡ mẫu của cuộc điều tra là bao nhiêu?

7. Bước 7: Kinh phí cho điều tra.

a) Ước lượng kinh phí để tiến hành cuộc điều tra bao gồm kinh phí cho điều tra viên, đi lại và các chi phí khác.

b) Kiểm tra xem những nguồn lực nào như điều tra viên, phương tiện đi lại và các nguồn khác đã có sẵn ở địa phương.

Để bảo đảm đủ kinh phí tiến hành điều tra, cần phải ước lượng số điều tra viên và các chi phí hậu cần khác.

Kinh phí điều tra phụ thuộc vào nhiều yếu tố như cỡ mẫu điều tra, thời gian xác định và tiếp cận đối tượng phỏng vấn. Nhiều khi việc tìm kiếm hộ gia đình đã được chọn không phải là dễ và đôi khi họ đi vắng. Thời gian phỏng vấn mỗi một người có thể từ vài phút đến vài giờ tùy theo cuộc điều tra. Đối với các cuộc điều tra nhỏ, nhân viên y tế địa phương có thể tự tiến hành được, nhưng đối với các cuộc điều tra lớn, đòi hỏi phải có sự tham gia giúp đỡ của những người ở các nơi khác đến. Những bước chủ yếu để dự trù kinh phí cho một cuộc điều tra là:

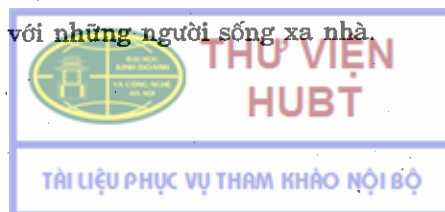
- Ước lượng số cuộc phỏng vấn mà mỗi điều tra viên có thể hoàn thành trong một ngày hay một tuần. Tốt nhất nên tiến hành thử nghiệm trước một vài cuộc phỏng vấn và ghi lại thời gian cần thiết để hoàn thành một cuộc phỏng vấn.

- Ước lượng cỡ mẫu cần điều tra. Cỡ mẫu là số đối tượng cần được phỏng vấn hay số bệnh nhân cần thăm khám tùy thuộc vào mục đích của cuộc điều tra. Một nguyên tắc là cỡ mẫu tối thiểu cho mỗi cuộc điều tra là 50.

- Ước lượng thời gian cần thiết để hoàn thành cuộc điều tra.

Những chi phí cần thiết để tiến hành cuộc điều tra là:

- Tiền lương cho điều tra viên, giám sát viên và những người tổ chức điều tra.
- Tiền tàu xe vận chuyển điều tra viên.
- Tiền ăn ở cho điều tra viên đối với những người sống xa nhà.



- Tiền giấy và in ấn bộ câu hỏi điều tra.
- Tiền xử lý, phân tích và phân phát các kết quả điều tra.

Ở bước này, người lập kế hoạch phải trả lời một số câu hỏi sau:

- Phương tiện đi lại có cần thiết không? Các phương tiện giao thông công cộng có phù hợp không?
- Những chi phí chính cho cuộc điều tra là gì? Có bao gồm kinh phí cho những việc mà ta chưa biết hay không?
- Kinh phí cho cuộc điều tra có thể được cung cấp từ nguồn nào?

8. Bước 8: Những quyết định cuối cùng về cuộc điều tra.

Người lập kế hoạch phải đưa ra những quyết định cuối cùng về:

- Những thông tin cần thiết phải thu thập.
- Phạm vi và mức độ của cuộc điều tra.
- + Cuộc điều tra sẽ được tiến hành ở những vùng địa dư nào?
- + Cần phải giới hạn những vấn đề nghiên cứu hay không?
- + Những thông tin có độ chính xác thấp có đủ để giải quyết những nhu cầu của cộng đồng hay không?
- + Những thông tin thu thập được có giúp đỡ việc xác định được các nhu cầu của cộng đồng không? Nếu không, có loại bỏ việc thu thập những thông tin này ra khỏi cuộc điều tra hay không?

Nếu không thiết kế cẩn thận, cuộc điều tra có thể trở thành quá lớn, quá tốn kém và quá phức tạp.

Bước chuẩn bị tiếp theo cho cuộc điều tra là phải bảo đảm chắc chắn rằng cuộc điều tra có thể được tiến hành với các nguồn lực hiện có (nhân lực, phương tiện đi lại, và tiền). Nếu các nguồn lực không đủ để tiến hành điều tra, cuộc điều tra trở nên vô ích và lãng phí. Nếu cuộc điều tra bị dừng do không đủ kinh phí thì kết quả thu được không đại diện cho toàn bộ quần thể và đó là những kết quả hoàn toàn sai. Do đó, tốt nhất là nên tiến hành một cuộc điều tra nhỏ hơn là tiến hành một cuộc điều tra lớn khi mà kinh phí cho cuộc điều tra nhỏ. Một cuộc điều tra không nên bắt đầu trừ khi đã có đủ kinh phí để tiến hành điều tra. Đối với những cuộc điều tra nhỏ, không đòi hỏi nhiều nhân lực tham gia, nhưng đòi hỏi một khoản kinh phí nhất định. Nếu có sự tham gia của các nhân viên y tế ở trạm y tế hay bệnh viện thì cuộc điều tra không đòi hỏi nhiều kinh phí lắm.

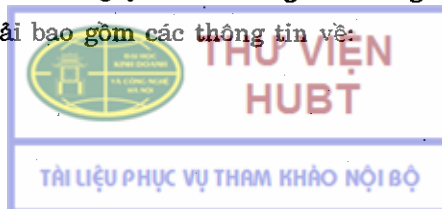
9. Bước 9: Chuẩn bị bộ câu hỏi điều tra.

Cần phải xác định là bộ câu hỏi bao gồm bao nhiêu câu hỏi và mục đích của mỗi câu hỏi là gì?

Để thu thập thông tin về một chủ đề nào đó, cần phải có một số câu hỏi ngắn gọn, rõ ràng và dễ hiểu. Mỗi chủ đề đòi hỏi một số câu hỏi nhất định. Quá trình thiết kế bộ câu hỏi sẽ được trình bày chi tiết ở chương "Thiết kế bộ câu hỏi".

Dưới đây trình bày một số nguyên tắc đơn giản nhưng rất quan trọng:

- Mỗi bộ câu hỏi phải bao gồm các thông tin về:



- Tên làng hay địa chỉ.
- Tên gia đình hay người được phỏng vấn
- Ngày tiến hành phỏng vấn.

Những thông tin này được đặt ở phần đầu tiên của bộ câu hỏi.

- b) Các câu hỏi phải đơn giản, ngắn gọn và dễ hiểu.
- c) Chỉ hỏi những câu hỏi quan trọng và cần thiết. Cố gắng giữ bộ câu hỏi càng ngắn càng tốt.
- d) Bộ câu hỏi được bắt đầu bằng những câu hỏi dễ trả lời và về những chủ đề quen thuộc.
- e) Dưới mỗi câu hỏi, dành một khoảng trống để điều tra viên có thể ghi chép những điều quan tâm. Nếu khoảng trống này quá hẹp, chữ viết sẽ nhỏ và khó đọc.
- f) Đặt các câu hỏi theo một trình tự hợp lý có nghĩa là những câu hỏi về một chủ đề được nhóm thành một phần/chương. Điều này sẽ giúp cho người được phỏng vấn hiểu cần phải cung cấp những thông tin gì.

Một bộ câu hỏi tốt đòi hỏi phải được thiết kế một cách cẩn thận. Việc thiết kế này có thể mất nhiều tuần. Nếu có thể được, nên tiến hành thiết kế bộ câu hỏi sớm nhiều tháng trước khi tiến hành điều tra chính thức.

Một số câu hỏi quan trọng phải được trả lời khi tiến hành bước chuẩn bị này:

- + Bộ câu hỏi bao gồm bao nhiêu câu hỏi.
- + Trong bộ câu hỏi, có chủ đề nào là khó trả lời đối với đối tượng phỏng vấn hay không?
- + Bộ câu hỏi có đủ thông tin để người điều tra có thể xác định và điều tra lại hộ gia đình hay một người nào đó không?
- + Có những chỉ dẫn nào giúp cho điều tra viên bắt đầu phỏng vấn ở những phần/chương khác nhau về các chủ đề khác nhau không?
- + Có đủ chỗ để khoanh tròn hay đánh dấu vào những câu trả lời đúng hay để viết thêm vào hay không?
- + Có đủ chỗ trống giữa các câu hỏi để cho điều tra viên dễ đọc và điền câu trả lời hay không?
- + Các câu hỏi có được xếp theo một trình tự hợp lý hay không?

10. Bước 10: Chuẩn bị những chỉ dẫn cho người điều tra.

Sau khi thiết kế xong bộ câu hỏi, cần phải chuẩn bị những chỉ dẫn ngắn gọn cho người điều tra. Những chỉ dẫn này cũng có thể được sử dụng khi tiến hành huấn luyện điều tra viên trước khi điều tra. Những chỉ dẫn này nhằm 3 mục tiêu chính:

- a. Giải thích một cách kỹ lưỡng cách chọn đối tượng điều tra (kỹ thuật chọn mẫu).

Trong mỗi cuộc điều tra, phải nhấn mạnh đến một số tình huống đặc biệt có thể xảy ra như sẽ phải làm gì khi đối tượng phỏng vấn được chọn không có nhà hoặc chuyển đi chỗ khác, hoặc không hiểu ngôn ngữ của người điều tra hoặc từ chối tham gia vào cuộc phỏng vấn...



b. Giải thích cách sử dụng và điền bộ câu hỏi cũng như làm thế nào để động viên đối tượng điều tra cộng tác tham gia vào cuộc phỏng vấn.

c. Là tài liệu tham khảo giúp cho điều tra viên xử lý khi gặp những tình huống khó khăn.

Ở một số cộng đồng, người dân nói nhiều thứ tiếng khác nhau. Đối tượng phỏng vấn có thể sử dụng một ngôn ngữ khác hoặc không hiểu ngôn ngữ của người điều tra có thể khắc phục vấn đề này bằng cách:

+ Mỗi nhóm điều tra viên bao gồm hai người, mỗi một người có thể nói và hiểu được các ngôn ngữ khác nhau trong cộng đồng.

+ Hướng dẫn điều tra viên cách nhận biết ngôn ngữ của đối tượng phỏng vấn và hẹn vào một ngày khác một người nào khác hiểu ngôn ngữ của đối tượng phỏng vấn sẽ quay lại để phỏng vấn. Điều này có thể nhờ sự giúp đỡ của một người hàng xóm nào đó.

Việc chuẩn bị bộ câu hỏi và các chỉ dẫn điều tra phải được tiến hành ngay cả trong trường hợp chỉ có một điều tra viên. Đối với các cuộc điều tra lớn hơn, những chỉ dẫn này càng trở nên quan trọng.

Những câu hỏi của bước lập kế hoạch này:

- Những điều tra viên có được chỉ dẫn một cách kỹ lưỡng cách bắt đầu cuộc phỏng vấn hay không?

- Liệu đối tượng điều tra có biết tại sao họ được lựa chọn để phỏng vấn hay không? Điều tra viên có được hướng dẫn cách giải thích mục đích của cuộc điều tra cho đối tượng điều tra hay không?

- Thời gian và địa điểm phỏng vấn có đóng vai trò quan trọng hay không? Nếu có, có phổ biến cho người điều tra biết hay không? Họ có được hướng dẫn cách lựa chọn thời gian và địa điểm phỏng vấn thích hợp hay không?

- Liệu đối tượng phỏng vấn có được bảo đảm chắc chắn rằng những câu trả lời của họ được giữ bí mật và không cho một người nào khác biết hay không?

B. TỔ CHỨC MỘT CUỘC ĐIỀU TRA

1. **Bước 1:** Chuẩn bị cộng đồng cho cuộc điều tra.

a) Thông báo cho các nhà lãnh đạo của cộng đồng biết kế hoạch điều tra và thảo luận để họ đồng ý và cộng tác tiến hành điều tra. Giải thích cho họ mục tiêu của cuộc điều tra.

b) Chuẩn bị cộng đồng cho cuộc điều tra. Những người sau phải được thông báo:

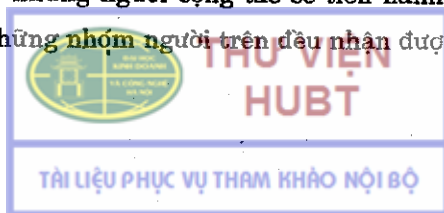
- Các nhân viên y tế và cán bộ lãnh đạo ở tuyến trên.

- Các nhân viên hành chính và y tế địa phương.

- Đối tượng điều tra hay người đại diện của họ.

- Những cán bộ hay những người cộng tác sẽ tiến hành cuộc điều tra.

Không phải tất cả những nhóm người trên đều nhận được những thông tin giống nhau.



Nhưng họ phải biết điều gì sẽ xảy ra.

c) Để có được sự cộng tác và hỗ trợ hoàn toàn cho cuộc điều tra của các nhà lãnh đạo, trước khi tiến hành điều tra, phải chọn thời điểm thích hợp thông báo cho họ về:

- Kế hoạch của cuộc điều tra, thời gian bắt đầu, và kết thúc cuộc điều tra.
- Mục tiêu của cuộc điều tra và ai sẽ nhận được báo cáo kết quả và kết luận của cuộc điều tra.
- Kinh phí cần thiết cho điều tra, và liệu họ có thể giúp đỡ một phần kinh phí được không.

d) Để có được sự cộng tác hoàn toàn của cộng đồng mà ta nghiên cứu, mọi người có liên quan phải được thông báo rằng:

- Sẽ có một cuộc điều tra và một số người trong số họ sẽ được chọn lựa để phỏng vấn.
- Mục đích của cuộc điều tra mà họ không cần phải biết một cách chi tiết.
- Cuộc điều tra sẽ có ích lợi gì đối với họ và cộng đồng của họ.
- Những thông tin mà họ cung cấp sẽ được giữ bí mật hoàn toàn và không một ai khác biết được những thông tin này.

2. Bước 2: Thử nghiệm phương pháp điều tra.

Bước này bao gồm:

- Thử nghiệm các câu hỏi và bộ câu hỏi.
- Nếu có thể được, tiến hành một nghiên cứu thử nhỏ.

Trước khi tiến hành điều tra chính thức, thử nghiệm bộ câu hỏi trên lớp tập huấn hay ở một số người trong cộng đồng. Sau khi tiến hành như vậy, thường một số câu hỏi cần phải sửa lại về cách hành văn. Điều tra thử sẽ làm bộc lộ những điểm sau:

a) Những câu hỏi khó hiểu không rõ ràng và có thể bị hiểu sai. Phải đặt lại câu hỏi sao cho rõ ràng và dễ hiểu.

b) Không đủ chỗ để điền các câu trả lời.

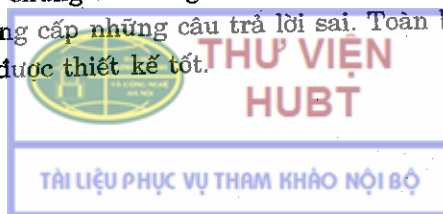
c) Đối tượng phỏng vấn có thể từ chối trả lời các câu hỏi hay đưa ra các câu trả lời không rõ ràng. Cố gắng đặt lại câu hỏi này và đặt nó vào phần cuối của cuộc phỏng vấn hay bỏ câu này đi.

d) Một số câu hỏi về các chủ đề mà người được phỏng vấn không biết hay không có kinh nghiệm. Loại bỏ những câu hỏi này đi, bởi vì chúng không cung cấp được những thông tin cần thiết.

e) Thời gian để hoàn thành một cuộc phỏng vấn.

Đối tượng phỏng vấn có thể mệt mỏi và chán nản nếu như bộ câu hỏi quá dài hay họ không hiểu gì về nó. Nếu vậy, chúng ta phải làm cho bộ câu hỏi ngắn đi bằng cách loại bỏ những câu hỏi không cần thiết.

Bộ câu hỏi đã được thiết kế sẵn là một phương tiện tốt nhất để thu thập thông tin. Nếu nó không được thiết kế tốt, chúng ta không thể thu thập được những thông tin đúng và người được phỏng vấn sẽ cung cấp những câu trả lời sai. Toàn bộ cuộc điều tra sẽ bị thất bại nếu bộ câu hỏi không được thiết kế tốt.



Những câu hỏi phải được trả lời trong bước lập kế hoạch này là:

- Bộ câu hỏi này có được thử nghiệm trước không? Kết quả của thử nghiệm này có được dùng để làm cho các câu hỏi rõ ràng hơn sắp xếp theo một trật tự hợp lý và loại bỏ những câu không cần thiết hay không?

- Nghiên cứu thử có thể tiến hành được không? Nghiên cứu thử có thể được tiến hành ở đâu?

- Nghiên cứu thử sẽ được tiến hành ở quy mô nào? Ai sẽ tiến hành nghiên cứu thử này?

- Cuộc phỏng vấn sẽ kéo dài bao lâu?

- Có cần thay đổi kế hoạch điều tra sau khi tiến hành nghiên cứu thử hay không?

- Kết quả của nghiên cứu thử có được đưa vào kết quả của điều tra hay không?

3. Bước 3: Huấn luyện điều tra viên.

Huấn luyện điều tra viên là bước cuối cùng trước khi tiến hành điều tra.

Những vấn đề cần huấn luyện điều tra viên là:

a) Giải thích mục tiêu điều tra và phạm vi của nó.

Nếu điều tra viên hiểu được mục đích của cuộc điều tra thì họ có thể xử lý được tốt hơn các tình huống bất thường xảy ra trong quá trình điều tra. Họ sẽ quan tâm hơn tới cuộc điều tra khi họ hiểu tại sao họ sẽ tiến hành cuộc điều tra.

b) Giải thích những chỉ dẫn đã được viết sẵn về những người họ sẽ điều tra và làm thế nào để tiếp cận được những người đó.

Điều tra viên có thể sẽ không làm theo những chỉ dẫn nếu họ không hiểu được tầm quan trọng của những chỉ dẫn đó.

c) Cùng với điều tra viên, nghiên cứu từng câu hỏi một của bộ câu hỏi. Tiến hành tập sử dụng bộ câu hỏi này để phỏng vấn các điều tra viên khác trên lớp hay những thành viên khác của nhóm điều tra hay các nhân viên y tế ở trạm y tế. Điều này sẽ làm cho điều tra viên làm quen với bộ câu hỏi và họ sẽ sử dụng bộ câu hỏi để phỏng vấn theo một phương pháp thống nhất đã được vạch ra.

d) Cung cấp cho điều tra viên những lời khuyên về cách xử trí những tình huống thường xảy ra ở thực địa, bao gồm cả cách xử trí nếu một đối tượng điều tra từ chối cộng tác hay không có nhà hay đã chuyển đi nơi khác.

e) Giải thích cho điều tra viên cách tốt nhất để bắt đầu và kết thúc cuộc phỏng vấn. Bắt đầu cuộc phỏng vấn không phải lúc nào cũng dễ dàng. Nếu mở đầu phỏng vấn không tốt sẽ làm cho đối tượng phỏng vấn nghi ngờ và không cộng tác. Trong trường hợp này, người điều tra viên sẽ không thu thập được những thông tin mà họ mong muốn. Đôi khi, đối tượng phỏng vấn chủ ý trả lời sai vì họ không tin tưởng ở người điều tra.

Những câu hỏi phải được trả lời cho bước này là:

- Lớp huấn luyện điều tra viên được tổ chức như thế nào?

- Điều tra viên có hiểu kỹ tầm quan trọng của việc tuân thủ những chỉ dẫn về chọn mẫu hay không?



- Điều tra viên có phải viết lại những điều nên làm trong cuộc điều tra hay không? Họ có được cung cấp những ví dụ về những điều nên làm trong cuộc điều tra hay không?

- Điều tra viên có được phổ biến về cách phỏng vấn và nhấn mạnh các câu hỏi hay không?

- Điều tra viên có hiểu rằng họ phải phỏng vấn theo những câu hỏi đã được in sẵn trong bộ câu hỏi và không được thay đổi cách hành văn của những câu hỏi đó hay không?

- Có tiến hành phỏng vấn thử bộ câu hỏi trên lớp hay không?

- Điều tra viên có được cung cấp danh sách ngắn những câu hỏi mà đối tượng phỏng vấn có thể hỏi họ hay không? Họ có được cung cấp những chỉ dẫn để trả lời những câu hỏi đó hay không?

4. Bước 4: Bắt đầu tiến hành điều tra.

a) Bố trí các cuộc họp và thảo luận thường xuyên với các điều tra viên.

b) Kiểm tra lại những bộ câu hỏi mà các điều tra viên đã hoàn thành.

Một điều rất quan trọng là phải họp với các điều tra viên mỗi ngày một lần. Trong những lần họp này, phải tiến hành những công việc sau:

- Thảo luận với các điều tra viên về những khó khăn và tình huống bất thường mà họ gặp phải trong quá trình điều tra. Cần phải cung cấp cho họ những chỉ dẫn và hỗ trợ để họ có thể xử trí được khi gặp lại những tình huống đó.

- Cùng với điều tra viên, kiểm tra các bộ câu hỏi đã hoàn thành. Một số câu hỏi có thể bị diễn sai, bỏ sót. Nếu những điều này được phát hiện sớm sau khi phỏng vấn, người điều tra có thể nhớ lại được điều gì đã xảy ra. Trong một số cuộc điều tra, điều tra viên có thể sẽ phải quay lại hộ gia đình để bổ sung những thông tin còn thiếu.

Nếu việc lập kế hoạch và chuẩn bị điều tra được tiến hành kỹ lưỡng cuộc điều tra sẽ không gặp phải vấn đề gì. Những khó khăn không dự đoán trước như lụt lội, hay người điều tra viên bị ốm đôi khi có thể xảy ra. Trong trường hợp này, người tổ chức điều tra phải sắp xếp thời gian biểu điều tra sao cho phù hợp để đối phó với những khó khăn này. Nhưng nhìn chung, cuộc điều tra phải được tiến hành như đã dự định.

Một số điều tra viên sẽ có nhiều khó khăn với người được phỏng vấn, và họ từ chối không trả lời. Một điều quan trọng là phải phát hiện ngay những người điều tra viên như thế và phải thay thế một người khác phù hợp hơn.

Những câu hỏi cho bước điều tra này là:

- Điều tra viên có gặp phải vấn đề đặc biệt gì không?

Đối tượng phỏng vấn có cộng tác trả lời các câu hỏi điều tra không?

- Có cần phải quay lại để điều tra người vắng nhà hay người mà điều tra viên gặp nhiều khó khăn trong cuộc phỏng vấn đầu tiên hay không?

- Có tính đến khả năng là lúc ban đầu có thể cuộc điều tra sẽ gặp nhiều sai sót và tiến trình của cuộc điều tra sẽ chậm hơn dự tính không?

5. Bước 5: Xử lý thông tin thu được.

Sau khi hoàn thành các cuộc phỏng vấn, các bộ câu hỏi đã hoàn thành được thu thập

và kiểm tra. Bước tiếp theo là xử lý các thông tin thu thập được từ các bộ câu hỏi. Những thông tin này chỉ có ích khi nó được rút ra từ các bộ câu hỏi và được trình bày theo một cách dễ hiểu.

Trước khi tiến hành xử lý thông tin cần phải tiến hành hai việc sau:

- Sắp xếp các bộ câu hỏi một cách ngăn nắp và theo thứ tự.

Các bộ câu hỏi có thể bị làm xáo trộn trong quá trình tách thông tin từ các bộ câu hỏi, và sau đó chúng phải được sắp xếp lại theo thứ tự trước khi tiến hành các bước tiếp theo.

- Liệt kê tóm tắt những thông tin cần lấy ra từ bộ câu hỏi.

Bắt đầu tách những thông tin đơn giản nhất, và những số liệu quan trọng nhất có ý nghĩa là những thông tin có liên quan trực tiếp đến mục đích của cuộc điều tra. Có hai phương pháp cơ bản để mô tả phân tích và trình bày những thông tin tóm tắt này:

+ Tóm tắt và trình bày thông tin dưới dạng các bảng số liệu.

+ Mô tả và nhận xét những thông tin đã được trình bày trong các bảng đó.

Trong báo cáo điều tra, cả hai phương pháp sẽ được sử dụng.

Giữa các bảng số liệu là những đoạn văn mô tả lý do thu thập những số liệu này và các kết luận rút ra từ các bảng đó.

Đối với các cuộc điều tra nhỏ, không cần thiết đòi hỏi phải có các kiến thức và kỹ thuật thống kê để phân tích và xử lý kết quả.

Đối với hầu hết các cuộc điều tra nhỏ, chỉ cần bốn quy trình thống kê đơn giản như sau:

+ Tính toán số trung bình

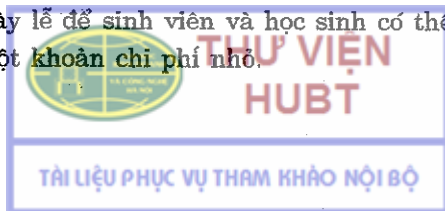
+ Tính toán các tỷ lệ phần trăm.

+ Tính toán các chỉ số như tỷ số, tỷ lệ, tỷ suất và các tỷ lệ khác nếu thấy cần thiết như tỷ lệ chết.

+ Trình bày các số liệu thành các bảng.

Các kết quả quan trọng nhất của cuộc điều tra sẽ được trình bày thành các bảng thống kê. Bốn quy trình này sẽ được thảo luận chi tiết hơn trong chương "Trình bày kết quả điều tra".

Các thầy giáo địa phương, đặc biệt là các thầy giáo dạy số học sẽ đóng vai trò quan trọng trong bước xử lý thông tin này. Tất cả các phép tính toán phải được một người khác làm lại để bảo đảm kết quả chính xác. Người thứ hai này không được biết kết quả của người thứ nhất cho đến khi người đó tính xong. Kết quả của hai người này sau đó được so sánh và kiểm tra lại bởi một người thứ ba. Nếu có sai khác, thầy giáo là người tốt nhất phát hiện và sửa chữa những sai sót này. Một số thầy giáo quan tâm đến công việc này có thể đề nghị một số học trò giỏi của họ tham gia giúp đỡ thực hiện bước xử lý thông tin này. Công việc này phải được giám sát rất cẩn thận. Sinh viên của các trường đại học và cao đẳng cũng có thể tham gia vào công việc này. Công việc xử lý số liệu nên tiến hành vào các ngày lễ để sinh viên và học sinh có thể tham gia được. Nó không tốn kém, nếu có chỉ là một khoản chi phí nhỏ.



Trong quá trình lập kế hoạch điều tra có thể bỏ sót một số khía cạnh nào đó mà chỉ có thể phát hiện được trong khi xử lý số liệu và viết báo cáo.

Có một phương pháp đơn giản để đề phòng sự bỏ sót. Ngay trong những bước lập kế hoạch đầu tiên, phải liệt kê ra các mục sau:

- Tên của các phần trong bảng báo cáo.
- Các bảng và thông tin cần biết ở mỗi phần.

Bảng liệt kê này sẽ giúp cho người lập kế hoạch nhớ các mục mà anh ta đã quên đưa vào bộ câu hỏi. Nếu sự bỏ sót được phát hiện trước khi điều tra chính thức, chúng vẫn có thể được bổ sung vào bộ câu hỏi.

Xử lý thông tin từ bộ câu hỏi có thể được tiến hành bằng một trong ba phương pháp sau:

- Phương pháp ghi chép bằng tay.
- Phương pháp dùng máy cơ học như máy lựa chọn phiếu lỗ, máy chữ đánh bảng.
- Máy vi tính.

Đối với các cuộc điều tra nhỏ, phương pháp ghi chép bằng tay thường là thích hợp vì nó rẻ, không đòi hỏi trang thiết bị gì đặc biệt và không đòi hỏi người sử dụng phải được đào tạo ở trình độ cao. Phương pháp này dựa vào một phiếu xử lý dễ dàng được tạo ra bởi một người thư ký đánh máy. Nó cũng có thể được vẽ bằng tay sử dụng bút và thước kẻ. Dưới đây là một ví dụ về phía xử lý thông tin.

Hình 1- Phiếu xử lý thông tin

Số lần đến khám bệnh ở trạm y tế xã		
Số lần		Tổng cộng
0	### ##	13
1	///	3
2	/	1
3		0
4	/	1
	Tổng cộng	19

Kết quả chính

- Số gia đình được phỏng vấn 19
- Số gia đình có ít nhất 1 lần đến khám bệnh ở trạm Y tế xã trong 4 tuần vừa qua 6
- Số lần đến khám ở trạm y tế xã 10

Đối với mỗi biến số, hay mỗi mục thông tin, các đơn vị được ghi ở bên lề. Ví dụ ở hình 1, các mục được ghi là số lần đến trạm y tế xã khám được biểu thị bằng 0,1,2,3,4 lần. Đối với các biến số khác như cân nặng và chiều cao của trẻ, cần phải lựa chọn các khoảng giá trị kinh điển và viết vào lề bên trái. Mỗi biến số đòi hỏi một phiếu xử lý thông tin riêng.

Quá trình xử lý được tiến hành như sau: Mỗi câu hỏi đều có câu trả lời hay một giá trị nào đó. Chẳng hạn ta lấy ví dụ một câu hỏi như "Trong 4 tuần vừa qua, bạn hay người trong gia đình của bạn đến khám ở trạm y tế bao nhiêu lần?" Nếu câu trả lời là "không", ta sẽ vạch một đường ở hàng có số lần là 0. Nếu bộ câu hỏi sau có câu trả lời là 2 thì ta lại vạch một đường ở hàng có số lần đến khám là 2 trên cùng một phiếu xử lý. Cứ tiếp tục như vậy cho các bộ câu hỏi sau cho đủ 5 vạch (4 vạch đều là vạch thẳng đứng, và vạch thứ 5 là vạch chéo, 卅). Điều này làm cho việc tính toán trở nên dễ dàng hơn, vì các vạch ngang và chéo làm thành từng nhóm 5 một. Quá trình này cứ được tiếp tục như thế cho đến khi toàn bộ các câu trả lời của bộ câu hỏi đã được tóm tắt trên phiếu xử lý. Sau đó tính tổng của từng hàng trong phiếu và viết vào lề phía bên phải và cuối cùng tính tổng cộng toàn bộ các hàng.

6. Bước 6: Viết báo cáo và phân phát báo cáo kết quả của cuộc điều tra.

6.1 *Viết báo cáo.*

Đây là bước cuối cùng của cuộc điều tra. Báo cáo có thể được viết thành nhiều dạng khác nhau. Hầu hết các báo cáo được chia thành nhiều phần và thường có ít nhất 5 phần riêng biệt như sau:

a) Tóm tắt những kết quả và kết luận chính của cuộc điều tra. Phần này không được dài quá 2 - 3 trang.

b) Phần giới thiệu. Phần này mô tả các vấn đề và nhu cầu dẫn đến việc tiến hành cuộc điều tra. Cuối phần này sẽ trình bày một cách rõ ràng các mục tiêu của cuộc điều tra.

c) Phương pháp và quy trình điều tra: Phần này mô tả phương pháp điều tra bao gồm kế hoạch chọn mẫu, huấn luyện điều tra viên, và các khía cạnh có liên quan và ảnh hưởng đến cuộc điều tra.

d) Kết quả và kết luận. Phần này thường dài nhất và bao gồm các bảng và những thông tin thu được khi phân tích kết quả điều tra.

e) Cảm ơn. Phần này có thể được đặt ở đầu hay ở cuối báo cáo. Nội chung, nó thường được đặt ở phần đầu của bản báo cáo.

Ngoài ra bản báo cáo có thể bao gồm các phần sau:

- Phần phụ lục, bao gồm bộ câu hỏi đã được sử dụng trong cuộc điều tra và những chỉ dẫn cho điều tra viên. Phần này rất quan trọng cho những ai muốn tiến hành những nghiên cứu tương tự ở cộng đồng của họ.

- Danh sách tài liệu tham khảo, bao gồm các báo cáo, tài liệu đã được công bố có liên quan đến cuộc điều tra và những vấn đề của cuộc điều tra. Danh sách này không được quá dài và chỉ bao gồm những tài liệu mà người tác giả của cuộc điều tra đã đọc và nhận thấy rằng nó hay và có ích.



Báo cáo điều tra phải rõ ràng và dễ đọc. Mỗi khía cạnh của các vấn đề thảo luận trong báo cáo phải được trình bày dưới những đề mục nhỏ cùng với các bảng số liệu có liên quan. Việc đánh máy và trình bày bản báo cáo sáng sủa cũng giúp cho bản báo cáo rõ ràng, lôi cuốn và dễ đọc hơn.

6.2 Những bản báo cáo đặc biệt.

Bản báo cáo chính thường mô tả đầy đủ các phương pháp điều tra, quy trình tiến hành điều tra, các kết quả và đề nghị. Tuy nhiên bản báo cáo dạng đầy đủ này không phải lúc nào cũng phù hợp đối với những người khác và những nhà lãnh đạo vì các lý do sau:

- Các nhà lãnh đạo, thường là ở các cấp cao hơn, có các trách nhiệm và nhiệm vụ riêng của họ. Do đó, chỉ có phần kết quả điều tra là quan trọng đối với họ. Họ sẽ không cảm thấy hứng thú khi phải đọc những bản báo cáo không phù hợp với công việc của họ.

- Những người khác, thường là những người có uy tín và ảnh hưởng trong cộng đồng, có thể không đủ kiến thức y học và văn hóa để có thể hiểu được bản báo cáo chi tiết mà có những phần mà họ không quan tâm lắm như quy trình tiến hành điều tra. Để động viên họ ủng hộ cho những đề nghị của bản báo cáo, bản báo cáo phải khuấy động được mối quan tâm của họ và phải được viết sao cho phù hợp với mức độ mà họ có thể có liên quan tới.

Do những lý do trên, bên cạnh bản báo cáo chính phải có 2 bản báo cáo đặc biệt rất ngắn bổ sung.

a) Bản báo cáo cho những nhà lãnh đạo ở cấp cao hơn phải bao gồm:

- Tóm tắt ngắn gọn những lý do tiến hành điều tra và những kết quả chính.
- Một ít thông tin về kết quả của cuộc điều tra và những đề nghị phù hợp với mối quan tâm của họ.
- Một số chỉ dẫn cho những nhà lãnh đạo nơi họ có thể có được bản báo cáo chi tiết nếu họ muốn xem nó.

b) Bản báo cáo cho những người có ảnh hưởng và uy tín trong cộng đồng phải thông báo và động viên sự hỗ trợ của họ đối với những đề nghị của cuộc điều tra đối với cộng đồng của họ. Nếu không có những hỗ trợ mạnh mẽ của họ, cộng đồng sẽ không chấp nhận sự thay đổi theo đề nghị của cuộc điều tra. Để có được sự hỗ trợ của những người lãnh đạo trong cộng đồng, một điều quan trọng là họ phải hiểu được lý do tại sao phải thay đổi và cộng đồng sẽ có ích lợi gì khi thay đổi. Do đó bản báo cáo phải

- Rất rõ ràng và rất ngắn.
- Giải thích một cách dễ hiểu những lý do tại sao lại đưa ra những đề nghị thay đổi đó.
- Giải thích một cách thành thực những ích lợi đối với cộng đồng. Đồng thời nó cũng phải nêu lên những khó khăn khi tiến hành những thay đổi đó.

Cuối cùng, những bản báo cáo này phải phản ánh những suy nghĩ và động cơ của người lập kế hoạch điều tra. Khi cuộc điều tra được tiến hành một cách sáng tạo, nó sẽ giúp cho cộng đồng giải quyết nhiều vấn đề.



SỬ DỤNG THÔNG TIN SẴN CÓ

I. MỤC TIÊU

- 1- Biết cách khai thác và tìm kiếm các thông tin sẵn có
- 2- Đánh giá đúng giá trị thông tin thu thập được.
- 3- Sử dụng những thông tin sẵn có.

II. NỘI DUNG

A. TỔ CHỨC KHAI THÁC, THẢO LUẬN VỀ CÁC VẤN ĐỀ LIÊN QUAN ĐẾN CUỘC ĐIỀU TRA

Lý thuyết điều tra cung cấp những chỉ dẫn lập kế hoạch và tổ chức một cuộc điều tra, xác định cỡ mẫu điều tra, thiết kế bộ câu hỏi; xử lý thông tin và báo cáo kết quả.

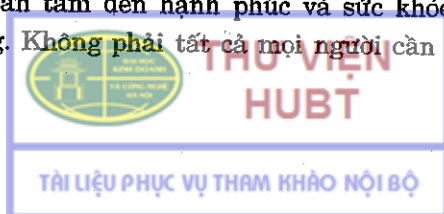
Các cuộc điều tra đòi hỏi phải có thực tế và kinh nghiệm. Do đó các nhà lập kế hoạch và tổ chức điều tra phải gặp gỡ những người đang tiến hành điều tra để thảo luận về phương pháp và so sánh các kết quả. Một điều quan trọng là nên học hỏi từ những người khác tại sao họ lại tiến hành cuộc điều tra như thế này mà không phải là như thế khác.

Khi kết thúc mỗi cuộc điều tra, người tổ chức điều tra thường nghĩ rằng họ có thể cải tiến phương pháp điều tra và tiến hành nghiên cứu theo một cách khác. Nhiều điều có thể học được khi thảo luận với những người đã tiến hành các cuộc điều tra.

Tổ chức một cuộc "hội thảo về điều tra" là một cách tốt nhất để mọi người có cùng một mối quan tâm có thể ngồi lại cùng với nhau thảo luận và học hỏi lẫn nhau. Thời gian cho một cuộc hội thảo như thế có thể kéo dài từ 1 đến 5 ngày. Tại hội thảo, các báo cáo viên sẽ trình bày những việc họ đã làm, những khó khăn mà họ gặp phải và cách giải quyết của họ. Hội thảo này phải cần nhiều thời gian để thảo luận. Hội thảo nên chia ra các phiên thảo luận sáng và chiều, với những chủ đề đặc biệt khác nhau.

B. TÌM KIẾM VÀ ĐÁNH GIÁ CÁC THÔNG TIN SẴN CÓ

Tất cả những ai quan tâm đến hạnh phúc và sức khỏe của cộng đồng đều đòi hỏi thông tin về cộng đồng. Không phải tất cả mọi người cần thông tin giống nhau. Nhưng



nhà quản lý và lập kế hoạch cần những thông tin khác với những thông tin mà các thầy thuốc, y tá và nhân viên y tế xã cần.

Trên thực tế những thông tin cần thiết thường không sẵn có. Những thông tin đó thường chưa được thu thập hay chưa được công bố hoặc quá cũ. Ngoài ra những thông tin sẵn có trong các hồ sơ biểu mẫu thường không phù hợp với vấn đề mà ta cần nghiên cứu. Ngay cả khi những thông tin đó đã có, nó thường phân tán rải rác ở các tạp chí, và các báo cáo mà ta không biết.

Tìm kiếm những thông tin đã có về các vấn đề nghiên cứu không phải là một vấn đề dễ dàng. Khi mà tất cả những thông tin cần thiết hiện chưa có, vẫn có thể tìm được ít nhất một số thông tin cần thiết nhất có liên quan với các vấn đề y tế và xã hội. Cần phải tổ chức những nghiên cứu tích cực chủ động thì mới thu thập được những thông tin này. Phương pháp tốt nhất để thu thập những thông tin này bao gồm thảo luận với đồng nghiệp và bạn bè, nói chuyện với bệnh nhân, phỏng vấn những nhân viên làm ở các cơ quan, các ngành, bộ. Tuy nhiên, những thông tin thu thập được từ những hồ sơ, sổ sách, báo cáo thường không đáp ứng được những thông tin mà ta cần có. Những thông tin đó thường không đầy đủ, không toàn diện, cũ, ở dưới dạng không sử dụng được và không phù hợp.

Tuy nhiên, ngay cả khi những thông tin hiện có không hoàn toàn đáp ứng được mục tiêu nghiên cứu, nó vẫn rất có giá trị và phải được sử dụng như điểm khởi đầu để thu thập thêm những thông tin khác.

Hầu hết các nghiên cứu và các cuộc điều tra có thể được rút ngắn lại hoặc đơn giản đi vì đã có sẵn một số thông tin và không cần thiết phải tiến hành điều tra lại.

Sử dụng một cách khôn ngoan những thông tin cũng quan trọng như việc thu thập nó. Các thông tin về sức khỏe thường có thể thu thập được từ các bài báo, các báo cáo và các bảng thống kê.

Để giải quyết những vấn đề sức khỏe của cộng đồng, những thông tin sẵn có phải được kiểm tra dựa trên 4 tiêu chuẩn đơn giản sau:

- Tính phù hợp: Những thông tin hiện có phải phù hợp với vấn đề mà ta đang nghiên cứu.
- Yếu tố thời gian: Những thông tin phải là những thông tin mới, gần thời điểm mà ta nghiên cứu.
- Tính có thể so sánh được: Những thông tin hiện có liên quan đến cộng đồng, nhóm tuổi, môi trường, và các vấn đề khác phải có thể so sánh được và tương tự như những thông tin về các vấn đề hiện tại của cộng đồng.
- Tính tin cậy: Những thông tin thu thập được theo một phương pháp đúng đắn và đáng tin cậy.

Những tiêu chuẩn trên sẽ giúp các nhà quản lý, các thầy thuốc và những người khác có thể lựa chọn phân loại và chỉ sử dụng những thông tin sẵn có hữu ích vào mục đích nghiên cứu. Sau đó họ sẽ quyết định xem có cần phải thu thập thêm những thông tin nào, và liệu cuộc điều tra có thể cung cấp những thông tin đó hay không.

Các thông tin về sức khỏe của cộng đồng thường được sử dụng theo một trong hai cách sau:

- Mô tả những tính chất và sự kiện cơ bản về cộng đồng và tình hình sức khỏe của nó.

- Chỉ ra rằng đã có sự thay đổi về các điều kiện xã hội và sức khỏe của cộng đồng.

Những thông tin quan trọng thường dùng để mô tả các sự kiện, các vấn đề và những thay đổi đang diễn ra trong cộng đồng. Dưới đây là một ví dụ minh họa thu thập được từ một trung tâm kế hoạch hóa gia đình và chăm sóc trẻ em (KHHGD và CSTE).

Bảng 1- SỐ LIỆU VỀ SỐ TRẺ ĐẸ SỐNG VÀ SỐ TRẺ ĐẸ CHẾT DƯỚI 1 TUỔI Ở MỘT SỐ TRUNG TÂM KHHGD VÀ CSTE

Năm	Số phụ nữ có chồng Tuổi 15-44	Số trẻ đẻ sống	Số trẻ dưới 1 tuổi chết trong năm
1970	980	203	54
1980	1032	191	23

Số liệu trên mô tả tình hình thực tế diễn ra vào các năm 1970 và 1980 trình bày số phụ nữ đến khám tại trung tâm, số trẻ đẻ sống trong năm và số trẻ dưới 1 tuổi chết trong năm. Nó cũng chỉ ra rằng việc chăm sóc trẻ em ở trung tâm này đã làm giảm tỷ suất chết ở trẻ dưới 1 tuổi từ 27% (54/203) xuống còn 12% (23/191). Đồng thời bảng số liệu trên cũng cho thấy là chương trình kế hoạch hóa gia đình chỉ làm giảm nhẹ tỷ suất sinh. Trong năm 1970, tỷ suất trẻ đẻ sống là 21% (203/980), đến năm 1980, tỷ suất này giảm xuống còn 19% (191/1032).

Những thông tin này thường có liên quan chặt chẽ với việc ô nhiễm nước và thực phẩm. Những thông tin như số người trong một gia đình cung cấp nước, bảo quản và vệ sinh thực phẩm thường không có sẵn mà chỉ có thể thu thập được qua điều tra. Thông tin về thái độ và quan điểm của người dân về nhu cầu của họ đóng một vai trò rất quan trọng. Do đó, cần phải thảo luận với mọi người dân trong cộng đồng để phát hiện ra:

- Cộng đồng quan niệm như thế nào về nhu cầu của họ.
- Cộng đồng có những khó khăn gì.
- Cộng đồng mong muốn gì.
- Cộng đồng trông đợi gì ở các cơ quan y tế.
- Ý kiến của cộng đồng làm gì để nâng cao chất lượng các dịch vụ y tế.

Việc áp dụng những tiêu chuẩn đánh giá các thông tin hiện có vào các vấn đề thực tế đòi hỏi phải có kỹ năng và sự xem xét cẩn thận. Người nghiên cứu phải tự đặt ra nhiều câu hỏi về các thông tin hiện có. Những câu hỏi này mang tính đặc thù và thay đổi tùy theo loại vấn đề nghiên cứu loại thông tin hiện có, và cách sử dụng thông tin.



C. SỬ DỤNG THÔNG TIN HIỆN CÓ ĐỂ THU THẬP NHỮNG THÔNG TIN BỔ SUNG

Những tổ chức lớn như các chuyên ngành khác nhau trong y tế thường thu thập và ghi chép nhiều thông tin có liên quan đến các hoạt động của họ. Những tổ chức này có thể cung cấp cho ta các thông tin một cách nhanh chóng theo 3 cách.

- Họ cho phép tiếp cận và sử dụng những số liệu hiện có.
- Họ có thể đồng ý nói với nhân viên của họ xử lý và phân tích thông tin có trong hồ sơ sổ sách và cung cấp cho người cần nó.
- Đôi khi họ có thể được thuyết phục tham gia vào việc thu thập những thông tin bổ sung trong một thời gian ngắn.

Những người lãnh đạo của những tổ chức này phải được thông báo về các lý do cần có những thông tin đó và phải được thuyết phục rằng những thông tin mà họ cung cấp rất có ích lợi cho cộng đồng, và người điều tra sẽ xử lý thông tin một cách có trách nhiệm và tôn trọng tính bí mật của những thông tin được cung cấp.

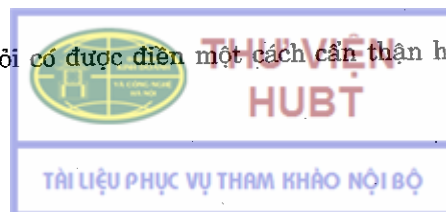
Khi yêu cầu các tổ chức cung cấp thông tin hiện có, một điều cần nhớ là việc thu thập thông tin bổ sung hay xử lý thông tin theo một cách khác với cách hiện đang sử dụng có thể sẽ làm gián đoạn việc ghi chép thông tin thường xuyên và làm tăng khối lượng công việc của các nhân viên của các tổ chức đó. Do đó, những người chịu trách nhiệm ghi chép và thu thập thông tin phải hiểu và được thuyết phục rằng họ cần phải thay đổi những phương pháp thông thường của họ và cần phải thu thập nhiều thông tin hơn. Nếu chỉ có sự đồng ý của người lãnh đạo mà không có sự đồng ý tham gia của các nhân viên thì chưa đủ. Những nhân viên ghi chép số liệu thường luôn có nhiều công việc khác và họ có ít thời gian để tham gia vào các công việc tính toán và văn phòng khác. Họ cần phải được động viên, khuyến khích tham gia vào công việc này.

Cần phải tổ chức việc huấn luyện và hướng dẫn các nhân viên về cách thu thập thông tin mới hoặc sử dụng những biểu mẫu mới để điền các số liệu vào. Những công việc này phải được giám sát một cách cẩn thận. Những thông tin hiện có thường không đáp ứng được về mặt chất lượng bởi vì những nhân viên làm công việc thu thập và ghi chép số liệu không được đào tạo đầy đủ, không được theo dõi giám sát và không được động viên, khuyến khích. Người nghiên cứu không nên chỉ đưa ra một số chỉ dẫn rồi vài ngày sau quay lại để lấy số liệu mà họ phải tham gia tích cực và liên tục trong các giai đoạn của quá trình thu thập và ghi chép thông tin.

Trước khi sử dụng những thông tin hiện có, cần phải kiểm tra chất lượng của nó. Việc sử dụng những thông tin đã được công bố phải dựa trên những mô tả về cách thu thập số liệu và các bước đã làm để bảo đảm tính chính xác và tính đại diện của số liệu. Đồng thời người nghiên cứu cũng phải kiểm tra tính tin cậy của những số liệu đó.

Người nghiên cứu cần phải xem xét và trả lời các câu hỏi sau:

- Những thông tin thu thập được có đại diện cho quần thể hay một bệnh mà ta nghiên cứu hay không?
- Những biểu mẫu và bộ câu hỏi có được điền một cách cẩn thận hay không? Tất cả



đã được điền vào biểu mẫu hay một số số liệu đã bị mất hay bỏ đi?

c) Những bộ câu hỏi đã hoàn thành có được kiểm tra và bảo đảm chắc chắn không bị bỏ sót một câu hỏi nào và những câu trả lời là phù hợp với nhau hay không?

d) Có sự khác nhau nào giữa các biểu mẫu mà do những người khác nhau điền vào hay không?

e) Những người điều tra có được huấn luyện và chỉ dẫn trước khi thu thập và ghi chép thông tin hay không? Những thuật ngữ dùng trong biểu mẫu điều tra có được giải thích cẩn thận hay không?

Số liệu được gọi là "số liệu mềm" (Soft data) khi nó không hoàn toàn đại diện hoặc những thuật ngữ dùng không được định nghĩa rõ ràng, hoặc những biến số hay đặc trưng mà chúng ta quan tâm không được đo lường một cách chính xác. Ngược lại, những thông tin chắc chắn có chất lượng cao được gọi là "số liệu cứng" (hard data).

"Số liệu mềm" vẫn có ích lợi khi không có thông tin hiện có nào tốt hơn. Cần phải đặc biệt chú ý khi sử dụng "số liệu mềm". Nó có thể dễ đưa đến những nhận định sai. Do đó nó chỉ được dùng để cung cấp các chỉ dẫn và gợi ý không được dùng làm cơ sở để đưa ra các quyết định.

D. NHỮNG HỖ TRỢ ĐƠN GIẢN VỀ MẶT SỐ HỌC ĐỂ HIỂU NHỮNG THÔNG TIN HIỆN CÓ

Những số liệu về dân số học và sức khỏe được công bố thường được trình bày dưới dạng tần số, có nghĩa là tổng số các trường hợp được chia thành nhiều lớp hay nhiều nhóm nhỏ. Những chỉ số đơn giản sau có thể tính được từ những tổng số ấy.

- Số trung bình
- Tỷ lệ phần trăm
- Các tỷ suất
- Các tỷ lệ chết và tỷ lệ mắc, sử dụng các số liệu sẵn có về nhân khẩu học.

1. Các số trung bình

a) *Tính trung bình trong nhiều năm.*

Số trung bình cung cấp những hình ảnh đáng tin cậy và ổn định hơn so với tổng số của từng năm. Tổng số của từng năm có thể dao động nhiều đặc biệt khi tần số hàng năm là nhỏ. Điều này xảy ra đối với những bệnh hiếm hay có sự xuất hiện bất thường, ví dụ như ung thư gan hay tự tử. Không nên tính số trung bình trong thời kỳ dài hơn 5 năm. Nếu trong một thời kỳ quá dài, chiều hướng tăng giảm rõ rệt và những thay đổi khác có thể xảy ra và do đó quá trình tính toán số trung bình có thể làm lu mờ chiều hướng và những thay đổi đó đi. Đối với những bệnh phổ biến xảy ra ở những cộng đồng tách biệt hay ở những vùng ít dân cư sinh sống, tổng số hàng năm có thể nhỏ. Tính trung bình trong một thời kỳ 3 - 5 năm cũng hoàn toàn phù hợp trong những trường hợp này.

b) *Tính trung bình theo đơn vị hành chính.*

Số liệu các đơn vị hành chính hay những tổ chức nhỏ như bệnh viện, trạm y tế, trường



học, làng xã thường khác nhau rất nhiều. Do đó không nên tính toán số liệu của từng đơn vị hành chính riêng biệt mà tính trung bình của các đơn vị đó.

c) Tỷ lệ phần trăm.

Việc tính toán tỷ lệ phần trăm cũng rất có ích trong việc so sánh giữa các vùng, các cơ quan khác nhau, và trong các thời gian khác nhau. Điều này sẽ được minh họa bằng số liệu về tỷ lệ tử vong của trẻ em dưới 1 tuổi do viêm phổi, được trình bày ở bảng 2.

Bảng này cho thấy có sự giảm rõ rệt tỷ lệ phần trăm chết do viêm phổi trong từng thập kỷ.

Bảng 2 - TỶ LỆ CHẾT Ở TRẺ DƯỚI 1 TUỔI Ở SCOTLEN

	1950	1960	1970	1980
Số chết do viêm phổi	418	242	195	19
Số chết do tất cả các nguyên nhân	1438	825	599	296
Tỷ lệ phần trăm chết do viêm phổi	29,1%	29,3%	32,6%	6,4%

Hai hàng đầu của bảng này cho thấy rằng số trẻ chết dưới 1 tuổi giảm đi trong thời kỳ 1950 đến 1980, đối với chết do viêm phổi cũng như đối với chết do tất cả các nguyên nhân. Tuy nhiên hàng thứ ba lại cho thấy rằng tỷ lệ phần trăm chết do viêm phổi không thay đổi trong thời kỳ từ 1950 đến 1970, vào khoảng 30%. Năm 1980, mười năm sau, tỷ lệ phần trăm chết do viêm phổi giảm đi một cách đáng kể xuống còn 6,4%, có thể là do sử dụng kháng sinh đường hô hấp mới và nâng cao chất lượng của việc điều trị và chăm sóc tích cực.

Không nên tính tỷ lệ phần trăm đối với các mẫu số nhỏ hơn 30.

Một số nhà thống kê còn cho rằng mẫu số phải lớn hơn 50.

d) Tỷ suất (Ratio).

Tỷ suất là một thương số, là kết quả của việc chia một số này cho một số khác. Tỷ suất cho ta biết mối quan hệ lẫn nhau giữa hai tần số hay hai số lượng. Ví dụ, chiều hướng tăng rõ rệt số người già thu được qua phân tích số liệu tổng điều tra dân số và qua việc tính toán các tỷ suất.

Bảng 3- SỐ LIỆU ĐIỀU TRA DÂN SỐ Ở CÁC NHÓM TUỔI Ở SCOTLEN, 1931-1981.

Dân số	1931	1951	1961	1971	1981
Tuổi ≥ 65	353244	506895	548908	644452	710794
Tuổi 15-64	3184702	3333419	3219338	3229709	3246592
Tỷ suất	0,11	0,15	0,17	0,20	0,22
Trên 65/15-64					

(Tổng điều tra dân số năm 1941 không tiến hành được vì có chiến tranh)



THƯ VIỆN
HUBT

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

Số tuyệt đối những người trên 65 tuổi cho phép ta ước lượng được nhu cầu về các dịch vụ y tế và xã hội của những người cao tuổi. Để đáp ứng nhu cầu của họ, cần phải thành lập các bệnh viện cho người cao tuổi, nhà dưỡng lão, v.v... Điều này đặc biệt cần thiết đối với các nước công nghiệp phát triển là nơi mà thiếu sự giúp đỡ của những người khác trong gia đình đối với những người già cả.

Tuy nhiên, mọi sự giúp đỡ đối với người già vẫn trông cậy vào nhóm tuổi 15-64, là nhóm lao động tích cực sản xuất ra của cải vật chất của xã hội. Do tầm quan trọng về mặt kinh tế và xã hội, cần phải xác định xem những gánh nặng của việc hỗ trợ người già đè lên vai những người trẻ tuổi hơn 15-64 tuổi, có thay đổi không và thay đổi như thế nào. Một chỉ số đơn giản để xác định sự thay đổi này là tính toán tỷ suất số người trên 65 tuổi/số người 15-64 tuổi. Bảng trên đã cho thấy ở Scotlen tỷ suất này đã tăng gấp đôi trong thời kỳ từ 1931 đến 1981.

Chúng ta có thể sử dụng một số nguyên tắc là.

a) Sử dụng tổng số tuyệt đối để trả lời các câu hỏi liên quan tới yêu cầu đòi hỏi các dịch vụ y tế hay các câu hỏi liên quan tới các nguồn lực hiện có.

b) Sử dụng tỷ suất để xác định những thay đổi tương đối của các yếu tố ảnh hưởng đến cung cấp và nhu cầu.

c) Các tỷ lệ (rates).

Tỷ lệ đóng một vai trò rất quan trọng trong việc nghiên cứu các vấn đề sức khỏe và vạch ra các đường lối chính sách về y tế.

Tỷ lệ là một dạng đặc biệt của tỷ suất trong đó số trường hợp xảy ra trong một năm chia cho số người phơi nhiễm trong năm đó. Ví dụ, tỷ lệ sinh được định nghĩa là số trẻ đẻ sống trong năm chia cho số phụ nữ có chồng dưới 45 tuổi trong năm đó. Để tính toán tỷ lệ sinh đòi hỏi phải có những thông tin sau:

- Số trẻ đẻ sống trong một thời kỳ nhất định (thường là một năm). Thông tin này thường không sẵn có vì nó đòi hỏi một hệ thống đăng ký và ghi chép số liệu hoạt động tốt. Việc định nghĩa "đẻ sống" về lâm sàng cũng gặp nhiều khó khăn. Ngoài ra ở một số nước đang phát triển, những trẻ sơ sinh chết sau vài ngày hay vài tuần sau khi đẻ ra, có thể được báo cáo là chết sơ sinh hay không báo cáo.

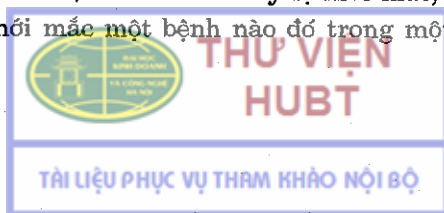
- Số phụ nữ có chồng ở lứa tuổi 15 - 45 (hay 50). Người ta thường không biết số này. Số này cũng thay đổi trong năm do có thêm số phụ nữ mới cưới chồng và số phụ nữ chết đi. Do đó, người ta thường tính số trung bình phụ nữ ở tuổi 15 - 45 (hay 50) ở khoảng giữa năm. Do có những khó khăn trên, một số người khi tính tỷ lệ sinh, họ tính gộp luôn cả số trẻ đẻ hoang và không phân biệt phụ nữ có chồng và phụ nữ chưa có chồng.

Khi số người phơi nhiễm không biết, người ta chỉ có thể tính được tỷ lệ gần đúng sử dụng số liệu thu thập được từ các cuộc điều tra hay từ các người khác.

d) Tỷ lệ hiện mắc và tỷ lệ mới mắc (Prevalence và incidence)

Tỷ lệ liên quan đến các bệnh được chia thành tỷ lệ hiện mắc và tỷ lệ mới mắc.

- Tỷ lệ mới mắc (incidence) để tính toán tỷ lệ mới mắc, trước hết chúng ta phải đếm tất cả các trường hợp mới mắc một bệnh nào đó trong một khoảng thời gian nhất định



thường là 12 tháng. Cần phải loại trừ các trường hợp bệnh đã được chẩn đoán trước thời điểm bắt đầu nghiên cứu cho dù là đã được điều trị khỏi hay chưa. Sau đó, tổng số mới mắc này sẽ được chia cho tổng số người phơi nhiễm đối với yếu tố nguy cơ mắc bệnh này trong cùng thời gian đó. Người ta cũng có thể tính tỷ lệ mới mắc theo giới và theo nhóm tuổi và người ta gọi nó là "tỷ lệ mới mắc riêng phần theo tuổi và giới".

- Tỷ lệ hiện mắc (Prevalence): Để tính tỷ lệ hiện mắc, trước hết chúng ta đếm số các trường hợp bệnh đã được chẩn đoán và phát hiện hoặc điều trị trong một khoảng thời gian, thường là 12 tháng. Do đó, tỷ lệ hiện mắc bao gồm cả những bệnh nhân đã được chẩn đoán trước khi bắt đầu thời điểm nghiên cứu nhưng trong thời gian nghiên cứu họ vẫn bị bệnh. Sau đó, tổng số này sẽ được chia cho tổng số người phơi nhiễm với yếu tố nguy cơ mắc bệnh trong cùng thời gian đó. Tỷ lệ hiện mắc cũng được tính theo các nhóm tuổi và giới.

Đôi lúc, tỷ lệ hiện mắc được tính như là "tỷ lệ hiện mắc kỳ" vì nó đo lường tỷ lệ hiện mắc trong một khoảng thời gian, thường là 12 tháng. Khi thời kỳ nghiên cứu quá ngắn, ví dụ trong một ngày, tỷ lệ hiện mắc được gọi là "tỷ lệ hiện mắc điểm".

E. TÌM KIẾM CÁC THÔNG TIN HIỆN CÓ

Không có quy tắc hoàn toàn đúng nào để có thể giúp ta tìm kiếm các thông tin hiện có, mà chỉ có một số chỉ dẫn chung. Người tìm kiếm thông tin phải tự đặt ra hai câu hỏi đơn giản sau:

1- Những thông tin nào là cần thiết nhất.

2- Ai đang quan tâm đến những thông tin tương tự hoặc ai đang làm việc trong lĩnh vực có liên quan đến thông tin này?

Ví dụ: nếu người nghiên cứu muốn xem xét các vấn đề điều kiện nhà ở; những chủ đề quan trọng sau, cần phải được nghiên cứu và thu thập.

- Số dân và cấu trúc tuổi của quần thể.

- Loại nhà và số lượng nhà hiện có.

- Khoảng cách từ nhà đến chợ, trung tâm thành phố, trường học và trạm y tế, vv...

- Cung cấp nước.

- Vệ sinh nước thải và rác.

- Phương tiện giao thông, điều kiện đường xá, vv...

Dựa vào bảng liệt kê những thông tin này, chúng ta có thể biết những ai cũng quan tâm và liên quan đến những thông tin này.

Ví dụ, đối với các chủ đề trên, ta có thể liệt kê những nguồn thông tin như sau:

a) Số dân và cấu trúc tuổi của quần thể

- Các báo cáo về điều tra dân số và những vụ, cục chịu trách nhiệm thu thập và xử lý các thông tin về điều tra dân số.

- Phòng đăng ký sinh, tử, kết hôn và di dân.



- Sở giáo dục, bộ giáo dục nếu cần có các thông tin về thanh thiếu niên.

b) Loại nhà và số lượng nhà hiện có.

Các thông tin này có thể thu thập từ các nhà lãnh đạo chính quyền của huyện, thị xã hay thành phố. Các thành phố lớn thường có những cán bộ chuyên trách thu thập và lưu trữ các thông tin về nhà ở và điều kiện nhà ở. Ở một số nước, việc sắp xếp vị trí nhà ở thuộc về trách nhiệm của các cơ quan nhà nước. Ở một số nước khác, các công ty thương mại và công nghiệp đóng vai trò quan trọng trong việc cung cấp nhà ở. Việc ban hành chính sách sắp xếp vị trí nhà ở đòi hỏi phải có nhiều thông tin có liên quan.

c) Khoảng cách từ nhà ở đến chợ, trường học, v.v...

Các nhà lãnh đạo của địa phương thường có các thông tin về chủ đề này. Nó thường ở dưới dạng bản đồ. Ngoài ra họ cũng cung cấp cho người nghiên cứu những thông tin bổ sung có thể điền thêm trên bản đồ. Sau đó, khi đi khảo sát và điều tra thực địa, người nghiên cứu có thể so sánh những thông tin mà các nhà lãnh đạo cung cấp và những thông tin mà họ phát hiện ra và sẽ biết được thực trạng của tình hình hiện tại.

Ở những nơi mà các điều kiện môi trường và xã hội thay đổi nhanh, những điều kiện và các vấn đề sức khỏe của cộng đồng đôi lúc sẽ trở nên phức tạp và nó không được phản ánh đúng trong các thông tin hiện có từ các nguồn chính thức.

d) Cung cấp nước và vệ sinh môi trường.

Nhiều nhà lãnh đạo chính quyền quan tâm đến vấn đề cung cấp nước và vệ sinh môi trường. Thông tin về chủ đề này thường có thể thu thập được từ các nguồn:

- Công ty môi trường.
- Công ty cấp thoát nước.
- Trung tâm y tế huyện, thành phố.

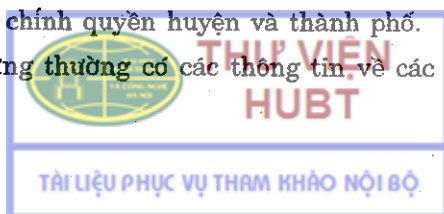
Đôi khi, một số công ty sẵn sàng giúp đỡ cung cấp thông tin nhưng hầu hết họ không muốn cung cấp thông tin đó vì sợ ảnh hưởng xấu đến lợi ích nghề nghiệp của họ. Một số nước có các bộ, vụ, ngành chịu trách nhiệm về các vấn đề thủy lợi, đê đập, các nơi dự trữ nước và các nguồn nước. Những cơ quan này có thể có những thông tin bổ ích về cung cấp nước, sự thay đổi cung cấp nước theo thời gian trong năm và theo các nhu cầu trong công nghiệp và tiêu dùng. Tuy nhiên, những thông tin này thường liên quan đến những vùng rộng lớn và không thích hợp cho các nghiên cứu sâu về các điều kiện ở mức cộng đồng.

e) Đường sá và giao thông.

Đường sá, giao thông, trường học, chợ, cửa hàng không nhất thiết lúc nào cũng nằm trong nghiên cứu về điều kiện nhà ở. Tuy nhiên con người sẽ sống và làm việc trong những điều kiện khó khăn hơn nếu đường sá giao thông, chợ, cửa hàng, trung tâm y tế, trường học, v.v... là không phù hợp và thiếu. Đường sá và giao thông có ảnh hưởng rất lớn đến các khả năng làm nâng cao hạnh phúc và phúc lợi của cộng đồng. Thông tin về đường sá và giao thông có thể được thu thập từ các thông tin sẵn có từ ba nguồn sau:

- Các nhà lãnh đạo chính quyền huyện và thành phố.

Sở kỹ thuật xây dựng thường có các thông tin về các điều kiện đường sá, kế hoạch



xây dựng đường mới, việc tắc nghẽn giao thông trong khu vực. Thông tin về các dịch vụ giao thông cũng có thể thu thập được từ các nhà lãnh đạo huyện, thành phố. Thông tin về thời gian chạy, số lượng xe ô tô khách, số lái xe và số người sử dụng dịch vụ này (hay ít nhất là số vé bán ra) cũng có thể do họ cung cấp. Nếu giao tiếp với họ một cách khéo léo, người nghiên cứu có thể được cung cấp được những thông tin này một cách dễ dàng.

- Các cơ quan của chính phủ hay của khu vực.

Hầu hết các quốc gia đều có các báo cáo về đường bộ, đường sắt, giao thông đường thủy và các dịch vụ vận tải khác. Một số thông tin có thể được thu thập từ các báo cáo chính thức. Một số thông tin bổ sung khác sẽ được cung cấp nếu người nghiên cứu đề nghị. Những thông tin cơ bản này có thể được so sánh với thực trạng đường xá và giao thông ở địa phương.

- Các công ty điều hành các dịch vụ giao thông vận tải.

Ở một số nước, việc vận chuyển bằng xe tải, xe khách do một số công ty đảm nhiệm mà chính phủ có thể sẽ hỗ trợ một phần kinh phí. Những công ty này nếu hoạt động có hiệu quả sẽ có nhiều thông tin và một số công ty sẽ sẵn sàng cung cấp thông tin cho người nghiên cứu.

F. SỬ DỤNG NHỮNG THÔNG TIN ĐÃ THU THẬP

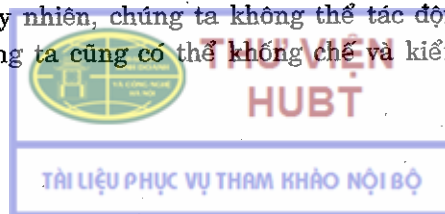
Việc thu thập thông tin là rất tốn kém ngay cả khi chính phủ và các cơ quan tham gia vào việc đó. Do đó sử dụng đúng đắn những thông tin này là rất quan trọng. Chúng ta đều biết rằng việc lập kế hoạch và đưa ra các quyết định dựa trên những thông tin sai là rất phiêu lưu và sẽ dẫn đến những kế hoạch không thể thực hiện được và những quyết định sai.

Đôi lúc, việc sử dụng những thông tin này là cực kỳ đơn giản và đôi lúc lại không thể sử dụng được chúng. Nếu sử dụng thông tin có hiệu quả, chúng ta có thể:

- a) Làm ảnh hưởng đến tình trạng sức khỏe và các điều kiện môi trường và xã hội.
- b) Làm cho mọi người sẵn sàng chấp nhận những thay đổi được đề xuất ở mọi cấp như các nhà lãnh đạo, người lập kế hoạch, cộng đồng và từng cá nhân trong cộng đồng.
- c) Có được các nguồn lực cần thiết để thay đổi và cải thiện các điều kiện của cộng đồng.

Ví dụ, mọi người đều biết rằng tăng huyết áp làm tăng nguy cơ bị nhồi máu cơ tim và các bệnh mạch vành tim. Ngày nay với những thuốc điều trị mới, người ta có thể điều trị tăng huyết áp một cách có hiệu quả. Tuy nhiên, việc phát hiện tất cả những trường hợp tăng huyết áp không phải là dễ. Một số người gợi ý là tiến hành khám đo huyết áp sàng lọc nhưng đòi hỏi phải có một chương trình sàng lọc rất tốn kém, với các cán bộ điều tra được đào tạo tốt và với sự sẵn sàng cộng tác hưởng ứng chương trình của bệnh nhân.

Đối với một số bệnh khác, rất khó áp dụng những chỉ dẫn nêu trên. Ví dụ, nguy cơ mắc ung thư tăng lên theo tuổi. Tuy nhiên, chúng ta không thể tác động được vào yếu tố tuổi. Không phải yếu tố nào chúng ta cũng có thể không chế và kiểm soát được. Do



đó chúng ta chỉ có thể tác động một cách gián tiếp vào các yếu tố đó. Ví dụ, ở một nhóm tuổi nào đó, ta tiến hành chẩn đoán thường xuyên và kỹ càng hơn để phát hiện sớm những triệu chứng ban đầu của ung thư.

Một ví dụ khác nữa là mọi người đều biết về ảnh hưởng có hại của việc hút thuốc lá. Ảnh hưởng lớn nhất của hút thuốc lá làm tăng nguy cơ ung thư phổi và viêm phế quản. Ngoài ra, hút thuốc lá còn gây nhiều ảnh hưởng khác nữa mà mọi người cũng đều biết rõ. Nhưng tại sao những thông tin này lại không được sử dụng để ngăn ngừa việc hút thuốc lá như cấm trồng thuốc lá, hay đưa ra những biện pháp khác. Có nhiều lý do để giải thích điều này. Ở mức độ cá nhân, anh ta không muốn bỏ hút thuốc lá. Ở mức độ cộng đồng, có nhiều ý kiến phản đối. Chủ của hàng sợ thua lỗ trong buôn bán, công nhân trong các nhà máy sản xuất thuốc lá sợ mất việc làm. Ngoài ra đối với một số nước, việc đánh thuế vào việc nhập khẩu thuốc lá đóng một vai trò quan trọng làm tăng ngân sách thu nhập, thì họ lại càng phản đối việc cấm hút thuốc lá. Để giải quyết vấn đề này, lại càng phải đòi hỏi sử dụng thông tin bỏ hút thuốc, chúng ta cần phải tìm hiểu về hiệu quả của các chương trình giáo dục sức khỏe chống hút thuốc lá. Chúng ta cũng phải cần biết các thông tin về việc liệu tăng giá thuốc lá có làm giảm đáng kể số người hút thuốc lá hay không, về ảnh hưởng của việc cấm hút thuốc lá đối với công ăn việc làm của công nhân và đối với người thu nhập qua đánh thuế nhập khẩu thuốc lá. Việc cấm hút thuốc lá không phải lúc nào cũng có hiệu quả vì nó có thể khuyến khích việc buôn lậu thuốc lá và những phương pháp đối phó của những người nghiện hút thuốc lá.

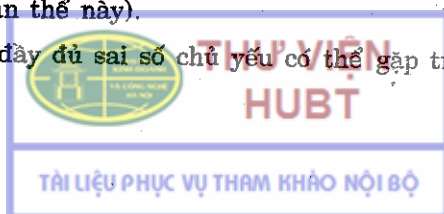
Để sử dụng thông tin một cách có hiệu quả, người nghiên cứu phải trả lời các câu hỏi sau:

1. Anh (chị) biết gì về vấn đề y tế cần phải giải quyết.
2. Làm thế nào để có thể áp dụng thông tin này một cách tốt nhất.
3. Cần phải có những thông tin bổ sung nào để giải quyết vấn đề y tế này?
4. Thông tin bổ sung này hiện đã có sẵn ở một nơi nào đó hay không?
 - a) Nếu có, làm thế nào để có được thông tin này?
 - b) Nếu không, liệu một cuộc điều tra đặc biệt để thu thập thông tin này có thể được tiến hành trong thời gian thích hợp và với một kinh phí hợp lý hay không?

CHỌN MẪU

I. MỤC TIÊU

1. Xác định được đúng quần thể định danh cho nghiên cứu (đặc biệt là xác định được tính đồng nhất của quần thể này).
2. Hình dung được đầy đủ sai số chủ yếu có thể gặp trong thể loại nghiên cứu của



đánh giá được nguy hại của những sai số đó, và biết cách để hạn chế chúng.

3. Phân biệt được độ lệch chuẩn và sai số chuẩn - ý nghĩa của chúng.

4. Biết dùng sai số chuẩn tính các giới hạn tin cậy 95% (đối với cả trung bình lẫn tỷ lệ) và biết phiên giải về các giới hạn tin cậy của khoảng tin cậy.

II. NỘI DUNG

Phần lớn các nghiên cứu dịch tễ học đều tiến hành trong những quần thể rộng lớn, ta không thể (và không cần) nghiên cứu toàn bộ quần thể đó, mà sẽ nghiên cứu một mẫu nào đó, với một điều kiện khác nghiệt là mẫu phải đại diện quần thể.

Mẫu được kể là đại diện quần thể được đánh giá trên 2 điểm chủ yếu: Chọn mẫu như thế nào và cỡ mẫu là bao nhiêu. Bài này đề cập về vấn đề chọn mẫu.

A. NHỮNG KIẾN THỨC CƠ BẢN CẦN CÓ PHỤC VỤ CHỌN MẪU

Ngoài những kiến thức quen thuộc về các chỉ số tập trung và các chỉ số phân tán, cần phải hiểu đúng và nắm vững những khái niệm sau:

1. Quần thể từ đó đã rút ra mẫu nghiên cứu: quần thể định danh.

Quần thể là một tập hợp các đơn vị thống kê, mà trên mỗi đơn vị thống kê này, một giá trị về tiêu thức nghiên cứu sẽ được lấy ra và được kể là một số hạng trong một chuỗi thống kê.

Đơn vị thống kê trong các nghiên cứu dịch tễ thường là 1 người (một cá thể). Và quần thể từ đó muốn rút mẫu nghiên cứu ra là quần thể định danh, nghĩa là một quần thể được gọi tên ra một cách cụ thể, thí dụ: quần thể phụ nữ 15-49 tuổi có 3 con, nữ y tá phòng mổ, nam công nhân nhà máy cao su tuổi nghề từ 15 năm trở lên... Trong những quần thể như thế, ta chỉ lấy một số cá thể vào mẫu nghiên cứu, rồi từ kết quả nghiên cứu được, ta sẽ quy ra cho toàn bộ quần thể này.

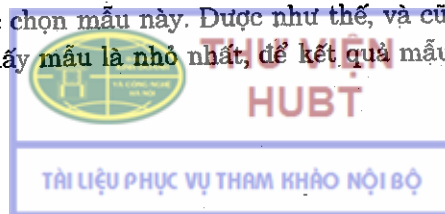
2. Những sai số có thể gặp phải trong nghiên cứu mẫu và cách khắc phục.

Có rất nhiều. Tùy loại hình mẫu nghiên cứu mà loại sai số nào là nghiêm trọng nhiều hay ít, có thể tránh được, hay hạn chế được đến đâu... Những loại sai số quan trọng, chung nhất, thường gặp phổ biến là: (xin xem thêm ở sách đã dẫn)

a) Sai số lấy mẫu.

Sai số lấy mẫu là sai số sẽ gây ra sự khác biệt giữa kết quả mẫu và kết quả thực của quần thể. Trên thực tế, kết quả thực của quần thể ta không được biết (vì không nghiên cứu quần thể toàn bộ), ta thường gọi nó là kết quả lý thuyết (trung bình lý thuyết, hoặc tỷ lệ lý thuyết), nên ta cũng không biết được kết quả mẫu có khác biệt không, và nếu có thì khác biệt đến đâu, khác biệt như thế nào đối với kết quả thực của quần thể.

Ta sẽ phải dựa vào các thủ tục chọn mẫu thích hợp, một cách chặt chẽ, tuân thủ nghiêm ngặt và đầy đủ các thủ tục chọn mẫu này. Được như thế, và cũng chỉ có như thế, chúng ta mới làm được cho sai số lấy mẫu là nhỏ nhất, để kết quả mẫu gần kết quả thực



của quần thể.

Sai số lấy mẫu bao gồm ít nhất 2 loại sai số phổ biến sau: Sai số chọn.

Sai số do các biến thiên ngẫu nhiên.

Sai số chọn là sai số nảy sinh khi chúng ta chọn không đúng các đối tượng vào trong mẫu, thí dụ chúng ta đã chọn phải bộ phận không đại diện của quần thể, như khi ta nghiên cứu tỷ lệ một bệnh nào đó trong quần thể học sinh, lại không bao gồm số học sinh nghỉ học (có nhiều khả năng nghỉ học chính ta định tìm) mà chỉ thăm khám trong số học sinh có đến trường chẳng hạn. Sai số chọn không thể tính được, chỉ có bản thân người nghiên cứu mới biết được cần làm như thế nào để tránh được.

Còn sai số sẽ do các biến thiên ngẫu nhiên nảy sinh phụ thuộc vào may rủi từ mẫu này sang mẫu khác, dù ta tuân thủ triệt để các thủ tục chọn mẫu thì cũng thật khó hy vọng lặp lại những mẫu như nhau khi lấy mẫu nhiều lần. Người ta có thể hạn chế được sai số này bằng hai cách: Quần thể định danh từ đó rút mẫu ra phải đồng nhất (định nghĩa về quần thể định danh phải chặt chẽ, rõ ràng) và tính ngẫu nhiên trong quá trình chọn các cá thể vào mẫu phải đảm bảo đúng đắn, đầy đủ từ đầu đến cuối ở tất cả các bước cần phải áp dụng kỹ thuật ngẫu nhiên.

b) Sai số có tính chất hệ thống.

Cũng có nhiều loại: sai số chẩn đoán, sai số từ chối, sai số nhỏ lại, sai số bỏ cuộc, sai số thu thập thông tin, sai số xếp lần... các loại sai số này quan trọng nhiều hay ít tùy loại hình nghiên cứu, và có những cách khác phục riêng biệt nhau (xin xem thêm sách đã dẫn).

c) Sai số do các yếu tố nhiễu gây nên.

Trong nghiên cứu mẫu, các yếu tố gây nhiễu đóng vai trò rất quan trọng. Có những yếu tố nhiễu rõ ràng có thể tránh được chung trong nhiều nghiên cứu, những cũng còn có những yếu tố nhiễu tiềm ẩn khác, mà chỉ chính nhà nghiên cứu về vấn đề nào đó mới có thể thấy được nhờ vào những hiểu biết về cơ chế bệnh sinh, về các yếu tố nguy cơ có thể có tác dụng với bệnh, về các hiểu biết mới phát minh của y học... mới có thể loại bỏ được hoặc hạn chế được các sai số này, cùng với việc áp dụng các kỹ thuật ghép cặp, các kỹ thuật chuẩn hóa...

3. Tính ngẫu nhiên trong việc chọn mẫu.

Tính ngẫu nhiên ở đây được toán học quy định, những nhà y học chúng ta hiểu theo cách gần đúng là không hề có một ý kiến chủ quan nào của bất cứ một ai, ở bất kỳ bước nào của việc chọn mẫu.

Một mẫu được gọi là mẫu ngẫu nhiên khi các cá thể trong quần thể định danh được chọn vào mẫu với cùng một cơ hội may rủi như nhau, nói theo ngôn ngữ thống kê là mọi người đều có cùng một xác suất phản ánh vào trong mẫu.

Mẫu được lấy ngẫu nhiên có hai ưu điểm:

- Giảm tối thiểu được sai số.
- Xác định được tính chính xác của kết quả nghiên cứu mẫu.

Nếu mẫu được lấy ngẫu nhiên thực sự, thì đáng ngại chỉ còn là có sai số ngẫu nhiên



(mà chúng ta có thể làm giảm thiểu bằng xác định tính đồng nhất của quần thể định danh như trên đã nói, và mở rộng cỡ mẫu hợp lý).

4. Sai số chuẩn

Cần phải phân biệt độ lệch chuẩn với sai số chuẩn. Độ lệch chuẩn là số đo độ phân tán của các trị số của tiêu thức nghiên cứu nghĩa là các trị số thu được ở từng đơn vị nghiên cứu (đơn vị thống kê) trong dãy thống kê so với số trung bình của tiêu thức của dãy đó. Thí dụ: lượng cholesterol của từng người so với lượng cholesterol trung bình của quần thể đó. Độ lệch chuẩn, như vậy, để xác định khoảng dao động của các số đo đó, để cho ta khái niệm về dãy thống kê đó phân tán nhiều hay ít.

Còn sai số chuẩn thuộc phạm trù quan niệm khác, sai số chuẩn nói lên tính biến thiên của các kết quả thống kê của mẫu (tính biến thiên của số trung bình mẫu hoặc tính biến thiên của tỷ lệ mẫu). Cần phải hiểu tính biến thiên này là tính biến thiên của kết quả nghiên cứu từ một mẫu này sang một mẫu khác trong cùng một quần thể nghiên cứu (khi ta làm nhiều mẫu cùng cỡ lặp lại trong một quần thể, thì các kết quả mẫu đó sẽ khác nhau: chúng ta nói là kết quả mẫu có tính biến thiên). Sai số chuẩn càng nhỏ thì tính biến thiên càng nhỏ, và kết quả nghiên cứu càng nhiều tin cậy. Sai số chuẩn dùng trước hết là để xây dựng các khoảng tin cậy.

5. Giới hạn tin cậy.

Khoảng tin cậy dùng để tổng quát hóa thành kết quả quần thể từ kết quả nghiên cứu mẫu. Khoảng tin cậy có hai giới hạn: cận trên và cận dưới. Khoảng tin cậy như vậy là một khoảng số nào đó có giới hạn rõ ràng, để cho giá trị thực của quần thể (là một con số có thực, nhất định có, nhưng ta không có được vì không điều tra toàn bộ quần thể), nhất định là rơi vào trong khoảng đó, ở một điểm nào đó. Khi nếu khoảng tin cậy bao giờ cũng phải nêu rõ khoảng tin cậy bao nhiêu phần trăm, điều này gắn liền với xác suất tin cậy. Thí dụ khi ta nêu 95% khoảng tin cậy, thì là chúng ta đã nói rằng có tới 95% khả năng là khoảng đó có chứa đựng giá trị thực của quần thể. Đối với những cỡ mẫu đủ lớn hợp lý thì khoảng tin cậy 95% được biểu thị như sau:

95% CI = Kết quả mẫu \pm 2 sai số chuẩn.

Confidence Interval CI $CI = \bar{X} \pm 1.96 SE$

III. BÀI TẬP

Bài tập 1.

Muốn mô tả tỷ lệ mổ lấy thai, người ta đã lấy một mẫu ngẫu nhiên hồi cứu trong số bệnh án sản khoa của hai bệnh viện lớn trong một thành phố thấy tỷ lệ mẫu là 20%. Trong khi đó, thì một nghiên cứu lấy mẫu ngẫu nhiên hồi cứu từ các bệnh án của tất cả các bệnh viện sản khoa trong thành phố thì lại thấy tỷ lệ đó chỉ từ 10% đến 12%.

Câu hỏi.

1. Nên chọn quần thể định danh của nghiên cứu trên là quần thể nào?
2. Nghiên cứu trên đã chọn ngẫu nhiên bệnh án sản khoa, tại sao lại có kết quả sai lệch như vậy.



Bài tập 2.

Người ta đã ghi lại thời gian từ lúc xuất hiện triệu chứng đầu tiên về tim đến đợt đầu tiên bị nhồi máu cơ tim của 330 bệnh án. Họ đều có khai rằng đã đi khám về triệu chứng tim trước khi bị nhồi máu, với thời gian như sau:

Thời gian	Số lượng bệnh nhân
1 ngày	4
2-6 ngày	12
1-3 tuần	20
1-2 tháng	28
3-6 tháng	54
7-11 tháng	45
1-2 năm	50
3 năm	23
4 năm	22
5 năm	16
6 năm	12
7 năm	8
8 năm	12
9 năm	4
10 năm trở lên	20
Cộng:	330

Câu hỏi.

1. Nên xác định quần thể định danh cho nghiên cứu này như thế nào?
2. Những trường hợp bệnh nào có khả năng đã không thu thập được thông tin cần nghiên cứu?
 3. Đề cập nào sau đây có thể thu được thông tin chính xác:
 - a) Hỏi các bệnh nhân nhớ lại triệu chứng tim sau khi bị nhồi máu cơ tim.
 - b) Theo dõi bệnh nhân từ khi họ xuất hiện triệu chứng tim đầu tiên (bao gồm cả những người không bị nhồi máu cơ tim).
 4. Có thể bình luận gì về những thời gian quá dài như trong bảng.

Trả lời:

Trả lời bài tập 1.

1. Phải là hồ sơ của tất cả các sản phụ của thành phố trong thời gian đó.
2. Tuy có chọn ngẫu nhiên, nhưng lại chỉ chọn của hai bệnh viện thì sao lại đại diện cho cả thành phố, và chẳng các bệnh viện lớn lại thường nhận nhiều trường hợp nặng và khó, lại rất nhiều tiền chi phí, nhiều người không thể vào viện này.

Trả lời bài tập 2.



1. Quần thể định danh nên là tất cả mọi trường hợp bị nhồi máu cơ tim của thành phố đó với các triệu chứng trước đó.

2. Nguồn thất thoát thông tin chủ yếu là những trường hợp không khai thác được triệu chứng trước khi xuất hiện NMCT, có thể là:

a) Chết ngay ở cơn NMCT đầu tiên.

b) Thiếu các bệnh nhân biết và khai báo về triệu chứng trước đó.

c) Nhớ lại triệu chứng không chính xác.

d) Sự thất thoát này còn có thể xảy ra khi thời gian giữa hai sự kiện liên quan kể trên quá ngắn và bệnh nhân không kịp đến khám triệu chứng tim.

3. Đề cập b) Sẽ cung cấp những thông tin chính xác nhất. Thật khó có thể hy vọng thu thập được thông tin chính xác về các chứng đã xảy ra sau khi bệnh nhân đã bị nhồi máu cơ tim, vì bản thân bệnh nhân khó còn khả năng nhớ chính xác, và nếu bệnh nhân tử vong thì càng không thể khai thác những thông tin đó từ người thân, trong số bạn bè, gia đình.

4. Khoảng cách giữa hai sự kiện đó càng dài thì càng dễ mắc sai lầm, vì khó mà có thể gán hai sự kiện đó với nhau, và bản thân bệnh nhân cũng không thể nhớ xác định được.

CƠ MẪU

I. MỤC TIÊU

1. Hiểu được cận kề những ý niệm về cơ mẫu, độ chính xác, lực mẫu, và quan hệ của chúng với nhau.

2. Chọn được một cơ mẫu thích hợp để đạt được một độ chính xác nhất định cho từng thể loại nghiên cứu.

3. Chọn được một cơ mẫu thích hợp để đạt tới được một lực mẫu nhất định cho từng thể loại nghiên cứu.

II. NỘI DUNG

Bất cứ cuộc nghiên cứu nào, chúng ta đều phải xác định cần bao nhiêu người trong nghiên cứu để đạt được mục tiêu nghiên cứu, nếu quá nhỏ kết quả sẽ không chính xác hoặc không phát hiện được kết quả nghiên cứu, nếu lớn quá sẽ tốn kém nhiều nguồn lực.

Trong thực hành, tùy thuộc nguồn lực ta có, và muốn với nguồn lực đó ta phải đạt được: liệu ước lượng nghiên cứu đủ chính xác? liệu có ý nghĩa thống kê? Phải tính toán đến cỡ mẫu.

Trước hết, ta đề cập đến cỡ mẫu ngẫu nhiên đơn để mô tả một hiện tượng sức khỏe (DTH mô tả).

Đơn giản nhất là ta chỉ nghiên cứu một biến trong một cuộc điều tra.

Ước lượng một tỷ lệ

Thí dụ ta muốn tiến hành một cuộc điều tra ở một khu vực nhất định để tìm tỷ lệ hiện mắc sốt rét ở trẻ em dưới 5 tuổi.

Có thể sử dụng độ chính xác tuyệt đối:

Thí dụ tỷ lệ hiện mắc sốt rét, được biết quần thể trẻ dưới 5 tuổi ở khu vực ta muốn nghiên cứu là khoảng 20% (thường gọi là tỷ lệ ước định, vì không biết chắc chắn lắm) và chúng ta muốn rằng tỷ lệ chúng ta sẽ nghiên cứu thấy (gọi là ước lượng kết quả) nằm trong khoảng 5% ở hai phía của tỷ lệ thực (nghĩa là từ 15% đến 25%) thì chúng ta lấy độ chính xác tuyệt đối là 5%.

Cũng có thể dùng độ chính xác tương đối:

Vẫn với thí dụ trên, nay chúng ta muốn ước lượng kết quả tìm được chỉ sai khác với tỷ lệ thực của quần thể đến 30% của tỷ lệ đó (nghĩa là 30% của tỷ lệ được biết 20%). Trường hợp này nếu quy ra độ chính xác tuyệt đối sẽ là 6% ($0,3 \times 0,2 = 0,06$) ở hai phía của tỷ lệ thực, như vậy là ta đã muốn phát hiện ra rằng tỷ lệ nghiên cứu được sẽ nằm trong khoảng từ 14% đến 26%).

Tính tỷ lệ: Nếu gọi số trẻ mắc SR là m, số trẻ được thăm khám là n, thì tỷ lệ mắc SR ở khu vực đó sẽ là $p = \frac{m}{n}$

Ước lượng này tất nhiên sẽ có sai số nhưng chúng ta sẽ có thể tin đến một mức tin cậy nào đó, thí dụ 95%, là tỷ lệ thực sẽ nằm trong khoảng:

$$p \pm 1,96 SE(p)$$

$$\text{hay } p \pm 1,96 \sqrt{\frac{pq}{n}}$$

trong đó SE(p) là sai số chuẩn của tỷ lệ p tìm được

$$q = 1-p$$

n = tổng trẻ em được thăm khám, ở đây là cỡ mẫu

pq: phương sai của tỷ lệ p

cụm số $1,96 \cdot SE(p)$ gọi là độ chính xác của tỷ lệ ở mức tin cậy 95% (viết dưới dạng chung là $w = Z_{1-\alpha/2} \cdot SE(p)$)

do đó công thức để tính cỡ mẫu n:

$$n = \frac{Z_{1-\alpha/2}^2 pq}{w^2}$$

$$CI = p \pm 1,96 \cdot SE$$

$$= p \pm 1,96 \sqrt{\frac{pq}{n}}$$

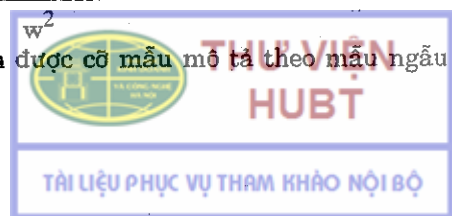
$$w = 1,96 \sqrt{\frac{pq}{n}}$$

$$= Z_{1-\alpha/2} \sqrt{\frac{pq}{n}}$$

$$n = \frac{Z_{1-\alpha/2}^2 \cdot pq}{w^2}$$

Như vậy để tính được cỡ mẫu mô tả theo mẫu ngẫu nhiên đơn, ta phải có sẵn và ước

$$w = Z_{1-\alpha/2} \cdot SE$$



lượng những thông tin sau:

1. Tỷ lệ ước lượng về tiêu thức (định nghiên cứu) trong quần thể.

Tỷ lệ ước lượng này càng chính xác với tỷ lệ thực lúc đó trong quần thể càng tốt. Nhưng cũng không cần chính xác tuyệt đối, vì nếu có ước lượng không sát, thì kết quả tỷ lệ nghiên cứu vẫn đáng tin cậy, không hề thay đổi, mà chỉ có độ lệch chuẩn của tỷ lệ là bị ảnh hưởng: ước lượng càng sát, độ lệch chuẩn kết quả càng nhỏ, ứng với tính chính xác càng cao.

2. Phải ấn định một ngưỡng xác suất sai lầm α loại I nhất định, gọi là sai lầm (ứng với sai lầm α là độ tin cậy $1 - \alpha$) mới có được giới hạn chuẩn $Z_{1-\alpha/2}$ của sai lầm đó:

$$\alpha = 0,05 \rightarrow Z_{1-\alpha/2} = 1,96$$

$$\alpha = 0,10 \rightarrow Z_{1-\alpha/2} = 1,645$$

$$\alpha = 0,01 \rightarrow Z_{1-\alpha/2} = 2,58$$

Tùy thuộc tính chất của nghiên cứu mà ta chọn α ở mức độ nào. Trong các nghiên cứu thông thường, người ta có thể chọn $\alpha = 0,05$.

3. Tính SE(p) sơ bộ từ n có thể chấp nhận được cho nghiên cứu:

$$SE(p) = \sqrt{\frac{pq}{n}}$$

Nhưng rõ ràng ở đây n chưa biết, nên nhiều trường hợp không biết được n là bao nhiêu thì có thể chấp nhận được, nên thường người ta ấn định ra một độ chính xác, trong đó người ta khuyên nên áp dụng độ chính xác tương đối cho dễ làm: độ chính xác tương đối có thể lấy từ 0,1 đến 0,5 của tỷ lệ ước định, tùy theo chúng ta cần chính xác nhiều hoặc ít. Khi chúng ta có khái niệm tương đối rõ ràng về tỷ lệ thực, thì dùng độ chính xác tuyệt đối đơn giản hơn.

Ước lượng một trung bình

Chúng ta cũng đề cập như trên, chỉ có ở đây thì:

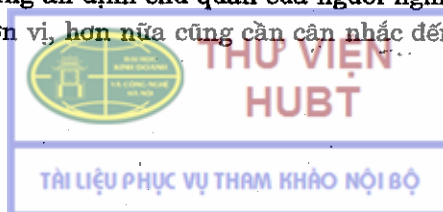
$$SE(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

$$\text{và } n = \frac{Z_{1-\alpha/2}^2 \cdot \sigma^2}{w^2} \text{ với } w = Z_{1-\alpha/2} \cdot SE(\bar{x})$$

Cũng như trên, để tính cỡ mẫu này, ta phải có được sai số chuẩn ước định (từ một nghiên cứu trước, hoặc phải thăm dò, trường hợp đặc biệt không cần chính xác lắm, ta có thể ước lượng thô bằng cách lấy khoảng phân tán của dãy thống kê biến liên tục chia cho 4, với khoảng phân tán là hiệu số của số đo lớn nhất với số đo nhỏ nhất của dãy số).

Chú ý:

Các công thức để tính cỡ mẫu trên đây cần được nhìn nhận chỉ như là một quy trình cung cấp cho ta một ước lượng thô của cỡ mẫu tối thiểu vừa đủ cần đòi hỏi (vì chúng đều dựa trên những ước định, và những ấn định chủ quan của người nghiên cứu) cho nên không cần lấy chính xác đến hàng đơn vị, hơn nữa cũng cần cân nhắc đến điều muốn đạt



được với khả năng thực thi hợp lý nữa. Ngoài ra, trong hầu hết các nghiên cứu mô tả, người ta còn cần tăng cỡ mẫu lên, khoảng 10%, để phòng những trường hợp không trả lời.

Bây giờ ta đề cập đến cỡ mẫu cho việc so sánh hai nhóm (DTH phân tích).

Việc so sánh hai nhóm trong dịch tễ học phân tích có thể thực hiện bằng: nghiên cứu thuần tập (tương lai hoặc hồi cứu).

Nghiên cứu bệnh chứng (chỉ là hồi cứu).

Nghiên cứu can thiệp (trong đó có thử nghiệm lâm sàng).

Tuy các cỡ mẫu cho các thể loại nghiên cứu này có khác nhau ít nhiều, song đều được xây dựng trên cùng những nguyên lý và nguyên tắc nhất định sau đây:

Ta hãy lấy một thí dụ về thử nghiệm lâm sàng cho dễ nhận biết: giả sử ta đem so sánh hiệu quả khỏi bệnh của loại thuốc A mới phát minh với loại thuốc B vẫn dùng từ trước đến nay (hoặc một loại placebo):

Thuốc	Kết quả	Khỏi	Không khỏi	Cộng Tỷ lệ khỏi
Thuốc mới A	a	b	a+b	$a/(a+b) = p_A$
Thuốc vẫn dùng B	c	d	c+d	$d/(c+d) = p_B$

Để tính cỡ mẫu cho việc so sánh hai nhóm, ngoài các thông tin cần có như đối với việc tính cỡ mẫu cho việc mô tả, còn cần có thêm:

Khái niệm về lực mẫu:

Giả sử ta có được hai tỷ lệ khỏi của hai loại thuốc đó: p_A của thuốc mới, và p_B của thuốc cũ, và khác biệt thực của hai tỷ lệ đó là D:

$$D = p_A - p_B$$

Kết quả nghiên cứu mẫu sẽ cho ta thấy được một khác biệt mẫu $d = p_A - p_B$, chúng ta có thể trắc nghiệm bằng những trắc nghiệm ý nghĩa thông thường, và tùy thuộc chủ yếu vào cỡ mẫu chúng ta đã chọn, khác biệt mẫu d có thể ý nghĩa thống kê hoặc không có ý nghĩa thống kê. Lực mẫu lúc này chính là xác suất để đảm bảo cho kết quả đạt tới ý nghĩa thống kê mong muốn, nếu sự khác biệt này là có thật. Lực mẫu được tính bằng $1-\beta$.

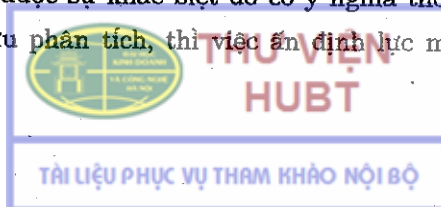
Lực mẫu phụ thuộc vào:

1. Các tỷ lệ thực khác nhau nhiều hay ít: Khác biệt càng lớn thì lực mẫu càng lớn.
2. Cỡ mẫu lớn hay nhỏ: Cỡ mẫu càng lớn thì lực mẫu càng lớn.

Nói một cách chung, thì lực mẫu là khả năng, là xác suất đảm bảo cho chúng ta có thể đạt được kết quả mong muốn: phát hiện ra được sự khác biệt có ý nghĩa thống kê).

Như vậy, mục tiêu của chúng ta là làm thế nào để chọn được một cỡ mẫu sao cho, nếu quả thật có sự khác biệt quan trọng về lâm sàng mà chúng ta cảm nhận thấy, chúng ta sẽ có cơ may tìm ra được sự khác biệt đó có ý nghĩa thống kê.

Trong các nghiên cứu phân tích, thì việc ấn định lực mẫu rất quan trọng. Và trong



các công thức tính cỡ mẫu, thì ứng với mỗi lực mẫu nhất định, cũng sẽ có một giá trị tới hạn Z_β nhất định. Những lực mẫu thường dùng là:

Nếu $\beta = 0,05$ thì $1 - \beta = 0,95$

$\beta = 0,10$ thì $1 - \beta = 0,90$

$\beta = 0,20$ thì $1 - \beta = 0,80$

$$1 - \beta = 0,95$$

Giá trị tới hạn Z_β của các lực mẫu thường dùng trên:

$1 - \beta = 0,80$ thì $Z_{1-\beta} = -Z_\beta = 0,86$

$1 - \beta = 0,90$ thì $Z_{1-\beta} = -Z_\beta = 1,28$

$1 - \beta = 0,95$ thì $Z_{1-\beta} = -Z_\beta = 1,645$

$$Z_{1-\beta} = 0,86 \quad 1 - \beta = 0,80$$

Có nhiều công thức tính cỡ mẫu cho các nghiên cứu phân tích tùy theo thể loại nghiên cứu (nghiên cứu thuần tập tương lai, thuần tập hồi cứu, nghiên cứu bệnh chứng, thử nghiệm lâm sàng, thử nghiệm cộng đồng...), đều sử dụng đến giá trị Z , ngoài các dữ kiện cần có cho mọi công thức tính cỡ mẫu khác.

Để tính cỡ mẫu trong các nghiên cứu phân tích, chúng ta cần chú ý tới các công thức tính phù hợp, với các ấn định hợp lý. Nói chung, phải tiến hành tuân tự như sau:

1. Ấn định một sự khác biệt về tỷ lệ khởi đàng kể trên lâm sàng.

Vấn đề này tùy thuộc vào thầy thuốc lâm sàng, vào bệnh nghiêm trọng hay không, phổ biến hay không...

Thí dụ: $80\% - 70\% = 10\%$

2. Ấn định một lực mẫu để có thể phát hiện ra sự khác biệt đó.

Thí dụ: 90%

3. Ấn định một mức ý nghĩa muốn dùng để kiểm định sự khác biệt tỷ lệ giữa hai nhóm (chính là giá trị p)

Thí dụ: $0,05$

4. Tính cỡ mẫu cho mỗi nhóm (nhóm chủ cứu và nhóm đối chứng). Theo công thức đơn giản nhất sau đây:

$$n = \frac{2pqF}{d^2}$$

trong đó $p = \frac{PA+PB}{2}$

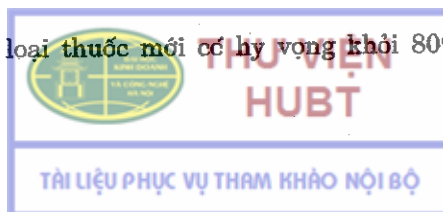
$$q = 1 - p$$

$$n = \frac{2pqF}{d^2} \quad p = \frac{p_1 + p_2}{2} \quad F = (Z_{\alpha/2} + Z_\beta)^2$$

F: Tính sẵn ở bảng kết hợp tương ứng giữa lực mẫu và mức ý nghĩa (xem ở bài thử nghiệm lâm sàng ngẫu nhiên trong sách này).

Khi muốn nghiên cứu ở những mức ý nghĩa khác, ở những α khác, ở những β khác, cần tìm ở các bảng tính sẵn các giá trị của chúng. Ở bảng này, người ta tính sẵn cho ta giá trị $F = (Z_{\alpha/2} + Z_\beta)^2$ là những trị số rất hay dùng đến cho việc tính các cỡ mẫu phân tích.

Thí dụ: ta muốn áp dụng một loại thuốc mới có hy vọng khởi 80% so với thuốc cũ



khỏi 70%, và muốn chắc chắn đến 90% là đạt được kết quả có ý nghĩa thống kê (lực mẫu 90%) và thỏa mãn với mức ý nghĩa $p < 0,05$, thì:

$$n = \frac{2(0,75 \times 0,25) \times 10,5}{(0,10)^2} = 394 \text{ cho mỗi nhóm}$$

Trường hợp so sánh 2 trung bình, cũng theo những nguyên lý và nguyên tắc trên, với công thức.

$$n = \frac{\sigma_1^2 + \sigma_2^2}{D^2} \times F$$

với $D = \bar{x}_1 - \bar{x}_2$

Chú ý: Các công thức kể trên đây là dành cho các cuộc nghiên cứu bằng mẫu ngẫu nhiên đơn hoặc chỉ định ngẫu nhiên đơn, dùng trong các nghiên cứu ngang (dùng ngẫu nhiên đơn), trong các thử nghiệm lâm sàng (chỉ định ngẫu nhiên đơn) và các nghiên cứu dọc tương lai.

Riêng đối với các nghiên cứu mẫu nhiều bậc, hoặc mẫu chùm thì cỡ mẫu tính được (theo các công thức trên) phải đem nhân lên với yếu tố hiệu chỉnh để đảm bảo hiệu quả thiết kế, vì tính chính xác của ước lượng bị giảm đi do ta sử dụng thiết kế nhiều bậc (chứ không phải ngẫu nhiên đơn).

Còn trong trường hợp ta dùng mẫu tầng, thì nên chỉ định vào các tầng theo tỷ lệ với tầng quần thể (gọi là chỉ định tầng theo tỷ lệ) và sẽ được độ chính xác tối ưu. Nhưng nếu các kết quả của mỗi tầng là rất cần thiết, thì nên tính cỡ mẫu riêng cho từng tầng.

Một điểm nữa cũng cần phải chú ý đến trong việc tính cỡ mẫu là, nếu ta ước lượng sự biến thiên của trị số nghiên cứu từ một cuộc điều tra nghiên cứu trước đó, thì cần phải biết độ lệch chuẩn của sự biến thiên đó. Nếu nghi ngờ rằng cuộc điều tra nghiên cứu trước là không chính xác, tuy vẫn dùng độ lệch chuẩn đó, nhưng người ta có thể gấp đôi cỡ mẫu lên, nếu cần.

Cũng cần nói thêm rằng, cỡ mẫu đã tính được như trên là cho các mẫu rút ra từ một quần thể vô hạn trong các cuộc điều tra nghiên cứu khác nhau. Còn trong trường hợp quần thể là hữu hạn, ta phải dùng yếu tố hiệu chỉnh cho cỡ mẫu trong quần thể hữu hạn. Có nhiều phương pháp để điều chỉnh cỡ mẫu trong trường hợp này, đơn giản nhất là:

$$n_f = \frac{n_i N_f}{n_i + N_f}$$

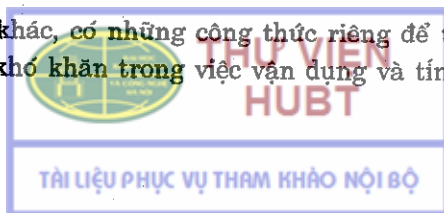
trong đó N_f là kích thước của quần thể hữu hạn.

n_f là cỡ mẫu sẽ rút ra từ quần thể hữu hạn này.

n_i cỡ là mẫu tính được từ quần thể vô hạn theo những công thức đã trình bày trên đây.

Ngoài phương pháp tính cỡ mẫu theo công thức như đã làm trên đây người ta còn dùng các bảng tính sẵn, tiện lợi hơn.

Đối với các loại mẫu khác, có những công thức riêng để tính cho mỗi loại, tuy chính xác hơn, song có những khó khăn trong việc vận dụng và tính toán không đề cập ở đây.



III. BÀI TẬP

Bài tập 1.

Muốn nghiên cứu tỷ lệ được khám thai trong 3 tháng đầu, với mức tin cậy 95%, và tỷ lệ ước định là 40%, và muốn không sai số quá 5% của tỷ lệ thật.

Câu hỏi: cỡ mẫu n cần bao nhiêu?

Trả lời: Dùng công thức mô tả: ước lượng tỷ lệ (theo độ chính xác tương đối)

$$n = \frac{4 \times 0,40 \times 0,60}{(0,05 \times 0,4)^2} = 2400 \text{ thai phụ}$$

Bài tập 2.

Muốn nghiên cứu mức tăng thể trọng chuột con được nhận chế độ dinh dưỡng đặc biệt, với mức tin cậy 95%, và một nghiên cứu trước đã cho biết $x = 100g$ với $s=20$, và muốn kết quả không có sai số vượt quá 5% (hệ số chính xác tương đối 0,05), người ta phải cần bao nhiêu chuột?

Câu hỏi: cỡ mẫu n?

Trả lời:

Dùng công thức mô tả: ước lượng trung bình (độ chính xác tương đối)

$$n = \frac{Z_{1-\alpha/2}^2 s^2}{(0,05x)^2} = \frac{(1,96)^2 \times 20^2}{(0,05)^2 (100)^2} = 62 \text{ chuột con.}$$

Bài tập 3.

Muốn tìm hiểu tác dụng độc của một chất được coi là độc trong môi trường, người ta đã đem xử lý trên chuột xem tỷ lệ chuột sống sót sau một tuần, với mức tin cậy 95%, một nghiên cứu tương tự đã cho biết tỷ lệ sống sót $p = 0,20$ và độ chính xác tuyệt đối $e = 0,03$, phải có n là bao nhiêu?

Câu hỏi: cỡ mẫu n?

Trả lời:

Dùng công thức mô tả: ước lượng tỷ lệ (độ chính xác tuyệt đối)

$$n = \frac{Z_{1-\alpha/2}^2 pq}{e^2} = \frac{(1,96)^2 \times 0,2 \times 0,8}{(0,03)^2} = 683 \text{ chuột}$$

Bài tập 4.

Trị số albumin huyết thanh vẫn được coi là chuẩn trước nay ở mức 5,0 g/100ml. Nay muốn so sánh trị số trung bình albumin huyết thanh của các bệnh nhân mắc bệnh M, với $\alpha = \beta = 0,05$, và độ lệch chuẩn được biết là $s = 0,3g$, và với điều kiện trị số đó khác biệt với trị số chuẩn ít nhất là $d \geq 0,2g$.

Hỏi: cỡ mẫu n?

Trả lời: Dùng công thức so sánh trung bình mẫu với trung bình chuẩn:

$$n = \frac{[(Z_{1-\alpha/2}) - Z_{\beta}]^2}{\left(\frac{d}{s}\right)^2} = \frac{[1,96 - (-1,645)]^2}{\left(\frac{0,2}{0,3}\right)^2} = 30 \text{ bệnh nhân mắc bệnh M}$$



Chú ý: Ở đây cần đến khái niệm về sai lầm beta (sai lầm loại II: Chấp nhận giả thuyết H_0 trong khi nó sai) và khái niệm khác biệt d có ý nghĩa trên thực tế lâm sàng (khác biệt rõ ràng với trị số chuẩn tối thiểu là bao nhiêu mới được coi là mắc bệnh M)

Bài tập 5.

Trước nay vẫn coi là chuẩn một tỷ lệ khối bệnh $p_1 = 30\%$. Nay muốn nghiên cứu một can thiệp mới, với hy vọng tỷ lệ khối tăng lên ít nhất là 5% (khác biệt d so với tỷ lệ chuẩn, có ý nghĩa trên thực tế lâm sàng, vận dụng vào công thức tính là độ chính xác tuyệt đối) với $\alpha = \beta = 0,05$.

Hỏi: cỡ mẫu n ?

Trả lời:

Dùng công thức tỷ lệ mẫu với tỷ lệ chuẩn:

$$\begin{aligned} n &= \frac{(Z_{1-\alpha/2}\sqrt{p_1q_1} - Z_\beta\sqrt{p_2q_2})^2}{d^2} \\ &= \frac{[1,96\sqrt{0,3\cdot0,7} - (-1,645\sqrt{0,35\cdot0,65})]^2}{(0,05)^2} \\ &= 1.133 \text{ can thiệp mới} \end{aligned}$$

Bài tập 6.

Có hai chế độ ăn cho trẻ suy dinh dưỡng, nay muốn so sánh hai chế độ ăn đó bằng cách cân lại trẻ sau một thời gian nhất định, với mong muốn khác biệt có ý nghĩa thực tế lâm sàng là 3 kg ở mức $\alpha = \beta = 0,05$.

Được biết rằng độ lệch chuẩn (biến thiên về lên cân) chấp nhận là 4kg.

Hỏi: cỡ mẫu n ?

Trả lời:

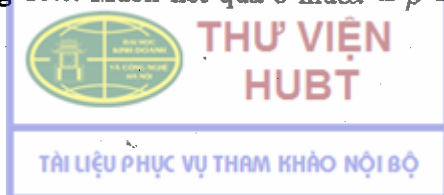
Ta sẽ dùng công thức so sánh hai trung bình ở hai nhóm nghiên cứu mẫu:

$$\begin{aligned} n &= \frac{(Z_{1-\alpha/2} - Z_\beta)^2}{\frac{d^2}{s_1^2 + s_2^2}} \\ &= \frac{[1,96 - (-1,645)]^2}{\frac{3^2}{4^2 + 4^2}} \\ &= 46 \text{ trẻ ở mỗi nhóm.} \end{aligned}$$

Bài tập 7.

Có một loại thuốc mới điều trị bệnh M, muốn đem so sánh tác dụng của nó với tác dụng của loại thuốc vẫn dùng và sẽ được coi là tác dụng tốt hơn, nếu tỷ lệ khối khác biệt nhau ít nhất từ 0,10 trở lên mới có ý nghĩa thực tế lâm sàng. Được biết thuốc vẫn dùng có tỷ lệ khối p_1 khoảng 40%. Muốn kết quả ở mức $\alpha = \beta = 0,05$.

Hỏi: cỡ mẫu n ?



Trả lời:

Dùng công thức so sánh hai tỷ lệ ở hai nhóm nghiên cứu mẫu:

$$n = \frac{(Z_{1-\alpha/2} - Z_{\beta})^2 \cdot 2pq}{d^2}$$
$$= \frac{[1,96 - (-1,645)]^2 \times 2 \times 0,45 \times 0,55}{(0,5 - 0,4)^2}$$
$$= 645 \text{ bệnh nhân cho mỗi nhóm.}$$

Bài tập 8.

Người ta mới áp dụng phương pháp điều trị A có thể có hiệu lực gấp 2 lần phương pháp điều trị B đang dùng với tỷ lệ khỏi 17,5%. Muốn kết quả có ý nghĩa với $\alpha = 0,05$ và $\beta = 0,10$.

Câu hỏi: Cần phải có cỡ mẫu n là bao nhiêu cho mỗi nhóm?

Trả lời: Dùng công thức nghiên cứu thuần tập tương lai:

$$n = \frac{(Z_{1-\alpha/2} \sqrt{2PQ} + Z_{1-\beta} \sqrt{P_1Q_1 + P_2Q_2})^2}{(P_1 - P_2)^2}$$

Cũng có thể dùng công thức đơn giản.

$$n = \frac{2PQF}{(P_1 - P_2)^2} = \frac{2 \times 0,2625 \times 0,7375 \times 10,5}{(0,35 - 0,175)^2} = 135$$

Vậy mỗi phương pháp A và B đều phải theo dõi ở mỗi nhóm gồm 135 người bệnh.

Bài tập 9.

Người ta muốn đánh giá hiệu lực của vaccin BCG ở trẻ em bằng cách so sánh tỷ lệ BCG ở nhóm trẻ mắc lao và nhóm trẻ không mắc làm đối chứng. Biết rằng có tới 30% trẻ trong quần thể không tiêm vaccin. Ấn định $\alpha = 0,05$; $\beta = 0,20$ và muốn có OR ≥ 2 (mới đạt khác biệt có ý nghĩa).

Hỏi: cỡ mẫu n?

Trả lời:

Dùng công thức của nghiên cứu bệnh chứng (kiểm định giả thuyết đối với tỷ suất chênh lệch OR).

$$n = \frac{(Z_{1-\alpha/2} \sqrt{2PQ} + Z_{1-\beta} \sqrt{P_1Q_1 + P_2Q_2})^2}{(P_1 - P_2)^2}$$
$$= 130 \text{ trẻ cho mỗi nhóm.}$$

Bài tập 10.

Giả sử dùng lâu dài thuốc tránh thai có làm tăng nguy cơ mắc bệnh mạch vành. Nay muốn phát hiện xem có thật gia tăng nguy cơ tương đối RR $\geq 1,3$ bằng nghiên cứu bệnh chứng, với $\alpha = \beta = 0,05$, và được biết rằng tỷ lệ nữ có dùng thuốc tránh thai ở nhóm đối chứng (lấy trong quần thể chung) là 0,30.



THƯ VIỆN
HUBT

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

Hỏi: cỡ mẫu n?

Trả lời:

Dùng công thức đơn giản của nghiên cứu bệnh chứng.

$$n = \frac{(Z_{1-\alpha/2} - Z_{\beta})^2 2pq}{(p - p_2)^2} = \frac{p_1 + p_2}{2}$$
$$= \frac{(1,96 + 1,645)^2 \times 2 \times 0,2225 \times 0,7775}{(0,245 - 0,2)^2} = 2220 \text{ cho mỗi nhóm}$$

Bài tập 11.

Câu hỏi:

Người ta đã nghiên cứu ở một mẫu tính được là 30 phụ nữ để tìm số con trung bình của họ ở một quần thể 600 phụ nữ ở tuổi sinh đẻ, thấy $\bar{x} = 3,6$, $s_x = 2,12$.

Nay, ta chỉ có quần thể 320 phụ nữ thì muốn thấy kết quả tương tự phải nghiên cứu một mẫu là bao nhiêu?

Hỏi: cỡ mẫu n?

Trả lời:

Dùng công thức điều chỉnh cỡ mẫu ở quần thể hữu hạn:

$$nf = \frac{n_i \times N_f}{n_i + N_f} = \frac{30 \times 320}{30 + 320} = 28 \text{ phụ nữ.}$$

MẪU NGẪU NHIÊN ĐƠN

I. MỤC TIÊU

1. Hiểu tầm quan trọng của ngẫu nhiên trong nghiên cứu mẫu.
2. Vận dụng đúng kỹ thuật ngẫu nhiên để chọn mẫu.
3. Làm được một nghiên cứu với mẫu ngẫu nhiên đơn.

II. NỘI DUNG

Nhắc lại một số khái niệm cần thiết:

1. Tính ngẫu nhiên

Tính ngẫu nhiên phải được tuân thủ theo khái niệm của toán học. Các số ngẫu nhiên



định chọn ra, nên dựa vào bảng số ngẫu nhiên, đơn giản và tiện dùng. Cần chú ý, trước khi vào bảng số ngẫu nhiên, phải định ra quy ước một cách rõ ràng, quy ước này sẽ phải được duy trì thống nhất từ đầu đến cuối quá trình lấy số ngẫu nhiên.

Trong mẫu ngẫu nhiên đơn, thì đơn vị thống kê là đơn vị để rút ngẫu nhiên. Nói khác đi, tính ngẫu nhiên ở mẫu ngẫu nhiên đơn rơi vào chính đơn vị thống kê (đơn vị trên đó các đặc trưng nghiên cứu được đo, đếm) như một người, một người bệnh, một súc vật thí nghiệm...

Mẫu có đại diện cho quần thể nghiên cứu hay không, thì tính ngẫu nhiên chịu trách nhiệm một phần quan trọng.

2. Các tình huống sử dụng mẫu ngẫu nhiên đơn.

Trước hết, số đơn vị thống kê ở quần thể nghiên cứu không quá nhiều. Trong các tài liệu hiện có, không tài liệu nào nói đến giới hạn này. Tuy nhiên, qua những kinh nghiệm của nhiều cuộc nghiên cứu, con số này không bao giờ vượt quá 10.000 (còn giới hạn dưới sẽ tùy thuộc vào cỡ mẫu, và sao cho việc lấy ngẫu nhiên là có ý nghĩa). Nếu số đơn vị thống kê vượt quá con số này, cần phải sử dụng hoặc mẫu chùm, hoặc mẫu tầng.

Bất kỳ đặc trưng nào cũng có thể sử dụng mẫu ngẫu nhiên đơn để tiến hành nghiên cứu. Tuy nhiên quần thể ở đây thường nhỏ, nên có thể sẽ ít tốn kém và dễ dàng hơn khi phải sử dụng những kiến thức cần kè, những sinh vật phẩm, nguyên vật liệu, trang thiết bị tinh vi, đắt tiền để khảo sát các đặc trưng định nghiên cứu. Phần lớn các tổng kết lâm sàng đơn giản, các cuộc điều tra thăm dò nhỏ đều sử dụng mẫu ngẫu nhiên đơn.

3. Cỡ mẫu ngẫu nhiên.

Cỡ mẫu ngẫu nhiên đơn lớn, nhỏ tùy thuộc vào tỷ lệ P của tiêu thức nghiên cứu trong quần thể (nếu là nghiên cứu tỷ lệ) hoặc vào độ lệch chuẩn sơ qua của tiêu thức nghiên cứu (nếu là nghiên cứu số trung bình). Đối với các nghiên cứu mô tả thì tính theo công thức:

$$n = \frac{Z^2_{1-\alpha/2} PQ}{e^2} \quad \text{hoặc} \quad n = \frac{Z^2_{1-\alpha/2} \sigma^2}{e^2}$$

Cần chú ý sử dụng các công thức để tính cỡ mẫu:

- Tỷ lệ tiêu thức nghiên cứu trong quần thể không quá nhỏ. Theo lý thuyết thì tỷ lệ 0,10 trở lên mới có thể dễ dàng kéo về phân phối chuẩn. Tuy nhiên, có thể sử dụng được đối với tỷ lệ tiêu thức 0,05. Dưới tỷ lệ này, khái niệm ít chính xác hoặc không chính xác sẽ được đưa ra nhận định về cỡ mẫu tính theo những công thức này. Đối với các nghiên cứu trung bình, khái niệm về độ lệch chuẩn sigma cũng được áp dụng tương ứng thông lệ này.

- Việc chọn độ chính xác e là cực kỳ quan trọng. Cần phải hiểu rõ, và nắm vững ý nghĩa của nó mới vận dụng được đúng đắn trong việc tính cỡ mẫu.

Cũng cần lưu ý điều chỉnh cỡ mẫu, nếu nghiên cứu được tiến hành trong quần thể hữu hạn, theo công thức:

$$n_f = \frac{n_i \times N_f}{n_i + N_f}$$



Có thể tra cỡ mẫu từ các bảng tính sẵn theo từng mức tin cậy và sai lầm α là một phía hay hai phía. (Đối với các vấn đề khác có liên quan, xin xem chi tiết ở bài "cỡ mẫu" trong sách Dịch tễ học đại cương, quyển 2, Đại học Y Hà Nội).

4. Một vài điểm cần chú khác.

Trên đây là đề cập đến mẫu ngẫu nhiên đơn dùng để mô tả một hiện tượng sức khỏe, hoặc để ước lượng một tỷ lệ, hoặc để ước lượng một số trung bình trong một quần thể nhất định, tất cả đều là trường hợp tiến hành với một mẫu duy nhất trong quần thể.

Cũng vẫn nằm trong việc mô tả, nhưng khi ta muốn ước lượng về một sự khác biệt nào đó của hai tỷ lệ (hoặc của hai số trung bình) ở hai quần thể khác nhau, thì mẫu ngẫu nhiên đơn cũng được dùng, nhưng công thức để tính cỡ mẫu cho mỗi nhóm ở một quần thể sẽ khác đi.

Đối với các nghiên cứu phân tích, thì kỹ thuật ngẫu nhiên vẫn phải sử dụng và mẫu ngẫu nhiên đơn cũng vẫn được dùng, nhưng công thức tính cỡ mẫu sẽ khác đi, và phải tuân theo những thủ tục chọn mẫu riêng cho từng loại mẫu của nghiên cứu phân tích.

III. CÁC BƯỚC TIẾN HÀNH NGHIÊN CỨU BẰNG MẪU NGẪU NHIÊN ĐƠN

Mẫu ngẫu nhiên đơn là hình ảnh cơ bản nhất của các loại mẫu ngẫu nhiên. Dưới đây chỉ đề cập đến mẫu ngẫu nhiên duy nhất dùng để mô tả một hiện tượng sức khỏe trong một quần thể cho dễ hiểu.

1. Xác định vấn đề nghiên cứu.

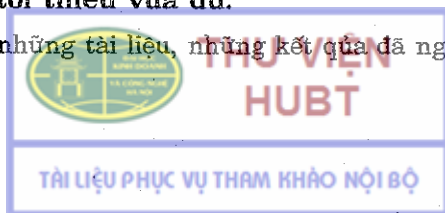
Mô tả bằng mẫu ngẫu nhiên đơn thuộc loại hình nghiên cứu thực tế ứng dụng. Nó phục vụ cần thiết trước mắt cho các vấn đề y tế (quản lý, chủ trương, điều hành... y tế) và cũng có mục tiêu đề xuất một giả thuyết nhân quả cho các vấn đề y học, là gợi ý cho nhiều nghiên cứu sau đó. Chọn vấn đề nghiên cứu nhỏ, tùy thuộc ưu tiên của từng ngành, chuyên môn y học, nhưng đều có chung một điểm quan trọng là bao giờ cũng phải mô tả cùng một lúc hiện tượng sức khỏe với một nguy cơ quan tâm của hiện tượng sức khỏe đó.

2. Định nghĩa hiện tượng sức khỏe và nguy cơ của nó.

Để đảm bảo tính chặt chẽ của nghiên cứu, một trong những công việc đầu tiên quan trọng là đưa ra được định nghĩa thật rõ ràng, chính xác của hiện tượng sức khỏe nghiên cứu (thí dụ: một bệnh phổi bao gồm những triệu chứng, dấu hiệu gì ở các triệu chứng, dấu hiệu đó được xác định bằng phương pháp, loại xét nghiệm nào? và ở mức độ là bao nhiêu? như đo thân nhiệt thì đặt ở đâu? trong bao lâu? bao nhiêu độ thì coi là sốt trong trường hợp đó?) và đồng thời cũng đưa ra được định nghĩa của nguy cơ cần nghiên cứu cùng lúc với hiện tượng sức khỏe đó (thí dụ được kể là có khám thai thì là khám ở đâu? người khám có trình độ nào đã đi khám bao nhiêu lần? khám vào những tháng thứ mấy của quá trình mang thai?...

3. Tính cỡ mẫu tối thiểu vừa đủ.

Cần phải tìm đọc những tài liệu, những kết quả đã nghiên cứu về vấn đề đó, và khảo



sát tình hình cụ thể hiện tại trong quần thể dự định tiến hành nghiên cứu, để ít nhất là có một ý niệm về P quần thể (nếu muốn mô tả một tỷ lệ) hoặc sigma (nếu muốn mô tả một trung bình) để làm cơ sở cho việc tính cỡ mẫu sát hợp. Cần nhắc lại rằng những ý niệm thu thập được kể trên được ước lượng càng chính xác càng tốt, vì nó sẽ ảnh hưởng đến cỡ mẫu sẽ nghiên cứu, còn đối với kết quả nghiên cứu, nó không làm sai lệch kết quả của bản thân tỷ lệ (hoặc trung bình) tìm thấy, nhưng lại có ảnh hưởng quan trọng đến độ phân tán của các kết quả đó: ước lượng càng ít chính xác thì phương sai càng lớn.

4. Thiết lập một khung mẫu.

Đó là bản liệt kê danh mục các đơn vị thống kê của quần thể mà nghiên cứu sẽ được tiến hành trong quần thể đó. Nếu đơn vị thống kê là người bệnh chẳng hạn, thì bản danh mục sẽ bao gồm tên của từng người bệnh gắn liền mỗi tên người bệnh với một số thứ tự của họ (số thứ tự này sẽ được sử dụng làm mã số cho từng người, thường được gọi là nhãn mẫu theo ý niệm thống kê).

5. Chọn ngẫu nhiên đơn vị thống kê vào mẫu.

Nếu dùng bảng số ngẫu nhiên, phải lấy các số ngẫu nhiên có cùng số chữ số (số ký tự) với mã số (số thứ tự) lớn nhất trong bảng danh mục khung mẫu. Có nhiều cách lấy các số ngẫu nhiên này. Có thể mỗi lần "vào bảng" thì được một số, như vậy ít nhất cũng phải vào bảng một số lần bằng với cỡ mẫu tính được. Nên người ta thường quy ước luôn cả hướng ngẫu nhiên: Thí dụ có thể quy ước sẽ lấy số ngẫu nhiên gồm 4 chữ số liền nhau về phía bên phải con số "vào bảng", rồi lần lượt đi xuống phía dưới chẳng hạn, nếu gặp số nào bằng và nhỏ hơn số thứ tự cuối cùng thì số đó được chọn vào mẫu, còn khi gặp số nào lớn hơn số đó, hoặc những số trùng lặp thì đi tiếp để lấy số khác, đến bao giờ đủ được số ngẫu nhiên bằng với số cỡ mẫu đã tính. Nếu có điều kiện, chúng ta sẽ rút các số ngẫu nhiên này từ máy vi tính, nhanh chóng và đơn giản. Còn đối với quần thể tham khảo nhỏ có thể dùng kỹ thuật rút thăm hoặc một vài kỹ thuật khác.

6. Xây dựng đề cương và quy trình nghiên cứu (hoặc bộ câu hỏi).

Mỗi ngành chuyên môn, mỗi vấn đề nghiên cứu sẽ có quy trình nghiên cứu riêng của mình được xây dựng đáp ứng với mục tiêu nghiên cứu cụ thể. Nhưng chúng đều có chung phần quan trọng là "vật liệu và phương pháp nghiên cứu" phải được miêu tả và quy định một cách cụ thể, chi tiết và chính xác, và có quy định thời điểm, lịch trình từng bước nghiên cứu. Các bản tổng hợp cũng được xây dựng lúc này.

7. Tổ chức, triển khai tiến hành cuộc nghiên cứu.

Phần chuẩn bị các nguồn lực là rất cơ bản, cùng với cả sự chuẩn bị về tổ chức, tuyên truyền giáo dục... thật chu đáo. Ở bước này, điều quan trọng để tránh một số loại sai số là tuân thủ triệt để quy trình nghiên cứu, thực hiện đúng và đủ, có giám sát chặt chẽ, khi đã tiến hành là không thay đổi, sửa chữa, bổ sung quy trình nghiên cứu.

8. Tập hợp, xử lý kết quả như thường lệ.

IV MỘT BÀI TẬP MẪU NGẪU NHIÊN ĐƠN ĐỂ MÔ TẢ

Ở một xã 10.000 dân, có tỷ lệ sinh 0,02, báo cáo trong những năm gần đây có tình



trạng gia tăng số trẻ đẻ thiếu cân. Nay muốn mô tả tỷ lệ đó, ở ngưỡng xác suất 0,05 với độ chính xác tuyệt đối 0,10, biết rằng tỷ lệ hiện nay có thể lên tới 30%.

Bài làm

1. Xác định vấn đề: trẻ thiếu cân = trẻ cân nặng dưới 2500g lúc sinh (mô tả nguy cơ kèm theo: sản phụ giảm ăn về lượng gạo đến 25%).

2. Tính cỡ mẫu: 80 trẻ.

3. Lấy hồ sơ sơ sinh của xã trong 12 tháng qua, có cân lúc sinh gồm 200 cháu. Lập danh sách, chọn ngẫu nhiên ra 80 trẻ bằng bảng số ngẫu nhiên.

4. Đếm số trẻ thiếu cân: giả sử thấy 20 trẻ.

5. Tính tỷ lệ: 0,25

sai số chuẩn: 0,05

Vậy tỷ lệ trẻ đẻ thiếu cân ở xã đó trong 12 tháng qua là:

$$p = 0,25 \pm 0,1$$

ở ngưỡng xác suất 5%

V. BÀI TẬP

(Chứa ở các buổi thảo luận nhóm nhỏ)

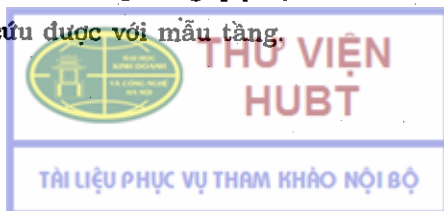
Hãy viết đề cương chi tiết về quy trình nghiên cứu theo mẫu ngẫu nhiên đơn để mô tả (không hạn chế quy mô nghiên cứu):

1. Tình trạng suy dinh dưỡng ở trẻ 5 tuổi.
2. Phụ nữ áp dụng biện pháp tránh thai bằng bao cao su.
3. Sơ sinh được bú sữa non.
4. Tiêu chuẩn ở trẻ con dưới 2 tuổi và dưới 5 tuổi.

MẪU TẦNG

I. MỤC TIÊU

1. Chia được quần thể nghiên cứu thành các tầng sát hợp.
2. Thiết kế được mẫu tầng đúng quy định.
3. Làm nghiên cứu được với mẫu tầng.



II. NỘI DUNG

Chúng ta hãy nhắc lại một số khái niệm cần thiết sau:

1. Mẫu tầng thường được thực thi ở quần thể lớn.

Trong đó tiêu thức định nghiên cứu khác nhau nhiều, giữa các khu vực khác nhau (như ở đặc điểm dân số học khác nhau, ở không gian khác nhau).

2. Cân nầm vững định nghĩa của tầng với tính chất quy định của nó: mỗi tầng là một tập hợp đơn vị thống kê có tiêu thức nghiên cứu sàn sàn bằng nhau (để đảm bảo phương sai nội tầng nhỏ). Định nghĩa này kéo theo một quy định nữa là tiêu thức nghiên cứu giữa tầng nọ với tầng kia là rất khác nhau (phương sai liên tầng lớn).

3. Số tầng trong mẫu tầng.

Về nguyên lý mà nói, để cứ có tiêu thức nghiên cứu khác nhau là được chia sang một tầng khác, nghĩa là chúng ta có thể chia quần thể nghiên cứu thành nhiều tầng bao nhiêu cũng được, để đảm bảo cho tiêu thức nghiên cứu gần như đồng nhất trong mỗi tầng. Nhưng về thực hành mà nói, không thể chia quần thể nghiên cứu thành quá nhiều tầng, dù ở quần thể có kích thước như thế nào, tiêu thức nghiên cứu phân tán như thế nào, số tầng trong mẫu cũng không vượt quá 10 tầng trong mẫu tầng (để dễ phiên giải kết quả nghiên cứu và có nhiều lợi ích thực tế trong thực hành y học và y tế).

4. Do tính chất chọn mẫu tầng có thể coi mẫu tầng là một tập hợp nhiều mẫu ngẫu nhiên đơn nhỏ, để tiến hành và cho kết quả chính xác.

5. Cơ mẫu tầng. Được tính như với mẫu ngẫu nhiên đơn.

6. Phân chia cơ mẫu vào cho các tầng.

Có thể chia đều cho số tầng đã định (ít chính xác hơn). Thường người ta phân chia cơ mẫu đã tính được vào các tầng theo tỷ lệ phần trăm đơn vị thống kê của từng tầng: tầng nào lớn sẽ khảo sát nhiều đơn vị thống kê hơn tầng nhỏ.

Cần chú ý là trên đây ta đề cập đến việc chia cơ mẫu tỷ lệ theo phần trăm thống kê của từng tầng (chứ không chỉ là chia theo phần trăm số dân). Nhưng để cho dễ tiến hành, nếu số đơn vị thống kê thường tỷ lệ với số dân (thí dụ số đơn vị thống kê gồm các cá thể ở một lứa tuổi nhất định) thì có thể phân chia cơ mẫu vào các tầng tỷ lệ với số dân.

Một điểm nữa cũng cần phải chú ý là khi cỡ mẫu nhỏ, mà ta lại chia thành quá nhiều tầng, sẽ có thể nảy sinh tình trạng có những tầng có số đơn vị thống kê được nghiên cứu nhỏ hơn 30, và như trên đây đã nói, thực chất ở mỗi tầng là một mẫu ngẫu nhiên đơn thành phần, nên số đơn vị thống kê nhỏ hơn 30 ở một tầng là không được chấp nhận. trong trường hợp này, người ta buộc phải hoặc gộp tầng lại cho số tầng ít đi hoặc tăng cỡ mẫu (vẫn đảm bảo phân chia vào tầng theo tỷ lệ) sao cho tầng nhỏ nhất cũng có được số đơn vị thống kê được nghiên cứu ít nhất là 30.

7. Chọn đơn vị thống kê nghiên cứu trong từng tầng.

Sau khi đã chia ấn định số đơn vị thống kê được nghiên cứu vào từng tầng, các đơn vị này sẽ được chọn theo kỹ thuật ngẫu nhiên trong tổng số đơn vị thống kê của tầng đó (cách làm như đối với mẫu ngẫu nhiên đơn). Cộng tất cả lại sẽ được cỡ mẫu tầng.



8. Tiến hành điều tra, thăm khám, mô tả đặc trưng nghiên cứu theo quy trình nghiên cứu hoặc bộ câu hỏi đối với từng đơn vị thống kê, rồi tập hợp, xử lý kết quả như thường lệ. Để cho đơn giản, và cũng là có ý nghĩa, người ta chỉ chú trọng đến phương sai nội tầng.

III. CÁCH TIẾN HÀNH NGHIÊN CỨU THEO MẪU TẦNG

Có thể có một chi tiết có yêu cầu khá cao hơn với người nghiên cứu mẫu tầng là phải đọc nhiều tài liệu xung quanh tiêu thức định nghiên cứu để có được khái niệm rõ ràng về sự khác biệt của tiêu thức nghiên cứu giữa tầng nọ với tầng kia (thí dụ như tử vong hoặc mắc bệnh có thật khác nhau theo lứa tuổi, nhóm tuổi, nghề nghiệp, trình độ văn hóa, khu vực địa lý nào, cũng như sự khác nhau đó như thế nào... để định liệu việc chia tầng cho hợp lý nhất có thể được). Những sự khác biệt giữa nông thôn thành thị, dùng nước máy nước ao hồ, nông dân có nghề phụ, không có nghề phụ... đều là những sự khác biệt rõ rệt, và nghiên cứu theo tầng là cần thiết, để nhận biết, còn những chi tiết khác khó nhận biết hơn, nhà nghiên cứu cần có những căn cứ nhất định để chia tầng.

- Sau khi đã xác định vấn đề nghiên cứu cần và có thể chia tầng hợp lý, nhà nghiên cứu cần rà soát lại ở quần thể mình sẽ tiến hành nghiên cứu việc chia tầng dự kiến như thế có phù hợp và khả năng thực thi như thế nào.

- Chia tầng quần thể.

Tùy theo mục tiêu và vấn đề nghiên cứu, xem có thể làm mẫu tầng không tỷ lệ được không. Thông thường thì người ta tiến hành mẫu tầng theo tỷ lệ với số đơn vị thống kê của từng tầng (chứ không phải bao giờ cũng chia tỷ lệ với số dân). Chỉ khi nào số đơn vị thống kê. Thường tỷ lệ với số dân (như lứa tuổi chẳng hạn) thì mới chia mấy tầng tỷ lệ với số dân mà thôi.

- Tính cỡ mẫu: như đối với mẫu ngẫu nhiên đơn (nếu là nghiên cứu mô tả) theo mẫu tầng).

- Chia cỡ mẫu vào các tầng.

Trường hợp chia cỡ mẫu tỷ lệ với số dân, thì trước hết lấy số dân theo tầng rồi tính ra phần trăm cho mỗi tầng, rồi căn cứ vào phần trăm thành phần đó mà chia cỡ mẫu vào các tầng, tầng nào đông dân thì phần cỡ mẫu ở đó sẽ lớn.

- Chọn các đơn vị thống kê trong mỗi tầng: tuân thủ kỹ thuật ngẫu nhiên như đối với mẫu ngẫu nhiên đơn.

- Tiến hành điều tra, khảo sát, mô tả đặc trưng nghiên cứu theo quy trình nghiên cứu, rồi tập hợp, xử lý kết quả như thường lệ.

IV. MỘT BÀI TẬP VỀ MẪU TẦNG DÙNG TRONG NGHIÊN CỨU MÔ TẢ

Một thầy thuốc muốn nghiên cứu tình trạng suy dinh dưỡng của trẻ em dưới 5 tuổi ở



một huyện với 100.000 dân có tỷ lệ sinh là 0,02, ở rải rác thành 20 xã, tương đối đều nhau về nhiều khía cạnh. Theo kinh nghiệm của mình và trao đổi với các bạn đồng nghiệp, thầy thuốc thấy cần mô tả tình trạng suy dinh dưỡng riêng rẽ ở những trẻ là con đầu lòng, và là con thứ ba trong gia đình. Ở đây mẫu tầng đã được sử dụng sau nhiều cân nhắc thỏa đáng: tầng con đầu lòng và tầng con thứ ba sau.

- Trước hết, tính cỡ mẫu, căn cứ vào:

- + Tỷ lệ ước lượng SDD trẻ dưới 5 tuổi: 30%
- + Ấn định mức tin cậy của nghiên cứu: 95%
- + Ấn định độ chính xác tương đối: 0,10 của p

Tra bảng tính sẵn: cỡ mẫu tối thiểu cần thiết 900 trẻ dưới 5 tuổi.

- Chia cỡ mẫu vào các tầng:

Theo những thu thập có được trong 5 năm gần đây trung bình có khoảng 30% phụ nữ độ tuổi sinh đẻ sinh con thứ 3 (họ vẫn có thể có con đầu lòng dưới 5 tuổi, khoảng 20%).

Như vậy, trong số trẻ dưới 5 tuổi, có thể có khoảng 3 trẻ là con thứ ba và 9 trẻ là con đầu lòng tương ứng lần lượt là 25% và 75%. Vậy ta sẽ điều tra:

$900 \times 0,25 = 225$ trẻ dưới 5 tuổi là con thứ ba, và

$900 \times 0,75 = 675$ trẻ dưới 5 tuổi là con đầu lòng.

- Chọn ngẫu nhiên trẻ vào nghiên cứu:

Mỗi xã như vậy có trung bình 5.000 dân thì sẽ có 100 trẻ 0 - 1 tuổi và 500 trẻ dưới 5 tuổi. Ta sẽ chọn 2 xã để nghiên cứu là đáp ứng được cỡ mẫu. Lập danh mục các xã của huyện rồi bắt thăm ra hai xã một cách ngẫu nhiên. Số trẻ đầu lòng và số trẻ là con thứ ba trong gia đình sẽ được nghiên cứu tất cả ở hai xã đó.

- Giả sử thu được kết quả như sau:

Thứ hạng con/gia đình	Tình trạng dinh dưỡng	Có SDD	Không SDD	Cộng
Là con thứ 3		182	78	260
Là con đầu lòng		183	517	700
Cộng		365	595	960

+ Trước hết tìm ý nghĩa thống kê của kết quả: $\chi^2 = 152,89$ tính được lớn hơn nhiều lần χ^2 trong bảng ở một bậc tự do, như vậy kết quả có ý nghĩa thống kê ở ngưỡng xác suất $p < 0,05$ dự định.

+ Tính tỷ lệ SDD chung của mẫu nghiên cứu $p = 0,38$ và sai số chuẩn theo mẫu tầng tỷ lệ $SEp = 0,016$ trong đó, ở ngưỡng xác suất 0,05.

Trẻ SDD ở tầng con thứ 3: $p_1 = 0,70$ $SEp_1 = 0,03$

Trẻ SDD ở tầng con đầu lòng: $p_2 = 0,26$ $SE = 0,02$

Vì các trẻ được nghiên cứu ở mẫu tầng này được chọn theo tỷ lệ vào mẫu một cách ngẫu nhiên nên có thể dẫn đến kết luận sau:

- **Kết luận:** Tình hình SDD ở huyện được nghiên cứu trên số trẻ dưới 5 tuổi là $P = 0,38 \pm 0,016$ (với độ chính xác tương đối là 0,1P và ở ngưỡng xác suất $p < 0,05$) trong

đó tỷ lệ SDD ở trẻ là con thứ 3 trong gia đình là 0,67 đến 0,73 và ở trẻ là con đầu lòng chỉ là 0,24 đến 0,28.

Ghi chú

Cần nói thêm rằng, nghiên cứu này chỉ nhằm mục đích mô tả, và một giả thuyết được nêu lên (như tất cả các nghiên cứu mô tả với những loại mẫu ngẫu nhiên khác) là bước đầu nhận thấy tỷ lệ SDD ở trẻ là con thứ 3 trong gia đình cao hơn nhiều so với tỷ lệ SDD ở trẻ là con đầu lòng. Muốn kết luận sự khác biệt này, phải làm tiếp một nghiên cứu phân tích, trong đó nhóm chủ cứu nên lấy số trẻ là con thứ 3 và nhóm đối chứng là con đầu lòng, và cỡ mẫu phải được tính với công thức của nghiên cứu phân tích.

V. BÀI TẬP

(Thảo luận nhóm nhỏ để chữa bài tập)

1. Những vấn đề sức khỏe nào sau đây nên tiến hành nghiên cứu mô tả theo mẫu tầng? tại sao?: uống vaccin Sabin, tiêu chảy, viêm gan virus, thiếu dinh dưỡng protein-năng lượng, còi xương, bấu cổ đơn thuần, tai nạn lao động công nghiệp.

2. Hãy viết đề cương chi tiết về quy trình nghiên cứu theo mẫu tầng về một hiện tượng sức khỏe (tự đề xuất quy mô và phạm vi nghiên cứu) thuộc chuyên ngành của mình, hoặc về một hiện tượng sức khỏe, bệnh tật... mà mình quan tâm.

MẪU CHÙM

Gọi đầy đủ thì là mẫu ngẫu nhiên chùm hoặc mẫu chùm ngẫu nhiên, nhưng chúng ta thường quen dùng "mẫu chùm".

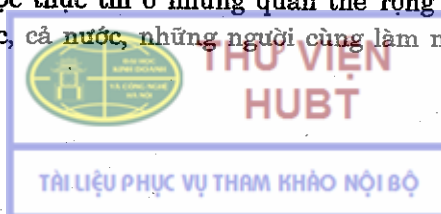
I. MỤC TIÊU

1. Thiết kế được mẫu chùm đúng quy định.
2. Vận dụng được những hiện tượng sức khỏe có thể tiến hành nghiên cứu bằng mẫu chùm.

II. NỘI DUNG

Một số khái niệm cơ bản cần thiết nhắc lại:

1. Mẫu chùm được thực thi ở những quần thể rộng lớn, như ở khu vực hành chính: huyện, tỉnh, khu vực, cả nước, những người cùng làm một nghề nghiệp nhất định: công



nhân đường sắt, công nhân cao su, nông dân, giáo viên, những người ở một lứa tuổi nhất định: trẻ em 12-23 tháng, trẻ em dưới 5 tuổi... với số lượng lớn cần nghiên cứu, đánh giá.

2. Cần nắm vững định nghĩa của chùm với tính chất đặc biệt của nó: chùm là một tập hợp cá thể, có đặc trưng nghiên cứu ở chùm này so với chùm khác phải ở mức sàn sàn bằng nhau, không chênh lệch nhau quá nhiều, (mà không cần biết đến trong nội bộ mỗi chùm thì đặc trưng nghiên cứu đó như thế nào). Cỡ chùm cũng cần xấp xỉ như nhau.

3. Cần tránh quan niệm cho rằng lấy đơn vị hành chính làm chùm là "tiện lợi" nhất giống như chương trình tiêm chủng mở rộng, khi đánh giá tỷ lệ tiêm chủng đã lấy xã là một chùm, ở đây lấy xã làm chùm là đúng vì các xã đều được huấn luyện, trang bị, cung cấp... về TCMR như nhau, do đó có thể cho rằng đặc trưng nghiên cứu sàn sàn như nhau. Và chẳng, đánh giá tỷ lệ tiêm chủng là đánh giá trên số trẻ 12 - 23 tháng, nghĩa là đối tượng nghiên cứu dựa theo lứa tuổi, nó luôn luôn tỷ lệ với số dân của xã..., còn trong những trường hợp khác, khi đối tượng nghiên cứu không tỷ lệ với số dân của xã (như thu nhập bình quân đầu người, tỷ lệ trẻ con bỏ học...) hoặc như đặc trưng nghiên cứu khác nhau quá xa như trong việc mô tả tình hình cạnh thị của học sinh cấp 2 bằng mẫu chùm chẳng hạn, nếu coi xã là chùm khi tình trạng trường học của các xã quá khác nhau (xã có trường học, có xã không có, xã có trường xây cửa kính có đèn điện, lại có xã có trường che liếp không có điện...) thì kết quả sẽ là không tin cậy được.

4. Trong mẫu chùm thì tính ngẫu nhiên rơi vào chùm, nói khác đi, các chùm đều có xác suất phản ánh vào mẫu nghiên cứu như nhau.

5. Có hai loại hình mẫu chùm hiện dùng: mẫu chùm đích thực và mẫu chùm cải tiến. Vì đều là mẫu chùm, nên bắt buộc tính ngẫu nhiên rơi vào chùm, nghĩa là các chùm phải được rút ra bằng kỹ thuật ngẫu nhiên có lập.

Ở mẫu chùm đích thực, sau khi các chùm đã được chọn ra theo cách ngẫu nhiên, thì tất cả mọi đơn vị thống kê trong các chùm đó đều được nghiên cứu về đặc trưng nghiên cứu của chúng. Loại hình này cho kết quả chính xác hơn loại mẫu chùm cải tiến sau đây.

Ở mẫu chùm cải tiến, sau khi các chùm đã được chọn ra theo cách ngẫu nhiên, người ta không khảo sát đặc trưng nghiên cứu trên tất cả đơn vị thống kê của từng chùm, mà ở mỗi chùm này chỉ khảo sát đặc trưng nghiên cứu ở một số đơn vị thống kê thôi, số đơn vị thống kê này phải bằng nhau ở mọi chùm ngẫu nhiên đã được rút ra, và cũng được rút thăm một cách ngẫu nhiên.

Vận dụng loại hình mẫu chùm nào là tùy thuộc vào đặc trưng nghiên cứu, mục tiêu và quy mô nghiên cứu, v.v... trong đó mẫu chùm cải tiến dễ tính toán trong thiết kế hơn.

Trong chương trình TCMR, người ta theo kinh nghiệm, quy định số chùm được rút ra một cách ngẫu nhiên là con số 30, đối với mẫu chùm cải tiến, và chúng ta có thể áp dụng loại hình này cho nhiều nghiên cứu y học và y tế khác.

Tuy nhiên, muốn chính xác (vì bản thân mẫu chùm ngẫu nhiên không chính xác bằng mẫu ngẫu nhiên đơn) chúng ta có thể tính số chùm c sẽ được rút ngẫu nhiên ra theo công thức sau đây, để đảm bảo cả hiệu quả thiết kế nghiên cứu:

$$c = \frac{pq[1+(b-1)r]}{s^2 b}$$

Cũng cần lưu ý thêm ở đây là với cùng một cỡ mẫu thì trong mẫu chùm nghiên cứu,



số chòm càng nhiều, càng trải rộng trong quần thể nghiên cứu thì kết quả nghiên cứu càng chính xác.

7. Cỡ chòm.

Việc quyết định quy mô nào được coi là một chòm là một vấn đề quan trọng: một xã, một làng hay chỉ là một xóm, một xí nghiệp, một phân xưởng hay chỉ là công đoạn, thì cần cân nhắc kỹ lưỡng, trên những nguyên lý chủ yếu sau:

Cỡ chòm không được lớn đến nỗi chỉ một số ít chòm đã chứa đựng hết cả đơn vị thống kê của cỡ mẫu mà như đã đề cập ở trên, số chòm nghiên cứu càng nhiều kết quả càng chính xác hơn điều đó kéo theo khái niệm cỡ chòm phải đủ nhỏ, nhất là đối với loại hình mẫu chòm đích thực.

8. Cỡ mẫu chòm.

Tính cỡ mẫu như đối với mẫu ngẫu nhiên đơn, rồi đem nhân đôi lên.

9. Mẫu chòm, có thể dùng trong dịch tễ học mô tả để mô tả một hiện tượng sức khỏe cùng với yếu tố nguy cơ của nó, hoặc dùng trong dịch tễ học phân tích để kiểm định một giả thuyết. Về nguyên lý và phương pháp tiến hành là hoàn toàn giống nhau, cố nhiên mỗi trường hợp đều phải tuân thủ những quy định riêng của mục tiêu mô tả hay phân tích, mặc dầu về hình thức đều là mẫu chòm ngẫu nhiên.

III. CÁCH TIẾN HÀNH MỘT NGHIÊN CỨU NGANG BẰNG MẪU CHÒM

- Sau khi xác định vấn đề nghiên cứu giống như các nghiên cứu mẫu khác, cần xác định sẽ dùng loại hình mẫu chòm nào, và đặc biệt là phải xác định chòm là đơn vị nào (xã, thôn, trường, lớp, bệnh viện, khoa...) sao cho cỡ chòm là hợp lý (cân đối giữa cỡ mẫu, cỡ chòm, số chòm).

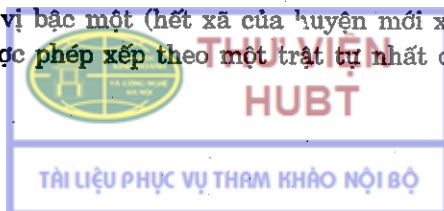
- Tính cỡ mẫu như đối với mẫu ngẫu nhiên đơn (nếu là nghiên cứu mô tả) và số chòm nếu cần.

- Lập danh mục chòm:

Thường người ta kẻ thành một bảng để tiện dùng cho cả mẫu chòm đích thực và mẫu chòm cải tiến, như sau:

TT (mã chòm)	Tên đơn vị chòm (xã, trường, bv...)	Số đơn vị thống kê từng chòm	Tổng dồn số đơn vị thống kê	Chòm được chọn vào mẫu

Trong việc lập danh mục chòm cần tuân thủ một số quy định quan trọng, mang tính quyết định đối với tính chính xác của mẫu chòm, đặc biệt là đối với mẫu chòm cải tiến khi ta lấy ngẫu nhiên theo cách hệ thống. Vì mẫu chòm là một loại mẫu hai bậc, nên khi trong quần thể nghiên cứu có nhiều đơn vị bậc một (nhiều huyện, nhiều xí nghiệp được phẩm, nhiều trường...) thì các đơn vị bậc hai (xã, phân xưởng, lớp học) phải được lập danh mục theo từng đơn vị bậc một (hết xã của huyện mới xếp liền sang xã của huyện khác) nhưng lại không được phép xếp theo một trật tự nhất định (thí dụ như huyện nào cũng



xếp theo trật tự có tỷ lệ tiêm chủng cao rồi dần thấp xuống). Nếu xếp theo trật tự nhất định như thế thì tiêu thức nghiên cứu sẽ mang tính tuần hoàn trong khung mẫu, và mẫu chòm trong trường hợp này sẽ là rất xấu, đặc biệt là trường hợp có khoảng cách chòm tương ứng với số đơn vị bậc hai tương đối bằng nhau trong khung chòm của mẫu.

- Chọn ngẫu nhiên chòm:

Nếu dùng mẫu chòm đích thực, chỉ cần rút thăm ra một số số ngẫu nhiên (có số ký tự bằng số ký tự của con số chòm lớn nhất, tương ứng với số chòm cần thiết, các chòm có mã trùng với các số ngẫu nhiên đó sẽ được lấy vào mẫu nghiên cứu, rồi khảo sát tất cả đơn vị thống kê của các chòm được lấy r ngẫu nhiên này.

Nếu dùng mẫu chòm cải tiến, thường tính ngẫu nhiên được làm theo cách lấy hệ thống nghĩa là chỉ cần lấy một số ngẫu nhiên lúc đầu để xác định chòm đầu tiên sau đó cứ theo trật tự hệ thống mà xác định các chòm tiếp theo cho đủ số chòm quy định. Người ta sẽ chỉ khảo sát ở mỗi chòm một số đơn vị thống kê bằng với thương số của cỡ mẫu với số chòm. Số đơn vị thống kê này cũng phải được chọn ngẫu nhiên trong số đơn vị của chòm được chọn (vì không khảo sát tất cả mọi đơn vị thống kê của các chòm này). Số đơn vị thống kê này bằng nhau ở các chòm.

- Tiến hành điều tra, khảo sát, mô tả đặc trưng nghiên cứu theo quy trình nghiên cứu hoặc các bộ câu hỏi, và tập hợp xử lý kết quả như thường lệ. Để cho đơn giản, người ta có thể chỉ tính phương sai liên chòm.

IV. MỘT BÀI TẬP VỀ MẪU CHÒM ĐÍCH THỰC VỚI MỤC ĐÍCH MÔ TẢ

Một nhà sản khoa muốn đánh giá công tác chăm sóc sức khỏe ban đầu đối với các phụ nữ mang thai xem họ có được khám thai trong 3 tháng đầu ở một tỉnh đồng bằng (có tỷ lệ sinh là 0,02, với 2,9 triệu dân sinh sống trên một mật độ tương đối đều nhau ở 16 huyện, mỗi huyện có 20 xã, mỗi xã có 4 làng) bằng một nghiên cứu ngang theo mẫu chòm:

1. Xác định vấn đề nghiên cứu: tỷ lệ phụ nữ có được khám thai trong 3 tháng đầu là bao nhiêu?

2. Tính cỡ mẫu:

Tỷ lệ khám thai 3 tháng đầu theo báo cáo là 60%

Lấy độ chính xác tương đối là 5%

Mức tin cậy của nghiên cứu ấn định là 95%

Tính theo mẫu NND, cỡ mẫu sẽ tính được là 1050 phụ nữ có mang từ 3 tháng trở lên, nên cỡ mẫu chòm này sẽ là 2100.

3. Tính thử số phụ nữ có mang tại một xã (dự định lấy xã làm chòm) 6.000 dân x 0,02 = 120 trẻ ra đời trong năm, giả sử số sinh được phân bố đều trong 12 tháng thì mỗi tháng có khoảng 10 phụ nữ mới mang thai, thì lúc nào cũng có thể gặp khoảng 60 phụ nữ mang thai trên 3 tháng.

4. Có thể tiến hành mẫu chòm đích thực với $2100 : 60 = 35$ chòm (trong nghiên cứu này không cần tính chính xác số chòm theo công thức).



5. Lập khung chùm theo tên huyện, trong mỗi huyện xếp theo vần chữ cái tên các xã và mã số của từng xã.

Huyện	TT xã	Xã	Đề ghi kết quả		
			Số phụ nữ có thai > 3 tháng	Số được khám 3 tháng đầu	Tỷ lệ được khám
A	1	A N			
	2	B X			
	30				
B	31	A H			
	32	C D			
	33	D C			

6. Rút ngẫu nhiên chùm (xã) theo kỹ thuật ngẫu nhiên hệ thống:

- Tính khoảng cách chùm: $480 \text{ xã} : 35 \text{ chùm} = 13$

- Lấy một số ngẫu nhiên nhỏ hơn khoảng cách chùm:

Thí dụ đó là số 2 thì xã BX là chùm đầu tiên được nghiên cứu. Tiếp đó, đem số ngẫu nhiên ban đầu là số 2 này cộng với một khoảng cách mẫu là số 13. Ta được chùm thứ hai (xã số $2+13=15$) rồi cứ tiếp tục cộng thêm với khoảng cách mẫu ta được chùm thứ 3 (xã số $15+2=17$), chùm thứ tư ($17+2=\text{xã số } 19$)... ta sẽ có đủ 35 xã được nghiên cứu.

7. Nhà sản khoa đó đã tiến hành nghiên cứu ở 35 xã nữa mới được rút ngẫu nhiên, tất cả các phụ nữ có thai 3 tháng trở lên trong các xã đó đều trả lời, giả sử có kết quả tổng hợp như sau:

Có 1150 người được khám trong số 2100 người được điều tra.

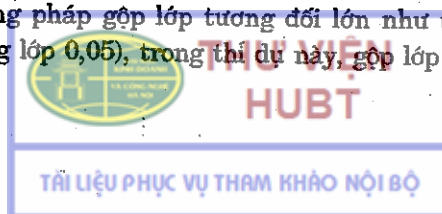
Như vậy, có tỷ lệ: $1150/2100=0,55$

Đây là mẫu chùm có cùng kích thước, nên để tính độ lệch bình phương tỷ lệ mẫu chùm này, ta lập bảng sau:

Tỷ lệ được khám thai	Tiêu thức gộp lớp	Trị số giữa	Tần số chùm	$p_i - p$	$\sum (p_i - p)$
270/300	0,85-0,94	0,90	5	+0,35	+1,75
144/185	0,75-0,84	0,80	3	+0,25	+0,75
84/120	0,65-0,74	0,70	4	+0,15	+0,60
36/60	0,55-0,64	0,60	1	0	0
80/160	0,45-0,54	0,50	4	-0,05	-0,20
240/600	0,35-0,44	0,40	10	-0,15	-1,50
144/480	0,25-0,34	0,30	8	-0,25	-2,00
1150/2100			35		-0,6

Tính $V_p = 0,0005$

Tính theo phương pháp gộp lớp tương đối lớn như thế này, ít chính xác hơn gộp lớp nhỏ (thí dụ gộp từng lớp 0,05), trong thí dụ này, gộp lớp tương đối lớn chỉ để cho dễ hiểu.



Và kết quả thu được mẫu chùm này là:

$$p = 0,55 \pm 0,01$$

Với mức ý nghĩa 0,05 và độ chính xác tương đối là 0,05p. Thực ra trong mẫu chùm, người ta thường quan tâm đến phương sai liên chùm nhiều hơn.

Thêm vào đó, cũng trong kết quả điều tra cũng chỉ ra rằng: nguyên nhân của không khám thai là:

- Thấy không cần khám (15% chiếm 82% sản phụ có mang thai (nhiều lần)).
- Sợ tiêm uốn ván thai phụ (5%)
- Bận công việc quá (17%).
- Không nghĩ là có mang thai (3%).

V. MỘT BÀI TẬP VỀ MẪU CHÙM CẢI TIẾN DÙNG TRONG DỊCH TỄ HỌC MÔ TẢ

Chúng ta hãy xem thí dụ đối với mẫu chùm cải tiến thường dùng trong việc đánh giá tỷ lệ tiêm chủng của chương trình TCMR (gọi là mẫu 30 chùm 7 trẻ). Trong nghiên cứu này sau nhiều tính toán và kinh nghiệm người ta quy định cỡ mẫu là 210 trẻ, số chùm là 30, do đó ở mỗi chùm chỉ khảo sát 7 trẻ để xem tình hình tiêm chủng của một tỉnh đồng bằng gồm 2,4 triệu dân phân bố đều trên 24 huyện, và 400 xã, mỗi xã trên dưới 6000 dân, và đều có tỷ lệ sinh là 0,02. Làm như sau:

1. Xác định chùm: ở đây có thể coi xã là một chùm
2. Lập danh mục chùm, theo bảng sau

TT (mã số chùm)	Tên xã (chùm)	Số dân	Cộng dồn số dân	Chùm được chọn

Cũng như với mẫu chùm đích thực, các chùm cũng được xếp theo thứ tự vần chữ cái trong một huyện, hết huyện nọ đến huyện kia (tránh tiêu thức bị tuần hoàn). Ở đây phải lấy số dân làm mốc tỷ lệ: xã nào đông dân có thể có 2 chùm sẽ rơi vào đó. Và chằng ở đây đơn vị thống kê là trẻ em 12-23 tháng, nghĩa là căn cứ vào lứa tuổi, mà lứa tuổi lại tỷ lệ với số dân (số dân đông thì số trẻ 12-23 tháng cũng nhiều), cho nên ở nghiên cứu TCMR này căn cứ vào số dân (chính ra phải căn cứ vào số trẻ 12-23 tháng) để chọn chùm nghiên cứu sẽ dễ có số liệu hơn, sát hợp hơn.

3. Chọn chùm vào mẫu:

a) Tìm khoảng cách mẫu.

Lấy tổng số dân chia cho 30 (chùm) sẽ được khoảng cách mẫu. Khoảng cách mẫu ở đây là tính theo số dân nên thường là những số lớn. Trong thí dụ này, khoảng cách mẫu sẽ là: $2.400.000 : 30 = 80.000$.



THƯ VIỆN
HUBT

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

b) Rút lấy một số ngẫu nhiên nhỏ hơn khoảng cách mẫu. Ở đây khoảng cách mẫu là 80.000 là con số gồm 5 chữ số (có nhiều người thích gọi là 5 ký tự). Thí dụ ta rút được số ngẫu nhiên đầu tiên này là 43210.

c) Xác định các chòm được chọn vào nghiên cứu:

- Số ngẫu nhiên đầu tiên này quy định chòm 1 được chọn vào nghiên cứu: ta căn cứ vào cột tổng dồn số dân, ở hàng nào có tổng dồn số dân bắt đầu lớn hơn số ngẫu nhiên này là xã tương ứng được chọn là chòm thứ nhất của mẫu.

- Chọn chòm thứ hai: đem số ngẫu nhiên 43210 cộng với khoảng cách mẫu 80.000 ta sẽ được con số 123.210. Vẫn căn cứ vào cột tổng dồn số dân, ở hàng nào có tổng dồn số dân bắt đầu lớn hơn số 123.210 thì xã tương ứng ở hàng đó là được chọn làm chòm thứ hai vào mẫu nghiên cứu.

- Chọn các chòm tiếp theo: cũng theo nguyên tắc trên, cho đến $43210 + [(80.000 (30-1))]$ sẽ là chòm cuối cùng, được chọn vào mẫu. Chú ý đối với những xã có số dân lớn, có thể có hai chòm cùng rơi vào một xã.

d) Xác định đơn vị thống kê được nghiên cứu.

Vẫn phải áp dụng nguyên tắc ngẫu nhiên. Ở đây có nhiều cách chọn ngẫu nhiên.

- Lập danh mục trẻ con 12-23 tháng của xã, rồi chọn ngẫu nhiên ra 7 đứa trẻ bằng kỹ thuật ngẫu nhiên đơn, hoặc ngẫu nhiên hệ thống. Phương pháp này tốt, nhưng khó thực hiện, nên người ta thường áp dụng phương pháp "cồng gần cồng" với những quy định ngẫu nhiên nhiều bậc trước đó, như lấy ngẫu nhiên thôn, ngẫu nhiên xóm, ngẫu nhiên ngõ và ngẫu nhiên nhà đầu tiên, rồi từ đó tìm theo quy định cho đủ 7 trẻ, ở đứa trẻ cuối cùng nếu có anh chị em sinh đôi sinh ba, thì phải khai thác cả (và chòm này sẽ có 8 hoặc 9 trẻ).

e) Tiến hành khai thác các đặc trưng trên các trẻ đó theo phiếu quy định.

f) Tính kết quả:

$$p = \frac{168 \text{ trẻ nhận đủ, đúng 6 loại vaccin}}{210 \text{ trẻ điều tra}} = 80\%$$

Muốn tính chuẩn, vì trong nghiên cứu này người ta lấy mẫu sai số chòm theo tỷ lệ ở các chòm có kích thước khác nhau, nên cách làm cũng như đối với mẫu chòm đích thực trên đây, sau khi tính thấy $SEp=0,03$ và kết quả được ghi là:

Tỷ lệ trẻ 12-23 tháng trong năm 1990 ở tỉnh X được nhận đủ và đúng 6 loại vaccin trong chương trình TCMR là:

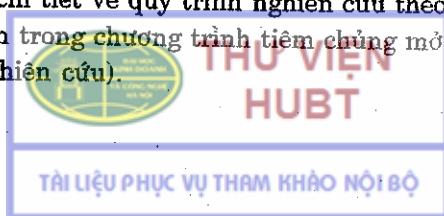
$$p = 0,80 \pm 0,06$$

với ý nghĩa 0,05 và độ chính xác tương đối là 0,05p. Cũng như trên đã nêu, ở mẫu chòm người ta thường quan tâm nhiều đến phương sai liên chòm.

III. BÀI TẬP

(Chứa bài tập trong các buổi thảo luận nhóm nhỏ)

Hãy viết đề cương chi tiết về quy trình nghiên cứu theo mẫu chòm để mô tả: uốn ván sơ sinh, hiệu lực vaccin trong chương trình tiêm chủng mở rộng (tự đề xuất phạm vi, quy mô, typ mẫu chòm nghiên cứu).



THIẾT KẾ BỘ CÂU HỎI, KỸ THUẬT PHÒNG VẤN VÀ ĐIỀN CÂU HỎI

I. MỤC TIÊU

1. Hiểu rõ ý nghĩa và vai trò của từng dạng câu trong bộ câu hỏi để có thể vận dụng đúng từng dạng câu hỏi trong chuyên ngành của mình.
2. Biết tác dụng của kỹ thuật phỏng vấn và cách điền bộ câu hỏi.
3. Xây dựng được một bộ câu hỏi đơn giản phục vụ cho một mục tiêu điều tra cộng đồng theo ngành chuyên môn của mình.

II. NỘI DUNG Ý NGHĨA, VAI TRÒ, PHƯƠNG PHÁP XÂY DỰNG MỘT BỘ CÂU HỎI

Mục tiêu của tất cả các cuộc điều tra là thu thập thông tin và phương tiện để thu thập là bộ câu hỏi. Như vậy bộ câu hỏi phải thật rõ ràng, phản ánh rõ mục tiêu của việc thu thập số liệu. Trước khi thiết kế bộ câu hỏi người nghiên cứu phải thật rõ về:

- a) Tại sao cần phải tiến hành cuộc điều tra?
- b) Kết quả của cuộc điều tra sẽ được dùng thế nào?

Sau khi hai câu hỏi này được trả lời rõ ràng, thì lúc đó người nghiên cứu bắt tay vào việc thiết kế bộ câu hỏi. Chương này khu trú vào việc làm thế nào để thiết kế một bộ câu hỏi hoàn chỉnh, cho phép thu thập thông tin cần thiết một cách chính xác.

Một bộ câu hỏi điều tra gồm một loạt những câu hỏi để hỏi đối tượng nghiên cứu. Những đối tượng nghiên cứu chỉ có thể trả lời đầy đủ và chính xác những câu hỏi này nếu từng câu hỏi này được thiết kế thật cẩn thận cả về người lẫn từng từ sử dụng trong đó. Phải cân nhắc thật cẩn thận những điểm sau:

1. Loại cấu trúc câu hỏi: Liệu câu hỏi là mở hay đóng.
2. Sắp xếp từ và câu trong các câu hỏi: có rõ và dễ hiểu.
3. Trình bày câu hỏi: có đủ chỗ cho người trả lời điền không.
4. Giúp người trả lời: cân nhắc những yếu tố tác động tới khả năng của người trả lời trong việc trả lời chính xác.
5. Giúp và huấn thị với người hỏi: Nhắc nhở họ điều gì họ cần nói và làm ở những

giai đoạn khác nhau của cuộc phỏng vấn.

Những chủ đề chính sau đây cần thảo luận trong chương này:

1. Bắt đầu bộ câu hỏi.
2. Cấu trúc câu hỏi.
3. Làm cho bộ câu hỏi đúng.
4. Trình bày bộ câu hỏi.
5. Nhớ và kể lại.
6. Một số điểm chung cần cân nhắc

Thứ tự, chúng ta sẽ đi qua từng chủ đề một.

A. BẮT ĐẦU BỘ CÂU HỎI.

Có nhiều cách bắt đầu thiết kế một bộ câu hỏi, cách sau đây là cách thực tế và rất có ích để bắt đầu:

Bước 1: Viết ra thật ngắn một, hai hoặc ba câu, hoặc là dưới dạng một danh sách ngắn, những mục đích quan trọng nhất của nghiên cứu đang chuẩn bị tiến hành.

Bước 2. Lập một danh sách những thông tin tương ứng trực tiếp và cần thiết phù hợp với những mục tiêu nghiên cứu.

Bước 3. Quyết định những phần chính trong bộ câu hỏi như:

- (a) Phần về hành chính.
- (b) Những phần riêng cho mỗi khía cạnh chính của nghiên cứu.

Bước 4. Trong mỗi phần và đối với mỗi khoản thông tin cần thu thập, rút ra những câu hỏi để hỏi và cho phép thu thập được những thông tin này.

Bước 5. Kiểm tra từng câu hỏi đề ra ở bước 4 xem có phản ánh đúng những mục tiêu nghiên cứu đề ra ở bước 1 không. Loại bỏ những phần hoặc câu hỏi không đúng.

Bước 6. Kiểm tra xem danh sách những câu hỏi ở bước 5 có cho phép thu thập và ghi lại toàn bộ những thông tin cần thiết cho nghiên cứu không.

Bước 7. Kiểm tra mỗi câu hỏi ở bước 5 và 6 và tự hỏi liệu:

- (a) Câu đó đã rõ ràng và không quá tham không.

Người trả lời sẽ hiểu câu hỏi này không?

- (b) Đơn giản và ngắn.

Câu hỏi dài và dùng từ không thông dụng có thể làm nhiều người trả lời nhầm.

- (c) Câu hỏi có hợp lý khi hỏi không?

Những người trả lời sẽ có đủ hiểu biết và kinh nghiệm cần thiết để trả lời câu hỏi không?

- (d) Là câu hỏi mở hay đóng.

Bước 8. Kiểm tra lại để đảm bảo cuộc phỏng vấn không quá dài.

Một số điểm cần cân nhắc thêm:



THƯ VIỆN
HUBT

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

Khi xây dựng bộ câu hỏi, thường có xu hướng thêm ngoài phần câu hỏi cần thiết cho mục tiêu cuộc điều tra.

Xu hướng này cần phải tránh để khỏi gây những hậu quả có thể tránh được cho cả người hỏi lẫn người được hỏi và nếu bộ câu hỏi quá dài sẽ làm giá thành cuộc điều tra tăng lên. Thông thường thì những câu hỏi thêm này không được phân tích sử dụng khi kết thúc cuộc điều tra và vì vậy toàn bộ chi phí, thời gian và cố gắng là phí phạm.

Tầm quan trọng của việc chia bộ câu hỏi thành những phần thích hợp cần thiết phải được nhận thức rõ, ở đầu mỗi phần này cần ghi rõ đề mục theo thứ tự logic nhằm giúp cho người lập kế hoạch điều tra, người phỏng vấn và cả người trả lời trong quá trình phỏng vấn. Một sự tiến triển logic của thứ tự các câu hỏi trong từng phần và trong các phần của bộ câu hỏi sẽ cho phép người trả lời những câu dài và khó. Ví dụ, trong một cuộc nghiên cứu chăm sóc bà mẹ trẻ em, đầu tiên nên bắt đầu bằng những thông tin về hành chính, tên, tuổi, địa chỉ, ngày tháng năm sinh,... nghề nghiệp, sau đó phỏng vấn bà mẹ trước khi khám cho đứa trẻ, rồi tiếp theo đó mới là những ghi nhận về điều kiện môi trường trong nhà ở.

Bước 6 cũng cần được nhấn mạnh, và một điểm rất ngạc nhiên thường gặp trong các cuộc điều tra là một số thông tin cơ bản không được hỏi tới hoặc không được ghi nhận lại. Những thiếu sót như vậy sẽ giới hạn rất nhiều tính hữu ích của nghiên cứu và vì vậy nó làm cho nghiên cứu không đạt được mục tiêu đề ra. Phương pháp tốt nhất để tránh tình trạng này là nên có một nhóm điền vào bản nhập bộ câu hỏi và phân tích một số bộ câu hỏi đã được điền đầy đủ, tương tự như công việc chúng ta thường phải làm ở giai đoạn cuối của người nghiên cứu. Việc làm này cho phép phát hiện ra những mục quan trọng đã bị bỏ qua khi thiết kế bộ câu hỏi.

Câu kiểm tra thật kỹ để đảm bảo rằng những thông tin hành chính phải được ghi nhận cẩn thận và chính xác như tên, họ, tuổi, giới, địa chỉ vì những thông tin này là đặc biệt quan trọng trong những nghiên cứu tương lai. Vì những nghiên cứu này đòi hỏi phải tìm lại được những đối tượng nghiên cứu để xem xem có sự kiện gì xảy ra hoặc thay đổi trên những đối tượng này so với lần tiếp xúc trước không. Một số nghiên cứu tương lai người ta thường thu thập một số dữ kiện để có thể dễ dàng tìm lại được đối tượng nghiên cứu ví dụ như: nghề nghiệp, tên, địa chỉ của họ hàng, bạn bè thân.

B. CẤU TRÚC CÂU HỎI

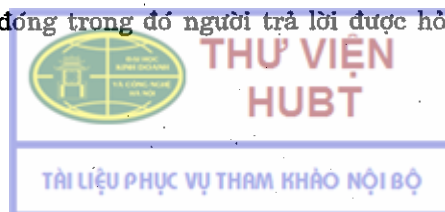
Những câu hỏi mở và đóng

Có hai cách cơ bản để xây dựng một câu hỏi điều tra. Một câu hỏi có thể là *đóng* hoặc *mở*.

1. Câu hỏi đóng.

Câu hỏi đóng đưa ra một danh sách những câu trả lời có thể xảy ra. Từ đó người trả lời phải chọn tối thiểu là một câu. Những câu trả lời không có trong danh sách đó là không được phép.

Sau đây là một ví dụ về câu hỏi đóng đó người trả lời được hỏi về 4 loại thức ăn.



Câu hỏi: Chúng tôi muốn biết xem anh đã ăn gì trong bữa chính ngày hôm qua. Những thứ đó có bao gồm:

- (1) Cá hoặc thịt có không
- (2) Trứng có không
- (3) Sữa hoặc fomát có không
- (4) Đậu, đỗ có không

2. Câu hỏi mở.

Câu hỏi mở đòi hỏi người trả lời trả lời bằng ngôn ngữ của mình. Đặc biệt, bộ câu hỏi không đưa ra một danh sách câu trả lời có thể có để người trả lời chọn lựa.

Câu hỏi đóng giới thiệu ở trên có thể đặt lại thành câu hỏi mở như sau:

Câu hỏi: chúng tôi muốn biết xem anh đã ăn gì trong bữa chính ngày hôm qua. Hãy liệt kê những thứ anh đã ăn trong bữa chính hôm qua ra.

Vì câu hỏi mở không liệt kê sẵn danh sách những câu trả lời, nên khi muốn nhấn mạnh tới khía cạnh nào đó đặc biệt quan tâm, câu hỏi mở có thể nêu lên điều quan tâm đó theo một nghĩa chung. Lại lấy lại ví dụ trên với quan tâm của ta lúc này là thức ăn protein.

Câu hỏi: Chúng tôi muốn biết anh đã ăn những gì trong bữa chính ngày hôm qua, và chúng có bao gồm những thức ăn như trứng hoặc thịt hay không, xin hãy mô tả bữa chính ngày hôm qua có gì? Chú ý rằng khi thiết kế câu hỏi mở, thì sau mỗi câu hỏi phải để đủ chỗ trống để điền câu trả lời. Một điều nữa cũng cần chú ý là những khó khăn có thể gặp phải khi phân tích/mã hóa câu hỏi mở ở giai đoạn cuối của nghiên cứu.

3. Nhắc lại những câu hỏi quan trọng.

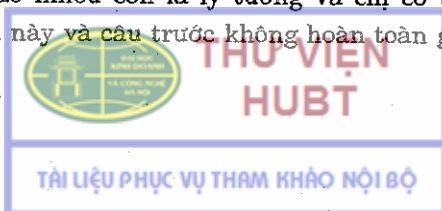
Nhiều cuộc điều tra sử dụng cả câu hỏi mở lẫn câu hỏi đóng. Nhiều câu hỏi đôi khi được nhắc lại, lúc thì dưới dạng đóng ở một điểm khác của cuộc điều tra, để kiểm tra sự nhất quán của người trả lời hoặc khi câu hỏi này là quan trọng trong cuộc điều tra. Những ví dụ về loại câu hỏi này thường thấy trong những nghiên cứu về dân số kế hoạch hóa gia đình hoặc sinh đẻ khi nó là câu hỏi quan trọng để xác định thái độ của những người vợ trẻ về kích thước gia đình. Sau đây là hai câu hỏi tương tự hỗ trợ nhau.

"Chị muốn sinh bao nhiêu cháu"

- (a) Không muốn sinh con
- (b) Chỉ 1 thôi
- (c) 2
- (d) 3-4
- (e) 5-6
- (f) trên 6
- (g) Không khẳng định/không biết.

Câu hỏi thứ hai (mở)

"Chị cho rằng có bao nhiêu con là lý tưởng và chị có thể cho tôi biết tại sao chị lại chọn như thế". Câu hỏi này và câu trước không hoàn toàn giống nhau nhưng mỗi câu đều



hỏi về thái độ của người vợ về kích thước gia đình mà họ mong muốn.

Câu hỏi sau kiểm tra xem liệu người trả lời có nghiêm túc và nhất quán trong việc trả lời về số con mong muốn hay không. Tuy nhiên, câu hỏi còn khai thác được nhiều thông tin hơn câu trước, nó tạo cơ hội cho người trả lời nói thoải mái về kích thước gia đình theo nghĩa chung là lý do tại sao chị ta lại muốn có kích thước đó. Câu hỏi thứ hai này phải được đặt ở phần khác của bộ câu hỏi và kế tiếp luôn với câu thứ nhất.

4. danh sách câu trả lời nên ngắn.

Số câu trả lời liệt kê ra cho các câu hỏi đóng không nên nhiều quá, thường không nên quá 8 câu nếu có thể được, bình thường người ta thường xây dựng danh sách khoảng từ 2 tới 6 câu. Nếu danh sách này quá dài, người trả lời thường quên một số câu đặc biệt là những câu ở giữa, danh sách này càng dài, người trả lời càng có khuynh hướng chọn những câu ở đầu hoặc cuối danh sách.

Trong câu hỏi đóng về kích thước gia đình, có 6 câu trả lời bắt đầu bằng "không có con nào" và kết thúc bằng "có trên 6 con" cộng thêm câu cuối cùng "không khẳng định/không biết" có thể đặt lại bằng cách khác như:

- (a) có hai hoặc ít hơn.
- (b) có 3 tới 5
- (c) có hơn 5
- (d) không khẳng định/ không biết.

Trong hai danh sách trả lời này thì danh sách nào tốt hơn? Thực sự thì không có luật nào để đánh giá cả. Trong ví dụ đặc biệt này, câu trả lời sẽ phụ thuộc vào:

(I) Mục đích của nghiên cứu: nếu mục đích chính của nghiên cứu là để phát hiện xem có bao nhiêu người muốn có một gia đình đông đúc (trên 5 con chẳng hạn) thì lúc đó danh sách ngắn hơn sẽ tốt hơn. Nếu nghiên cứu đòi hỏi chi tiết thái độ của gia đình thì danh sách chi tiết sẽ tốt hơn.

(II) Văn hóa/Tôn giáo của cộng đồng. Ở nhiều nước cặp vợ chồng nào cũng muốn có con vì vậy câu trả lời "không muốn có con" là không cần thiết.

(III) So sánh với những nghiên cứu khác. Nếu có nhu cầu so sánh những kết quả điều tra với những nghiên cứu khác thì điều quan trọng là giữ danh sách giống như của nghiên cứu kia. Nếu không giữ như vậy thì việc so sánh sẽ không thể tiến hành được.

5. Những ưu điểm của câu hỏi đóng.

- (I) Những câu trả lời có thể được đưa ra và ghi lại nhanh.
- (II) Người trả lời chú ý vào những điểm chủ yếu của cuộc điều tra.
- (III) Danh sách câu trả lời có những điểm quan trọng mà người trả lời có thể không nhớ mà kể ra.
- (IV) Phân tích dễ dàng hơn vì đã được mã hóa trước.

6. Những nhược điểm của câu hỏi đóng.

(I) Danh sách trả lời có thể không phù hợp với ý định người trả lời vì toàn bộ câu trả lời có thể không tương tự như điều người trả lời định diễn tả.

(II) Người trả lời có thể bị tác động khi trả lời vì người ta dùng cho chị ta một danh sách để chọn câu trả lời và chị ta có thể cảm thấy rằng người hỏi muốn chị ta trả lời như vậy.

7. Những ưu điểm của câu hỏi mở.

(I) Câu hỏi mở cho phép người trả lời diễn đạt theo kiểu riêng của mình và vì vậy chị ta không bị tác động. Điều này có thể dẫn tới sự thật trong câu trả lời hơn.

(II) Câu hỏi mở không giới hạn người trả lời vào những câu trả lời đặc biệt, chị ta có thể nói về những vấn đề không định trước, trong những nghiên cứu về thái độ, tâm lý và sức khỏe, người trả lời có nhiều cơ hội để phát biểu cởi mở và chi tiết về những điều quan trọng với chị ta và chị ta cảm nhận thế nào về những vấn đề của riêng mình.

8. những nhược điểm của câu hỏi mở.

(I) Câu trả lời của những câu hỏi mở rất dài nên người phỏng vấn sẽ mệt khi phải viết dài quá.

(II) Người trả lời có thể sẽ nói về những vấn đề không quan trọng cho nghiên cứu. Một người phỏng vấn có kinh nghiệm có thể lái câu chuyện trở về đề tài quan tâm nhưng tất nhiên là mất thời gian.

(III) Người trả lời có thể không nhớ, khi phỏng vấn toàn bộ những sự kiện tác động tới chị ta hoặc có tầm quan trọng với chị ta khi không có gì để gợi ý chị ta nhớ lại.

(IV) Câu hỏi mở có thể khó phân tích, mỗi người trả lời lại dùng từ ngữ riêng, theo cách diễn đạt riêng, đưa ra những thông tin về nhiều khía cạnh không tương ứng. Người trả lời cũng có thể đưa những thông tin khó hiểu và khó phân tích.

C. ĐỂ CÓ ĐƯỢC NHỮNG CÂU HỎI ĐÚNG

1. Những từ và câu trong câu hỏi.

Những luật sau đây sẽ giúp cho việc xây dựng các câu hỏi rõ ràng và không quá tham vọng (chỉ có 1 nghĩa).

(I) Hỏi ngắn, câu hỏi đặc hiệu.

(II) Sử dụng từ đơn giản, thường dùng hàng ngày.

(III) Mỗi lần hỏi chỉ hỏi một câu hỏi.

Mỗi câu hỏi phải hỏi và chỉ một khoản. Những câu hỏi muốn có nhiều câu trả lời đôi khi được gọi là những câu hỏi phức hợp. Những câu hỏi loại này thường gây nhầm lẫn cho cả người hỏi lẫn người trả lời. Ví dụ:

"Bạn có mấy con và mấy đứa đã đi học"

Câu này gồm hai câu hỏi riêng biệt. Trong câu hỏi này sẽ rất khó phân tích nếu như người trả lời đáp rằng "không". Liệu có thể phân tích là chị ta không có con hay không có con đi học.

Vì vậy câu này phải tách thành hai câu:

- Chị có mấy con?



- Máy con đang đi học?

Bây giờ hai câu hỏi phải trả lời tách riêng và không còn sự nhầm lẫn nào nữa.

Trong hội thoại hàng ngày, những câu hỏi phức tạp như vậy thường chấp nhận được vì trong quá trình hội thoại bất kỳ sự hiểu nhầm nào cũng đều có thể làm rõ bằng những câu hỏi sâu hơn, trả lời và thảo luận. Một cuộc điều tra không tạo được cơ hội như vậy vì, trong quá trình phỏng vấn điều quan trọng là phải đặt cùng những câu hỏi một cách giống nhau cho toàn bộ những người trả lời.

2. Những chỉ dẫn về nhảy cóc câu hỏi.

Ta hãy xem ví dụ sau:

Câu hỏi đầu: Chị có mấy con?

Câu hỏi hai: Máy con đang đi học?

Nếu câu hỏi đầu có câu trả lời là "không có" thì tất nhiên ta sẽ không hỏi câu thứ hai (nhảy cóc qua câu hai) mà hỏi tiếp các câu khác trong bộ câu hỏi. Như vậy, trong bộ câu hỏi ta sẽ thiết kế như sau:

Câu hỏi 1: Chị có mấy con? Nếu "không có", thì hỏi sang câu 3 (bỏ qua câu 2).

3. Câu trả lời ngược ngặc cho những câu hỏi ngược ngặc.

Có một câu ngạn ngữ Anh: "Hỏi một câu hỏi ngược ngặc, bạn sẽ nhận được một câu trả lời ngược ngặc". Điều này thực sự đúng trong những cuộc điều tra vì chúng ta thường không thực sự cẩn thận khi xây dựng bộ câu hỏi và người tổ chức cuộc điều tra không thực sự nhận thức được những khó khăn của người trả lời khi trả lời một số bộ câu hỏi.

Áp dụng những luật sau đây sẽ cho phép tránh được những câu hỏi dễ gây nhầm lẫn và ngược ngặc.

(I) Tránh những câu hỏi quá chung, chỉ dùng những câu đặc hiệu.

(II) Tránh những câu hỏi giả định (như những câu hỏi về những điều có thể xảy ra).

(III) Tránh hỏi những câu hỏi về tham khảo và so sánh trừ khi người trả lời có kinh nghiệm về điều đang được so sánh.

Điều quan trọng là hỏi những người trả lời về những điều họ đã biết và chúng có tác động tới công việc của họ, gia đình, sức khỏe và những quan tâm của họ.

Sau đây là những ví dụ về những câu hỏi trong đó những luật này không được áp dụng.

Ví dụ câu hỏi 1: "Có phải hầu hết mọi người anh biết đều thỏa mãn với những chăm sóc y tế ở trung tâm y tế của anh không?"

Câu hỏi này quá chung chung vì những lý do sau:

(I) Người trả lời có thể không nghĩ nhiều về những dịch vụ tại trung tâm y tế và vì vậy sẽ không có ý kiến rõ ràng về vấn đề này.

(II) Người trả lời và gia đình không có kinh nghiệm cá nhân về trung tâm y tế.

(III) Câu "hầu hết những người anh biết" là rất chung; Có bao nhiêu bạn và người quen của người trả lời đã thảo luận về những dịch vụ trung tâm y tế? nếu có, có thể rất ít.



Ví dụ câu hỏi 2: "Bạn sẽ làm gì nếu nhà bạn hoặc nhà hàng xóm bị cháy?"

Đây là một câu hỏi rất kém vì hai lý do:

(I) Đây là câu hỏi phức tạp gồm hai câu riêng rẽ.

- Hành động phải làm nếu bản thân nhà mình bị cháy và hành động phải làm nếu nhà hàng xóm bị cháy.

(II) Đây là câu hỏi giả định. Làm thế nào người ta lại có thể cho một câu trả lời và điều họ có thể làm trong một tình huống đáng sợ như vậy? Người trả lời chỉ có thể hình dung điều họ có thể làm trong những tình huống như vậy.

D. TRÌNH BÀY BỘ CÂU HỎI

Trình bày bộ câu hỏi liên quan tới khoảng cách và vị trí của câu hỏi và câu trả lời, hoặc là được mã hóa trước hoặc mở, và vị trí của những dòng chỉ dẫn người phỏng vấn.

Trình bày bộ câu hỏi rất quan trọng vì khi trình bày tốt thì bộ câu hỏi sẽ:

(I) Làm bộ câu hỏi dễ sử dụng trong quá trình phỏng vấn.

(II) Tăng cường chất lượng phỏng vấn vì chỉ dẫn chính xác.

(III) Ngăn ngừa nhầm lẫn.

(IV) Tránh quên câu hỏi, câu trả lời hoặc câu chỉ dẫn.

(V) Làm dễ dàng hơn cho người tổ chức điều tra trong giai đoạn phân tích.

Khoảng cách đủ để viết ra câu trả lời là đặc biệt quan trọng trong câu hỏi mở.

Câu chỉ dẫn và câu hỏi nên riêng biệt nhau để khỏi lẫn có thể bằng cách những câu chỉ dẫn in bằng kiểu chữ khác hoặc in đậm hơn để tránh nhầm lẫn, hoặc cũng có thể để trong ngoặc. Câu chỉ dẫn nên để cạnh câu cần thiết phải có câu chỉ dẫn đó. Trong những cuộc điều tra cần phải có nhiều câu chỉ dẫn thì trong trình bày bộ câu hỏi nên để lề rộng để chỉ ghi câu chỉ dẫn vào đó.

1. Phân chia thành từng phần.

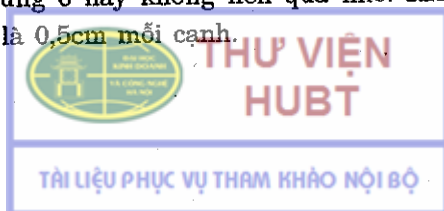
Nhiều bộ câu hỏi sẽ được chia ra làm nhiều phần, trong mỗi phần, những câu hỏi sẽ liên quan với một số khía cạnh đặc biệt của nghiên cứu như dinh dưỡng, loại nhà ở, lịch sử y tế, vv...

Bộ câu hỏi sẽ được sử dụng dễ hơn nếu.

(I) Giữa những phần khác nhau nên để khoảng cách.

(II) Các bắt đầu mỗi phần, nên có những câu mào đầu của người phỏng vấn để chuyển sang một phần khác. Hứng thú và quan tâm của người trả lời sẽ tăng lên nếu có một quãng nghỉ ngắn giữa lúc hỏi cũng như nếu ta nói với họ điều gì đó, có thể rất ít thời cũng được, về việc tại sao ta lại quan tâm tới những thông tin ta hỏi họ.

Trong nhiều bộ câu hỏi, những câu hỏi đóng thường có những ô vuông nhỏ bên cạnh những ý trả lời. Người phỏng vấn chỉ việc đánh dấu vào những ô tương ứng trong quá trình phỏng vấn. Những ô này không nên quá nhỏ. Kích thước thường dùng nhất của những ô này thường là 0,5cm mỗi cạnh.



Ví dụ về một câu hỏi về "một cuộc điều tra về có thai và chăm sóc trẻ sơ sinh". Bộ câu hỏi này mô tả nhiều điểm đã trình bày trong các phần trước.

1. Tính tới lúc sinh cháu chị có thai được mấy tuần?

Câu thăm dò: (i) Cháu đủ hay thiếu tháng?

(ii) Sớm hay muộn bao lâu?

Dưới 34 tuần

34 < 38 tuần

38 < 41 tuần

> 41 tuần

Thời gian có thai bình thường là 40 tuần

2. Chị sinh thế nào?

Bình thường

Forceps

Mô đẻ không dự định

Mô đẻ có dự định

Hút

Loại khác

3. Có vấn đề gì trước, trong và sau khi sinh không?

Không thực ra chẳng có vấn đề gì. Chuyển dạ kéo dài khoảng 2 tiếng, sau đó tôi rất mệt nhưng mọi thứ thực sự tiến triển rất tốt.

4. Cháu được mấy cân? Khoảng 2,5kg. (người trả lời hy vọng có một trẻ nặng hơn và con trai chứ không phải con gái)

5. Khi sinh ra cháu có vấn đề gì không?

Có

Không

Câu thăm dò: Có điều gì bất thường hoặc không hoàn toàn bình thường khi sinh cháu.

Nếu có, những vấn đề gì?

Chú ý rằng câu 1 là câu hỏi đóng với 4 câu trả lời riêng biệt và bao quát. Một ô vuông được đặt ngay cạnh mỗi câu trả lời; người điều tra sẽ đánh dấu vào ô tương xứng với câu trả lời của người trả lời. Những ô này có kích thước tương ứng.

Câu thăm dò hướng dẫn người phỏng vấn hỏi câu sâu hơn mà người trả lời có thể không chắc chắn trong quá trình mang thai.

(i) Hỏi người trả lời nhớ lại xem đứa trẻ sinh ra quá sớm hay hơi muộn.

(II) Nhằm khuyến khích câu trả lời đặc hiệu, chính xác hơn. Câu đóng trong khung

Có thai bình thường là 40 tuần

có hai chức năng.

(a) Nó nhắc người phỏng vấn một số thông tin quan trọng mà người này cần biết khi hỏi câu hỏi này.

(b) Nó cho phép người này kiểm tra xem liệu câu trả lời có hợp lý không.

Xem câu 1 sẽ thấy cách trình bày rất sáng sủa, dễ dàng cho người phỏng vấn khi điền vào các ô và vì vậy khi phân tích kết quả sẽ dễ dàng hơn.

Câu 2. Ngắn gọn và là câu hỏi trực tiếp, khó có thể nhầm lẫn được khi điền vào các ô và bản thân người trả lời cũng khó ỡ thể nhầm lẫn ý nghĩa từng câu trả lời.

Câu 3: bao gồm 2 câu hỏi 3(a) là câu hỏi đóng ngắn. Câu 3(b) là câu hỏi mở và có đủ khoảng cách để người hỏi có thể đủ chỗ để ghi toàn bộ câu trả lời.

Câu 4: Người phỏng vấn ghi "khoảng 2,5kg" và cạnh đó ghi thêm: Người trả lời hy vọng có con trai và có cân nặng hơn thế. Đó là những thông tin tốt để phân tích sau này.

Câu 5: Kèm theo bằng một câu hỏi điều kiện. Nếu câu trả lời là có khó khăn khi sinh thì người điều tra được chỉ dẫn tiếp tục hỏi về những khó khăn đó, và chỉ việc khoanh tròn vào những tình huống đã viết vào trong khung.

Câu hỏi 5 này là một câu hỏi phức tạp nhưng được trình bày và chỉ dẫn rõ ràng và nó nhắc người phỏng vấn từng bước phải theo.

Để cho cuộc điều tra phỏng vấn thành công, người trả lời phải:

(a) Hiểu cuộc điều tra là về vấn đề gì và tại sao lại hỏi chị/anh ta những câu hỏi đó.

(b) Được khuyến khích để tham gia vào cuộc phỏng vấn.

(c) Quan tâm trong suốt cuộc điều tra. Điều này có nghĩa là không nên hỏi quá nhiều câu hỏi, trong quá trình phỏng vấn nên cung cấp thêm cho chị/anh ta thêm thông tin để duy trì hào hứng và sẵn sàng giúp đỡ.

(d) Hào hứng với cuộc điều tra sao cho cuối cuộc điều tra chị/anh ta cảm thấy cuộc thảo luận là đáng giá, quan trọng và có ích lợi.

Vì vậy cuộc điều tra phỏng vấn phải được lên kế hoạch một cách tổng thể chứ không thể coi như chỉ là một loạt những câu hỏi riêng rẽ.

2. Quá trình phỏng vấn.

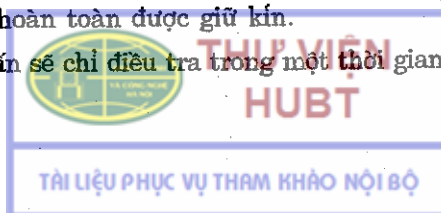
Thứ tự của việc cung cấp thông tin cho người trả lời, thứ tự từng phần (chủ đề) và thứ tự các câu hỏi có tầm quan trọng lớn nếu như muốn người trả lời thích thú với cuộc phỏng vấn. Thứ tự này khác nhau trong những cuộc điều tra khác nhau nhưng thông thường là như sau:

(I) Người phỏng vấn tự giới thiệu, và giải thích ngắn gọn về nghiên cứu.

(II) Người trả lời được đảm bảo rằng:

(a) Thông tin là hoàn toàn được giữ kín.

(b) Cuộc phỏng vấn sẽ chỉ điều tra trong một thời gian nào đó (15 - 20 phút chẳng hạn).



- (III) Hỏi về phần thủ tục hành chính như tên, tuổi, địa chỉ.
- (IV) Người phỏng vấn giới thiệu mỗi phần mới của bộ câu hỏi bằng những câu ngắn gọn nhằm:
- (a) Giới thiệu lý do của việc muốn thu thập thông tin về chủ đề tiếp theo.
 - (b) Cho người trả lời có thời gian nghỉ trong cuộc phỏng vấn.
 - (c) Tạo cho người trả lời chuyển sự chú ý từ phần cũ sang phần mới.
 - (v) Trong mỗi phần, những câu hỏi được xếp đặt theo một thứ tự logic và thường được phát triển từ đơn giản hơn, dễ trả lời hơn tới khó hơn.
 - (VI) Những câu hỏi về những sự kiện khó, gây bối rối cho người trả lời phải được đặt vào phần phù hợp của bộ câu hỏi.
 - (VII) Kết thúc cuộc phỏng vấn bằng những câu cảm ơn lịch sự gồm:
 - (a) Nhấn mạnh về tầm quan trọng của nghiên cứu
 - (b) Nhắc lại rằng thông tin sẽ được giữ bí mật.

E. NHỚ VÀ KẾ LẠI

1. Giúp cho người trả lời nhớ lại.

a) *Kế lại*: Nhớ lại những sự kiện trong quá khứ, là một vấn đề chủ yếu trong điều tra. Hầu hết những cuộc điều tra y tế đều hỏi những câu hỏi không chỉ về những hoạt động và tình trạng sức khỏe hiện tại mà cả về những sự kiện và điều kiện trong quá khứ.

Nhớ lại, kể lại quá khứ thường khó và vì vậy sai số (error) của việc kể lại này thường gặp. Những sai lầm của việc nhớ lại có thể có nhiều dạng khác nhau:

(I) Người trả lời thường không thu thập lại được toàn bộ những sự kiện hoặc những điều đã trải qua trong quá khứ vì họ hoàn toàn quên.

(II) Chỉ nhớ lại từng phần, họ chỉ nhớ lại những khía cạnh nào đó đã xảy ra chứ không phải toàn bộ.

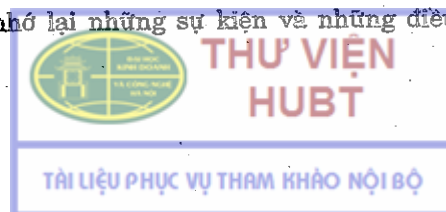
(III) Hai hoặc nhiều sự kiện trong quá khứ xa nhau được nối lại, đẩy liền vào nhau, vì vậy người trả lời nghĩ về một số sự kiện riêng biệt xảy ra ở những thời điểm khác nhau, như là một sự kiện độc nhất.

(IV) Những sự kiện này được nhớ lại chính xác, nhưng người trả lời lẫn về thời gian và địa điểm sự kiện xảy ra.

Nhiều yếu tố tác động tới trí nhớ nhưng sau đây là những lý do chung cho việc kể lại không chính xác:

- (I) Tuổi của người trả lời, người già thường hay quên hơn.
- (II) Thời gian trôi qua kể từ khi những sự kiện đã xảy ra.
- (III) Sự quan trọng của các sự kiện, hoặc nổi đau và mối nguy hiểm gây ra với cá nhân và gia đình anh ta.

Người trả lời có thể được giúp nhớ lại những sự kiện và những điều kiện trong quá khứ theo 3 cách:



- (a) Nhắc nhở
- (b) Thăm dò
- (c) Những câu hỏi mào đầu.

b) Nhắc nhở:

Người phỏng vấn nhắc nhở, ví dụ người này có thể đọc ra một danh sách có thể các câu trả lời từ đó người trả lời phải chọn. Những câu nhắc nhở trong câu hỏi số 2 và phần hai của câu 5 trong bộ câu hỏi ví dụ thuộc loại này.

c) Thăm dò:

Người phỏng vấn nhắc nhở, ví dụ cô ta hỏi những câu hỏi thêm điều đó giúp cho việc làm rõ ràng câu trả lời đối với câu hỏi chính, hai câu thăm dò trong câu hỏi 1 và phần đầu của câu 5 thuộc loại này.

d) Những câu hỏi mào đầu

Bộ câu hỏi có thể được thiết kế để đặt một loạt những câu hỏi ngắn, mỗi câu chuẩn bị cho người trả lời và dẫn cô ta tới câu hỏi quan trọng. Người phỏng vấn chỉ đọc những câu hỏi y nguyên như được in trong bộ câu hỏi; Không nhắc nhở hoặc thăm dò sâu hơn trừ khi người phỏng vấn được phép làm như vậy như phần câu hỏi mào đầu.

2. Những nghiên cứu kế lại và tương lai

Dựa trên kế lại theo trí nhớ hoặc trên những ghi chép lâm sàng thì không bao giờ thỏa mãn hoàn toàn, nhưng người ta thường tiến hành vì chúng thường là những phương pháp nhanh nhất và rẻ nhất đối với người nghiên cứu; Đôi khi nó lại là phương pháp duy nhất có thể thực hiện được. Tuy nhiên, biết được những nguy hiểm và những giới hạn gắn liền với những câu hỏi phụ thuộc vào trí nhớ thì nên lập kế hoạch cho một *nghiên cứu tương lai*. Nghiên cứu tương lai và cơ bản là một nghiên cứu dọc và một nghiên cứu so sánh trong đó những thông tin tương xứng thu thập ở lúc bắt đầu nghiên cứu và rồi được so sánh với số liệu tương tự được thu thập ở giai đoạn sau này trên cùng các cá thể đó. Khoảng cách giữa việc thu thập lần đầu và lần sau thường rất dài thường là một năm hoặc hơn. Những phương pháp thu thập thông tin cần thiết thay đổi, nhưng bao gồm những cuộc điều tra và ghi chép lâm sàng. Những yêu cầu cơ bản là:

(I) Cũng những trường hợp (bệnh nhân) được nghiên cứu ở cả hai thời điểm, vì vậy những thay đổi về những điều kiện hoặc hành vi của cá nhân đó có thể được ghi nhận lại.

(II) Toàn bộ, hoặc tối thiểu là một tỷ lệ cao những trường hợp được nghiên cứu chủ dẫn có thể còn tìm thấy và được nghiên cứu ở giai đoạn trong tương lai.

Để minh họa, hãy cân nhắc nhu cầu phát hiện một chương trình giáo dục cộng đồng về chăm sóc bà mẹ trẻ em, về cai sữa mẹ khi nào và như thế nào, v.v... đã thành công hay không. Về cơ bản có 3 cách để tiến hành nghiên cứu.

(a) Một cuộc điều tra phỏng vấn có thể được tiến hành trong số những người mẹ trẻ có hai hay trên hai con. Những người mẹ này có thể được hỏi về cách nuôi con hiện nay của họ và họ đã nuôi con trước đây thế nào

(b) Một điều tra phỏng vấn có thể được tiến hành trong số những người mẹ trẻ và những thông tin thu được, được so sánh với những ghi nhận lâm sàng của lần sinh và nuôi con lần trước.

(c) Một nghiên cứu tương lai có thể được thiết lập trong đó những câu hỏi đã được



lựa chọn phù hợp được hỏi và những ghi chép được giữ lại ở mỗi lần sinh: Nói cách khác, tiến hành một nghiên cứu tương lai phù hợp.

Những phương pháp (a) và (b) sẽ dựa chủ yếu vào trí nhớ của các bà mẹ và kể lại. Nhiều bà mẹ sẽ không nhớ rõ ràng trọng lượng khi sinh của những đứa con trước, độ dài của thời kỳ thai nghén, tuổi khi cai sữa và loại thức ăn trong thời kỳ cai sữa. Khi những ghi chép lâm sàng còn chúng có thể giúp thu được một số thông tin trước đó, có thể là không nhiều. Nhưng ghi chép lâm sàng được giữ để chăm sóc bệnh nhân, chẩn đoán và điều trị và trong các nhu cầu khác về mặt hành chính, chứ không phải để đáp ứng những câu hỏi nghiên cứu. Vì vậy, những ghi chép lâm sàng, thậm chí cả khi tìm thấy, cũng có thể rất giới hạn khi sử dụng.

Phương pháp thứ 3, nghiên cứu tương lai, là lý tưởng, vì nó đặt ít tin tưởng lên trí nhớ người mẹ. Một nghiên cứu tương lai yêu cầu:

(I) Kế hoạch phải dài.

(II) Các nguồn lực để theo dõi phải đủ.

(III) Thông tin đầy đủ phải được ghi nhận ở lần phỏng vấn đầu tiên hoặc kiểm tra cho phép người tổ chức toàn bộ, hoặc hầu như toàn bộ các trường hợp sau đó, việc theo dõi có thể là 1 năm hoặc hơn nhưng thông thường là những khoảng cách cố định sau lần phỏng vấn ban đầu.

F. MỘT SỐ QUAN NIỆM CHUNG

1. Việc sử dụng hình ảnh.

Việc sử dụng tranh, ảnh hoặc hình vẽ có thể rất có ích trong một cuộc điều tra. người điều tra được hướng dẫn chỉ một bức tranh đặc biệt cho người trả lời ở một giai đoạn đặc biệt của cuộc phỏng vấn và sau đó hỏi người trả lời về bức tranh đó. Bức tranh này phải có kích thước phù hợp và chất lượng tranh phải tốt. Tranh màu thường tốt hơn. Kích thước và chất lượng có tầm quan trọng vì không phải tất cả người trả lời đều nhìn tốt như nhau.

Phương pháp này đặc biệt hữu ích khi các câu hỏi hỏi về những tình huống khó mô tả thành lời. Mục đích của kiểu hỏi này là:

(I) Thử xem hiểu biết của người trả lời về sức khỏe và bệnh tật.

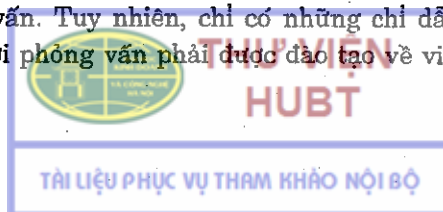
(II) Tìm hiểu thái độ và ý kiến người trả lời liên hệ với những tình huống nhất định như đã chỉ ra trong những bức tranh này.

(III) Tìm hiểu sở thích của người trả lời giữa những lựa chọn sẵn có trong bức tranh.

Một ví dụ về thử nghiệm sử dụng tranh, ảnh là ảnh của những bệnh nhân có những triệu chứng điển hình và phát triển thiếu hụt dinh dưỡng chung như Kwashiorkor. Sau đó câu hỏi kèm theo xem người trả lời có nhận ra là bệnh gì không? Cái gì gây ra bệnh đó? Làm gì để phòng bệnh đó?

2. Giúp đỡ người điều tra.

Một bộ câu hỏi được trình bày tốt, chỉ dẫn rõ ràng và chính xác sẽ giúp người phỏng vấn rất nhiều trong quá trình phỏng vấn. Tuy nhiên, chỉ có những chỉ dẫn in trong bộ câu hỏi thì chưa đủ, tất cả những người phỏng vấn phải được đào tạo về việc sử dụng bộ



câu hỏi trước khi tiến hành điều tra. Họ cần được dạy cách tiếp cận người trả lời, cách đọc câu hỏi, nói gì lúc bắt đầu và lúc kết thúc cuộc phỏng vấn. Họ cũng cần được hướng dẫn về cách thực hiện, nói gì khi người trả lời không sẵn lòng tham gia hoặc tiếp tục cuộc phỏng vấn. Nhiều cuộc phỏng vấn được tiến hành tốt nhất chỉ bằng cách chỉ hỏi người trả lời mà không có sự xuất hiện của bất cứ người thứ ba nào. Bố trí điều này có thể khó khăn nhưng người điều tra có thể được hướng dẫn một vài cách để đạt được điều này. Những tình huống có thể gặp trong cuộc điều tra cũng cần được hướng dẫn cho điều tra viên.

3. Những nghiên cứu thử (pilot studies).

Một bộ câu hỏi không nên được sử dụng nếu không được thử ít nhất là một lần qua những nghiên cứu thử. Một nghiên cứu thử sẽ luôn bộc lộ những khó khăn về mặt thực hành trong khi trình bày cũng như dùng những từ có nhiều nghĩa trong các câu hỏi cũng như câu chỉ dẫn. Những nghiên cứu thử chỉ khoảng 10-15 đối tượng sẽ cho thấy ngay những khó khăn không thấy trước khi xây dựng bộ câu hỏi. Người trả lời sẽ trả lời những câu không đúng dự đoán và như vậy sẽ phải sửa đổi bộ câu hỏi.

Một bộ câu hỏi sẽ luôn cần được thảo ra và thử và rồi lại thảo lại. Trong một cuộc điều tra phức tạp, bộ câu hỏi có thể cần được thử lại bốn tới năm lần trước khi dùng. Người điều tra phải tham gia vào những nghiên cứu thử này nếu có thể được. Việc thử này giúp cho những người lập kế hoạch cho cuộc điều tra trở nên nhận thức được những sai sót trong thiết kế bộ câu hỏi.

Nghiên cứu thử nhỏ này cũng sẽ chỉ ra cuộc phỏng vấn sẽ kéo dài bao lâu, cần bao lâu để đi từ cuộc phỏng vấn này sang cuộc phỏng vấn khác. Những thông tin này là vô giá trong việc quyết định xem cần bao nhiêu điều tra viên trong thời gian điều tra. Không có thông tin này, thì không thể ước lượng phù hợp những nguồn lực cần thiết để tiến hành cuộc điều tra.

Nghiên cứu thử cũng sẽ cung cấp những thông tin và liệu người trả lời có mệt mỏi trong quá trình phỏng vấn không. Người trả lời thường mệt, và không chú ý nghe vì một vài lý do như cuộc phỏng vấn quá dài, những câu hỏi không lý thú, người phỏng vấn không cung cấp cho họ đủ thông tin để thu hút sự chú ý của họ. Chú ý tới những phản ứng không thuận lợi của người trả lời trong các nghiên cứu thử sẽ cho những người tổ chức điều tra cơ hội để tìm ra tại sao người trả lời lại phản ứng không thuận lợi để cải thiện việc phỏng vấn, phương pháp điều tra chung và bộ câu hỏi.

THU THẬP VÀ TRÌNH BÀY KẾT QUẢ NGHIÊN CỨU

I. MỤC TIÊU

- Nêu và giải thích ba giai đoạn cơ bản liên quan đến việc đánh giá số liệu dịch tễ học.



- Sử dụng được các nguồn gốc thu thập số liệu dịch tễ học.
- Nêu được các nguyên lý cấu trúc một bảng số liệu.
- Giải thích, xây dựng và sử dụng được các bảng, biểu sau:
 - + Bảng tần số (tuyệt đối, tương đối và lũy tích)
 - + Bảng hai chiều và nhiều chiều.
 - + Đồ thị hình cột.
 - + Đa giác tần số.
 - + Biểu đồ hình chữ nhật.
 - + Biểu đồ hình quạt.
 - + Đồ thị tần số lũy tích.
 - + Bản đồ.
 - + Biểu đồ hình chấm.
- Trình bày một tập hợp số liệu vào một bảng thích hợp.
- Vẽ đồ thị và biểu đồ thích hợp từ một bảng số liệu.

II. NỘI DUNG

Có ba giai đoạn liên quan tới đánh giá các số liệu dịch tễ học thu thập từ các nguồn khác nhau.

Thu thập số liệu (Trên thực địa hay từ bất kỳ một cơ sở y tế nào) một cách chính xác tối đa và ghi nhận lại kết quả. Đây là khâu quan trọng nhất trong quá trình thống kê học, mọi thử nghiệm thống kê không bù được sự thiếu khoa học của việc chọn mẫu và thu thập số liệu. Chỉ khi nào khâu này làm tốt thì việc phân tích thống kê mới có giá trị khoa học.

Tập hợp và trình bày số liệu.

Số liệu được tập hợp và trình bày theo cách dễ hiểu và đơn giản nhất, thường là các bảng, biểu đồ và đồ thị.

Phân tích số liệu và rút ra những kết luận

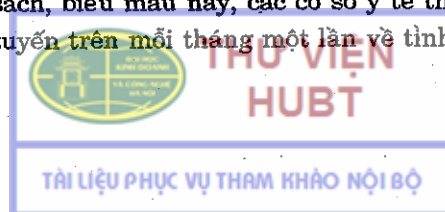
Giai đoạn này đòi hỏi phải sử dụng những kỹ thuật phân tích khác nhau và sử dụng khái niệm xác suất trong việc rút ra những kết luận.

1. Thu thập số liệu dịch tễ học.

Các thông tin hay số liệu dịch tễ học thông thường được thu thập từ:

a) Hệ thống thu thập số liệu sự vụ thường xuyên, định kỳ theo nội dung khoảng cách giữa các kỳ báo cáo, theo các biểu mẫu đã quy định của Bộ y tế.

Hiện nay theo quy định của Bộ y tế, và các chương trình y tế; Chúng ta có các sổ sách, biểu mẫu cho các cơ sở y tế (trạm y tế xã, các trung tâm y tế huyện, v.v...) để ghi nhận các thông tin sức khỏe. Theo các sổ sách, biểu mẫu này, các cơ sở y tế thường xuyên ghi chép theo hướng dẫn và báo cáo lên tuyến trên mỗi tháng một lần về tình hình sức khỏe,



bệnh tật ở tuyến mình quản lý.

Ví dụ:

- Biểu mẫu thống kê tình hình truyền nhiễm dịch tễ.
- Biểu mẫu báo cáo thống kê tiêm phòng đại.
- Biểu mẫu báo cáo bệnh hô hấp cấp tính ở trẻ em.
- Biểu mẫu báo cáo bệnh sốt rét: (ARI)

b) Từ các nghiên cứu điều tra quần thể hay từ các điều tra vụ dịch. Điều tra dịch tễ học để thu thập thông tin không có sẵn ở các cơ sở hệ thống thu thập số liệu thường xuyên, có thể là các cuộc điều tra nghiên cứu chuyên biệt hoặc sự xem xét lại kỹ lưỡng những số liệu đã thu thập được từ hệ thống thu thập số liệu thường xuyên. Số liệu thu thập từ các cuộc điều tra này có thể phục vụ cho những mục đích hành chính hoặc mục đích nghiên cứu.

Ví dụ:

- Một cuộc điều tra nhân lực y tế quốc gia.
- Một cuộc điều tra đánh giá tỷ lệ trẻ em suy dinh dưỡng ở một quần thể dân cư nhất định, v.v...

2. Trình bày số liệu.

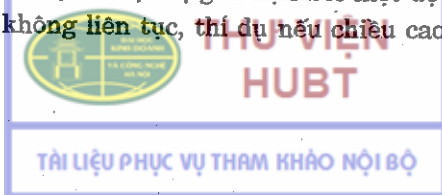
a) Các loại biến số.

Trình bày và tóm tắt số liệu trước tiên yêu cầu phải hiểu được các loại biến số thường gặp trong nghiên cứu dịch tễ học, và cách đo lường những biến số này.

Có hai loại cơ bản: Biến rời rạc và biến liên tục.

- Biến rời rạc, đó là các số đo chỉ là những số nguyên, không có số lẻ trung gian. Có biến rời rạc 2 bậc bao gồm giống (Nam hoặc nữ), sống hoặc chết, và tình trạng phơi nhiễm (có phơi nhiễm hoặc không phơi nhiễm). Các biến rời rạc nhiều bậc như nhóm máu (A, B, AB, O), màu da (trắng, đen, vàng). Tình trạng hôn nhân (chưa có gia đình, có vợ hoặc có chồng, ly dị, góa, ly thân),... Các giai đoạn của bệnh (giai đoạn I đến IV đối với ung thư vú), mức hút thuốc lá (không hút, hút ít, hút vừa phải, nghiện). Cuối cùng, một dạng biến rời rạc phức tạp nhất gọi là biến rời rạc lấy giá trị bằng số, những biến loại này không phải là biến định tính, như ở các ví dụ trên mà nó là biến định lượng (đặc tính định lượng không liên tục) như số lần đẻ của một bà mẹ, số con trai của một gia đình. Phân vào biến rời rạc, vì mặc dù giá trị là số, nhưng các giá trị của đặc tính chỉ được biểu thị bằng những con số nguyên, không lấy được bất kỳ số lẻ trung gian nào (nghĩa là không thể có một phân số của số lần đẻ của một bà mẹ).

- Biến liên tục. Là những số đo mà ngoài những phần nguyên, người ta có quyền lấy đến các số lẻ bao nhiêu cũng được. Ví dụ: Chiều cao, cân nặng, huyết áp, lượng cholesterol trong máu, nhiệt độ. Đối với chiều cao chẳng hạn, trong khoảng từ 152 cm đến 154cm, ta có thể gặp các trị số 152,1cm, 152,2cm, 153,3cm.... 153,8cm, 153,9cm, và trong khoảng rất bé từ 152,1cm đến 152,2cm, ta vẫn có thể gặp các trị số 152,11cm, 152,12cm... Tuy nhiên, trong thực tế do sự đo đạc bị giới hạn bởi một độ chính xác nhất định, nếu ta vẫn chỉ có những giá trị không liên tục, thì dù nếu chiều cao được đo với độ chính xác 0,5cm



thì trong khoảng từ 152 đến 153cm chỉ có 3 trị số: 152,5cm, 153cm và 153,5 cm.

- Sự phân loại thành những biến liên tục hay những biến rời rạc cho ta một số rất tiện lợi trong việc trình bày và *phân tích số liệu*. Đối với những biến rời rạc hoặc là định tính hoặc là định lượng, cơ sở phân tích là sự so sánh tỷ lệ những đối tượng ở những loại khác nhau. Như vậy, câu hỏi quan tâm đến có thể là liệu tỷ lệ những bệnh nhân bị ung thư máu sống được 5 năm do nhận phương pháp điều trị A có khác gì với những tỷ lệ bệnh nhân ung thư máu sống được 5 năm do nhận phương pháp điều trị B không? (so sánh tác dụng của phương pháp điều trị), liệu những người nhóm máu O có thể bị bệnh loét hành tá tràng nhiều hơn những người nhóm máu A, hoặc tỷ lệ những người nam giới có triệu chứng đường hô hấp ở những người nghiện thuốc lá có nhiều hơn ở những người không hút thuốc lá không?

- Đối với những biến liên tục, các nhóm thường được so sánh với nhau về những giá trị trung bình của các biến số. Ví dụ: Cân nặng trung bình của những trẻ sinh từ những bà mẹ hút thuốc lá so với cân nặng trung bình của những trẻ sinh từ những bà mẹ không hút thuốc lá. Tương tự, trong một thiết kế thử nghiệm, kiểm tra hiệu lực một loại thuốc trong việc làm giảm cho cholesterol máu, thì lượng cholesterol trung bình ở trong nhóm được điều trị được so sánh với nhóm không được điều trị hoặc placebo.

b) Giảm ước số liệu.

Một số biến vì bản chất của chúng, nếu chỉ phân loại được theo một cách ví dụ như giới của các đối tượng nghiên cứu phân loại: Nam hoặc nữ, nhưng có nhiều biến có thể được phân loại theo nhiều cách như tuổi, một người 58 1/2 tuổi có thể được phân loại: 55 - 59 tuổi, trên 50 tuổi hoặc tuổi trung niên. Mức chi tiết hoặc độ chính xác cần thiết cho phân tích sẽ phụ thuộc vào mục tiêu nghiên cứu. Ví dụ, trong đánh giá những yếu tố nguy cơ đối với gãy xương hông, việc gộp lại từng 10 tuổi một là cần thiết vì gãy xương này rất hiếm xảy ra ở dưới 40 tuổi và tăng lên ở tuổi trung niên, đặc biệt rất hay gặp ở những người trên 70 tuổi, nên tuổi được gộp thành: dưới 40 tuổi, 40 - 50, 50 - 60, 60-70 và trên 70 tuổi. Ngược lại khi so sánh tỷ lệ bệnh mạch vành ở những phụ nữ trước và sau mãn kinh thì tuổi lại chỉ phân thành 2 lớp, dưới 54 tuổi hoặc trên 54 tuổi, tuổi mãn kinh trung bình ở Mỹ. Thậm chí, trong cùng một nghiên cứu, việc phân loại tuổi để đánh giá một giả thuyết có thể không giống nhau khi yêu cầu đánh giá khác nhau. Ví dụ, trong một nghiên cứu về tuổi của bà mẹ và nguy cơ sinh đẻ khác nhau, nếu đánh giá hội chứng Down thì tuổi của bà mẹ có thể phân thành: dưới 30, 30-34, 35-39 và trên 40 tuổi là thích hợp vì thực tế lâm sàng hội chứng này hiếm gặp trong những trẻ em được sinh ra từ những bà mẹ trẻ nhưng lại hay gặp ở những trẻ sinh ra từ những bà mẹ trên 30 tuổi nhưng sự phân loại tuổi này sẽ không thích hợp khi đánh giá những tai biến trong sản khoa vì những bà mẹ trẻ (tuổi 15-16) có nguy cơ đặc biệt mà sự phân loại tuổi như trên có thể sẽ không nhận biết được.

Điều quan trọng luôn luôn phải nhớ rằng, việc giảm ước các số liệu thì thường làm cho việc trình bày số liệu dễ dàng hơn và có thể làm đơn giản sự phân tích thống kê sau đó, nhưng luôn luôn có sự kèm theo là sự mất chi tiết từ những sự quan sát ban đầu, mà

điều này có thể che giấu những thông tin hoặc những chiều hướng quan trọng. Hơn nữa, trong khi số liệu luôn có thể được giản ước thành dạng đơn giản hơn thì việc xây dựng lại những thông tin chi tiết hơn từ số liệu đã được ghi nhận đầu tiên chỉ trong sự phân loại thì đó là điều không bao giờ làm được. Ví dụ, trong một nghiên cứu về bệnh mạch vành, đánh giá những nguy cơ kết hợp với sự thay đổi những lượng cholesterol máu toàn phần là sẽ không có khả năng nếu những giá trị cholesterol đã được ghi nhận từ ban đầu chỉ là 2 loại bình thường hoặc cao. Tuy nhiên nếu, những mức (lượng) cholesterol thực sự đã được ghi nhận sự so sánh phân tích tiếp theo những cá nhân có mức cao hoặc thấp có thể dễ dàng làm được. Bởi vậy điều quan trọng xem xét cẩn thận tất cả những mục tiêu nghiên cứu trước khi bắt đầu thu thập số liệu, sao cho mức chi tiết cần thiết có thể đạt được. Nếu cần thiết hoặc mong muốn, giản ước số liệu khi đó có thể được làm trong phân tích của nghiên cứu.

c) Trình bày số liệu.

Sau giai đoạn thu thập số liệu, chúng ta có một đám số liệu có khi rất lớn, thường là hỗn độn, nếu công việc đầu tiên là phải sắp xếp lại một cách hợp lý, rõ ràng, và giữ nguyên được những thông tin gốc, càng nhiều càng tốt. Do vậy các số liệu dịch tễ học thu thập từ các nguồn phải được nhật ra và rồi tóm tắt lại dưới dạng các bảng số liệu, từ các bảng ta vẽ các đồ thị hoặc biểu đồ.

Xây dựng một bảng số liệu.

Bảng số liệu là một bảng trong đó các số liệu được sắp xếp theo các hàng và các cột. Tác dụng của một bảng là biểu thị tần số xuất hiện các sự kiện ở các nhóm hay thứ nhóm khác nhau đối với một biến số đáng xét. Hầu hết các thông tin về lượng có thể được trình bày vào bảng, và đây là phương pháp thống kê tần số phổ biến nhất. Chính từ các bảng số liệu này ta mới có thể tiến hành vẽ đồ thị và các biểu đồ.

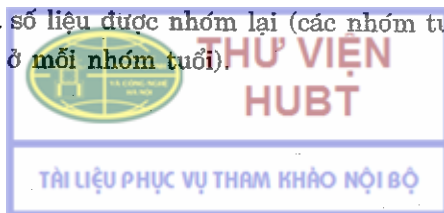
Các bảng phải càng đơn giản càng tốt. Thường thì hai hay ba bảng nhỏ thích hợp hơn một bảng lớn chứa đựng tất cả các chi tiết cho các bảng trên.

Nói chung, trong một bảng số liệu số tối đa có thể dễ dàng đọc và phân tích được là 3 biến số.

Tất cả các bảng phải có:

- Tên của bảng phải rõ, chính xác, đầy đủ, nó phải trả lời được các câu hỏi: cái gì, khi nào và ở đâu.
- Tên của mỗi hàng, mỗi cột phải chính xác và rõ ràng. Thêm vào đó phải ghi các đơn vị đo lường, các cột phải được tách biệt nhau bằng các đường kẻ đứng.
- Phải ghi tổng số cột và tổng số hàng.
- Các thuật ngữ viết tắt, mã hóa phải được giải thích ở phần ghi chú bên dưới.
- Nguồn gốc các dữ kiện cũng phải được ghi dưới bảng.

Dạng bảng phân phối tần số đơn giản nhất là bảng chỉ có 2 cột (một biến). Cột thứ nhất liệt kê các lớp mà số liệu được nhóm lại (các nhóm tuổi). Cột thứ 2 ghi tần số cho mỗi lớp (số trường hợp ở mỗi nhóm tuổi).



Ví dụ: Phân bố tuổi của mẫu nghiên cứu 100 người dân.

Nhóm tuổi (năm)	Số lượng
0-4	19
5-14	25
15-44	40
45 ⁺	16
Tổng cộng	100

Vì bảng chỉ có một biến, đó là tuổi, ta gọi bảng đó là *bảng một chiều*. Trong ví dụ này tuổi được xử lý như là một chỉ số đo lường rời rạc.

Ví dụ: về *bảng 2 chiều*

Phân bố tuổi - Giới của mẫu nghiên cứu 100 người dân.

Nhóm tuổi (năm)	Nam	Nữ	Tổng cộng
0-4	10	9	19
5-14	12	13	25
15-44	20	20	40
45 ⁺	7	9	16
Cộng	49	51	100

Các bảng hai hay nhiều chiều bao gồm việc dùng đến ít nhất là 2 biến trở lên. Dạng đơn giản nhất là bảng 2 x 2, bao gồm 2 hàng và 2 cột (không kể cột tổng hàng và tổng cột). Bảng 9-2 cho ví dụ về một bảng 2 x 2, người ta thường sử dụng bảng này khi cả 2 biến đều được phân loại thành "có" hoặc "không", như khi trình bày những số liệu rời rạc, đối với những biến 2 bậc: "có bệnh/không có bệnh" và "có phơi nhiễm/không phơi nhiễm".

Bảng 9-2. BẢNG 2 X 2 TÓM TẮT SỐ LIỆU TỪ MỘT NGHIÊN CỨU BỆNH CHỨNG VỀ DÙNG THUỐC TRÁNH THAI VÀ UNG THƯ VÚ

Dùng thuốc tránh thai	Nhóm bệnh	Ung thư vú nhóm chứng	Tổng số
Có	273	2641	2914
Không	716	7260	7976
Tổng số	989	9901	10.890

- Bảng phân phối tần số được sử dụng để trình bày những biến rời rạc, như bảng 9-2; Đồng thời bảng này cũng được dùng để tóm tắt những biến liên tục về lượng. Trong thực

tế khi số liệu tăng lên cả về số lượng đối tượng nghiên cứu và cả khối lượng thông tin ghi nhận, thì cần phải phân nhóm các giá trị và ghi nhận số lượng quan sát được vào mỗi nhóm. Ví dụ bảng 9-3. Cột 1 là số liệu đã được phân lớp, cột 2 ghi số lượng trong mỗi lớp. Như vậy 2 cột đầu của bảng cho ta thấy sự phân phối tần số axit uric trong huyết thanh của 267 nam giới cho máu. Chia số lượng trong mỗi lớp cho tổng số (267) ta được tần số tương đối (còn gọi là tần suất), như chỉ ra ở cột 3 của bảng, được biểu thị bằng tỷ lệ %.

Bảng 9-3. SỰ PHÂN PHỐI LƯỢNG AXIT URIC TRONG HUYẾT THANH
(267 nam giới khỏe mạnh cho máu)

Axit uric (mg/100ml)	Tần số tuyệt đối	Tần số tương đối (%)	Tần số lũy tích (% lũy tích)
3,0-3,4	2	0,8	0,8
3,5-3,9	15	5,6	6,4
4,0-4,4	33	12,4	18,7
4,5-4,9	40	15,0	33,7
5,0-5,4	54	20,2	53,9
5,5-5,9	47	17,6	71,5
6,0-6,4	38	14,2	85,8
6,5-6,9	16	6,0	91,8
7,0-7,4	15	5,6	97,4
7,5-7,9	3	1,1	98,5
8,0-8,4	1	0,4	98,9
8,5-8,9	3	1,1	100,0
Tổng số	267	100,0	

Cột 4 cho ta thấy sự phân phối tần số lũy tích. Cột này cho phép ta xác định tỷ lệ cá thể có lượng axit uric trong huyết thanh trên hoặc dưới một giới hạn nhất định nào đó. Ví dụ: Chúng ta có thể nói rằng khoảng 92% số người trong bảng 9-3 có lượng axit uric dưới 7mg/100ml.

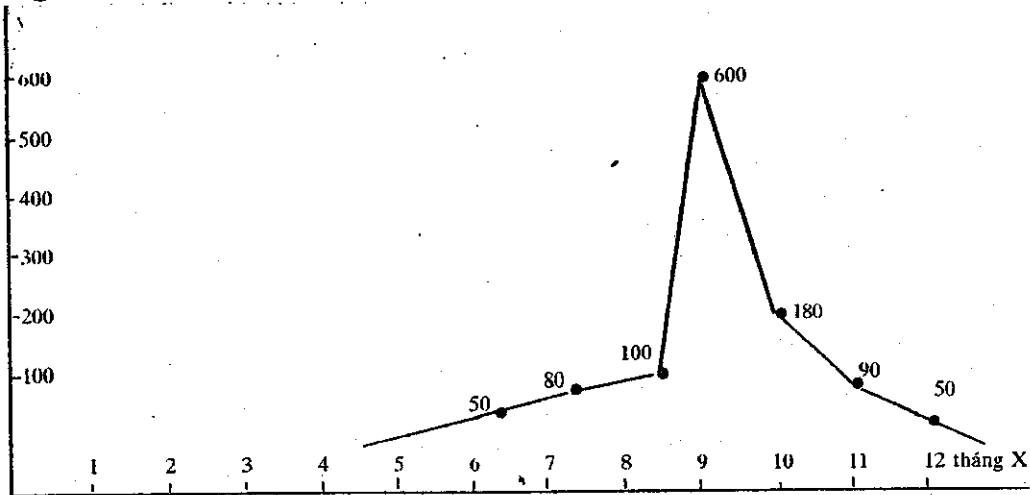
Xây dựng đồ thị.

Vẽ đồ thị là một phương pháp biểu thị các giá trị của một đặc tính định lượng bằng việc sử dụng một hệ thống tọa độ. Có nhiều loại đồ thị khác nhau, loại phổ biến nhất và thường sử dụng nhiều nhất là đồ thị số học.

+ Đồ thị số học (đồ thị thường).

Đồ thị số học là loại gồm 2 đường thẳng vuông góc với nhau. Đường nằm ngang thường được ký hiệu là X; Trục này biểu thị cho biến độc lập (phương pháp phân loại, chẳng hạn như thời gian hay tuổi), ngược lại biến phụ thuộc được biểu thị trên trục đứng, thường được ký hiệu là Y (tần số lưu hành của một sự kiện, chẳng hạn như số mới mắc, tỷ lệ % v.v...). Một dạng đồ thị số học hay được dùng để biểu thị sự lưu hành của bệnh theo thời gian là đồ thị dạng đoạn thẳng.

Ví dụ: Phân bố tần số bệnh Dengue xuất huyết ở quận Đống Đa, Hà Nội qua các tháng trong năm 1977.



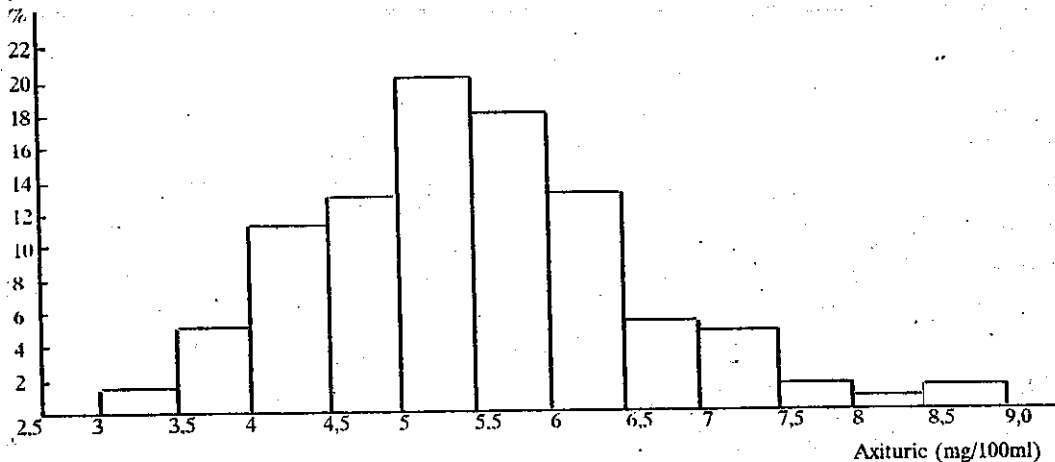
+ Đồ thị hình cột.

Đồ thị hình cột là một loại đồ thị minh họa sự phân bố tần số, nét đặc biệt của nó là:

- Chiều rộng của cột đứng tỷ lệ với chiều rộng của các khoảng cách lớp được sử dụng.
- Chiều cao của cột đứng ở một khoảng cách lớp tỷ lệ với tần số xảy ra (lưu hành) của sự kiện ở khoảng cách lớp đó. Do đặc điểm này, nếu loại giảm đồ hình cột thường được xây dựng theo cách các khoảng cách lớp bằng nhau.

Và cũng vì thế mà diện tích một cột chiếm phải bằng đúng số lượng các trường hợp ở trên trục đứng.

Ví dụ:



Đồ thị phân phối axit uric trong huyết thanh của 267 nam khỏe mạnh.

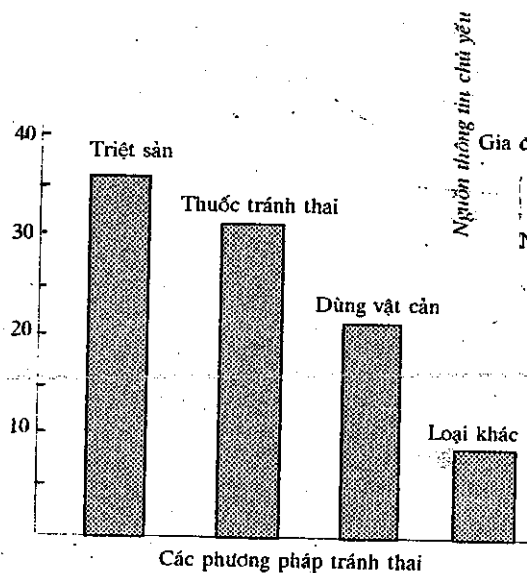
+ Biểu đồ

Biểu đồ là các phương pháp để minh họa thông tin về thống kê chỉ sử dụng một trục tọa độ. Loại này đặc biệt thích hợp cho việc so sánh độ lớn của các sự kiện khác nhau

hoặc các thành phần trong một tổng số.

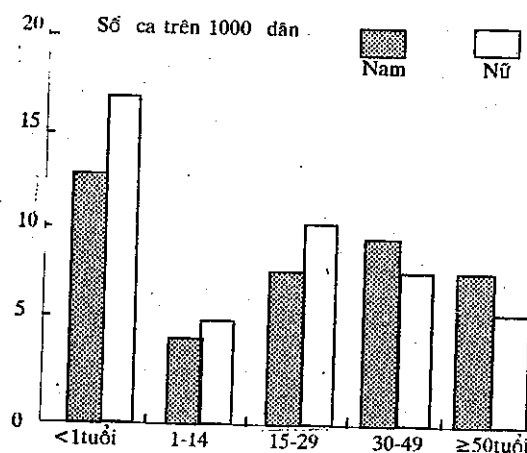
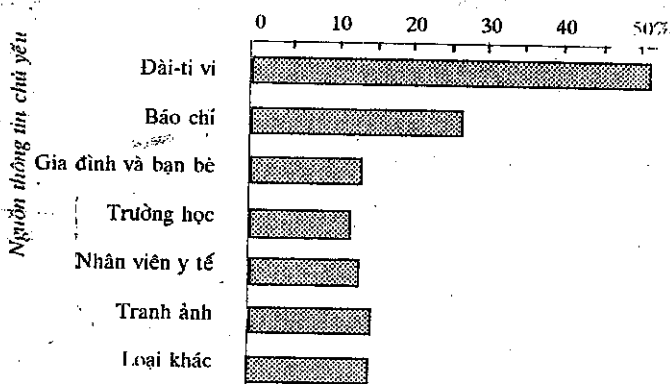
- *Biểu đồ cột chữ nhật.*

Loại biểu đồ hình chữ nhật có nhiều cột, tất cả các cột có chiều rộng như nhau (chiều rộng ta chọn tùy ý) cách biệt nhau bởi các khoảng cách nhất định. Một cột lại có thể có một hoặc nhiều ngăn hình chữ nhật, và trong trường hợp này, tốt nhất giữa các ngăn phải được tách biệt nhau một khoảng cách và phải được minh họa riêng biệt. Sự phân biệt phải được nêu trong chú giải. Chiều dài của mỗi cột tỷ lệ với tần số xảy ra của sự kiện trong bảng số liệu. Các hình chữ nhật có thể được vẽ nằm ngang hay thẳng đứng và người ta có thể biểu diễn các hình chữ nhật đó theo thứ tự tăng hay giảm để dễ phân tích. Những hình chữ nhật có thể để trống, hay có những nét gạch, hay tô đậm để xác định rõ hơn những nhóm khác nhau trên trục X. Ví dụ:



- Biểu đồ cột chữ nhật minh họa sự phân bố các cặp vợ chồng sử dụng các biện pháp tránh thai.

- Biểu đồ cột chữ nhật biểu diễn nguồn thông tin y tế chủ yếu mà người dân tiếp nhận được

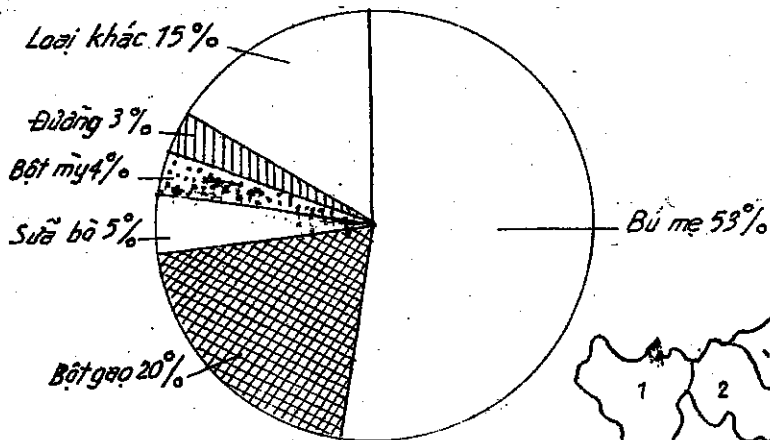


- Biểu đồ cột chữ nhật minh họa tỷ lệ tấn công theo tuổi và giới trong vụ dịch sốt xuất huyết ở Zaia năm 1978.

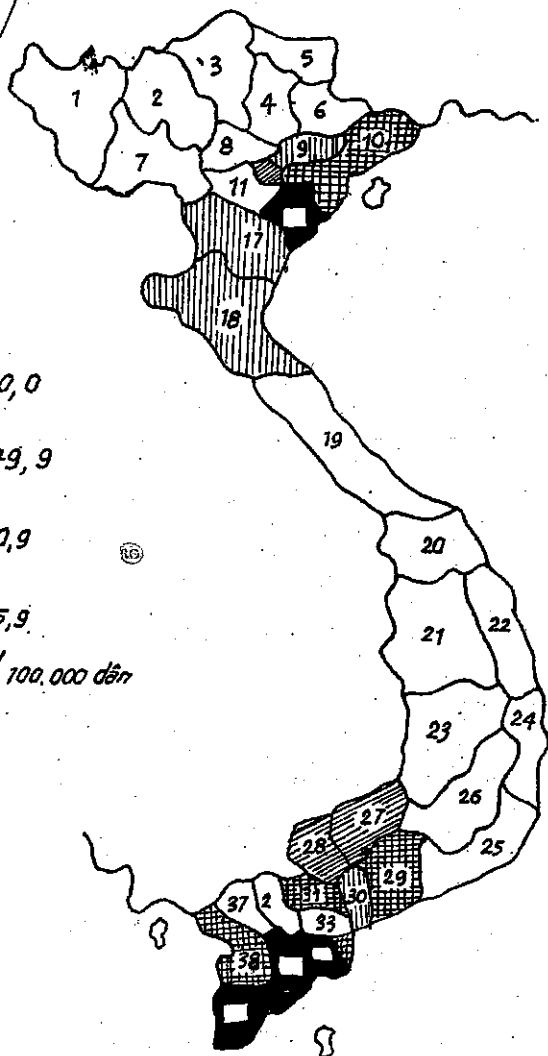
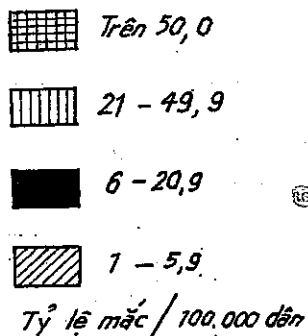
- Biểu đồ hình quạt.

Đây là loại biểu đồ hình tròn cắt ra thành một số phần, tượng trưng cho sự phân phối tần số của các nhóm khác nhau của một biểu đồ mô tả. Biểu đồ hình quạt thường dùng tỷ lệ phần trăm.

Nguồn calo chủ yếu cho trẻ em từ 6-12 tháng tuổi.



- Biểu đồ tọa độ địa lý.



Tình hình dịch tã ở Việt Nam năm 1976.



THƯ VIỆN
HUBT

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

Các biểu đồ tọa độ địa lý là những loại biểu đồ biểu thị sự lưu hành của một sự kiện bằng cách dùng các bản đồ. Thường sử dụng phổ biến là các bản đồ chấm và bản đồ khu vực.

- Bản đồ chấm dùng các chấm hay các thuật ngữ để đánh dấu vị trí sự kiện xảy ra hay vị trí bệnh đang lưu hành.

- Bản đồ khu vực cho thấy tỷ lệ mắc của một sự kiện (bằng cách tô màu hoặc đánh bóng khác nhau) ở các khu vực khác nhau, hoặc phân bố theo khu vực của một bệnh nào đó.

Bản đồ chấm được tạo ra một cách đơn giản bằng cách ghim hoặc đính một ký hiệu nào đó lên bản đồ, ở tại vị trí tương ứng với nơi bệnh hoặc sự kiện xảy ra. Trường hợp nhiều sự kiện xảy ra tại một nơi, phải dùng các ký hiệu sao cho diễn giải được dải tần số xuất hiện của các sự kiện.

Giá trị của loại bản đồ chấm là miêu tả được sự phân bố theo địa lý của một lớp hay một loại sự kiện. Loại bản đồ này không cung cấp cho ta một chỉ số nào đo lường nguy cơ xảy ra sự kiện ở một vị trí cụ thể mặc dầu có thể cho biết số lượng sự kiện xuất hiện ở vị trí đó, vì rằng nó không xét đến kích thước của quần thể nơi sự kiện xảy ra.

III. BÀI TẬP

Bài tập 1. Kết quả điều tra albumin huyết thanh (g/l) của 50 phụ nữ huyện X tháng 10-1990 như sau:

42	41	42	44	44	36	38	41	42	44
42	39	49	40	45	32	34	43	37	39
41	39	48	42	43	33	43	35	32	34
39	35	43	44	47	40	39	42	41	46
37	49	41	39	43	42	47	48	51	52

Sử dụng 6 khoảng cách bằng nhau, từ bảng số liệu này có thể được trình bày thành một bảng phân phối tần số.

Bảng phân phối tần số albumin huyết thanh.

Albumin huyết thanh (g/lit)	Tần số
30-33	3
34-37	7
38-41	14
42-45	17
46-49	7
50-53	2
Tổng	50



- Qua bảng này ta thấy được sự phân phối albumin huyết thanh trong 50 đối tượng nghiên cứu. sự thay đổi từ 30 tới 53, và thực tế mức albumin huyết thanh giữa 38 và 45g/lít là tập trung nhiều, thấy rõ được.

- Số liệu albumin huyết thanh ở bài tập 1 ta vẽ đồ thị minh họa như đồ thị hình cột.

Đồ thị hình cột của số liệu albumin huyết thanh

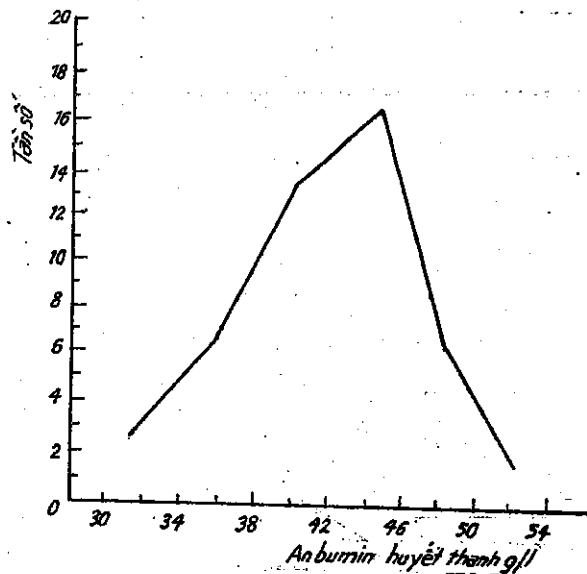
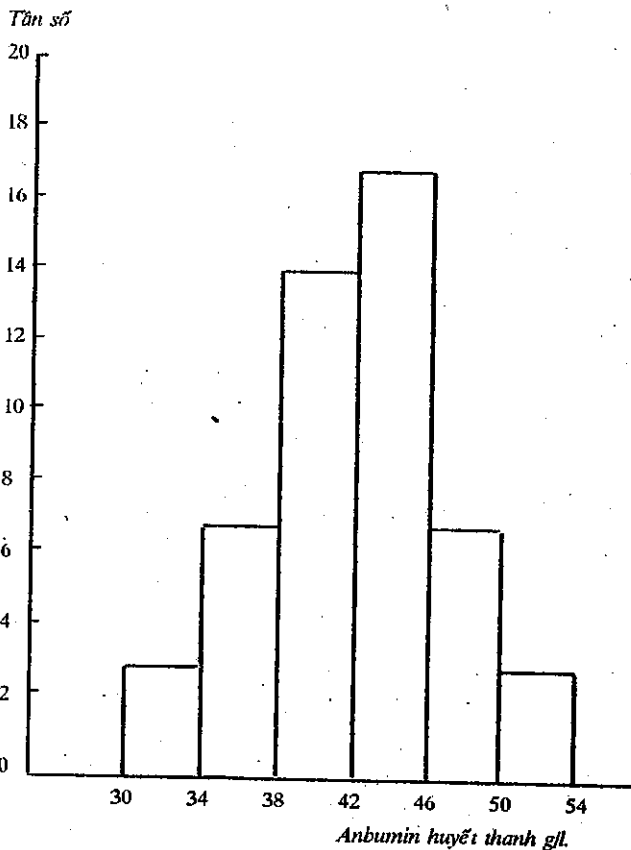
- Số liệu albumin từ bài tập 1 ta có thể trình bày đồ thị như đồ thị đa giác tần số.

Bài tập 2.

Trong một nghiên cứu về mối quan hệ giữa viêm miệng và nghề nghiệp, nghề nghiệp của các đối tượng nghiên cứu được ghi nhận: P(công nhân chuyên nghiệp), S(công nhân lành nghề, và U(công nhân không lành nghề). Có 88 người bị bệnh viêm miệng (+) và 100 người không bị bệnh viêm miệng. Danh sách bệnh và nghề nghiệp của các đối tượng nghiên cứu được trình bày ở bảng sau:

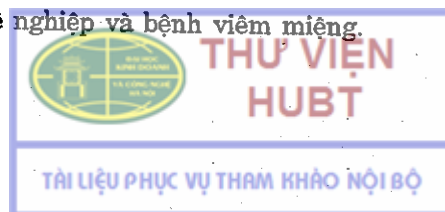
Bảng số liệu về viêm miệng.

Người	Viêm miệng	Nghề nghiệp
1	-	P
2	+	U
3	+	U
4	-	-
...
186	+	S
187	-	U
188	-	P



Số liệu này có thể trình bày dưới dạng bảng 2 chiều của 2 biến số: Viêm miệng (có/không) và nghề nghiệp (P, S, U) như sau:

Sự phân phối của 188 người theo nghề nghiệp và bệnh viêm miệng.



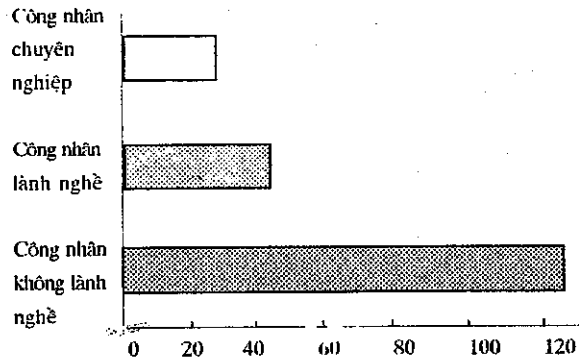
Viêm miệng	Nghề nghiệp			Tổng số
	P	S	U	
Có	5	13	70	88
Không	20	30	50	100
Tổng số	25	43	120	188
% mắc bệnh	20,0	30,2	58,3	46,8

Bảng này cho thấy tần số tương đối về bệnh viêm miệng ở những loại nghề nghiệp khác nhau.

Sự phân phối nghề nghiệp của 188 người ở bài tập 2 có thể trình bày thành biểu đồ hình chữ nhật.

Biểu đồ hình chữ nhật về nghề nghiệp.

Một biểu đồ hình quạt trình bày sự phân phối nghề nghiệp ở bài tập 2. Tần số phải được biến đổi thành những góc tương ứng.



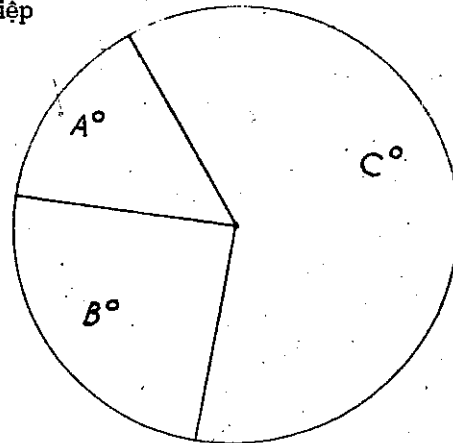
Nghề nghiệp	Tần số	Góc hình quạt
Công nhân chuyên nghiệp	25	A
Công nhân lành nghề	43	B
Công nhân không lành nghề	120	C
Tổng số	188	360°

Biểu đồ hình quạt trình bày số liệu nghề nghiệp

$$\frac{A}{360^\circ} = \frac{25}{188} \quad \text{vậy} \quad A = \frac{25 \times 360^\circ}{188} = 48^\circ$$

$$B = \frac{43 \times 360^\circ}{188} = 82^\circ$$

$$C = \frac{120 \times 360^\circ}{188} = 230^\circ$$



Tóm tắt những phương pháp thích hợp trình bày số liệu tùy thuộc vào từng tình huống.

Bảng phân phối tần số

Số liệu

- Số liệu nhiều, chỉ có một biến
- Số liệu nhiều, có từ 2 biến trở lên

Bảng

- Bảng tần số
- Bảng hai hay nhiều chiều

Đồ thị phù hợp với bảng phân phối tần số

Bảng số liệu

- Bảng tần số, biến định lượng, một bộ số liệu
- Bảng tần số, biến định lượng, hai bộ số liệu
- Bảng tần số, số liệu phân loại

Đồ thị

- Đồ thị hình cột hoặc Đa giác tần số
- Đa giác tần số
- Biểu đồ hình chữ nhật hoặc biểu đồ hình quạt.

123 BÀI TẬP TỰ ĐÁNH GIÁ

(Có trả lời ở cuối sách)

HÃY CHỌN MỘT HOẶC NHIỀU CÂU TRẢ LỜI ĐÚNG.

(Sinh viên không tự ý đánh dấu vào các câu này, cho các năm sau học)

Một vụ dịch viêm dạ dày xảy ra trên một chiếc tàu thủy.

Số liệu dưới đây đã được thu thập ngay sau khi dịch xảy ra bằng cách phỏng vấn tất cả mọi người trên tàu:

Bảng 1. SỐ LIỆU VỀ MỘT VỤ DỊCH VIÊM DẠ DÀY TRÊN TÀU.

Thực phẩm	Những người có ăn thực phẩm		Những người không ăn thực phẩm	
	Viêm dạ dày	Khỏe	Viêm dạ dày	Khỏe
Cá trích	200	800	100	900
Cà	650	350	100	900
Trứng rán	200	800	500	500
Sò	300	700	400	600
Kem	600	400	450	550

Bảng số liệu này cho cả bài tập số 1 và 2

1- Loại thức ăn nào có khả năng gây viêm dạ dày trên tàu?

a/ Cá trích.

b/ Gà.

c/ Trứng rán.

d/ Sò.

e/ Hến.

2- Nguy cơ tương đối gây viêm dạ dày của cá trích là bao nhiêu?

a/ 0,5

b/ 2,0

c/ 2,3

d/ 8.0

e/ Không thể tính được từ bảng đã cho.

3- Các nhà thống kê cho biết rằng tai nạn thường xảy ra ở các ô tô màu xanh nhiều hơn so với các ô tô màu khác. Người ta kết luận rằng, trong khi lái xe màu xanh, người lái xe có nguy cơ bị tai nạn cao hơn so với khi lái xe ô tô cơ màu khác. Kết luận này là:

a/ Đúng

b/ Không đúng vì so sánh không dựa trên tỷ lệ.

c// Không đúng vì không sử dụng nhóm chứng hay nhóm so sánh.

d/ Không đúng vì không làm kiểm định ý nghĩa thống kê.

e/ Không đúng vì tỷ lệ hiện mắc được sử dụng chứ không phải tỷ lệ mới mắc.

4- Trong một nghiên cứu 500 bệnh nhân và 500 người đối chứng, người ta đã tìm ra một yếu tố bệnh căn nghi ngờ ở 400 bệnh nhân và 100 người đối chứng. Nguy cơ tuyệt đối, tỷ lệ mới mắc ở những người có yếu tố này là:

a/ 80%

b/ 40%

c/ 16%

d/ 20%

e/ Không thể tính được từ số liệu đã cho

5- Năm 1945, có 1000 phụ nữ làm việc ở nhà máy vẽ mặt đồng hồ. Tỷ lệ mới mắc (incidence) ung thư xương của những phụ nữ này vào năm 1975 đã được so sánh với 1000 phụ nữ làm việc ở tổng đài điện thoại vào năm 1945. 20 phụ nữ vẽ mặt đồng hồ và 4 phụ nữ ở tổng đài điện thoại đã bị ung thư xương trong khoảng thời gian từ 1945 - 1975. Nguy cơ tương đối phát triển ung thư xương ở những người vẽ mặt đồng hồ là:

a/ 2

b/ 4

c/ 5

d/ 8

	20.996	20.00	4	20	
ng. c. b.	400	100			
đ. c. h.	100	400			
	500	500			

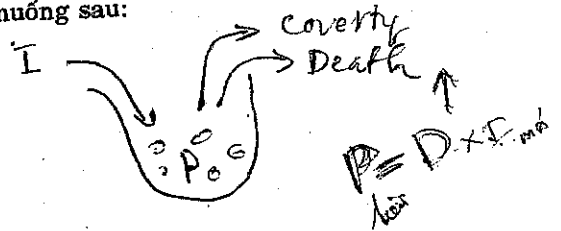
4.980	980	996	20	20
1000	1000	1000	20	20

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

- e/ Không thể tính được từ số liệu đã cho.
- 6- Dịch được định nghĩa là.
- a/ Một bệnh có tỷ lệ thấp nhưng thường xuyên xảy ra trong một cộng đồng hay một vùng.
- b/ Tỷ lệ tấn công (attack rate) vượt quá 10/1000 dân.
- c/ Sự xuất hiện hiện bệnh vượt quá số mong đợi trung bình ở quần thể đó trong thời gian đó.
- d/ Các bệnh đường hô hấp xảy ra theo mùa.
- e/ Tỷ lệ mắc hàng năm/100.000 dân.

7- Khi một loại thuốc hay một phương pháp điều trị có khả năng làm giảm tỷ lệ chết nhưng không làm khởi hẳn bệnh, sẽ dẫn đến tình huống sau:

- a/ Tỷ lệ hiện mắc của bệnh sẽ giảm.
- b/ Tỷ lệ mới mắc của bệnh sẽ tăng.
- c/ Tỷ lệ hiện mắc của bệnh sẽ tăng.
- d/ Tỷ lệ mới mắc của bệnh sẽ giảm.
- e/ Tỷ lệ mới mắc và hiện mắc của bệnh sẽ giảm.



Một cộng đồng A có 100.000 dân. Năm 1985 có 1.000 người chết do tất cả các nguyên nhân. Có 300 bệnh nhân lao, 200 nam và 100 nữ. Trong năm 1985, có 60 bệnh nhân lao chết, trong đó 50 là nam. Số liệu này được sử dụng cho các bài từ 8-12.

8- Tỷ lệ chết thô (Cude mortality rate) ở cộng đồng A là:

- a/ 300/100.000
- b/ 60/1.000
- c/ 10/1.000
- d/ 100/1.000
- e/ Không thể tính được từ số liệu đã cho.

	300	60	60
Chết	1000	200	100
Đông			
			100.000

Handwritten notes: "60/1000" and "100.000" are written below the table.

9- Tỷ lệ chết do lao là:

- a/ 20%
- b/ 30%
- c/ 6%
- d/ 3%
- e/ Không thể tính được từ số liệu đã cho.

10- Tỷ lệ chết mắc (case fatality rate) của lao là:

- a/ 6%
- b/ 20%
- c/ 2%
- d/ Như nhau ở nam và nữ.
- e/ Không thể tính được từ số liệu đã cho.

11- Tỷ lệ chết riêng phần theo nguyên nhân (cause-specific mortality rate) đối với lao là:

- a/ 60/100.000
- b/ 300/100.000
- c/ 200/1.000
- d/ 20%
- e/ Không thể tính được từ số liệu đã cho

12- Tỷ lệ chết riêng phần theo thời giới (sex-specific mortality rate) đối với lao là:

- a/ 0,5/1.000
- b/ 25%
- c/ Nam lớn hơn nữ
- d/ 50/300

$$\frac{50}{10000}$$

$$\frac{\sum \text{chết}}{\text{Số người}}$$

$$\frac{\sum \text{chết}}{2 \text{ đ/s}}$$

e/ Không thể tính được từ số liệu đã cho.

13- Cộng đồng P và Q có tỷ lệ tử vong đã được chuẩn hóa theo tuổi là như nhau. Cộng đồng P có tỷ lệ tử vong thô (crude mortality rate) thấp hơn so với cộng đồng Q. Người ta có thể kết luận rằng:

- a/ Hai cộng đồng có phân bố tuổi như nhau.
- b/ Chẩn đoán ở cộng đồng P chính xác hơn cộng đồng Q.
- c/ Cộng đồng P có dân chúng già hơn cộng đồng Q.
- d/ Chẩn đoán ở cộng đồng Q kém chính xác hơn so với cộng đồng P.
- e/ Cộng đồng P có dân chúng trẻ hơn cộng đồng Q.

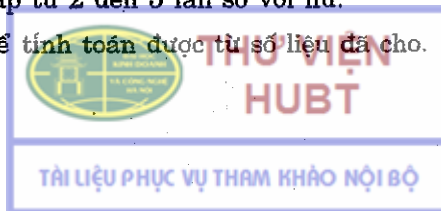
14- Có 3 đợt bệnh có phân bố theo giới như sau:

Bảng 2. PHÂN BỐ THEO GIỚI CỦA BA ĐỢT BỆNH.

Đợt bệnh	Bệnh nhân nam	Bệnh nhân nữ
1	200	100
2	250	50
3	450	150
Tổng cộng	900	300

Tỷ lệ mới mắc (incidence) của bệnh theo giới là:

- a/ Ở nam gấp đôi ở nữ.
- b/ Ở nam gấp ba lần so với nữ.
- c/ Ở nam gấp năm lần so với nữ.
- d/ Ở nam gấp từ 2 đến 5 lần so với nữ.
- e/ Không thể tính toán được từ số liệu đã cho.



Một nghiên cứu lớn về ung thư bàng quang và hút thuốc lá thu được những số liệu sau:

Bảng 3- TỶ LỆ UNG THƯ BÀNG QUANG Ở NHỮNG NGƯỜI CÓ HÚT VÀ KHÔNG HÚT THUỐC LÁ.

Tỷ lệ ung thư bàng quang / 100.000 nam giới	
Hút thuốc lá	48,0
Không hút thuốc lá	25,4

Bảng này được sử dụng cho bài 15 và 16.

15- Nguyên cơ tương đối phát triển ung thư bàng quang ở những nam giới hút thuốc lá so với nam giới không hút thuốc lá là:

a/ 48,0

b/ $48,0 - 25,4 = 22,6$

c/ $48,0/25,4 = 1,89$
 $48,0 - 25,4$

d/ $\frac{48,0}{48,0}$

e/ Không thể tính toán được từ số liệu đã cho.

	K	Đ	
Hút	48	25,4	1770
Không	22,4	1	1072

ad
bc

16. Nguyên cơ quy thuộc của ung thư bàng quang do hút thuốc lá ở nam giới hút thuốc lá là:

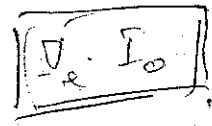
a/ $48,0/25,4 = 1,89$

b/ $48,0 - 25,4 = 22,6/100.000$

c/ 48,0

d/ $48,0/100.000 = 0,00048$

e/ Không thể tính toán được từ số liệu này.



Bảng 4. Trình bày tổng số những người ăn một trong hai loại thực phẩm đặc biệt có thể bị nhiễm Streptococci nhóm A.

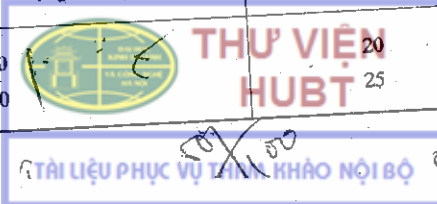
Bảng 5. Trình bày tổng số bệnh nhân (ví dụ: viêm họng cấp tính) là những người đã ăn hai loại thực phẩm đó.

Bảng 4- TỔNG SỐ NGƯỜI ĂN CÁC LOẠI THỰC PHẨM ĐẶC BIỆT.

	Ăn thịt gà lồi	Không ăn thịt gà lồi
Ăn trứng cá muối	100	100
Không ăn trứng cá muối	100	100

Bảng 5- TỔNG SỐ NGƯỜI ỐM VÌ ĂN CÁC LOẠI THỰC PHẨM ĐẶC BIỆT.

	Ăn thịt gà lồi	Không ăn thịt gà lồi
Ăn trứng cá muối	50	20
Không ăn trứng cá muối	50	25



AR/6

Hai bảng này được sử dụng cho bài 17 và 18.

17. Tỷ lệ tấn công bị viêm họng ở những người ăn cả hai loại thực phẩm thịt gà lồi và trứng cá muối là bao nhiêu?

- a/ 50/50
- b/ 50/70
- c/ 50/75
- d/ 50/100
- e/ 50/200

18. Theo kết quả trình bày ở bảng 4 và bảng 5, loại thực phẩm nào (hay kết hợp cả hai loại) là có khả năng bị nhiễm khuẩn nhiều nhất?

- a/ Gà lồi
- b/ Cá muối
- c/ Không phải gà lồi và cũng không phải cá muối.
- d/ Ăn cả thịt gà lồi và trứng cá muối.
- e/ Không thể tính toán được từ số liệu đã cho.

19. Những chỉ số thống kê sức khỏe nào sau đây có ích lợi nhất trong việc xác định các yếu tố nguy cơ trên cơ sở đó áp dụng những biện pháp can thiệp có hiệu quả nhất. Giả sử rằng các biện pháp can thiệp có hiệu quả như nhau nhưng thiếu các nguồn lực.

- a/ Tỷ lệ mới mắc bệnh ở những người có phơi nhiễm.
- b/ Nguy cơ quy thuộc
- c/ Tỷ lệ hiện mắc bệnh ở những người có phơi nhiễm.
- d/ Nguy cơ tương đối của bệnh.
- e/ Tỷ lệ tử vong theo tỷ lệ.

20. Để đánh giá mức độ kết hợp giữa một phơi nhiễm và một bệnh, những chỉ số thống kê sức khỏe sau là có ích lợi nhất.

- a/ Tỷ lệ mới mắc bệnh ở những người có phơi nhiễm.
- b/ Nguy cơ quy thuộc.
- c/ Tỷ lệ người hiện có phơi nhiễm.
- d/ Nguy cơ tương đối của bệnh
- e/ Tỷ lệ tử vong theo tỷ lệ.

21. Bảng dưới đây trình bày tỷ lệ tấn công bị ngộ độc thực phẩm do ăn một số loại thực phẩm đặc biệt.

Bảng 6- TỶ LỆ TẤN CÔNG DO ĂN MỘT SỐ LOẠI THỰC PHẨM.

	Tổng số	Ăn gà tây		Không ăn gà tây		
		Số ốm	Tỷ lệ tấn công (%)	Tổng số	Số ốm	Tỷ lệ tấn công (%)
Ăn khoai tây và nước thịt	127	92	72	0	0	-
Không ăn khoai tây và nước thịt	6	5	83	25	2	8



Nghiên cứu các bảng 4,5,6, những nhận xét nào sau đây là đúng liên quan với yếu tố gây nhiều?

1. Việc ăn hay không ăn khoai tây và nước thịt là yếu tố gây nhiều khi đánh giá nguy cơ nhiễm khuẩn do ăn thịt gà tây.

2. Việc ăn trứng cá muối hay không ăn trứng cá muối là yếu tố gây nhiều khi đánh giá nguy cơ nhiễm khuẩn do ăn thịt gà lôi.

- a/ 1 đúng, 2 sai.
- b/ 2 đúng, 1 sai
- c/ cả 1 và 2 đều đúng
- d/ Cả 1 và 2 đều sai.

e/ Không thể đánh giá yếu tố gây nhiều dựa trên các số liệu đã cho.

22. Ví dụ về tỷ lệ hiện mắc là như sau:

a/ Số lần bị viêm họng ở trẻ 3 tuổi hàng năm.

b/ Số các trường hợp mới bị ung thư tuyến tiền liệt hàng năm trên 100.000 đàn ông.

c/ Số bệnh nhân đái đường ở một trường đại học.

d/ Tổng số bệnh nhân bị xơ cứng lan tỏa trên 100.000 dân hàng năm.

23. 112 người bị ốm sau một cuộc dạo chơi dã ngoại trong tổng số 250 người (80 nam và 170 nữ). Trong số những người bị ốm có 76 nữ và 36 nam.

a/ Tỷ lệ tấn công theo giới đối với nam là 0,32

b/ Tỷ lệ tấn công theo giới đối với nam là 0,45

c/ Tỷ lệ tấn công theo giới đối với nữ là 0,68

d/ Tỷ lệ tấn công chung là 0,45

Trong đó:

$$36/112 = 0,32$$

$$76/112 = 0,68$$

$$36/250 = 0,14$$

$$76/250 = 0,30$$

$$36/80 = 0,45$$

$$76/170 = 0,45$$

$$122/250 = 0,45$$

24. Sau 5 năm tiến hành chương trình phát hiện và điều trị tích cực tỷ lệ hiện mắc của bệnh đái đường cao hơn so với một năm trước khi bắt đầu tiến hành chương trình. Những lý do nào sau đây có thể là đúng?

a- Có sự tăng số trường hợp bệnh mà trước đây không phát hiện được.

b- Giảm các yếu tố nguy cơ mắc bệnh đái đường

c- Giảm tỷ lệ chết so với số mắc bệnh đái đường.

d- Tăng tỷ lệ tử vong theo tỷ lệ của bệnh đái đường.

25.



Bảng 7- TỶ LỆ TỬ VONG THEO TUỔI Ở 2 BỆNH VIỆN CỦA MỘT THÀNH PHỐ NĂM 1988.

Tuổi (năm)	Bệnh viện A			Bệnh viện B		
	Số chết	Số bệnh nhân	Tỷ lệ chết trên 1000	Số chết	Số bệnh nhân	Tỷ lệ chết trên 1000
50	10	500	20	30	1000	30
49	50	1000	50	30	500	60
Tổng cộng	60	1500		60	1500	

Những giải thích nào sau đây đúng với bảng trên.

- a) Tỷ lệ chết thô là như nhau ở cả hai bệnh viện A và B.
- b) Tỷ lệ tử vong riêng phần theo tuổi ở bệnh viện B lớn hơn so với bệnh viện A.
- c) Bệnh viện A thu nhận bệnh nhân già hơn.
- d) Tỷ lệ tử vong chuẩn hóa theo tuổi ở bệnh viện B sẽ lớn hơn so với bệnh viện A.
- Một chương trình phát hiện hàng loạt tăng huyết áp đã được thực hiện ở một vùng nội thành. Kết quả huyết áp tâm trương ở 1500 đàn ông, tuổi từ 30-69 được trình bày ở bảng dưới đây.

Bảng 8- TẦN SỐ HUYẾT ÁP TÂM TRƯƠNG Ở NAM GIỚI TUỔI TỪ 30-69.

Huyết áp tâm trương (mmHg)	Tần số	Phần trăm
<65	60	4
65-74	270	18
75-84	540	36
85-94	420	28
95-104	150	10
105-115	45	3
>115	15	1
Tổng cộng	1500	100

Bảng số liệu này được sử dụng cho các bài từ 26 đến 29.

26. Nếu có một nghiệm pháp chẩn đoán xác định bệnh cao huyết áp cho những người có HATT từ 95 mmHg trở lên, thì nghiệm pháp chẩn đoán xác định này phải được áp dụng bắt đầu từ độ bách phân vị nào sau đây:

- a/ BPV 86 trở lên
- b/ BPV 97,5 trở lên
- c/ BPV 95 trở lên
- d/ BPV 14 trở xuống
- e/ BPV 28 trở xuống.

27. Giả sử rằng những người đàn ông được khám huyết áp đã đại diện cho tất cả đàn ông ở lứa tuổi từ 30-69 tuổi ở vùng này. Khả năng một người đàn ông được chọn ngẫu nhiên sẽ có huyết áp tâm trương ≥ 105 mmHg là:

- a/ 3%
- b/ 4%
- c/ 14%
- d/ 86%
- e/ 96%

28. Trung bình và trung vị của phân phối huyết áp là xấp xỉ 83 mmHg và độ lệch chuẩn là 12 mmHg. Những chỉ số này đã cho phép chúng ta rút ra những nhận xét từ một nhận xét nào sau đây:

- a/ Xấp xỉ 95% đàn ông có huyết áp tâm trương từ 59 đến 107 mmHg.
- b/ Phân phối gần như đối xứng.
- c/ Giới hạn với khoảng tin cậy 95% đối với huyết áp trung bình của tất cả đàn ông, tuổi từ 30-69 trong quần thể này là từ 59 đến 107 mmHg.
- d/ Xấp xỉ 50% đàn ông có huyết áp tâm trương > 83 mmHg.
- e/ Số trung bình không bị ảnh hưởng nhiều bởi những người có huyết áp rất cao trong ví dụ này.

29. Phân bố huyết áp tâm trương của 800 phụ nữ trong quần thể này gần như đối xứng với độ lệch chuẩn giống như ở nam giới. Huyết áp tâm trương trung bình của nữ là 79 mmHg. Do đó, chúng ta có thể kết luận là:

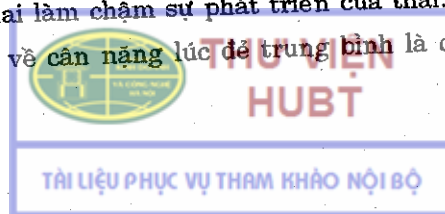
- a/ Giá trị trung vị giống như ở nam giới
- b/ Khoảng giới hạn bình thường của nữ giống như ở nam
- c/ Sai số chuẩn của giá trị trung bình giống như ở nam
- d/ Tỷ lệ tăng huyết áp ở nữ lớn hơn ở nam
- e/ Tỷ lệ tăng huyết áp ở nữ nhỏ hơn ở nam.

30. Tương quan giữa hai biến số đo lường mức độ mà các biến số đó.

- a/ Loại trừ nhau
- b/ Có liên hệ nhân quả
- c/ Kết hợp với nhau
- d/ Có ý nghĩa thống kê
- e/ Ảnh hưởng lẫn nhau

31. Cân nặng lúc đẻ trung bình của trẻ sơ sinh của 23 bà mẹ hút hơn 1 bao thuốc lá một ngày trong khi có thai thấp hơn 200g so với trẻ sơ sinh của 16 bà mẹ không bao giờ hút thuốc. Có sự khác biệt có ý nghĩa thống kê ở mức ý nghĩa 5% ($p < 0,05$). Điều này có nghĩa là:

- a/ Hút thuốc lá trong khi có thai làm chậm sự phát triển của thai.
- b/ Sự khác biệt quan sát được về cân nặng lúc đẻ trung bình là quá lớn nên không



thể là do may rủi.

- c/ Sự khác biệt quan sát được về cân nặng lúc đẻ trung bình có thể là do may rủi.
- d/ Số bệnh nhân được nghiên cứu đã không đủ để đưa ra một kết luận.
- e/ Hút thuốc lá khi có thai không ảnh hưởng đến sự phát triển của thai.

Một nghiệm pháp sàng tuyển phát hiện ung thư và đã được tiến hành trên 400 phụ nữ bị ung thư vú đã được xác định bằng giải phẫu bệnh và 400 phụ nữ không bị ung thư vú. Kết quả dương tính của nghiệm pháp này được thấy ở 100 bệnh nhân bị ung thư và 50 phụ nữ bình thường.

(Số liệu này được sử dụng chung cho các bài 32, 33 và 34)

32 Độ nhạy của nghiệm pháp là:

- a/ 87%
- b/ 67%
- c/ 25%
- d/ 33%
- e/ 12%

$$Se = \frac{a}{a+b}$$

	B	KB
+	100a	50b
-	300c	350d
	400	400

33. Độ đặc hiệu của nghiệm pháp là:

- a/ 87%
- b/ 67%
- c/ 25%
- d/ 33%
- e/ 12%

$$Sp = \frac{d}{c+d}$$

$$Se = \frac{a}{a+b}$$

$$Sp = \frac{d}{c+d}$$

34. Giá trị dự đoán dương tính của nghiệm pháp là:

- a/ 88%
- b/ 67%
- c/ 25%
- d/ 33%
- e/ 12%

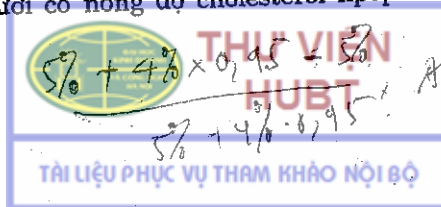
$$PV = \frac{p \cdot Se}{p \cdot Se + (1-p) \cdot (1-Sp)}$$

$$p \cdot Se + (1-p) \cdot (1-Sp)$$

35. 5% phụ nữ có thai có triệu chứng nhiễm khuẩn đường tiết niệu khi đến khám thai lần đầu tiên. 4% những phụ nữ không có triệu chứng nhiễm khuẩn đường tiết niệu lúc khám thai lần đầu đã phát triển nhiễm khuẩn trong thời gian sau và lúc đẻ. Khả năng một người phụ nữ bị nhiễm khuẩn đường tiết niệu trong khi có thai là:

- a/ $0,04 \times 0,95 = 0,038$
- b/ 0,05
- c/ $0,04 + 0,05 = 0,09$
- d/ $0,05 + (0,04) \cdot (0,95) = 0,088$
- e/ $0,04 \times 0,05 = 0,002$

36. Những người có nồng độ cholesterol lipoprotein cao (HDL) được phát hiện là có



$$AR\% = \frac{I_e - I_o}{I_e} \cdot 100$$

nguy cơ thấp mắc bệnh động mạch vành tim. Các yếu tố kết hợp với cholesterol đã được thu thập trong nghiên cứu 293 người đàn ông khỏe mạnh. Bảng dưới đây trình bày hệ số tương quan r và giá trị P có liên quan với từng test kiểm định tương quan giữa cholesterol và biến số khác. Nghiên cứu được tiến hành đồng thời ở 293 người.

Bảng 9. TƯƠNG QUAN GIỮA CHOLESTEROL, MÁU VÀ 6 BIẾN SỐ.

	r	P
Plasma Triglyceride	-0,42	< 0,001
Uống rượu	0,24	< 0,001
Glucose máu	- 0,19	< 0,001
Trọng lượng cơ thể	- 0,11	> 0,05
Huyết áp tâm trương	0,04	< 0,05

Những nhận xét nào sau đây có liên quan với nồng độ Triglyceride máu.

- a/ Khi làm tăng Triglyceride máu. Chúng ta có thể làm tăng nồng độ Cholesterol
- b/ Khi làm giảm Triglyceride máu, chúng ta có thể làm tăng nồng độ cholesterol.
- c/ Nồng độ cao cholesterol dễ được phát hiện ở những người đàn ông có triglyceride thấp.
- d/ Nồng độ thấp cholesterol HDL dễ được phát hiện ở những người đàn ông có Triglyceride máu thấp.
- e/ Giá trị P quá nhỏ nên người ta coi là không tương quan hơn là do may rủi.

37. Trong một nghiên cứu 50 bệnh nhân (cases) và 50 đối chứng (control) người ta đã xác định rằng sự khác nhau phát hiện được về một yếu tố nguy cơ là không có ý nghĩa thống kê. Người ta có thể kết luận rằng:

- a/ Không có sự kết hợp giữa yếu tố và bệnh
- b/ Sự khác biệt chỉ có ý nghĩa về lâm sàng.
- c/ Sự khác biệt có thể là do thay đổi cách chọn mẫu.
- d/ Tính so sánh giữa nhóm bệnh và nhóm chứng đã được khẳng định.
- e/ Sai lệch do người phỏng vấn và người quan sát đã được loại trừ.

38. Ở lứa tuổi 65, khả năng sống trong 5 năm sau đó là 0,8 đối với nam da trắng và 0,9 đối với phụ nữ da trắng. Đối với những cặp vợ chồng ở tuổi 65, khả năng người vợ sẽ góa chồng 5 năm sau đó là:

- a/ 90%
- b/ 20%
- c/ 18%
- d/ 10%
- e/ 8%

18/

	chết	sống	
♂	20	80	100
♀	10	90	100

39. Với số liệu của bài 38 trên đây, thì khả năng ít nhất một trong 2 người (vợ,

20% 100%

	chết	sống	
♂	20	80	100
♀	10	90	100

chồng) sẽ sống trong 5 năm sau đó là:

a/ 98%

b/ 90%

c/ 72%

d/ 28%

e/ 10%

$$0,8 + 0,9 - 0,2 = 1,5$$

40. Một kết quả xét nghiệm với giá trị trung bình là 18 g/100ml và độ lệch chuẩn là 1,5 có nghĩa là:

a/ Giá trị thật là từ 16,5 - 19,5g/100ml

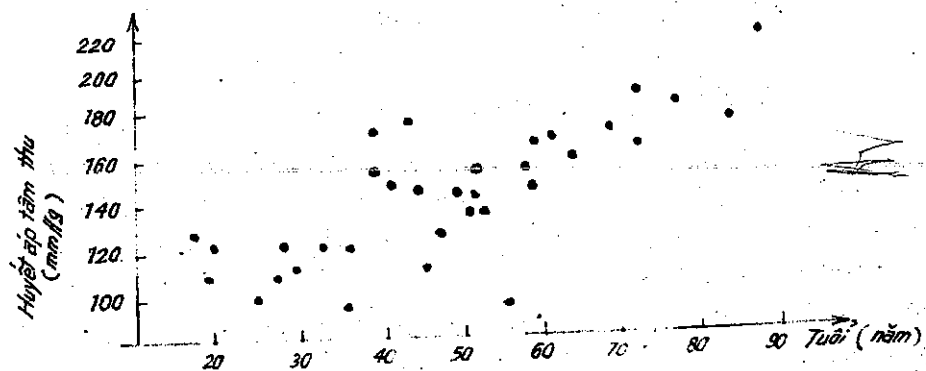
b/ Giá trị thật là từ 15,0 - 21,0 g/100ml

c/ Sai số quá lớn đến nỗi không xác định được bất kỳ một giá trị nào.

d/ Nếu tiến hành xét nghiệm lặp lại trên cùng một mẫu 95% các giá trị sẽ nằm trong khoảng 15,0-21,0 g/100ml.

e/ 5% giá trị thật nằm ngoài khoảng 16,5-19,5g/100ml (nhỏ hơn 16,5 hay lớn hơn 19,5 g/100 ml).

41. Đồ thị phân bố điểm (scattergram) dưới đây cho thấy mối quan hệ giữa huyết áp tâm thu và tuổi ở 33 phụ nữ.



Hình 1- Mối quan hệ giữa huyết áp tâm thu và tuổi của phụ nữ.

Hệ số tương quan nào sau đây giữa huyết áp và tuổi đã được xác định từ số liệu trên?

a/ + 1,20

b/ - 0,22

c/ + 0,70

d/ - 0,85

e/ 0

Bảng 10 dưới đây được sử dụng cho các bài từ 42 - 44.

Bảng 10- CÁC HỆ SỐ ĐỐC (b) LIÊN QUAN GIỮA THỜI KỲ THAI NGHÉN VÀ CÁC ĐẶC TRƯNG CỦA BÀ MẸ.

Biến số độc lập	Hệ số đốc(b)
Tuổi (năm)	- 0,0215
Chiều cao (Inch)	0,0279
Giáo dục (năm)	0,0424
Chủng tộc (Da trắng=0, Da đen =1)	-1,16
Số lần có thai trước đó	0,0103
Hệ số chặn (a) = 37,0 $R^2 = 0,0242$	

Bảng trên là kết quả của một nghiên cứu 883 phụ nữ có thai nhằm tính toán phương trình hồi quy đa bội mối liên quan giữa thời kỳ thai nghén. (tính theo tuần) và các đặc trưng của bà mẹ.

42. Hệ số chặn (a) cho biết:

- Thời gian thai nghén trung bình của các phụ nữ có thai này.
- Thời gian thai nghén dự tính khi các biến số độc lập bằng 0.
- Thời gian thai nghén sớm nhất của những phụ nữ có thai này.
- Thời gian thai nghén dài nhất của những phụ nữ có thai này
- Độ mạnh kết hợp giữa thời gian thai nghén và 5 biến số độc lập.

43. Hệ số đốc (b) là -1,16 đối với biến số chủng tộc có giá trị $p < 0,001$, do đó, những nhận xét nào sau đây là đúng.

- Ảnh hưởng của chủng tộc đến thời gian thai nghén chắc là do may rủi.
- Sự biến thiên về thời gian thai nghén không phải là do nghiên cứu đã bao gồm cả phụ nữ da trắng và phụ nữ da đen.
- 0,9% sự biến thiên về thời kỳ thai nghén có thể là do chủng tộc.
- Một phụ nữ da đen có thể sinh sớm hơn so với phụ nữ da trắng ở cùng lứa tuổi, chiều cao, giáo dục và số lần có thai trước đó.

c. Ảnh hưởng của chủng tộc đến thời gian thai nghén có thể là vì có sự khác biệt giữa người da trắng và người da đen về tuổi, chiều cao, giáo dục và số lần có thai trước đó.

44. Ở những phụ nữ có thai này, các biến số tuổi, chiều cao, chủng tộc, giáo dục và số lần có thai trước đó:

- Cung cấp sự dự đoán chính xác về độ dài thời gian thai nghén.
- Không liên quan gì với độ dài thời gian thai nghén.
- Thay đổi quá nhiều nên không thể được dùng như là các chỉ số dự đoán về thời gian thai nghén.
- Giải thích cho tỷ lệ lớn biến thiên thời gian thai nghén.



e/ Giải thích cho tỷ lệ nhỏ biến thiên thời gian thai nghén.

45) Khi phân phối của một thông số ở một quần thể khỏe mạnh chệch nhiều so với phân phối chuẩn:

a/ Khoảng biến thiên của giá trị thông số đó không thể xác định được.

b/ Tỷ lệ phần trăm phân phối (percentile) là 2,5 và 97,5 được xác định tốt nhất từ phân phối tích lũy (cộng dồn).

c/ Tỷ lệ phần trăm phân phối 2,5 và 97,5 không còn là giới hạn bình thường nữa.

d/ Giá trị trung bình đã bị ảnh hưởng của các giá trị bất thường.

46. Trong một mẫu ngẫu nhiên 500 học sinh ở thành phố A, 27% có cảm thụ với bệnh sỏi. Sai số chuẩn của tỷ lệ phần trăm có cảm thụ với sỏi là 2%. Từ đó, có thể kết luận được là:

a/ Với độ tin cậy là 95%, tỷ lệ phần trăm có cảm thụ đối với tất cả học sinh ở thành phố A là từ 25% - 29%.

b/ Có sai chệch trong việc chọn mẫu.

c/ Số liệu phải được chuẩn hóa theo tuổi.

d/ Với độ tin cậy là 95%. Tỷ lệ phần trăm có cảm thụ với sỏi của tất cả học sinh ở thành phố A là từ 23-31%.

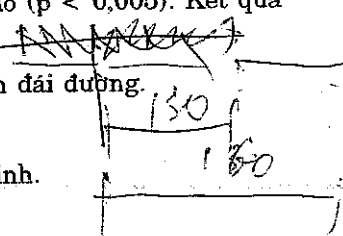
47. 50 bệnh nhân đái đường đã được so sánh với 50 người khỏe mạnh. Những bệnh nhân đái đường có tỷ lệ phản ứng thần kinh đối với bộ câu hỏi cao ($p < 0,005$). Kết quả này:

a/ Có thể do đặc tính của bệnh nhân không liên quan tới bệnh đái đường.

b/ Có thể bị ảnh hưởng do tác động của insulin.

c/ Có thể xảy ra nếu bệnh đái đường gây ra phản ứng thần kinh.

d/ Chắc là xuất hiện do may rủi.



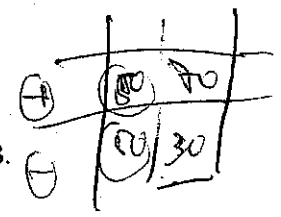
48. trong chương trình phát hiện bệnh đái đường, nồng độ đường máu ở mức sàng lọc đối với xét nghiệm A là 160mg/100ml và đối với xét nghiệm B là 130mg/100ml. Điều này có nghĩa là:

a/ Độ nhạy của xét nghiệm A lớn hơn so với xét nghiệm B.

b/ Độ đặc hiệu của xét nghiệm A lớn hơn so với xét nghiệm B.

c/ Số dương tính giả ở xét nghiệm A lớn hơn so với xét nghiệm B.

d/ Số âm tính giả ở xét nghiệm A lớn hơn so với xét nghiệm B.



49. Trong một lần khám sức khỏe thường kỳ, người ta tiến hành đo nồng độ axit uric của một người đàn ông 35 tuổi và được kết quả là 7,8mg/100ml. Khoảng biến thiên bình thường của axit uric trong xét nghiệm này là từ 3,4 - 7,5 mg/100ml. Nếu người này không có biểu hiện một triệu chứng hay dấu hiệu nào của bệnh gout những nhận xét sau có thể giải thích cho kết quả này:

a/ Người này nằm trong một tỷ lệ nhỏ những người khỏe mạnh có nồng độ axit uric máu cao.



b/ Người này nằm trong khoảng 2 độ lệch chuẩn so với số trung bình của những người khỏe mạnh.

c/ Những triệu chứng của bệnh gout chưa được biểu hiện.

d/ Giá trị nồng độ axit uric chệch khỏi khoảng biến thiên bình thường là không có ý nghĩa thống kê.

50. Phương trình hồi quy sau đã được phát triển từ một nghiên cứu 16 người mới được chẩn đoán là đái đường và đã dùng phenformin trong 1 năm:

$$L = -34 + 0,29W$$

Trong đó L là số cân nặng sút đi sau 1 năm điều trị, W là cân nặng lúc đầu của bệnh nhân. Trên cơ sở của thông tin này, chúng ta có thể kết luận như sau.

a/ Tất cả các bệnh nhân sút đi ít nhất 15,3kg sau năm đầu điều trị.

b/ Đường tương quan phù hợp với đồ thị phân bố điểm (scattergram) về sự sút cân ngược lại với cân nặng lúc đầu có đường thẳng nghiêng dương tính.

c/ Tương quan giữa sự sút cân và cân nặng lúc đầu là dương tính và rất mạnh.

d/ Những bệnh nhân nặng cân hơn những người khác vào lúc bắt đầu điều trị sẽ bị sút cân nhiều hơn trong năm đầu điều trị.

51. Trong việc nghiên cứu ngang về loét dạ dày ở một cộng đồng tỷ lệ những người có đủ triệu chứng để chẩn đoán là loét dạ dày là 80/100.000 đàn ông lứa tuổi 35 - 49 tuổi và 90/100.000 phụ nữ ở lứa tuổi 35-49. Người ta có thể kết luận rằng ở nhóm tuổi này, phụ nữ có nguy cơ cao bị loét dạ dày. Kết luận này là:

a/ Đúng

b/ Không đúng vì không thể phân biệt được giữa tỷ lệ mới mắc và tỷ lệ hiện mắc.

c/ Không đúng vì các tỷ lệ này đã được dùng để so sánh nam với nữ.

d/ Không đúng vì không nhận biết được ảnh hưởng của thuần tập.

e/ Không đúng vì không có nhóm so sánh hay nhóm đối chứng.

52. Các nghiên cứu dịch tễ học về vai trò của một yếu tố nghi ngờ về bệnh căn có thể là nghiên cứu quan sát hay thực nghiệm. Sự khác nhau cơ bản giữa các nghiên cứu thực nghiệm và quan sát là trong các nghiên cứu thực nghiệm thì:

a/ Nhóm nghiên cứu và nhóm chứng giống nhau về cỡ mẫu.

b/ Nghiên cứu là nghiên cứu tương lai

c/ Nhóm chứng và nhóm nghiên cứu luôn luôn so sánh được với nhau

d/ Nhà nghiên cứu quyết định ai sẽ phơi nhiễm với yếu tố nghi ngờ và ai không.

e/ Nhóm chứng được sử dụng.

53. Những yếu tố nào sau đây là quan trọng nhất đối với tính giá trị của các kết luận rút ra từ một thử nghiệm lâm sàng:

a/ Số những người nhận thuốc điều trị và nhận placebo là như nhau.

b/ Theo dõi được 100% cá thể nghiên cứu.

c/ Phân bố ngẫu nhiên các cá thể nghiên cứu.



d/ Tỷ lệ mới mắc tương đối cao của bệnh trong quần thể nghiên cứu.

e/ Các cá thể nghiên cứu gồm có cả 2 nhóm và các lứa tuổi.

54. Trong một nghiên cứu tương lai về 1 bệnh, nhóm thuần tập được lựa chọn gồm có:

a/ Những người bị bệnh.

b/ Những người không bị bệnh.

c/ Những người có phơi nhiễm với yếu tố mà ta nghiên cứu.

d/ Những người có tiền sử gia đình mắc bệnh.

e/ Những người không phơi nhiễm với yếu tố mà ta nghiên cứu.

55. Trong một nghiên cứu về các yếu tố dẫn đến phạm tội ở thanh thiếu niên người ta thu được kết quả sau:

Bảng 11. CÁC YẾU TỐ ẢNH HƯỞNG ĐẾN PHẠM TỘI Ở THANH THIẾU NIÊN.

	Phần trăm phạm tội theo loại gia đình	P
- Trẻ em ở các gia đình 14 tan vỡ - Trẻ em ở các gia đình 5 hạnh phúc	$p < 0,15$	Phần trăm phạm tội theo tình trạng kinh tế xã hội của gia đình $p < 0,05$
- Trẻ em ở các gia đình có điều kiện kinh tế xã hội cao - Trẻ em ở các gia đình có điều kiện kinh tế xã hội thấp	16 4	

Từ kết quả trên anh/ chị có kết luận gì?

a/ Những gia đình tan vỡ hình như là nguyên nhân gây phạm tội ở thanh thiếu niên.

b/ Cả gia đình tan vỡ và tình trạng kinh tế xã hội cao đều là nguyên nhân gây phạm tội ở thanh thiếu niên.

c/ Ảnh hưởng của gia đình tan vỡ phải được kiểm tra với mỗi tầng lớp kinh tế xã hội.

d/ Không thể có nhận xét trên bởi vì tỷ lệ là quá nhỏ.

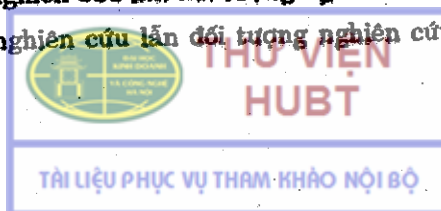
e/ Tình trạng kinh tế xã hội quyết định cả hai yếu tố là gia đình tan vỡ và phạm tội ở thanh thiếu niên.

56. Nghiên cứu "làm mù kép" một loại vaccin là một nghiên cứu trong đó:

a/ Nhóm nghiên cứu nhận vaccin và nhóm chứng nhận placebo

b/ Cả người nghiên cứu lẫn đối tượng nghiên cứu đều không biết bản chất của placebo.

c/ Cả người nghiên cứu lẫn đối tượng nghiên cứu không ai biết ai sẽ nhận vaccin, ai



sẽ nhận placebo.

d/ Cả nhóm nghiên cứu lẫn nhóm chứng không nhóm nào biết ai là nhà nghiên cứu.

e/ Những người ở nhóm chứng không biết những người ở nhóm nghiên cứu.

57. Nhược điểm cơ bản của các nghiên cứu hồi cứu về vai trò của yếu tố bệnh căn nghi ngờ khi so sánh và nghiên cứu tương lai là:

a/ Chúng tốn kém hơn và kéo dài hơn.

b/ Cơ thể có sai số hệ thống trong việc xác định sự có mặt hay không có mặt của yếu tố nguy cơ.

c/ Cơ thể có sai số hệ thống trong việc xác định sự có mặt hay không có mặt của hậu quả bệnh.

d/ Khó chọn nhóm đối chứng hơn.

e/ Gặp khó khăn hơn trong việc đảm bảo khả năng so sánh giữa nhóm bệnh và nhóm chứng.

58. Một nhóm thuần tập phải có:

a/ Cùng năm sinh

b/ Trải qua cùng một thời gian

c/ Cùng nơi sinh sống.

d/ Phơi nhiễm với cùng một bệnh

e/ Có cùng tiền sử bệnh.

59. Trong một nghiên cứu ở Mỹ, qua thăm khám sức khỏe sơ bộ đã phát hiện tỷ lệ bệnh trong mạch vành tim ở nam giới 30-44 tuổi là 5/1000 và ở nữ giới 30-44 tuổi là 5/1000.

Người ta kết luận rằng nguy cơ phát triển bệnh động mạch vành ở nam và nữ là như nhau. Kết luận này là:

a/ Đúng

b/ Không đúng vì không phân biệt giữa tỷ lệ mới mắc và tỷ lệ hiện mắc.

c/ Không đúng vì tỷ suất tỷ lệ (proportional ratio) được sử dụng để ủng hộ kết luận này.

d/ Không đúng vì không nhận biết được hiện tượng thuần tập (cohort phenomenon).

e/ Không đúng vì không có nhóm chứng và nhóm so sánh.

60. Những nhận xét nào dưới đây là ưu điểm của một nghiên cứu hồi cứu.

a/ Không có hay có ít sai chệch trong việc đánh giá phơi nhiễm với một yếu tố nguy cơ.

b/ Hậu quả nhiều bệnh xảy ra sau khi phơi nhiễm với một yếu tố nguy cơ có thể được nghiên cứu một cách dễ dàng.

c/ Loại trừ được sự phụ thuộc vào việc nhớ lại của đối tượng nghiên cứu.

d/ Có khả năng xác định được tỷ lệ mới mắc thật của bệnh.

e/ Nó có thể được sử dụng để nghiên cứu bệnh căn các bệnh hiếm gặp.



61. Một nhà ngoại khoa chỉnh hình công bố số liệu có liên quan tới việc điều trị 204 bệnh nhân gãy cổ xương đùi. Bảng dưới đây trình bày số liệu về tỷ lệ tử vong của các bệnh nhân có liên quan với điều trị ngoại khoa (bó bột mở có đóng đinh nội tủy) hay điều trị bảo tồn theo kinh điển (bó bột bất động).

Bảng 12- SO SÁNH TỶ LỆ TỬ VONG Ở CÁC BỆNH NHÂN ĐIỀU TRỊ NGOẠI KHOA VÀ ĐIỀU TRỊ KINH ĐIỂN.

	Điều trị ngoại khoa	Điều trị kinh điển	Tổng
Số bệnh nhân	139	65	204
Số khỏi	103	34	137
Số chết	36	31	67
Tỷ lệ tử vong	26,5%	47,7%	32,9%

$$\chi^2 = 8,57 \text{ p} < 0,01.$$

Giải thích hợp lý nhất cho kết quả trên là:

a/ Sự giảm có ý nghĩa tỷ lệ tử vong ở những bệnh nhân được điều trị ngoại khoa có nghĩa rằng phẫu thuật là phương pháp điều trị đối với gãy cổ xương đùi.

b/ Số lượng bệnh nhân ở nhóm điều trị ngoại khoa quá lớn và do đó làm sai lệch kết quả.

c/ Không thể lý giải được bởi vì thiếu thông tin về mất khả năng hoạt động của các bệnh nhân ở cả hai nhóm.

d/ Bệnh nhân được điều trị theo phương pháp kinh điển có thể già hơn và yếu hơn so với những người được điều trị ngoại khoa.

e/ Số liệu này không có ý nghĩa gì vì không cho biết tỷ lệ chết sau khi ra viện.

62. năm 1945, người ta đã xác định 1000 phụ nữ làm việc ở nhà máy vẽ radium lên mặt đồng hồ. Tỷ lệ mới mắc ung thư xương ở những phụ nữ này cho đến năm 1975 đã được so sánh với tỷ lệ ung thư xương của 1000 phụ nữ đã làm việc ở tổng đài điện thoại năm 1945. 20 phụ nữ vẽ radium lên mặt đồng hồ và 4 phụ nữ làm ở tổng đài điện thoại đã bị ung thư xương trong thời gian từ 1945 - 1975. Nghiên cứu này là một ví dụ về một:

a/ Nghiên cứu tương lai

b/ Nghiên cứu thực nghiệm.

c/ Thử nghiệm lâm sàng.

d/ Nghiên cứu ngang.

e/ Nghiên cứu hồi cứu.

63. Mọi người trong ngành y đều tin rằng sau khi phẫu thuật ngoại khoa, phụ nữ đẻ con là rất nguy hiểm. Một bác sĩ đã nghiên cứu hồ sơ bệnh án của những phụ nữ đã sống sót sau phẫu thuật. Đối với mỗi phụ nữ những thông tin về số con và số năm sau khi phẫu thuật đã được thu thập. Ông ta đã quan sát và thấy rằng, phụ nữ nào càng đẻ nhiều

con, sau phẫu thuật, người đó càng sống lâu hơn. Ông ta cũng kết luận rằng, quan niệm cũ là sai lầm. Kết luận này không đúng bởi vì:

a/ Cách chọn mẫu của người bác sĩ đã có sai lệch vì nó chỉ bao gồm những phụ nữ sống sót sau phẫu thuật.

b/ Những phụ nữ càng sống lâu hơn sau phẫu thuật, chị ta chắc sẽ càng có nhiều con hơn.

c/ Người bác sĩ này phải kết luận rằng quan niệm cũ đúng.

d/ Những phụ nữ nhiều tuổi hơn có thể đã hết tuổi sinh đẻ.

e/ Những phụ nữ nhiều tuổi hơn có thể không qua khỏi được sau phẫu thuật.

64. Người ta đã tiến hành một thử nghiệm vaccin như sau: 1000 trẻ em 2 tuổi đã được chọn ngẫu nhiên để nhận một loại vaccin phòng một bệnh nào đó và được theo dõi trong 10 năm. Trong số những trẻ em này, 80% trẻ đã không mắc bệnh. Trong những kết luận sau, kết luận nào là đúng nhất có liên quan tới hiệu quả của vaccin.

a/ Vaccin là rất tốt vì tỷ lệ trẻ được gây miễn dịch cao.

b/ Không thể kết luận được vì không nghiên cứu theo dõi những trẻ không được tiêm vaccin.

c/ Vaccin không có hiệu quả cao lắm vì nó phải tạo ra tỷ lệ trẻ có miễn dịch cao hơn nữa.

d/ Không thể kết luận được vì không làm kiểm dịch ý nghĩa thống kê.

e/ Số liệu có ý nghĩa là tỷ lệ mắc bệnh là: $100\% - 80\% = 20\%$

65. trong một nghiên cứu đo lường tần số các triệu chứng phụ do dùng một thuốc điều trị:

a/ Nhóm đối chứng không dùng thuốc là cần thiết để giải thích kết quả.

b/ Nhóm chứng nhận placebo là cần thiết để giải thích kết quả.

c/ Sự đưa thêm nhóm chứng vào có thể làm cho người nghiên cứu lầm lẫn, đặc biệt là nếu incidence của phản ứng phụ thấp.

d/ Sự mong muốn có nhóm chứng phụ thuộc vào loại phản ứng phụ sẽ xảy ra.

e/ Sự mong muốn có nhóm chứng phụ thuộc vào tuổi của đối tượng nghiên cứu.

66. Một nghiên cứu về tuổi và béo phì đã đưa ra kết quả sau:

Bảng 13. TUỔI VÀ BỆNH BÉO PHÌ.

Tuổi	Phần trăm béo phì
20-40	19
40-60	25
60-80	15
>80	5

Người ta kết luận rằng, tuổi càng cao, người ta càng gầy hơn. Kết luận này là:



TRƯỜNG ĐẠI HỌC
HUBT

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ

- a/ Đúng.
- b/ Không đúng, vì tỷ lệ này cần thiết để hỗ trợ sự quan sát.
- c/ Không đúng, vì không có nhóm chứng hay nhóm so sánh.
- d/ Không đúng, vì không thể kết luận được từ các số liệu của nghiên cứu ngang.
- e/ Không đúng, vì tỷ lệ hiện mắc đã được tính, trong khi ấy thì việc tính tỷ lệ mới mắc là cần thiết.

67. Loại nghiên cứu nào sau đây là nghiên cứu hồi cứu.

a/ Nghiên cứu tỷ lệ tử vong/hay mắc bệnh trước đây cho phép ước lượng tỷ lệ bệnh trong tương lai.

b/ Phân tích các nghiên cứu trước đây ở những nơi khác nhau trong những hoàn cảnh khác nhau cho phép đưa ra một giả thuyết dựa trên hiểu biết về tất cả các yếu tố nguy cơ đã biết sẽ dẫn đến bệnh mà ta nghiên cứu.

c/ Thu thập tiền sử và những thông tin khác từ một nhóm bệnh nhân đã biết và từ một nhóm so sánh để xác định tần số tương đối của các đặc trưng nghiên cứu ở những bệnh nhân đó (case)

d/ Nghiên cứu nguy cơ tương đối của ung thư ở những người đàn ông đã bỏ thuốc lá và những người vẫn đang hút thuốc lá.

e/ Điều tra prevalence của bệnh ở những tầng khác nhau của một quần thể.

68. Kỹ thuật ghép cặp được áp dụng trong nghiên cứu bệnh chứng để:

a/ Kiểm soát các biến số đã được biết là có ảnh hưởng đến sự phân bố của bệnh mà ta nghiên cứu ở cả hai nhóm bệnh và nhóm chứng.

b/ Có thể nghiên cứu được ảnh hưởng của các biến số được "đấu"

c/ Kết quả không được quy cho ảnh hưởng của các biến số được đấu.

d/ Kết quả nghiên cứu có thể bao gồm những kết luận về ảnh hưởng của việc đấu các biến số trước khi được lựa chọn.

69. Là một yếu tố có liên quan về nguyên nhân gây bệnh, một yếu tố bệnh cần phải thỏa mãn những điều kiện sau:

a/ Yếu tố phổ biến ở bệnh nhân hơn là ở những người không bị bệnh.

b/ Phơi nhiễm với yếu tố đó phải xảy ra trước khi phát triển bệnh.

c/ Loại trừ yếu tố đó đi sẽ làm giảm nguy cơ mắc bệnh.

d/ Yếu tố đó được thấy ở tất cả bệnh nhân.

70. Nhóm chứng cần thiết trong nghiên cứu bệnh chứng hồi cứu bởi vì.

a/ Chúng được đấu với nhóm bệnh về các yếu tố bệnh cần nghi ngờ.

b/ Chúng được theo dõi để xác định liệu có phát triển bệnh mà ta nghiên cứu hay không.

c/ Chúng làm tăng cỡ mẫu để có thể đạt được ý nghĩa thống kê.

d/ Chúng cho phép đánh giá liệu tần số của một đặc trưng hay phơi nhiễm trong quá khứ ở những bệnh nhân có khác so với ở những người ở nhóm so sánh trong quần thể

71
Mg
N/A
Be

mà không bị bệnh hay không.

71. Năm 1975, người ta đã tiến hành nghiên cứu hồ sơ bệnh án của một bệnh viện sản khoa nhằm xác định những phụ nữ đã sinh ra những trẻ em gái trong thời gian từ 1950 - 1959. Từ những hồ sơ bệnh án này, người ta cũng đã xác định những phụ nữ đã dùng diethylstilbestrol (DES) trong thời kỳ có thai và những phụ nữ khác không dùng DES trong thời kỳ có thai (nhóm chứng). Sau đó, người ta đã tiến hành phỏng vấn con gái của những bà mẹ đó (nếu họ còn sống) và khám sản phụ khoa nhằm phát hiện ung thư âm đạo. Người ta cũng tiến hành nghiên cứu giấy chứng nhận tử vong hay hồ sơ bệnh án của những cô gái đã chết nhằm xác định nguyên nhân tử vong.

Những thuật ngữ nào sau đây mô tả đúng nghiên cứu này:

- a/ Theo dõi quan sát.
- b/ Thục nghiệm.
- c/ Tương lai.
- d/ Hồi cứu.

72. Trong một thử nghiệm lâm sàng, những điều kiện nào dưới đây là phù hợp với việc loại trừ đối tượng nghiên cứu ra khỏi nghiên cứu.

- a/ Những người không tuân thủ việc khám xét sàng lọc trước khi được chọn ngẫu nhiên.
- b/ Những người không tuân thủ chế độ nghiên cứu sau khi đã được chọn ngẫu nhiên vào các nhóm nghiên cứu.
- c/ Những người đã nhận các biện pháp can thiệp trước khi được chọn ngẫu nhiên vào các nhóm nghiên cứu.
- d/ Những người ở nhóm chứng nhận các biện pháp can thiệp qua các bác sĩ tư của họ.

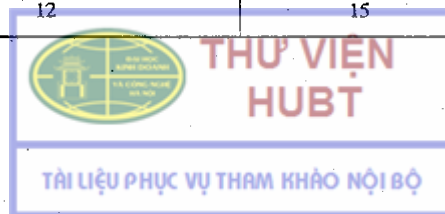
73. Nếu sự pha loãng của một xét nghiệm sàng lọc kết hợp với bệnh, ảnh hưởng chắc chắn nhất của việc giảm tiêu chuẩn pha loãng đối với một kết quả dương tính là phải:

- a/ Tăng độ nhạy và tăng độ đặc hiệu.
- b/ Tăng độ nhạy và giảm độ đặc hiệu.
- c/ Giảm độ nhạy và tăng độ đặc hiệu.
- d/ Tăng độ nhạy và tăng hoặc giảm độ đặc hiệu tùy từng hoàn cảnh.

74. Số bệnh nhân ung thư mới (trên 100.000 dân) một năm ở nam giới 55 - 59 tuổi ở hai quần thể như sau:

Bảng 14- UNG THƯ Ở NAM GIỚI TUỔI TỪ 55-59.

	Số trường hợp mới mắc trên 100.000 nam giới/năm	
	Quần thể A	Quần thể B
- Phổi	40	55
- Đại tràng và trực tràng	20	30
- Tuyến tiền liệt	12	15



Người ta kết luận rằng nam giới tuổi 55-59 ở quần thể B có xu hướng mắc ung thư phổi, đại tràng và trực tràng, tuyến tiền liệt cao hơn đàn ông 55-59 tuổi ở quần thể A. Kết luận này là:

- a/ Đúng
- b/ Không đúng vì không phân biệt được số mới mắc và số hiện mắc.
- c/ Không đúng vì một mình tỷ lệ tử vong theo tỷ lệ không cho ta ước lượng nguy cơ.
- d/ Không đúng vì không chuẩn hóa được sự khác nhau theo cấu trúc tuổi của hai quần thể này.
- e/ Không đúng vì không có nhóm chứng và nhóm so sánh.

75. Trong một điều tra về một vụ dịch ngộ độc thực phẩm ở một bữa tiệc lớn, những người ăn thịt bò nướng và ăn nấm có tỷ lệ tấn công cao. Kết hợp cả hai loại thực phẩm này được trình bày ở bảng sau.

Bảng 15- TỶ LỆ TẤN CÔNG ĐỐI VỚI VIỆC ĂN HAI LOẠI THỰC PHẨM.

	Ăn nấm			Không ăn nấm		
	Tổng số	Số ốm	Tỷ lệ tấn công (%)	Tổng số	Số ốm	Tỷ lệ tấn công (%)
Ăn thịt bò nướng	150	105	70	72	2	3
Không ăn thịt bò nướng	42	33	78	26	0	0

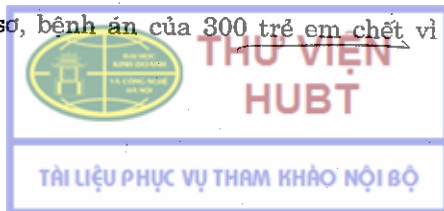
Loại thực phẩm gây ngộ độc thực phẩm là:

- a/ Nấm.
- b/ Thịt bò nướng.
- c/ Kết hợp ăn cả thịt bò nướng và nấm.
- d/ Ăn nấm mà không ăn thịt bò nướng.
- e/ Chẳng phải do thịt bò nướng, chẳng phải do nấm.

76. Trong một quảng cáo về việc cho thêm hương vị quả mâm xôi vào aureomycin người ta nói rằng, trong 1000 trẻ em bị bệnh nhiễm khuẩn đường hô hấp trên được điều trị bằng aureomycin có cho thêm hương vị quả mâm xôi, 970 trẻ đã không có triệu chứng trong vòng 72 giờ. Người ta kết luận rằng aureomycin có cho hương vị quả mâm xôi là thuốc điều trị tốt cho những trẻ em bị nhiễm khuẩn đường hô hấp cấp. Kết luận này là:

- a/ Đúng.
- b/ Không đúng vì sự so sánh không dựa trên tỷ lệ.
- c/ Không đúng vì không sử dụng nhóm chứng hay nhóm so sánh.
- d/ Không đúng vì không làm kiểm định ý nghĩa thống kê.
- e/ Không đúng vì ảnh hưởng thuần tập có thể xảy ra.

77. Nghiên cứu hồ sơ, bệnh án của 300 trẻ em chết vì một tai nạn ngộ độc ở thành



phố A cho thấy rằng số trẻ em có điều kiện kinh tế xã hội thấp chết nhiều gấp 5 lần trẻ em có điều kiện kinh tế xã hội cao. Người ta kết luận rằng tai nạn ngộ độc ở trẻ em của các gia đình có điều kiện kinh tế xã hội thấp gấp 5 lần so với gia đình có điều kiện kinh tế xã hội cao. Kết luận này là:

- a/ Đúng.
- b/ Không đúng vì các nhóm so sánh và nhóm nghiên cứu là không thích hợp và không thể so sánh được.
- c/ Không đúng vì có sự giải thích không đúng của kiểm định ý nghĩa thống kê.
- d/ Không đúng vì sự so sánh không dựa trên tỷ lệ.
- e/ Không đúng vì sự chênh lệch do người quan sát và người phỏng vấn có thể ảnh hưởng đến kết quả.

78. Cộng đồng A và cộng đồng B đều có tỷ lệ chết thô do bệnh động mạch vành là 4/1000. Tỷ lệ tử vong của bệnh động mạch vành tim đã được chuẩn hóa theo tuổi là 5/1000 đối với cộng đồng A và 3/1000 đối với cộng đồng B. Người ta có thể kết luận rằng:

- a/ Cộng đồng A có dân chúng trẻ hơn cộng đồng B.
- b/ Cộng đồng A có dân chúng già hơn cộng đồng B. $x = 5 + 5 \cdot (1 - x) = 4$
 $- 5x$
- c/ Hai cộng đồng có phân bố tuổi như nhau.
- d/ Chẩn đoán ở cộng đồng A chính xác hơn cộng đồng B
- e/ Chẩn đoán ở cộng đồng B chính xác hơn cộng đồng A.

79. Một nghiên cứu hồi cứu có các đặc trưng sau nhưng trừ một đặc trưng.

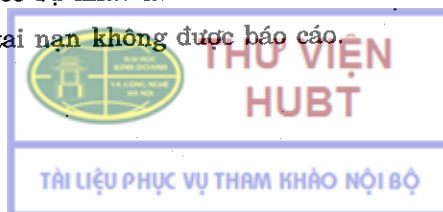
- a/ Nó tương đối không tốn kém.
- b/ Nguy cơ tương đối có thể ước lượng được từ các kết quả.
- c/ Tỷ lệ mới mắc có thể tính được.
- d/ Người ta có thể chọn nhóm chứng không bị bệnh.
- e/ Đánh giá các phơi nhiễm trong quá khứ có thể bị sai lệch.

80. Nguy cơ mắc bệnh có thể được đo lường bằng:

- a/ Tỷ lệ mới mắc.
- b/ Tỷ lệ mới mắc nhân với thời gian trung bình của bệnh.
- c/ Tỷ lệ mới mắc chia cho tỷ lệ hiện mắc.
- d/ Tỷ lệ hiện mắc.
- e/ Tỷ lệ hiện mắc nhân với thời gian trung bình của bệnh.

81. Nghiên cứu về an toàn giao thông ở một thành phố cho thấy rằng 61% những người bị tai nạn năm ngoái đã có kinh nghiệm trên 10 năm lái xe, 21% có kinh nghiệm 10 năm lái xe, và 17% có kinh nghiệm 1-5 năm lái xe. Các chuyên gia về giao thông kết luận rằng kinh nghiệm lái xe làm cho người lái xe tự mãn và cầu thả hơn. Nhận xét nào sau đây nói lên lý do tại sao kết luận này không đúng:

- a/ Tỷ lệ đã không được chuẩn hóa theo sự khác nhau về tuổi.
- b/ Số liệu không đầy đủ vì có nhiều tai nạn không được báo cáo.



50%

$$\frac{I_e - I_o}{I_e}$$

$$\frac{I_e}{I_o}$$

c/ Cần phải so sánh với số liệu tương tự ở những người lái xe không bị tai nạn.

d/ Không làm kiểm định ý nghĩa thống kê.

e/ Người ta đã tính tỷ lệ hiện mắc chứ không phải là tỷ lệ mới mắc.

82. Mức độ kết hợp giữa một yếu tố nguy cơ và bệnh được đo lường tốt nhất bằng:

a/ Thời kỳ ủ bệnh.

b/ Tỷ lệ mới mắc của bệnh trong toàn bộ dân chúng.

c/ Sự phổ biến của yếu tố nguy cơ.

d/ Nguy cơ quy thuộc.

e/ Nguy cơ tương đối.

$$\begin{cases} x - a + (1-x)b = 4 \\ a \cdot \frac{1}{3} + (1-x)b \cdot \frac{2}{3} = 5 \end{cases}$$

83. Tần số tương đối của các trường hợp ung thư ở các vị trí cơ quan ở hai quần thể được trình bày ở bảng sau:

$$\frac{2}{3} = \frac{1}{3} = 5$$

Bảng 16- TẦN SỐ TƯƠNG ĐỐI CỦA CÁC UNG THƯ Ở HAI QUẦN THỂ

Vị trí ung thư	Phần trăm so với tổng số	
	Quần thể A	Quần thể B
- Phổi	10,0	6,7
- Vú	30,0	20,0
- Tử cung	25,0	16,7
- Các nơi khác	35,0	56,6
- Tổng các vị trí	100,0	100,0

Người ta kết luận rằng quần thể A có xu hướng bị ung thư phổi, vú và tử cung cao hơn so với quần thể B. Kết luận này là:

a/ Đúng.

b/ Không đúng vì không thể phân biệt số mới mắc và số hiện mắc.

c/ Không đúng vì người ta đã dùng tỷ suất tỷ lệ để hỗ trợ kết luận này.

d/ Không đúng vì không thể biết được hiện tượng thuần tập (cohort phenomenon)

e/ Không đúng vì không có nhóm chứng hay nhóm so sánh.

84. Khi điều tra một vụ dịch ỉa chảy người ta thấy rằng tỷ lệ bệnh nhân ăn ở cửa hàng A là 85%, cửa hàng B là 15%, cửa hàng C là 55%, và tỷ lệ dùng nước công cộng là 95%. Liên quan với các số liệu trên. Kết luận nào sau đây là có giá trị.

a/ Nguồn gốc của vụ dịch là ở cửa hàng A vì nó có tỷ lệ bệnh nhân cao nhất trong các cửa hàng.

b/ Nguồn gốc của vụ dịch không phải là ở cửa hàng B vì nó có tỷ lệ bệnh thấp nhất.

c/ Nguồn gốc của vụ dịch là cung cấp nước vì có tỷ lệ bệnh nhân cao nhất.

d/ Nguồn gốc của vụ dịch có thể là hoặc cửa hàng A, hoặc cửa hàng C, hoặc cung cấp nước.

e/ Không kết luận được vì không tiến hành so sánh những nguồn có phơi nhiễm và



điều tra khác cũng ở Anh, 30 trong số 99 bệnh nhân bại liệt đã trả lời là có nuôi vẹt trong nhà. Người ta kết luận rằng có một mối liên hệ giữa sự có mặt của vẹt trong nhà và sự xuất hiện bệnh bại liệt ở những người trong gia đình đó. Kết luận này là:

- a/ Đúng.
- b/ Không đúng vì không phân biệt được tỷ lệ mới mắc và tỷ lệ hiện mắc.
- c/ Không đúng vì người ta dùng tỷ suất tỷ lệ để hỗ trợ kết luận này chứ không phải là tỷ lệ.
- d/ Không đúng vì không thể biết được hiện tượng thuần tập (cohort phenomenon).
- e/ Không đúng vì không có nhóm chứng hay nhóm so sánh.

90. Một cuộc tranh luận xảy ra giữa những người đề nghị điều trị bằng thuốc và điều trị bằng đọc cho những bệnh nhân bị chứng loạn đọc (dyslexia). Để hỗ trợ cho quan điểm của mình, một nhóm công bố rằng "Trong 119 bệnh nhân bị chứng loạn đọc, 97 bệnh nhân thấy có tiến triển tốt sau khi được áp dụng biện pháp đọc điều trị". Người ta kết luận rằng, ở bệnh nhân bị chứng loạn đọc, biện pháp đọc điều trị có thể được áp dụng. Kết luận này là:

- a/ Đúng.
- b/ Không đúng vì sự so sánh không dựa trên tỷ lệ.
- c/ Không đúng vì không sử dụng nhóm chứng hay nhóm so sánh.
- d/ Không đúng vì không làm kiểm chứng hay nhóm so sánh.
- e/ Không đúng vì không làm kiểm định ý nghĩa thống kê.
- e/ Không đúng vì có thể có ảnh hưởng thuần tập (cohort effect).

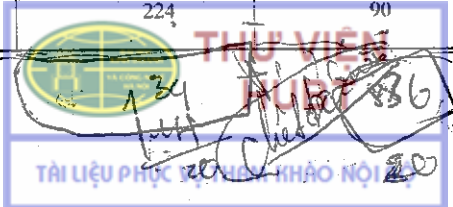
91. Một nhà nghiên cứu xác định được hệ số tương quan giữa nồng độ triglyceride và mức độ xơ mỡ động mạch là +1,67. Trên cơ sở số liệu này ta có thể kết luận là:

- a/ Nồng độ triglyceride là chỉ số dự đoán tốt xơ mỡ động mạch.
- b/ Nồng độ triglyceride không phải là chỉ số dự đoán tốt xơ mỡ động mạch.
- c/ Nồng độ triglyceride cao gây xơ mỡ động mạch.
- d/ Xơ mỡ động mạch gây nên tăng nồng độ triglyceride.
- e/ Người nghiên cứu này đã xác định không đúng hệ số tương quan.

Bảng dưới đây trình bày sự sống sót của các bệnh nhân ung thư sau khi điều trị một thuốc mới.

Bảng 17- SỰ SỐNG SỐT CỦA BỆNH NHÂN UNG THƯ SAU KHI ĐIỀU TRỊ THUỐC.

Bệnh nhân Khoảng thời gian	Bệnh nhân sống lúc bắt đầu khoảng thời gian	Bệnh nhân chết trong khoảng thời gian	Tỷ lệ những người sống lúc bắt đầu khoảng thời gian nhưng chết trong khoảng thời gian
0-3 tháng	1000	300	0,3
3-6 tháng	700	140	0,2
6-12 tháng	560	112	0,2
12-18 tháng	448	224	0,5
18-24 tháng	224	90	0,4



Handwritten calculations: $\frac{1000 - 880}{1000} = \frac{120}{1000} = 0,12$

Sử dụng bảng này cho các bài từ 94 đến 96.

92. Khả năng mà một bệnh nhân sống sót sau hai năm điều trị sẽ sống sót sau 6 tháng là:

a/ $134/560 = 0,24$

b/ $426/560 = 0,76$

c/ $134/700 = 0,19$

d/ $426/700 = 0,61$

e/ Không thể xác định được từ số liệu này.

93. Khả năng mà một bệnh nhân sống sót 3 tháng sau khi bắt đầu điều trị là:

a/ 0,3

b/ 0,4

c/ 0,5

d/ 0,6

e/ 0,7

94. Khả năng mà một bệnh nhân chết trong 2 năm sau điều trị là:

a/ $90/224=0,4$

b/ $866/1000=0,866$

c/ $90/1000 = 0,09$

d/ $224/1000 = 0,224$

e/ Không thể xác định được từ số liệu này.

95. Người ta đã tiến hành đánh giá một phương pháp điều trị ngoại khoa mới để làm giảm số mới mắc của các biến chứng sau phẫu thuật. Tỷ lệ biến chứng là 40% ở 25 bệnh nhân được điều trị theo phương pháp mới và 60% ở 20 bệnh nhân được điều trị theo phương pháp cũ. Sự khác biệt này là không có ý nghĩa thống kê. Do đó, ta có thể kết luận rằng:

a/ Phương pháp điều trị có hiệu quả làm giảm các biến chứng sau mổ.

b/ Phương pháp điều trị mới không có hiệu quả làm giảm các biến chứng sau mổ.

c/ Có sai lệch trong chọn mẫu.

d/ Kết quả có ý nghĩa về lâm sàng.

e/ Không đủ bằng chứng để cho rằng phương pháp điều trị mới có hiệu quả làm giảm các biến chứng sau mổ.

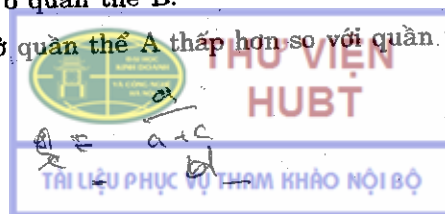
96. Một xét nghiệm sàng lọc với độ nhạy và độ đặc hiệu đã biết được áp dụng ở hai cộng đồng. Tỷ lệ hiện mắc của bệnh được sàng lọc là 10% ở quần thể A và 1% ở quần thể B. Nhận xét nào sau đây là đúng.

a/ Tỷ lệ phần trăm âm tính giả ở quần thể A nhỏ hơn so với quần thể B.

b/ Tính đặc hiệu ở quần thể A thấp hơn ở quần thể B.

c/ Độ tin cậy ở quần thể A cao hơn ở quần thể B.

d/ Tỷ lệ phần trăm dương tính giả ở quần thể A thấp hơn so với quần thể B.



Handwritten marks and a table:

#	10	A
---	----	---

e/ Độ nhảy ở quần thể A cao hơn so với quần thể B.

97. Người ta đã thử tiến hành hai kế hoạch điều trị theo dõi những người mới được phát hiện tăng huyết áp ở một cộng đồng. Kế hoạch A được thực hiện ở các quận, phía đông và nam, trong khi đó, kế hoạch B được thực hiện ở các quận phía tây và phía bắc. 3 năm sau khi bắt đầu thực hiện các kế hoạch này, người ta đã thu thập được các số liệu sau:

Bảng 18- TỶ LỆ THÀNH CÔNG CỦA HAI KẾ HOẠCH ĐIỀU TRỊ TĂNG HUYẾT ÁP.

	Kế hoạch A	Kế hoạch B
- Số người tăng huyết áp	2200	1900
- Tỷ lệ phần trăm những người tăng huyết áp được điều trị khỏi	41	45

Sự khác nhau về tỷ lệ điều trị khỏi ở hai kế hoạch này là có ý nghĩa thống kê ($p < 0,01$). Tuy nhiên, các nhân viên y tế đã quyết định không thay đổi chuyển sang kế hoạch B ở các quận phía đông và phía nam vì mức độ khác nhau là quá nhỏ. Điều này ngụ ý là:

a/ Họ cho là sự khác nhau về tỷ lệ điều trị khỏi chỉ là do ngẫu nhiên.

b/ Họ cảm thấy cỡ mẫu quá nhỏ nên không thể biện hộ cho quyết định chuyển sang kế hoạch điều trị B.

c/ Họ cảm thấy rằng giá trị p quá nhỏ nên không thể biện hộ cho quyết định chuyển sang kế hoạch điều trị B.

d/ Họ đã phân biệt được ý nghĩa thống kê và tầm quan trọng về mặt thực hành của sự khác nhau về tỷ lệ điều trị khỏi.

e/ Họ cảm thấy rằng, tỷ lệ điều trị khỏi phải cao hơn 50%.

98. Người ta tiến hành xét nghiệm nồng độ cholesterol máu ở 4 người đàn ông khỏe mạnh. Xác suất mà cả 4 người đàn ông này có nồng độ cholesterol máu ở dưới 10 phần trăm phân bố tần số (percentile), cholesterol máu của những người khỏe mạnh là:

a/ 0,4

b/ $1 - (0,1)^4$

c/ $(0,1)^4$

d/ $(0,9)^4$

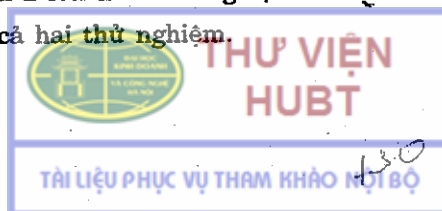
e/ Không thể xác định được từ số liệu trên.

99. trong một chương trình phát hiện bệnh đái đường. Mức sàng lọc đối với glucose máu ở thử nghiệm 1 là 160 mg/100ml và ở thử nghiệm 2 là 130mg/100ml. Độ nhảy :

a/ Ở thử nghiệm 1 lớn hơn thử nghiệm 2.

b/ Ở thử nghiệm 2 lớn hơn thử nghiệm 1.

c/ Như nhau ở cả hai thử nghiệm.



Handwritten notes and signatures, including the number 180 and a signature.

123. Tỷ lệ mới mắc tại một thời điểm của một bệnh ở nam giới gấp 5 lần tỷ lệ này ở phụ nữ, những tỷ lệ mới mắc bệnh lại cho thấy không có sự khác nhau về giới tính. Sự giải thích, thích hợp nhất ở đây là:

- a/ Tỷ lệ chết vì bệnh này ở phụ nữ lớn hơn.
- b/ Tỷ lệ chết vì bệnh này ở nam giới cao hơn.
- c/ Thời gian kéo dài trung bình của bệnh này ở phụ nữ lớn hơn.
- d/ Phụ nữ được điều trị không đầy đủ đối với bệnh này.
- e/ Chẩn đoán thường bỏ sót nhiều hơn ở phụ nữ.

$$P = D \cdot I \uparrow (D) \text{ cũ}$$

Hiện mới

$$P = D \cdot I \text{ cũ}$$

MỤC LỤC

	<i>Trang</i>
Lời nói đầu	3
X Tỷ lệ mắc (HT Hoa)	5
X Tỷ lệ tử vong (HT Hoa, DD Thiện)	12
X Đo lường nguy cơ (HT Hoa)	22
Biến sinh học (DD Thiện)	27
X Độ tập trung và độ phân tán (DD Thiện)	37
Ý nghĩa thống kê-Sai số chuẩn-Tự đối chiếu cặp (DD Thiện)	58
X Xác suất (PH Dũng)	71
X Kỹ thuật sàng tuyển phát hiện hàng loạt (PH Dũng)	75
Kết hợp và căn nguyên (NT Hiến)	80
Điều tra xử lý dịch (HT Hoa)	87
X Nghiên cứu bệnh chứng (HT Hoa, DD Thiện)	99
X Nghiên cứu thuần tập (HT Hoa, DD Thiện)	108
X Thử nghiệm lâm sàng ngẫu nhiên (NT Hiến, DD Thiện)	117
Điều tra nghiên cứu sức khỏe cộng đồng - Nghiên cứu ngang (DD Thiện)	126
Lập kế hoạch và tổ chức một cuộc điều tra cộng đồng (NT Hiến)	133
Sử dụng thông tin sẵn có (NT Hiến)	149
Chọn mẫu (DD Thiện)	159
Cỡ mẫu (DD Thiện)	164
Mẫu ngẫu nhiên đơn (DD Thiện)	173
Mẫu tầng (DD Thiện)	177
Mẫu chùm (DD Thiện)	181
Thiết kế bộ câu hỏi, kỹ thuật phỏng vấn và điền bộ câu hỏi (LV Anh)	188
Thu thập và trình bày kết quả nghiên cứu (TV Chí)	201
Bài tập tự đánh giá (NT Hiến, HT Hoa)	214



**THƯ VIỆN
HUBT**

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ



**THƯ VIỆN
HUBT**

TÀI LIỆU PHỤC VỤ THAM KHẢO NỘI BỘ