

LẬP KẾ HOẠCH VÀ THỰC HIỆN PHÂN TÍCH DỮ LIỆU

DESIGNING PLAN
AND PERFORMING DATA ANALYSIS

MỤC TIÊU (OBJECTIVES)

Sau khi đọc xong bài này, học viên sẽ:

1. Lập được kế hoạch phân tích cho dữ liệu thu thập
2. Thực hành nhập liệu phần mềm Epidata
3. Thực hiện phân tích các dữ liệu dựa trên phần mềm stata

1. Lập kế hoạch phân tích (Designing plan)

a/ Xây dựng ma trận đo lường:

+ Phân loại biến

+ Mục tiêu nghiên cứu

Ví dụ ma trận đo lường bộ câu hỏi WHO

Số thứ tự người phỏng vấn	Đặc điểm đối tượng					
	Tuổi	Giới	TDHV	Nghề	CN
1						
2						
3						
.....						

**Số thự tự
người
phỏng vấn**

Các yếu tố nguy cơ

Hút thuốc

Uống rượu

Dinh dưỡng

Hoạt động
thể lực

.....

Q1

Q2

Q3

Q4

Q5

Q6

Q7

Q8

1

2

.....

1. Lập kế hoạch phân tích (tt)

b/ Dự trù các kiểm định cho bộ câu hỏi:

- + Mục tiêu nghiên cứu
- + Thiết kế nghiên cứu
- + Bảng tóm tắt kiểm định

Chọn kiểm định phù hợp đặc điểm biến số

Biến phụ thuộc	Biến độc lập		
	Nhị giá	Danh định (hoặc thứ tự)	Định lượng, đa biến (hoặc thứ tự)
Định lượng phân phối bình thường	T test	Anova một chiều	Tương quan Pearson Hồi qui tuyến tính
Định lượng phân phối không bình thường, không cần kiểm định giả thuyết	Wilcoxon rank sum test (Mann – Whitney)	Kruskal – Wallis	Tương quan Spearman
Nhị giá	X^2 OR, RR	X^2 X^2 khuynh hướng	Hồi qui logistic
Sống còn	Wilcoxon tổng quát Logrank	Wilcoxon tổng quát Logrank	Hồi qui Cox

Bảng phân loại biến và kiểm định bộ câu hỏi WHO

Biến phụ thuộc	Biến độc lập			
	Tuổi	Giới	TDHV
Bệnh (nhị giá)	Hồi qui logistic	χ^2	χ^2 Tendency χ^2	
.....				

Bài tập

Lớp chia làm 8 nhóm thảo luận 20 phút các vấn đề liên quan đến bộ câu hỏi WHO như sau:

- 1/ Liệt kê biến độc lập và biến phụ thuộc
- 2/ Xác định các loại kiểm định phù hợp

2. Nhập liệu bằng phần mềm Epidata

Lớp thực hành nhập liệu bộ câu hỏi WHO (phỏng vấn hộ gia đình và phỏng vấn cá nhân). Mỗi học viên phải tự mình thao tác nhập liệu 1 bộ câu hỏi mẫu trong vòng 20 phút.

3. Thống kê mô tả (Descriptive statistics)

Biến định lượng	Biến định tính
<ul style="list-style-type: none">- Trung bình (Mean)- Trung vị (Median)- Yếu vị (Mode)- Giá trị tối đa (Maximum)- Giá trị tối thiểu (Minimum)- Độ lệch chuẩn (Standard deviation)- Khoảng tứ phân vị (25%, 50%, 75%)	<ul style="list-style-type: none">- Tần suất (Frequency)- Tỷ lệ phần trăm (Percentage)

Lệnh stata

1. Thống kê mô tả biến định tính:

tab1 varname

Ví dụ: tab1 gioi honnhan dantoc....

2. Thống kê mô tả biến định lượng:

summarize varname

Ví dụ: summarize tuoi cannang chieucao...

3. Thống kê phân tầng theo nhóm:

bysort indepvar: sum depvar

Ví dụ: bysort gioi: sum suckhoe

N=

Tần số

Tỷ lệ (%)

Giới

Nam

Nữ

Trình độ học vấn

Tiểu học

THCS

THPT

Đại học

Sau đại học

.....

N=

Tuổi

Cân nặng

.....

Mean

Median

Mode

Maximum

Minimum

Sd

.....

4. Thống kê phân tích (Inferential statistics)

- T test
- Chi bình phương
- Phi tham số Wilcoxon
- Phi tham số Kruskal – Wallis
- Hệ số tương quan Pearson
- Phương trình hồi qui tuyến tính
- Anova
-

Thống kê phân tích biến định lượng bằng stata

1. So sánh 2 phương sai:

sdtest depvar, by(indepvar)

Ví dụ: `sdtest cannang, by(gioi)`

2. So sánh 2 trung bình sử dụng t test phương sai đồng nhất:

ttest depvar, by(indepvar)

Ví dụ: `ttest cannang, by(gioi)`

3. So sánh 2 trung bình sử dụng t test phương sai không đồng nhất:

ttest depvar, by(indepvar) unequal

Ví dụ: `ttest cannang, by(gioi) unequal`

Thống kê phân tích biến định lượng bằng stata (tt)

4. So sánh 2 trung bình sử dụng phép kiểm phi tham số Mann Whitney:

ranksum depvar, by(indepvar)

Ví dụ: ranksum cannang, by(gioi)

5. So sánh trung bình của 1 biến định lượng ở nhiều nhóm: phân tích Anova:

oneway depvar indepvar,tab

Ví dụ: oneway cannang nhomtuoi,tab

Nếu Anova một chiều không thỏa, dùng kiểm định sau

kwallis varname, by(groupvar)

Ví dụ: kwallis cannang, by(nhomtuoi)

Thống kê phân tích biến định lượng bằng stata (tt)

6. Xác định hệ số tương quan

Ví dụ: xác định hệ số tương quan giữa cân nặng với nhóm tuổi

pwcorr cannang tuoi, sig star(5)

7. Xây dựng phương trình hồi qui

Ví dụ xây dựng phương trình hồi qui của cân nặng theo tuổi

regress cannang tuoi

Thống kê phân tích biến định tính bằng stata

1. Kiểm định chi bình phương:

tab depvar indepvar, row chi

Ví dụ: tab suckhoe gioi, row chi

Nếu vọng trị < 5 dùng kiểm định Fisher thay thế

tab depvar indepvar, expect

2. Tính OR trong nghiên cứu bệnh chứng:

cc depvar indepvar

Ví dụ: cc suckhoe gioi

3. Tính RR/PR:

cs depvar indepvar

Ví dụ: cs suckhoe gioi

Thống kê phân tích biến định tính bằng stata (tt)

4. Xác định biến gây nhiễu hay không:

cc depvar indepvar, by(factorvar)

Ví dụ: cc suckhoe gioi, by(nhomtuoi)

5. Khử biến gây nhiễu:

mhodds depvar indepvar factorvar

Ví dụ : mhodds suckhoe cannang gioi

6. Kiểm tra mối liên quan có tính chất khuynh hướng hay không dùng lệnh sau:

nptrend depvar, by(indepvar)

$p_{\text{trend}} < 0.05 \rightarrow$ có tính chất khuynh hướng

Đặc tính mẫu N=	Hút thuốc		Không hút thuốc		P (chung)	P (từng tầng)	PR (KTC 95%)
	N	%	N	%			

Nhóm tuổi

25-34

35-44

45-54

>55

Trình độ học vấn

Tiểu học

THCS

THPT

Đại học

Sau đại học

.....