

www.mientayvn.com

Dịch tiếng anh chuyên ngành khoa học tự nhiên và kỹ thuật.

Dịch các bài giảng trong chương trình học liệu mở của học viện MIT, Yale.

Tìm và dịch tài liệu phục vụ cho sinh viên làm seminar, luận văn.

Tại sao mọi thứ đều miễn phí và chuyên nghiệp ???

Trao i tr c tuy n t i:

www.mientayvn.com/chat_box_toan.html

CHƯƠNG 1

**DỮ LIỆU
và
THỐNG KÊ**

1

NỘI DUNG CHÍNH

- Thống kê và các ứng dụng trong kinh doanh và kinh tế
- Dữ liệu
- Nguồn dữ liệu
- Thống kê mô tả
- Thống kê suy diễn

2

THỐNG KÊ VÀ ỨNG DỤNG TRONG KINH DOANH VÀ KINH TẾ

Thống kê là một *Nghệ thuật và Khoa học về:*

- Thu thập
- Phân tích
- Trình bày
- Và giải thích DỮ LIỆU

3

THỐNG KÊ VÀ ỨNG DỤNG TRONG KINH DOANH VÀ KINH TẾ

Ứng dụng trong kinh doanh và kinh tế:

- Các ứng dụng của thống kê rất hiển nhiên trong nhiều lãnh vực
- Thống kê được sử dụng để:
 - Thông báo cho công chúng
 - Dự báo cho việc lập kế hoạch và ra quyết định

4

THỐNG KÊ VÀ ỨNG DỤNG TRONG KINH DOANH VÀ KINH TẾ

Ứng dụng trong kinh doanh và kinh tế:

- Các lãnh vực kinh doanh và kinh tế dựa trên các kỹ thuật và thông tin thống kê:
 - Kế toán
 - Tài chánh
 - Tiếp thị
 - Sản xuất
 - Kinh tế

5

THỐNG KÊ VÀ ỨNG DỤNG TRONG KINH DOANH VÀ KINH TẾ

Các phần mềm thống kê so với Excel

- Các phần mềm thống kê thường là “Hộp đen”
 - **MINITAB**: là một phần mềm thống kê có nguồn gốc được thiết kế tại Đại học Penn State đặc biệt dành cho sinh viên.
 - **SAS**: Statistical Analysis System – Hệ thống phân tích thống kê
 - **SPSS**: Statistical Package for the Social Science – Phần mềm thống kê cho khoa học xã hội

6

THỐNG KÊ VÀ ỨNG DỤNG TRONG KINH DOANH VÀ KINH TẾ

Các phần mềm thống kê so với Excel

Sử dụng Excel để phân tích thống kê bởi vì:

- Excel sẵn có ở các văn phòng
- Excel đủ mạnh để giải quyết các vấn đề thống kê thường gặp
- Người sử dụng có thể hiểu được ý nghĩa của các vấn đề thống kê

Các nhà quản lý và ra quyết định thành công là những người có thể hiểu và sử dụng các thông tin một cách hiệu quả nhất

DỮ LIỆU

Dữ liệu

- **Dữ liệu** là các sự kiện và con số được thu thập, phân tích và tổng kết để trình bày và giải thích
- **Tập dữ liệu** là tất cả các dữ liệu được thu thập cho một nghiên cứu cụ thể
- **Thang đo**
- **Dữ liệu định tính so với định lượng**
- **Dữ liệu chéo so với chuỗi thời gian**

DỮ LIỆU

Các phân tử, các biến và các quan sát

- **Phân tử** là toàn bộ thực thể dựa vào đó dữ liệu được thu thập
- **Biến** là các đặc tính được quan tâm đối với các phân tử
- **Quan sát** là tập các đại lượng đo lường được thu thập đối với một phân tử cụ thể

DỮ LIỆU

Company	Stock Exchange	Annual Sales(\$M)	Earn/ Sh.(\$)
Dataram	AMEX	73.10	0.86
EnergySouth	OTC	74.00	1.67
Keystone	NYSE	365.70	0.86
LandCare	NYSE	111.40	0.33
Psychemedics	AMEX	17.60	0.13

10

Câu hỏi?

Tham khảo Bảng 1.1:
BỘ DỮ LIỆU CỦA 5 LOẠI CHỨNG KHÓA:

- Có bao nhiêu phần tử trong bộ dữ liệu này ?
- Có bao nhiêu biến trong bộ dữ liệu này?
- Có bao nhiêu quan sát trong bộ dữ liệu này?
- Có bao nhiêu số lượng dữ liệu trong bộ dữ liệu này?

11

DỮ LIỆU

Thang đo
Xác định lượng thông tin có trong dữ liệu và chỉ ra sự tổng kết dữ liệu và phân tích thống kê nào là thích hợp nhất

- Thang đo chỉ danh
- Thang đo thứ tự
- Thang đo khoảng
- Thang đo tỉ lệ

12

DỮ LIỆU

▪ **Thang đo chỉ danh**

Sử dụng nhãn hiệu hoặc tên để nhận dạng một thuộc tính của phân tử → bằng số hoặc không bằng số

▪ **Thang đo thứ tự**

Có đặc tính của thang đo chỉ danh và có thể dùng để sắp hạng hoặc thứ tự dữ liệu → bằng số hoặc không bằng số

▪ **Thang đo khoảng**

Có đặc tính của thang đo thứ tự và khoảng cách giữa các quan sát được diễn tả dưới dạng các đơn vị đo lường cố định → luôn luôn bằng số

▪ **Thang đo tỉ lệ**

Có đặc tính của thang đo khoảng và tỉ lệ của 2 giá trị là có ý nghĩa → luôn luôn bằng số
(Chứa giá trị **Zero** → Có nghĩa là không có gì)

13

DỮ LIỆU

Dữ liệu định tính so với định lượng

▪ **Dữ liệu định tính**

- Dữ liệu định tính là các nhãn hiệu hay tên được dùng để nhận dạng và đặc trưng cho mỗi phân tử
- Biến định tính là biến với dữ liệu định tính
- Dữ liệu định tính sử dụng thang đo chỉ danh hoặc thang đo thứ tự; có thể đo bằng số hoặc không bằng số

14

DỮ LIỆU

Dữ liệu định tính so với định lượng

▪ **Dữ liệu định lượng**

- Dữ liệu định lượng là dữ liệu cho biết số lượng bao nhiêu của một đại lượng nào đó
- Biến định lượng là biến với dữ liệu định lượng
- Dữ liệu định tính sử dụng thang đo khoảng hoặc thang đo tỷ lệ; luôn đo bằng số

15

DỮ LIỆU

Dữ liệu định tính so với định lượng

- Sự khác nhau giữa dữ liệu định lượng và định tính
 - Các phép tính số học thông thường chỉ có ý nghĩa đối với dữ liệu định lượng
 - Tuy nhiên, khi dữ liệu định tính được ghi nhận như các giá trị bằng số thì các phép tính số học sẽ cho ra các kết quả không có ý nghĩa

16

Câu hỏi ?

Hãy phát biểu xem các biến sau đây biến nào là biến định tính, biến nào là biến định lượng và hãy chỉ ra thang đo thích hợp cho mỗi biến.

- Tuổi
- Giới tính
- Thứ hạng trong lớp
- Nhiệt độ
- Thu nhập

17

DỮ LIỆU

Dữ liệu chéo và dữ liệu chuỗi thời gian

- Dữ liệu chéo là các dữ liệu được thu thập trong cùng hay gần cùng một thời điểm
- Dữ liệu chuỗi thời gian là các dữ liệu được thu thập trong các thời điểm liên tiếp nhau

18

NGUỒN DỮ LIỆU

Nguồn dữ liệu có thể thu thập từ:

- Các nguồn hiện có:
 - Internet đã trở thành một nguồn dữ liệu quan trọng
- Các nghiên cứu thống kê:
 - Nghiên cứu thí nghiệm
 - Nghiên cứu quan sát

19

NGUỒN DỮ LIỆU

Các sai số của thu thập dữ liệu

- Một sai số trong thu thập dữ liệu xảy ra khi giá trị của dữ liệu thu thập được không bằng với giá đúng/thực có được từ một qui trình thu thập đúng
- Sử dụng dữ liệu sai có thể xấu hơn không sử dụng bất kỳ dữ liệu nào
- **GIGO** "Garbage In Garbage Out – Rác vào Rác Ra"

20

THỐNG KÊ MÔ TẢ

- **Thống kê mô tả:** Thu thập, Tổng kết và Mô tả dữ liệu
- **Các phương pháp** được sử dụng để tổng kết dữ liệu:
 - Lập Bảng
 - Đồ Thị
 - Bảng số

21

THỐNG KÊ MÔ TẢ

- **Thống kê mô tả:**
 - Các tham số thống kê
 - Tần số
 - Phân phối xác suất

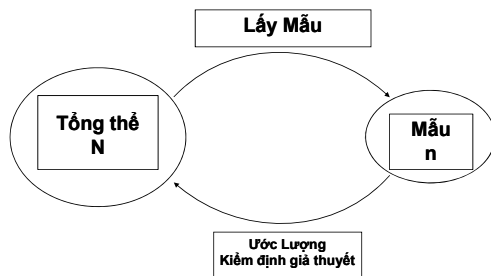
22

THỐNG KÊ SUY DIỄN

- **Tổng thể** là tập tất cả các phần tử cần quan tâm trong một nghiên cứu cụ thể
- **Mẫu** là một tập con của tổng thể
- **Thống kê suy diễn:** là quá trình sử dụng dữ liệu thu thập được từ mẫu để ước lượng hoặc kiểm định các giả thuyết thống kê về các đặc trưng của tổng thể

23

THỐNG KÊ SUY DIỄN



24

CHƯƠNG 2

**TRÌNH BÀY
BẰNG BẢNG và BIỂU ĐỒ**

25

NỘI DUNG CHÍNH

- Tổng kết dữ liệu định tính
- Tổng kết dữ liệu định lượng
- Phân tích dữ liệu khám phá: Trình bày dạng cảnh và lá
- Đồ thị phân tán điểm và bảng chéo

26

TỔNG KẾT DỮ LIỆU ĐỊNH TÍNH

Phân phối tần số

Phân phối tần số là một bảng tổng kết một tập dữ liệu trong đó trình bày tần số (hay số) của các giá trị quan sát có trong mỗi lớp của các lớp không trùng lên nhau

27

TỔNG KẾT DỮ LIỆU ĐỊNH TÍNH

DỮ LIỆU TỪ MỘT MẪU GỒM 50 LON NƯỚC GIẢI KHÁT

Coke Classic	Sprite	Pepsi-Cola
Diet Coke	Coke Classic	Coke Classic
Pepsi-Cola	Diet Coke	Coke Classic
Diet Coke	Coke Classic	Coke Classic
Coke Classic	Diet Coke	Pepsi-Cola
Coke Classic	Coke Classic	Dr.Pepper
Dr.Pepper	Sprite	Coke Classic
Diet Coke	Pepsi-Cola	Diet Coke
Pepsi-Cola	Coke Classic	Pepsi-Cola
Pepsi-Cola	Coke Classic	Pepsi-Cola
Coke Classic	Coke Classic	Pepsi-Cola
Dr.Pepper	Pepsi-Cola	Pepsi-Cola
Sprite	Coke Classic	Coke Classic
Coke Classic	Sprite	Dr.Pepper
Diet Coke	Dr.Pepper	Pepsi-Cola
Coke Classic	Pepsi-Cola	Sprite
Coke Classic	Diet Coke	

TỔNG KẾT DỮ LIỆU ĐỊNH TÍNH

PHÂN PHỐI TẦN SỐ CỦA LON NƯỚC GIẢI KHÁT

<i>Nước giải khát</i>	<i>Tần số</i>
Coke Classic	19
Diet Coke	8
Dr.Pepper	5
Pepsi-Cola	13
Sprite	5
<i>Tổng</i>	<i>50</i>

TỔNG KẾT DỮ LIỆU ĐỊNH TÍNH

Phân phối tần số tương đối và tần số phần trăm

- **Phân phối tần số tương đối:** Một bảng tổng kết tập một dữ liệu trong đó trình bày tần số tương đối –nghĩa là, tỉ số – của tổng số các giá trị quan sát có trong mỗi lớp của các lớp không trùng lên nhau
- Tần số tương đối của 1 lớp = Tần số của 1 lớp / n
- Tần số phần trăm = Tần số tương đối* 100

TỔNG KẾT DỮ LIỆU ĐỊNH TÍNH

Phân phối tần số tương đối và tần số phần trăm

- **Phân phối tần số tương đối:** Một bảng tổng kết tập một dữ liệu trong đó trình bày phần trăm của tổng số các giá trị quan sát có trong mỗi lớp của các lớp không trùng lên nhau

31

TỔNG KẾT DỮ LIỆU ĐỊNH TÍNH

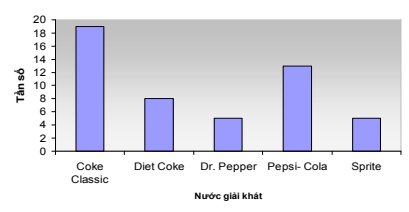
PHÂN PHỐI TẦN SỐ TƯƠNG ĐỐI và PHẦN TRĂM CỦA LỌN NƯỚC GIẢI KHÁT

Nước giải khát	Tần số tương đối	Tần số phần trăm
Coke Classic	.38	38
Diet Coke	.16	16
Dr.Pepper	.10	10
Peppi-Cola	.26	26
Sprite	.10	10
Tổng	1.00	100

32

TỔNG KẾT DỮ LIỆU ĐỊNH TÍNH

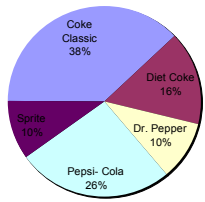
Biểu đồ hình thanh và biểu đồ hình tròn BIỂU ĐỒ HÌNH THANH CỦA NƯỚC GIẢI KHÁT



33

TỔNG KẾT DỮ LIỆU ĐỊNH TÍNH

Biểu đồ hình thanh và biểu đồ hình tròn
BIỂU ĐỒ HÌNH TRÒN CỦA NƯỚC GIẢI KHÁT



34

TỔNG KẾT DỮ LIỆU ĐỊNH LƯỢNG

Phân phối tần số

Phân phối tần số là một bảng tổng kết một tập dữ liệu trong đó trình bày tần số (hay số) của các giá trị quan sát có trong mỗi lớp của các lớp không trùng lên nhau

35

TỔNG KẾT DỮ LIỆU ĐỊNH LƯỢNG

Phân phối tần số

- Xây dựng một phân phối tần số
 - Thu thập dữ liệu mẫu
 - Xác định số lớp không trùng lấp
 - Xác định chiều rộng của mỗi lớp
 - Xác định các giới hạn của mỗi lớp
 - Đếm số các giá trị dữ liệu có trong mỗi lớp
 - Tổng kết các tần số của lớp vào trong một bảng phân phối tần số

36

TỔNG KẾT DỮ LIỆU ĐỊNH LƯỢNG

Phân phối tần số

▪ **Số lớp (K):** $5 \leq K \leq 20$

▪ **Chiều rộng lớp**

Chiều rộng lớp = (Giá trị lớn nhất – Giá trị nhỏ nhất) / K

▪ **Các giới hạn của lớp**

Các giới hạn của lớp là số lớn nhất và nhỏ nhất thuộc về lớp

• Giới hạn dưới của lớp = Số nhỏ nhất

• Giới hạn trên của lớp = Số lớn nhất

Sự khác biệt giữa giới hạn dưới của các lớp liên nhau sẽ cho ta chiều rộng của lớp

37

TỔNG KẾT DỮ LIỆU ĐỊNH LƯỢNG

Phân phối tần số

▪ **Các biên giới của lớp**

Các biên của lớp là các đường phân chia giữa các lớp

▪ **Điểm giữa của lớp**

Điểm giữa của lớp là giá trị nằm ở giữa các giới hạn dưới và giới hạn trên của lớp

38

TỔNG KẾT DỮ LIỆU ĐỊNH LƯỢNG

CÁC THỜI GIAN KIỂM TOÁN CUỐI NĂM
(Tính theo số ngày)

12	14	19	18
15	15	18	17
20	27	22	23
22	21	33	28
14	18	16	13

39

TỔNG KẾT DỮ LIỆU ĐỊNH LƯỢNG

PHÂN PHỐI TẦN SỐ ĐỐI VỚI DỮ LIỆU THỜI GIAN KIỂM TÓAN

Thời gian kiểm toán (ngày)	Tần số
10-14	4
15-19	8
20-24	5
25-29	2
30-34	1
Tổng	20

40

TỔNG KẾT DỮ LIỆU ĐỊNH LƯỢNG

Phân phối tần số tương đối và tần số phần trăm

- Tần số tương đối của 1 lớp = Tần số của 1 lớp / n
- Tần số phần trăm = Tần số tương đối * 100

41

TỔNG KẾT DỮ LIỆU ĐỊNH LƯỢNG

PHÂN PHỐI TẦN SỐ TƯƠNG ĐỐI VÀ TẦN SỐ PHẦN TRĂM ĐỐI VỚI DỮ LIỆU THỜI GIAN KIỂM TÓAN

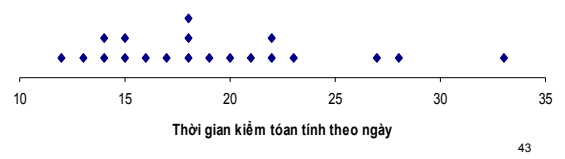
Thời gian (ngày)	Tần số tương đối	Tần số phần trăm
10-14	.20	20
15-19	.40	40
20-24	.25	25
25-29	.10	10
30-34	.05	5
Tổng	1.00	100

42

TỔNG KẾT DỮ LIỆU ĐỊNH LƯỢNG

Biểu đồ điểm

- Trục hoành trình bày miền các giá trị của dữ liệu.
- Mỗi giá trị được biểu thị bằng một điểm nằm trên trục



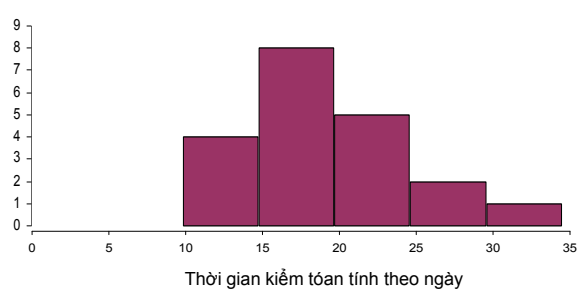
TỔNG KẾT DỮ LIỆU ĐỊNH LƯỢNG

Biểu đồ tần số

- Một biểu đồ tần số được xây dựng bằng đặt các biến quan tâm trên trục hoành và tần số, tần số tương đối, tần số phần trăm trên trục tung
- Biểu đồ tần số mô tả dạng của tập dữ liệu

44

TỔNG KẾT DỮ LIỆU ĐỊNH LƯỢNG



45

TỔNG KẾT DỮ LIỆU ĐỊNH LƯỢNG

Các phân phối tích lũy

Phân phối tần số tích lũy trình bày số các quan sát có giá trị nhỏ hơn hoặc bằng giới hạn trên của lớp của mỗi lớp

46

TỔNG KẾT DỮ LIỆU ĐỊNH LƯỢNG

CÁC PHÂN PHỐI TẦN SỐ TÍCH LŨY, TẦN SỐ TƯƠNG ĐỐI TÍCH LŨY VÀ TẦN SỐ PHẦN TRĂM TÍCH LŨY ĐỐI VỚI DỮ LIỆU THỜI GIAN KIỂM TỐAN

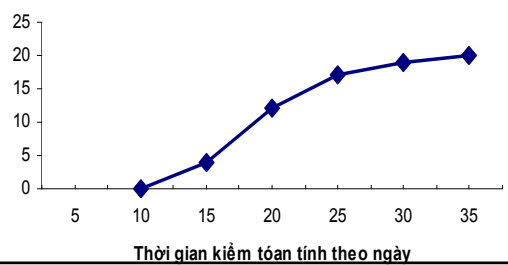
Thời gian (ngày)	Tần số Tích lũy	Tần số tương đối Tích lũy	Tần số % Tích lũy
Nhỏ hơn hoặc bằng 14	4	.20	20
Nhỏ hơn hoặc bằng 19	12	.60	60
Nhỏ hơn hoặc bằng 24	17	.85	85
Nhỏ hơn hoặc bằng 29	19	.95	95
Nhỏ hơn hoặc bằng 34	20	1.00	100

47

TỔNG KẾT DỮ LIỆU ĐỊNH LƯỢNG

Ogive

Ogive là đồ thị của phân phối tích lũy



48

**PHÂN TÍCH DỮ LIỆU KHÁM PHÁ:
TRÌNH BÀY DẠNG CẢNH VÀ LÁ**

Trình bày dạng cảnh và lá: Một kỹ thuật phân tích khám phá theo đó sắp hạng các thứ tự của dữ liệu định lượng và cho thấy sâu sắc về dạng của phân phối cùng một lúc

49

**PHÂN TÍCH DỮ LIỆU KHÁM PHÁ:
TRÌNH BÀY DẠNG CẢNH VÀ LÁ**

SỐ CÂU HỎI ĐƯỢC TRẢ LỜI ĐÚNG
Ở KỶ THI NĂNG KHIẾU

112	72	69	97	107
73	92	76	86	73
126	128	118	127	124
82	104	132	134	83
92	108	96	100	92
115	76	91	102	81
95	141	81	80	106
84	119	113	98	75
38	98	115	106	95
100	85	94	106	119

50

**PHÂN TÍCH DỮ LIỆU KHÁM PHÁ:
TRÌNH BÀY DẠNG CẢNH VÀ LÁ**

6	8	9												
7	2	3	3	5	6	6								
8	0	1	1	2	3	4	5	6						
9	1	2	2	2	4	5	5	6	7	8	8			
10	0	0	2	4	6	6	6	7	8					
11	2	3	5	5	8	9	9							
12	4	6	7	8										
13	2	4												
14	1													

51

**ĐỒ THỊ PHÂN TÁN ĐIỂM
và BẢNG CHÉO**

Bảng chéo

- Bảng chéo là một tổng kết dưới dạng bảng của dữ liệu gồm 2 biến. Các giá trị của một biến được trình bày theo các hàng. Các giá trị của một biến khác được trình bày theo các cột
- Bảng chéo được sử dụng rộng rãi trong việc xem xét mối quan hệ giữa hai biến

52

**ĐỒ THỊ PHÂN TÁN ĐIỂM
và BẢNG CHÉO**

BẢNG CHÉO VỀ ĐÁNH GIÁ CHẤT LƯỢNG VÀ GIÁ CỦA CÁC BỮA ĂN TẠI 300 NHÀ HÀNG Ở LOS-ANGELES

Chất lượng	Giá bữa ăn				Tổng
	\$10-19	\$20-29	\$30-39	\$40-49	
Tốt	42	40	2	0	84
Rất tốt	34	64	46	6	150
Xuất sắc	2	14	28	22	66
Tổng	78	118	76	28	300

53

**ĐỒ THỊ PHÂN TÁN ĐIỂM
và BẢNG CHÉO**

PHÂN TRĂM TÍNH THEO HÀNG ĐỐI VỚI MỖI LOẠI CHẤT LƯỢNG

Chất lượng	Giá bữa ăn				Tổng
	\$10-19	\$20-29	\$30-39	\$40-49	
Tốt	50.0	47.6	2.4	0.0	100
Rất tốt	22.7	42.7	30.6	4.0	100
Xuất sắc	3.0	21.2	42.4	33.4	100

54

**ĐỒ THỊ PHÂN TÁN ĐIỂM
và BẢNG CHÉO**

Đồ thị phân tán điểm và đường xu hướng

- Một đồ thị phân tán điểm là một trình bày dưới dạng đồ thị về mối quan hệ của hai biến. Một biến được trình bày trên trục hoành và biến khác được trình bày trên trục tung
- Một đường xu hướng là một đường cho thấy một cách gần đúng mối quan hệ giữa hai biến

55

**ĐỒ THỊ PHÂN TÁN ĐIỂM
và BẢNG CHÉO**

**DỮ LIỆU MẪU ĐỐI VỚI
CỬA HÀNG THIẾT BỊ STEREO VÀ ÂM THANH**

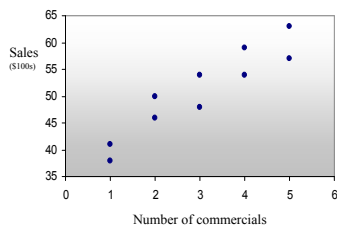
Tuần	Số thương vụ x	Doanh số (\$100s) y
1	2	50
2	5	57
3	1	41
4	3	54
5	4	54
6	1	38
7	5	63
8	3	48
9	4	59
10	2	46

56

**ĐỒ THỊ PHÂN TÁN ĐIỂM
và BẢNG CHÉO**

Đồ thị phân tán điểm

Đồ thị phân tán điểm đối với cửa hàng thiết bị Stereo và âm thanh



57

ĐỒ THỊ PHÂN TÁN ĐIỂM và BẢNG CHÉO

Đồ thị phân tán điểm
 Các loại quan hệ được miêu tả bằng đồ thị phân tán điểm

Quan hệ đồng biến Dường như không quan hệ Quan hệ nghịch biến

58

CÁC QUI TRÌNH BẢNG BIỂU VÀ ĐỒ THỊ ĐỐI VỚI TỔNG KẾT DỮ LIỆU

DỮ LIỆU

- Dữ liệu định tính
 - Phương pháp Bảng
 - Phân phối tần số
 - Phân phối tần số tương đối
 - Phân phối tần số phân trăm
 - Bảng chéo
 - Phương pháp Đồ thị
 - Biểu đồ hình thanh
 - Biểu đồ hình tròn
- Dữ liệu định lượng
 - Phương pháp Bảng
 - Phân phối tần số
 - Phân phối tần số tương đối
 - Phân phối tần số tích lũy
 - Phân phối tần số tương đối tích lũy
 - Cảnh và lá - Bảng chéo
 - Phương pháp Đồ thị
 - Biểu đồ điểm
 - Biểu đồ tần số
 - Biểu đồ tần số tích lũy (Ogive)
 - Đồ thị phân tán điểm

59

CHƯƠNG 3

CÁC ĐẠI LƯỢNG BẢNG SỐ

60

CÁC NỘI DUNG CHÍNH

- Giới thiệu
- Đại lượng về vị trí
- Đại lượng về sự biến thiên
- Đại lượng về dạng phân phối, vị trí tương đối và nhận dạng các điểm cá biệt
- Phân tích dữ liệu khám phá
- Đại lượng về sự liên hệ giữa 2 biến
- Trung bình có trọng số và xử lý dữ liệu nhóm

61

GIỚI THIỆU

- **Một đại lượng mô tả** là một con số đơn giản được tính toán từ dữ liệu mẫu để cung cấp thông tin về dữ liệu tổng thể
- Có hai loại đại lượng mô tả:
 - Đại lượng về vị trí
 - Đại lượng về sự biến thiên

62

GIỚI THIỆU

- **Tham số của tổng thể** (population parameter) là một giá trị bằng số được dùng như một đại lượng tổng kết đối với một dữ liệu của tổng thể
- **Các trị thống kê của mẫu** (sample statistics) được dùng như một đại lượng tổng kết đối với một mẫu

63

CÁC ĐẠI LƯỢNG VỀ VỊ TRÍ
(measure of location)

Một số các đại lượng về vị trí là:

- Số trung bình (Mean)
- Số trung vị (Median)
- Số yếu vị (Mode)
- Số phân vị (Percentiles)
- Số tứ phân (Quartiles)

64

CÁC ĐẠI LƯỢNG VỀ VỊ TRÍ

Số trung bình

- Số trung bình được sử dụng phổ biến nhất để đo lường vị trí

- Trung bình của tổng thể: $\mu = \frac{\sum x}{N}$

- Trung bình của mẫu: $\bar{x} = \frac{\sum x}{n}$

65

CÁC ĐẠI LƯỢNG VỀ VỊ TRÍ

Số yếu vị (Md)

Số trung vị là giá trị ở giữa tập dữ liệu đã được sắp xếp theo thứ tự

- n là số lẻ, Md là giá trị ở giữa tập dữ liệu
- n là số chẵn, Md là trung bình của hai giá trị ở giữa tập dữ liệu

66

CÁC ĐẠI LƯỢNG VỀ VỊ TRÍ

Số yếu vị (Mo)

Số yếu vị là giá trị dữ liệu xuất hiện với tần số lớn nhất

- Bimodal → có hai số yếu vị
- Multimodal → > two hai số yếu vị

67

CÁC ĐẠI LƯỢNG VỀ VỊ TRÍ

Số phân vị

- Số phân vị p^{th} là giá trị có ít nhất $p\%$ số hạng của tập dữ liệu có giá trị nhỏ hơn hoặc bằng giá trị này, và có ít nhất $(100-p)\%$ số hạng của tập dữ liệu có giá trị lớn hơn hoặc bằng giá trị này
- Phân vị 50^{th} là số trung vị

68

CÁC ĐẠI LƯỢNG VỀ VỊ TRÍ

Số phân vị

Xác định phân vị p^{th}

- Bước 1: Sắp xếp tập dữ liệu theo thứ tự tăng dần
- Bước 2: tính chỉ số $i: i = \left(\frac{p}{100}\right) * n$
- Bước 3:
 - Nếu i không là số nguyên → làm tròn lên trên. Số nguyên kế tiếp > i sẽ chỉ vị trí của phân vị p^{th} .
 - Nếu i là số nguyên, phân vị p^{th} là trung bình của 2 giá trị dữ liệu ở vị trí i và $i + 1$

69

CÁC ĐẠI LƯỢNG VỀ VỊ TRÍ

Số tứ phân

Số tứ phân chỉ đơn thuần là các số phân vị cụ thể, sẽ chia tập dữ liệu ra làm 4 phần, được gọi tên là:

- Q_1 = số tứ phân thứ nhất = $P_{25\%}$
- Q_2 = số tứ phân thứ hai = $P_{50\%}$ = Median
- Q_3 = số tứ phân thứ ba = $P_{75\%}$

70

CÁC ĐẠI LƯỢNG VỀ SỰ BIẾN THIÊN

- **Đại lượng về sự biến thiên** được sử dụng để mô tả xu hướng của các giá trị dữ liệu phân tán xung quanh giá trị trung bình.
- Một số đại lượng về sự biến thiên:
 - Khoảng biến thiên (Range)
 - Khoảng biến thiên nội tứ phân (Interquartile Range)
 - Phương sai (Variance)
 - Độ lệch chuẩn (Standard Deviation)
 - Hệ số biến thiên (Coefficient of variation)

71

CÁC ĐẠI LƯỢNG VỀ SỰ BIẾN THIÊN

- **Khoảng biến thiên**
 - Range = Giá trị lớn nhất – Giá trị nhỏ nhất hay
 - Range = Max – Min
- **Khoảng biến thiên nội tứ phân (IQR)**
 - IQR = $Q_3 - Q_1$

72

CÁC ĐẠI LƯỢNG VỀ SỰ BIẾN THIÊN

▪ Phương sai

• Phương sai của tổng thể: $\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}$

• Phương sai của mẫu: $s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$

73

CÁC ĐẠI LƯỢNG VỀ SỰ BIẾN THIÊN

▪ Độ lệch chuẩn

Độ lệch chuẩn là căn bậc hai của phương sai. Độ lệch chuẩn và phương sai được sử dụng phổ biến để đo lường sự biến thiên

$\sigma = \sqrt{\sigma^2}$ $s = \sqrt{s^2}$

▪ Hệ số biến thiên

$CV = \frac{\text{Standard Deviation}}{\text{Mean}} * 100 = \frac{s}{\bar{X}} * 100$

74

CÁC ĐẠI LƯỢNG VỀ DẠNG PHÂN PHỐI, VỊ TRÍ TƯƠNG ĐỐI VÀ NHẬN DẠNG CÁC ĐIỂM CÁ BIỆT

▪ Dạng phân phối

- Độ lệch (Skewness) là đại lượng về dạng của phân phối của tập dữ liệu
 - Đối với dữ liệu lệch về bên trái, độ lệch sẽ âm
 - Đối với dữ liệu lệch về bên phải, độ lệch sẽ dương
 - Nếu dữ liệu đối xứng, độ lệch sẽ bằng 0
- Đối với phân phối đối xứng, số trung bình và số trung vị sẽ bằng nhau

75

CÁC ĐẠI LƯỢNG VỀ DẠNG PHÂN PHỐI, VỊ TRÍ TƯƠNG ĐỐI VÀ NHẬN DẠNG CÁC ĐIỂM CÁ BIỆT

- **Trị thống kê Z (Z-Scores)**
 Trị thống kê Z thường được gọi là giá trị chuẩn hóa

$$Z_i = \frac{x_i - \bar{x}}{s}$$

Z_i : là số độ lệch chuẩn mà X_i cách xa giá trị trung bình

76

CÁC ĐẠI LƯỢNG VỀ DẠNG PHÂN PHỐI, VỊ TRÍ TƯƠNG ĐỐI VÀ NHẬN DẠNG CÁC ĐIỂM CÁ BIỆT

- **Định lý Chebyshev**

Định lý Chebyshev được sử dụng để phát biểu về phần trăm của các số hạng sẽ nằm trong một con số cụ thể của độ lệch chuẩn tính từ giá trung bình

77

CÁC ĐẠI LƯỢNG VỀ DẠNG PHÂN PHỐI, VỊ TRÍ TƯƠNG ĐỐI VÀ NHẬN DẠNG CÁC ĐIỂM CÁ BIỆT

- **Định lý Chebyshev**

- Tối thiểu $(1-1/Z^2)$ của các số hạng có trong mọi tập dữ liệu sẽ phải nằm trong Z độ lệch chuẩn tính từ số trung bình, khi $Z > 1$.

hay

- $\text{Prob}(\bar{x} - zs < x < \bar{x} + zs) \geq 1 - \frac{1}{Z^2}$

78

CÁC ĐẠI LƯỢNG VỀ DẠNG PHÂN PHỐI, VỊ TRÍ TƯƠNG ĐỐI VÀ NHẬN DẠNG CÁC ĐIỂM CÁ BIỆT

▪ **Định lý Chebyshev**

Đối với mọi tập dữ liệu

- Prob $(\bar{x} - 2s < x < \bar{x} + 2s) \geq 75\%$
- Prob $(\bar{x} - 3s < x < \bar{x} + 3s) \geq 89\%$
- Prob $(\bar{x} - 4s < x < \bar{x} + 4s) \geq 94\%$

79

CÁC ĐẠI LƯỢNG VỀ DẠNG PHÂN PHỐI, VỊ TRÍ TƯƠNG ĐỐI VÀ NHẬN DẠNG CÁC ĐIỂM CÁ BIỆT

▪ **Quy tắc kinh nghiệm**

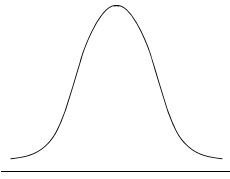
Đối với mọi tập dữ liệu có phân phối dạng hình chuông:

- Prob $(\bar{x} - 1s < x < \bar{x} + 1s) \geq 68\%$
- Prob $(\bar{x} - 2s < x < \bar{x} + 2s) \geq 95\%$
- Prob $(\bar{x} - 3s < x < \bar{x} + 3s) \geq 99.7\%$

80

CÁC ĐẠI LƯỢNG VỀ DẠNG PHÂN PHỐI, VỊ TRÍ TƯƠNG ĐỐI VÀ NHẬN DẠNG CÁC ĐIỂM CÁ BIỆT

MỘT PHÂN PHỐI DẠNG HÌNH CHUÔNG ĐỐI XỨNG



81

CÁC ĐẠI LƯỢNG VỀ DẠNG PHÂN PHỐI, VỊ TRÍ TƯƠNG ĐỐI VÀ NHẬN DẠNG CÁC ĐIỂM CÁ BIỆT

- **Nhận dạng các điểm cá biệt (outliers)**
 - Các điểm cá biệt là các giá trị thái cực (lớn khác thường hoặc nhỏ khác thường)
 - Sử dụng Z để nhận dạng điểm cá biệt: mọi giá trị dữ liệu với Z nhỏ hơn -3 hoặc lớn hơn +3 là điểm cá biệt

82

PHÂN TÍCH DỮ LIỆU KHÁM PHÁ

- **Mô hình 5-điểm**
 - Giá trị nhỏ nhất = Min
 - Số tứ phân thứ nhất = Q_1
 - Số tứ phân thứ hai = $Q_2 = \text{Median}$
 - Số tứ phân thứ ba = Q_3
 - Giá trị lớn nhất = Max

Một cách gần đúng, 25% của các giá trị dữ liệu ở giữa các số kế nhau trong mô hình 5-điểm

83

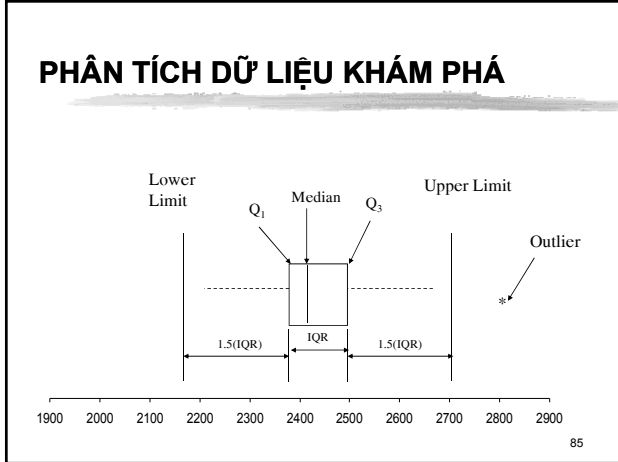
PHÂN TÍCH DỮ LIỆU KHÁM PHÁ

- **Biểu đồ hộp (Box plot):**

Biểu đồ hộp là tổng kết dữ liệu bằng đồ thị

- Outliers
- Whiskers

84



ĐẠI LƯỢNG VỀ SỰ LIÊN HỆ GIỮA 2 BIẾN

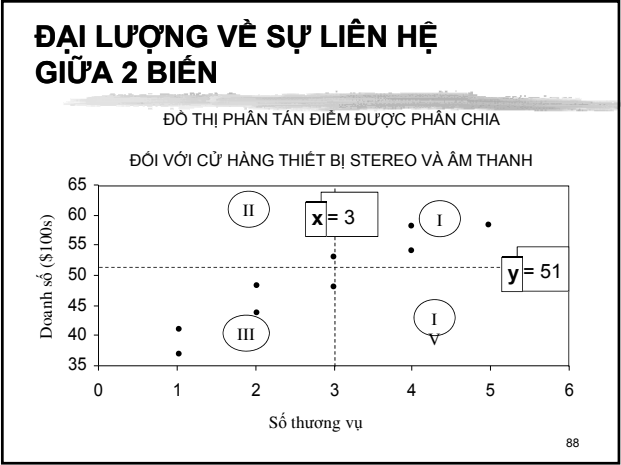
- **Đồng phương sai (Covariance)**
 - Đồng phương sai đo lường sự liên hệ tuyến tính giữa 2 biến.
 - Đồng phương sai của tổng thể:
$$\sigma_{xy} = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{N}$$
 - Đồng phương sai của mẫu:
$$s_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

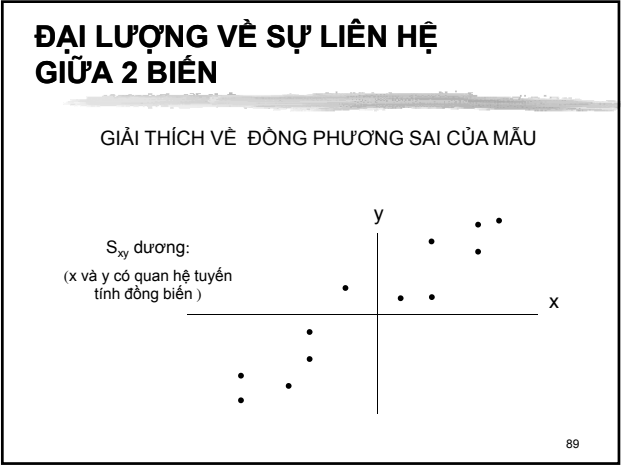
86

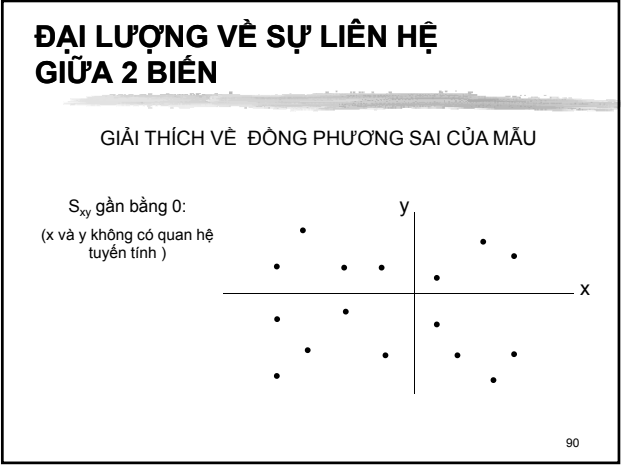
ĐẠI LƯỢNG VỀ SỰ LIÊN HỆ GIỮA 2 BIẾN

- **Đồng phương sai**
 - $s_{xy} > 0$ → Quan hệ đồng biến
 - $s_{xy} < 0$ → Quan hệ nghịch biến
 - Giá trị của đồng phương sai phụ thuộc đơn vị đo lường của x và y

87







ĐẠI LƯỢNG VỀ SỰ LIÊN HỆ GIỮA 2 BIẾN

GIẢI THÍCH VỀ ĐỒNG PHƯƠNG SAI CỦA MẪU

S_{xy} âm:
(x và y có quan hệ tuyến tính nghịch biến)

91

ĐẠI LƯỢNG VỀ SỰ LIÊN HỆ GIỮA 2 BIẾN

- **Hệ số tương quan (Correlation Coefficient)**
 - Một đại lượng bằng số đo lường mối quan hệ tuyến tính giữa 2 biến
 - Hệ số tương quan Pearson
 - Tổng thể: $\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$
 - Mẫu: $r_{xy} = \frac{S_{xy}}{S_x S_y}$

$$r = r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

92

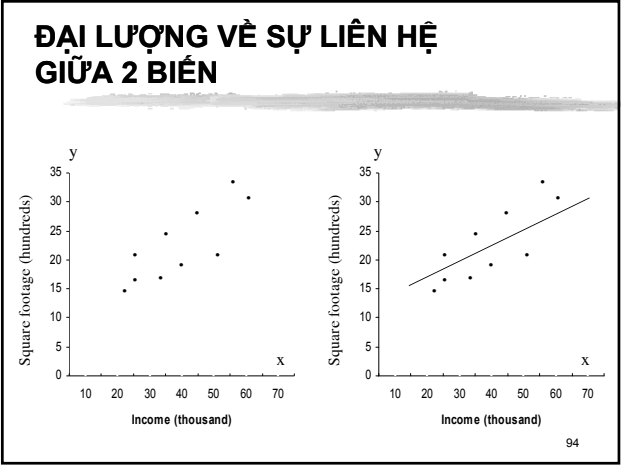
ĐẠI LƯỢNG VỀ SỰ LIÊN HỆ GIỮA 2 BIẾN

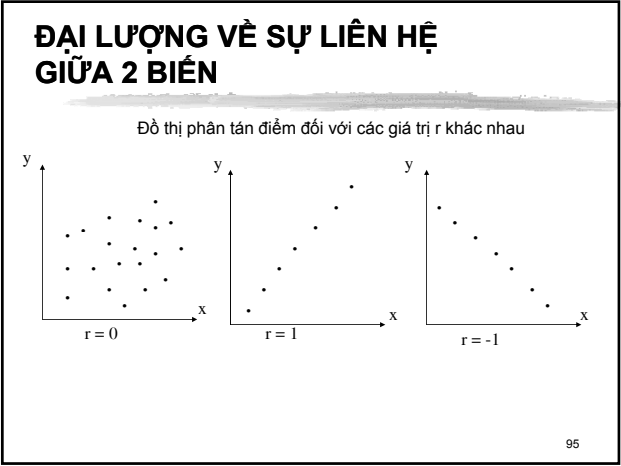
- **Hệ số tương quan**

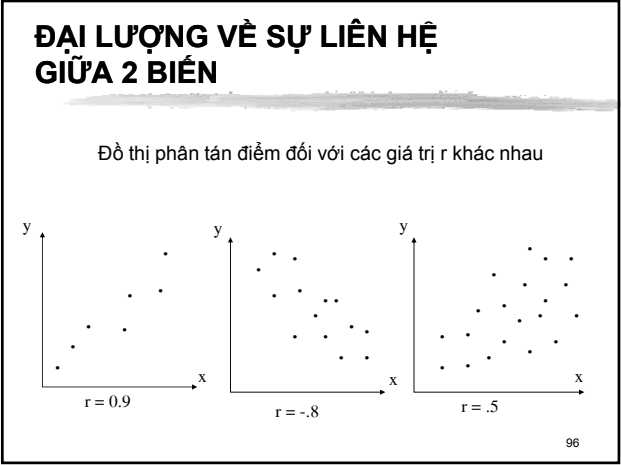
Các tính chất quan trọng của r:

- $-1 \leq r \leq 1$
- $|r|$ càng lớn thì mối quan hệ tuyến tính càng mạnh.
- $r = 0$ -> không có quan hệ tuyến tính giữa X và Y
- $r = 1$ hoặc $r = -1$ → X và Y tương quan tuyến tính hoàn toàn
- Dấu của r cho thấy mối quan hệ giữa X và Y là đồng biến hay nghịch biến

93







TRUNG BÌNH CÓ TRỌNG SỐ VÀ XỬ LÝ DỮ LIỆU NHÓM

- **Trung bình có trọng số** (The weighted Mean)
 Trung bình của tập dữ liệu có được bằng cách gán mỗi giá trị dữ liệu một trọng số phản ảnh tầm quan trọng của nó trong tập dữ liệu

$$\bar{x} = \frac{\sum w_i * x_i}{\sum w_i}$$

97

TRUNG BÌNH CÓ TRỌNG SỐ VÀ XỬ LÝ DỮ LIỆU NHÓM

- **Dữ liệu nhóm** (Grouped data)
 Dữ liệu có sẵn trong các lớp được tổng kết bằng phân phối tần số. Các giá trị riêng của tập dữ liệu gốc sẽ không được ghi nhận

98

TRUNG BÌNH CÓ TRỌNG SỐ VÀ XỬ LÝ DỮ LIỆU NHÓM

- **Dữ liệu nhóm**
 Trung bình của dữ liệu nhóm

- Tổng thể $\mu = \frac{\sum f_i * M_i}{N}$

- Mẫu $\bar{x} = \frac{\sum f_i * M_i}{n}$

99

TRUNG BÌNH CÓ TRỌNG SỐ VÀ XỬ LÝ DỮ LIỆU NHÓM

▪ Dữ liệu nhóm

Phương sai của dữ liệu nhóm

{ Tổng thể $\sigma^2 = \frac{\sum f_i * (M_i - \mu)^2}{N}$

{ Mẫu $s^2 = \frac{\sum f_i * (M_i - \bar{x})^2}{n - 1}$

100

CHƯƠNG 4

**GIỚI THIỆU
về
XÁC SUẤT**

101

NỘI DUNG CHÍNH

- Thí nghiệm, qui tắc đếm và xác định xác suất
- Biến cố và xác suất của biến cố
- Một số mối quan hệ căn bản của xác suất
- Xác suất có điều kiện
- Định lý Bayes

102

THÍ NGHIỆM, QUI TẮC ĐẾM và XÁC ĐỊNH XÁC SUẤT

- **Thí nghiệm (Experiment)**
 - Thí nghiệm là mọi quá trình tạo ra kết quả đã được định nghĩa rõ ràng
- **Không gian mẫu (Sample space)**
 - Điểm của mẫu là một kết quả cụ thể của một thí nghiệm
 - Không gian mẫu là tập hợp của tất cả các điểm có thể có của mẫu (các kết quả của thí nghiệm)

103

THÍ NGHIỆM, QUI TẮC ĐẾM và XÁC ĐỊNH XÁC SUẤT

- **Qui tắc đếm**
 - Sơ đồ cây là một phương tiện đồ thị rất hữu ích trong việc xác định các điểm của mẫu của một thí nghiệm có liên quan đến nhiều bước.
 - Qui tắc đếm đối với thí nghiệm nhiều bước
Số kết quả của thí nghiệm = $(n_1) \times (n_2) \times \dots \times (n_k)$
 - Qui tắc đếm đối với tổ hợp
Số tổ hợp của N phần tử được chọn n trong một lần là:

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

104

THÍ NGHIỆM, QUI TẮC ĐẾM và XÁC ĐỊNH XÁC SUẤT

- **Yêu cầu căn bản của xác suất**

Gọi E_i là kết quả của thí nghiệm

 - $0 \leq P(E_i) \leq 1$
 - $\sum P(E_i) = 1$
- **Các phương pháp xác định xác suất**
 - Phương pháp cổ điển
 - Phương pháp tần số tương đối
 - Phương pháp chủ quan

105

THÍ NGHIỆM, QUI TẮC ĐẾM và XÁC ĐỊNH XÁC SUẤT

▪ **Phương pháp cổ điển**

Một phương pháp xác định xác suất thích hợp khi tất cả các kết quả của thí nghiệm có cùng khả năng xảy ra

▪ **Phương pháp tần số tương đối**

Một phương pháp xác định xác suất thích hợp khi có sẵn dữ liệu (dữ liệu lịch sử) để ước lượng tỉ lệ của số lần kết quả thí nghiệm sẽ xảy ra nếu thí nghiệm được lặp lại với một số lần đủ lớn

106

THÍ NGHIỆM, QUI TẮC ĐẾM và XÁC ĐỊNH XÁC SUẤT

▪ **Phương pháp chủ quan**

- Một phương pháp xác định xác suất dựa trên cơ sở phán đoán
- Một xác suất chủ quan là một mức độ tin tưởng của cá nhân đối với việc xảy ra một kết quả của thí nghiệm

107

BIẾN CỐ và XÁC SUẤT CỦA BIẾN CỐ

▪ **Biến cố**

Một biến cố là một tập hợp của các kết quả của thí nghiệm

▪ **Xác suất của biến cố**

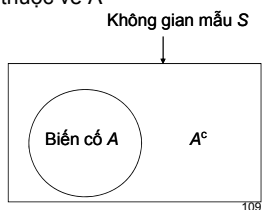
Xác suất của một biến cố bất kỳ sẽ bằng với tổng các xác suất của các kết quả của thí nghiệm

108

MỘT SỐ MỐI QUAN HỆ CĂN BẢN CỦA XÁC SUẤT

▪ **Phần bù/phụ của biến cố**

- Phần phụ của biến cố A là biến cố chứa tất cả các kết quả của mẫu mà không thuộc về A
- $P(A) = 1 - P(A^c)$

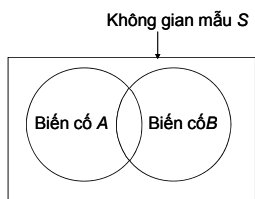


109

MỘT SỐ MỐI QUAN HỆ CĂN BẢN CỦA XÁC SUẤT

▪ **Biến cố HỢI của 2 biến cố: $A \cup B$**

$A \cup B$ là biến cố chứa tất cả các kết quả của thí nghiệm thuộc A hoặc B, hoặc cả hai

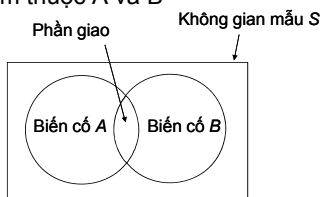


110

MỘT SỐ MỐI QUAN HỆ CĂN BẢN CỦA XÁC SUẤT

▪ **Biến cố GIAO của 2 biến cố: $A \cap B$**

$A \cap B$ là biến cố chứa tất cả các kết quả của thí nghiệm thuộc A và B



111

MỘT SỐ MỐI QUAN HỆ CĂN BẢN CỦA XÁC SUẤT

▪ **Phép cộng xác suất**

• $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

• Biện cố cách biệt

• Hai biến cố được gọi là cách biệt nếu hai biến cố không có các điểm ở phần giao.

• A và B là hai biến cố cách biệt: $P(A \cap B) = 0$

• Phép cộng xác suất đối với hai biến cố cách biệt

• $P(A \cup B) = P(A) + P(B)$

112

XÁC SUẤT CÓ ĐIỀU KIỆN

▪ **Xác suất có điều kiện**

hay $P(A \setminus B) = \frac{P(A \cap B)}{P(B)}$

$P(B \setminus A) = \frac{P(A \cap B)}{P(A)}$

▪ **Các biến cố độc lập**

Nếu A và B là hai biến cố độc lập thì:

$P(A \setminus B) = P(A)$ hay $P(B \setminus A) = P(B)$

113

XÁC SUẤT CÓ ĐIỀU KIỆN

▪ **Phép nhân xác suất**

• $P(A \cap B) = P(B) \cdot P(A \setminus B) = P(A) \cdot P(B \setminus A)$

• Phép nhân xác suất đối với hai biến cố độc lập

$P(A \cap B) = P(A) \cdot P(B)$

114

ĐỊNH LÝ BAYES

- **Các xác suất tiên nghiệm:** Các ước lượng ban đầu về xác suất của các biến cố
- **Xác suất hậu nghiệm:** Các xác suất được sửa lại của các biến cố dựa trên các thông tin bổ sung
- **Định lý Bayes**

$$P(A_i | B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)} = \frac{P(A_i \cap B)}{P(B)}$$

115

CHƯƠNG 5

PHÂN PHỐI XÁC SUẤT RỜI RẠC

116

NỘI DUNG CHÍNH

- Các biến ngẫu nhiên
- Các phân phối xác suất rời rạc
- Giá trị kỳ vọng và phương sai của biến rời rạc
- Phân phối xác suất nhị thức
- Phân phối xác suất Poisson

117

BIẾN NGẪU NHIÊN

▪ **Biến ngẫu nhiên**

Một biến ngẫu nhiên là một mô tả bằng số của kết quả của thí nghiệm

▪ **Biến ngẫu nhiên rời rạc**

Một biến ngẫu nhiên rời rạc có thể nhận một số đếm được của các giá trị trong một khoảng

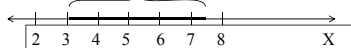
118

BIẾN NGẪU NHIÊN

▪ **Biến ngẫu nhiên liên tục**

Một biến ngẫu nhiên liên tục được giả định có thể nhận mọi giá trị trong một khoảng

Mọi giá trị có thể có trong khoảng này



X: height, in feet

119

PHÂN PHỐI XÁC SUẤT RỜI RẠC

▪ **Phân phối xác suất** đối với một biến ngẫu nhiên sẽ mô tả làm thế nào các xác suất được phân phối theo các giá trị của biến ngẫu nhiên

▪ **Một phân phối xác suất đối với một biến ngẫu nhiên rời rạc X** là một danh sách các giá trị có thể có của biến X và các xác suất tương ứng

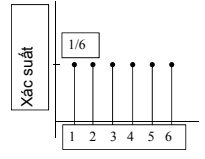
120

PHÂN PHỐI XÁC SUẤT RỜI RẠC

Một phân phối xác suất có thể được trình bày dưới dạng:

- Bảng
- Đồ thị (Đồ thị tần số)
- Công thức (Hàm số)

x	1	2	3	4	5	6
P(x)	1/6	1/6	1/6	1/6	1/6	1/6



121

PHÂN PHỐI XÁC SUẤT RỜI RẠC

Hàm xác suất rời rạc $f(x)$ là một hàm xác định xác suất đối với mỗi giá trị của biến X

$$f(x) = \text{Prob}(X=x)$$

- Các điều kiện yêu cầu đối với hàm xác suất rời rạc
 - $0 \leq f(x) \leq 1$
 - $\sum f(x) = 1$

122

PHÂN PHỐI XÁC SUẤT RỜI RẠC

Hàm phân phối xác suất rời rạc đều

$$f(x) = 1/n$$

n = số các giá trị có thể có của biến ngẫu nhiên rời rạc

123

GIÁ TRỊ KỲ VỌNG và PHƯƠNG SAI của BIẾN NGẪU NHIÊN RỜI RẠC

- **Giá trị kỳ vọng**

$$E(x) = \mu = \sum x \cdot f(x)$$

- **Phương sai**

$$\text{Var}(x) = \sigma^2 = \sum (x - \mu)^2 \cdot f(x)$$

or

$$\sigma^2 = \sum x^2 \cdot f(x) - \mu^2$$

- **Độ lệch chuẩn**

$$\sigma = \sqrt{\sigma^2}$$

124

PHÂN PHỐI XÁC SUẤT NHỊ THỨC

- **Một thí nghiệm nhị thức**

Một thí nghiệm nhị thức có 4 tính chất:

- Thí nghiệm gồm có một chuỗi n lần thử tương tự
- Hai kết quả có thể có cho mỗi lần thử: thành công và thất bại
- Xác suất của thành công, p, không thay đổi ở lần thử này sang lần thử khác. Vì vậy, xác suất của thất bại, 1-p, không thay đổi ở lần thử này sang lần thử khác
- Các lần thử độc lập với nhau

125

PHÂN PHỐI XÁC SUẤT NHỊ THỨC

- **Hàm xác suất nhị thức**

$$f(x) = \binom{n}{x} p^x (1-p)^{(n-x)}$$

- **Giá trị kỳ vọng và phương sai của phân phối xác suất nhị thức**

- Giá trị kỳ vọng: $E(x) = \mu = np$
- Phương sai: $\sigma^2 = np(1-p)$

126

PHÂN PHỐI XÁC SUẤT POISSON

- Các tính chất của **Thí nghiệm Poisson**
 - Xác suất của một sự kiện sẽ giống nhau cho bất kỳ 2 khoảng có cùng độ dài
 - Việc xảy ra hay không xảy ra trong 1 khoảng bất kỳ sẽ độc lập với việc xảy ra hay không xảy ra trong 1 khoảng bất kỳ khác

127

PHÂN PHỐI XÁC SUẤT POISSON

- Hàm xác suất Poisson $f(x) = \frac{\mu^x e^{-\mu}}{x!}$
- μ = Giá trị kỳ vọng hay số trung bình của sự kiện trong một khoảng.
- Giá trị kỳ vọng và phương sai của phân phối xác suất Poisson
 - Giá trị kỳ vọng: $E(x) = \mu$
 - Phương sai: $Var(x) = \mu$

128

CHƯƠNG 6

**PHÂN PHỐI XÁC SUẤT
LIÊN TỤC**

129

NỘI DUNG CHÍNH

- Giới thiệu
- Phân phối xác suất đều
- Phân phối xác suất chuẩn
- Tính gần đúng phân phối chuẩn cho phân phối nhị thức

130

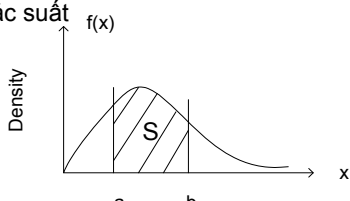
GIỚI THIỆU

- Một biến ngẫu nhiên liên tục là một giá trị ngẫu nhiên có thể nhận bất kỳ giá trị nào trong một khoảng hay tập hợp các khoảng
- Một Phân phối xác suất đối với một biến ngẫu nhiên liên tục được đặc trưng bởi một Hàm mật độ xác suất (Probability Density Function – PDF)

131

GIỚI THIỆU

- Các diện tích dưới đường cong mật độ xác suất là các xác suất



$$P(a < X < b) = S = \int_a^b f(x)dx$$

132

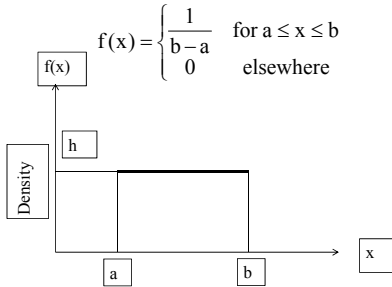
GIỚI THIỆU

- Một số các phân phối xác suất phổ biến đối với biến liên tục:
 - Phân phối đều (Uniform Distribution)
 - Phân phối chuẩn (Normal Distribution)

133

PHÂN PHỐI XÁC SUẤT ĐỀU

- Hàm mật độ xác suất của phân phối đều



134

PHÂN PHỐI XÁC SUẤT ĐỀU

- Giá trị kỳ vọng và phương sai của phân phối đều

$$E(x) = \mu = \int_a^b x \cdot f(x) dx = \frac{a+b}{2}$$

$$\text{Var}(x) = \sigma^2 = \int_a^b (x - \mu)^2 f(x) dx = \frac{(b-a)^2}{12}$$

135

PHÂN PHỐI XÁC SUẤT CHUẨN

- Hàm mật độ xác suất của phân phối chuẩn

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Với

- μ = Trung bình
- σ = Độ lệch chuẩn
- $\pi = 3.14159$
- $e = 2.71828$
- $X \sim N(\mu, \sigma^2)$

136

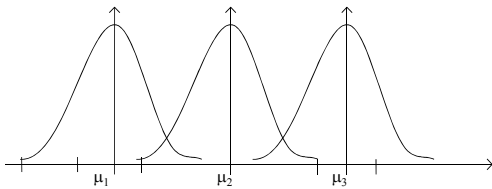
PHÂN PHỐI XÁC SUẤT CHUẨN

- Đường cong chuẩn
 - Dạng của $f(x)$ đối xứng, giống dạng hình chuông
 - Đường cong chuẩn có 2 tham số, μ và σ . Chúng xác định vị trí và dạng của phân phối

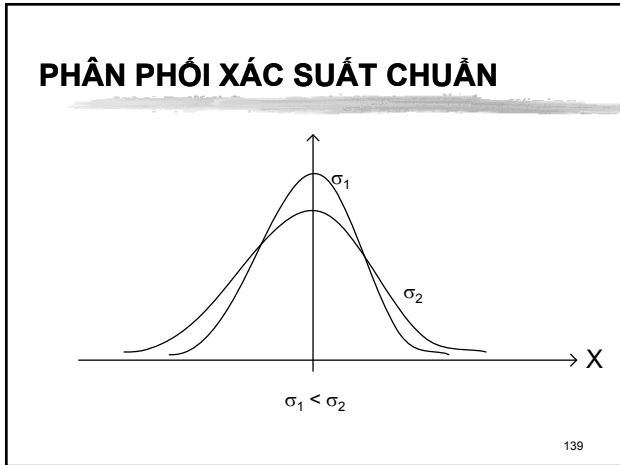
137

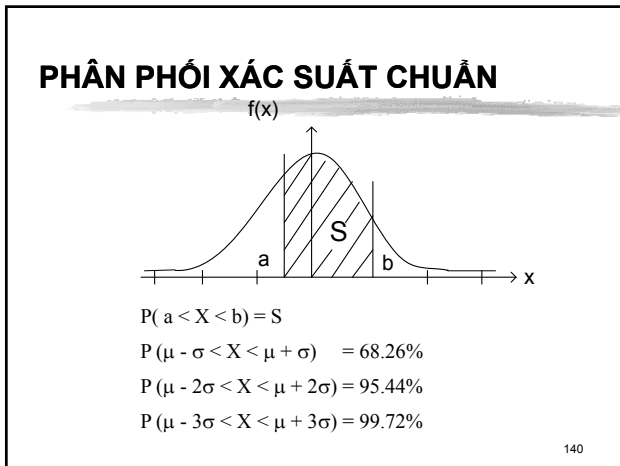
PHÂN PHỐI XÁC SUẤT CHUẨN

$$\mu_1 < \mu_2 < \mu_3$$



138





PHÂN PHỐI XÁC SUẤT CHUẨN

- Phân phối xác suất chuẩn chuẩn hóa
 - Phân phối xác suất chuẩn chuẩn hóa là một phân phối chuẩn có trung bình bằng 0 và phương sai bằng 1
 - Một biến ngẫu nhiên chuẩn chuẩn hóa Z là một biến tuân theo phân phối xác suất chuẩn chuẩn hóa

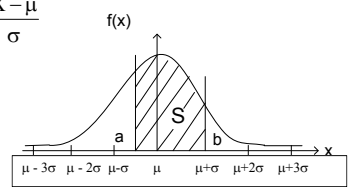
$Z \sim N(0, 1^2)$

141

PHÂN PHỐI XÁC SUẤT CHUẨN

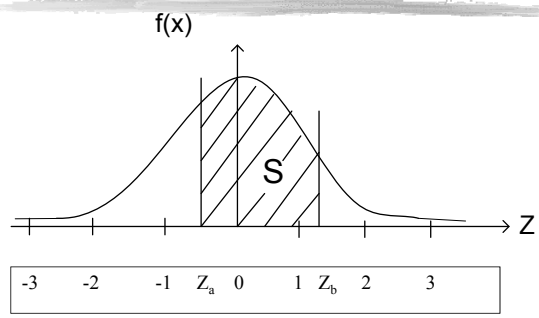
- Một biến chuẩn chuẩn hóa
- Nếu $X \sim N(\mu, \sigma^2)$ thì biến chuẩn chuẩn hóa Z có trung bình bằng 0, phương sai bằng 1 và $Z \sim N(0, 1^2)$

$$Z = \frac{X - \mu}{\sigma}$$



142

PHÂN PHỐI XÁC SUẤT CHUẨN



143

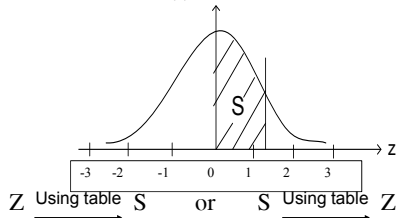
PHÂN PHỐI XÁC SUẤT CHUẨN

- $X \sim N(\mu, \sigma^2) \xrightarrow{Z = \frac{X - \mu}{\sigma}} Z \sim N(0, 1^2)$
 - $P(a < X < b) = P(Z_a < Z < Z_b) = S$
- $$Z_a = \frac{a - \mu}{\sigma} \qquad Z_b = \frac{b - \mu}{\sigma}$$

144

PHÂN PHỐI XÁC SUẤT CHUẨN

- Sử dụng bảng diện tích của đường cong chuẩn để tìm giá trị của S f(x)



145

TÍNH GẦN ĐÚNG PHÂN PHỐI CHUẨN CHO PHÂN PHỐI NHỊ THỨC

- Khi chúng ta gặp một vấn đề của phân phối nhị thức với số lần thử lớn, chúng ta có thể muốn tính gần đúng xác suất của phân phối nhị thức
 - $n > 20$
 - $np \geq 5$
 - $n(1-p) \geq 5$
- Với n cho trước, tính gần đúng phân phối chuẩn cho phân phối nhị thức sẽ tốt nhất khi $p = 0.5$

146

TÍNH GẦN ĐÚNG PHÂN PHỐI CHUẨN CHO PHÂN PHỐI NHỊ THỨC

- Khi sử dụng phân phối chuẩn tính gần đúng cho phân phối nhị thức, chúng ta đặt

$$\mu = np$$

$$\sigma = \sqrt{np(1-p)}$$

Vào trong định nghĩa của đường cong chuẩn

147

**TÍNH GẦN ĐÚNG PHÂN PHỐI CHUẨN
CHO PHÂN PHỐI NHỊ THỨC**

- Nhân tố điều chỉnh liên tục là giá trị 0.5, nghĩa là cộng hoặc trừ vào giá trị của X khi sử dụng phân phối xác suất chuẩn liên tục để tính gần đúng cho phân phối xác suất nhị thức rời rạc

148

CHƯƠNG 7

**LẤY MẪU
và
PHÂN PHỐI MẪU**

149

NỘI DUNG CHÍNH

- Giới thiệu vấn đề lấy mẫu
- Lấy mẫu ngẫu nhiên đơn giản
- Ước lượng điểm
- Giới thiệu phân phối mẫu
- Phân phối mẫu của trung bình mẫu
- Phân phối mẫu của tỉ lệ mẫu
- Các tính chất của ước lượng điểm
- Các phương pháp lấy mẫu khác

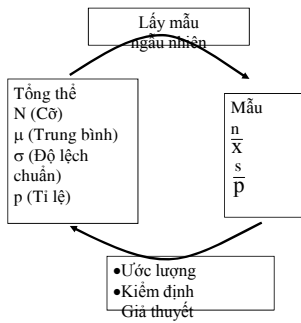
150

GIỚI THIỆU VẤN ĐỀ LẤY MẪU

- Một **Tổng thể** là tập hợp tất cả các phần tử cần quan tâm trong một nghiên cứu.
- Một **Mẫu** là một tập hợp con của tổng thể.
- **Mục đích của thống kê suy diễn** là thu thập thông tin về tổng thể từ các thông tin có trong mẫu.

151

GIỚI THIỆU VẤN ĐỀ LẤY MẪU



152

GIỚI THIỆU VẤN ĐỀ LẤY MẪU

- **Các trị thống kê mẫu:** Một đặc trưng của mẫu, như là trung bình mẫu \bar{x} , độ lệch chuẩn mẫu s , tỉ lệ mẫu \bar{p} . Giá trị của trị thống kê mẫu được dùng để ước lượng giá trị tham số của tổng thể

153

LẤY MẪU NGẪU NHIÊN ĐƠN GIẢN

- Định nghĩa của mẫu ngẫu nhiên đơn giản và quá trình lựa chọn một thể mẫu ngẫu nhiên đơn giản tùy thuộc vào tổng thể là hữu hạn hay vô hạn.
- **Tổng thể hữu hạn** thường được định nghĩa bằng một danh sách.
- **Tổng thể vô hạn** thường được định nghĩa là một quá trình đang diễn ra. Các phần tử của tổng thể vô hạn có thể không liệt kê được

154

LẤY MẪU NGẪU NHIÊN ĐƠN GIẢN

- **Lấy mẫu từ tổng thể hữu hạn**
 - Một mẫu ngẫu nhiên đơn giản cỡ mẫu n từ tổng thể hữu hạn cỡ N là một mẫu được chọn sao cho mỗi mẫu có thể với cỡ mẫu n đều có cùng xác suất được chọn
 - Số mẫu ngẫu nhiên đơn giản cỡ mẫu n khác nhau từ tổng thể hữu hạn cỡ N là:

$$\frac{N!}{n!(N - n)!}$$

155

LẤY MẪU NGẪU NHIÊN ĐƠN GIẢN

- **Lấy mẫu từ tổng thể hữu hạn**
 - **Lấy mẫu không thay thế:** Khi một phần tử được chọn vào mẫu thì nó được lấy ra khỏi tổng thể và không thể được chọn lần thứ hai
 - **Lấy mẫu có thay thế:** Khi một phần tử được chọn vào mẫu thì nó được bỏ trở lại tổng thể. Một phần tử được lựa chọn lần trước thì nó có thể được lựa chọn lần nữa và vì vậy phần tử đó có thể xuất hiện trong mẫu hơn một lần

156

LẤY MẪU NGẪU NHIÊN ĐƠN GIẢN

▪ **Lấy mẫu từ tổng thể vô hạn**

Một mẫu ngẫu nhiên đơn giản từ một tổng thể vô hạn là một một được chọn phải thỏa mãn các điều kiện sau:

- Mỗi phần tử được chọn phải đến từ cùng một tổng thể
- Mỗi phần tử được chọn một cách độc lập

157

ƯỚC LƯỢNG ĐIỂM

- Trong Ước lượng điểm chúng ta sử dụng dữ liệu từ mẫu để tính một giá trị của trị thống kê mẫu và dựa vào đó cung cấp một ước lượng về một tham số của tổng thể
- Ước lượng điểm là một trị thống kê mẫu, như là \bar{x} , s hay \bar{p} cung cấp ước lượng điểm về tham số của tổng thể, μ , σ và p .

158

GIỚI THIỆU PHÂN PHỐI MẪU

- Phân phối xác suất của bất kỳ trị thống kê mẫu cụ thể được gọi là phân phối mẫu của trị thống kê.
- Phân phối xác suất của \bar{x} được gọi là phân phối mẫu của \bar{x} . Kiến thức về phân phối mẫu này và các tính chất của nó sẽ cho phép chúng ta phát biểu về xác suất để cho trung bình của mẫu \bar{x} gần bằng với trung bình của tổng thể μ .
- Trong thực tế, chúng ta chỉ chọn một mẫu ngẫu nhiên đơn giản từ tổng thể

159

PHÂN PHỐI MẪU CỦA \bar{X}

- Phân phối mẫu của \bar{X}

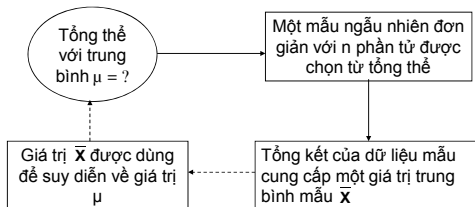
Phân phối mẫu của \bar{X} là phân phối xác suất của tất cả các giá trị có thể của trung bình mẫu \bar{X}

- Giá trị kỳ vọng của \bar{X}

$$E(\bar{X}) = \mu$$

160

PHÂN PHỐI MẪU CỦA \bar{X}



161

PHÂN PHỐI MẪU CỦA \bar{X}

- Độ lệch chuẩn của \bar{X}

- Tổng thể vô hạn hay không biết N $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

- Tổng thể hữu hạn hay biết N $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$

- Với $\sqrt{\frac{N-n}{N-1}}$ là nhân tố điều chỉnh tổng thể hữu hạn

162

PHÂN PHỐI MẪU CỦA \bar{X}

- Độ lệch chuẩn của \bar{X}
 - Bỏ qua nhân tố điều chỉnh tổng thể hữu hạn khi $n/N \leq 0.05$
 - Sai số chuẩn là độ lệch chuẩn của một ước lượng điểm
 - $\sigma_{\bar{x}}$ được xem như sai số chuẩn của trung bình

163

PHÂN PHỐI MẪU CỦA \bar{X}

- Phân phối của \bar{X}
 - Câu hỏi: Phân phối xác suất của \bar{X} là gì?
- Định lý giới hạn trung tâm
 - Phân phối của tổng thể được biết là phân phối chuẩn
$$X \sim N(\mu, \sigma^2) \longrightarrow \bar{X} \sim N(\mu, \sigma^2/n)$$

164

PHÂN PHỐI MẪU CỦA \bar{X}

- Định lý giới hạn trung tâm
 - Trong việc chọn các mẫu ngẫu nhiên đơn giản cỡ mẫu n từ một tổng thể, phân phối mẫu của trung bình mẫu \bar{X} có thể gần đúng tuân theo phân phối chuẩn khi cỡ mẫu đủ lớn.
- $X \sim$ Bất kỳ phân phối nào
 - Không biết phân phối xác suất tổng thể
 - Cỡ mẫu lớn ($N > 30$)
$$\longrightarrow \bar{X} \sim N(\mu, \sigma^2/n)$$

165

PHÂN PHỐI MẪU CỦA \bar{X}

- $\bar{X} \sim N(\mu, \sigma^2/n)$ với $Z \sim N(0,1^2)$

$$\sigma_{\bar{x}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

166

PHÂN PHỐI MẪU CỦA \bar{p}

- Phân phối mẫu của \bar{p}

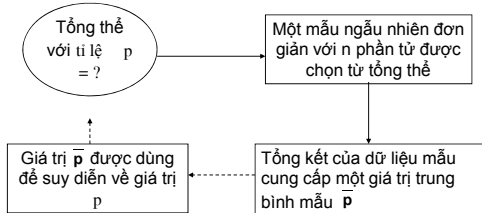
Phân phối mẫu của \bar{p} là phân phối xác suất của tất cả các giá trị có thể của tỉ lệ mẫu \bar{p}

- Giá trị kỳ vọng của \bar{p}

$$E(\bar{p}) = p$$

167

PHÂN PHỐI MẪU CỦA \bar{p}



168

PHÂN PHỐI MẪU CỦA \bar{p}

- Độ lệch chuẩn của \bar{p}
 - Tổng thể vô hạn: $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$
 - Tổng thể hữu hạn: $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}}$
- Bỏ qua nhân tử điều chỉnh tổng thể hữu hạn $\sqrt{\frac{N-n}{N-1}}$ khi $n/N < 0.05$

169

PHÂN PHỐI MẪU CỦA \bar{p}

- Dạng phân phối mẫu của \bar{p}

Phân phối mẫu của \bar{p} có thể gần đúng tuân theo phân phối xác suất chuẩn khi cỡ mẫu lớn

- $np \geq 5$
- $n(1-p) \geq 5$

170

TÍNH CHẤT CỦA ƯỚC LƯỢNG ĐIỂM

Gọi

- θ = tham số tổng thể được quan tâm
- $\hat{\theta}$ = trị thống kê mẫu hay ước lượng điểm của θ

- **Không thiên lệch**
 Trị thống kê mẫu $\hat{\theta}$ là một ước lượng không thiên lệch của tham số tổng thể θ nếu $E(\hat{\theta}) = \theta$
- **Độ thiên lệch (Bias)**
 $Bias(\hat{\theta}) = E(\hat{\theta}) - \theta$

171

TÍNH CHẤT CỦA ƯỚC LƯỢNG ĐIỂM

- **Độ hữu hiệu**

Cho hai ước lượng điểm không thiên lệch của tham số tổng thể, ước lượng điểm với độ lệch chuẩn nhỏ hơn $\hat{\theta}_1$ được xem là hiệu quả hơn $\hat{\theta}_2$ nếu:

$$\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2)$$

- **Độ nhất quán**

Một tính chất của ước lượng điểm được trình bày khi các cỡ mẫu lớn hơn sẽ cung cấp các ước lượng điểm gần với tham số của tổng thể

172

CÁC PHƯƠNG PHÁP LẤY MẪU KHÁC

- **Lấy mẫu hệ thống**

Một phương pháp lấy mẫu xác suất theo đó chúng ta sẽ chọn một cách ngẫu nhiên một trong k phần tử đầu tiên và sau đó chọn mỗi phần tử thứ k kế tiếp

- **Lấy mẫu thuận tiện**

Một phương pháp lấy mẫu phi xác suất theo đó các phần tử được chọn vào mẫu dựa trên cơ sở thuận tiện

173

CÁC PHƯƠNG PHÁP LẤY MẪU KHÁC

- **Lấy mẫu phán đoán**

Một phương pháp lấy mẫu phi xác suất theo đó các phần tử được chọn vào mẫu dựa trên sự phán đoán của người thực hiện nghiên cứu

174

CHƯƠNG 8

ƯỚC LƯỢNG KHOẢNG

175

NỘI DUNG CHÍNH

- Ước lượng khoảng của trung bình tổng thể: biết σ
- Ước lượng khoảng của trung bình tổng thể: không biết σ
- Xác định cỡ mẫu
- Ước lượng khoảng của tỉ lệ tổng thể

176

ƯỚC LƯỢNG KHOẢNG CỦA TRUNG BÌNH TỔNG THỂ: BIẾT σ

- Ước lượng khoảng là một ước lượng của một tham số của tổng thể theo đó cung cấp một khoảng được tin là sẽ chứa giá trị của tham số
- Trường hợp cỡ mẫu lớn: $n \geq 30$
- Trường hợp cỡ mẫu nhỏ: $n < 30$

177

ƯỚC LƯỢNG KHOẢNG CỦA TRUNG BÌNH TỔNG THỂ: BIẾT σ

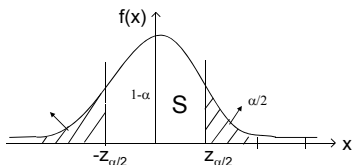
- Dạng tổng quát của ước lượng khoảng là:

Ước lượng điểm \pm Biên của sai số

- Biên của sai số là giá trị cộng và trừ vào ước lượng điểm để tạo ra một khoảng tin cậy
- Để tạo ra một khoảng tin cậy của μ , thì cả σ và s phải được sử dụng để tính biên của sai số

178

ƯỚC LƯỢNG KHOẢNG CỦA TRUNG BÌNH TỔNG THỂ: BIẾT σ



$$P(Z > Z_{\alpha/2}) = \alpha/2$$

$$P(Z < -Z_{\alpha/2}) = \alpha/2$$

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha$$

$Z_{\alpha/2}$: là giá trị của biến phân phối chuẩn chuẩn hóa tương ứng với một diện tích $\alpha/2$ ở dưới đuôi phía trên của phân phối

179

ƯỚC LƯỢNG KHOẢNG CỦA TRUNG BÌNH TỔNG THỂ: BIẾT σ

$$P\left(-Z_{\alpha/2} < \frac{x - \mu}{\sigma / \sqrt{n}} < Z_{\alpha/2}\right) = 1 - \alpha$$

$$P\left(\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

180

ƯỚC LƯỢNG KHOẢNG CỦA TRUNG BÌNH TỔNG THỂ: BIẾT σ

- Tính ước lượng khoảng: biết σ

$$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Với:

- $(1-\alpha)$ là độ tin cậy
- \bar{x} là ước lượng điểm của μ
- $\pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ là biên của sai số
- Cỡ mẫu lớn ($n \geq 30$) \rightarrow dùng công thức này

181

ƯỚC LƯỢNG KHOẢNG CỦA TRUNG BÌNH TỔNG THỂ: BIẾT σ

CÁC GIÁ TRỊ CỦA $Z_{\alpha/2}$ ĐỐI VỚI CÁC MỨC TIN CẬY ĐƯỢC SỬ DỤNG PHỔ BIẾN NHẤT

Mức tin cậy	α	$\alpha/2$	$Z_{\alpha/2}$
90%	.10	.050	1.645
95%	.05	.025	1.960
99%	.01	.005	2.576

182

ƯỚC LƯỢNG KHOẢNG CỦA TRUNG BÌNH TỔNG THỂ: BIẾT σ

- Tính ước lượng khoảng: biết σ
 - Biên của sai số là giá trị cộng và trừ vào ước lượng điểm để tạo ra một khoảng tin cậy
 - Khoảng tin cậy: Một khoảng tin cậy $100(1 - \alpha)\%$ đối với trung bình của phân phối chuẩn μ là

$$\left[\bar{x} - Z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}} \right), \bar{x} + Z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}} \right) \right]$$

183

ƯỚC LƯỢNG KHOẢNG CỦA TRUNG BÌNH TỔNG THỂ: KHÔNG BIẾT σ

- Nếu không biết σ , độ lệch chuẩn của mẫu s được dùng để ước lượng độ lệch chuẩn của tổng thể σ và khoảng tin cậy thích hợp sẽ dựa trên một phân phối xác suất được gọi là phân phối t
- Trị thống kê t :

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

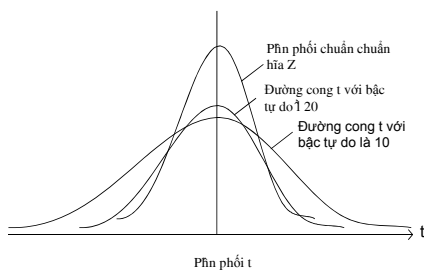
184

ƯỚC LƯỢNG KHOẢNG CỦA TRUNG BÌNH TỔNG THỂ: KHÔNG BIẾT σ

- Trị thống kê t sẽ tuân theo một Phân phối Student's t , với độ tự do d_f
- $$d_f = n - 1$$
- Phân phối t thường được dùng với phân phối cỡ mẫu nhỏ của \bar{x}
 - If $n \rightarrow N$ then $t \# Z$

185

ƯỚC LƯỢNG KHOẢNG CỦA TRUNG BÌNH TỔNG THỂ: KHÔNG BIẾT σ



186

ƯỚC LƯỢNG KHOẢNG CỦA TRUNG BÌNH TỔNG THỂ: KHÔNG BIẾT σ

- Ước lượng khoảng của một trung bình tổng thể: không biết σ

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

- Cỡ mẫu nhỏ ($n < 30$) và tổng thể tuân theo một phân phối chuẩn hoặc gần chuẩn \rightarrow cũng dùng công thức này

187

TÓM TẮT CÁC THỦ TỤC ƯỚC LƯỢNG KHOẢNG ĐỐI VỚI TRUNG BÌNH TỔNG THỂ

```

    graph TD
      A{Có thể giả sử độ lệch chuẩn của tổng thể đã biết?} -- Có --> B[Dùng  
x̄ ± z_{α/2} * σ / √n  
Trường hợp biết σ]
      A -- Không --> C[Dùng độ lệch chuẩn của mẫu s để ước lượng σ]
      C --> D[Dùng  
x̄ ± t_{α/2} * s / √n  
Trường hợp không biết σ]
  
```

188

XÁC ĐỊNH CỠ MẪU

- Gọi E = biên của sai số kỳ vọng

$$E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- Cỡ mẫu đối với ước lượng khoảng của một trung bình của tổng thể $n = \frac{(Z_{\alpha/2})^2 \sigma^2}{E^2}$
- Cỡ mẫu đối với không biết σ $n = \frac{(Z_{\alpha/2})^2 s^2}{E^2}$

189

ƯỚC LƯỢNG KHOẢNG CỦA TỈ LỆ TỔNG THỂ

$$E(\bar{p}) = p$$

$$\sigma_p = \sqrt{p(1-p)/n}$$

- Ước lượng khoảng của tỉ lệ tổng thể

hay
$$\bar{p} \pm Z_{\alpha/2} \sigma_{\bar{p}}$$

$$\bar{p} \pm Z_{\alpha/2} \sqrt{\bar{p}(1-\bar{p})/n}$$

190

ƯỚC LƯỢNG KHOẢNG CỦA TỈ LỆ TỔNG THỂ

- Xác định cỡ mẫu
Gọi E = biên của sai số kỳ vọng

$$E = Z_{\alpha/2} \sqrt{p(1-p)/n}$$

$$n = \frac{(Z_{\alpha/2})^2 p(1-p)}{E^2}$$

191

CHƯƠNG 9

KIỂM ĐỊNH GIẢ THUYẾT

192

NỘI DUNG CHÍNH

- Phát triển giả thuyết không và giả thuyết khác
- Các sai lầm loại I và loại II
- Kiểm định một-phía về trung bình của tổng thể: biết σ
- Kiểm định hai-phía về trung bình của tổng thể: biết σ
- Kiểm định về trung bình của tổng thể: không biết σ
- Kiểm định về tỉ lệ của tổng thể

193

PHÁT TRIỂN GIẢ THUYẾT KHÔNG và GIẢ THUYẾT KHÁC

Giả thuyết

- Giả thuyết là một giả sử hay phát biểu về các tham số của tổng thể; Nó có thể đúng hoặc sai

Giả thuyết Không (H_0)

- H_0 là một phát biểu (đẳng thức hoặc bất đẳng thức) liên quan đến tham số của tổng thể
- H_0 là một giả định đúng trong thủ tục kiểm định giả thuyết
- Một tuyên bố của nhà sản xuất thường bị nghi ngờ và được phát biểu trong H_0

194

PHÁT TRIỂN GIẢ THUYẾT KHÔNG và GIẢ THUYẾT KHÁC

Giả thuyết khác (H_a)

- H_a là phát biểu ngược với H_0
- H_a được kết luận là đúng nếu H_0 bị bác bỏ
- Nhà nghiên cứu mong muốn ủng hộ H_a và nghi ngờ H_0

Tổng kết các dạng của giả thuyết Không và giả thuyết khác

- $H_0 : \mu = \mu_0$ or $H_0 : \mu \leq \mu_0$ or $H_0 : \mu \geq \mu_0$
- $H_a : \mu \neq \mu_0$ $H_a : \mu > \mu_0$ $H_a : \mu < \mu_0$

Nhiệm vụ của tất cả kiểm định giả thuyết hoặc là bác bỏ H_0 hay không bác bỏ H_0 (\neq Accept H_0)

195

CÁC SAI LẦM LOẠI I VÀ LOẠI II

- Sai lầm loại I là sai lầm của việc bác bỏ H_0 khi nó đúng
- Sai lầm loại II là sai lầm của việc không bác bỏ H_0 khi nó sai

{ CÁC KẾT LUẬN ĐÚNG VÀ SAI TRONG KIỂM ĐỊNH GIÁ THUYẾT

{ Điều kiện của tổng thể
 { H_0 đúng H_0 sai

CÁC SAI LẦM LOẠI I VÀ LOẠI II

• α là xác suất của sai lầm loại I

- $\alpha = P(\text{Bác bỏ } H_0 / H_0 \text{ đúng}) = P(\text{Sai lầm loại I})$
- α được gọi là mức ý nghĩa của kiểm định, $0.01 < \alpha < 0.1$
- Thường chọn $\alpha = 0.05$

• β là xác suất của sai lầm loại II

- $\beta = P(\text{Không bác bỏ } H_0 / H_0 \text{ sai}) = \text{Sai lầm loại II}$
- $(1-\beta) = P(\text{Bác bỏ } H_0 / H_0 \text{ sai}) = \text{Năng lực của kiểm định}$
- α càng nhỏ thì β càng lớn

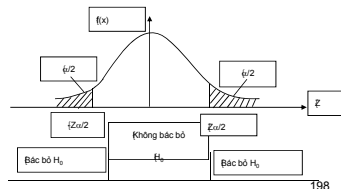
197

MIỀN BÁC BỎ

Một miền bác bỏ R định rõ các giá trị của trị thống kê sẽ chỉ dẫn cho chúng ta bác bỏ H_0

Kiểm định 2-phía

$H_0 : \mu = \mu_0$
 $H_a : \mu \neq \mu_0$



198

MIỀN BÁC BỎ

Kiểm định 1-phía

$H_0 : \mu \geq \mu_0$
 $H_a : \mu < \mu_0$

$H_0 : \mu \leq \mu_0$
 $H_a : \mu > \mu_0$

199

KIỂM ĐỊNH 1-PHÍA VỀ TRUNG BÌNH CỦA TỔNG THỂ: BIẾT σ

Giả thuyết

<p>Trường hợp 1</p> <p>$H_0 : \mu \geq \mu_0$ $H_a : \mu < \mu_0$</p>	<p>Trường hợp 2</p> <p>$H_0 : \mu \leq \mu_0$ $H_a : \mu > \mu_0$</p>
---	---

Trị thống kê

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

200

KIỂM ĐỊNH 1-PHÍA VỀ TRUNG BÌNH CỦA TỔNG THỂ: BIẾT σ

Phương pháp p-value

- **p-value**
p-value là xác suất, được tính từ trị thống kê, đo lường mức độ ủng hộ (hay không ủng hộ) cung cấp bởi mẫu đối với giả thuyết H_0
- **Tiêu chí p-value đối với kiểm định giả thuyết**
Bác bỏ H_0 nếu $p\text{-value} < \alpha$

201

KIỂM ĐỊNH 1-PHÍA VỀ TRUNG BÌNH CỦA TỔNG THỂ: BIẾT σ

Phương pháp giá trị tới hạn
(Qui tắc bác bỏ)

Bác bỏ H_0 nếu $Z < -Z_\alpha$ Bác bỏ H_0 nếu $Z > Z_\alpha$

202

KIỂM ĐỊNH 2-PHÍA VỀ TRUNG BÌNH CỦA TỔNG THỂ: BIẾT σ

Giả thuyết:
 $H_0 : \mu = \mu_0$
 $H_a : \mu \neq \mu_0$

Trị thống kê:

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

203

KIỂM ĐỊNH 2-PHÍA VỀ TRUNG BÌNH CỦA TỔNG THỂ: BIẾT σ

p-value đối với kiểm định 2-phía

- Trong kiểm định 2-phía, p-value được tính bằng cách nhân đôi diện tích ở phần đuôi của phân phối
- Vì diện tích được nhân đôi nên p-value có thể so sánh trực tiếp với α và qui tắc bác bỏ vẫn giống như trước
- **Bác bỏ H_0 nếu p-value $< \alpha$**

204

KIỂM ĐỊNH 2-PHÍA VỀ TRUNG BÌNH CỦA TỔNG THỂ: BIẾT σ

Phương pháp giá trị tới hạn
(Qui tắc bác bỏ)

Bác bỏ H_0 nếu $Z < -Z_{\alpha/2}$ Bác bỏ H_0 nếu $Z > Z_{\alpha/2}$

205

KIỂM ĐỊNH 2-PHÍA VỀ TRUNG BÌNH CỦA TỔNG THỂ: BIẾT σ

Mối liên hệ giữa ước lượng khoảng và kiểm định giả thuyết
Một phương pháp khoảng tin cậy để kiểm định giả thuyết dưới dạng:

$H_0 : \mu = \mu_0$
 $H_a : \mu \neq \mu_0$

Chọn một mẫu ngẫu nhiên đơn giản từ tổng thể và dùng giá trị của trung bình của mẫu để phát triển khoảng tin cậy đối với μ .

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Nếu khoảng tin cậy chứa giá trị được giả thuyết μ_0 , thì không bác bỏ H_0 . Nếu không chứa thì bác bỏ H_0

206

CÁC BƯỚC KIỂM ĐỊNH GIẢ THUYẾT

- Bước 1: Phát triển H_0 và H_a
- Bước 2: Định mức ý nghĩa α
- Bước 3: Thu thập dữ liệu mẫu và tính trị thống kê kiểm định
- Phương pháp p-value**
- Bước 4: Dùng giá trị của trị thống kê kiểm định để tính p-value
- Bước 5: Bác bỏ H_0 nếu p-value $< \alpha$
- Phương pháp giá trị tới hạn**
- Bước 4: Dùng α để xác định giá trị tới hạn và qui tắc bác bỏ
- Bước 5: Dùng giá trị của trị thống kê kiểm định và qui tắc bác bỏ để xác định xem có bác bỏ H_0 hay không

207

KIỂM ĐỊNH VỀ TRUNG BÌNH CỦA TỔNG THỂ: KHÔNG BIẾT σ

- s được dùng để ước lượng σ
- Phân phối t có thể được dùng để suy diễn về μ
- Trị thống kê kiểm định là

$$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$$

- df = n-1
- Cỡ mẫu nhỏ (n < 30) và tổng thể tuân theo một phân phối chuẩn hoặc gần chuẩn → cũng dùng công thức này

208

KIỂM ĐỊNH VỀ TRUNG BÌNH CỦA TỔNG THỂ: KHÔNG BIẾT σ

Kiểm định 1-phía

$H_0 : \mu \geq \mu_0$

$H_0 : \mu \leq \mu_0$

$H_a : \mu < \mu_0$

$H_a : \mu > \mu_0$

Bác bỏ H_0 nếu $t < -t_{\alpha, n-1}$

Bác bỏ H_0 nếu $t > t_{\alpha, n-1}$

Kiểm định 2-phía

$H_0 : \mu = \mu_0$

$H_a : \mu \neq \mu_0$

Bác bỏ H_0 nếu $t < -t_{\alpha/2, n-1}$ hay nếu $t > t_{\alpha/2, n-1}$

209

KIỂM ĐỊNH VỀ TRUNG BÌNH CỦA TỔNG THỂ: KHÔNG BIẾT σ

p-value và phân phối t

$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$ Dùng bảng t $\xrightarrow{df = n-1}$ p-value

Bác bỏ H_0 nếu p-value < α

210

KIỂM ĐỊNH VỀ TỈ LỆ CỦA TỔNG THỂ

Gọi p : tỉ lệ của tổng thể
 p_0 : giá trị cụ thể của giả thuyết đối với tỉ lệ của tổng thể

Giả thuyết

$H_0 : p \geq p_0$ $H : p \leq p_0$ $H_0 : p = p$
 $H_a : p < p_0$ $H : p > p_0$ $H_a : p \neq p_0$

Trị thống kê kiểm định

$$Z = \frac{p - p_0}{\sigma_{\bar{p}}}$$

$$\sigma_{\bar{p}} = \sqrt{\frac{p_0(1-p_0)}{n}}$$

211

KIỂM ĐỊNH VỀ TỈ LỆ CỦA TỔNG THỂ

Quy tắc bác bỏ

211

CHƯƠNG 10

HỒI QUI TUYẾN TÍNH ĐƠN

HỒI QUI TUYẾN TÍNH ĐƠN

- Mô hồi qui tuyến tính đơn
- Phương pháp bình phương tối thiểu
- Hệ số xác định
- Các giả định của mô hình
- Kiểm định mức ý nghĩa
- Sử dụng mô hình hồi qui ước lượng để ước lượng và dự đoán

214

MÔ HÌNH HỒI QUI TUYẾN TÍNH

- Phương trình mô tả y liên hệ với x như thế nào và một số hạng sai số được gọi là mô hình hồi qui.
- Mô hình hồi qui tuyến tính đơn là:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Với:

- { β_0 và β_1 được gọi là các tham số của mô hình,
- { ε là biến ngẫu nhiên được gọi là số hạng sai số.

215

PHƯƠNG TRÌNH

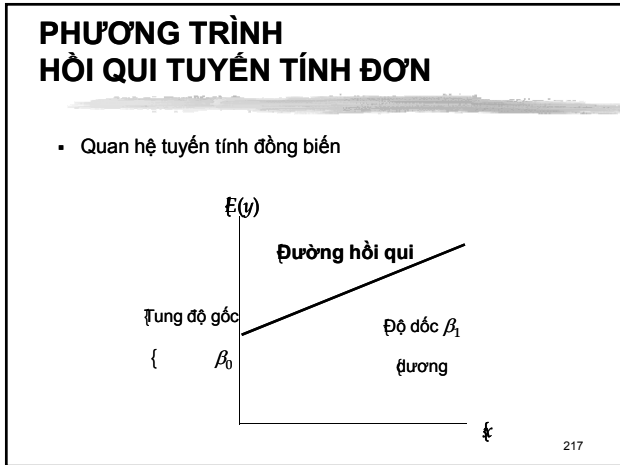
HỒI QUI TUYẾN TÍNH ĐƠN

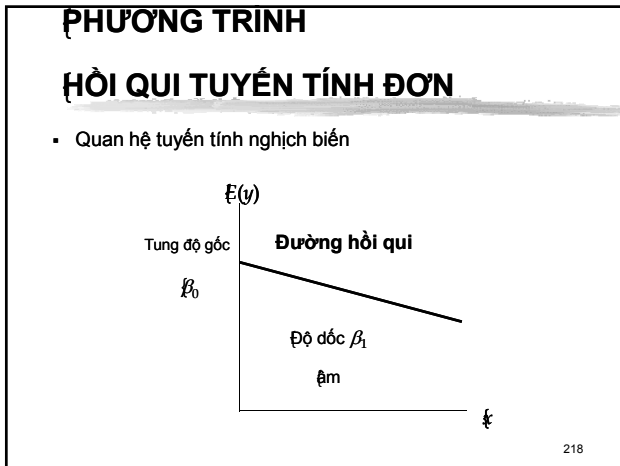
- Phương trình hồi qui tuyến tính đơn là:

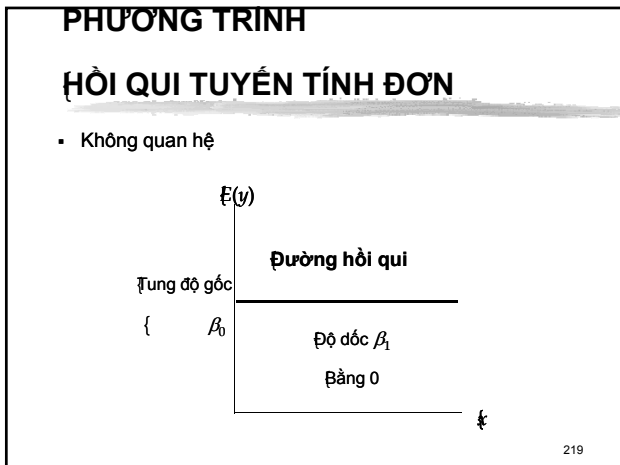
$$E(y) = \beta_0 + \beta_1 x$$

- Đồ thị của phương trình hồi qui là đường thẳng.
- β_0 là tung độ gốc của đường hồi qui
- β_1 là độ dốc của đường hồi qui
- $E(y)$ là giá trị kỳ vọng của y đối với giá trị x cho trước.

216







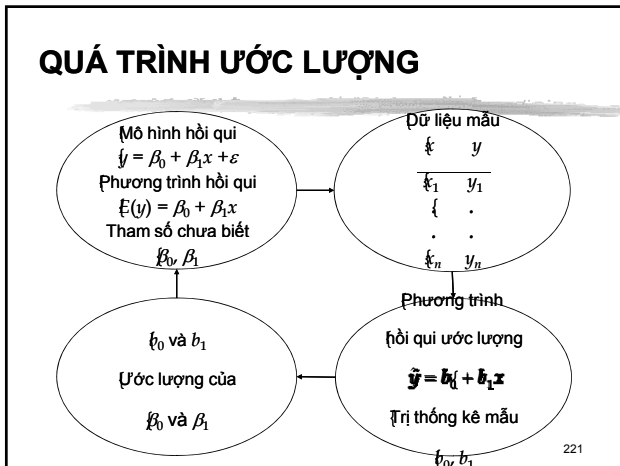
PHƯƠNG TRÌNH HỒI QUI TUYẾN TÍNH ĐƠN ƯỚC LƯỢNG

- Phương trình hồi qui tuyến tính đơn ước lượng

$$\hat{y} = b_0 + b_1x$$

- Đồ thị được gọi là đường hồi qui ước lượng.
- b_0 là tung độ gốc của đường.
- b_1 là độ dốc của đường
- \hat{y} là giá trị ước lượng của y đối với giá trị x cho trước.

220



PHƯƠNG PHÁP BÌNH PHƯƠNG TỐI THIỂU

- Tiêu chí bình phương tối thiểu

$$\min \sum (y_i - \hat{y}_i)^2$$

Với:

- y_i = giá trị quan sát của biến phụ thuộc
- \hat{y}_i = giá trị ước lượng của biến phụ thuộc đối với quan sát thứ i

222

PHƯƠNG PHÁP BÌNH PHƯƠNG TỐI THIỂU

- Độ dốc của phương trình hồi qui ước lượng

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

223

PHƯƠNG PHÁP BÌNH PHƯƠNG TỐI THIỂU

- Tung độ gốc của phương trình hồi qui ước lượng

$$b_0 = \bar{y} - b_1 \bar{x}$$

Với:

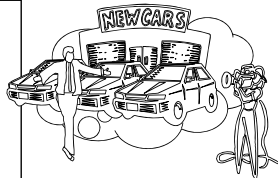
- x_i = giá trị của biến độc lập đối với quan sát thứ i
- y_i = giá trị của biến phụ thuộc đối với quan sát thứ i
- \bar{x} = giá trị trung bình của biến độc lập
- \bar{y} = giá trị trung bình của biến phụ thuộc
- n = tổng số quan sát

224

HỒI QUI TUYẾN TÍNH ĐƠN

- Ví dụ: Doanh số xe hơi

Quảng cáo	Doanh số
TV	xe hơi
1	14
3	24
2	18
1	17



225

PHƯƠNG TRÌNH HỒI QUI ƯỚC LƯỢNG



- Độ dốc của phương trình hồi qui ước lượng

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{20}{4} = 5$$

- Tung độ gốc của phương trình hồi qui ước lượng

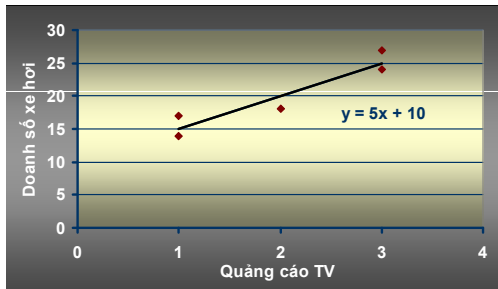
$$b_0 = \bar{y} - b_1 \bar{x} = 20 - 5(2) = 10$$

- Phương trình hồi qui ước lượng

$$\hat{y} = 10 + 5x$$

226

ĐỒ THỊ PHÂN TÁN ĐIỂM VÀ ĐƯỜNG XU HƯỚNG



227

HỆ SỐ XÁC ĐỊNH

- Mối liên hệ giữa SST, SSR, SSE

$$SST = SSR + SSE$$

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2$$

Với:

- SST = Tổng bình phương toàn phần
- SSR = Tổng bình phương hồi qui
- SSE = Tổng bình phương sai số

228

HỆ SỐ XÁC ĐỊNH

- Hệ số xác định là:

$$f^2 = SSR/SST$$

Với:

SSR = Tổng bình phương hồi qui

SST = Tổng bình phương toàn phần

229

HỆ SỐ XÁC ĐỊNH



$$f^2 = SSR/SST = 100/114 = .8772$$

Mối quan hệ hồi qui rất mạnh; 88%

Sự biến thiên của doanh số xe hơi có thể được giải thích bởi mối quan hệ tuyến tính giữa số quảng cáo trên TV và doanh số xe hơi.

230

HỆ SỐ TƯƠNG QUAN MẪU

$$r_{xy} = (\text{sign of } b_1) \sqrt{\text{Coefficient of Determination}}$$

$$r_{xy} = (\text{sign of } b_1) \sqrt{r^2}$$

Với

b_1 = độ dốc của phương trình hồi qui ước lượng

$$\hat{y} = b_0 + b_1x$$

231

HỆ SỐ TƯƠNG QUAN MẪU

$$r_{xy} = (\text{sign of } b_1) \sqrt{r^2}$$

Dấu của b_1 trong phương trình $\hat{y} = 10 + 5x$ là "+".

$$r_{xy} = +\sqrt{0.9772}$$

$$r_{xy} = +.9366$$

232

CÁC GIẢ ĐỊNH VỀ SỐ HẠNG SAI SỐ ε

1. Sai số ε là biến ngẫu nhiên với trung bình bằng 0
2. Phương sai của ε , ký hiệu σ^2 , sẽ giống nhau đối với tất cả các giá trị của biến độc lập.
3. Các giá trị của ε là độc lập.
4. Sai số ε là biến ngẫu nhiên tuân theo phân phối chuẩn

233

KIỂM ĐỊNH Ý NGHĨA

Để kiểm định ý nghĩa của mối quan hệ hồi qui, chúng ta phải tiến hành kiểm định giả thuyết xác định xem giá trị của β_1 có bằng 0 hay không.

{Hai kiểm định được dùng phổ biến là:

t Test và F Test

{Cả kiểm định t và F đều yêu cầu một ước lượng của σ^2 , phương sai của ε trong mô hình hồi qui

234

KIỂM ĐỊNH Ý NGHĨA

- Ước lượng của σ

Trung bình của sai số bình phương (MSE) sẽ cung cấp
 ước lượng của σ^2 , và s^2 cũng được sử dụng.

$$\hat{\sigma}^2 = \text{MSE} = \text{SSE} / (n - 2)$$

Với

$$\text{SSE} = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - b_0 - b_1 x_i)^2$$

235

KIỂM ĐỊNH Ý NGHĨA

- Ước lượng của σ

- Để ước lượng σ chúng ta lấy căn bậc hai của σ^2 .
- Kết quả s được gọi là sai số chuẩn của ước lượng

$$s = \sqrt{\text{MSE}} = \sqrt{\frac{\text{SSE}}{n - 2}}$$

236

KIỂM ĐỊNH Ý NGHĨA: KIỂM ĐỊNH t

- Giả thuyết

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

- Trị thống kê

$$t = \frac{b_1}{s_{b_1}}$$

237

KIỂM ĐỊNH Ý NGHĨA: KIỂM ĐỊNH t

- Qui tắc bác bỏ

Bác bỏ H_0 nếu $p\text{-value} \leq \alpha$

Với $t \leq -t_{\alpha/2}$ hay $t \geq t_{\alpha/2}$

$t_{\alpha/2}$ được dựa trên phân phối t

với bậc tự do là $n - 2$

238

KIỂM ĐỊNH Ý NGHĨA: KIỂM ĐỊNH t



1. Xác định giả thuyết.

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

2. Xác định mức ý nghĩa.

$$\alpha = .05$$

3. Lựa chọn trị thống kê kiểm định.

$$t = \frac{b_1}{s_{b_1}}$$

4. Qui tắc bác bỏ.

Bác bỏ H_0 nếu $p\text{-value} \leq .05$

hay $|t| > 3.182$ (với độ tự do là 3)

239

KIỂM ĐỊNH Ý NGHĨA: KIỂM ĐỊNH t



5. Tính giá trị của trị thống kê kiểm định.

$$t = \frac{b_1}{s_{b_1}} = \frac{5}{1.08} = 4.63$$

6. Xác định có bác bỏ H_0 hay không

$t = 4.541$ cho một diện tích .01 ở phần đuôi

phía trên. Vì vậy, $p\text{-value}$ nhỏ hơn .02. (Cũng vậy,

$t = 4.63 > 3.182$.) chúng ta có thể bác bỏ H_0 .

240

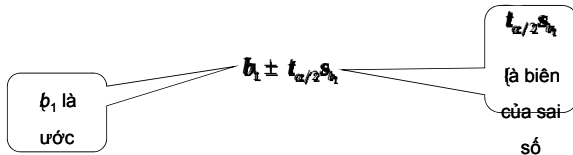
KHOẢNG TIN CẬY CỦA β_1

- Chúng ta có thể dùng một khoảng tin cậy 95% của β_1 để kiểm định các giả thuyết vừa mới dùng trong
 - { kiểm định t.
 - H_0 bị bác bỏ nếu giá trị được giả thuyết của β_1 không nằm trong khoảng tin cậy của β_1 .

241

KHOẢNG TIN CẬY CỦA β_1

- Công thức của khoảng tin cậy đối với β_1 là:



Với $t_{\alpha/2}$ là giá trị t cho một diện tích là $\alpha/2$ trong phần đuôi phía trên của phân phối t với độ tự do là $n - 2$

242

KHOẢNG TIN CẬY CỦA β_1



- Qui tắc bác bỏ
Bác bỏ H_0 nếu 0 không nằm trong khoảng tin cậy của β_1 .
- Khoảng tin cậy 95% của β_1
 $\hat{\beta}_1 \pm t_{\alpha/2} \cdot s_{\hat{\beta}_1} = 5 \pm /- 3.182(1.08) = 5 \pm /- 3.44$
tức 1.56 to 8.44
- Kết luận
0 không nằm trong khoảng tin cậy .
Bác bỏ H_0

243

KIỂM ĐỊNH Ý NGHĨA: KIỂM ĐỊNH F

- Giả thuyết

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

- Trị thống kê

$$f = MSR/MSE$$

244

KIỂM ĐỊNH Ý NGHĨA: KIỂM ĐỊNH F

- Quy tắc bác bỏ

Bác bỏ H_0 nếu

$$p\text{-value} \leq \alpha$$

Với:

$$\text{hay } F \geq F_\alpha$$

F_α dựa trên phân phối F với độ tự do là 1

ở tử số và độ tự do là $n - 2$ ở mẫu số

245

KIỂM ĐỊNH Ý NGHĨA: KIỂM ĐỊNH F

- Xác định giả thuyết.

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

- Xác định mức ý nghĩa.

$$\alpha = .05$$

- Lựa chọn trị thống kê kiểm định.

$$f = MSR/MSE$$

- Quy tắc bác bỏ.

Bác bỏ H_0 nếu $p\text{-value} \leq .05$

hay $F \geq 10.13$ (với d.f ở tử số

là 1 và d.f ở mẫu số là 3)

246



KIỂM ĐỊNH Ý NGHĨA: KIỂM ĐỊNH F



5. Tính giá trị của trị thống kê kiểm định.

$$F = MSR/MSE = 100/4.667 = 21.43$$

6. Xác định có bác bỏ H_0 hay không

{ $F = 17.44$ cho một diện tích là .025 ở phần đuôi

phía trên. Vì vậy, p -value tương ứng với $F = 21.43$

sẽ nhỏ hơn $2(.025) = .05$. Vì vậy, chúng ta bác bỏ

Chứng cứ thống kê đủ để kết luận có mối quan hệ
có ý nghĩa giữa số lượng quảng cáo TV và doanh số
xe hơi

247

MỘT VÀI LƯU Ý VỀ GIẢI THÍCH CÁC KIỂM ĐỊNH Ý NGHĨA

▪ Bác bỏ $H_0: \beta_1 = 0$ và kết luận rằng mối quan hệ giữa x và y có ý nghĩa, chúng ta không thể kết

luận có mối quan hệ nhân quả giữa x và y

▪ Bởi vì chúng ta có thể bác bỏ $H_0: \beta_1 = 0$ và

{ chúng ta có ý nghĩa về thống kê, chúng ta không thể kết

{ luận có mối quan hệ tuyến tính giữa x và y

248

SỬ DỤNG PHƯƠNG TRÌNH HỒI QUI ƯỚC LƯỢNG ĐỂ ƯỚC LƯỢNG VÀ DỰ ĐOÁN

• Khoảng tin cậy của $E(y_p)$

$$\hat{y}_p \pm t_{\alpha/2} s_{\hat{y}_p}$$

• Dự đoán khoảng tin cậy của y_p

$$\hat{y}_p \pm t_{\alpha/2} s_{\text{pred}}$$

Với:

hệ số tin cậy là $1 - \alpha$ và
 $t_{\alpha/2}$ dựa trên phân phối t
với độ tự do là $n - 2$

249

ƯỚC LƯỢNG ĐIỂM



Nếu 3 quảng cáo TV được tiến hành trước khi bán, chúng ta kỳ vọng trung bình doanh số xe hơi là:

$$\hat{y} = 10 + 5(3) = 25 \text{ xe hơi}$$

250

KHOẢNG TIN CẬY CỦA $E(y_p)$



• Kết quả khoảng tin cậy từ Excel

D	E	F	G
1	CONFIDENCE INTERVAL		
2	x_p	3	
3	\bar{x}	2.0	
4	$x_p - \bar{x}$	1.0	
5	$(x_p - \bar{x})^2$	1.0	
6	$\sum (x_p - \bar{x})^2$	4.0	
7	Variance of \hat{y}	2.1000	
8	Std. Dev of \hat{y}	1.4491	
9	t Value	3.1824	
10	Margin of Error	4.6118	
11	Point Estimate	25.0	
12	Lower Limit	20.39	
13	Upper Limit	29.61	

251

KHOẢNG TIN CẬY CỦA $E(y_p)$



Ước lượng khoảng tin cậy 95% của trung bình doanh số xe hơi khi 3 quảng cáo TV được thực hiện là:

$$\hat{y} \pm 4.61 = 20.39 \text{ to } 29.61 \text{ xe hơi}$$

252

KHOẢNG DỰ ĐOÁN CỦA y_p



- Kết quả khoảng dự đoán từ Excel

	H	I
1	PREDICTION INTERVAL	
2	Variance of y_{ind}	6.76667
3	Std. Dev. of y_{ind}	2.60128
4	Margin of Error	8.27845
5	Lower Limit	16.72
6	Upper Limit	33.28
7		

253

KHOẢNG DỰ ĐOÁN CỦA y_p



Ước lượng khoảng dự đoán 95% của doanh số xe hơi trong 1 tuần cụ thể khi 3 quảng cáo TV được thực hiện là:

$$25 \pm 8.28 = 16.72 \text{ to } 33.28 \text{ xe hơi}$$

254

PHÂN TÍCH PHẦN DƯ

- Nếu các giả định về số hạng sai số ϵ không đảm bảo thì các kiểm định giả thuyết về ý nghĩa của mối quan hệ hồi qui và các kết quả ước lượng khoảng không còn
- Các phần dư sẽ cho thông tin tốt nhất về ϵ .
- Phần dư của quan sát thứ i

$$y_i - \hat{y}_i$$

- Rất nhiều phân tích phần dư dựa trên việc khảo sát đồ thị phần dư

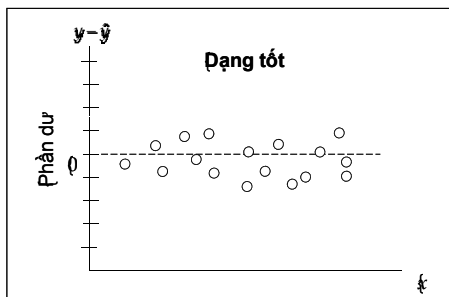
255

ĐỒ THỊ PHẦN DƯ THEO X

- Nếu giả định phương sai của ε giống nhau đối với tất cả các giá trị của x được thỏa, và mô hình hồi qui giả định là một biểu diễn đầy đủ của mối quan hệ giữa các biến, thì
 - { Đồ thị phần dư sẽ cho một ấn tượng
 - { tổng thể về giải bằng các điểm nằm ngang
 - {

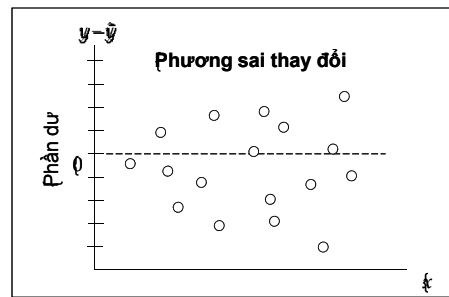
256

ĐỒ THỊ PHẦN DƯ THEO X



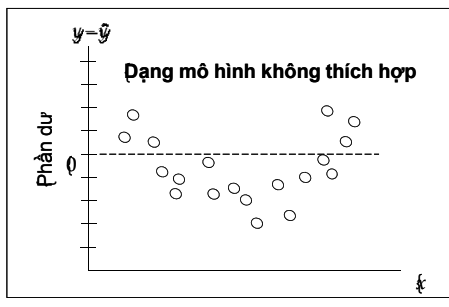
257

ĐỒ THỊ PHẦN DƯ THEO X



258

ĐỒ THỊ PHẦN DƯ THEO X



259

ĐỒ THỊ PHẦN DƯ THEO X

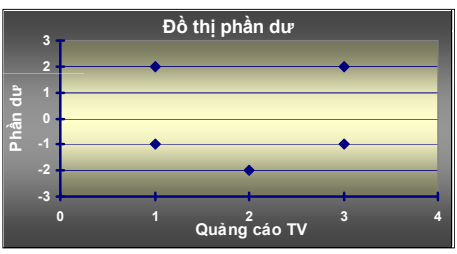


- Phần dư

Observation	Predicted Cars Sold	Residuals
1	15	-1
2	25	-1
3	20	-2
4	15	2
5	25	2

260

ĐỒ THỊ PHẦN DƯ THEO X



261
