Thống kê ứng dụng kinh doanh

Trần Tuấn Anh
Giảng viên chính
Khoa QTKD - Trường đại học Mở TPHCM
anh.tt@ou.edu.vn

Thống kê ứng dụng trong kinh doanh
Trần Tuấn Anh

Mục tiêu môn học

- Hiểu rõ các khái niệm cở bản của xác suất và thống kê ứng dụng trong lĩnh vực kinh doanh.
- Nắm vững phương pháp xác suất và thống kê cơ bản như: các phương pháp tính xác suất và các trường hợp sử dụng thích hợp, các phương pháp tóm tắt và trình bày dữ liệu của thống kê mô tả và một số phương pháp cơ bản của thống kê suy diễn. Đặc biệt là các phương pháp ước lượng, kiểm định giả thuyết, phân tích phương sai, tương quan và hồi qui tuyến tính.
- Có đủ kiến thức để học tiếp các môn liên quan đến phương pháp định lượng.

2

Nội dung chính

3

Phương pháp giảng dạy

- Phương pháp diễn giảng và hướng dẫn sinh viên tự nghiên cứu thêm dựa trên tài liệu học tập.
- Hướng dẫn sinh viên vận dụng lý thuyết để giải các bài tập.
- Hướng dẫn sinh viên ứng dụng phần mềm Excel và SPSS trong một số phân tích thống kê cơ bản.

4

Thời lượng & đánh giá

- Thời lượng môn học: 50 tiết.
- Phương pháp đánh giá:
  - Kiểm tra giữa kỳ & quá trình học: 30% tổng điểm.
  - Kiểm tra cuối kỳ: 70% tổng điểm.

5

Tài liệu học tập – tham khảo

- Trần Tuấn Anh, Bài giảng Thống kê ứng dụng trong kinh doanh.
- Trần Tuấn Anh, Thống kê ứng dụng trong kinh doanh, NXB Thống kê, 2011.
- Trần Bá Nhẫn, Đinh Thái Hoàng, Thống kê ứng dụng trong quản trị, kinh doanh và nghiên cứu kinh tế, NXB Thống kê, 2007.
- Trần Bá Nhẫn, Đinh Thái Hoàng, Bài tập Thống kê ứng dụng trong quản trị, kinh doanh và nghiên cứu kinh tế, NXB Thống kê, 2007.

6

"Trong cách học, phải lấy tự học làm cốt"

Hồ Chí Minh

7

Nhập môn Thống kê ứng dụng kinh doanh

Chương 1

Thống kê ứng dụng trong kinh doanh
Trần Tuấn Anh

Nội dung chính

- Thống kê ứng dụng là gì?
- Thống kê mô tả và thống kê suy diễn là gì?
- Sự khác nhau giữa biến định tính và biến định lượng.
- Sự khác nhau giữa biến rời rạc và biến liên tục.
- Hiểu được 4 loại thang đo: danh nghĩa, thứ bậc, khoảng và thứ tự.

2

Giới thiệu Thống kê ứng dụng trong kinh doanh

Thống kê ứng dụng trong kinh doanh là môn học về thu thập, tổ chức, trình bày, phân tích và diễn giải dữ liệu nhằm hỗ trợ cho việc ra quyết định trong lĩnh vực kinh doanh và kinh tế.

3

Ứng dụng của thống kê

Kỹ thuật thống kê được ứng dụng nhiều trong lĩnh vực marketing, kiểm soát chất lượng, nghiên cứu người tiêu dùng, tài chính, kế toán, quản trị…

4

Thống kê mô tả & thống kê suy diễn

Thống kê mô tả - phương pháp tổ chức, tóm tắt và trình bày dữ liệu nêu bậc được thông tin quan trọng.

Thí dụ : Một nghiên cứu cho thấy có 49% người tiêu dùng biết đến thương hiệu Phở 24 . Số thống kê 49 cho thấy có 49 người trong số 100 người được khảo sát biết đến thương hiệu này.

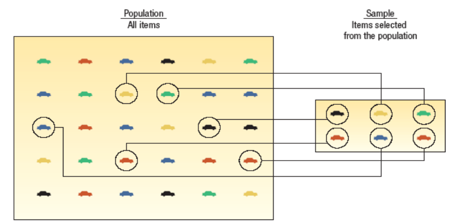
Thống kê suy diễn: phương pháp dựa vào dữ liệu của <u>mẫu</u> để ước lượng, dự báo, ra quyết định về <u>tổng thể</u>.

Thí dụ : Từ tỷ lệ 49% người tiêu dùng trong mẫu biết thương hiệu phở 24, ta ước lượng tỷ lệ người tiêu dùng biết đến thương hiệu này trên tổng thể nghiên cứu.

Tổng thể và mẫu

<u>Tổng thể</u> là tập hợp tất cả các đối tượng, cá nhân hay số đo cần nghiên cứu.

<u>Mẫu</u> là một tập con, một phần của tổng thể đang nghiên cứu.



Các loại biến

A. Biến định tính - đặc trưng không có ý nghĩa số học.

Thí dụ : Giới tính, tôn giáo, hiệu xe máy, nơi sinh.

B. Biến định lượng – đặc trưng có ý nghĩa là con số.

Thí dụ : Số trẻ em trong hộ, thời gian chờ tính tiền tại siêu thị.

Phân loại biến định lượng

Biến định lượng được chia làm 2 loại: Biến rời rạc và biến liên tục.

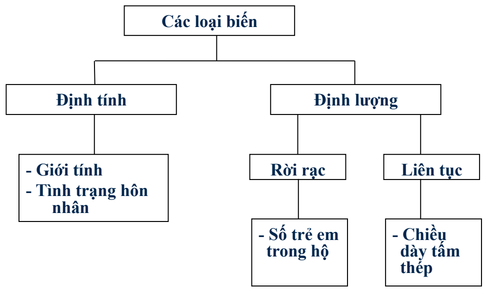
A. Biến rời rạc : biến có giới hạn các giá trị và có các “khoảng trống” giữa các giá trị.

Thí dụ: số phòng ngủ trong một căn hộ, số nhân viên đi trễ trong 1 ca sản xuất.

B. Biến liên tục có giá trị bất kỳ trong một khoảng.

Thí dụ : Áp suất nồi hơi, trọng lượng xe tải, chiều dày tấm thép.

Phân loại biến trong thống kê



9

Bốn loại thang đo

Thang đo danh nghĩa – là thang đo định tính. Nó được dùng để phân loại dữ liệu. Người ta còn gọi nó là thang đo định danh.
Thí dụ: Giới tính.

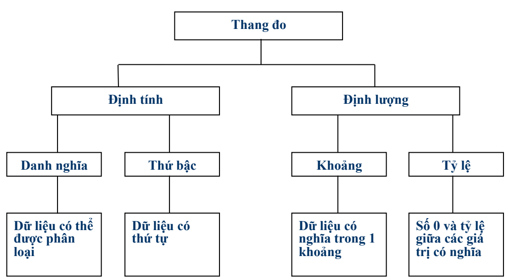
Thang đo thứ tự – là thang đo định tính. Nó được dùng để phân loại và cho biết mức độ hơn kém của các mục dữ liệu.
Thí dụ : Xếp hạng thi đua cuối năm: A, B, C, D.

Thang đo khoảng – là thang đo định lượng. Các giá trị của thang đo có ý nghĩa trong 1 khoảng.
Thí dụ: : Nhiệt độ.

Thang đo tỷ lệ - là thang đo định lượng. Nó là sự mở rộng của thang đo khoảng. trong thang đo tỷ lệ, số 0 có nghĩa và nhờ đó ta xác định được quan hệ tỷ số giữa các giá trị.
Thí dụ: Số trẻ em trong hộ.

10

Tóm tắt các loại thang đo

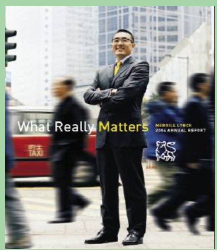


11

Hết chương 1

12

Trình bày dữ liệu :
Bảng tần số, phân phối tần số và biểu đồ tần số



Chương 2

Thống kê ứng dụng trong kinh doanh
Trần Tuấn Anh

Nội dung chính

- Sắp xếp dữ liệu định tính vào bảng tần số.
- Biểu đồ thanh và biểu đồ tròn.
- Sắp xếp dữ liệu định lượng vào bảng tần số.
- Biểu đồ histogram, đa giác tần số và đa giác tần số tích lũy.
- Biểu đồ nhánh và lá.
- Biểu đồ tương quan.

2

Trình bày dữ liệu định tính

Thí dụ 2.1: Tập dữ liệu khách hàng của một cửa hàng kinh doanh

STT	TÊN KHÁCH HÀNG	TUỔI	GIỚI TÍNH	NGHỀ NGHIỆP
1	HỒ THỊ BẠCH KIM	49	NỮ	KINH DOANH
2	VÕ VĂN VIÊN	46	NAM	NHÂN VIÊN
3	VŨ THỊ HOÀNG YẾN	33	NỮ	CNV
4	NGUYỄN VĂN PHI	41	NAM	NHÂN VIÊN
5	NGUYỄN THỊ HỒNG TƯƠI	29	NỮ	NHÂN VIÊN
6	NGUYỄN THỊ OANH	36	NỮ	TỰ DO
7	GIANG THỊ THÀNH	26	NAM	BUÔN BÁN
8	NGUYỄN ĐÌNH TUẤN	43	NAM	CNV
9	NGUYỄN THỊ VÂN	30	NỮ	CNV
10	TRẦN QUAN TRUNG KIÊN	23	NAM	TỰ DO
11	NGUYỄN VAN TRƯỜNG	34	NAM	CNV
12	ĐỖ THÀNH HƯNG	21	NAM	CNV
13	PHẠM THỊ HƯƠNG	38	NỮ	TỰ DO
14	NGUYỄN HOÀNG LONG	46	NAM	BUÔN BÁN
15	PHẠM BÁ QUỐC	27	NAM	NHÂN VIÊN
16	TRẦN VĂN LÝ	54	NAM	NHÂN VIÊN
17	NGUYỄN THUỘC	70	NAM	KINH DOANH
18	PHẠM THỊ HƯƠNG	37	NỮ	CNV
19	PHẠM THỊ MINH THƠ	38	NỮ	CNV
20	TRỊNH THỊ THANH HIỀN	20	NỮ	SINH VIÊN

3

Bảng tần số

Bảng 2.1: Tần số của biến giới tính

Giới tính	Tần số
Nam	11
Nữ	9

Trong bảng tần số, ta có 2 cột: cột thứ nhất là các nhóm tách biệt nhau và cột thứ hai là số quan sát tương ứng với mỗi nhóm.

4

Tần số tương đối

- **Tần số tương đối** là tỷ số giữa tần số của một nhóm và tổng số quan sát.

Bảng 2.3: Tần số tương đối của biến giới tính

Giới tính	Tần số	Tần số tương đối
Nam	11	0,55
Nữ	9	0,45
Cộng	20	

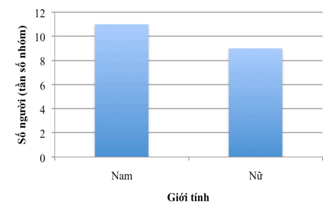
Bảng 2.4: Tần số phần trăm của biến giới tính

Giới tính	Tần số	Tần số phần trăm
Nam	11	55%
Nữ	9	45%
Cộng	20	

5

Biểu đồ thanh

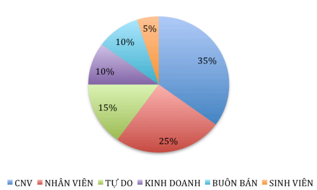
Biểu đồ thanh là biểu đồ mà trong đó, các nhóm được biểu diễn ở trục ngang. Tần số các nhóm được biểu diễn ở trục đứng. Chiều cao của thanh biểu diễn tần số của mỗi nhóm.



6

Biểu đồ tròn

Biểu đồ tròn là biểu đồ mà trong đó, tần số của mỗi nhóm tương ứng với 1 phần diện tích của hình tròn. Người ta thường dùng tần số phần trăm để biểu diễn trên biểu đồ tròn.



7

Trình bày dữ liệu định lượng

Thí dụ 2.2.a: Một lớp học ứng dụng thống kê trong kiểm soát quá trình sản xuất có kết quả kiểm tra cuối khóa của 45 học viên như sau:

4	10	5	7	3
5	6	7	8	5
8	9	3	8	7
6	2	5	1	6
6	7	7	4	10
8	6	4	8	
8	5	9	4	
5	6	6	3	
4	3	6	6	
7	6	6	7	

Yêu cầu: bạn hãy lập bảng tần số

8

Trình bày dữ liệu định lượng

Ta có tập dữ liệu về hệ số P/E của 57 công ty trên sàn giao dịch chứng khoán SG.

Yêu cầu: bạn hãy lập bảng tần số

8	20	15	11	21	18
12	25	17	13	29	23
14	9	20	16	11	11
17	13	25	17	14	14
19	15	11	21	16	16
24	17	13	28	18	19
8	20	16	11	22	24
12	25	17	14	11	16
14	10	20	16	14	18
17	13	27			

9

Các bước lập bảng tần số

Các bước lập bảng tần số

Bước 1: Sắp dữ liệu theo thứ tự tăng dần

Bước 2: Xác định số nhóm

Bước 3: Xác định độ rộng của mỗi nhóm

Bước 4: Đặt dữ liệu vào các nhóm tương ứng

Bước 5: Tính tần số tương đối và các giá trị khác

Công thức 2.1 - Công thức Sturges - xác định số nhóm

$k = 1 + 3{,}3\log(n)$

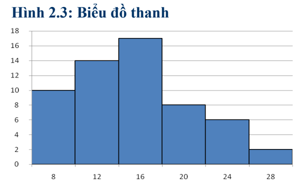
Công thức 2.2 - Xác định độ rộng mỗi nhóm

$$w = \frac{X_{\max} - X_{\min}}{k}$$

10

Biểu đồ thanh (histogram)

Hình 2.3: Biểu đồ thanh



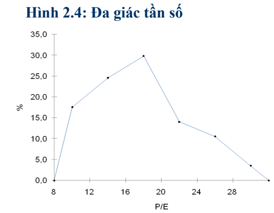
Nhóm	Tần số	Tần số tích lũy
8 – 12	10	10
12 – 16	14	24
16 – 20	17	41
20 – 24	8	49
24 – 28	6	55
28 – 32	2	57
Cộng	57	

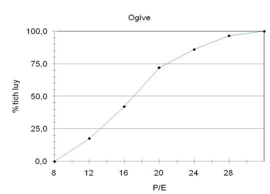
11

Đa giác tần số & biểu đồ Ogive

Nhóm	Tần số	Tần số tương đối	Tần số tương đối tích lũy
8 – 12	10	0,1754	0,1754
12 – 16	14	0,2456	0,4210
16 – 20	17	0,2982	0,7193
20 – 24	8	0,1404	0,8596
24 – 28	6	0,1053	0,9649
28 – 32	2	0,0351	1,0000
Cộng	57		

Hình 2.4: Đa giác tần số





Hình 2.5: Biểu đồ Ogive (tần số phần trăm tích lũy)

12

Biểu đồ nhánh và lá

Các bước tạo biểu đồ nhánh và lá

Bước 1: Khảo sát tập dữ liệu và chọn đơn vị cho nhánh và lá. Thông thường, bạn nên chọn sao cho số nhánh ít hơn 20.

Bước 2: Đặt các giá trị vào nhánh theo thứ tự từ nhỏ đến lớn theo chiều từ trên xuống.

Bước 3: Đặt các giá trị vào phần lá, tức là các hàng tương ứng trong biểu đồ.

Bước 4: Sắp xếp dữ liệu từ nhỏ đến lớn theo chiều từ trái sang phải cho các lá.

13

Biểu đồ nhánh và lá

Thí dụ 2.3a: Đây là số liệu thu thập của 31 ngày về số lượt khách hàng mang máy điện thoại di động đến bảo hành trong 1 ngày tại một trung tâm chăm sóc khách hàng.

37	21	14	33	21	14
33	20	14	32	20	12
29	19	12	9	19	28
6	18	28	18	23	22
18	18	22	22	16	15
21					

0	6 9
1	2 2 4 4 4 5 6 8 8 8 8 9 9
2	0 0 1 1 1 2 2 2 3 8 8 9
3	2 3 3 7

14

Biểu đồ nhánh và lá

Thí dụ : Ta có tập dữ liệu chiều dày tấm thép (mm) xuất xưởng trong 1 ca sản xuất như sau:

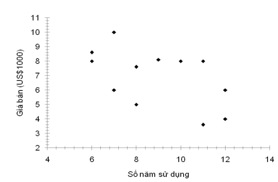
30,8	30,9	32,0	32,3	32,6	31,7	30,4	31,4	32,7	31,4
30,1	32,5	30,8	31,2	31,8	31,6	30,3	32,8	30,6	31,9
32,1	31,3	32,0	31,7	32,8	33,3	32,1	31,5	31,4	31,5
31,3	32,5	32,4	32,2	31,6	31,0	31,8	31,0	31,5	30,6
32,0	30,4	29,8	31,7	32,2	32,4	30,5	31,1	30,6	

Yêu cầu: lập biểu đồ nhánh và lá

15

Biểu đồ phân tán

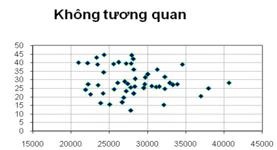
Biểu đồ phân tán là biểu đồ biểu diễn các cặp giá trị (x_1, y_1), (x_2, y_2), …, (x_n, y_n) trên 2 trục X,Y. Mỗi cặp giá trị được biểu diễn bằng 1 điểm trên biểu đồ.

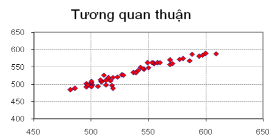


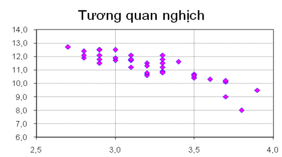
Xe	Số năm sử dụng	Giá bán (US $1000)
1	9	8,1
2	7	6,0
3	11	3,6
4	12	4,0
5	8	5,0
6	7	10,0
7	8	7,6
8	11	8,0
9	10	8,0
10	12	6,0
11	6	8,6
12	6	8,0

16

Biểu đồ phân tán







17

Hết chương 2

18

Cơ bản về xác suất

Chương 4

Thống kê ứng dụng trong kinh doanh
Trần Tuấn Anh

Nội dung chính

- Hiểu được các khái niệm cơ bản của xác suất.
- Phân biệt được các loại xác suất và ý nghĩa của từng loại.
- Áp dụng được các công thức tính xác suất cơ bản.
- Biết cách vận dụng các qui tắc cộng và nhân để tính xác suất trong các trường hợp phức tạp.
- Biết cách dùng cây xác suất để phân tích tình huống và tính xác suất.
- Biết cách dùng các qui tắc đếm trong tính toán xác suất.

2

Định nghĩa xác suất

Phép thử là một quá trình, một tác động dẫn đến một kết quả xảy ra trong số nhiều kết quả có thể xảy ra.

Kết cục là kết quả của một phép thử.

Không gian mẫu là tập hợp tất cả các kết cục có thể có của một phép thử.

Biến cố là tập hợp của một hoặc nhiều kết cục của một phép thử.

Xác suất của một biến cố là khả năng xảy ra của biến cố đó. Xác suất có giá trị trong khoảng [0,1]. Xác suất bằng 0 có nghĩa là biến cố không xảy ra. Xác suất bằng 1 có nghĩa là biến cố chắc chắn xảy ra.

3

Thí dụ minh họa

Phép thử	Tung xúc xắc	Tung 2 đồng xu (sấp/ngửa)
Tất cả các kết cục	mặt 1 chấm mặt 2 chấm mặt 3 chấm mặt 4 chấm mặt 5 chấm mặt 6 chấm	sấp – ngửa ngửa – sấp ngửa – ngửa sấp – sấp
Biến cố	mặt chẵn mặt có số chấm > 4	có ít nhất 1 mặt sấp có 2 mặt giống nhau

4

Tính xác suất

Tính xác suất theo cổ điển:

$$P(A) = \frac{m}{n} \qquad 0 \le P(A) \le 1$$

Tính xác suất theo thực nghiệm

$$P(A) = \frac{m}{n}$$

Xác suất chủ quan là giá trị xác suất được gán cho một biến cố nào đó dựa trên nhận định của chuyên gia từ những thông tin sẵn có.

5

Qui tắc cộng

Qui tắc cộng

$$P(A \cup B) = P(A) + P(B)$$

Thí dụ: Trong 1 cuộc khảo sát, ta có xác suất khách hàng tuổi dưới 18 là 0,15, xác suất khách hàng có tuổi trên 60 là 0,09. Khi đó, xác suất có khách hàng có tuổi dưới 18 hoặc trên 60 được tính như sau:

Qui tắc cộng 2 biến cố đối lập

$$P(A \cup \overline{A}) = P(A) + P(\overline{A}) = 1$$

Thí dụ 4.7: Tại một xưởng đóng gói bột giặt, người ta biết xác suất của 1 bao bột giặt thiếu cân là 0,025. Xác suất của 1 bao bột giặt dư cân là 0,075. Tìm xác suất của bao bột giặt đúng cân.

6

Qui tắc cộng

Qui tắc cộng trong trường hợp các biến cố không xung khắc nhau

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

Thí dụ : Khảo sát 200 khách tham quan công viên Văn hóa Đầm Sen, thấy có 50 khách hàng tham quan khu Thủy cung, 100 khách hàng tham quan khu Không gian, 30 khách tham quan Thủy cung và tham quan khu Không gian. Tính xác suất khách hàng tham quan khu Thủy cung hoặc khu Không gian.

7

Qui tắc nhân

Hai biến cố độc lập với nhau là 2 biến cố xảy ra mà không có sự ảnh hưởng lẫn nhau. Tức là sự xuất hiện của biến cố này không ảnh hưởng gì đến biến cố kia và ngược lại.

Qui tắc nhân 2 biến cố độc lập nhau

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

Thí dụ : Hãng hàng không Việt Nam Airline trong một nghiên cứu biết được 30% khách hàng đặt vé trực tuyến trong năm 2011 đã từng đặt vé trực tuyến trong năm 2010. Một người nghiên cứu chọn ngẫu nhiên 2 khách hàng đặt vé trực tuyến trong năm 2011. Vậy xác suất chọn đúng 2 khách hàng đã đặt vé trực tuyến trong năm 2010 là bao nhiêu ?

8

Qui tắc nhân

Biến cố điều kiện là biến cố xảy ra cần có sự xảy ra của biến cố khác. Biến cố B/A xảy ra chỉ khi biến cố A xảy ra.

Công thức xác suất điều kiện

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

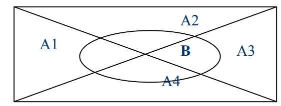
Qui tắc nhân 2 biến cố không độc lập nhau

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

Thí dụ 4.10 : Một quầy hàng trưng bày và bán áo thun có 12 cái áo, trong đó có 9 áo tốt và có 3 áo bị lỗi. 2 khách hàng lần lượt vào mua áo tại quầy. Tính xác suất để cả 2 khách hàng đó đều chọn áo tốt.

9

Công thức xác suất đầy đủ
Công thức Bayes



Công thức xác suất đầy đủ

$$P(B) = \sum_{i=1}^{n} P(A_i B)$$

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B/A_i)$$

TD: 1 cửa hàng bán máy vi tính 3 dòng máy A,B,C với thị phần: 50%; 30% và 20%. Tỷ lệ bảo hành trong 1 năm của 3 dòng máy A, B, C tương ứng là 10%, 20% và 25%. Một khách hàng mua máy bất kỳ tại cửa hàng, tìm xác suất để khách hàng đó mang máy đến bảo hành.

Công thức Bayes

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{P(B)} \qquad P(A_i/B) = \frac{P(A_i)P(B/A_i)}{\sum_{i=1}^{n} P(A_i)P(B/A_i)}$$

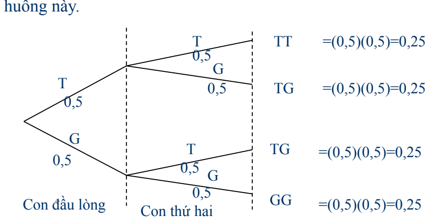
Tìm xác suất để máy mang đến bảo hành là dòng máy A

10

Cây xác suất

Cây xác suất là một sơ đồ liệt kê các xác suất xảy ra của các biến cố theo hệ thống.

Thí dụ : Một cặp vợ chồng mới cưới lên kế hoạch sinh con. Họ dự định có 2 con và băn khoăn không biết sẽ là trai hay gái. Ta có thể dùng sơ đồ cây để biểu diễn tình huống này.



11

Qui tắc đếm

Công thức nhân

Nếu có m cách chọn trong bước 1, có n cách chọn trong bước 2 thì kết hợp lại số cách chọn cho cả 2 bước là

$$\mathbf{m \times n}$$

Nếu có n_1 cách chọn trong bước 1, có n_2 cách chọn trong bước 2…có n_k cách chọn trong bước k thì số cách chọn trong k bước sẽ là:

$$\mathbf{n_1 \times n_2 \times \ldots \times n_k}$$

Thí dụ : Giả sử trong một công ty có 4 kho hàng được đặt tại các tỉnh Long an, Tiền Giang, Hậu Giang, Kiên Giang. Từ TPHCM đến kho Long An có 3 lộ trình. Từ kho Long An đến kho Tiền Giang có 4 lộ trình. Từ kho Tiền Giang đến kho Hậu Giang có 2 lộ trình. Từ kho Hậu Giang đến kho Kiên Giang có 4 lộ trình. Như vậy, từ TPHCM đi qua các kho Long An, Tiền Giang, Hậu Giang, Kiên Giang có số lộ trình là ?

12

Chỉnh hợp – hoán vị

Chỉnh hợp là một tập k phần tử có thứ thự được chọn ra từ n phần tử cho trước.

Số chỉnh hợp

$$A_n^k = \frac{n!}{(n-k)!}$$

Thí dụ : Trong 1 xưởng may, người ta có 8 máy may nhưng chỉ có 3 vị trí để đặt máy may. Vậy có bao nhiêu cách khác nhau để sắp đặt 8 máy may này vào 3 vị trí đó.

Số hoán vị

$$A_n^n = n!$$

Thí dụ : Trên kệ trưng bày có 6 chiếc máy tính xách tay. Có bao nhiêu cách trưng bày dựa trên sự thay đổi chỗ của 6 máy đó trên kệ.

13

Tổ hợp

Tổ hợp là một tập k phần tử không có thứ tự được chọn ra từ n phần tử cho trước.

Số tổ hợp

$$C_n^k = \frac{n!}{k!(n-k)!}$$

Thí dụ : Một chuỗi cửa hàng tiện lợi có 42 cửa hàng. Phòng kinh doanh của chuỗi cửa hàng muốn dùng 3 mã màu để đánh dấu các thùng đĩa CD chuyển xuống các cửa hàng. Yêu cầu ở đây là nếu 3 màu đã dùng cho cửa hàng này thì không thể dùng cho cửa hàng khác. Thí dụ màu xanh – tím – đỏ đã dùng cho cửa hàng thứ i rồi thì bộ ba màu đó dù có thứ tự khác cũng không được dùng cho các cửa hàng khác. Câu hỏi đặt ra là nếu có tổng cộng 7 màu thì có đủ dùng để phân biệt các thùng CD cho 42 cửa hàng không ?

14

Hết chương 4

15

Phân phối xác suất rời rạc

Chương 5

Thống kê ứng dụng trong kinh doanh
Trần Tuấn Anh

Nội dung chính

- Hiểu được định nghĩa biến ngẫu nhiên rời rạc và phân phối xác suất
- Hiểu các khái niệm giá trị kỳ vọng và phương sai của phân phối xác suất và biết cách sử dụng chúng.
- Nắm được các mô hình phân phối xác suất rời rạc, phân phối nhị thức và phân phối Poisson.
- Nhận diện mô hình phân phối xác suất phù hợp cho vấn đề cần giải quyết.

2

Biến ngẫu nhiên

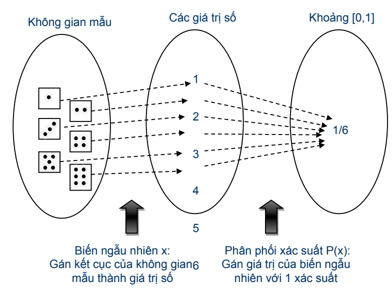
Biến ngẫu nhiên là một hàm hay một qui luật gán một giá trị số cho mỗi kết cục trong không gian mẫu của một thử nghiệm ngẫu nhiên.

Biến ngẫu nhiên "số nhân viên đi trễ" nhận các giá trị 0, 1, 2,…

Biến ngẫu nhiên rời rạc là biến ngẫu nhiên mà các giá trị của nó đếm được, tách rời nhau.

3

Phân phối xác suất của biến ngẫu nhiên rời rạc



Đặc điểm của phân phối rời rạc:

Nếu có n giá trị rời rạc của X ($x_1, x_2, \ldots, x_n$), ta có:

$$0 \le P(x_i) \le 1$$

$$\sum_{i=1}^{n} P(x_i) = 1$$

4

Thí dụ

Thí dụ Ta tung đồng xu 3 lần, phân phối xác suất như sau: x là số mặt sấp

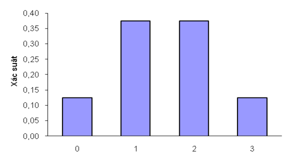
Bảng phân phối xác suất:

Biến cố	x	P(x)
NNN	0	1/8
SNN,NSN,NNS	1	3/8
SSN,SNS,NSS	2	3/8
SSS	3	1/8
Cộng		1

Tổng quát, ta có:

x	P(x)
x_1	$P(x_1)$
x_2	$P(x_2)$
…	…
x_n	$P(x_n)$
Cộng	1

Biểu đồ phân phối xác suất:



5

Giá trị kỳ vọng của PPXS rời rạc

Giá trị kỳ vọng của biến ngẫu nhiên rời rạc

$$E(X) = \mu = \sum_{i=1}^{n} x_i P(x_i)$$

Thí dụ : Một trạm dịch vụ bảo dưỡng xe máy tận nhà nhận cuộc gọi dịch vụ bảo dưỡng xe máy tận nhà qua điện thoại. x là số cuộc gọi nhận trong 1 ca trực. Ta có bảng phân phối xác suất của x như sau :

x	P(x)	xP(x)
0	0,05	0,00
1	0,10	0,10
2	0,30	0,60
3	0,25	0,75
4	0,20	0,80
5	0,10	0,50
Cộng	1	2,75

Tính E(X)

6

Phương sai & độ lệch chuẩn PPXS rời rạc

Phương sai của biến ngẫu nhiên rời rạc

$$V(X) = \sigma^2 = \sum (x_i - \mu)^2 P(x_i)$$

Độ lệch chuẩn của biến ngẫu nhiên rời rạc

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

Thí dụ: tính phương sai và độ lệch chuẩn của thí dụ trước.

7

Phân phối nhị thức

Phân phối nhị thức là phân phối của các biến có các phép thử ngẫu nhiên chỉ có 2 kết cục: thành công – không thành công.

Thí dụ: tung đồng xu có 2 kết cục sấp – ngửa, kiểm tra chất lượng sản phẩm có 2 kết cục đạt – không đạt, kết quả kỳ sát hạch lấy bằng lái xe ôtô C_1 là đạt – không đạt…2 kết cục này phải xung khắc hoàn toàn

8

Phân phối nhị thức – đặc điểm

- Các phép thử chỉ có 2 kết cục là thành công – không thành công, và 2 kết cục này phải xung khắc hoàn toàn.
- Giá trị của biến là kết quả việc đếm số thành công của mỗi phép thử.
- Xác suất thành công trong mọi phép thử là như nhau
- Các phép thử phải độc lập với nhau. Tức là kết quả của phép thử này không ảnh hưởng đến phép thử kia và ngược lại.

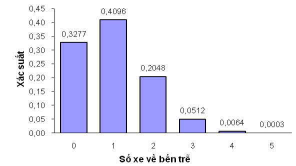
9

Phân phối nhị thức

Tính phân phối nhị thức

$$P(x) = C_n^x p^x (1-p)^{n-x}$$

Thí dụ: Tại bến xe miền đông, mỗi ngày có 5 chuyến xe từ Đắk Lắk về bến. Giả sử xác suất xe về bến trễ mỗi ngày là 0,2. Vậy xác suất để không có chuyến xe nào về bến trễ trong ngày là bao nhiêu?



10

Phân phối nhị thức – trung bình, phương sai & độ lệch chuẩn

Giá trị trung bình của phân phối nhị thức

$$\mu = np$$

Phương sai của phân phối nhị thức

$$\sigma^2 = np(1-p)$$

Độ lệch chuẩn của phân phối nhị thức

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

Tính trung bình, phương sai và độ lệch chuẩn của thí dụ trên

Tra bảng phân phối nhị thức?

11

Phân phối Poisson

Phân phối Poisson

Phân phối Poisson là phân phối mô tả số lần của biến cố xảy ra trong một khoảng nào đó. Khoảng ở đây có nghĩa là khoảng thời gian, khoảng cách, diện tích hoặc thể tích.

Số lỗi của việc nhập dữ liệu, số hỏng hóc của thiết bị trong sản xuất, số sản phẩm khuyết tật phát sinh trong thời gian bảo quản hàng hóa, số khách hàng chờ được phục vụ trong một tiệm rửa xe, số tai nạn giao thông trong khoảng thời gian nghiên cứu như ngày, tuần,

12

Phân phối Poisson – đặc điểm

- Biến ngẫu nhiên là số lần xảy ra của biến cố trong một khoảng (thời gian) xác định.
- Xác suất của biến cố tỷ lệ với độ lớn của khoảng (thời gian).
- Các khoảng (thời gian) không chồng lên nhau và hoàn toàn độc lập nhau.

13

Phân phối Poisson

Hàm xác suất của phân phối Poisson

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Thí dụ : Người ta nghiên cứu tình trạng thất lạc hành lý trong các chuyến bay. Khảo sát 1000 chuyến bay, người ta thấy có tổng cộng 300 hành lý bị thất lạc. Ta dùng công thức phân phối Poisson để tính xác suất chuyến bay không có hành lý bị thất lạc và xác suất chuyến bay có một hành lý bị thất lạc.

14

Phân phối Poisson – trung bình và phương sai

Giá trị trung bình của phân phối Poisson

$$\mu = \lambda$$

Phương sai của phân phối Poisson

$$\sigma^2 = \lambda$$

Tính trung bình, phương sai và độ lệch chuẩn của thí dụ trên

Tra bảng phân phối Poisson?

15

Hết chương 5

16

Phân phối xác suất liên tục

Chương 6

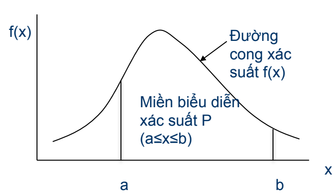
Thống kê ứng dụng trong kinh doanh
Trần Tuấn Anh

Nội dung chính

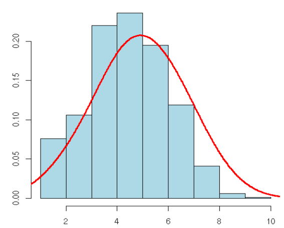
- Phân biệt sự khác biệt giữa biến ngẫu nhiên liên tục và biến ngẫu nhiên rời rạc.
- Nắm và sử dụng được các tính toán cơ bản trên phân phối đều, phân phối chuẩn và phân phối chuẩn chuẩn tắc.
- Biết cách chọn phân phối phù hợp và ứng dụng để tính toán trong từng trường hợp.
- Biết cách dùng phân phối chuẩn để xấp xỉ các phân phối nhị thức và phân phối Poisson.

2

Phân phối xác suất liên tục



Xác suất của một giá trị nằm trong khoảng [a,b] là diện tích miền mặt phẳng nằm dưới đường cong xác suất f(x)



3

PPXS liên tục – đặc điểm

Đặc điểm của phân phối xác suất liên tục

Phân phối xác suất liên tục hay đường cong xác suất có 2 đặc điểm sau:

1) $f(x) \geq 0 \quad \forall x$

2) $\int_{-\infty}^{+\infty} f(x)dx = 1$

Giá trị kỳ vọng của biến ngẫu nhiên liên tục

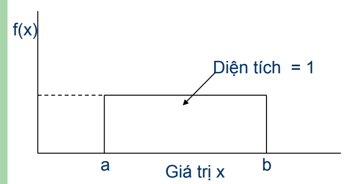
$$E(X) = \mu = \int_{-\infty}^{+\infty} x f(x)dx$$

Phương sai của biến ngẫu nhiên liên tục

$$V(X) = \sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx$$

4

Phân phối đều



Công thức tính giá trị trung bình và độ lệch chuẩn của phân phối đều

$$\mu = \frac{a+b}{2} \qquad \sigma = \frac{b-a}{\sqrt{12}}$$

Hàm mật độ xác suất của phân phối đều

$f(x) = \frac{1}{b-a}$ khi $a \le x \le b$

$f(x) = 0$ trong các trường hợp khác.

5

Thí dụ

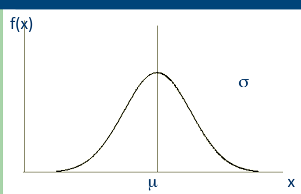
Một quản lý của một trung tâm thương mại đang phân tích số liệu thời gian chờ của khách hàng sử dụng thang máy trong trung tâm thương mại. Số liệu điều tra 100 trường hợp khách hàng chờ được lập thành biểu đồ tần số. Biểu đồ cho thấy khách hàng chờ trong khoảng từ 0 đến 4 phút và tần số của thời gian chờ là gần như nhau.

Tính giá trị trung bình và phương sai của thời gian chờ.

Tìm xác suất 1 khách hàng chờ tối thiểu 2,5 phút.

6

Phân phối chuẩn



Phân phối chuẩn có dạng **hình quả chuông, đối xứng** quanh giá trị trung bình

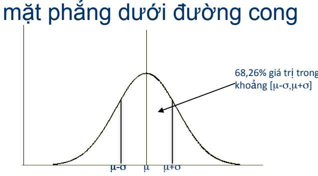
Hàm mật độ xác suất của phân phối chuẩn

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

7

Đặc điểm của phân phối chuẩn

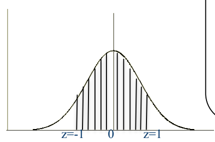
- Mỗi phân phối trong họ phân phối chuẩn được xác định bởi 2 giá trị cơ bản là giá trị trung bình μ và độ lệch chuẩn σ.
- Các giá trị trung bình, trung vị và mode trùng nhau và là trục đối xứng của đường cong chuẩn.
- Hai đuôi của đường cong chuẩn tiệm cận với trục ngang và tổng diện tích của miền mặt phẳng dưới đường cong là 1.



8

Phân phối chuẩn chuẩn tắc

- Phân phối chuẩn chuẩn tắc là một trường hợp đặc biệt của phân phối chuẩn khi $\mu = 0$ và $\sigma = 1$.
- $P(-1 \le z \le 0) = P(0 \le z \le 1) = 0,3413$
- Cách tra bảng.



Biến đổi biến từ phân phối chuẩn sang phân phối chuẩn chuẩn tắc

$$z = \frac{x - \mu}{\sigma}$$

9

Xấp xỉ phân phối nhị thức

Điều kiện để tính xấp xỉ phân phối nhị thức

$np \ge 10$ và

$n(1-p) \ge 10$

Đặc trưng của phân phối chuẩn xấp xỉ phân phối nhị thức

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

Thí dụ : Một bài thi trắc nghiệm có 32 câu theo kiểu đúng – sai. Nếu một thí sinh chọn ngẫu nhiên thì khả năng đáp đúng là 50%. Hãy tìm xác suất để có một bài thi có nhiều hơn 17câu có trả lời đúng đáp án nhờ chọn ngẫu nhiên.

10

Xấp xỉ phân phối Poisson

Điều kiện để tính xấp xỉ phân phối Poisson

$\lambda \ge 10$

Đặc trưng của phân phối chuẩn xấp xỉ phân phối Poisson

$\mu = \lambda$ $\qquad \sigma = \sqrt{\lambda}$

Thí dụ : Tại một trung tâm cấp cứu vào sáng thứ 7 từ 10 giờ đến 12 giờ, người ta nhận 42 ca cấp cứu/giờ. Ta cần tìm xác suất có nhiều hơn 50 ca cấp cứu/giờ.

11

Hết chương 6

12

Phương pháp chọn mẫu và phân phối mẫu

Chương 7

Thống kê ứng dụng trong kinh doanh
Trần Tuấn Anh

Nội dung chính

•Biết được lý do vì sao người ta dùng phương pháp chọn mẫu để nghiên cứu tổng thể.
•Nắm được các phương pháp chọn mẫu trong nghiên cứu thống kê.
•Biết được định nghĩa và cách lập phân phối mẫu của trung bình mẫu.
•Hiểu và giải thích được định lý giới hạn trung tâm.
•Sử dụng định lý giới hạn trung tâm để tìm xác xuất của một trung bình mẫu rút ra từ một tổng thể nghiên cứu.

2

Mẫu xác suất

- Một mẫu được chọn theo kiểu xác suất được gọi là mẫu xác suất. Trong cách chọn mẫu này, ta biết được khả năng các phần tử trong tổng thể nghiên cứu được chọn vào mẫu.

3

Lý do chọn mẫu

- Thời gian
- Chi phí
- Tính khả thi về mặt kỹ thuật
- Tính đặc thù của kiểm tra phá hũy
- Tính thỏa đáng của việc chọn mẫu

4

Phương pháp chọn mẫu ngẫu nhiên đơn giản

- Trong phương pháp này, khả năng các phần tử trong tổng thể được chọn vào mẫu là như nhau.

Thí dụ : Có 845 khách hàng tham gia vào chương trình khuyến mãi. Để chọn ngẫu nhiên 10 khách hàng trúng giải nhất, ta thường dùng phương pháp bốc thăm.

79610	45326	**969**02	82055	66636	62782	5058
99365	27467	**786**52	98849	17982	71963	67920
03789	82229	**514**22	26734	58672	90563	90331
14688	18585	**020**37	5362	2048	70781	37452
64752	96144	**893**85	72642	3007	62966	73396
80251	85642	**929**24	89544	8034	85349	14475
19931	71434	**373**19	10591	22222	07084	31602
13148	13656	**843**03	96536	60892	34501	73676
94682	55834	**390**48	62891	87226	48898	20534
84109	19689	**052**89	86097	93142	70626	74494
55071	83518	63110	24211	31632	10092	27528
97573	18562	62767	55351	94973	34148	01921
29383	93582	87087	78521	70990	71727	14890
44350	98928	79619	55140	66102	91205	60349
72354	53685	40746	63081	91327	58797	95749

Phương pháp chọn mẫu hệ thống

Chọn mẫu hệ thống

Trước tiên, ta tính hệ số k theo công thức: $k = \frac{N}{n}$

Trong đó: N là qui mô tổng thể và n: qui mô mẫu.

Sau đó, ta chọn ngẫu nhiên 1 số từ 1 đến k thí dụ là s. các phần tử được chọn vào mẫu sẽ có các thứ tự : s, s+k, s+2k, s+3k…

Thí dụ : Phòng bán hàng của một công ty có 1000 hóa đơn bán hàng trong tháng vừa qua. Trưởng phòng bán hàng muốn chọn ngẫu nhiên 100 hóa đơn trong số 2000 hóa đơn này.

Phương pháp chọn mẫu phân tầng

Tổng thể được chia làm nhiều nhóm nhỏ được gọi là tầng. các phần tử trong mẫu sẽ được chọn ngẫu nhiên từ các tầng này.

Gọi N là qui mô tổng thể. Giả sử ta có L tầng và mỗi tầng có số phần tử là N_1, N_2, N_3…,N_L
Ta có : $N = N_1 + N_2 + N_3 + \ldots + N_L$
Trọng số của mỗi tầng là $w_j = N_j/N$

Thí dụ: Bạn muốn chọn một mẫu gồm 200 công nhân trong khu công nghiệp để phỏng vấn. Trong khu công nghiệp có 10000 công nhân, trong đó có 5500 nam và 4500 nữ. Bạn sẽ chọn như thế nào?

Phương pháp chọn mẫu cụm

Tổng thể được chia làm nhiều cụm, trong đó mỗi cụm là một vùng địa lý tự nhiên hay được phân chia theo ranh giới hành chính. Sau đó, các cụm này được chọn ngẫu nhiên và mẫu sẽ được chọn ngẫu nhiên trong các cụm này.

Thí dụ: Bạn cần chọn mẫu 300 người tiêu dùng trong quận 5 TPHCM. Bạn sẽ chọn như thế nào?

Sai số chọn mẫu

Sai số chọn mẫu

Sai số chọn mẫu là sự khác biệt giữa giá trị thống kê mẫu và tham số tổng thể tương ứng.

Thí dụ: Một trung tâm cho thuê xe có số liệu 30 ngày hoạt động như sau :

0	2	3	2	3	4	2	3	4	7
3	4	4	4	7	0	5	3	6	2
3	2	3	6	0	4	1	1	3	3

Bạn thử chọn mẫu và tính sai số chọn mẫu.

9

Phân phối mẫu của trung bình mẫu

Phân phối mẫu của trung bình mẫu

Là phân phối xác suất của tất cả các trung bình mẫu có thể có với cùng một cỡ mẫu cho trước.

Thí dụ : Một đội thi công sửa chữa nhà gồm 7 người (ở đây là tổng thể). Tiền công theo ngày của mỗi thợ được cho như sau :

Thợ	Tiền công theo ngày (10.000 đ)
Bình	7
Minh	7
Kim	8
Mộc	8
Thủy	7
Hỏa	8
Thổ	9

Hãy lập phân phối trung bình mẫu của tổng thể này.

10

Thí dụ

Mẫu	Thợ	Tiền công trung bình
1		
2		
3		
4		
5		
6		
7		
8		
9		
…		
21		

Ta có phân phối trung bình mẫu như sau :

Trung bình mẫu	Số trung bình	Xác suất
7	3	0,1449
7,5	9	0,4285
8	6	0,2857
8,5	3	0,1429
	21	1,0000

11

Định lý giới hạn trung tâm

Định lý giới hạn trung tâm

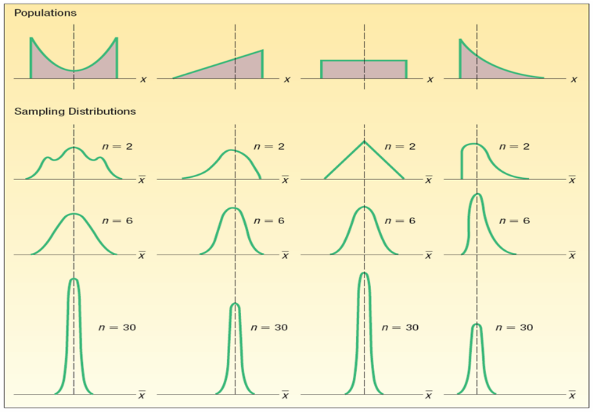
Nếu ta tập hợp tất cả các mẫu ứng với một qui mô mẫu được chọn từ một tổng thể nghiên cứu thì phân phối mẫu của trung bình mẫu sẽ có khuynh hướng có dạng phân phối chuẩn. Khi ta tăng qui mô mẫu lên thì phân phối mẫu của trung bình mẫu càng gần với phân phối chuẩn hơn.

Ta có: $\mu_{\bar{x}} = \mu$, tức là: giá trị trung bình của phân phối mẫu trung bình mẫu chính bằng giá trị trung bình của phân phối tổng thể. Và độ lệch chuẩn của phân phối mẫu này là:

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

12

Định lý giới hạn trung tâm



13

Thí dụ

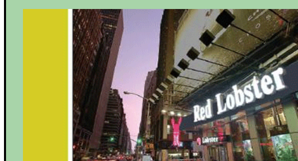
- Thí dụ: Trong một phân xưởng đóng chai của nhà máy hóa chất An Hòa, người ta duy trì lượng hóa chất trong chai có trọng lượng 31,2g và độ lệch chuẩn là 0,4g. Lượng hóa chất trong chai tại phân xưởng này là biến ngẫu nhiên có dạng phân phối chuẩn. Lượng hóa chất này trong chai quá cao hay quá thấp so với trong lượng trung bình đều được coi là không đạt yêu cầu kỹ thuật cho việc đóng chai. Trong ca sản xuất sáng nay, bộ phận KCS (kiểm tra chất lượng sản phẩm) lấy mẫu 16 chai để kiểm tra và tính được trọng lượng trung bình của mẫu này là 31,38g. Rõ ràng ở đây có sự sai biệt giữa trung bình của mẫu so với yêu cầu là 31,2g. Liệu sự sai biệt này có được chấp nhận hay không ? Liệu đây có phải là sự khác biệt bất thường ?

14

Hết chương 7

15

Ước lượng



Chương 8

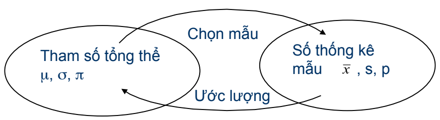
Thống kê ứng dụng trong kinh doanh
Trần Tuấn Anh

Nội dung chính

- Nắm được kiến thức về sử dụng số liệu của mẫu để ước lượng các giá trị tham số tổng thể.
- Biết cách tính ước lượng điểm.
- Biết cách tính ước lượng khoảng của trung bình tổng thể trong các trường hợp biết hoặc chưa biết độ lệch chuẩn của tổng thể.
- Tính khoảng tin cậy của tỷ lệ tổng thể dựa vào tỷ lệ mẫu.
- Biết cách tính hệ số điều chỉnh tổng thể hữu hạn và trường hợp sử dụng nó trong các phép tính ước lượng.
- Biết cách tính cỡ mẫu cho các nghiên cứu.

2

Ước lượng điểm & ước lượng khoảng



Ước lượng điểm là dùng giá trị thống kê của mẫu để ước lượng tham số tương ứng của tổng thể.

Ước lượng khoảng là khoảng giá trị tính được từ mẫu sao cho tham số tổng thể tương ứng có khả năng nằm trong khoảng này. Xác suất xảy ra khoảng ước lượng chứa tham số của tổng thể được gọi là độ tin cậy của ước lượng.

3

Khoảng tin cậy của giá trị trung bình

Khoảng tin cậy của trung bình tổng thể (biết trước σ)

$$\bar{x} \pm z\frac{\sigma}{\sqrt{n}}$$

Độ tin cậy	Z
90%	1,64
95%	1,96
99%	2,58

Thí dụ : Hiệp hội bán lẻ của một thành phố muốn điều tra thu nhập trung bình hàng năm của các quản lý cửa hàng bán lẻ trong các doanh nghiệp thuộc hiệp hội. Một mẫu ngẫu nhiên gồm 256 quản lý cửa hàng được chọn và thu nhập trung bình hàng năm là 75,42 triệu đồng với độ lệch chuẩn của tổng thể là 2,05. Bằng phương pháp ước lượng, hãy xác định khoảng tin cậy 95% của thu nhập trung bình của các quản lý cửa hàng trong các doanh nghiệp thuộc hiệp hội.

4

Khoảng tin cậy của giá trị trung bình

Khoảng tin cậy của trung bình tổng thể (chưa biết σ)

$$\bar{x} \pm t\frac{\sigma}{\sqrt{n}}$$

Thí dụ : Một nhà sản xuất muốn kiểm tra độ mòn của thiết bị nghiền đá sau thời gian sử dụng. Một mẫu gồm 10 trục nghiền được khảo sát. Sau thời gian sử dụng, nhà sản xuất đo được độ mòn là 0,32 cm và độ lệch chuẩn là 0,09cm. Hãy xác định khoảng tin cậy 95% độ mòn của thiết bị nghiền đá sau thời gian sử dụng.

Bậc tự do df = n – 1

5

Khoảng tin cậy cho tỷ lệ

Tỷ lệ là tỷ số hoặc phần trăm xác định phần của mẫu hay tổng thể có đặc trưng cần quan tâm.

Tỷ lệ mẫu

$$p = \frac{X}{n}$$

Khoảng tin cậy của tỷ lệ tổng thể

$$p \pm z\sqrt{\frac{p(1-p)}{n}}$$

Thí dụ : Một doanh nghiệp muốn nghiên cứu sự hài lòng của người tiêu dùng về chính sách hậu mãi của doanh nghiệp. Một khảo sát trên diện rộng toàn quốc được thực hiện với qui mô mẫu là 2000 khách hàng. Trong đó, có 1600 khách hàng đánh giá hài lòng. Hãy dùng dữ liệu trên để ước lượng tỷ lệ hài lòng của khách hàng của doanh nghiệp với độ tin cậy 95%.

6

Hệ số hiệu chỉnh cho tổng thể hữu hạn

Hệ số điều chỉnh cho tổng thể hữu hạn

$$FPC = \sqrt{\frac{N-n}{N-1}}$$

Qui mô mẫu	Tỷ lệ trên tổng thể	Hệ số điều chỉnh
10	0,01	0,9955
25	0,025	0,9879
50	0,05	0,9752
100	0,1	0,9492
200	0,2	0,8949
500	0,5	0,7075

7

Xác định cỡ mẫu

- Để xác định cỡ mẫu, ta cần 3 yếu tố sau:
  - Độ tin cậy cần có của nghiên cứu.
  - Sai số trong nghiên cứu.
  - Sự phân tán của tổng thể nghiên cứu.

Cỡ mẫu để ước lượng trung bình tổng thể

$$n = \left(\frac{z\sigma}{E}\right)^2$$

Cỡ mẫu để ước lượng tỷ lệ tổng thể

$$n = 0{,}25\left(\frac{z}{E}\right)^2$$

8

Thí dụ

Thí dụ : Một sinh viên ngành quản trị nhân lực muốn nghiên cứu tiền thưởng trung bình của các nhà quản lý làm trong các doanh nghiệp có vốn đầu tư nước ngoài tại các khu công nghiệp trong tỉnh A. Sai số trong phép ước lượng kỳ vọng ít hơn $100 với độ tin cậy 95%. Sinh viên này tìm thấy tài liệu trong một nghiên cứu tương tự cho thấy độ lệch chuẩn của tổng thể nghiên cứu này là $1000. Theo bạn, sinh viên này nên chọn mẫu bao nhiêu ?

9

Thí dụ

- Thí dụ : Để nghiên cứu tỷ lệ sinh viên sử dụng dịch vụ thư viện điện tử tại một trường đại học, người ta xác định sai số nghiên cứu không quá 0,1 và độ tin cậy của nghiên cứu là 90%. Trong trường hợp này, cần chọn cỡ mẫu cho nghiên cứu là bao nhiêu ?

10

Hết chương 8

11

Kiểm định giả thuyết một mẫu

Chương 9

Thống kê ứng dụng trong kinh doanh
Trần Tuấn Anh

Nội dung chính

•Nắm được định nghĩa về giả thuyết và kiểm định giả thuyết.
•Hiểu được mức ý nghĩa của kiểm định, các loại sai lầm trong kiểm định giả thuyết: sai lầm loại I và sai lầm loại II.
•Phân biệt được kiểm định một đuôi và kiểm định hai đuôi.
•Biết cách thực hiện kiểm định giả thuyết về trung bình tổng thể.
•Biết cách thực hiện kiểm định giả thuyết về tỷ lệ tổng thể.
•Hiểu và dùng được giá trị p trong kiểm định giả thuyết.

2

Căn bản về kiểm định giả thuyết

Giả thuyết là một phát biểu về một tham số của tổng thể nhằm kiểm định xem nó có bị bác bỏ hay không.

Kiểm định giả thuyết là thủ tục dựa trên các chứng cứ từ mẫu để đánh giá tính thuyết phục của giả thuyết. Kiểm định giả thuyết sẽ tìm chứng cứ xem giả thuyết đã phát biểu có thể bị bác bỏ hay không.

3

Các bước kiểm định giả thuyết

Thủ tục kiểm định giả thuyết

Bước 1: Phát biểu giả khuyết không H_0 và giả thuyết đối H_1.

Bước 2: Chọn mức ý nghĩa của kiểm định α.

Bước 3: Tính giá trị thống kê kiểm định.

Bước 4: Áp dụng qui tắc ra quyết định.

Bước 5: Ra quyết định về giả thuyết không dựa trên kết quả tính toán. Diễn giải kết quả kiểm định.

4

Giả thuyết H_0 và H_1

- H_0: giả thuyết không, H_1: giả thuyết đối.
- Hai giả thuyết H_0 và H_1 có tính đối lập nhau.
- H0 luôn được giả định là đúng.
- H1 là điều cần chứng minh.
- Dùng mẫu (n) để "bác bỏ" H_0.
- Khi ta kết luận "không bác bỏ H_0" thì không có nghĩa H_0 đúng. Nó chỉ đồng nghĩa không có đủ chứng cứ để bác bỏ H_0. Khi H_0 bị bác bỏ sẽ dẫn đến kết luận có khả năng H_1 đúng.
- Các quan hệ =, ≤, ≥ luôn xuất hiện trong H_0
- Các quan hệ ≠, <, > luôn xuất hiện trong H_0

5

Thiết lập các giả thuyết

- Trong thực tế, tình trạng ban đầu của vấn đề nghiên cứu được lập thành giả thuyết H0.
- Lời tuyên bố hay phát biểu được lập thành giả thuyết H_1. nên nhớ, giả thuyết H_1 là cái cần chứng minh.
- Các nội dung như có tuổi thọ lớn hơn, trọng lượng tối thiểu… sẽ được dùng để thiết lập dấu trong các giả thuyết.

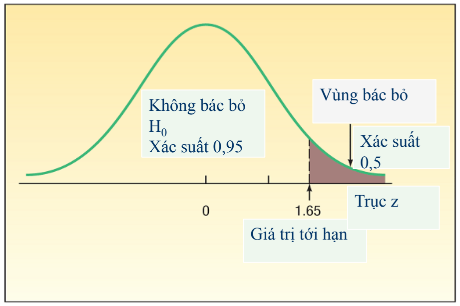
6

Sai lầm loại I và sai lầm loại II

Giả thuyết không	Không bác bỏ H_0	Bác bỏ H_0
H_0 đúng	Quyết định đúng	Sai lầm loại I
H_0 sai	Sai lầm loại II	Quyết định đúng

7

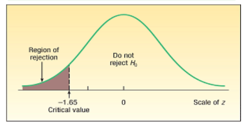
Các phần của phân phối trong kiểm định giả thuyết



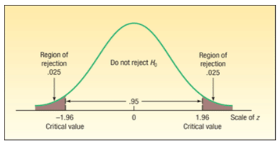
8

Kiểm định 1 đuôi, 2 đuôi

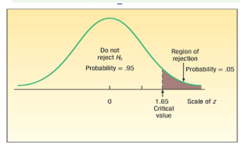
Kiểm định đuôi trái



Kiểm định 2 đuôi



Kiểm định đuôi phải



9

Thí dụ 1

- Thí dụ : Thanh Bình là xí nghiệp sản xuất các loại sản phẩm đồ gỗ. Trong đó, bộ sản phẩm mã số A325 là một trong những bộ sản phẩm chủ lực của xí nghiệp. Sản lượng hàng ngày của sản phẩm này là một phân phối chuẩn có giá trị trung bình là 200 bộ/ngày và độ lệch chuẩn là 16. Hiện nay, do có một số cải tiến trong xí nghiệp và sự thay đổi nhân sự giữa các xưởng trong xí nghiệp nên tính ổn định về năng lực sản xuất A325 bị ảnh hưởng. Giám đốc xí nghiệp muốn nghiên cứu xem liệu có sự thay đổi nào về sản lượng hàng ngày đối với sản phẩm A325 hay không. Chọn mẫu 50 ngày theo dõi tình hình sản xuất thấy sản lượng trung bình đạt 203,5 bộ. Giám đốc có thể kết luận là có sự thay đổi về sản lượng sản xuất sản phẩm A325 trong xí nghiệp hay không với mức ý nghĩa 0,01.

10

Thí dụ 2

- Thí dụ : Giám đốc kỹ thuật xưởng sản xuất vỏ xe ôtô mã số B825 tuyên bố tuổi thọ của vỏ xe này là 60000km. Để kiểm chứng lời tuyên bố của giám đốc này, người ta lấy 30 vỏ xe được xưởng này sản xuất để thử nghiệm tuổi thọ. Theo bạn, ta nên dùng loại kiểm định nào cho trường hợp này : 2 đuôi, đuôi trái, đuôi phải ?

11

Kiểm định trung bình tổng thể - σ đã biết

Bước 1 : Phát biểu giả thuyết H_0 và H_1

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$ (Trường hợp kiểm định 2 đuôi)

Bước 2 : Chọn mức ý nghĩa của kiểm định.

Bước 3 : Tính giá trị thống kê kiểm định theo công thức sau :

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

Bước 4 : áp dụng qui tắc ra quyết định

Với kiểm định 2 đuôi, ta có 2 giá trị tới hạn là và -

Giả thuyết H_0 bị bác bỏ nếu z ≥ $z_{\alpha/2}$ hoặc z ≤ - $z_{\alpha/2}$

Giả thuyến H_0 không thể bị bác bỏ nếu - $z_{\alpha/2} \le z <$ $z_{\alpha/2}$

12

z_α

Thí dụ

- Thực hiện kiểm định giả thuyết – thí dụ 1
- Lưu ý:

α	z_α	$z_{\alpha/2}$
0,1	1,29	1,65
0,05	1,65	1,96
0,01	2,33	2,58

13

Kiểm định trung bình tổng thể - σ chưa biết

- Dùng độ lệch chuẩn của mẫu thay cho độ lệch chuẩn của tổng thể trong công thức kiểm định và dùng phân phối t thay cho phân phối chuẩn.
- Giá trị thống kê kiểm định: $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$

14

Thí dụ 3

- Thí dụ : Công ty dịch vụ bảo hiểm Thanh Bình cho biết chi phí trung bình thực hiện dịch vụ kiểm định là $60. Nếu so với chi phí dịch vụ kiểm định của các doanh nghiệp khác trong ngành thì đây là mức khá cao. Giám đốc công ty Thanh Bình muốn thực hiện chương trình cắt giảm chi phí này. Sau thời gian triển khai chương trình cắt giảm chi phí, giám đốc công ty muốn đánh giá hiệu quả chương trình. Một mẫu 26 trường hợp kiểm định được thu thập trong tháng vừa rồi có kết quả sau (đơn vị $):

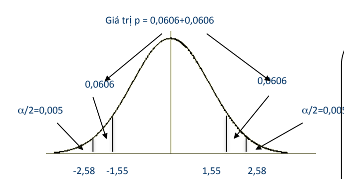
45	49	62	40	43	61
48	53	67	63	78	64
48	54	51	56	63	69
58	51	58	59	56	57
38	76				

Với mức ý nghĩa 0,01 liệu có thể kết luận chi phí kiểm định trung bình trong công ty đã giảm ít hơn $60 không ?

15

Giá trị p

Giá trị p là giá trị xác suất thể hiện mức độ dữ liệu của mẫu ủng hộ hoặc bác bỏ giả thuyết H_0. Giá trị p càng nhỏ khả năng giả thuyết H_0 bị bác bỏ càng cao.



nếu giá trị p nhỏ hơn:

- 0,1 ta có một ít chứng cứ bác bỏ giả thuyết H_0.
- 0,05 ta có chứng cứ mạnh để bác bỏ giả thuyết H_0.
- 0,01 ta có chứng cứ rất mạnh để bác bỏ giả thuyết H_0.
- 0,001 ta có chứng cứ cực mạnh để bác bỏ giả thuyết H_0.

16

Kiểm định tỷ lệ tổng thể

Bước 1: Phát biểu giả thuyết H_0 và H_1.

Thí dụ: $H_0: \pi = \pi_0$

$H_1: : \pi \neq \pi_0$

Bước 2: Chọn mức ý nghĩa của kiểm định α. Thông thường α có các giá trị 0,1 hoặc 0,05 hoặc 0,01.

Bước 3: Tính giá trị thống kê kiểm định $z = \dfrac{p - \pi_0}{\sqrt{\dfrac{\pi_0(1-\pi_0)}{n}}}$

Bước 4 : Để dùng phân phối chuẩn trong kiểm định này, ta cần thỏa điều kiện $n\pi_0 \geq 10$ và $n(1-\pi_0) \geq 10$. Giá trị tới hạn được xác định tùy theo loại điểm định một đuôi hay hai đuôi. Trong trường hợp kiểm định 2 đuôi, các giá trị tới hạn là $-z_{\alpha/2}$ và $z_{\alpha/2}$.

Bước 5 : Ta so sánh giá trị thống kê kiểm định và giá trị tới hạn để đưa đến kết luận có bác bỏ giả thuyết H_0 hay không.

17

Thí dụ

- Thí dụ : Nhãn hiệu sữa tươi X-Milk có mức độ người tiêu dùng ưa thích trong một tỉnh lên đến 80%. Để đánh giá lại mức độ này, một khảo sát gồm 2000 người tiêu dùng trong tỉnh được thực hiện và kết quả cho thấy có 1550 người ưa thích X-Milk. Có thể kết luận mức độ ưa thích của người tiêu dùng với X-milk trong tỉnh ít nhất là 80% với mức ý nghĩa là 0,05 hay không ?

18

Hết chương 9

19

Kiểm định giả thuyết hai mẫu

Chương 10

Thống kê ứng dụng trong kinh doanh
Trần Tuấn Anh

Nội dung chính

- Biết cách thực hiện kiểm định giả thuyết về sự khác biệt giá trị trung bình của 2 tổng thể.
- Biết cách thực hiện kiểm định giả thuyết về sự khác biệt giá trị tỷ lệ của 2 tổng thể.
- Biết cách thực hiện kiểm định giả thuyết về trung bình 2 tổng thể trường hợp chọn mẫu phụ thuộc.
- Hiểu và biết cách áp dụng sự khác biệt giữa mẫu phụ thuộc và mẫu độc lập trong kiểm định giả thuyết về trung bình tổng thể.

2

Kiểm định giả thuyết trung bình 2 tổng thể - mẫu độc lập

- Thí dụ: Một chuyên viên nghiên cứu về nhân sự đang tìm hiểu sự khác biệt về tiền thưởng tết trung bình của các nhân viên kỹ thuật trong hai khu công nghiệp Phước Lộc A và Phước Lộc B. Trong trường hợp này, ta có 2 tổng thể liên quan: tổng thể nhân viên kỹ thuật làm việc trong khu công nghiệp Phước Lộc A và tổng thể nhân viên kỹ thuật làm việc trong khu công nghiệp Phước Lộc B.

3

Thí dụ:

Mẫu	Phước Lộc A	Phước Lộc B	Chênh lệch
1	29,8 triệu	28,76	1,04
2	30,32	29,40	0,92
3	30,57	29,94	0,63
4	30,04	28,93	1,11
5	30,09	29,78	0,31
6	30,02	28,66	1,36
7	29,6	29,13	0,47
8	29,63	29,42	0,21
9	30,17	29,29	0,88
10	28,74	29,21	-0,47

4

Trường hợp phương sai 2 tổng thể đã biết

- Giá trị thống kê kiểm định: $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

Thí dụ :Một công ty kinh doanh xe máy chính hãng thực hiện dịch vụ bảo dưỡng xe máy theo qui trình bảo dưỡng chuẩn tại cửa hàng bán xe. Để mở rộng dịch vụ này, công ty triển khai thêm một dịch vụ bảo dưỡng xe tận nhà. Theo phản ánh của trưởng phòng dịch vụ, thời gian bảo dưỡng xe máy tại nhà lớn hơn thời gian bảo dưỡng xe máy tại cửa hàng trong cùng một qui trình bảo dưỡng. Giám đốc công ty đề nghị nghiên cứu xem liệu bảo dưỡng xe máy tại nhà có lâu hơn tại cửa hàng trong cùng qui trình bảo dưỡng không? 2 mẫu khảo sát được thực hiện. Một mẫu gồm 100 khách hàng bảo dưỡng tại cửa hàng cho kết quả thời gian trung bình là 5,3 phút với độ lệch chuẩn của tổng thể là 0,3 phút. Một mẫu khác gồm 50 khách hàng bảo dưỡng tận nhà cho kết quả thời gian bảo dưỡng trung bình là 5,5 phút với độ lệch chuẩn tổng thể là 0,4 phút. Với mức ý nghĩa 0,01 bạn hãy kiểm định sự sai biệt về thời gian bảo dưỡng trung bình của 2 loại hình dịch vụ này.

5

Trường hợp phương sai 2 tổng thể chưa biết và giả định 2 phương sai này bằng nhau

- Phương sai chung: $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

- Giá trị thống kê kiểm định : $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

6

Thí dụ

- Thí dụ : Tại một nhà máy lắp ráp máy gặt đập liên hợp, các kỹ sư đang bàn bạc về 2 phương án lắp thiết bị cắt vào thân máy trong một công đoạn của dây chuyền sản xuất. Có ý kiến cho rằng 2 phương pháp lắp ráp này có thời gian lắp ráp là như nhau. Để đánh giá 2 phương pháp này, người ta chọn mẫu để đo thời gian lắp ráp của từng phương pháp. Kết quả chọn mẫu và đo thời gian lắp ráp (đơn vị là phút) của 2 phương pháp được cho trong bảng sau:

Phương pháp I	Phương pháp II
2	3
4	7
9	5
3	8
2	4
	3

Với mức ý nghĩa là 0,1 có thể kết luận thời gian lắp ráp của 2 phương pháp là khác nhau hay không?

7

Trường hợp phương sai 2 tổng thể chưa biết và giả định 2 phương sai này khác nhau

- Giá trị thống kê kiểm định $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)}}$

- Bậc tự do của kiểm định t: $df = \dfrac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}}$

- Thí dụ : Một nhóm nghiên cứu muốn tìm hiểu xem có sự khác biệt về giá bán thuốc A tại các nhà thuốc trong khu vực nội thành và ngoại thành hay không. Khảo sát 16 nhà thuốc trong khu vực nội thành và 15 nhà thuốc trong khu vực ngoại thành ta có kết quả như sau (đơn vị giá: ngàn đồng)

8

Thí dụ

Với mức ý nghĩa 0,05 có bằng chứng nào cho thấy giá bán thuốc A của các nhà thuốc ở 2 khu vực nội và ngoại thành khác nhau hay không?

Khu vực nội thành		Khu vực ngoại thành	
Mã nhà thuốc	Giá bán A	Mã nhà thuốc	Giá bán A
T001	125,05	N001	145,32
T002	137,56	N002	131,19
T003	142,50	N003	151,65
T004	145,95	N004	141,55
T005	117,49	N005	125,99
T006	142,75	N006	126,29
T007	121,99	N007	139,19
T008	117,49	N008	156,00
T009	141,64	N009	137,56
T010	128,69	N010	154,10
T011	130,29	N011	126,41
T012	142,39	N012	114,00
T013	121,99	N013	144,99
T014	141,30		
T015	153,43		
T016	133,39		

9

Kiểm định giả thuyết tỷ lệ 2 tổng thể

- Thí dụ: Khoa trưởng của một khoa muốn so sánh tỷ lệ sinh viên vắng mặt trên lớp trong 1 học kỳ giữa sinh viên hệ đại học và sinh viên hệ cao đẳng.

Tỷ lệ chung của 2 mẫu

$$p_c = \frac{X_1 + X_2}{n_1 + n_2}$$

Giá trị thống kê kiểm định:

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_c(1-p_c)}{n_1} + \frac{p_c(1-p_c)}{n_2}}}$$

10

Thí dụ

- Thí dụ : Hãng nước hoa Mely gần đây chuẩn bị đưa ra thị trường sản phẩm nước hoa Melym. Một nghiên cứu gần đây cho thấy nước hoa này có lượng khách hàng tiềm năng khá lớn. Phòng kinh doanh của hãng Mely đang quan tâm đến việc liệu tỷ lệ khách hàng muốn mua nước hoa Melym có khác nhau giữa nhóm khách hàng trẻ và nhóm khác hàng lớn tuổi hơn không ? Nghiên cứu thực hiện chọn mẫu ngẫu nhiên 100 khách hàng nữ trẻ thấy có 19 người có ý định mua nước hoa Melym. Tương tự, một mẫu gồm 200 khách hàng nữ lớn tuổi được khảo sát cho thấy có 62 người thích mua sản phẩm này. Với mức ý nghĩa 0,05, liệu có sự khác biệt nào giữa tỷ lệ khách hàng nữ trẻ tuổi và khác hàng nữ lớn tuổi muốn mua nước hoa này ?

11

Kiểm định giả thuyết về trung bình 2 tổng thể - trường hợp mẫu phụ thuộc

- Thí dụ : Công ty tài chính S&A thuê công ty Thanh Trúc và công ty Hoàng Phong thẩm định giá nhà của 10 căn nhà. Bảng sau đây cho giá thẩm định (đơn vị triệu đồng) của 10 căn này. Với mức ý nghĩa 0,05, ta có thể kết luận giá nhà do 2 công ty Thanh Trúc và Hoàng Phong thẩm định là khác nhau hay không ?

Nhà	Thanh Trúc	Hoàng Phong
1	235	228
2	210	205
3	231	219
4	242	240
5	205	198
6	230	223
7	231	227
8	210	215
9	225	222
10	249	245

12

Công thức kiểm định

- Giá trị thống kê kiểm định: $t = \dfrac{\bar{d}}{s_d / \sqrt{n}}$

- Trong đó: $s_d = \sqrt{\dfrac{\sum (d - \bar{d})^2}{n-1}}$

Hết chương 10

Phân tích phương sai

Chương 11

Thống kê ứng dụng trong kinh doanh
Trần Tuấn Anh

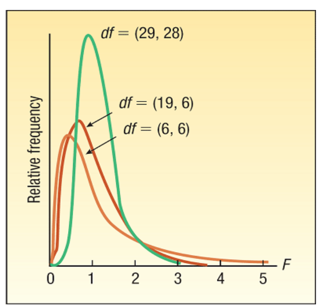
Nội dung chính

- Nắm được các đặc điểm cơ bản của phân phối F và cách sử dụng phân phối F.
- Biết cách thực hiện kiểm định giả thuyết về sự bằng nhau của phương sai 2 tổng thể.
- Nắm được các khái niệm cơ bản của phân tích phương sai.
- Biết cách tổ chức dữ liệu trong bảng ANOVA trong quá trình thực hiện phân tích phương sai.
- Biết cách dùng khoảng tin cậy của sự khác biệt trung bình tổng thể để so sánh tìm ra cặp tổng thể có trung bình tổng thể khác nhau.

2

Phân phối F

- Phân phối F là phân phối xác suất liên tục.
- Phân phối F không có giá trị âm, giá trị nhỏ nhất của phân phối F là 0.
- Phân phối F là phân phối có dạng nghiêng phải.
- Phân phối F tiệm cận với trục hoành nhưng không bao giờ tiếp xúc với trục này.
- Có nhiều phân phối F tùy vào 2 tham số: bậc tự do của tử số và bậc tự do của mẫu số.



3

So sánh phương sai 2 tổng thể

- Thí dụ : Minh Long là công ty sản xuất hàng thủ công mỹ nghệ tại tỉnh Long An. Công ty thường xuyên chuyển hàng từ công ty đến TPHCM theo 2 lộ trình L_1 và L_2. Giám đốc kho vận của công ty muốn nghiên cứu sự phân tán của thời gian vận chuyển hàng hóa trên 2 lộ trình này. Ông chọn mẫu 7 chuyến dùng lộ trình L_1 và 8 chuyến trên lộ trình L_2. Thời gian vận chuyển có đơn vị là phút của 2 mẫu được cho trong bảng sau:

Lộ trình L_1	Lộ trình L_2
52	59
67	60
56	61
45	51
70	56
54	63
64	57
	65

Với mức ý nghĩa 0,1 ta có thể kết luận có sự khác nhau của phương sai thời gian vận chuyển trên 2 lộ trình này không?

4

Công thức kiểm định

- H_0: $\sigma_1^2 = \sigma_2^2$
- H_1: $\sigma_1^2 \neq \sigma_2^2$

Giá trị thống kê kiểm định: $F = \frac{s_1^2}{s_2^2}$

Trong công thức này, giá trị phương sai nào của mẫu lớn hơn sẽ được đặt ở tử số. Do đó, F luôn lớn hơn hoặc bằng 1 và ta chỉ cần quan tâm đến đuôi phải khi kiểm định giả thuyết.

Giá trị tới hạn được xác định trên phân phối F với n_1-1 và n_2-1 bậc tự do và mức ý nghĩa của kiểm định. Giả thuyết H_0 bị bác bỏ nếu giá trị thống kê kiểm định lớn hơn hoặc bằng giá trị tới hạn.

5

Công thức kiểm định

Lưu ý:

Trong kiểm định 2 đuôi, ta chia đôi mức ý nghĩa α khi tra bảng F. Đối với kiểm định 1 đuôi, ta giữ nguyên giá trị α khi tra bảng F.

6

Phân tích phương sai

Để thực hiện phân tích phương sai, tình huống kiểm định của bạn phải thỏa một số điều kiện sau :

- Các tổng thể tuân theo luật phân phối chuẩn.
- Phương sai của các tổng thể bằng nhau.
- Các tổng thể độc lập với nhau.

7

Thí dụ

- Thí dụ : Giám đốc bộ phận chăm sóc khách hàng của một siêu thị muốn đo lường năng suất làm việc của các nhân viên chăm sóc khách hàng trong phòng chăm sóc khách hàng của siêu thị. Chỉ tiêu đo năng suất là số khách hàng được chăm sóc trong ngày. Để đo lường năng suất của 3 nhân viên Tâm, Trí và Tài, Giám đốc này theo dõi số liệu trong 4 ngày làm việc. Kết quả thu được trong bảng sau :

Tâm	Trí	Tài
55	66	47
54	76	51
59	67	46
56	71	48

Với mức ý nghĩa 0,05, liệu có thể kết luận số khách hàng trung bình được phục vụ mỗi ngày của 3 nhân viên này là khác nhau không ?

8

Dạng tổng quát của ANOVA

A_1	A_2	…	A_k
y_{11}	y_{12}	…	y_{1k}
y_{21}	y_{22}	…	y_{2k}
y_{31}	y_{32}	…	y_{3k}
…	…	…	…
n_1 quan sát $\bar{y}_1$	n_2 quan sát $\bar{y}_2$	…	n_k quan sát $\bar{y}_k$
		…	

9

ANOVA

Logic của ANOVA

$$(y_{ij} - \bar{y}) = (\bar{y}_j - \bar{y}) + (y_{ij} - \bar{y}_j)$$

Biến động tổng cộng là tổng các độ lệch bình phương giữa các giá trị quan sát và trung bình toàn bộ.

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y})^2$$

Biến động giữa các nhóm là tổng các độ lệch bình phương giữa các giá trị trung bình mỗi nhóm và trung bình toàn bộ.

$$SSG = \sum_{j=1}^{k} n_j(\bar{y}_j - \bar{y})^2$$

10

ANOVA

Biến động trong nội bộ nhóm là tổng các độ lệch bình phương giữa các giá trị quan sát của nhóm và trung bình nhóm đó.

$$SSW = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2$$

Mối quan hệ giữa SST, SSG và SSW

SST = SSG + SSW

11

Một số công thức

$$n = n_1 + n_2 + \ldots + n_k$$

$$\bar{y}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} y_{ij}$$

$$\bar{y} = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j} y_{ij} = \frac{1}{n}\sum_{j=1}^{k} n_j \bar{y}_j$$

12

Các bước phân tích phương sai

- Bước 1 : Phát biểu giả thuyết không và giả thuyết đối.
- H_0 : $\mu_1 = \mu_2 = \ldots = \mu_k$
- H_1 : Không phải tất cả các trung bình tổng thể đều bằng nhau.
- Bước 2 : Xác định mức ý nghĩa của phân tích phương sai.
- Bước 3 : Tính giá trị tới hạn của phân tích phương sai dựa trên phân phối F. Trong đó :
- Bậc tự do của tử là k - 1.
- Bậc tự do của mẫu là n - k.
- Bước 4 : tính giá trị thống kê kiểm định. $F = \frac{MSG}{MSW}$
- Giả thuyết H_0 bị bác bỏ nếu F ≥ giá trị tới hạn được xác định trong bước 3.

13

Bảng ANOVA

Nguồn biến thiên	Tổng các độ lệch bình phương	Bậc tự do	Trung bình các độ lệch bình phương	Giá trị kiểm định F
Giữa các nhóm	SSG	k - 1	$MSG = \frac{SSG}{k-1}$	$F = \frac{MSG}{MSW}$
Nội bộ các nhóm	SSW	n - k	$MSW = \frac{SSW}{n-k}$	
Tổng cộng	SST = SSG + SSW			

14

Thí dụ

- Thí dụ : Giám đốc nhân sự công ty Thanh Bình đang cân nhắc việc đánh giá 4 chuyên viên dịch vụ của công ty. Kết quả khảo sát ý kiến của khách hàng đánh giá các nhân viên này được cho trong bảng sau.

Đông	Tây	Nam	Bắc
94	75	70	68
90	68	73	70
85	77	76	72
80	83	78	65
	88	80	74
		68	65
		65	

Con số trong bảng là tổng điểm đánh giá khách hàng với 100 điểm là điểm cao nhất.
Với mức ý nghĩa 0,01, ta có thể kết luận điểm trung bình khách hàng đánh giá các nhân viên này khác nhau hay không ?

15

Bảng ANOVA của thí dụ

Nguồn biến thiên	Tổng các độ lệch bình phương	Bậc tự do	Trung bình các độ lệch bình phương	Giá trị kiểm định F
Giữa các nhóm	890,69	3	296,9	8,99
Nội bộ các nhóm	594,41	18	33,02	
Tổng cộng	1485,1			

16

Xác định cặp trung bình khác nhau

- Làm sao xác định cặp trung bình tổng thể nào khác nhau sau khi phân tích phương sai đưa đến kết luận tồn tại ít nhất một cặp trung bình tổng thể khác nhau?

Khoảng tin cậy của sự khác biệt các trung bình tổng thể

$$\bar{x}_1 - \bar{x}_2 \pm t\sqrt{MSW\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Nếu khoảng tin cậy được tính được có chứa giá trị 0 thì ta kết luận trung bình 2 tổng thể so sánh không khác nhau. Ngược lại, ta kết luận trung bình 2 tổng thể khác nhau.

17

Thí dụ

- Trong thí dụ ở phần trên, giả sử ta muốn so sánh sự khác nhau giữa nhân viên Đông và nhân viên Bắc.
- Áp dụng công thức, ta có :

$$\bar{x}_D - \bar{x}_B \pm t\sqrt{MSW\left(\frac{1}{n_D} + \frac{1}{n_B}\right)} = (87{,}25 - 69) \pm 2{,}101\sqrt{33{,}02\left(\frac{1}{4} + \frac{1}{6}\right)} = 18{,}25 \pm 7{,}79$$

Như vậy, khoảng tin cậy là từ 10,46 đến 26,04 không chứa giá trị 0 nên ta kết luận điểm trung bình khách hàng đánh giá 2 nhân viên Đông và Bắc là khác nhau.

18

Hết chương 11

19

Tương quan & hồi qui tuyến tính

Chương 12

Thống kê ứng dụng trong kinh doanh
Trần Tuấn Anh

Nội dung chính

- Hiểu và áp dụng được các biến độc lập và biến phụ thuộc trong phân tích thống kê.
- Biết cách tính toán và diễn giải hệ số tương quan, hệ số xác định và sai số chuẩn của ước lượng.
- Biết cách thực hiện kiểm định giả thuyết để xác định sự tương quan của 2 tổng thể.
- Biết cách tính các thông số của đường hồi qui mẫu.
- Biết cách tính ước lượng các tham số của phương trình hồi qui tuyến tính tổng thể.
- Biết kiểm định giả thuyết về hồi qui tuyến tính của 2 tổng thể.
- Nắm được những nội dung cơ bản dự báo dựa vào quan hệ hồi qui tuyến tính.

2

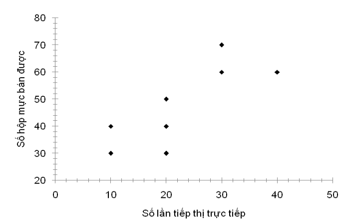
Tương quan

- Thí dụ : Một công ty kinh doanh thiết bị văn phòng muốn nghiên cứu mối liên hệ giữa số hộp mực máy in bán được trong tuần và số lần các tiếp thị ghé thăm các khách hàng. Giám đốc kinh doanh chọn mẫu 10 nhân viên bán hàng trực tiếp và ghi nhận số lần các nhân viên đến tiếp thị trực tiếp khách hàng và số hộp mực máy in bán được. Kết quả được cho trong bảng sau:

Nhân Viên	Số lần tiếp thị trực tiếp	Số hộp mực bán được
Thông	20	30
Hùng	40	60
Bình	20	40
Giang	30	60
Sinh	10	30
Nam	10	40
Cao	20	40
Kha	20	50
Minh	20	30
Toàn	30	70

3

Tương quan



4

Hệ số tương quan mẫu

Hệ số tương quan mẫu

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

$$ss_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$ss_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$ss_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Hệ số tương quan mẫu

$$r = \frac{ss_{xy}}{\sqrt{ss_{xx}}\sqrt{ss_{yy}}}$$

Thí dụ:

5

Đặc điểm của hệ số tương quan mẫu r

- Giá trị của r nằm trong khoảng từ -1 đến 1.
- Hệ số r gần 0 cho biết 2 yếu tố x và y rất ít tương quan với nhau. Nếu r = 0 thì x và y không có mối tương quan với nhau.
- Hệ số r gần 1 cho thấy có mối tương quan thuận mạnh. Tức là 2 yếu tố này cùng tăng, cùng giảm (đồng biến) với nhau.
- Hệ số r gần -1 cho thấy có mối tương quan nghịch mạnh. Tức là yếu tố này tăng thì yếu tố kia giảm và ngược lại (nghịch biến).

6

Kiểm định mối quan hệ tương quan

Giả thuyết : $H_0 : \rho = 0$

$H_1 : \rho \neq 0$

Giá trị thống kê kiểm định: $t = r\sqrt{\frac{n-2}{1-r^2}}$

Trong đó, phân phối t với bậc tự do là n - 2 được dùng để xác định giá trị tới hạn trong kiểm định giả thuyết.

Mức ý nghĩa của kiểm định giả thuyết là α.

Giả thuyết H_0 bị bác bỏ khi $t \leq -t_{n-1,\alpha/2}$ hoặc $t > t_{n-1,\alpha/2}$ trong trường hợp kiểm định 2 đuôi.

Thí dụ:

7

Hồi qui tuyến tính

Biến độc lập là biến cung cấp cơ sở cho ước lượng, dự báo. Nó còn được gọi là biến tiên đoán (thuật ngữ tiếng anh là preditor).

Biến phụ thuộc là biến được ước lượng, dự báo. Nó còn được gọi là biến đáp ứng (thuật ngữ tiếng anh là response).

8

Thí dụ

Thí dụ : Một khảo sát chọn mẫu 10 sinh viên ghi nhận số giờ ôn tập môn Tiếng Anh chuyên ngành I và điểm thi môn này được cho như sau :

Sinh viên	Số giờ ôn tập	Điểm thi
Thanh	1	53
Quang	5	74
Minh	7	59
Kỳ	8	43
Thông	10	56
Mẫn	11	84
Tuấn	14	96
Thành	15	69
Công	15	84
Phong	19	83

Mối quan hệ hồi qui tuyến tính giữa số giờ ôn tập và điểm thi?

9

Phương trình hồi qui tuyến tính

Phương trình hồi qui tuyến tính của tổng thể

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Phương trình hồi qui tuyến tính của mẫu

$$\hat{y}_i = b_0 + b_1 x_i$$

Sai số giữa giá trị quan sát và giá trị ước lượng

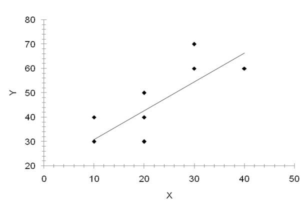
$$e_i = y_i - \hat{y}_i$$

10

Phương trình hồi qui tuyến tính

$$\sum_{i=1}^{n}(y_i - \hat{y}_i) = 0$$

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2$$



11

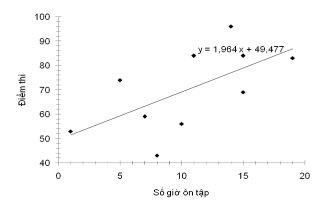
Phương trình hồi qui tuyến tính

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$b_1 = \frac{ss_{xy}}{ss_{xx}}$$

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

Thí dụ:



12

Hệ số xác định

- Trong phân tích hồi qui, hệ số xác định được dùng để chỉ phần trăm sự biến động của biến phụ thuộc được biến độc lập giải thích

Hệ số xác định

$$R^2 = 1 - \frac{SSE}{SST} = \frac{SSR}{SST}$$

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

SST = SSR + SSE

Thí dụ:

13

Khoảng tin cậy của β_1 và β_0

Sai số chuẩn của hồi qui

$$s_{yx} = \sqrt{\frac{SSE}{n-2}}$$

Sai số chuẩn của b_1

$$s_{b_1} = \frac{s_{yx}}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

Khoảng tin cậy của β_1

$$b_1 - t_{n-2} s_{b_1} \le \beta_1 \le b_1 + t_{n-2} s_{b_1}$$

14

Khoảng tin cậy của β_1 và β_0

Sai số chuẩn của b_0

$$s_{b_0} = s_{yx} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

Khoảng tin cậy của β_0

$$b_0 - t_{n-2} s_{b_0} \le \beta_0 \le b_0 + t_{n-2} s_{b_0}$$

Thí dụ:

15

Kiểm định giả thuyết về quan hệ hồi qui tuyến tính

Giả thuyết :H_0 : $\beta_1 = 0$ (không có mối quan hệ hồi qui tuyến tính)
H_1 : $\beta_1 \neq 0$ (có mối quan hệ hồi qui tuyến tính)

Giá trị kiểm định: $$t = \frac{b_1}{s_{b_1}}$$

Giả thuyết H_0 bị bác bỏ nếu $t \le -t_{n-2,\alpha/2}$ hoặc $t \ge t_{n-2,\alpha/2}$ trong trường hợp kiểm định 2 đuôi. Giá trị tới hạn được tra trong bảng phân phối t.

Thí dụ:

16

Khoảng tin cậy của Y

Khoảng ước lượng giá trị của y_i với độ tin cậy cho trước

$$\hat{y}_i \pm t_{n-2} s_{yx} \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

Khoảng ước lượng giá trị trung bình của y_i với độ tin cậy cho trước

$$\hat{y}_i \pm t_{n-2} s_{yx} \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

Thí dụ:

17

Hết chương 12

18