ÔN THI CUỐI KỲ
MÔN: XÁC SUẤT THỐNG KÊ
HỌC KỲ 2 NĂM HỌC 2023-2024
ĐH KHOA HỌC TỰ NHIÊN, ĐHQG-HCM

Nguyễn Văn Thùy, HCMUS

Nội dung:

1) *Phân phối chuẩn; xấp xỉ phân phối nhị thức bằng phân phối chuẩn;*
2) *Khoảng tin cậy cho trung bình;*
3) *Khoảng tin cậy cho tỷ lệ;*
4) *Tìm cỡ mẫu khi biết sai số ước lượng;*
5) *So sánh trung bình với 1 số;*
6) *So sánh tỷ lệ với 1 số;*
7) *So sánh hai trung bình;*
8) *So sánh hai tỷ lệ (hai mẫu độc lập);*
9) *Hồi quy: tìm đường hồi quy; ước lượng; tính R^2; hệ số tương quan r_{x;y}.*

VẤN ĐỀ 1. PHÂN PHỐI CHUẨN; XẤP XỈ PHÂN PHỐI NHỊ THỨC BẰNG PHÂN PHỐI CHUẨN

- ***Phân phối chuẩn tắc:*** *Đại lượng ngẫu nhiên Z được gọi là một ĐLNN có phân phối chuẩn tắc, ký hiệu* $Z \sim N(0;1)$ *nếu hàm mật độ của nó là*

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

- *Hàm phân phối của Z, ký hiệu bởi* $\Phi(z)$, *là*

$$\Phi(z) = \int_{-\infty}^{z} f(t)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt$$

- *Bảng giá trị của hàm Φ được lập sẵn với* $z > 0$. *Với* $z < 0$ *ta dùng công thức* $\Phi(-z) = 1 - \Phi(z)$
- *Ta có:*

$$P(Z < a) = \Phi(a); P(Z > a) = 1 - \Phi(a); P(a < Z < b) = \Phi(b) - \Phi(a)$$

- ***Phân phối chuẩn:*** *ĐLNN X được gọi là có phân phối chuẩn với tham số* μ *và* σ^2 *(ở đó* $\sigma > 0$*) nếu ĐLNN* $Z = \frac{X-\mu}{\sigma}$ *có phân phối chuẩn tắc. Khi đó ta ký hiệu* $X \sim N(\mu; \sigma^2)$.

Ex Var$_X$

- *Ta có*

$$P(X < a) = P\left(Z < \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(X > a) = 1 - P(X < a) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

- ***Xấp xỉ phân phối nhị thức bằng phân phối chuẩn:*** *Nếu X là biến ngẫu nhiên có phân phối nhị thức với hai tham số n và p,*

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

Là biến ngẫu nhiên xấp xỉ phân phối chuẩn. Để ước tính xác suất nhị thức bằng phân phối chuẩn, việc hiệu chỉnh tính liên tục được áp dụng như sau

$$P(X \le x) = P(X \le x + 0{,}5) \approx P\left(Z \le \frac{x + 0{,}5 - np}{\sqrt{np(1-p)}}\right)$$

$$P(X \ge x) = P(X \ge x - 0{,}5) \approx P\left(Z \ge \frac{x - 0{,}5 - np}{\sqrt{np(1-p)}}\right)$$

Xấp xỉ là tốt khi $np > 5$ *và* $n(1-p) > 5$

Bài 1.. Trọng lượng của một gói đường (đóng bằng máy tự động) có phân phối chuẩn. Trong 1000 gói đường có 70 gói có trọng lượng lớn hơn 1015g. Hãy ước lượng xem có bao nhiêu gói đường có trọng lượng ít hơn 1008g, biết rằng trọng lượng trung bình của 1000 gói đường là 1012g.

Bài 2. Trọng lượng sản phẩm của nhà máy H là biến ngẫu nhiên X có phân phối chuẩn với trọng lượng trung bình là 100 gam và độ lệch chuẩn là 0,5 gam. Tính xác suất để trong 10 sản phẩm của nhà máy H không có sản phẩm nào trọng lượng dưới 99 gam.

Bài 3. Một máy sản xuất sản phẩm với tỷ lệ loại I là 20%. Cho máy sản xuất 100 sản phẩm. Tính xác suất trong 100 sản phẩm đó có không dưới 18 sản phẩm loại I.

Bài 4. Một xạ thủ có xác suất bắn trúng của mỗi phát là 0,8. Xạ thủ này bắn 64 phát vào bia. Tính xác suất

a) Có từ 45 đến 52 phát trúng bia;

b) Không dưới 51 phát trúng bia.

Bài 5. Công ty Đất Xanh Miền Nam chính thức mở bán 926 căn hộ của Chung cư Sài Gòn Gateway Quận 9. Xác suất bán được của mỗi căn hộ là 0,6. Tính xác suất công ty bán được ít nhất 400 căn trong lần mở bán này.

VẤN ĐỀ 2. KHOẢNG TIN CẬY CHO TRUNG BÌNH

- *Có 3 trường hợp*
- *TH1. Phương sai σ^2 đã biết (Bổ sung thêm: $n \geq 30$ hoặc $n < 30$ và có phân phối chuẩn)*
- *TH2. Phương sai σ^2 chưa biết và cỡ mẫu $n \geq 30$*
- *TH3. Phương sai σ^2 chưa biết và cỡ mẫu $n < 30$ và có phân phối chuẩn*
- *B1) Chuẩn bị: Tìm trung bình mẫu $\bar{x}$ và phương sai mẫu s; xác định trường hợp áp dụng*
- *B2) Tìm phân vị $z_{1-\frac{\alpha}{2}}$ (TH1; TH2) hoặc $t^{n-1}_{1-\frac{\alpha}{2}}$ (TH3)*
- *B3) Tìm dung sai*

$$\varepsilon = \begin{cases} z_{1-\frac{\alpha}{2}} \dfrac{\sigma}{\sqrt{n}}; TH1 \\ z_{1-\frac{\alpha}{2}} \dfrac{s}{\sqrt{n}}; TH2 \\ t^{n-1}_{1-\frac{\alpha}{2}} \dfrac{s}{\sqrt{n}}; TH3 \end{cases}$$

- *B4) Kết luận: Khoảng tin cậy $100(1-\alpha)\%$ cho trung bình của tổng thể là $(\bar{x} - \varepsilon;\ \bar{x} + \varepsilon)$.*

Bài 6. Đo sức bền chịu lực của một loại ống thí nghiệm, người ta thu được bộ số liệu sau: 4500; 6500; 5200; 4800; 4900; 5125; 6200; 5375. Từ kinh nghiệm nghề nghiệp, người ta cũng biết rằng sức bền đó có phân phối chuẩn và độ lệch chuẩn là $\sigma = 300$. Hãy xây dựng khoảng tin cậy 90% cho sức bền trung bình của loại ống trên.

Bài 7. Sản lượng mỗi ngày của một phân xưởng là biến ngẫu nhiên có phân phối chuẩn. Kết quả thống kê của 9 ngày là: 27; 26; 21; 28; 25; 30; 26; 23; 26. Hãy xác định khoảng tin cậy 95% cho sản lượng trung bình.

Bài 8. Quan sát chiều cao X (cm) của một số người, ta được bảng số liệu sau.

$x\,(cm)$	140-145	145-150	150-155	155-160	160-165	165-170
Số người	1	3	7	9	5	2

Biết X có phân phối chuẩn. Hãy ước lượng chiều cao trung bình ở độ tin cậy 0,95.

Bài 9. Đo đường kính X (đơn vị: mm) của một loại chi tiết máy do xí nghiệp M sản xuất ta thu được bảng số liệu sau:

X	86-88	88-90	90-92	92-94	94-96	96-98	98-100
Số chi tiết	37	45	69	83	71	45	32

Hãy ước lượng đường kính trung bình của các chi tiết máy với độ tin cậy 96%, biết X có phân phối chuẩn.

Bài 10. Để đánh giá mức độ ảnh hưởng của vụ xì căng đan J làm giảm doanh thu của thương hiệu F, người ta điều tra doanh thu X (đơn vị: trăm triệu đồng/tháng) của một số cửa hàng được chọn ngẫu nhiên của thương hiệu này trong một tháng và thu được bảng số liệu sau:

X	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12
Số cửa hàng	25	39	65	82	96	89	78	56	36	18

Hãy tìm khoảng tin cậy cho doanh thu trung bình trong 1 tháng sau vụ J của các cửa hàng thuộc thương hiệu F với độ tin cậy 99%.

VẤN ĐỀ 3. KHOẢNG TIN CẬY CHO TỶ LỆ

- ***Bài toán:*** *Tìm khoảng tin cậy cho tỷ lệ p của tổng thể với độ tin cậy* $1-\alpha$ *từ mẫu ngẫu nhiên*
- ***Phương pháp:***
- *B1) Chuẩn bị: Tìm tỷ lệ mẫu* f *và kiểm tra các điều kiện* $nf > 5; n(1-f) > 5$
- *B2) Tìm phân vị* $z_{1-\frac{\alpha}{2}}$ *từ độ tin cậy* $1-\alpha$
- *B3) Tìm dung sai*

$$\varepsilon = z_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}$$

- *B4) Kết luận: Khoảng tin cậy* $100(1-\alpha)\%$ *cho tỷ lệ của tổng thể là* $(f-\varepsilon;\ f+\varepsilon)$.

Bài 11. Quan sát trong 3 giờ ở sân bay Tân Sơn Nhất có 27 máy bay đến trễ. Số phút trễ của các chuyến bay được liệt kê trong bảng sau:

Số phút trễ (X)	0-10	10-20	20-30	30-40	40-50	50-60
Số chuyến	8	6	4	3	4	2

Hãy ước lượng tỷ lệ chuyến bay trễ trên 30 phút với độ tin cậy 95%.

Bài 12. Một máy sản xuất hoạt động bình thường đóng gói các sản phẩm có khối lượng trung bình là 1kg. Nghi ngờ máy hoạt động không bình thường, người ta khảo sát khối lượng của 100 sản phẩm thì thấy như sau

Khối lượng (kg)	0,94-0,96	0,96-0,98	0,98-1	1-1,02	1,02-1,04	1,04-1,06
Số sản phẩm	9	31	40	15	3	2

Tìm khoảng tin cậy của tỷ lệ sản phẩm do máy đóng gói có khối lượng dưới 1kg với độ tin cậy 97%.

VẤN ĐỀ 4. TÌM CỠ MẪU KHI BIẾT SAI SỐ ƯỚC LƯỢNG

- ***Bài toán:*** *Tìm cỡ mẫu tối thiểu sao cho dung sai (trong bài toán ước lượng trung bình; ước lượng tỷ lệ) không vượt quá ngưỡng cho trước*
- ***Phương pháp:***
- *B1) Chuẩn bị (như Vấn đề 2+Vấn đề 3)*
- *B2) Tìm phân vị*
- *B3) Viết biểu thức của sai số ước lượng*
- *B4) Giải bất phương trình: sai số ước lượng **nhỏ hơn hoặc bằng** ngưỡng cho trước*
- ***Chú ý:***
- *Sai số ước lượng trong bài toán ước lượng trung bình*

$$\varepsilon = \begin{cases} z_{1-\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}; TH1 \\ z_{1-\frac{\alpha}{2}}\dfrac{s}{\sqrt{n}}; TH2 \\ t_{1-\frac{\alpha}{2}}^{n-1}\dfrac{s}{\sqrt{n}}; TH3 \end{cases}$$

- *Sai số ước lượng trong bài toán ước lượng cho tỷ lệ*

$$\varepsilon = z_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}$$

- ***Để ý: nếu f chưa biết thì dùng tính chất***

$$f(1-f) = -f^2 + f = -\left(f - \frac{1}{2}\right)^2 + \frac{1}{4} \le \frac{1}{4}; \forall f \in \mathbb{R}$$

Bài 13. Quan sát tuổi thọ X (giờ) của một số bóng đèn do xí nghiệp A sản xuất, ta được bảng số liệu

X (giờ)	1000	1100	1200	1300	1400	1500	1600	1700	1800
Số trường hợp	10	14	16	17	18	16	16	12	9

Nếu muốn sai số khi ước lượng tuổi thọ trung bình của bóng đèn không quá 30 giờ ở độ tin cậy 0,99 thì quan sát mẫu ít nhất mấy trường hợp?

Bài 14. Để ước lượng xác suất mắc bệnh gan với độ tin cậy 90% và sai số không vượt quá 2% thì cần phải khám ít nhất bao nhiêu người, biết rằng tỷ lệ mắc bệnh gan thực nghiệm đã cho bằng 0,9.

Bài 15. Ta muốn ước lượng tỷ lệ viên thuốc bị sức mẻ p trong một lô thuốc rất nhiều. Nếu ta muốn sai số ước lượng không quá 0,01 ở độ tin cậy 0,95 thì phải quan sát ít nhất mấy viên?

VẤN ĐỀ 5. SO SÁNH TRUNG BÌNH VỚI MỘT SỐ

- ***Có 3 trường hợp***
- *TH1. Phương sai σ^2 đã biết (Bổ sung thêm: $n \geq 30$ hoặc $n < 30$ và có phân phối chuẩn)*
- *TH2. Phương sai σ^2 không biết và cỡ mẫu lớn: $n \geq 30$*
- *TH3. Phương sai σ^2 không biết và cỡ mẫu nhỏ: $n < 30$*
- ***Phương pháp***
- *B1) Phát biểu giả thuyết và đối thuyết: có 3 trường hợp*

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}; \quad \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}; \quad \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$$

- *B2) Tìm phân vị*
- *B3) Tính giá trị thống kê kiểm định (hoặc p-giá trị)*
- *B4) Dựa vào miền bác bỏ, kết luận: Với mức ý nghĩa α, ta bác bỏ H_0 hoặc không đủ cơ sở để bác bỏ H_0. Nghĩa là ...(trả lời theo yêu cầu của bài toán)*
- ***Miền bác bỏ***

Đối thuyết	*Bác bỏ H_0 khi*	
	TH1; TH2	*TH3*
$\mu \neq \mu_0$	$\lvert z_0\rvert = \dfrac{\lvert\bar{x} - \mu_0\rvert\sqrt{n}}{\sigma} > z_{1-\frac{\alpha}{2}}$	$\lvert t_0\rvert = \dfrac{\lvert\bar{x} - \mu_0\rvert}{s}\sqrt{n} > t^{n-1}_{1-\frac{\alpha}{2}}$
$\mu > \mu_0$	$z_0 = \dfrac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} > z_{1-\alpha}$	$t_0 = \dfrac{(\bar{x} - \mu_0)\sqrt{n}}{s} > t^{n-1}_{1-\alpha}$
$\mu < \mu_0$	$z_0 = \dfrac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} < -z_{1-\alpha}$	$t_0 = \dfrac{(\bar{x} - \mu_0)\sqrt{n}}{s} < -t^{n-1}_{1-\alpha}$

- ***Phương pháp p-giá trị***

Đối thuyết	*Bác bỏ H_0 khi p-giá trị $< \alpha$*	
	TH1; TH2	*TH3*
$\mu \neq \mu_0$	$p = 2[1 - \Phi(\lvert z_0\rvert)]$	$p = 2P[t(n-1) \geq \lvert t_0\rvert]$
$\mu > \mu_0$	$p = 1 - \Phi(z_0)$	$p = P[t(n-1) \geq t_0]$
$\mu < \mu_0$	$p = \Phi(z_0)$	$p = P[t(n-1) \leq t_0]$

Bài 16. Một máy đóng gói sản phẩm của công ty A phải dừng hoạt động để điều chỉnh nếu trọng lượng trung bình của một gói đóng ra khác 100 gam. Người phụ trách máy cân ngẫu nhiên một số gói đã đóng ra và thu được bảng số liệu

Trọng lượng (gam)	97-98	98-99	99-100	100-101	101-102	102-203	103-104
Số sản phẩm	24	34	36	42	31	27	20

Với mức ý nghĩa 1%, người phụ trách máy có phải dừng máy để điều chỉnh hay không? Tính p-giá trị.

Bài 17. Mẫu điều tra về giá bán của cổ phiếu A trên thị trường (đơn vị: 1000 đồng) trong các phiên giao dịch cho kết quả trong bảng

Giá của A	[11-13]	(13;15]	(15;17]	(17;19]	(19;21]	(21:23]	(23:25]	(25:27]
Số phiên	5	18	25	40	38	27	16	3

Giả sử giá cổ phiếu có phân phối chuẩn.

Trước đây trung bình giá cổ phiếu A trong mỗi phiên là 19,3 ngàn đồng. Với mức ý nghĩa 5% hãy xem trung bình giá cổ phiếu A có giảm không? Tính p-giá trị.

Bài 18. Trọng lượng của một sản phẩm theo quy định là 6 kg. Sau một thời gian sản xuất, người ta tiến hành kiểm tra 121 sản phẩm và tính được trung bình mẫu là 5,975 kg và phương sai mẫu hiệu chỉnh là 5,7596 kg^2. Sản xuất được xem là bình thường nếu các sản phẩm có trọng lượng trung bình bằng trọng lượng quy định. Với mức ý nghĩa 5%, hãy kết luận về tình hình sản xuất.

Bài 19. Trọng lượng trung bình khi xuất chuồng ở một trại chăn nuôi trước là 2,8 kg/con. Năm nay người ta sử dụng một loại thức ăn mới. Cân thử 25 con khi xuất chuồng, người ta tính được trung bình mẫu là 3,2 kg và phương sai mẫu hiệu chỉnh là 0,25 kg^2. Biết trọng lượng gà ở trang trại này có phân phối chuẩn.

a) Với mức ý nghĩa 5%, hãy kết luận xem loại thức ăn này có làm tăng trọng lượng trung bình của đàn gà hay không?

b) Nếu trại chăn nuôi báo cáo trọng lượng trung bình khi xuất chuồng là 3,3 kg/con thì có chấp nhận được không với mức ý nghĩa 5%?

Bài 20. Tuổi thọ X (đơn vị : giờ) của một loại sản phẩm do một dây chuyền sản xuất là biến ngẫu nhiên có phân phối chuẩn với tuổi thọ trung bình là 1000 giờ. Sau một thời gian sản xuất người ta nghi ngờ dây chuyền sản xuất hoạt động không bình thường. Kiểm tra ngẫu nhiên 29 sản phẩm do dây chuyền này sản xuất ta thu được tuổi thọ trung bình của 29 sản phẩm này là 990 giờ và độ lệch chuẩn mẫu hiệu chỉnh là 25 giờ. Hãy kết luận về nghi ngờ nói trên với mức ý nghĩa 5%.

VẤN ĐỀ 6. SO SÁNH TỶ LỆ VỚI MỘT SỐ

- ***Bài toán:*** *So sánh tỷ lệ p (của các phần tử có đặc điểm A từ tổng thể X) với một số p_0 cho trước*
- ***Phương pháp***
- *B1) Chuẩn bị: tìm n; f rồi kiểm tra các điều kiện:* $nf \geq 5$; $n(1-f) \geq 5$
- *B2) Phát biểu giả thuyết và đối thuyết: có 3 trường hợp*

$$\begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases}; \quad \begin{cases} H_0: p = p_0 \\ H_1: p > p_0 \end{cases}; \quad \begin{cases} H_0: p = p_0 \\ H_1: p < p_0 \end{cases}$$

- *B3) Tìm phân vị: tùy theo đối thuyết*
- *B4) Tính giá trị thống kê kiểm định*

$$z_0 = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}}$$

- *B5) Dựa vào điều kiện bác bỏ giả thuyết H_0, ta kết luận: Với mức ý nghĩa α, bác bỏ H_0 hoặc không đủ cơ sở để bác bỏ H_0. Nghĩa là ...(trả lời theo yêu cầu của bài toán)*
- ***Điều kiện bác bỏ***

Đối thuyết	*Bác bỏ H_0 khi*
$p \neq p_0$	$\lvert z_0\rvert = \frac{\lvert f - p_0\rvert\sqrt{n}}{\sqrt{p_0(1 - p_0)}} > z_{1-\frac{\alpha}{2}}$
$p > p_0$	$z_0 = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}} > z_{1-\alpha}$
$p < p_0$	$z_0 = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}} < -z_{1-\alpha}$

Bài 21. Một thành phố A, trong 300 người hút thuốc lá có 36 người hút ít nhất hai gói thuốc trong một ngày. Có ý kiến cho rằng tỷ lệ người hút thuốc ít nhất hai gói một ngày ở thành phố A không lớn hơn 8%. Hãy cho nhận xét về ý kiến trên với mức ý nghĩa 3%.

Bài 22. Một khách hàng nhận được lô hàng từ một nhà máy. Lô hàng sẽ bị từ chối nếu có trên 4% sản phẩm không đạt yêu cầu. Khách hàng kiểm tra ngẫu nhiên 450 sản phẩm và thấy 29 sản phẩm không đạt yêu cầu. Với mức ý nghĩa 5%, khách hàng có thể từ chối lô hàng được không?