



2015

ÔN TẬP PHẦN THỐNG KÊ

- Phần này chỉ ôn lại một số dạng toán cơ bản thường gặp.
- Các dạng toán còn lại sinh viên/ học viên tự xem trong quyển bài tập XSTK.



KHẮC CỐT GHI TÂM

- Một câu hỏi về Thống kê thường chỉ thuộc 1 trong 2 dạng: *ước lượng* hoặc *kiểm định*.
- **Nếu là ước lượng:** UL trung bình hay UL tỷ lệ.... UL điểm (**không cho độ tin cậy**) hay UL khoảng (**có cho độ tin cậy**).
- Xem thuộc dạng toán nào trong 3 dạng toán đã học (có 3 tham số; biết 2 tham số, tìm 1 tham số còn lại). UL tỷ lệ thêm 2 dạng toán (biết M tìm N, biết N tìm M).
- Nếu là UL trung bình thì xem mẫu lớn hay nhỏ, biết σ hay chưa biết σ .
- Nếu là UL tỷ lệ thì mẫu phải lớn.

KHẮC CỐT GHI TÂM (tt)

- **Nếu là kiểm định:** KĐ trung bình hay KĐ tỷ lệ....
- Xem con số cần kiểm định là con số nào.
- Xem kiểm định 2 phía hay 1 phía.
- Nếu là KĐ trung bình thì xem mẫu lớn hay nhỏ, biết σ hay chưa biết σ .
- Nếu là KĐ tỷ lệ thì mẫu phải lớn.
- **Làm toán về thống kê phải chú ý là đưa về cùng đơn vị tính/ đo.**

KHẮC CỐT GHI TÂM (phần bổ sung)

- *Nếu là ước lượng phương sai thì xem biết kỳ vọng hay chưa biết kỳ vọng.*
- *Nếu là KĐ tính độc lập của 2 dấu hiệu thì phải xác định số điểm (khoảng) chia của từng dấu hiệu quan sát.*
- *Nếu là KĐ Quy luật phân phối thì phải biết phân phối gì, các tham số của phân phối biết hay chưa, số điểm (khoảng) chia của mẫu quan sát.*

BÀI 1:

Nhà trường muốn đánh giá số giờ tự học của sinh viên trong tuần. Để biết điều này, phòng đào tạo chọn ngẫu nhiên 25 sinh viên và được kết quả sau:

9	8	7	6	7
8	9	4	7	6
6	4	11	5	4
3	7	8	8	7
6	2	2	8	6

5

- 1) Ước lượng số giờ tự học trung bình của sinh viên trong tuần?
- 2) Ước lượng số giờ tự học trung bình của sinh viên trong tuần, với độ tin cậy 95%? (Giả thiết số giờ tự học của sinh viên trong tuần tuân theo luật phân phối chuẩn)
- 3) Một báo cáo trong quá khứ nói rằng: số giờ tự học trung bình của sinh viên trong tuần là 8. Với mức ý nghĩa 5%, hãy so sánh kết quả mỗi điều tra này với kết quả trong quá khứ?

6

Hướng dẫn :

Số giờ tự học	2	3	4	5	6	7	8	9	11
Số sinh viên	2	1	3	1	5	5	5	2	1

$n = 25, \sum n_i x_i = 158, \sum n_i x_i^2 = 1118, \bar{x} = \frac{1}{n} \sum n_i x_i = 158/25 = 6,32$
 $s^2 = \frac{1}{n-1} (\sum n_i x_i^2 - n(\bar{x})^2) = (1118 - 25(6,32)^2)/24 = 4,9767$
 $s = \sqrt{s^2} = 2,2309$
1) Gọi μ là số giờ tự học trung bình của sinh viên trong tuần của toàn trường. Ta dùng \bar{x} để ước lượng μ
Vậy số giờ tự học trung bình trong tuần của sinh viên là 6,32 giờ.

7

- 2) $n = 25 < 30, \sigma$ chưa biết
 $\gamma = 95\% \Rightarrow \alpha = 1 - \gamma = 5\% \Rightarrow t_{\alpha/2}(n-1) = t_{0,025}(24) = 2,064$
 $\varepsilon = t_{\alpha/2}(n-1) \frac{s}{\sqrt{n}} = (2,064)(2,2309)/\sqrt{25} = 0,9209$
Khoảng tin cậy là (5,3991 ; 7,2409)
Vậy số giờ tự học TB của SV là (5,3991 giờ ; 7,2409 giờ)
- 3) Lập giả thiết $H_0 : \mu = 8 ; H_1 : \mu \neq 8$
 μ : số giờ tự học TB của SV trong tuần hiện nay
 $\mu_0 = 8$: số giờ tự học TB của SV trong tuần trong quá khứ
 $\alpha = 5\% \Rightarrow t_{0,025}(24) = 2,064$
 $t = \frac{(\bar{x} - \mu_0)\sqrt{n}}{s} = (6,32 - 8)\sqrt{25}/2,2309 = -3,7653$
Ta có: $|t| > t_{0,025}(24)$: bác bỏ H_0
Vậy số giờ tự học TB của sinh viên trong tuần hiện nay ít hơn trong quá khứ (do $\bar{x} = 6,32 < \mu_0 = 8$)

8

BÀI 2

Số liệu thống kê về doanh số bán của 1 siêu thị trong một số ngày cho ở bảng sau:

Doanh số (triệu đ/ ngày)	Số ngày
24	5
30	12
36	25
42	35
48	24
54	15
60	12
65	10
70	6

9

- 1) Ước lượng doanh số bán trung bình trong 1 ngày của siêu thị, với độ tin cậy 95%?
- 2) Những ngày có doanh số bán từ 60 triệu đ trở lên là những ngày "bán đắt hàng". Hãy ước lượng tỷ lệ những ngày bán đắt hàng của siêu thị?
- 3) Ước lượng tỷ lệ những ngày "bán đắt hàng" của siêu thị, với độ tin cậy 99%?
- 4) Ước lượng doanh số bán trung bình của 1 ngày "bán đắt hàng" ở siêu thị, với độ tin cậy 95%? (Giả thiết doanh số bán của những ngày bán đắt hàng là đại lượng ngẫu nhiên phân phối theo quy luật chuẩn)



10

5) Trước đây doanh số bán trung bình của siêu thị là 35 triệu đ/ngày. Số liệu ở bảng trên được thu thập sau khi siêu thị áp dụng 1 phương thức bán hàng mới. Hãy cho *nhận xét* về phương thức bán hàng mới, với mức ý nghĩa 5%?

11

Hướng dẫn :

Ta có $n = 144$, $\sum n_i x_i = 6602$, $\sum n_i x_i^2 = 321706$,
 $\bar{x} = 45,8472$, $s^2 = 133,0275$, $s = 11,5338$

- 1) Gọi μ là doanh số bán trung bình trong 1 ngày của siêu thị
 $\gamma = 95\% \Rightarrow t_{\alpha/2} = 1,96$
 $\varepsilon = \frac{t_{\alpha/2} s}{\sqrt{n}} = (1,96).(11,5338) / \sqrt{144} = 1,8839$
Khoảng tin cậy $(43,9633 < \mu < 47,7311)$
- 2) Tỷ lệ những ngày bán đắt hàng theo mẫu:
 $f = (12+10+6)/ 144 = 0,1944$
Gọi p là tỷ lệ những ngày bán đắt hàng của siêu thị
Vậy tỷ lệ những ngày bán đắt hàng của siêu thị là 19,44%

12

3) $\gamma = 99\% \Rightarrow t_{\alpha/2} = 2,58$

$$\varepsilon = \frac{t_{\alpha/2} \sqrt{f(1-f)}}{\sqrt{n}} = (2,58) \sqrt{(0,1944)(1-0,1944) / \sqrt{144}}$$

= 0,0851

KTC (0,1093 ; 0,2795) hay 10,93% <p < 27,95%

4) Lập bảng sau:

Doanh số	60	65	70
Số ngày	12	10	6

Ta có $n = 28$, $\sum n_i x_i = 1790$, $\sum n_i x_i^2 = 114850$,
 $\bar{x} = 1790/28 = 63,9286$; $s = 3,9335$
 $s^2 = (114850 - 28(63,9286)^2) / 27 = 15,4724$

13

4) Gọi μ' là doanh số bán TB trong 1 ngày bán đất hàng của siêu thị

$$t_{0,025}(27) = 2,0518$$
$$\varepsilon = (2,0518).(3,9335) / \sqrt{28} = 1,5254$$

Khoảng tin cậy $(62,4032 < \mu' < 65,454)$

5) Lập giả thiết $H_0 : \mu = 35$; $H_1 : \mu \neq 35$

μ : doanh số bán trung bình *hiện nay*

$\mu_0 = 35$: doanh số bán trung bình *trước đây*

$\alpha = 5\% \Rightarrow t_{\alpha/2} = 1,96$

$$t = (45,8472 - 35) \sqrt{144} / 11,5338 = 11,2856$$

Ta có $|t| > t_{\alpha/2}$: bác bỏ H_0

Vậy phương thức bán hàng mới tốt hơn (do $\bar{x} = 45,8472 > \mu_0 = 35$)

14

BÀI 3

Cho X là năng suất lúa ở 1 khu vực (đơn vị tính tạ/ha). Điều tra ở 1 số thửa ruộng ta có:

X	30 – 35	35 – 40	40 – 45	45 – 50	50 – 55
N	6	18	28	40	16

N: số thửa ruộng.

Ví dụ: có 18 thửa ruộng, năng suất của mỗi thửa là (35 – 40) tạ/ha

15

1) Hãy ước lượng năng suất lúa trung bình của toàn vùng, với độ tin cậy 96%?

2) Những thửa ruộng đạt năng suất trên 45 tạ/ha là những thửa ruộng đạt năng suất cao. Hãy ước lượng tỷ lệ những thửa ruộng đạt năng suất cao của vùng này, với độ tin cậy 95%?

3) Nếu muốn ước lượng năng suất lúa trung bình của toàn vùng đạt được độ chính xác là 1,4 tạ /ha thì độ tin cậy là bao nhiêu?

4) Người ta nhận định: tỷ lệ những thửa ruộng đạt năng suất cao *lớn hơn* 50%. Theo bạn nhận định đó đúng không, $\alpha = 5\%$?



16

5) Nếu muốn ước lượng năng suất lúa trung bình với độ chính xác 0,5 tạ/ha và độ tin cậy 99% dựa trên mẫu đã cho thì phải điều tra thêm bao nhiêu thửa ruộng nữa?

Hướng dẫn :

1) Lập bảng sau:

X	32,5	37,5	42,5	47,5	52,5
N	6	18	28	40	16

$n = 108, \sum n_i x_i = 4800, \sum n_i x_i^2 = 216575$
 $\bar{x} = 44,4444, s^2 = 30,2999, s = 5,5045$

17

Hướng dẫn :

- 1) $\gamma = 96\% \Rightarrow t_{\alpha/2} = 2,05$
 $\varepsilon = (2,05) (5,5045) / \sqrt{108} = 1,0879$
Khoảng tin cậy (43,3565 ; 45,5323)
- 2) Tỷ lệ thửa ruộng đạt năng suất cao theo mẫu:
 $f = (40+16)/108 = 0,5185$
Gọi p là tỷ lệ thửa ruộng đạt năng suất cao của vùng
 $\varepsilon = (1,96) \sqrt{(0,5185)(1-0,5185)} / \sqrt{108} = 0,0942$
Khoảng tin cậy (0,4243 < p < 0,6127)
- 3) $t_{\alpha/2} = \frac{\varepsilon \sqrt{n}}{s} = (1,4) \sqrt{108} / 5,5045 = 2,64$
 $\Rightarrow \phi(t_{\alpha/2}) = 0,4959 \Rightarrow \gamma = 2\phi(t_{\alpha/2}) = 2(0,4959)$
 $= 0,9918 = 99,18\%$

18

- 4) Lập giả thiết $H_0: p = 0,5 ; H_1: p > 0,5$
 p : tỷ lệ thửa ruộng đạt năng suất cao *thực tế*
 $p_0 = 0,5$: tỷ lệ thửa ruộng đạt năng suất cao *theo nhận định*
 $\alpha = 5\% \Rightarrow t_\alpha = 1,65$
 $t = \frac{(f-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} = (0,5185-0,5)\sqrt{108} / \sqrt{(0,5)(1-0,5)} = 0,3845$
Ta có $t < t_\alpha$: chấp nhận H_0 .
Vậy nhận định trên sai
 $n = \left(\frac{t_{\alpha/2} s}{\varepsilon}\right)^2 = (2,58 \times 5,5045)^2 / (0,5)^2 = 806,7429 \approx 807$
Vậy cần điều tra thêm $807-108 = 699$ thửa ruộng nữa

19

- **BÀI 4.**
- Để thăm dò nhu cầu về một loại hàng ở một TP, người ta đã tiến hành phỏng vấn 500 hộ gia đình thì thấy có 200 hộ có nhu cầu về loại hàng này.
- 1) Hãy ước lượng *số hộ* gia đình có nhu cầu về mặt hàng này ở thành phố, với độ tin cậy 96%? (Biết tổng số hộ gia đình của thành phố là 20 000 hộ)
- 2) Nếu muốn ước lượng tỷ lệ hộ gia đình có nhu cầu về mặt hàng này đạt được độ chính xác 4% thì độ tin cậy là bao nhiêu?
- 3) Nếu muốn ước lượng tỷ lệ hộ gia đình có nhu cầu về mặt hàng này đạt được độ chính xác 5% và độ tin cậy 99% thì cần điều tra bao nhiêu hộ?

20

Hướng dẫn :

- 1) Tỷ lệ hộ có nhu cầu về loại hàng này theo mẫu $f = 200/500 = 0,4$
Gọi p là tỷ lệ hộ có nhu cầu về loại hàng này ở TP
 $\varepsilon = 2,05 \times \sqrt{0,4(1-0,4)} / \sqrt{500} = 0,045$
Vậy $0,355 < p < 0,445$
Số hộ gia đình có nhu cầu về loại hàng này ở TP là:
 $0,355 \times 20\,000 < \text{Số hộ} < 0,445 \times 20\,000$
- 2) $t_{\alpha/2} = 0,04 \times \sqrt{500} / \sqrt{0,4(1-0,4)} = 1,83$
 $\Rightarrow \phi(t_{\alpha/2}) = 0,4664 \Rightarrow \gamma = 2\phi(t_{\alpha/2})$
- 3) $n = (2,58)^2(0,4)(1-0,4) / (0,05)^2 = 639,014 \approx 640$ hộ

21

BÀI 5: Kết quả quan sát về hàm lượng Vitamin C của một loại trái cây cho ở bảng sau:

Hàm lượng Vitamin C (%)	3 - 7	8 - 10	11 - 13	14 - 16	17 - 19	20 - 24
Số trái	5	10	20	35	25	5

- 1) Nếu nói hàm lượng Vitamin C trung bình trong 1 trái *nhỏ hơn* 16% thì tin được không, với $\alpha = 4\%$?
- 2) Qui ước những trái có hàm lượng Vitamin C từ 17% trở lên là trái loại I. Nếu muốn độ chính xác khi ước lượng *trung bình* hàm lượng vitamin C là 0,5% và độ chính xác khi ước lượng *tỷ lệ* trái loại I là 5%, với cùng độ tin cậy 95% thì cần mẫu có kích thước tối thiểu là bao nhiêu?

22

Hướng dẫn:

- 1) $H_0: \mu = 16$; $H_1: \mu < 16$
 μ : hàm lượng Vitamin C TB trong 1 trái *thực tế*
 $\mu_0 = 16$: hàm lượng Vitamin C TB trong 1 trái *theo nhận định*
 $n = 100$, $\bar{x} = 14,4$, $s = 3,8586$
 $t = \frac{(14,4 - 16) \cdot \sqrt{100}}{3,8586} = -4,1466$
 $\alpha = 4\% \rightarrow t_\alpha = 1,75$
Ta có $t = -4,1466 < -t_\alpha = -1,75$: bác bỏ H_0
Vậy nhận định trên tin được.

23

2) Ước lượng trung bình:

$$n_1 = \left(\frac{1,96 * 3,8586}{0,5} \right)^2 = 229$$

Ước lượng tỷ lệ:

Tỷ lệ mẫu $f = (25+5)/100 = 0,3$

$$n_2 = \left(\frac{1,96}{0,05} \right)^2 * 0,3 * 0,7 = 323$$

Cỡ mẫu $n = \max\{229, 323\} = 323$

Vậy cần khảo sát tối thiểu 323 trái.

24

BÀI 6. (MẪU 2 CHIỀU)

X(%) và Y(kg/mm²) là 2 chỉ tiêu chất lượng của 1 loại sản phẩm. Điều tra ở 1 số sản phẩm về (X,Y) ta có kết quả:

(2,5)	(8,15)	(4,15)	(4,10)	(2,10)	(8,25)
(2,5)	(6,10)	(4,10)	(8,20)	(6,10)	(8,15)
(6,10)	(6,15)	(4,15)	(6,15)	(8,20)	(6,15)
(6,20)	(6,10)	(6,20)	(6,15)	(6,25)	(8,20)
(6,15)	(6,20)	(8,15)	(6,15)	(8,25)	(8,15)

25

- 1) Ước lượng trung bình chỉ tiêu Y, với độ tin cậy 98%?
- 2) Có tài liệu nói: Trung bình chỉ tiêu X là 6,5% .
Cho *nhận xét* với mức ý nghĩa 5%?
- 3) Quy ước: Sản phẩm có chỉ tiêu $Y \leq 15(\text{kg/mm}^2)$ và $X \leq 6(\%)$ là sản phẩm loại A. Ước lượng trung bình chỉ tiêu Y của sản phẩm loại A, với độ tin cậy 99%?
(Biết rằng chỉ tiêu Y có quy luật phân phối chuẩn)
- 4) Ước lượng tỷ lệ sản phẩm loại A, với $\alpha = 5\%$?

26

Hướng dẫn :

1) Lập bảng sau:

X \ Y	Y				
	5	10	15	20	25
2	2	1			
4		2	2		
6		4	6	3	1
8			4	3	2

Bảng tần số quan sát của X và Y như sau:

X	2	4	6	8
n	3	4	14	9

Y	5	10	15	20	25
n	2	7	12	6	3

27

- 1) $n = 30$, $\sum n_y y = 455$, $\sum n_y y^2 = 7725$, $\bar{y} = 15,1667$
 $s_y^2 = \frac{1}{n-1}[\sum n_y y^2 - n(\bar{y})^2] = 28,4185$, $s_y = 5,3309$
 $\gamma = 98\% \Rightarrow t_{\alpha/2} = 2,33$
 $\varepsilon = \frac{t_{\alpha/2} s_y}{\sqrt{n}} = 2,33 \times 5,3309 / \sqrt{30} = 2,2639$
Vậy khoảng tin cậy $\bar{y} \pm \varepsilon$ là (12,9028 ; 17,4306)
- 2) $n = 30$, $\sum n_x x = 178$, $\sum n_x x^2 = 1156$, $\bar{x} = 5,9333$
 $s_x^2 = \frac{1}{n-1}[\sum n_x x^2 - n(\bar{x})^2] = 3,4441$, $s_x = 1,8558$

28

2) Lập giả thiết $H_0 : \mu = 6,5$; $H_1 : \mu \neq 6,5$
 μ : trung bình chỉ tiêu X thực tế
 $\mu_0 = 6,5$: trung bình chỉ tiêu X theo nhận định
 $\alpha = 5\% \Rightarrow t_{\alpha/2} = 1,96$
 $t = (5,9333 - 6,5) \sqrt{30} / 1,8558 = -1,6726$
 $|t| < t_{\alpha/2}$: chấp nhận H_0
3) Lập bảng sau:

X \ Y	Y		
	5	10	15
2	2	1	
4		2	2
6		4	6

29

3) $n = 17$, $\sum n_y y = 200$, $\sum n_y y^2 = 2550$,
 $\bar{y} = 11,7647$, $s_y^2 = 12,3163$, $s_y = 3,5095$
 $\gamma = 99\% \rightarrow \alpha = 1\% \Rightarrow t_{0,005}(16) = 2,921$
 $\varepsilon = 2,921 \times 3,5095 / \sqrt{17} = 2,4863$
Vậy khoảng tin cậy $\bar{y} \pm \varepsilon$ là (9,2784 ; 14,251)
4) Tỷ lệ sản phẩm loại A theo mẫu: $f = 17/30 = 0,5667$
 $\alpha = 5\% \Rightarrow t_{\alpha/2} = 1,96$
 $\varepsilon = 1,96 \times \sqrt{0,5667(1-0,5667) / 30} = 0,1773$
Khoảng tin cậy $f \pm \varepsilon$ là (0,3894 ; 0,744)

30

QUY ƯỚC: QUYỀN (*) LÀ QUYỀN:
BÀI TẬP XSTK. THS. LÊ KHÁNH LUẬN & GVC.
NGUYỄN THANH SƠN & THS. PHẠM TRÍ CAO,
NXB ĐHQG TP HCM 2013.

- Các bạn thân mến!
- Bạn đã “*thưởng thức*” xong “*món*” XSTK.
Bạn cảm thấy còn “*thòm thềm, chưa đã*” ư!
- Vậy còn chần chờ gì nữa!
Hãy đọc ngay quyền (*) !
- **MỘT PHÚT DÀNH CHO QUẢNG CÁO !!!**

31

Mời ghé thăm trang web:

- ❖ <https://sites.google.com/a/ueh.edu.vn/phamtricao/>
- ❖ <https://sites.google.com/site/phamtricao/>

32