

## ÔN THI CUỐI KỲ TKMT & UD 2023

Ngày 01/12/2023

Cấu trúc: 3 câu, tự luận, cho dùng tài liệu (không giới hạn); thời gian: 60 – 75 phút.

Cần mang theo bảng tra z-scores + máy tính bỏ túi (chức năng thống kê).

- Câu 1: ước lượng.
- Câu 2: kiểm định.
- Câu 3: hồi quy.

1) **Vấn đề ước lượng:** từ một mẫu  $\rightarrow$  có các tính toán để ước lượng được các tham số quần thể.

- Ước lượng cho  $\mu$  (trung bình).
- Ước lượng cho  $p$  (tỷ lệ).

a. **Từ trung bình mẫu:**  $\bar{x} \rightarrow$  ta ước lượng trung bình của quần thể  $\bar{x} - E < \mu < \bar{x} + E$  với  $E$  = biên độ lỗi. Biên độ lỗi sẽ tính được từ độ tin cậy và độ lệch chuẩn cho trước.

**độ tin cậy:**  $(1-\alpha).100\% \Leftrightarrow \text{mức ý nghĩa} = \alpha / 2$ .

90%  $\rightarrow$  mức ý nghĩa là  $\alpha / 2 = 0.05 \rightarrow z_{\alpha/2} = 1.65$ ,

95%  $\rightarrow$  mức ý nghĩa là  $\alpha / 2 = 0.025 \rightarrow z_{\alpha/2} = 1.96$ ,

99%  $\rightarrow$  mức ý nghĩa là  $\alpha / 2 = 0.005 \rightarrow z_{\alpha/2} = 2.57$ .

Bỏ qua sai số, ta dùng z cho mọi tình huống thay vì dùng phân phối t.

Có công thức tính  $E = z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$  trong đó  $s$  = độ lệch chuẩn của mẫu. Như thế, từ một bảng dữ liệu mẫu, ta có thể thu được:  $n$ ,  $s$ ,  $\bar{x}$ .

VD: cho bảng tần suất:

Lượng sữa (lít/ngày)	7	9	11	13	15
Số con bò	10	32	80	39	8

Với độ tin cậy 95%, có thể nói lượng sữa trung bình vắt được từ mỗi con bò là ...

Khi dùng máy tính bỏ túi, ta bật chế độ Frequency = On  $\rightarrow$  tính các giá trị kia:

- tổng số bò  $n = 10 + 32 + \dots + 8 = \dots$
- trung bình  $\bar{x} = (7.10 + 9.32 + \dots + 15.8) / n = \dots$
- độ lệch chuẩn  $s$  (vì đây là mẫu nên  $\sigma_{n-1}$ ) = ...

Đôi khi, bảng tần suất sẽ cho ở dạng khoảng giá trị, thay vì 7, 9, 11, 13, 15 như trên, ta sẽ có bảng:

[6 - 8], [8 - 10], [11 - 13], [14 - 16]

→ ta chỉ cần lấy giá trị trung bình của 2 đầu mút làm giá trị đại diện cho khoảng và quy mọi thao tác tính toán về dạng như trên.

### b. Ước lượng tỷ lệ.

Có  $p$  là tỷ lệ mẫu, ta cần ước lượng tỷ lệ quần thể:

$$p - E < p < p + E.$$

Số  $p$  sẽ tính được trước theo thông tin đề cho, và  $E = z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}$ . Các thao tác tương tự trên.

Lượng sữa (lít/ngày)	7	9	11	13	15
Số con bò	10	32	80	39	8

Với độ tin cậy 95%, có thể nói lượng sữa trung bình vắt được trong ngày là...

VD: ước lượng tỷ lệ số bò cho < 10 lít sữa / ngày với độ tin cậy là 99%.

Số bò thỏa mãn = 10+32 = 42 nên tỷ lệ  $p = 42 / 169 = 0.25$ .

Ta có độ tin cậy  $(1-\alpha).100\% = 99\%$  nên  $z_{\alpha/2} = 2.57$  nên có biên độ lỗi = 2,57.  $\sqrt{\frac{0.25 \cdot (1-0.25)}{169}}$

từ đó thay vào có được khoảng ước lượng của tỷ lệ.

### 2) Kiểm định giả thuyết:

#### a) Kiểm định trung bình $\mu$ .

Giả thuyết gốc  $H_0: \mu = \mu_0$  cho trước.

Giả thuyết thay thế:  $\mu \neq \mu_0$  (kiểm định 2 phía) hoặc  $\mu > \mu_0$  /  $\mu < \mu_0$  (kiểm định 1 phía).

Dựa theo thông tin của đề bài (khác / lớn / bé) để xác định cho đúng.

Quy trình thực hiện: tính ra giá trị  $z$  và tính ra giá trị  $p$ -value rồi so sánh.

Số  $p$ -value chính là các số 1.65, 1.96 hay 2.57 được suy ra từ mức ý nghĩa / độ tin cậy.

Ta tính giá trị  $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$  trong đó  $\mu$  được cho trước,  $\bar{x}$  sẽ được cho trước;  $\sigma$  là độ lệch chuẩn (nếu không cho trước thì dùng luôn độ lệch chuẩn của mẫu là  $s$ ).

VD:

Vì dịch bệnh đã giảm nên nhà trường thực hiện việc học trực tiếp đối với học sinh toàn trường.

a/ Nhà trường dự toán chi phí phòng bệnh mỗi lớp tối đa 450 ngàn/tuần, độ lệch chuẩn 45 ngàn. Nhưng khi xem xét số tiền đã chi cho 40 lớp thì chi phí phòng dịch trung bình mỗi lớp thực tế là 475 ngàn/tuần. Hãy kiểm định dự toán của nhà trường có đúng hay không với độ tin cậy 99%?

a) Kiểm định giả thuyết  $H_0: \mu \leq 450$ ;

Giả thuyết thay thế  $H_1: \mu > 450 \rightarrow$  kiểm định 1 phía.

Ta có:  $\mu = 450, \sigma = 45, n = 40, \bar{x} = 475 \rightarrow z = \frac{475 - 450}{45 / \sqrt{40}} = 3,51$ .

Độ tin cậy = 99%  $\rightarrow$  p-value = 2.57, so sánh: z và p-value:

- $z > \text{p-value}$ : bác bỏ  $H_0$ .
- $z \leq \text{p-value}$ : chấp nhận  $H_0$ .

Ở đây, ta có  $z = 3.51 > 2.57 = \text{p-value}$  nên bác bỏ  $H_0 \Leftrightarrow$  dự đoán của nhà trường không còn đúng với độ tin là 99%.

b) Kiểm định tỷ lệ  $p$ .

Khối	1	2	3	4	5
Số PH	10	35	60	30	15

Nhà trường cho rằng tổng số số phụ huynh các khối 3, 4, 5 đồng ý chiếm 80% phụ huynh của trường đồng ý. Với mức ý nghĩa 2,5% hãy kiểm định ý kiến của nhà trường đúng không?

Ta có:  $p.n \geq 5$  và  $(1-p).n \geq 5$  thì ta áp dụng được phân phối z vào việc kiểm định.

Giả thuyết gốc:  $p = 0.8$ .

Giả thuyết thay thế:  $p \neq 0.8$  (kiểm định 2 phía).

Ta có:  $n = 10 + 35 + 60 + 30 + 15 = 150$  và  $p^{\wedge} = (60 + 30 + 15) / 150 = 105 / 150 = 0.7$ .

Mức ý nghĩa là 2.5% = 0.025  $\rightarrow z_{\alpha/2} = 1.96 \rightarrow \text{p-value} = 2 * 1.96 = 3.92$ .

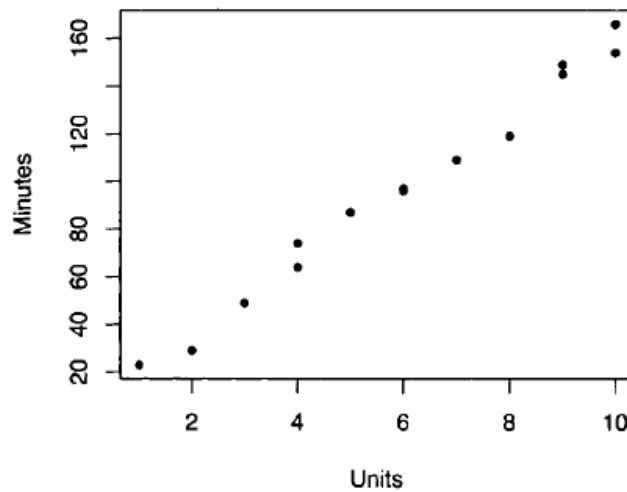
Công thức tính  $z = \frac{p - p^{\wedge}}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.7 - 0.8}{\sqrt{\frac{0.8(1-0.8)}{150}}} \rightarrow$  so sánh z và p-value là được.

### 3) Hồi quy tuyến tính.

Ta sẽ được cho trước 1 bảng dữ liệu có 2 chiều (X, Y), ta cần tính:

- Hệ số tương quan Pearson = hệ số tương quan (r).
- Hiệp phương sai;
- Hệ số a, b của PT hồi quy: shift 1  $\rightarrow$  các giá trị thống kê có liên quan (Reg).

Vẽ sơ đồ biểu diễn cho dữ liệu: chọn trục hoành, trục tung, biểu diễn các giá trị lên đó thành các điểm.



Số điếu/ngày	0	6	20	2	5	6	20	0	4	3
Tuổi thọ	90	68	55	70	60	65	50	82	75	63

PT đường hồi quy:  $y = A + Bx = 76.5 - 1.32x$ .

Tính được hệ số tương quan Pearson là  $-0.804 < 0$ , các tính chất:

- + Hệ số tương quan thuộc  $[-1;1]$ .
- +  $r = 0 \rightarrow$  2 đại lượng hầu như không có liên quan;
- +  $r > 0$  và càng gần 1: càng liên quan mạnh với nhau và giá trị này có sự củng cố, bổ sung cho giá trị kia.
- +  $r < 0$  và càng gần -1: càng liên quan mạnh với nhau và giá trị này có sự đối lập với giá trị kia.

Do đó, ở BT này, ta thấy hệ số tương quan là  $< 0$  và khá gần -1 nên việc hút thuốc sẽ có ảnh hưởng nhiều đến tuổi thọ, và có xu hướng đối lập nhau.

Để tính hiệp phương sai  $=$  hệ số tương quan \* độ lệch chuẩn của x \* độ lệch chuẩn của y  
 $= r * s_x * s_y$

(cần làm thế này vì trong máy tính casio không tính được rõ đại lượng này).