

## CÁC PHƯƠNG PHÁP ĐỊNH LƯỢNG 1

### LỜI GIẢI GỢI Ý BÀI TẬP 4

Ngày phát: 23/11/2023

Hạn nộp: 8h20, 28/11/2023

Bài làm được yêu cầu chỉ nộp bản điện tử trên Microsoft Teams

---

**Câu 1 (30đ):** Hãy trả lời những câu hỏi sau đây:

a. Giải thích  $H_0$ ,  $H_1$ ,  $\alpha$  trong cách tiếp cận kiểm định giả thuyết thống kê.

Học viên tự giải thích theo cách hiểu của mình dựa theo lý thuyết đã học.

b. Với mức ý nghĩa 5% cho kiểm định hai phía, bạn sẽ quyết định như thế nào khi giá trị  $Z = 2.21$ .

Với  $\alpha = 5\% \rightarrow Z_{\frac{\alpha}{2}} = NORM.S.INV(0.975,1) = 1.96$

$Z = 2.21 > 1.96$  nên chúng ta có thể bác bỏ giả thuyết  $H_0$  với mức ý nghĩa  $\alpha = 5\%$ .

c. Tính giá trị  $p$ -value cho giá trị  $Z$  trên và giải thích ý nghĩa của  $p$ -value.

$P(Z > 2.21) = 1 - NORM.S.DIST(2.21,1) = 0.01355$

$p\text{-value} = P(Z < -2.21) + P(Z > 2.21) = 2 \times 0.01355 = 0.0271$

Ý nghĩa  $p$ -value: Xác suất chúng ta mắc sai lầm khi bác bỏ  $H_0$  (khi  $H_0$  đúng) và chấp nhận  $H_1$  là 2.71%

**Câu 2 (45đ):** Điểm thi tốt nghiệp THPT trung bình môn Toán của tỉnh A năm 2022 là 6.7. Kết quả cuối cùng điểm thi năm 2023 chưa biết, nhưng biết được một mẫu ngẫu nhiên 30 em có trung bình mẫu là 6.3 và độ lệch chuẩn mẫu là 1.4. Tỉnh A muốn tìm hiểu xem từ mẫu này có thể suy luận ra điểm thi tốt nghiệp THPT môn Toán năm 2023 có thấp hơn năm 2022 hay không. Hãy thực hiện các bước sau:

a. Hãy viết  $H_0$  và  $H_1$  cho kiểm định một phía.

$H_0$ : Điểm trung bình môn Toán năm 2023 không thấp hơn năm 2022:  $\mu_{2023} \geq 6.7$

$H_1$ : Điểm trung bình môn Toán năm 2023 thấp hơn năm 2022:  $\mu_{2023} < 6.7$

b. Tính giá trị kiểm định, bạn sử dụng phân phối nào cho kiểm định này?

Chúng ta sử dụng phân phối  $t$  để thực hiện kiểm định này vì chưa biết giá trị phương sai của tổng thể

c. Với mức ý nghĩa 5%, hãy thực hiện kiểm định và đưa ra kết luận.

Từ dữ kiện của đề bài, trị kiểm định  $t$  là:

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{6.3 - 6.7}{1.4/\sqrt{30}} = -1.5649$$

Với  $\alpha = 5\% \rightarrow t_{\alpha,df} = t_{0.05,29} = T.INV(0.05,29) = -1.6991$

Vì  $t_{0.05,29} < t$  nên chưa đủ cơ sở để bác bỏ  $H_0$ , chấp nhận  $H_1$ , nghĩa là chưa đủ cơ sở để kết luận rằng điểm trung bình môn Toán năm 2023 thấp hơn năm 2022 với độ tin cậy 95%

d. Với mức ý nghĩa 1%, kết quả thay đổi thế nào? Tính  $p$ -value cho kiểm định này.

$$t = -1.5649 \rightarrow p\text{-value} = T.DIST(-1.5649, 29, TRUE) = 0.06422$$

Với  $p\text{-value} = 6.42\% > 1\%$  nên chưa đủ cơ sở để bác bỏ  $H_0$ , chấp nhận  $H_1$ , kết quả không thay đổi so với câu c

e(\*). Tính khoảng tin cậy 99% của trung bình tổng thể tỉnh A năm 2023. Bạn có liên hệ gì với kết quả câu d hay không.

$$\text{Với } \alpha = 1\% \rightarrow t_{\frac{\alpha}{2}, df} = t_{0.005, 29} = T.INV(0.995, 29) = 2.75638$$

Khoảng tin cậy 99% của trung bình tổng thể kết quả thi của tỉnh A năm 2023 là:

$$\mu \pm t_{0.005, 29} \times \sqrt{\frac{\sigma^2}{n}} = 6.3 \pm 2.75638 \times \sqrt{\frac{1.4^2}{30}} = [5.5954; 7.0045] \text{ (điểm)}$$

Như vậy, khi thực hiện lấy mẫu nhiều lần thì 99% giá trị điểm trung bình tổng thể của tỉnh A năm 2023 sẽ nằm trong khoảng ước lượng trên. Khoảng ước lượng này chứa giá trị trung bình  $\mu = 6.7$  của mẫu quan sát nên chưa thể khẳng định được điểm trung bình môn Toán năm 2023 ở tỉnh A là thấp hơn năm 2022. Điều này có sự tương đồng với kết luận ở câu d

f. Nếu đây là kiểm định hai phía, các kết quả kiểm định có thay đổi ở mức 5% và 1% hay không?

$$\text{Nếu kiểm định 2 phía: } p\text{-value} = 0.06422 \times 2 = 0.12845$$

Lúc này  $p$ -value sẽ càng lớn hơn các mức  $\alpha = 5\%$  hay  $\alpha = 1\%$ , kết quả không có sự thay đổi

g. Tỉnh B có một mẫu gồm 50 học sinh có trung bình và độ lệch chuẩn mẫu lần lượt là 6.8 và 1.6. Hãy thực hiện kiểm định để so sánh điểm trung bình tổng thể năm 2023 của hai tỉnh từ hai mẫu quan sát.

$$H_0: \text{Điểm trung bình môn Toán tỉnh A bằng tỉnh B: } \mu_A = \mu_B$$

$$H_1: \text{Điểm trung bình môn Toán tỉnh A khác tỉnh B: } \mu_A \neq \mu_B$$

$$\text{Trị điểm định } t = \frac{(n+m-2)^{\frac{1}{2}}(\bar{A}-\bar{B})}{\left(\frac{1}{n} + \frac{1}{m}\right)^{\frac{1}{2}}((n-1)S_A^2 + (m-1)S_B^2)^{\frac{1}{2}}} = \frac{(30+50-2)^{\frac{1}{2}}(6.3-6.8)}{\left(\frac{1}{30} + \frac{1}{50}\right)^{\frac{1}{2}}((30-1) \times 1.4^2 + (50-1) \times 1.6^2)^{\frac{1}{2}}} = -1.416$$

Giả sử mức ý nghĩa  $\alpha = 5\%$

$$\rightarrow t_{\frac{\alpha}{2}, df} = t_{0.025, 78} = T.INV(0.025, 78) = -1.9908$$

Vì  $t_{0.025, 78} < t$  nên chưa đủ cơ sở để kết luận rằng có sự khác biệt về điểm trung bình tổng thể năm 2023 của hai tỉnh này từ mẫu quan sát với độ tin cậy 95%

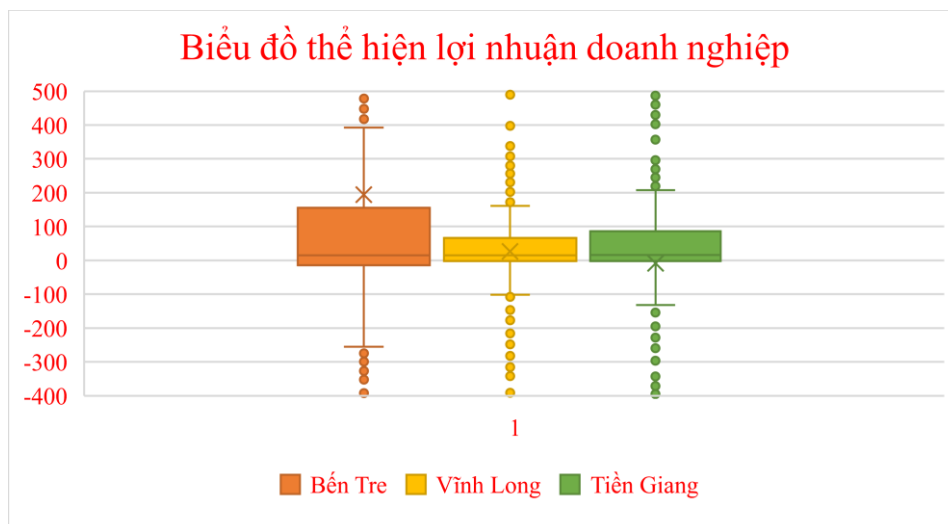
**Câu 3 (25đ):** Lợi nhuận kế toán trước thuế của một số doanh nghiệp tại 3 tỉnh Tiền Giang, Bến Tre và Trà Vinh được cung cấp ở file Excel *Assignment 4.xlsx* (gồm 2 sheet Data và Description). Hãy trả lời những câu hỏi sau:

a. Vẽ một bảng thể hiện các tham số: số lượng quan sát, trung bình, phương sai và độ lệch chuẩn lợi nhuận cho ba tỉnh.

Học viên sử dụng công cụ Data Analysis hoặc các hàm thông dụng trong Excel để thực hiện thống kê theo yêu cầu của đề bài

Tham số	Tiền Giang	Bến Tre	Vĩnh Long
Số lượng quan sát	433	457	451
Trung bình	-8.941	193.852	26.755
Phương sai	1646303	36242752	3544443
Độ lệch chuẩn	1283.083	6020.195	595.351

b. Vẽ biểu đồ hộp (box plot) cho lợi nhuận doanh nghiệp của ba tỉnh này. Bạn có nhận xét gì từ biểu đồ này không?



Học viên tự đưa ra nhận định phù hợp liên quan đến độ dãn trải (range), phân phối, các giá trị trung bình hoặc tứ phân vị về lợi nhuận doanh nghiệp của ba tỉnh

c. Khi thực hiện phân tích ANOVA kiểm định giả thuyết – xem lợi nhuận trung bình tổng thể của các doanh nghiệp giữa tỉnh này có bằng nhau hay không – thì  $H_0$  và  $H_1$  là gì?

$H_0$ : Không có sự khác biệt về lợi nhuận trung bình tổng thể của các doanh nghiệp giữa các tỉnh:  $\mu_{Tiền\ Giang} = \mu_{Bến\ Tre} = \mu_{Vĩnh\ Long}$

$H_1$ : Có sự khác biệt về lợi nhuận trung bình tổng thể của các doanh nghiệp giữa các tỉnh:  $\mu_{Tiền\ Giang} \neq \mu_{Bến\ Tre} \neq \mu_{Vĩnh\ Long}$

d. Từ kết quả phân tích ANOVA, bạn có kết luận thế nào về trung bình (tổng thể) lợi nhuận doanh nghiệp của ba tỉnh này.

Học viên sử dụng công cụ Data Analysis để thực hiện kiểm định ANOVA, ta có kết quả với mức ý nghĩa  $\alpha = 5\%$ :

**ANOVA**

Source of Variation	SS	df	MS	F	P-value	F critical
Between Groups	10545495.5	2	5272747.75	0.4055	0.6667	3.0024
Within Groups	17397396983	1338	13002538.9			
Total	17407942478	1340				

Vì  $p\text{-value} > \alpha$  nên ta chưa đủ bằng chứng để khẳng định rằng có sự khác biệt về lợi nhuận trung bình tổng thể của các doanh nghiệp giữa ba tỉnh với độ tin cậy 95%