ĐẠI HỌC QUỐC GIA HÀ NỘI

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

KHOA TOÁN - CƠ - TIN HỌC

o0o

THỐNG KÊ ỨNG DỤNG

Mã lớp học phần: **MAT2406**

Sinh viên: **LƯU VĂN VIỆT**

Lớp: **A2K65 TOÁN - TIN**

Hà Nội, tháng 6 năm 2022

Mục lục

1 ƯỚC LƯỢNG THAM SỐ **5**

I Khoảng tin cậy cho kỳ vọng 5

1 *Phương sai đã biết* 5

2 *Phương sai chưa biết,* $n > 30$ 5

3 *Phương sai chưa biết,* $n < 30$ 5

II Khoảng tin cậy cho tỉ lệ 6

III Khoảng tin cậy cho sự khác biệt giữa 2 giá trị trung bình với mẫu độc lập. 7

1 Phương sai đã biết 7

2 Phương sai chưa biết, $n_i > 30$ 7

3 Phương sai chưa biết, $n_i < 30$ 7

IV Khoảng tin cậy cho phương sai 7

2 Kiểm định giả thiết thống kê **9**

I Kiểm định giả thiết về giá trị trung bình 9

II Kiểm định cho tỉ lệ 10

III Kiểm định về giá trị của nhiều tỉ lệ 10

3 Bài toán so sánh **13**

I So sánh hai giá trị trung bình 13

1 *Hai mẫu độc lập* 13

2 *Hai mẫu phụ thuộc* 14

II Tiêu chuẩn phi tham số 15

III So sánh hai tỉ lệ 16

4 Phân tích phương sai **19**

I Phân tích phương sai một nhân tố 19

II Phân tích phương sai hai nhân tố 22

5 Phân tích tương quan và hồi quy 23

I Phân tích tương quan tuyến tính 23

II Kiểm tra tính độc lập 23

III Phân tích tương quan phi tuyến 25

IV Phân tích hồi quy tuyến tính 27

6 Các phân bố thường gặp 31

Chương 1

ƯỚC LƯỢNG THAM SỐ

I Khoảng tin cậy cho kỳ vọng

1 *Phương sai đã biết*

Giả sử $X \sim N(\mu, \sigma^2)$ trong đó σ^2 đã biết. Với độ tin cậy $1-\alpha$ đã cho, giả sử $z(\alpha)$ là giá trị thỏa mãn $\Phi(z(\alpha)) = 1-\alpha$. Khi đó khoảng tin cậy cho EX là:

$$\left(\overline{X} - z\left(\frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}}; \overline{X} + z\left(\frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}}\right)$$

Một số giá trị thông dụng của $z\left(\frac{\alpha}{2}\right)$:

Nếu $\alpha = 0,1$ thì $z(0,05) = 1,645$

Nếu $\alpha = 0,05$ thì $z(0,025) = 1,96$

Nếu $\alpha = 0,02$ thì $z(0,01) = 2,33$

Nếu $\alpha = 0,01$ thì $z(0,005) = 2,58$

2 *Phương sai chưa biết,* $n > 30$

Khoảng tin cậy $1-\alpha$ của EX là:

$$\left(\overline{X} - z\left(\frac{\alpha}{2}\right)\frac{s}{\sqrt{n}}; \overline{X} + z\left(\frac{\alpha}{2}\right)\frac{s}{\sqrt{n}}\right)$$

3 *Phương sai chưa biết,* $n < 30$

Khoảng tin cậy $1-\alpha$ của EX là:

$$\left(\overline{X} \pm t_{\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}}\right)$$

Ví dụ 1: Tìm khoảng tin cậy cho chiều cao trung bình của sinh viên dựa trên một mẫu có kích thước $n = 36$ với trung bình mẫu $\overline{X} = 66$. Giả sử độ lệch tiêu chuẩn của chiều cao là 3.

Ta có $\sigma = 3, n = 36, \alpha = 1 - 0,95 = 0,05$ nên khoảng tin cậy 95% cho chiều cao trung bình là:

$$\overline{X} \pm z\left(\frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}} = 66 \pm 1,96\frac{3}{\sqrt{36}} = 66 \pm 0,98$$

Ví dụ 2: Để ước lượng chiều cao trung bình của thanh niên trong một vùng A nào đó, một mẫu ngẫu nhiên gồm 16 thanh niên được chọn. Chiều cao của các thanh niên này đo được như sau:

172, 173, 173, 174, 174, 175, 176, 166, 166, 167, 165, 173, 171, 170, 171, 170

Hãy tìm khoảng tin cậy 95% cho chiều cao trung bình.

Từ các số liệu trên, ta có:

$$\overline{X} = 171; s = 3,4254; \alpha = 1 - 0,95 = 0,05; t_0,025(15) = 2,131$$

nên khoảng tin cậy 95% cho chiều cao trung bình là:

$$\overline{X} \pm t_{\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}} = 171 \pm 2,131\frac{3,4254}{\sqrt{16}} = 171 \pm 1,885$$

II Khoảng tin cậy cho tỉ lệ

Xét $p = p(A)$ chưa biết, ta cần ước lượng tỉ lệ này.

Giả sử trong mẫu cỡ n có k lần xuất hiện biến cố A, $f = \frac{k}{n}$. Khi đó, nếu $nf > 10, n(1-f) > 10$ thì khoảng tin cậy $1 - \alpha$ cho p là:

$$\left(f \pm z\left(\frac{\alpha}{2}\right)\sqrt{\frac{f(1-f)}{n}}\right)$$

Ví dụ 3: Trước ngày bầu cử tổng thống, một cuộc thăm dò dư luận được tiến hành. Người ta chọn ngẫu nhiên 100 người để hỏi ý kiến thì có 60 người nói rằng họ sẽ bỏ phiếu cho ông A. Tìm khoảng tin cậy 90% cho tỉ lệ cử tri bỏ phiếu cho ông A.

Ta có: $\alpha = 1 - 0.9 = 0.1; n = 100; k = 60 \Rightarrow f = \frac{k}{n} = 0,6$

$nf = 100.0,6 = 60 > 10; n(1-f) = 100.0,4 = 40 > 10$ nên khoảng tin cậy 90% cho p là:

$$f \pm z\left(\frac{\alpha}{2}\right)\sqrt{\frac{f(1-f)}{n}} = 0.6 \pm 1.64\sqrt{\frac{0,6.0,4}{100}} = 0,6 \pm 0,08$$

III Khoảng tin cậy cho sự khác biệt giữa 2 giá trị trung bình với mẫu độc lập.

1 Phương sai đã biết

Giả sử X, Y là 2 biến ngẫu nhiên có phân bố chuẩn với giá trị trung bình μ_1, μ_2. Phương sai σ_1^2, σ_2^2 đã biết. Đặt $D = \mu_1 - \mu_2$. Khi đó, khoảng tin cậy $1-\alpha$ cho ED là:

$$\left((\overline{X} - \overline{Y}) \pm z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\sigma_1^2}{2} + \frac{\sigma_2^2}{2}}\right)$$

2 Phương sai chưa biết, $n_i > 30$

Khoảng tin cậy $1-\alpha$ cho DX là:

$$\left((\overline{X} - \overline{Y}) \pm z\left(\frac{\alpha}{2}\right)\sqrt{\frac{s_1^2}{2} + \frac{s_2^2}{2}}\right)$$

3 Phương sai chưa biết, $n_i < 30$

Khoảng tin cậy $1-\alpha$ cho DX là:

$$\left((\overline{X} - \overline{Y}) \pm t_{\frac{\alpha}{2}}(n_1 + n_2 - 2)s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

trong đó:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

IV Khoảng tin cậy cho phương sai

Nếu tổng thể X có phân bố chuẩn thì khoảng tin cậy $1-\alpha$ cho phương sai DX là:

$$\left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}; \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right)$$

Ví dụ 4: Tìm khoảng tin cậy 95% cho độ lệch tiêu chuẩn của X biết rằng quan sát X 11 lần thấy phương sai mẫu $s = 1,549$.

Ta có: $\alpha = 1 - 0,95 = 0,05; s^2 = 1,549$ nên khoảng tin cậy 95% cho phương sai DX là:

$$DX \in \left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}; \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right) = \left(\frac{10.1,549}{\chi^2_{0,025}(10)}; \frac{10.1,549}{\chi^2_{0,975}(10)}\right) = (0,756; 4,77) \Rightarrow \sigma_X \in (0,87; 2,18)$$

Chương 2

Kiểm định giả thiết thống kê

I Kiểm định giả thiết về giá trị trung bình

Giả sử X là ĐLNN có phân bố chuẩn. Tập hợp chính ở đây là tập hợp tất cả các giá trị có thể có của X. Một mẫu kích thước n là một tập hợp gồm n giá trị $x_1, x_2, ..., x_n$ thu được từ n quan sát độc lập về X. Ta muốn kiểm định giả thiết về μ.

Giả thiết: $H_0 : \mu = \mu_0$

	$H_1 : \mu \neq \mu_0$	$H_1 : \mu > \mu_0$	$H_1 : \mu < \mu_0$	Test thống kê
Phương sai đã biết	$\lvert T\rvert > z\left(\dfrac{\alpha}{2}\right)$	$T > z(\alpha)$	$T < -z(\alpha)$	$T = \dfrac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma}$
Phương sai chưa biết $n > 30$	$\lvert T\rvert > z\left(\dfrac{\alpha}{2}\right)$	$T > z(\alpha)$	$T < -z(\alpha)$	$T = \dfrac{(\overline{X} - \mu_0)\sqrt{n}}{s}$
Phương sai chưa biết, $n < 30$	$\lvert T\rvert > t_{\frac{\alpha}{2}}(n-1)$	$T > t_\alpha(n-1)$	$T < -t_\alpha(n-1)$	$T = \dfrac{(\overline{X} - \mu_0)\sqrt{n}}{s}$

Ví dụ 5: Một tay đua xe đạp nói rằng mỗi ngày trung bình anh ta đạp xe ít nhất 5 dặm. Chọn ngẫu nhiên 8 ngày trong sổ tay anh ta thì thấy các số liệu ghi quãng đường anh ta đi được như sau:

$$5,3 \quad 4,5 \quad 4,8 \quad 5,1 \quad 4,3 \quad 4,8 \quad 4,9 \quad 4,7$$

Với mức ý nghĩa $\alpha = 5\%$ có thể cho rằng anh ta nói đúng hay không.

Giả thiết $H_0 : \mu = 5$

Đối thiết $H_1 : \mu < 5$

Từ dữ kiện đề bài, ta có: $\overline{X} = 4,8; \sigma = 0,2958$.

Test thống kê:

$$T = \frac{(4,8-5)\sqrt{8}}{0,2958} = -1,91$$

$$t_{0,05}(7) = 1,895$$

Do $T = -1,91 < -1,895$ nên ta không có cơ sở bác bỏ H_0. Vậy có thể nói anh ta đạp xe trung bình ít nhất 5 dặm 1 ngày.

II Kiểm định cho tỉ lệ

Xét một phép thử ngẫu nhiên G và một biến cố A liên kết với G. Xác suất xuất hiện A khi phép thử được thực hiện là p chưa biết. Ta muốn kiểm định giả thiết $p = p_0$ ở đó p_0 là một số đã cho. Tiến hành phép thử G n lần một cách độc lập và ta quan sát thấy biến cố A xuất hiện k lần. Tần suất xuất hiện của A là $f = \dfrac{k}{n}$ cho ta một hình ảnh xấp xỉ của p. Bài toán kiểm định:

Giả thiết: $H_0 : p = p_0$

$H_1 : p \neq p_0$	$H_1 : p > p_0$	$H_1 : p < p_0$	Test thống kê
$\lvert T\rvert > z\left(\dfrac{\alpha}{2}\right)$	$T > z(\alpha)$	$T < -z(\alpha)$	$T = \dfrac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}}$

Ví dụ 6: Một đảng chính trị trong một cuộc bầu cử tổng thống ở Mỹ tuyên bố rằng 45% cử tri sẽ bỏ phiếu cho ông A. Chọn ngẫu nhiên 200 cử tri để thăm dò ý kiến cho thấy 80 người trong số đó tuyên bố bỏ phiếu cho ông A. Với mức ý nghĩa $\alpha = 5\%$, hãy kiểm định xem dự đoán của đảng trên có đúng không.

Giả thiết: $H_0 : p = 0,45$

Đối thiết: $H_1 : p \neq 0,45$

Ta có: $f = \dfrac{80}{200} = 0,4 \Rightarrow$ Test thống kê $T = \dfrac{(0,4 - 0,45)\sqrt{200}}{\sqrt{0,45.0,55}} = -1,43$

Với mức ý nghĩa $\alpha = 0,05$ thì $z(0,025) = 1,96$

Do $|T| = 1,43 < 1,96$ nên ta không có cơ sở bác bỏ H_0. Dự đoán của đảng trên có thể đúng.

III Kiểm định về giá trị của nhiều tỉ lệ

Xét một phép thử ngẫu nhiên $\mathcal{T}$ và một hệ đầy đủ các biến cố $B_1, B_2, ..., B_k$ liên kết với $\mathcal{T}$. Điều đó có nghĩa là với mỗi kết quả của $\mathcal{T}$, dù là kết quả nào đi chăng nữa, luôn luôn có một và chỉ một biến cố trong các biến cố $B_1, B_2, ..., B_n$ xảy ra. Giả sử rằng ta quan tâm tới các xác suất của các biến cố B_i này. Giả thiết cần kiểm định là:

$$\begin{aligned} H_0: \quad & P(B_1) = p_1 \\ & P(B_2) = p_2 \\ & ... \\ & P(B_k) = p_k \end{aligned}$$

Tiến hành phép thử $\mathcal{T}$ n lần một cách độc lập. Giả sử rằng có n_i lần xảy ra các biến cố B_i, $\sum_{i=1}^{k} n_i = n$. Các số n_i được gọi là các tần số quan sát. Ta trình bày tỉ lệ cần kiểm định và tần số quan sát thành bảng sau:

Biến cố	B_1	B_2	...	B_k	Tổng
Tỉ lệ cần kiểm định	p_1	p_2	...	p_k	1
Tần số quan sát	n_1	n_2	...	n_k	n

Khi đó, Test thống kê được cho như sau:

$$T = \sum_{i=1}^{k} \frac{(n_i - n_i p_i)^2}{n_i p_i} = \frac{1}{n} \sum_{i=1}^{k} \frac{n_i^2}{p_i} - n$$

Miền bác bỏ giả thiết: $T > \chi^2_\alpha(n-1)$

Ví dụ 7: Gieo một con xúc sắc 600 lần. Số lần ra các mặt được cho trong bảng sau. Với mức ý nghĩa $\alpha = 5\%$, có thể coi con xúc sắc được chế tạo cân đối (tức là xác suất xuất hiện mỗi mặt là $\frac{1}{6}$) được không?

1	2	3	4	5	6	Tổng
106	92	97	105	88	112	600

Giả thiết $H_0 : p_1 = p_2 = ... = p_6 = \frac{1}{6}$

Đối thiết $H_1 : \exists i \quad p_i \neq \frac{1}{6}$

Test thống kê:

$$T = \frac{1}{600} \sum_{i=1}^{k} 6.n_i^2 - 600 = 4,22$$

Tra bảng ta được: $\chi^2_{0,05}(5) = 11,07$

Do $T < 11,07$ nên ta không có cơ sở bác bỏ H_0. Vậy có thể nói con xúc sắc này là cân đối.

Chương 3

Bài toán so sánh

I So sánh hai giá trị trung bình

1 *Hai mẫu độc lập*

Giả sử X và Y là hai ĐLNN có phân bố chuẩn; chúng ta muốn so sánh giá trị trung bình của X và Y. Giả sử $\{x_1, x_2, ..., x_n\}$ là một mẫu ngẫu nhiên kích thước n rút ra từ tập chính, bao gồm tập hợp tất cả các giá trị có thể có của X, và $\{y_1, y_2, ..., y_m\}$ là một mẫu ngẫu nhiên kích thước m rút ra từ tập chính, bao gồm tất cả các giá trị có thể có của Y. Hai giá trị mẫu trên độc lập với nhau. Ta muốn kiểm định giả thiết:

$$H_0 : \mu_1 = \mu_2$$

	$H_1 : \mu_1 \neq \mu_2$	$H_1 : \mu_1 > \mu_2$	$H_1 : \mu_1 < \mu_2$	Test thống kê
TH1	$\lvert T\rvert > z\left(\frac{\alpha}{2}\right)$	$T > z(\alpha)$	$T < -z(\alpha)$	$T = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$
TH2	$\lvert T\rvert > z\left(\frac{\alpha}{2}\right)$	$T > z(\alpha)$	$T < -z(\alpha)$	$T = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$
TH3	$\lvert T\rvert > t_{\frac{\alpha}{2}}(n_1 + n_2 - 2)$	$T > t_\alpha(n_1 + n_2 - 2)$	$T < -t_\alpha(n_1 + n_2 - 2)$	$T = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{s^2}{n_1} + \dfrac{s^2}{n_2}}}$

trong đó:

TH1: Phương sai đã biết

TH2: Phương sai chưa biết, $n_i > 30$

TH3: Phương sai chưa biết, $n_i < 30$, $s^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

Ví dụ 8: Người ta ghi lại sản lượng lúa mì, tính bằng tạ trên hecta của các mảnh ruộng đã bón lót 50 và 100 đơn vị đạm trên 1 hecta.

Bón 50 đơn vị: $47,2$ $43,1$ $35,7$ $47,0$ $45,7$ $42,6$ $46,7$ $42,3$

Bón 100 đơn vị: $47,9 \quad 48,9 \quad 43,5 \quad 53,1 \quad 50,8 \quad 46,1 \quad 41,1 \quad 43,0 \quad 41,0 \quad 48,5 \quad 47,7$

Với mức ý nghĩa $\alpha = 5\%$, có thể kết luận rằng bón lót 100 đơn vị đạm cho năng suất cao hơn bón lót 50 đơn vị đạm hay không?

Gọi μ_1 là sản lượng trung bình khi bón lót 100 đơn vị đạm và μ_2 là sản lượng trung bình khi bón lót 50 đơn vị đạm.

Giả thiết: $H_0 : \mu_1 = \mu_2$

Đối thiết: $H_1 : \mu_1 > \mu_2$

Từ giả thiết ta tính được: $\overline{X} = 46,54; \overline{Y} = 43,85; s = 3,84$

Test thống kê:

$$T = \frac{46,54 - 43,85}{3,84\sqrt{\dfrac{1}{8} + \dfrac{1}{11}}} = 1,49$$

Tra bảng ta được: $t_{0,05}(17) = 1,74$

Do $T < 1,74$ nên ta không có cơ sở bác bỏ H_0. Vậy chưa thể nói rằng bón lót 100 đơn vị đạm tốt hơn bón lót 50 đơn vị đạm.

2 *Hai mẫu phụ thuộc*

Giả sử (X, Y) là một cặp gồm hai đại lượng ngẫu nhiên phụ thuộc nhau với $EX = \mu_1$ và $EY = \mu_2$. Chúng ta muốn so sánh μ_1 và μ_2.

Giả sử $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ là n quan sát độc lập của (X, Y). Khi đó, ta có hai mẫu cùng kích thước

$$\{x_1, x_2, ..., x_n\}, \quad \{y_1, y_2, ..., y_n\}$$

Để giải quyết bài toán này, ta xét hiệu số

$$D = X - Y$$

Khi đó, giá trị trung bình của D là $\mu = \mu_1 - \mu_2$ và các giá trị $d_i = x_i - y_i$ cho ta một mẫu gồm n quan sát các giá trị của D. Giả thiết muốn kiểm định là

$$H_0 : \mu_1 = \mu_2 \text{ hay } \mu = \mu_1 - \mu_2 = 0$$

Khi đó, ta đưa bài toán so sánh và bài toán kiểm định giả thiết về giá trị trung bình.

Ví dụ 8: Để khảo sát tác dụng của việc bón thêm 1 loại phân mới A, người ta chia mỗi thửa ruộng thí nghiệm làm hai mảnh. Một mảnh đối chứng (Không bón phân A), mảnh kia có bón 70 đơn vị phân A. Sản lượng của 17 thửa ruộng được ghi lại như sau:

Thửa ruộng	Mảnh đối chứng	Mảnh bón phân A	Hiệu số
1	55.8	60.4	4.6
2	53.3	58.7	5.4
3	30.1	28.9	−1.2
4	51.0	48.0	−3.0
5	37.8	39.7	1.9
6	68.6	68.8	0.2
7	57.7	57.5	−0.2
8	59.1	70.4	11.3
9	49.4	56.8	7.4
10	35.4	40.6	5.2
11	53.4	57.3	3.9
12	42.7	44.3	1.6
13	21.2	32.2	11.0
14	28.3	47.7	19.4
15	57.3	77.0	19.7
16	42.4	55.1	12.7
17	61.4	66.1	4.7

Với mức ý nghĩa 5% hãy nhận định xem việc bón phân có tác dụng hay không?

Giả thiết: $H_0 : \mu = 0$

Đối thiết: $H_1 : \mu > 0$

Từ số liệu ta tìm được: $\overline{D} = 6,15; s = 6,694$

Test thống kê:

$$T = \frac{(6,15 - 0)\sqrt{17}}{6,694} = 3,79$$

Tra bảng ta được: $t_{0,05}(16) = 1,746$

Do $T > 1,746$ nên ta có cơ sở bác bỏ H_0. Vậy có thể nói rằng việc bón phân có tác dụng.

II Tiêu chuẩn phi tham số

Tham khảo giáo trình, trang 146-157

III So sánh hai tỉ lệ

Xét hai tập hợp chính I và II và một đặc tính A mà mỗi cá thể của hai tập hợp chính đó có thể có hay không. Ta muốn so sánh tỉ lệ cá thể có đặc tính A của tập chính I với tỉ lệ cá thể có đặc tính A của tập chính II. Gọi p_1 và p_2 tương ứng là các tỉ lệ cá thể có đặc tính A trong tập chính I và II. Giả thiết H_0 mà ta muốn kiểm định là:

$$H_0 : p_1 = p_2$$

Giả sử n_1 và n_2 là kích thước của hai mẫu rút ra từ tập chính I và II. k_1 và k_2 tương ứng là số các cá thể có đặc tính A trong mẫu lấy từ tập chính I và II.

Ta có các tần suất $f_1 = \dfrac{k_1}{n_1}$ và $f_2 = \dfrac{k_2}{n_2}$ là các ước lượng cho p_1 và p_2. Tần suất chung $f = \dfrac{k_1+k_2}{n_1+n_2}$

$H_1 : p_1 \neq p_2$	$H_1 : p_1 > p_2$	$H_1 : p_1 < p_2$	Test thống kê
$\|T\| > z\left(\dfrac{\alpha}{2}\right)$	$T > z(\alpha)$	$T < -z(\alpha)$	$T = \dfrac{f_1 - f_2}{\sqrt{f(1-f)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

Ví dụ 9: Trong một cuộc thăm dò trước ngày bầu cử, 42 trong tổng số 100 cử tri nam được hỏi cho biết sẽ bỏ phiếu cho ứng cử viên A. Trong khi đó 92 trong số 200 cử tri nữ cho biết sẽ bỏ phiếu cho ông A.

Với mức ý nghĩa 5% kiểm định xem tỉ lệ cử tri nam bầu cho ông A với tỉ lệ cử tri nữ bầu cho ông A có như nhau hay không?

Gọi p_1 và p_2 là tỉ lệ cử tri nam và nữ bỏ phiếu cho ông A.

Giả thiết: $H_0 : p_1 = p_2$

Đối thiết: $H_1 : p_1 \neq p_2$

Từ các dữ kiện đề bài ta có:

$$n_1 = 100, k_1 = 42, f_1 = 0,42;$$

$$n_2 = 200, k_2 = 92, f_2 = 0,46$$

$$\text{Tần suất chung: } f = \frac{42+92}{100+200} = 0,447$$

Test thống kê:

$$T = \frac{0,42 - 0,46}{\sqrt{0,447.0,553(0,01+0,005)}} = -0,66$$

Do $|T| = 0,66 < 1,96$ nên ta không có cơ sở bác bỏ H_0. Vậy có thể kết luận rằng tỉ lệ cử tri nam bầu cho ông A và tỉ lệ cử tri nữ bầu cho ông A là như nhau.

Chương 4

Phân tích phương sai

I Phân tích phương sai một nhân tố

Giả sử ta có k ĐLNN có phân bố chuẩn $X_1, X_2, ..., X_k$, trong đó $X_i \sim N(\mu_i, \sigma_i^2)$

Các giá trị trung bình μ_i và phương sai σ_i^2 đều chưa biết. Tuy nhiên chúng ta giả thiết rằng các phương sai bằng nhau. Chúng ta muốn kiểm định xem các giá trị trung bình có bằng nhau hay không.

Giả thiết:

$$H_0 : \mu_1 = \mu_2 = ... = \mu_k$$

Giả sử $\{x_{1i}, x_{1i}, ..., x_{n_i i}\}$ là một mẫu có kích thước n_i rút ra từ tập hợp chính các giá trị X_i. Các số liệu thu được sẽ được trình bày thành bảng ở dạng sau:

	Các mức nhân tố				
	1	2	...	k	
	x_{11}	x_{12}	...	n_{1k}	
	x_{21}	x_{22}	...	x_{2k}	
	...	...	...	...	
	$x_{n_1 1}$	$x_{n_2 2}$	...	$x_{n_k k}$	$n = \sum_{i=1}^{k} n_1$
Tổng số	T_1	T_2	...	T_k	$T = \sum_{i=1}^{k} T_k$
Trung bình	$\overline{x}_1$	$\overline{x}_2$	...	$\overline{x}_k$	$\overline{x} = \frac{T}{n}$

Ta đưa ra một số ký hiệu sau:

(1) Trung bình của mẫu thứ i (tức là mẫu ở cột thứ i trong bảng trên):

$$\overline{X_i} = \frac{T_i}{n_i} = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i}$$

(2) Trung bình chung:

$$\overline{X} = \frac{T}{n} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_j} x_{ij}}{\sum_{i=1}^{k} n_i}$$

(3) Tổng bình phương chung ký hiệu là SST:

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_j}(x_{ij} - \overline{X})^2 = \sum_{i,j} x_{ij}^2 - \frac{T^2}{n}$$

(4) Tổng bình phương do nhân tố ký hiệu là SSF:

$$SSF = \sum_{i=1}^{k} n_i(\overline{X_i} - \overline{X})^2 = \sum_{i=1}^{k} \frac{T_i^2}{n_i} - \frac{T^2}{n}$$

(5) Tổng bình phương do sai số ký hiệu là SSE:

$$SSE = \sum_{i,j}(\overline{X_{ij}} - \overline{X_i})^2 = SST - SSF$$

(6) Trung bình bình phương của nhân tố ký hiệu là MSF:

$$MSF = \frac{SSF}{k-1}$$

(7) Trung bình bình phương của sai số ký hiệu là MSE:

$$MSE = \frac{SSE}{n-k}$$

(8) Tỉ số F được tính bởi công thức:

$$F = \frac{MSF}{MSE}$$

Các kết quả trên được trình bày trong bảng sau đây gọi là bảng ANOVA:

Nguồn	Tổng bình phương	Bậc tự do	Trung bình bình phương	Tỷ số F
Nhân tố	SSF	k – 1	MSF	$\frac{MSF}{MSE}$
Sai số	SSE	n – k	MSE	
Tổng	SST	n – 1		

Miền bác bỏ giả thiết:

$$\{F > F_\alpha(k-1, n-k)\}$$

Các bước lập bảng phân tích phương sai ANOVA:

Bước 1: Tính SSF

Bước 2: Tính SST

Bước 3: Tính SSE = SST - SSF

Bước 4: Tính MSF

Bước 5: Tính MSE

Bước 6: Tính F

Bước 7: Tra bảng phân bố Fisher và đưa ra kết luận

Ví dụ 10 Điểm thi của 12 sinh viên học các giáo sư A, B, C được cho trong bảng sau. Với mức ý nghĩa 5%, kiểm định xem liệu điểm thi trung bình của các sinh viên theo học các giáo sư A, B, C có giống nhau hay không?

Giáo sư A	Giáo sư B	Giáo sư C
79	71	82
86	77	68
94	81	70
89	83	76

Gọi X_1, X_2, X_3 lần lượt là các nhân tố ứng với điểm của các sinh viên theo học các giáo sư A, B, C và μ_1, μ_2, μ_3 là các giá trị trung bình tương ứng.

Giả thiết:

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

Từ dữ kiện đề bài, ta tính được:

$$T_1 = 348, \quad T_2 = 312, \quad T_3 = 296 \Rightarrow \quad T = 956$$

$$n_1 = n_2 = n_3 = 4, \quad n = 12$$

Lập bảng phân tích phương sai ANOVA:

Bước 1: $SSF = \frac{348^2}{4} + \frac{312^2}{4} + \frac{296^2}{4} - \frac{956^2}{12} = 354,67$

Bước 2: $SST = 79^2 + 86^2 + ... + 70^2 + 76^2 - \frac{956^2}{12} = 676,67$

Bước 3: $SSE = SST - SSF = 322$

Bước 4: $MSF = \frac{354,67}{2} = 177,34$

Bước 5: $MSE = \frac{322}{9} = 35,78$

Bước 6: $F = \frac{177,34}{35,78} = 4,96$

Bước 7: Tra bảng ta được $F_{0,05}(2,9) = 4,26$

Từ đó ta có bảng phân tích phương sai ANOVA:

Nguồn	Tổng bình phương	Bậc tự do	Trung bình bình phương	Tỷ số F
Nhân tố	354,67	2	177,34	4,96
Sai số	322	9	35.78	
Tổng	676,67	11		

Do $F > 4,26$ nên ta có cơ sở bác bỏ H_0. Vậy với mức ý nghĩa 5% có thể nói rằng, điểm thi trung bình của các sinh viên theo học các giáo sư A, B, C là khác nhau.

II Phân tích phương sai hai nhân tố

Tham khảo giáo trình, trang 194-202

Chương 5

Phân tích tương quan và hồi quy

I Phân tích tương quan tuyến tính

Để đo mức độ phụ thuộc tuyến tính giữa 2 ĐLNN X và Y, người ta đưa ra khái niệm hệ số tương quan. Hệ số tương quan lý thuyết của X và Y, ký hiệu là ρ, được định nghĩa bởi công thức sau:

$$\rho = \frac{E(X-\mu_X).(Y-\mu_Y)}{\sigma_X \sigma_Y}$$

Trong đó: μ_X, σ_X lần lượt là giá trị trung bình và độ lệch tiêu chuẩn của X; μ_Y, σ_Y lần lượt là giá trị trung bình và độ lệch tiêu chuẩn của Y. Người ta đã chứng minh được $\rho \in [-1;1]$. Khi $\rho = 0$ thì không có tương quan tuyến tính giữa X và Y. Khi $|\rho|$ càng gần 1 thì sự phụ thuộc tuyến tính của X và Y càng mạnh.

Đại lượng sau đây được sử dụng để ước lượng cho ρ căn cứ trên một mẫu quan sát $(x_1, y_1), (x_2, y_2), ..., (x$ các giá trị của (X, Y).

$$r = \frac{n(\sum xy) - (\sum x).(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

II Kiểm tra tính độc lập

Xét bài toán kiểm định tính độc lập của hai dấu hiệu định tính A và B. Ta chia dấu hiệu A làm r mức độ $A_1, A_2, ..., A_r$ và chia đặc tính B làm k mức $B_1, B_2, ..., B_k$. Xét một ngẫu nhiên gồm n cá thể. Mỗi cá thể mang dấu hiệu A ở mức A_i nào đó và mang dấu hiệu B ở mức B_j nào đó. Giả sử n_{ij} là số các cá thể có các dấu hiệu A_i và B_j. Các số liệu n_{ij} được ghi trong bảng sau gọi là bảng liên hợp các dấu hiệu.

B \ A	B_1	B_2	...	B_k	Tổng
A_1	n_{11}	n_{12}	...	n_{1k}	n_{10}
A_2	n_{21}	n_{22}	...	n_{2k}	n_{20}
...	...	...	...	...	...
A_r	n_{r1}	n_{r2}	...	n_{rk}	n_{r0}
Tổng	n_{o1}	n_{o2}	...	n_{ok}	n

trong đó ta ký hiệu: $n_{i0} = \sum_{j=1}^{k} n_{ij}, \quad n_{0j} = \sum_{i=1}^{r} n_{ij}$

Test thống kê:

$$T = n\left(\sum_{i,j} \frac{n_{ij}^2}{n_{i0}n_{0j}} - 1\right)$$

Miền bác bỏ giả thiết:

$$\{T > \chi^2_\alpha[(k-1)\times(r-1)]\}$$

Ví dụ 11: Ở các cây ngọc trâm lá có hai dạng **lá phẳng** hoặc **lá nhăn**, hoa có hai dạng là **hoa bình thường** hoặc **hoa hoàng hậu**. Quan sát một mẫu gồm 560 cây ngọc trâm ta thu được kết quả như bảng sau. Với mức ý nghĩa 5%, kiểm định tính độc lập giữa 2 đặc tính trên.

Lá \ Hoa	Bình thường	Hoàng hậu	Tổng số
Phẳng	328	122	450
Nhăn	77	33	110
Tổng số	405	155	560

Giả thiết H_0 : Hai đặc tính trên độc lập

Test thống kê:

$$T = 560\left(\frac{328^2}{450.405} + \frac{122^2}{450.155} + \frac{77^2}{110.405} + \frac{33^2}{110.155} - 1\right) = 0,369$$

Tra bảng ta được:$\chi^2_{0,05}(1) = 3,841$

Do $T < 3,841$ nên ta không có cơ sở bác bỏ H_0. Vậy với mức ý nghĩa 5% có thể nói rằng hai đặc tính trên độc lập với nhau.

III Phân tích tương quan phi tuyến

Để đo mức độ phụ thuộc nói chung của ĐLNN Y vào ĐLNN X, người ta đưa ra khái niệm tỉ số tương quan. Tỉ số tương quan lý thuyết của Y theo X ký hiệu bởi:

$$\eta^2_{Y/X} = 1 - \frac{E(Y - E(Y/X))^2}{DY}$$

Người ta đã chứng minh được rằng:

$$0 \leq \eta^2_{Y/X} \leq 1 \text{ và } \rho^2 \leq \eta^2_{Y/X}$$

Hiệu số $\eta^2_{Y/X} - \rho^2$ đo mức độ phụ thuộc phi tuyến giữa Y và X. Hiệu số này càng lớn có nghĩa là sự tương quan phi tuyến càng mạnh.

Giả sử $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ là một mẫu gồm n quan sát độc lập rút ra từ tập chính (X, Y). Ta sẽ trình bày dãy số liệu (x_i, y_i) thành bảng sau đây gọi là bảng tương quan.

Y \ X	$x_{(1)}$	$x_{(2)}$	...	$x_{(k)}$	
	y_{11}	y_{12}	...	y_{1k}	
	y_{21}	y_{22}	...	y_{2k}	
	...	...	...	...	
	$y_{n_1 1}$	$y_{n_2 2}$	...	$y_{n_k k}$	
	n_1	n_2	...	n_k	$n = \sum n_i$
	T_1	T_2	...	T_k	$n = \sum T_i$

Phân tích phương sai:

(i) Tổng bình phương chung SST:

$$SST = \sum\sum y_{ij}^2 - \frac{T^2}{n}$$

(ii) Tổng bình phương do nhân tố SSF:

$$SSF = \sum_{i=1}^{k} \frac{T_i^2}{n_i} - \frac{T^2}{n}$$

(iii) Đại lượng sau đây được sử dụng như một ước lượng cho tỉ số tương quan lý thuyết $\eta^2_{Y/X}$:

$$\widehat{\eta^2_{Y/X}} = \frac{SSF}{SST}$$

$\widehat{\eta^2_{Y/X}}$, viết gọn là $\widehat{\eta^2}$ được gọi là tỉ số tương quan của Y đối với X. Người ta đã chứng minh được rằng:

$$0 \leq r^2 \leq \widehat{\eta^2}$$

Trong đó r là hệ số tương quan. Bình phương của hệ số tương quan r^2 được gọi là hệ số xác định.

Hiệu số $\eta^2 - \rho^2$ giữa tỉ số tương quan lý thuyết và hệ số xác định lý thuyết cho ta một hình ảnh về sự phụ thuộc phi tuyến của Y đối với X. Nếu hiệu số đó bằng 0 thì điều đó có nghĩa là chỉ có tương quan tuyến tính giữa Y và X. Bài toán kiểm định.

Giả thiết: $H_0 : \eta^2 - \rho^2 = 0$

Đối thiết: $H_1 : \eta^2 - \rho^2 > 0$

Test thống kê:

$$F = \frac{(\widehat{\eta^2} - r^2)(n-k)}{(1-\widehat{\eta^2})(k-2)}$$

Miền bác bỏ giả thiết:

$$\{F > F_\alpha(k-2, n-k)\}$$

Ví dụ 12: Cho mẫu quan sát sau đây của cặp ĐLNN (X, Y):

(8 ; 82) ; (8 ; 78) ; (12 ; 65) ; (12 ; 50) ; (20 ; 60) ;
(20 ; 47) ; (24 ; 52) ; (24 ; 41) ; (8 ; 87) ; (8 ; 58) ;
(8 ; 70) ; (12 ; 62) ; (12 ; 55) ; (12 ; 52) ; (20 ; 44) ;
(20 ; 66) ; (20 ; 41) ; (24 ; 57) ; (24 ; 50) ; (24 ; 47) ;
(8 ; 65) ; (12 ; 49) ; (20 ; 57) ; (24 ; 65).

Hãy tính hệ số tương quan, hệ số xác định và tỉ số tương quan của Y đối với X. Kiểm tra xem liệu có tương quan phi tuyến giữa X và Y hay không?.

Trước hết, ta trình bày các số liệu trên dưới dạng bảng tương quan sau:

Y \ X	8	12	20	24	
	82	65	60	52	
	78	50	47	41	
	87	62	44	57	
	58	55	66	50	
	70	52	41	47	
	65	49	57	63	
n_i	6	6	6	6	Σ = 24
T_i	440	333	315	310	T = 1398

Hệ số tương quan:

$$r = \frac{n(\sum xy) - (\sum x).(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}} = \frac{24.21256 - 384.1398}{\sqrt{24.71041 - 384^2}\sqrt{24.84909 - 1398^2}} = 0,6089$$

$\Rightarrow$ Hệ số xác định $r^2 = 0,6089^2 = 0,37$

Ta có:

$SST = \sum\sum y_{ij}^2 - \dfrac{T^2}{n} = 84908 - \dfrac{1398^2}{24} = 3474,5$

$SSF = \sum_{i=1}^{k} \dfrac{T_i^2}{n_i} - \dfrac{T^2}{n} = \dfrac{440^2 + 333^2 + 315^2 + 310^2}{6} - \dfrac{1398^2}{24} = 1868,83$

$\Rightarrow$ Tỉ số tương quan $\widehat{\eta^2} = \dfrac{SSF}{SST} = 0,5378$

Test thống kê:

$$F = \frac{(\widehat{\eta^2} - r^2)(n-k)}{(1-\widehat{\eta^2})(k-2)} = \frac{(0,5378 - 0,37)(24-4)}{(1-0,5378)(4-2)} = 3,63$$

Tra bảng ta được:

$$F_{0,05}(2,20) = 3,49$$

Do $F > 3,49$ nên ta bác bỏ H_0. Vậy với mức ý nghĩa 5%, có thể nói rằng có tồn tại mối tương quan phi tuyến của Y đối với X

IV Phân tích hồi quy tuyến tính

Giả sử X là một biến nào đó (có thể là biến ngẫu nhiên hay không ngẫu nhiên), còn Y là một ĐLNN phụ thuộc vào X theo cách sau đây. Nếu X nhận giá trị x thì Y sẽ có kỳ vọng là $\alpha x + \beta$, ở đó α và β là hằng số và phương sai là σ^2. Khi đó, ta nói Y *có hồi quy tuyến tính theo* X, và đường thẳng

$$y = \alpha x + \beta$$

được gọi là *đường thẳng hồi quy lý thuyết của Y đối với X*. Các hệ số α, β được gọi là các *hệ số hồi quy lý thuyết*. X được gọi là *biến độc lập* và Y được gọi là *biến phụ thuộc*.

Các hệ số α, β được xác định theo công thức sau

$$a = \frac{n\sum xy - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$
$$b = \frac{\sum y - a\sum x}{n}$$

Ngoài việc ước lượng hệ số hồi quy α và β, ta còn quan tâm tới ước lượng σ^2 là một con số đo sự phân tán của Y xung quanh đường thẳng hồi quy. Ước lượng cho σ^2, ký hiệu bởi $s^2_{Y/X}$ được xác định theo công thức sau:

$$s^2_{Y/X} = \frac{\sum y^2 - a\sum xy - b\sum y}{n-2}$$

$s_{Y/X}$ được gọi là sai số tiêu chuẩn của đường hồi quy. Nó cho ta số đo sự phân tán của đám mây điểm (x_i, y_i) xung quanh đường thẳng hồi quy.

Bây giờ dựa vào phương trình đường thẳng hồi quy tìm được, ta có thể dự báo được giá trị của Y nếu biết giá trị của X. Giá trị được dự báo của Y khi $X = x_0$ sẽ là:

$$y_0 = \alpha x_0 + \beta$$

Đây đồng thời cũng là giá trị dự báo cho kì vọng của Y tương ứng với $X = x_0$:

$$\mu_{x_0} = a x_0 + b$$

Tiếp theo, ta xét bài toán tìm khoảng tin cậy cho giá trị dự báo của Y, cũng như tìm khoảng tin cậy cho giá trị dự báo của μ_{x_0}

Công thức để tìm khoảng tin cậy cho giá trị dự báo của Y khi $X = x_0$ với độ tin cậy $1-\alpha$ là:

$$\left(y_0 \pm t_{\frac{\alpha}{2}}(n-2) s_{Y/X} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{X})^2}{\sum x^2 - \frac{(\sum x)^2}{n}}} \right)$$

Công thức để tìm khoảng tin cậy cho giá trị dự báo của μ_{x_0} khi $X = x_0$ với độ tin cậy $1-\alpha$ là:

$$\left(y_0 \pm t_{\frac{\alpha}{2}}(n-2) s_{Y/X} \sqrt{\frac{1}{n} + \frac{x_0 - \overline{X}}{\sum x^2 - \frac{(\sum x)^2}{n}}} \right)$$

Một bài toán quan trọng khác là kiểm tra xem hệ số hồi quy lý thuyết α có khác 0 hay không. Giả thiết:

$$H_0 : \alpha = 0$$

Người ta đã chứng minh được rằng hệ số hồi quy α có độ lệch tiêu chuẩn là:

$$s_\alpha = \frac{s_{Y/X}}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}}$$

Test thống kê:

$$T = \frac{\alpha}{s_\alpha}$$

Miền bác bỏ giả thiết:

$$\{T > t_{\frac{\alpha}{2}}(n-2)\}$$

Ví dụ 12: Các số liệu về số trang của một cuốn sách X và giá bán Y được cho trong bảng sau đây:

Tên sách	X	Y (nghìn)
A	400	44
B	600	47
C	500	48
D	600	48
E	400	43
F	500	46

(i) Tìm đường thẳng hồi quy của Y theo X căn cứ trên số liệu trên

(ii) Tính sai số tiêu chuẩn của đường hồi quy

(iii) Với độ tin cậy 95% hãy dự đoán giá bán của một cuốn sách với 450 trang và giá bán trung bình của tất cả cuốn sách có 450 trang.

(iv) Với mức ý nghĩa 5% hãy kiểm định xem hệ số góc của đường thẳng hồi quy có bằng 0 hay không?

Lời giải

(i) Sử dụng máy tính bỏ túi, ta tính được:

$$\alpha = 0,02 \quad \beta = 36$$

Vậy đường thẳng hồi quy là:

$$y = 0,02x + 36$$

(ii) Sai số tiêu chuẩn:

$$s_{Y/X}^2 = \frac{\sum y^2 - a\sum xy - b\sum y}{n-2} = \frac{12718 - 0,02.138800 - 36.276}{4} = 1,5 \Rightarrow s_{Y/X} = 1,22$$

(iii) Giá của cuốn sách đó được dự báo là:

$$y_0 = 0,02.450 + 36 = 45$$

Tra bảng, ta được:

$$t_{0,025}(4) = 2,776$$

Khoảng tin cậy 95% cho giá của một cuốn sách 450 trang là:

$$\left(45 \pm 2,776.1,22\sqrt{1 + \frac{1}{6} + \frac{(450-500)^2}{1540000 - \frac{(3000)^2}{6}}}\right) = (45 \pm 3,77)$$

Khoảng tin cậy 95% cho giá bán trung bình của tất cả cuốn sách 450 trang là:

$$\left(45 \pm 2,776.1,22\sqrt{\frac{1}{6} + \frac{(450-500)^2}{1540000 - \frac{(3000)^2}{6}}}\right) = (45 \pm 1,63)$$

(iv) Giả thiết:

$$H_0 : \alpha = 0$$

Ta có:

$$s_\alpha = \frac{s_{Y/X}}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}} = \frac{1,22}{\sqrt{1540000 - \frac{(3000)^2}{6}}} = 0,0061$$

Test thống kê:

$$T = \frac{0,02}{0,0061} = 3,33$$

Tra bảng ta được:

$$t_{0,025}(4) = 2,776$$

Do $T > 2,776$ nên ta có cơ sở bác bỏ H_0. Vậy với mức ý nghĩa 5% có thể nói rằng hệ số góc của đường hồi quy khác 0.

Chương 6

Các phân bố thường gặp

STATISTICAL TABLES

Cumulative normal distribution

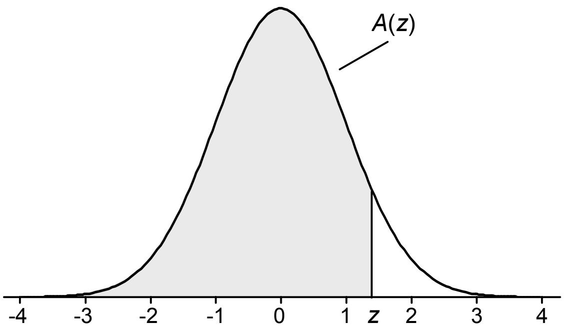
Critical values of the *t* distribution

Critical values of the *F* distribution

Critical values of the chi-squared distribution

TABLE A.1

Cumulative Standardized Normal Distribution



$A(z)$ is the integral of the standardized normal distribution from $-\infty$ to z (in other words, the area under the curve to the left of z). It gives the probability of a normal random variable not being more than z standard deviations above its mean. Values of z of particular importance:

z	$A(z)$	
1.645	0.9500	Lower limit of right 5% tail
1.960	0.9750	Lower limit of right 2.5% tail
2.326	0.9900	Lower limit of right 1% tail
2.576	0.9950	Lower limit of right 0.5% tail
3.090	0.9990	Lower limit of right 0.1% tail
3.291	0.9995	Lower limit of right 0.05% tail

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999							

Table A.2

t Distribution: Critical Values of *t*

		Significance level					
Degrees of freedom	*Two-tailed test*:	10%	5%	2%	1%	0.2%	0.1%
	One-tailed test:	5%	2.5%	1%	0.5%	0.1%	0.05%
1		6.314	12.706	31.821	63.657	318.309	636.619
2		2.920	4.303	6.965	9.925	22.327	31.599
3		2.353	3.182	4.541	5.841	10.215	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
6		1.943	2.447	3.143	3.707	5.208	5.959
7		1.894	2.365	2.998	3.499	4.785	5.408
8		1.860	2.306	2.896	3.355	4.501	5.041
9		1.833	2.262	2.821	3.250	4.297	4.781
10		1.812	2.228	2.764	3.169	4.144	4.587
11		1.796	2.201	2.718	3.106	4.025	4.437
12		1.782	2.179	2.681	3.055	3.930	4.318
13		1.771	2.160	2.650	3.012	3.852	4.221
14		1.761	2.145	2.624	2.977	3.787	4.140
15		1.753	2.131	2.602	2.947	3.733	4.073
16		1.746	2.120	2.583	2.921	3.686	4.015
17		1.740	2.110	2.567	2.898	3.646	3.965
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
21		1.721	2.080	2.518	2.831	3.527	3.819
22		1.717	2.074	2.508	2.819	3.505	3.792
23		1.714	2.069	2.500	2.807	3.485	3.768
24		1.711	2.064	2.492	2.797	3.467	3.745
25		1.708	2.060	2.485	2.787	3.450	3.725
26		1.706	2.056	2.479	2.779	3.435	3.707
27		1.703	2.052	2.473	2.771	3.421	3.690
28		1.701	2.048	2.467	2.763	3.408	3.674
29		1.699	2.045	2.462	2.756	3.396	3.659
30		1.697	2.042	2.457	2.750	3.385	3.646
32		1.694	2.037	2.449	2.738	3.365	3.622
34		1.691	2.032	2.441	2.728	3.348	3.601
36		1.688	2.028	2.434	2.719	3.333	3.582
38		1.686	2.024	2.429	2.712	3.319	3.566
40		1.684	2.021	2.423	2.704	3.307	3.551
42		1.682	2.018	2.418	2.698	3.296	3.538
44		1.680	2.015	2.414	2.692	3.286	3.526
46		1.679	2.013	2.410	2.687	3.277	3.515
48		1.677	2.011	2.407	2.682	3.269	3.505
50		1.676	2.009	2.403	2.678	3.261	3.496
60		1.671	2.000	2.390	2.660	3.232	3.460
70		1.667	1.994	2.381	2.648	3.211	3.435
80		1.664	1.990	2.374	2.639	3.195	3.416
90		1.662	1.987	2.368	2.632	3.183	3.402
100		1.660	1.984	2.364	2.626	3.174	3.390
120		1.658	1.980	2.358	2.617	3.160	3.373
150		1.655	1.976	2.351	2.609	3.145	3.357
200		1.653	1.972	2.345	2.601	3.131	3.340
300		1.650	1.968	2.339	2.592	3.118	3.323
400		1.649	1.966	2.336	2.588	3.111	3.315
500		1.648	1.965	2.334	2.586	3.107	3.310
600		1.647	1.964	2.333	2.584	3.104	3.307
∞		1.645	1.960	2.326	2.576	3.090	3.291

Table A.3

F Distribution: Critical Values of F (5% significance level)

v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.36	246.46	247.32	248.01
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.42	19.43	19.44	19.45
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.71	8.69	8.67	8.66
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.87	5.84	5.82	5.80
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.64	4.60	4.58	4.56
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.96	3.92	3.90	3.87
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.53	3.49	3.47	3.44
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.24	3.20	3.17	3.15
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.03	2.99	2.96	2.94
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.86	2.83	2.80	2.77
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.74	2.70	2.67	2.65
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.64	2.60	2.57	2.54
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.55	2.51	2.48	2.46
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.48	2.44	2.41	2.39
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.42	2.38	2.35	2.33
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.37	2.33	2.30	2.28
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.33	2.29	2.26	2.23
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.29	2.25	2.22	2.19
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.26	2.21	2.18	2.16
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.22	2.18	2.15	2.12
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.20	2.16	2.12	2.10
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.17	2.13	2.10	2.07
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.15	2.11	2.08	2.05
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.13	2.09	2.05	2.03
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.11	2.07	2.04	2.01
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.09	2.05	2.02	1.99
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.08	2.04	2.00	1.97
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.06	2.02	1.99	1.96
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.05	2.01	1.97	1.94
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.04	1.99	1.96	1.93
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	2.04	1.99	1.94	1.91	1.88
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.95	1.90	1.87	1.84
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.89	1.85	1.81	1.78
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.86	1.82	1.78	1.75
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.89	1.84	1.79	1.75	1.72
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.88	1.82	1.77	1.73	1.70
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.86	1.80	1.76	1.72	1.69
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.85	1.79	1.75	1.71	1.68
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.78	1.73	1.69	1.66
150	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.82	1.76	1.71	1.67	1.64
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.80	1.74	1.69	1.66	1.62
250	3.88	3.03	2.64	2.41	2.25	2.13	2.05	1.98	1.92	1.87	1.79	1.73	1.68	1.65	1.61
300	3.87	3.03	2.63	2.40	2.24	2.13	2.04	1.97	1.91	1.86	1.78	1.72	1.68	1.64	1.61
400	3.86	3.02	2.63	2.39	2.24	2.12	2.03	1.96	1.90	1.85	1.78	1.72	1.67	1.63	1.60
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.77	1.71	1.66	1.62	1.59
600	3.86	3.01	2.62	2.39	2.23	2.11	2.02	1.95	1.90	1.85	1.77	1.71	1.66	1.62	1.59
750	3.85	3.01	2.62	2.38	2.23	2.11	2.02	1.95	1.89	1.84	1.77	1.70	1.66	1.62	1.58
1000	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.76	1.70	1.65	1.61	1.58

TABLE A.3 (continued)

***F* Distribution: Critical Values of *F* (5% significance level)**

v_2 \ v_1	25	30	35	40	50	60	75	100	150	200
1	249.26	250.10	250.69	251.14	251.77	252.20	252.62	253.04	253.46	253.68
2	19.46	19.46	19.47	19.47	19.48	19.48	19.48	19.49	19.49	19.49
3	8.63	8.62	8.60	8.59	8.58	8.57	8.56	8.55	8.54	8.54
4	5.77	5.75	5.73	5.72	5.70	5.69	5.68	5.66	5.65	5.65
5	4.52	4.50	4.48	4.46	4.44	4.43	4.42	4.41	4.39	4.39
6	3.83	3.81	3.79	3.77	3.75	3.74	3.73	3.71	3.70	3.69
7	3.40	3.38	3.36	3.34	3.32	3.30	3.29	3.27	3.26	3.25
8	3.11	3.08	3.06	3.04	3.02	3.01	2.99	2.97	2.96	2.95
9	2.89	2.86	2.84	2.83	2.80	2.79	2.77	2.76	2.74	2.73
10	2.73	2.70	2.68	2.66	2.64	2.62	2.60	2.59	2.57	2.56
11	2.60	2.57	2.55	2.53	2.51	2.49	2.47	2.46	2.44	2.43
12	2.50	2.47	2.44	2.43	2.40	2.38	2.37	2.35	2.33	2.32
13	2.41	2.38	2.36	2.34	2.31	2.30	2.28	2.26	2.24	2.23
14	2.34	2.31	2.28	2.27	2.24	2.22	2.21	2.19	2.17	2.16
15	2.28	2.25	2.22	2.20	2.18	2.16	2.14	2.12	2.10	2.10
16	2.23	2.19	2.17	2.15	2.12	2.11	2.09	2.07	2.05	2.04
17	2.18	2.15	2.12	2.10	2.08	2.06	2.04	2.02	2.00	1.99
18	2.14	2.11	2.08	2.06	2.04	2.02	2.00	1.98	1.96	1.95
19	2.11	2.07	2.05	2.03	2.00	1.98	1.96	1.94	1.92	1.91
20	2.07	2.04	2.01	1.99	1.97	1.95	1.93	1.91	1.89	1.88
21	2.05	2.01	1.98	1.96	1.94	1.92	1.90	1.88	1.86	1.84
22	2.02	1.98	1.96	1.94	1.91	1.89	1.87	1.85	1.83	1.82
23	2.00	1.96	1.93	1.91	1.88	1.86	1.84	1.82	1.80	1.79
24	1.97	1.94	1.91	1.89	1.86	1.84	1.82	1.80	1.78	1.77
25	1.96	1.92	1.89	1.87	1.84	1.82	1.80	1.78	1.76	1.75
26	1.94	1.90	1.87	1.85	1.82	1.80	1.78	1.76	1.74	1.73
27	1.92	1.88	1.86	1.84	1.81	1.79	1.76	1.74	1.72	1.71
28	1.91	1.87	1.84	1.82	1.79	1.77	1.75	1.73	1.70	1.69
29	1.89	1.85	1.83	1.81	1.77	1.75	1.73	1.71	1.69	1.67
30	1.88	1.84	1.81	1.79	1.76	1.74	1.72	1.70	1.67	1.66
35	1.82	1.79	1.76	1.74	1.70	1.68	1.66	1.63	1.61	1.60
40	1.78	1.74	1.72	1.69	1.66	1.64	1.61	1.59	1.56	1.55
50	1.73	1.69	1.66	1.63	1.60	1.58	1.55	1.52	1.50	1.48
60	1.69	1.65	1.62	1.59	1.56	1.53	1.51	1.48	1.45	1.44
70	1.66	1.62	1.59	1.57	1.53	1.50	1.48	1.45	1.42	1.40
80	1.64	1.60	1.57	1.54	1.51	1.48	1.45	1.43	1.39	1.38
90	1.63	1.59	1.55	1.53	1.49	1.46	1.44	1.41	1.38	1.36
100	1.62	1.57	1.54	1.52	1.48	1.45	1.42	1.39	1.36	1.34
120	1.60	1.55	1.52	1.50	1.46	1.43	1.40	1.37	1.33	1.32
150	1.58	1.54	1.50	1.48	1.44	1.41	1.38	1.34	1.31	1.29
200	1.56	1.52	1.48	1.46	1.41	1.39	1.35	1.32	1.28	1.26
250	1.55	1.50	1.47	1.44	1.40	1.37	1.34	1.31	1.27	1.25
300	1.54	1.50	1.46	1.43	1.39	1.36	1.33	1.30	1.26	1.23
400	1.53	1.49	1.45	1.42	1.38	1.35	1.32	1.28	1.24	1.22
500	1.53	1.48	1.45	1.42	1.38	1.35	1.31	1.28	1.23	1.21
600	1.52	1.48	1.44	1.41	1.37	1.34	1.31	1.27	1.23	1.20
750	1.52	1.47	1.44	1.41	1.37	1.34	1.30	1.26	1.22	1.20
1000	1.52	1.47	1.43	1.41	1.36	1.33	1.30	1.26	1.22	1.19

TABLE A.3 (continued)

F Distribution: Critical Values of F (1% significance level)

v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6106.32	6142.67	6170.10	6191.53	6208.73
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.44	99.44	99.45
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.92	26.83	26.75	26.69
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.25	14.15	14.08	14.02
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.77	9.68	9.61	9.55
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.60	7.52	7.45	7.40
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.36	6.28	6.21	6.16
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.56	5.48	5.41	5.36
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	5.01	4.92	4.86	4.81
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.60	4.52	4.46	4.41
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.29	4.21	4.15	4.10
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.05	3.97	3.91	3.86
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.86	3.78	3.72	3.66
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.70	3.62	3.56	3.51
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.56	3.49	3.42	3.37
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.45	3.37	3.31	3.26
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.35	3.27	3.21	3.16
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.27	3.19	3.13	3.08
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.19	3.12	3.05	3.00
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.13	3.05	2.99	2.94
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.07	2.99	2.93	2.88
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	3.02	2.94	2.88	2.83
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.97	2.89	2.83	2.78
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.93	2.85	2.79	2.74
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.89	2.81	2.75	2.70
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.86	2.78	2.72	2.66
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.82	2.75	2.68	2.63
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.79	2.72	2.65	2.60
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.77	2.69	2.63	2.57
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.74	2.66	2.60	2.55
35	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88	2.74	2.64	2.56	2.50	2.44
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.56	2.48	2.42	2.37
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.56	2.46	2.38	2.32	2.27
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.39	2.31	2.25	2.20
70	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59	2.45	2.35	2.27	2.20	2.15
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55	2.42	2.31	2.23	2.17	2.12
90	6.93	4.85	4.01	3.53	3.23	3.01	2.84	2.72	2.61	2.52	2.39	2.29	2.21	2.14	2.09
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.37	2.27	2.19	2.12	2.07
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.23	2.15	2.09	2.03
150	6.81	4.75	3.91	3.45	3.14	2.92	2.76	2.63	2.53	2.44	2.31	2.20	2.12	2.06	2.00
200	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.27	2.17	2.09	2.03	1.97
250	6.74	4.69	3.86	3.40	3.09	2.87	2.71	2.58	2.48	2.39	2.26	2.15	2.07	2.01	1.95
300	6.72	4.68	3.85	3.38	3.08	2.86	2.70	2.57	2.47	2.38	2.24	2.14	2.06	1.99	1.94
400	6.70	4.66	3.83	3.37	3.06	2.85	2.68	2.56	2.45	2.37	2.23	2.13	2.05	1.98	1.92
500	6.69	4.65	3.82	3.36	3.05	2.84	2.68	2.55	2.44	2.36	2.22	2.12	2.04	1.97	1.92
600	6.68	4.64	3.81	3.35	3.05	2.83	2.67	2.54	2.44	2.35	2.21	2.11	2.03	1.96	1.91
750	6.67	4.63	3.81	3.34	3.04	2.83	2.66	2.53	2.43	2.34	2.21	2.11	2.02	1.96	1.90
1000	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.20	2.10	2.02	1.95	1.90

.

TABLE A.3 (continued)

F Distribution: Critical Values of *F* (1% significance level)

v_2 \ v_1	25	30	35	40	50	60	75	100	150	200
1	6239.83	6260.65	6275.57	6286.78	6302.52	6313.03	6323.56	6334.11	6344.68	6349.97
2	99.46	99.47	99.47	99.47	99.48	99.48	99.49	99.49	99.49	99.49
3	26.58	26.50	26.45	26.41	26.35	26.32	26.28	26.24	26.20	26.18
4	13.91	13.84	13.79	13.75	13.69	13.65	13.61	13.58	13.54	13.52
5	9.45	9.38	9.33	9.29	9.24	9.20	9.17	9.13	9.09	9.08
6	7.30	7.23	7.18	7.14	7.09	7.06	7.02	6.99	6.95	6.93
7	6.06	5.99	5.94	5.91	5.86	5.82	5.79	5.75	5.72	5.70
8	5.26	5.20	5.15	5.12	5.07	5.03	5.00	4.96	4.93	4.91
9	4.71	4.65	4.60	4.57	4.52	4.48	4.45	4.41	4.38	4.36
10	4.31	4.25	4.20	4.17	4.12	4.08	4.05	4.01	3.98	3.96
11	4.01	3.94	3.89	3.86	3.81	3.78	3.74	3.71	3.67	3.66
12	3.76	3.70	3.65	3.62	3.57	3.54	3.50	3.47	3.43	3.41
13	3.57	3.51	3.46	3.43	3.38	3.34	3.31	3.27	3.24	3.22
14	3.41	3.35	3.30	3.27	3.22	3.18	3.15	3.11	3.08	3.06
15	3.28	3.21	3.17	3.13	3.08	3.05	3.01	2.98	2.94	2.92
16	3.16	3.10	3.05	3.02	2.97	2.93	2.90	2.86	2.83	2.81
17	3.07	3.00	2.96	2.92	2.87	2.83	2.80	2.76	2.73	2.71
18	2.98	2.92	2.87	2.84	2.78	2.75	2.71	2.68	2.64	2.62
19	2.91	2.84	2.80	2.76	2.71	2.67	2.64	2.60	2.57	2.55
20	2.84	2.78	2.73	2.69	2.64	2.61	2.57	2.54	2.50	2.48
21	2.79	2.72	2.67	2.64	2.58	2.55	2.51	2.48	2.44	2.42
22	2.73	2.67	2.62	2.58	2.53	2.50	2.46	2.42	2.38	2.36
23	2.69	2.62	2.57	2.54	2.48	2.45	2.41	2.37	2.34	2.32
24	2.64	2.58	2.53	2.49	2.44	2.40	2.37	2.33	2.29	2.27
25	2.60	2.54	2.49	2.45	2.40	2.36	2.33	2.29	2.25	2.23
26	2.57	2.50	2.45	2.42	2.36	2.33	2.29	2.25	2.21	2.19
27	2.54	2.47	2.42	2.38	2.33	2.29	2.26	2.22	2.18	2.16
28	2.51	2.44	2.39	2.35	2.30	2.26	2.23	2.19	2.15	2.13
29	2.48	2.41	2.36	2.33	2.27	2.23	2.20	2.16	2.12	2.10
30	2.45	2.39	2.34	2.30	2.25	2.21	2.17	2.13	2.09	2.07
35	2.35	2.28	2.23	2.19	2.14	2.10	2.06	2.02	1.98	1.96
40	2.27	2.20	2.15	2.11	2.06	2.02	1.98	1.94	1.90	1.87
50	2.17	2.10	2.05	2.01	1.95	1.91	1.87	1.82	1.78	1.76
60	2.10	2.03	1.98	1.94	1.88	1.84	1.79	1.75	1.70	1.68
70	2.05	1.98	1.93	1.89	1.83	1.78	1.74	1.70	1.65	1.62
80	2.01	1.94	1.89	1.85	1.79	1.75	1.70	1.65	1.61	1.58
90	1.99	1.92	1.86	1.82	1.76	1.72	1.67	1.62	1.57	1.55
100	1.97	1.89	1.84	1.80	1.74	1.69	1.65	1.60	1.55	1.52
120	1.93	1.86	1.81	1.76	1.70	1.66	1.61	1.56	1.51	1.48
150	1.90	1.83	1.77	1.73	1.66	1.62	1.57	1.52	1.46	1.43
200	1.87	1.79	1.74	1.69	1.63	1.58	1.53	1.48	1.42	1.39
250	1.85	1.77	1.72	1.67	1.61	1.56	1.51	1.46	1.40	1.36
300	1.84	1.76	1.70	1.66	1.59	1.55	1.50	1.44	1.38	1.35
400	1.82	1.75	1.69	1.64	1.58	1.53	1.48	1.42	1.36	1.32
500	1.81	1.74	1.68	1.63	1.57	1.52	1.47	1.41	1.34	1.31
600	1.80	1.73	1.67	1.63	1.56	1.51	1.46	1.40	1.34	1.30
750	1.80	1.72	1.66	1.62	1.55	1.50	1.45	1.39	1.33	1.29
1000	1.79	1.72	1.66	1.61	1.54	1.50	1.44	1.38	1.32	1.28

.

TABLE A.3 (continued)

F Distribution: Critical Values of *F* (0.1% significance level)

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	4.05e05	5.00e05	5.40e05	5.62e05	5.76e05	5.86e05	5.93e05	5.98e05	6.02e05	6.06e05	6.11e05	6.14e05	6.17e05	6.19e05	6.21e05
2	998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37	999.39	999.40	999.42	999.43	999.44	999.44	999.45
3	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62	129.86	129.25	128.32	127.64	127.14	126.74	126.42
4	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47	48.05	47.41	46.95	46.60	46.32	46.10
5	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65	27.24	26.92	26.42	26.06	25.78	25.57	25.39
6	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03	18.69	18.41	17.99	17.68	17.45	17.27	17.12
7	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	14.33	14.08	13.71	13.43	13.23	13.06	12.93
8	25.41	18.49	15.83	14.39	13.48	12.86	12.40	12.05	11.77	11.54	11.19	10.94	10.75	10.60	10.48
9	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	10.11	9.89	9.57	9.33	9.15	9.01	8.90
10	21.04	14.91	12.55	11.28	10.48	9.93	9.52	9.20	8.96	8.75	8.45	8.22	8.05	7.91	7.80
11	19.69	13.81	11.56	10.35	9.58	9.05	8.66	8.35	8.12	7.92	7.63	7.41	7.24	7.11	7.01
12	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.48	7.29	7.00	6.79	6.63	6.51	6.40
13	17.82	12.31	10.21	9.07	8.35	7.86	7.49	7.21	6.98	6.80	6.52	6.31	6.16	6.03	5.93
14	17.14	11.78	9.73	8.62	7.92	7.44	7.08	6.80	6.58	6.40	6.13	5.93	5.78	5.66	5.56
15	16.59	11.34	9.34	8.25	7.57	7.09	6.74	6.47	6.26	6.08	5.81	5.62	5.46	5.35	5.25
16	16.12	10.97	9.01	7.94	7.27	6.80	6.46	6.19	5.98	5.81	5.55	5.35	5.20	5.09	4.99
17	15.72	10.66	8.73	7.68	7.02	6.56	6.22	5.96	5.75	5.58	5.32	5.13	4.99	4.87	4.78
18	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.56	5.39	5.13	4.94	4.80	4.68	4.59
19	15.08	10.16	8.28	7.27	6.62	6.18	5.85	5.59	5.39	5.22	4.97	4.78	4.64	4.52	4.43
20	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24	5.08	4.82	4.64	4.49	4.38	4.29
21	14.59	9.77	7.94	6.95	6.32	5.88	5.56	5.31	5.11	4.95	4.70	4.51	4.37	4.26	4.17
22	14.38	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.99	4.83	4.58	4.40	4.26	4.15	4.06
23	14.20	9.47	7.67	6.70	6.08	5.65	5.33	5.09	4.89	4.73	4.48	4.30	4.16	4.05	3.96
24	14.03	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80	4.64	4.39	4.21	4.07	3.96	3.87
25	13.88	9.22	7.45	6.49	5.89	5.46	5.15	4.91	4.71	4.56	4.31	4.13	3.99	3.88	3.79
26	13.74	9.12	7.36	6.41	5.80	5.38	5.07	4.83	4.64	4.48	4.24	4.06	3.92	3.81	3.72
27	13.61	9.02	7.27	6.33	5.73	5.31	5.00	4.76	4.57	4.41	4.17	3.99	3.86	3.75	3.66
28	13.50	8.93	7.19	6.25	5.66	5.24	4.93	4.69	4.50	4.35	4.11	3.93	3.80	3.69	3.60
29	13.39	8.85	7.12	6.19	5.59	5.18	4.87	4.64	4.45	4.29	4.05	3.88	3.74	3.63	3.54
30	13.29	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39	4.24	4.00	3.82	3.69	3.58	3.49
35	12.90	8.47	6.79	5.88	5.30	4.89	4.59	4.36	4.18	4.03	3.79	3.62	3.48	3.38	3.29
40	12.61	8.25	6.59	5.70	5.13	4.73	4.44	4.21	4.02	3.87	3.64	3.47	3.34	3.23	3.14
50	12.22	7.96	6.34	5.46	4.90	4.51	4.22	4.00	3.82	3.67	3.44	3.27	3.41	3.04	2.95
60	11.97	7.77	6.17	5.31	4.76	4.37	4.09	3.86	3.69	3.54	3.32	3.15	3.02	2.91	2.83
70	11.80	7.64	6.06	5.20	4.66	4.28	3.99	3.77	3.60	3.45	3.23	3.06	2.93	2.83	2.74
80	11.67	7.54	5.97	5.12	4.58	4.20	3.92	3.70	3.53	3.39	3.16	3.00	2.87	2.76	2.68
90	11.57	7.47	5.91	5.06	4.53	4.15	3.87	3.65	3.48	3.34	3.11	2.95	2.82	2.71	2.63
100	11.50	7.41	5.86	5.02	4.48	4.11	3.83	3.61	3.44	3.30	3.07	2.91	2.78	2.68	2.59
120	11.38	7.32	5.78	4.95	4.42	4.04	3.77	3.55	3.38	3.24	3.02	2.85	2.72	2.62	2.53
150	11.27	7.24	5.71	4.88	4.35	3.98	3.71	3.49	3.32	3.18	2.96	2.80	2.67	2.56	2.48
200	11.15	7.15	5.63	4.81	4.29	3.92	3.65	3.43	3.26	3.12	2.90	2.74	2.61	2.51	2.42
250	11.09	7.10	5.59	4.77	4.25	3.88	3.61	3.40	3.23	3.09	2.87	2.71	2.58	2.48	2.39
300	11.04	7.07	5.56	4.75	4.22	3.86	3.59	3.38	3.21	3.07	2.85	2.69	2.56	2.46	2.37
400	10.99	7.03	5.53	4.71	4.19	3.83	3.56	3.35	3.18	3.04	2.82	2.66	2.53	2.43	2.34
500	10.96	7.00	5.51	4.69	4.18	3.81	3.54	3.33	3.16	3.02	2.81	2.64	2.52	2.41	2.33
600	10.94	6.99	5.49	4.68	4.16	3.80	3.53	3.32	3.15	3.01	2.80	2.63	2.51	2.40	2.32
750	10.91	6.97	5.48	4.67	4.15	3.79	3.52	3.31	3.14	3.00	2.78	2.62	2.49	2.39	2.31
1000	10.89	6.96	5.46	4.65	4.14	3.78	3.51	3.30	3.13	2.99	2.77	2.61	2.48	2.38	2.30

TABLE A.3 (continued)

***F* Distribution: Critical Values of *F* (0.1% significance level)**

v_2 \ v_1	25	30	35	40	50	60	75	100	150	200
1	6.24e05	6.26e05	6.28e05	6.29e05	6.30e05	6.31e05	6.32e05	6.33e05	6.35e05	6.35e05
2	999.46	999.47	999.47	999.47	999.48	999.48	999.49	999.49	999.49	999.49
3	125.84	125.45	125.17	124.96	124.66	124.47	124.27	124.07	123.87	123.77
4	45.70	45.43	45.23	45.09	44.88	44.75	44.61	44.47	44.33	44.26
5	25.08	24.87	24.72	24.60	24.44	24.33	24.22	24.12	24.01	23.95
6	16.85	16.67	16.54	16.44	16.31	16.21	16.12	16.03	15.93	15.89
7	12.69	12.53	12.41	12.33	12.20	12.12	12.04	11.95	11.87	11.82
8	10.26	10.11	10.00	9.92	9.80	9.73	9.65	9.57	9.49	9.45
9	8.69	8.55	8.46	8.37	8.26	8.19	8.11	8.04	7.96	7.93
10	7.60	7.47	7.37	7.30	7.19	7.12	7.05	6.98	6.91	6.87
11	6.81	6.68	6.59	6.52	6.42	6.35	6.28	6.21	6.14	6.10
12	6.22	6.09	6.00	5.93	5.83	5.76	5.70	5.63	5.56	5.52
13	5.75	5.63	5.54	5.47	5.37	5.30	5.24	5.17	5.10	5.07
14	5.38	5.25	5.17	5.10	5.00	4.94	4.87	4.81	4.74	4.71
15	5.07	4.95	4.86	4.80	4.70	4.64	4.57	4.51	4.44	4.41
16	4.82	4.70	4.61	4.54	4.45	4.39	4.32	4.26	4.19	4.16
17	4.60	4.48	4.40	4.33	4.24	4.18	4.11	4.05	3.98	3.95
18	4.42	4.30	4.22	4.15	4.06	4.00	3.93	3.87	3.80	3.77
19	4.26	4.14	4.06	3.99	3.90	3.84	3.78	3.71	3.65	3.61
20	4.12	4.00	3.92	3.86	3.77	3.70	3.64	3.58	3.51	3.48
21	4.00	3.88	3.80	3.74	3.64	3.58	3.52	3.46	3.39	3.36
22	3.89	3.78	3.70	3.63	3.54	3.48	3.41	3.35	3.28	3.25
23	3.79	3.68	3.60	3.53	3.44	3.38	3.32	3.25	3.19	3.16
24	3.71	3.59	3.51	3.45	3.36	3.29	3.23	3.17	3.10	3.07
25	3.63	3.52	3.43	3.37	3.28	3.22	3.15	3.09	3.03	2.99
26	3.56	3.44	3.36	3.30	3.21	3.15	3.08	3.02	2.95	2.92
27	3.49	3.38	3.30	3.23	3.14	3.08	3.02	2.96	2.89	2.86
28	3.43	3.32	3.24	3.18	3.09	3.02	2.96	2.90	2.83	2.80
29	3.38	3.27	3.18	3.12	3.03	2.97	2.91	2.84	2.78	2.74
30	3.33	3.22	3.13	3.07	2.98	2.92	2.86	2.79	2.73	2.69
35	3.13	3.02	2.93	2.87	2.78	2.72	2.66	2.59	2.52	2.49
40	2.98	2.87	2.79	2.73	2.64	2.57	2.51	2.44	2.38	2.34
50	2.79	2.68	2.60	2.53	2.44	2.38	2.31	2.25	2.18	2.14
60	2.67	2.55	2.47	2.41	2.32	2.25	2.19	2.12	2.05	2.01
70	2.58	2.47	2.39	2.32	2.23	2.16	2.10	2.03	1.95	1.92
80	2.52	2.41	2.32	2.26	2.16	2.10	2.03	1.96	1.89	1.85
90	2.47	2.36	2.27	2.21	2.11	2.05	1.98	1.91	1.83	1.79
100	2.43	2.32	2.24	2.17	2.08	2.01	1.94	1.87	1.79	1.75
120	2.37	2.26	2.18	2.11	2.02	1.95	1.88	1.81	1.73	1.68
150	2.32	2.21	2.12	2.06	1.96	1.89	1.82	1.74	1.66	1.62
200	2.26	2.15	2.07	2.00	1.90	1.83	1.76	1.68	1.60	1.55
250	2.23	2.12	2.03	1.97	1.87	1.80	1.72	1.65	1.56	1.51
300	2.21	2.10	2.01	1.94	1.85	1.78	1.70	1.62	1.53	1.48
400	2.18	2.07	1.98	1.92	1.82	1.75	1.67	1.59	1.50	1.45
500	2.17	2.05	1.97	1.90	1.80	1.73	1.65	1.57	1.48	1.43
600	2.16	2.04	1.96	1.89	1.79	1.72	1.64	1.56	1.46	1.41
750	2.15	2.03	1.95	1.88	1.78	1.71	1.63	1.55	1.45	1.40
1000	2.14	2.02	1.94	1.87	1.77	1.69	1.62	1.53	1.44	1.38

Table A.4

χ^2 (Chi-Squared) Distribution: Critical Values of χ^2

	Significance level		
Degrees of freedom	5%	1%	0.1%
1	3.841	6.635	10.828
2	5.991	9.210	13.816
3	7.815	11.345	16.266
4	9.488	13.277	18.467
5	11.070	15.086	20.515
6	12.592	16.812	22.458
7	14.067	18.475	24.322
8	15.507	20.090	26.124
9	16.919	21.666	27.877
10	18.307	23.209	29.588

.

Thống kê bằng phần mềm R

I, Bài toán ước lượng tham số

1, Tìm khoảng tin cậy cho kỳ vọng

1.1, Phân phối chuẩn, phương sai chưa biết

x = c(165, 166, 168, 159, 149, 167, 174, 158, 168)

t.test(x)

One Sample t-test

data: x

t = 66.896, df = 8, p-value = 2.775e-12

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

158.1322 169.4234

sample estimates:

mean of x

163.7778

- Thay đổi độ tin cậy: ***conf.level***

t.test(x, conf.level = 0.99)

- Tìm khoảng tin cậy 1 phía: ***alternative***

t.test(x, alternative = "less")

t.test(x, conf.level = 0.99, alternative = "greater")

One Sample t-test

data: x

t = 66.896, df = 8, p-value = 1.387e-12

alternative hypothesis: true mean is greater than 0

99 percent confidence interval:

156.6866   Inf

sample estimates:

mean of x

163.7778

- Giải thích kết quả:

+ Kiểm định t-test một mẫu với dữ liệu là x

+ Test Thống kê t = 66.896, bậc tự do df = 8, p-value = 1.387e-12

+ Đối thuyết H_1: Giá trị trung bình của x lớn hơn 0

+ Khoảng tin cậy 99% cho giá trị trung bình của x là (156.6866, Inf)

+ Giá trị trung bình của x là 163.7778

+ Bài toán kiểm định: $H_0: \mu = 0$, $H_1: \mu > 0$

+ Do pvalue = 1.387e-12 < 0.01 nên bác bỏ H_0

+ Kết luận: Với mức ý nghĩa 1% có cơ sở để nói rằng giá trị trung bình của x lớn hơn 0

1.2 Phân phối chuẩn, phương sai đã biết

- Sử dụng hàm z.test trong gói lệnh bsda

```
install.packages("BSDA")
library(BSDA)
z.test(x, sigma.x = 0.5)
z.test(x, sigma.x = 0.5, conf.level = 0.9)
z.test(x, sigma.x = 0.5, alternative = "less")
```

One-sample z-Test

data: x

z = 982.67, p-value = 1

alternative hypothesis: true mean is less than 0

95 percent confidence interval:

NA 164.0519

sample estimates:

mean of x

163.7778

- Giải thích kết quả:

+ Kiểm định z-test một mẫu với dữ liệu là x

+ Test Thống kê: z = 982.67, bậc tự do df = 8, p-value = 1

+ Đối thuyết H_1: Giá trị trung bình của x nhỏ hơn 0

+ Khoảng tin cậy 95% cho giá trị trung bình của x là (NA, 164.0519)

+ Giá trị trung bình của x là 163.7778

+ Bài toán kiểm định: H_0: $\mu = 0$, H_1: $\mu < 0$

+ Do p_value = 1 > 0,05 nên chấp nhận H_0

+ Kết luận: Với mức ý nghĩa 5% không có cơ sở để nói rằng giá trị trung bình của x lớn hơn 0

2, Tìm khoảng tin cậy cho tỉ lệ

- Cú pháp: ***prop.test(k,n)***: n là số lần quan sát, k là số lần xuất hiện biến cố A

prop.test(83, 100)

prop.test(83, 100, conf.level = 0.9, alternative = "greater")

prop.test(83, 100, conf.level = 0.99, alternative = "less")

1-sample proportions test with continuity correction

data: 83 out of 100, null probability 0.5

X-squared = 42.25, df = 1, p-value = 1

alternative hypothesis: true p is less than 0.5

99 percent confidence interval:

0.0000000 0.9036452

sample estimates:

p

0.83

- Giải thích kết quả:

+ Kiểm định tỉ lệ một mẫu với dữ liệu là 83 lần xảy ra biến cố A trong 100 lần thực hiện phép thử

+ Test Thống kê: X-squared = 42.25, bậc tự do df = 1, p-value = 1

+ Đối thuyết H_1: p < 0,5

+ Khoảng tin cậy 99% cho p là (0.0000000, 0.9036452)

+ Giá trị cuả p từ mẫu là 0,83

+ Bài toán kiểm định: H_0: p = 0.5, H_1: p < 0.5

+ Do p_value = 1 > 0.01 nên chấp nhận H_0

+ Kết luận: Với mức ý nghĩa 1% không có cơ sở để nói rằng p lớn hơn 0.01

II, Kiểm định giả thiết thống kê

1, Kiểm định cho giá trị trung bình

attach(cars)

t.test(dist, mu = 45, conf.level = 0.9, alternative = "greater")

One Sample t-test

data: dist

t = -0.55428, df = 49, p-value = 0.709

alternative hypothesis: true mean is greater than 45

90 percent confidence interval:

38.24575 Inf

sample estimates:

mean of x

42.98

- Giải thích kết quả:

+ Kiểm định t-test một mẫu với dữ liệu là dist

+ Test Thống kê t = -0.55428, bậc tự do df = 35, pvalue = 0.709

+ Đối thuyết H_1: Giá trị trung bình của dist lớn hơn 45

+ Khoảng tin cậy 90% cho giá trị trung bình của dist là (38.24575, Inf)

+ Giá trị trung bình mẫu của dist là 42.98

+ Bài toán: H_0: $\mu = 45$, H_1: $\mu > 45$

+ Do p-value =0.709 > 0.1 nên chấp nhận H_0

+ Kết luận: Với mức ý nghĩa 10% không có cơ sở để nói rằng giá trị trung bình của dist lớn hơn 45

2, Kiểm định cho tỉ lệ

prop.test(length(speed[speed < 13]), length(speed), 0.4, conf.level = 0.98, alternative = "greater")

1-sample proportions test with continuity correction

data: length(speed[speed < 13]) out of length(speed), null probability 0.4

X-squared = 1.6875, df = 1, p-value = 0.903

alternative hypothesis: true p is greater than 0.4

98 percent confidence interval:

0.1787247 1.0000000

sample estimates:

p

0.3

- Giải thích kết quả:

+ Kiểm định tỉ lệ 1 mẫu với bộ dữ liệu là speed

+ Test Thống kê X-squared = 1.6875, bậc tự do df = 1, pvalue = 0.903

+ Gọi p là tỷ lệ số xe có vận tốc nhỏ hơn 13

+ Đối thuyết H_1: p > 0.4

+ Khoảng tin cậy 98% cho p là (0.1787247, 1.0000000)

+ Tỷ lệ số xe có vận tốc nhỏ hơn 13 trong mẫu là 0.3

+ Bài toán: H_0: p = 0.4, H_1: p > 0.4

+ Do p-value =0.903 > 0.02 nên chấp nhận H_0

+ Kết luận: Với mức ý nghĩa 2% không có cơ sở để nói rằng tỷ lệ số xe có tốc độ nhỏ hơn 13 lớn hơn 0.4

III, Bài toán so sánh

1, So sánh hai giá trị trung bình

- Nếu X và Y có phân phối chuẩn thì dùng t.test()

- Nếu X và Y chưa biết phân phối thì dùng wilcox.test()

1.1, X và Y độc lập

data1 = rnorm(100, 165, 5)

data2 = rnorm(100, 163, 8)

chisq.test(data1, data2)

Pearson's Chi-squared test

data: data1 and data2

X-squared = 9900, df = 9801, p-value = 0.239

- Giải thích kết quả:

+ Kiểm định sự độc lập với dữ liệu là data1 và data2

+ Test thống kê: X-squared = 9900, bậc tự do df = 9801, p-value = 0.239

+ Bài toán: H_0: data1 và data2 độc lập, H_1: data1 và data2 phụ thuộc

+ Do p-value = 0.239 > 0.05 nên chấp nhận H_0.

+ Kết luận: data1 và data2 độc lập

shapiro.test(data1)

Shapiro-Wilk normality test

data: data1

W = 0.98766, p-value = 0.4837

Giải thích kết quả:

+ Kiểm định phân phối chuẩn với dữ liệu là data1

+ Test thống kê: W = 0.98766, p-value = 0.4837

+ Bài toán: H_0: data1 tuân theo phân phối chuẩn, H_1: data1 không tuân theo phân phối chuẩn

+ Do p-value = 0.4837 > 0.05 nên chấp nhận H_0

+ Kết luận: data1 tuân theo phân phối chuẩn

t.test(data1, data2)

Welch Two Sample t-test

data: data1 and data2

t = 4.9344, df = 167.03, p-value = 1.932e-06

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

2.798166 6.530668

sample estimates:

mean of x mean of y

165.5651 160.9007

- Giải thích kết quả:

+ Kiểm định t_test 2 mẫu, dữ liệu là data1 và data2

+ Test Thống kê: t = 2.5679, bậc tự do df = 168.31, p_value = 0.011

+ Gọi μ_1 và μ_2 lần lượt là giá trị trung bình tổng thể của data1 và data2

+ Đối thuyết H_1: μ_1 khác μ_2

+ Khoảng tin cậy 95% cho μ_1-μ_2 là (0.5871073, 4.4913553)

+ Giá trị trung bình mẫu của data1 và data2 lần lượt là 165.0587, 162.5195

+ Bài toán: H_0: $\mu_1 = \mu_2$, H_1: μ_1 khác μ_2

+ Do p_value = 0.011 < 0.05 nên bác bỏ H_0

+ Kết luận: Với mức ý nghĩa 5% có thể nói rằng có sự khác nhau về giá trị trung bình của data1 và data2

- Thay đổi độ tin cậy và sự so sánh dùng ***conf.level*** và ***alternative***

t.test(data1, data2, alternative = "greater", conf.level = 0.99)

Welch Two Sample t-test

data: data1 and data2

t = 4.9344, df = 167.03, p-value = 9.658e-07

alternative hypothesis: true difference in means is greater than 0

99 percent confidence interval:

2.44405 Inf

sample estimates:

mean of x mean of y

165.5651 160.9007

- Giải thích kết quả:

+ Kiểm định t_test 2 mẫu, dữ liệu là data1 và data2

+ Test Thống kê: t = 4.9344, df = 167.03, p-value = 9.658e-07

+ Gọi μ_1 *và* μ_2 *lần lượt là giá trị trung bình tổng thể của data1 và data2*

+ Đối thuyết H_1*:* μ_1 *khác* μ_2

+ Khoảng tin cậy 99% cho μ_1-μ_2 *là (2.44405, Inf)*

+ Giá trị trung bình mẫu của data1 và data2 lần lượt là 165.0587, 162.5195

+ Bài toán: H_0: $\mu_1 = \mu_2$, H_1: $\mu_1 > \mu_2$

+ Do p_value = 9.658e-07 < 0.05 nên bác bỏ H_0

+ Kết luận: Với mức ý nghĩa 5% có thể nói rằng giá trị trung bình của data1 lớn hơn giá trị trung bình của data2.

- Nếu X và Y không tuân theo phân phối chuẩn thì dùng ***wilcox.test()***

x = c(161, 150, 122, 135, 133, 199, 2000, 1010, 2, 44)

y = c(1, 1000, 22, 834, 126, 44)

shapiro.test(x)

Shapiro-Wilk normality test

data: x

W = 0.61295, p-value = 7.893e-05

shapiro.test(y)

Shapiro-Wilk normality test

data: y

W = 0.74826, p-value = 0.01916

wilcox.test(x, y, alternative = "greater", conf.level = 0.9)

Wilcoxon rank sum test with continuity correction

data: x and y

W = 38.5, p-value = 0.1926

alternative hypothesis: true location shift is greater than 0

- Giải thích kết quả:

+ Kiểm định wilcox.test 2 mẫu, dữ liệu là x và y

+ Test Thống kê: W = 38.5, p-value = 0.1926

+ Gọi μ_1 và μ_2 lần lượt là giá trị trung bình tổng thể của x và y

+ Bài toán: H_0: $\mu_1 = \mu_2$, H_1: $\mu_1 > \mu_2$

+ Do p_value = 0.1926 > 0.1 nên chấp nhận H_0

+ Kết luận: Với mức ý nghĩa 1% không thể nói rằng giá trị trung bình của x lớn hơn giá trị trung bình của y.

1.2, X và Y phụ thuộc

before = c(190, 202, 177, 160, 225, 180, 196, 208, 185, 177)

after = c(185, 197, 185, 152, 205, 184, 184, 200, 187, 170)

- before: trọng lượng trước ăn kiêng

- after: trọng lượng sau ăn kiêng

shapiro.test(before)

Shapiro-Wilk normality test

data: before

W = 0.97881, p-value = 0.9585

+ Vì p_value = 0.9585 > 0.05 nên before tuân theo phân bố chuẩn

shapiro.test(after)

Shapiro-Wilk normality test

data: after

W = 0.90298, p-value = 0.2362

+ Vì p_value = 0.2362 > 0.05 nên after tuân theo phân bố chuẩn

t.test(before, after, paired = T, alternative = "greater")

Paired t-test

data: before and after

t = 1.9837, df = 9, p-value = 0.0393

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

0.3870813   Inf

sample estimates:

mean of the differences

5.1

- Giải thích kết quả:

+ Kiểm định t-test 2 mẫu theo cặp với dữ liệu là before và after

+ Test thống kê: t = 1.9837, df = 9, p-value = 0.0393

+ Đặt D = before – after

+ Đối thiết H_1: $\mu_D > 0$

+ Khoảng tin cậy 95% cho μ_D *là (0.3870813,   Inf)*

+ Bài toán: H_0: $\mu_D = 0$, H_1: $\mu_D > 0$

+ Vì p_value = 0.0393 < 0.05 nên bác bỏ H_0

+ Kết luận: Với mức ý nghĩa 5% có thể kết luận chế độ ăn kiêng có tác dụng giảm trọng lượng

2. So sánh 2 tỉ lệ

```
yes = c(301, 729)

sum = c(1432, 1656)

prop.test(yes, sum, conf.level = 0.99)
```

```
2-sample test for equality of proportions with continuity correction


data:  yes out of sum

X-squared = 181.75, df = 1, p-value < 2.2e-16

alternative hypothesis: two.sided

99 percent confidence interval:

 -0.2725838 -0.1874599

sample estimates:

   prop 1    prop 2

0.2101955 0.4402174
```

- Giải thích kết quả:

+ So sánh 2 tỉ lệ với dữ liệu là yes và sum

+ Test thống kê: X-squared = 181.75, df = 1, p-value < 2.2e-16

+ Gọi p_1 và p_2 lần lượt là tỉ lệ số người trả lời có năm ngoái và số người trả lời có năm nay

+ Đối thiết: H_1: p_1 khác p_2

+ Khoảng tin cậy 99% cho $p_1 - p_2$ là (-0.2725838 -0.1874599)

+ Tỉ lệ mẫu số người trả lời có năm ngoái và năm nay lần lượt là: 0.2101955 và 0.4402174

+ Bài toán: H_0: $p_1 = p_2$, H_1: p_1 khác p_2

+ Vì p-value = 2.2e-16 < 0.05 nên bác bỏ H_0

+ Kết luận: Vậy với mức ý nghĩa 1% có cơ sở để nói rằng tỉ lệ số người trả lời có năm ngoái và năm nay khác nhau.

3. So sánh nhiều tỉ lệ

hsg = c(45, 79, 20)

hs = c(120, 200, 90)

prop.test(hsg, hs)

3-sample test for equality of proportions without continuity correction

data: hsg out of hs

X-squared = 8.5526, df = 2, p-value = 0.01389

alternative hypothesis: two.sided

sample estimates:

prop 1 prop 2 prop 3

0.3750000 0.3950000 0.2222222

- Giải thích kết quả:

+ So sánh 3 tỉ lệ với dữ liệu là hsg và hs

+ Test thống kê: X-squared = 8.5526, df = 2, p-value = 0.01389

+ Gọi p_1, p_2 *và* p_3 *lần lượt là tỉ lệ số học sinh giỏi trong các năm 2020, 2021, 2022*

+ Đối thiết: H_1: p_1, p_2 *và* p_3 *không bằng nhau*

+ Tỉ lệ mẫu số số học sinh giỏi trong các năm 2020, 2021, 2022 lần lượt là 0.3750000, 0.3950000 và 0.2222222

+ Bài toán: H_0: $p_1 = p_2 = p_3$, H_1: p_1, p_2 *và* p_3 *không bằng nhau*

+ Vì p-value = 0.01389 < 0.05 nên bác bỏ H_0

+ Kết luận: Vậy với mức ý nghĩa 5% có cơ sở để nói rằng tỉ lệ số số học sinh giỏi trong các năm 2020, 2021, 2022 là khác nhau.

4, Kiểm định sự phù hợp

Thực hiện tung 12 con xúc xắc 26306 lần. Thành công nếu xúc xắc ra mặt 5 hoặc 6 => Xác suất thành công là p = 1/3. Gọi X là số lần thành công trong 12 lần. Khi X tuân theo phân phối nhị thức:

k = 0:12

p = dbinom(k, 12, 1/3) #p(x=k) khi x ~ B(12, 1/3)

p

[1] 7.707347e-03 4.624408e-02 1.271712e-01 2.119520e-01 2.384460e-01 1.907568e-01 1.112748e-01 4.768921e-02

[9] 1.490288e-02 3.311751e-03 4.967626e-04 4.516023e-05 1.881676e-06

Giải thích kết quả:

+ *Xác suất để 0 lần thành công: 7.707347e-03*

+ *Xác suất để 1 lần thành công: 4.624408e-02*

+ ....

+ *Xác suất để 12 lần thành công là: 1.881676e-06*

Binom = round(26306 * p)

Binom

[1] 203 1216 3345 5576 6273 5018 2927 1255 392   87   13   1   0

Giải thích kết quả:

+ *Khi thực hiện 26306 lần có 203 lần không ra mặt 5 hoặc 6*

+ *Khi thực hiện 26306 lần có 1216 lần không ra mặt 5 hoặc 6*

+ ...

+ *Khi thực hiện 26306 lần có 0 lần không ra mặt 5 hoặc 6*

names(Binom) = k

Binom #Lý thuyết

0   1   2   3   4   5   6   7   8   9   10   11   12

203 1216 3345 5576 6273 5018 2927 1255 392   87   13   1   0

Weldon = c(185, 1149, 3265, 5475, 6114, 5294, 3067, 1131, 405, 103, 14, 4, 0) #Thực tế

names(Weldon) = k

Weldon

0   1   2   3   4   5   6   7   8   9   10   11   12

185 1149 3265 5475 6114 5294 3067 1131 405 103   14   4   0

data.frame(Binom, Weldon, Diff = Weldon - Binom)

counts = cbind(Binom, Weldon)

cweldon = c(Weldon[1:10], sum(Weldon[11:13])) #Gộp 3 giá trị cuối thành 1 giá trị

cweldon

0   1   2   3   4   5   6   7   8   9

185 1149 3265 5475 6114 5294 3067 1131 405 103   18

cbinom = c(Binom[1:10], sum(Binom[11:13])) #Gộp 3 giá trị cuối thành 1 giá trị

cbinom

0   1   2   3   4   5   6   7   8   9

203 1216 3345 5576 6273 5018 2927 1255 392   87   14

probs = c(p[1:10], 1 - sum(p[1:10]))

probs #Xác suất theo lý thuyết khi số lần thành công tuân theo phân phối nhị thức

[1] 0.0077073466 0.0462440798 0.1271712194 0.2119520323 0.2384460363 0.1907568291 0.1112748170 0.0476892073

[9] 0.0149028773 0.0033117505 0.0005438045

chisq.test(cweldon, p = probs)

Chi-squared test for given probabilities

data: cweldon

X-squared = 51.158, df = 10, p-value = 1.633e-07

- Giải thích kết quả:

+ Giả thiết H_0: X tuân theo phân phối nhị thức

+ Đối thiết H_1: X không tuân theo phân phối nhị thức

+ Do p_value = 1.633e-07 < 0.05 nên bác bỏ H_0

+ Kết luận: X không tuân theo phân phối nhị thức

IV, Phân tích phương sai một nhân tố

Bảng dưới đây so sánh độ galactoso trong 3 nhóm bệnh nhân

- Nhóm 1 gồm 9 bệnh nhân với bệnh nhân Crohn
- Nhóm 2 gồm 11 bệnh nhân với bệnh ruột kết (coltins)
- Nhóm 3 gồm 20 đối tượng không có bệnh (gọi là nhóm đối chứng)

Câu hỏi đặt ra là độ galactoso giữa 3 nhóm bệnh nhân có khác nhau hay không

Giả thiết: H_0: $\mu_1 = \mu_2 = \mu_3$

Đối thiết: H_1: Có sự khác biệt giữa giá trị trung bình 3 nhóm

group = c(rep(1,9), rep(2, 11), rep(3,20))

group = as.factor(group) #Định nghĩa group là 1 nhân tố

galactoso = c(1343, 1393, 1420, 1641, 1897, 2160, 2169, 2279, 2890, 1264, 1314, 1399, 1605, 2385, 2511, 2514, 2767, 2827, 2895, 3011, 1809, 2850, 1926, 2964, 2283, 2973, 2384, 3171, 2447, 3257, 2479, 3271, 2495, 3288, 2525, 3358, 2541, 3643, 2769, 3657)

data = data.frame(group, galactoso)

View(data)

	group	galactoso
1	1	1343
2	1	1393
3	1	1420
4	1	1641
5	1	1897
6	1	2160
7	1	2169
8	1	2279
9	1	2890
10	2	1264
11	2	1314
12	2	1399
13	2	1605
14	2	2385
15	2	2511
16	2	2514
17	2	2767
18	2	2827
19	2	2895
20	2	3011
21	3	1809
22	3	2850
23	3	1926
24	3	2964
25	3	2283

- Phân tích phương sai (galactoso là 1 hàm của group)

analysis = lm(galactoso ~ group)

anova(analysis)

Analysis of Variance Table

Response: galactoso

```
          Df   Sum Sq Mean Sq F value    Pr(>F)
group      2  5683620 2841810  8.6655 0.0008191 ***
Residuals 37 12133923  327944
```

- Giải thích kết quả

+ Giả thiết: H_0: $\mu_1 = \mu_2 = \mu_3$

+ Đối thiết: H_1: Có sự khác biệt giữa giá trị trung bình 3 nhóm

+ Do p_value = 0.00088191 < 0.05 nên bác bỏ H_0

+ Kết luận: Có thể nói có sự khác biệt về độ galactoso giữa 3 nhóm bệnh nhân

Cách 2:

```
res = aov(galactoso ~ group)
summary(res)
            Df   Sum Sq Mean Sq F value   Pr(>F)
group        2  5683620 2841810   8.666 0.000819 ***
Residuals   37 12133923  327944
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
TukeyHSD(res)
Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = galactoso ~ group)

$group
        diff        lwr        upr     p adj
2-1 316.3232 -312.09857   944.745 0.4439821
3-1 894.2778  333.07916  1455.476 0.0011445
3-2 577.9545   53.11886  1102.790 0.0281768
```

- Giải thích kết quả:

+ Khoảng tin cậy 95% cho sự khác biệt giá trị trung bình giữa nhóm 2 và nhóm 1 là (-312.09857, 944.745)

+ Khoảng tin cậy 95% cho sự khác biệt giá trị trung bình giữa nhóm 3 và nhóm 1 là (333.07916, 1455.476)

+ Khoảng tin cậy 95% cho sự khác biệt giá trị trung bình giữa nhóm 3 và nhóm 2 là (53.11886, 1102.790)

V, Phân tích tương quan và hồi quy

1, Hệ số tương quan

```
data = read.csv("C:\\Users\\thoidaipc\\downloads\\bmi.csv",header = T)
attach(data)
chol = data$Chol
age = data$Age
cor(age, chol) # ( = cor(chol, age)), -1 < r < 1
[1] 0.9367261 # r > 0.8 => Hệ số tương quan mạnh
cor.test(age, chol)
Pearson's product-moment correlation

data:  age and chol
t = 10.704, df = 16, p-value = 1.058e-08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8350463 0.9765306
sample estimates:
      cor
0.9367261
```

<u>- Giải thích kết quả:</u>

+ Kiểm định hệ số tương quan với dữ liệu là age và chol

+ Test thống kê: t = 10.704, df = 16, p-value = 1.058e-08

+ Đối thiết H_1: ρ khác 0

+ Khoảng tin cậy 95% cho ρ là (0.8350463, 0.9765306)

+ Hệ số tương quan mẫu r = 0.9367261

+ Bài toán: Giả thiết H_0: $\rho = 0$, Đối thiết H_1: ρ khác 0

+ Do p-value = 1.058e-08 < 0.05 nên bác bỏ H_0

+ Kết luận: Với mức ý nghĩa 5% có thể nói rằng hệ số tương quan giữa age và chol khác 0

2. Phân tích hồi quy tuyến tính

lm(chol ~ age)

```
Call:
lm(formula = chol ~ age)

Coefficients:
(Intercept)          age
    1.08922      0.05779
```

- Giải thích kết quả:

*+ Mô hình hồi quy tuyến tính của chol theo age: chol = 0.05779*age + 1.08922*

model =lm(chol ~ age)

summary(model)

```
Call:
lm(formula = chol ~ age)

Residuals:
     Min       1Q   Median       3Q      Max
-0.40729 -0.24133 -0.04522  0.17939  0.63040

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.089218   0.221466   4.918 0.000154 ***
age         0.057788   0.005399  10.704 1.06e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual standard error: 0.3027 on 16 degrees of freedom

Multiple R-squared: 0.8775,   Adjusted R-squared: 0.8698

F-statistic: 114.6 on 1 and 16 DF, p-value: 1.058e-08

- Giải thích kết quả:

*+ chol = A*age + B + E*

+ Phần dư Residuals: E ~ N(0, σ^2)

+ Bài toán: Giả thiết H_0: A = 0, Đối thiết H_1: A khác 0

+ Do p-value = 1.06e-08 < 0.05 nên bác bỏ H_0. Vậy A khác 0

+ Bài toán: Giả thiết H_0: B = 0, Đối thiết H_1: B khác 0

+ Do p-value = 0.000154 < 0.05 nên bác bỏ H_0. Vậy B khác 0

re = model$residuals

shapiro.test(re)

Shapiro-Wilk normality test

data: re

W = 0.9493, p-value = 0.4142

- Giải thích kết quả:

+ Kiểm tra xem phần dư có tuân theo phân phối chuẩn hay không

+ Giả thiết H_0: Phần dư tuân theo phân phối chuẩn

+ Đối thiết H_1: Phần dư không tuân theo phân phối chuẩn

+ Do p-value = 0.4142 > 0.05 nên chấp nhận H_0

+ Kết luận: Phần dư tuân theo phân phối chuẩn

t.test(re)

One Sample t-test

data: re

t = 2.7851e-17, df = 17, p-value = 1

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

-0.1460145 0.1460145

sample estimates:

mean of x

1.927471e-18

- Giải thích kết quả:

+ *Kiểm tra xem trung bình phần dư có bằng 0 hay không*

+ *Giả thiết H_0: $\mu_E = 0$*

+ *Đối thiết H_1: μ_E khác 0*

+ *Do p-value = 1 > 0.05 nên chấp nhận H_0*

+ *Kết luận: Trung bình phần dư tuân theo phân phối chuẩn*