

# Lý thuyết Xác suất và Thống kê Toán học

PGS.TS. Trần Lộc Hùng

Tp. Hồ Chí Minh, 10/ 2013

Ngày 27 tháng 10 năm 2013

TRƯỜNG ĐẠI HỌC TÀI CHÍNH-MARKETING  
**KHOA CƠ BẢN, BỘ MÔN TOÁN-THỐNG KÊ**

— — — — —

PGS. TS. TRẦN LỘC HÙNG

LÝ THUYẾT XÁC SUẤT VÀ THỐNG KÊ TOÁN HỌC

Tp. Hồ Chí Minh, 10/2013

# Lý thuyết Xác suất và Thống kê Toán học

PGS.TS. Trần Lộc Hùng

Tp. Hồ Chí Minh, 10/ 2013

Ngày 27 tháng 10 năm 2013

## Chương 8. Kiểm định phi tham số

# Từ khóa (Key Words)

- Kiểm định phi tham số

# Từ khóa (Key Words)

- Kiểm định phi tham số
- Cặp giả thuyết thống kê

# Từ khóa (Key Words)

- Kiểm định phi tham số
- Cặp giả thuyết thống kê
- Hàm kiểm định

# Từ khóa (Key Words)

- Kiểm định phi tham số
- Cặp giả thuyết thống kê
- Hàm kiểm định
- Mức ý nghĩa

# Từ khóa (Key Words)

- Kiểm định phi tham số
- Cặp giả thuyết thống kê
- Hàm kiểm định
- Mức ý nghĩa
- Miền bác bỏ



# Từ khóa (Key Words)

- Kiểm định phi tham số
- Cặp giả thuyết thống kê
- Hàm kiểm định
- Mức ý nghĩa
- Miền bác bỏ
- Miền chấp nhận

## Chương 8. Kiểm định phi tham số

### 1 Đặt vấn đề

## Chương 8. Kiểm định phi tham số

- 1 Đặt vấn đề
- 2 Kiểm định sự phù hợp giữa thực nghiệm và lý thuyết

## Chương 8. Kiểm định phi tham số

- 1 Đặt vấn đề
- 2 Kiểm định sự phù hợp giữa thực nghiệm và lý thuyết
- 3 Kiểm định tính độc lập

## Chương 8. Kiểm định phi tham số

- 1 Đặt vấn đề
- 2 Kiểm định sự phù hợp giữa thực nghiệm và lý thuyết
- 3 Kiểm định tính độc lập
- 4 Kiểm định nhiều tỷ lệ

## Chương 8. Kiểm định phi tham số

- 1 Đặt vấn đề
- 2 Kiểm định sự phù hợp giữa thực nghiệm và lý thuyết
- 3 Kiểm định tính độc lập
- 4 Kiểm định nhiều tỷ lệ
- 5 Phương pháp hạng Mann Whitney (đọc thêm)

## Chương 8. Kiểm định phi tham số

- 1 Đặt vấn đề
- 2 Kiểm định sự phù hợp giữa thực nghiệm và lý thuyết
- 3 Kiểm định tính độc lập
- 4 Kiểm định nhiều tỷ lệ
- 5 Phương pháp hạng Mann Whitney (đọc thêm)
- 6 Phương pháp dấu Wilcoxon (đọc thêm)

## Chương 8. Kiểm định phi tham số

- 1 Đặt vấn đề
- 2 Kiểm định sự phù hợp giữa thực nghiệm và lý thuyết
- 3 Kiểm định tính độc lập
- 4 Kiểm định nhiều tỷ lệ
- 5 Phương pháp hạng Mann Whitney (đọc thêm)
- 6 Phương pháp dấu Wilcoxon (đọc thêm)
- 7 Bài tập



## 8.1 Đặt vấn đề

Giả sử  $\omega_n = \{X_1, X_2, \dots, X_n\}$  là một mẫu ngẫu nhiên sinh bởi biến ngẫu nhiên  $X$  nhưng chưa xác định được quy luật xác suất  $F(x, \theta)$  của  $X$ . Đây là sự khác biệt cơ bản giữa kiểm định tham số và kiểm định phi tham số. Vấn đề đặt ra là:

- 1 Kiểm định giả thuyết thống kê về sự phù hợp giữa các dữ liệu nhận được và quy luật  $F(x, \theta)$ .

## 8.1 Đặt vấn đề

Giả sử  $\omega_n = \{X_1, X_2, \dots, X_n\}$  là một mẫu ngẫu nhiên sinh bởi biến ngẫu nhiên  $X$  nhưng chưa xác định được quy luật xác suất  $F(x, \theta)$  của  $X$ . Đây là sự khác biệt cơ bản giữa kiểm định tham số và kiểm định phi tham số. Vấn đề đặt ra là:

- 1 Kiểm định giả thuyết thống kê về sự phù hợp giữa các dữ liệu nhận được và quy luật  $F(x, \theta)$ .
- 2 Xây dựng thuật toán (các thủ tục) để kiểm định

# Vấn đề khác

Giả sử  $X$  và  $Y$  là hai biến ngẫu nhiên, có các đặc tính  $x_1, x_2, \dots, x_m$  và  $y_1, y_2, \dots, y_n$ . Vấn đề đặt ra là:

- 1 Liệu các biến ngẫu nhiên  $X$  và  $Y$  có độc lập hay không?

# Vấn đề khác

Giả sử  $X$  và  $Y$  là hai biến ngẫu nhiên, có các đặc tính  $x_1, x_2, \dots, x_m$  và  $y_1, y_2, \dots, y_n$ . Vấn đề đặt ra là:

- ❶ Liệu các biến ngẫu nhiên  $X$  và  $Y$  có độc lập hay không?
- ❷ Một vài ví dụ

## 8.2 Kiểm định tính phù hợp giữa thực nghiệm và lý thuyết

### 1 Bài toán

## 8.2 Kiểm định tính phù hợp giữa thực nghiệm và lý thuyết

- 1 Bài toán
- 2 Tiêu chuẩn khi-bình phương (Quy tắc Pearson)

## 8.2 Kiểm định tính phù hợp giữa thực nghiệm và lý thuyết

- 1 Bài toán
- 2 Tiêu chuẩn khi-bình phương (Quy tắc Pearson)
- 3 Thủ tục kiểm định

## 8.2 Kiểm định tính phù hợp giữa thực nghiệm và lý thuyết

- 1 Bài toán
- 2 Tiêu chuẩn khi-bình phương (Quy tắc Pearson)
- 3 Thủ tục kiểm định
- 4 Ví dụ



# Kiểm định tính phù hợp giữa thực nghiệm và lý thuyết

## Bài toán 1

Giả sử  $\omega_n = \{X_1, X_2, \dots, X_n\}$  là một mẫu dữ liệu sinh từ biến ngẫu nhiên  $X$ . Liệu có thể nói dữ liệu nhận được có phù hợp với quy luật xác suất  $F(x, \theta)$  của biến  $X$  hay không?

Tức là

- Kiểm định giả thuyết thống kê  $(H_0 : X \in F(x, \theta) \mid \overline{H_0} : X \notin F(x, \theta))$

# Kiểm định tính phù hợp giữa thực nghiệm và lý thuyết

## Bài toán 1

Giả sử  $\omega_n = \{X_1, X_2, \dots, X_n\}$  là một mẫu dữ liệu sinh từ biến ngẫu nhiên  $X$ . Liệu có thể nói dữ liệu nhận được có phù hợp với quy luật xác suất  $F(x, \theta)$  của biến  $X$  hay không?

Tức là

- Kiểm định giả thuyết thống kê ( $H_0 : X \in F(x, \theta) \mid \overline{H_0} : X \notin F(x, \theta)$ )
- Nếu biến  $X$  rời rạc, thì tập số liệu  $\omega_n = \{X_1, X_2, \dots, X_n\}$  có thể chia thành  $k$  khoảng với các cỡ  $n_1, n_2, \dots, n_k$ , sao cho

$$n = n_1 + n_2 + \dots + n_k$$

# Kiểm định tính phù hợp giữa thực nghiệm và lý thuyết

## Bài toán 1

Giả sử  $\omega_n = \{X_1, X_2, \dots, X_n\}$  là một mẫu dữ liệu sinh từ biến ngẫu nhiên  $X$ . Liệu có thể nói dữ liệu nhận được có phù hợp với quy luật xác suất  $F(x, \theta)$  của biến  $X$  hay không?

Tức là

- Kiểm định giả thuyết thống kê ( $H_0 : X \in F(x, \theta) \mid \overline{H_0} : X \notin F(x, \theta)$ )
- Nếu biến  $X$  rời rạc, thì tập số liệu  $\omega_n = \{X_1, X_2, \dots, X_n\}$  có thể chia thành  $k$  khoảng với các cỡ  $n_1, n_2, \dots, n_k$ , sao cho

$$n = n_1 + n_2 + \dots + n_k$$

- Nếu biến  $X$  liên tục, thì xét số liệu trong các khoảng  $[x_j, x_{j+1}], j = 0, 1, 2, \dots, n$

# Cặp giả thuyết

Cặp  $(H_0 \mid \overline{H_0})$

$(H_0 : X \in F(x, \theta) \mid \overline{H_0} : X \notin F(x, \theta))$

# Hàm kiểm định

## ① Thống kê

$$\chi^2 = \sum_{j=1}^k \frac{(n_j - np_j)^2}{np_j}$$

# Hàm kiểm định

## 1 Thống kê

$$\chi^2 = \sum_{j=1}^k \frac{(n_j - np_j)^2}{np_j}$$

- 2 Nếu giả thuyết đầu  $H_0$  đúng thì  $\chi^2$  có phân phối khi bình phương với  $(k-r-1)$  bậc tự do

$$\chi^2 \in \chi^2(k - r - 1)$$

# Hàm kiểm định

## 1 Thống kê

$$\chi^2 = \sum_{j=1}^k \frac{(n_j - np_j)^2}{np_j}$$

- 2 Nếu giả thuyết đầu  $H_0$  đúng thì  $\chi^2$  có phân phối khi bình phương với  $(k-r-1)$  bậc tự do

$$\chi^2 \in \chi^2(k - r - 1)$$

- 3  $r$  là số tham số của phân phối  $F(x, \theta)$

# Mức ý nghĩa

- Với mức ý nghĩa  $\alpha \in (0, 1)$  cho trước, tra bảng khi-bình phương với  $(k-r-1)$  bậc tự do, xác định

$$\chi_{\alpha}^2(k - r - 1)$$



# Miền bác bỏ

Với phân vị khi bình phương  $\chi^2_\alpha(k - r - 1)$  đã xác định, miền bác bỏ có dạng sau:

- 1 Miền bác bỏ  $W_\alpha = \left( \chi^2_\alpha(k - r - 1), +\infty \right)$  đối với kiểm định  $(H_0 : X \in F(x, \theta) \mid \overline{H_0} : X \notin F(x, \theta))$ .

# Kiểm định

So sánh thống kê  $\chi^2$  với phân vị khi bình phương  $\chi_\alpha^2(k - r - 1)$ , ta có kết luận:

- 1 Nếu  $\chi^2 \in (\chi_\alpha^2(k - r - 1), +\infty)$ , thì bác bỏ giả thuyết  $H_0$ .

# Kiểm định

So sánh thống kê  $\chi^2$  với phân vị khi bình phương  $\chi_\alpha^2(k - r - 1)$ , ta có kết luận:

- 1 Nếu  $\chi^2 \in (\chi_\alpha^2(k - r - 1), +\infty)$ , thì bác bỏ giả thuyết  $H_0$ .
- 2 Nếu  $\chi^2 \notin (\chi_\alpha^2(k - r - 1), +\infty)$ , thì chấp nhận giả thuyết  $H_0$ .

# Kiểm định khi bình phương

## Ví dụ 1

Quan sát số người vào giao dịch ngân hàng (rút tiền, chuyển khoản, sao kê, ..) tại một máy ATM trong một thời gian xác định, có số liệu thống kê sau

$$X_j : 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6$$

$$n_j : 17 \quad 22 \quad 26 \quad 20 \quad 11 \quad 2 \quad 2$$

- 1 X số người vào giao dịch
- 2  $n_j, j = 0, 1, 2, 3, 4, 5, 6$  là số lần quan sát
- 3 Với mức ý nghĩa  $\alpha = 0.01$ , có thể coi X có phân phối Poisson được không? Cho biết  $\chi^2_{0.01}(5) = 15.1$

# Lời giải

- 1 Gọi  $X$  là số người vào giao dịch tại một máy ATM trong một thời gian xác định

# Lời giải

- 1 Gọi  $X$  là số người vào giao dịch tại một máy ATM trong một thời gian xác định
- 2 Ta có bài toán kiểm định sự phù hợp giữa thực nghiệm và lý thuyết

$$(H_0 : X \in P(\lambda) \mid \overline{H_0} : X \notin P(\lambda))$$

# Lời giải

- 1 Gọi  $X$  là số người vào giao dịch tại một máy ATM trong một thời gian xác định
- 2 Ta có bài toán kiểm định sự phù hợp giữa thực nghiệm và lý thuyết

$$(H_0 : X \in P(\lambda) \mid \overline{H_0} : X \notin P(\lambda))$$

- 3 Cần phải xác định giá trị tham số  $\lambda$ . Chọn  $\lambda = \overline{X} = 2$  (Tại sao?)

# Lời giải

- 1 Gọi  $X$  là số người vào giao dịch tại một máy ATM trong một thời gian xác định
- 2 Ta có bài toán kiểm định sự phù hợp giữa thực nghiệm và lý thuyết

$$(H_0 : X \in P(\lambda) \mid \overline{H_0} : X \notin P(\lambda))$$

- 3 Cần phải xác định giá trị tham số  $\lambda$ . Chọn  $\lambda = \overline{X} = 2$  (Tại sao?)
- 4 Như vậy cặp giả thuyết thống kê sẽ là

$$(H_0 : X \in P(2) \mid \overline{H_0} : X \notin P(2))$$



## Lời giải (tiếp tục)

- ① Nếu giả thuyết  $H_0 : X \in P(2)$  đúng, thì các xác suất

$$p_j = P(X = j) = \frac{e^{-2}2^j}{j!}, \quad j = 0, 1, \dots, 6$$

## Lời giải (tiếp tục)

- ① Nếu giả thuyết  $H_0 : X \in P(2)$  đúng, thì các xác suất

$$p_j = P(X = j) = \frac{e^{-2}2^j}{j!}, \quad j = 0, 1, \dots, 6$$

- ② Khi đó, hàm kiểm định có giá trị

$$\chi^2 = \sum_{j=0}^6 \frac{(n_j - np_j)^2}{np_j} = 3.76925$$

## Lời giải (tiếp tục)

- 1 Với mức ý nghĩa  $\alpha = 0.01$ , tra bảng  $\chi^2$  xác định  $\chi^2_{0.01}(7 - 1 - 1) = 15.1$

## Lời giải (tiếp tục)

- 1 Với mức ý nghĩa  $\alpha = 0.01$ , tra bảng  $\chi^2$  xác định  $\chi^2_{0.01}(7 - 1 - 1) = 15.1$
- 2 Miền bác bỏ  $W_{0.01} = (15.1, +\infty)$

## Lời giải (tiếp tục)

- 1 Với mức ý nghĩa  $\alpha = 0.01$ , tra bảng  $\chi^2$  xác định  $\chi^2_{0.01}(7 - 1 - 1) = 15.1$
- 2 Miền bác bỏ  $W_{0.01} = (15.1, +\infty)$
- 3 Do  $\chi^2 \notin W_{0.01}$ , nên chấp nhận giả thuyết ban đầu  $H_0 : X \in P(2)$

## Lời giải (tiếp tục)

- 1 Với mức ý nghĩa  $\alpha = 0.01$ , tra bảng  $\chi^2$  xác định  $\chi^2_{0.01}(7 - 1 - 1) = 15.1$
- 2 Miền bác bỏ  $W_{0.01} = (15.1, +\infty)$
- 3 Do  $\chi^2 \notin W_{0.01}$ , nên chấp nhận giả thuyết ban đầu  $H_0 : X \in P(2)$
- 4 Kết luận: số người vào giao dịch tại trạm ATM có phân phối Poisson với tham số 2 (với mức ý nghĩa 0.01)

# Sử dụng ngôn ngữ R

- 1  $X \leftarrow -c(0 : 6)$  - "tạo vector X có giá trị từ 0 tới 6"

# Sử dụng ngôn ngữ R

- 1  $X <- c(0 : 6)$  - "tạo vector X có giá trị từ 0 tới 6"
- 2  $N <- c(17, 22, 26, 20, 11, 2, 2)$  - "tạo vector N"



# Sử dụng ngôn ngữ R

- 1  $X <- c(0 : 6)$  - "tạo vector X có giá trị từ 0 tới 6"
- 2  $N <- c(17, 22, 26, 20, 11, 2, 2)$  - "tạo vector N"
- 3  $tbm <- sum(X * N) / sum(N)$  - "tính trung bình mẫu"

# Sử dụng ngôn ngữ R

- 1  $X < -c(0 : 6)$  - "tạo vector X có giá trị từ 0 tới 6"
- 2  $N < -c(17, 22, 26, 20, 11, 2, 2)$  - "tạo vector N"
- 3  $tbm < -sum(X * N)/sum(N)$  - "tính trung bình mẫu"
- 4  $p_j = P(X = x_j) = \frac{e^{-2}(2)^j}{j!}$  - "tính các xác suất thành phần"

# Sử dụng ngôn ngữ R

- 1  $X < -c(0 : 6)$  - "tạo vector X có giá trị từ 0 tới 6"
- 2  $N < -c(17, 22, 26, 20, 11, 2, 2)$  - "tạo vector N"
- 3  $tbm < -sum(X * N) / sum(N)$  - "tính trung bình mẫu"
- 4  $p_j = P(X = x_j) = \frac{e^{-2}(2)^j}{j!}$  - "tính các xác suất thành phần"
- 5  $P < -c(p0, p1, p2, p3, p4, p5, p6)$  - "Tạo vector xác suất P"

# Sử dụng ngôn ngữ R

- 1  $X < -c(0 : 6)$  - "tạo vector X có giá trị từ 0 tới 6"
- 2  $N < -c(17, 22, 26, 20, 11, 2, 2)$  - "tạo vector N"
- 3  $tbm < -sum(X * N) / sum(N)$  - "tính trung bình mẫu"
- 4  $p_j = P(X = x_j) = \frac{e^{-2}(2)^j}{j!}$  - "tính các xác suất thành phần"
- 5  $P < -c(p0, p1, p2, p3, p4, p5, p6)$  - "Tạo vector xác suất P"
- 6  $chi1 < -sum((N - sum(N) * P)^2 / (sum(N) * P))$  - "tính giá trị hàm kiểm định"

# Sử dụng ngôn ngữ R

- 1  $X < -c(0 : 6)$  - "tạo vector X có giá trị từ 0 tới 6"
- 2  $N < -c(17, 22, 26, 20, 11, 2, 2)$  - "tạo vector N"
- 3  $tbm < -sum(X * N) / sum(N)$  - "tính trung bình mẫu"
- 4  $p_j = P(X = x_j) = \frac{e^{-2}(2)^j}{j!}$  - "tính các xác suất thành phần"
- 5  $P < -c(p0, p1, p2, p3, p4, p5, p6)$  - "Tạo vector xác suất P"
- 6  $chi1 < -sum((N - sum(N) * P)^2 / (sum(N) * P))$  - "tính giá trị hàm kiểm định"
- 7  $chi1 = 3.76925$  - "Kết quả"

## 8.3 Kiểm định tính độc lập

### Bài toán 2

Giả sử biến ngẫu nhiên  $X$  có  $r$  đặc tính và  $Y$  có  $s$  đặc tính,  $s \neq r$ . Biểu diễn các biến ngẫu nhiên  $X$  và  $Y$  qua  $X = (x_1, x_2, \dots, x_r)$  và  $Y = (y_1, y_2, \dots, y_s)$ . Cần kiểm định giả thuyết đầu  $H_0 : X, Y$  độc lập

# Chú ý

- Biến ngẫu nhiên  $X$  có  $r$  đặc tính, biến ngẫu nhiên  $Y$  có  $s$  đặc tính. Có thể biểu diễn qua bảng (rs) phần tử

# Chú ý

- Biến ngẫu nhiên  $X$  có  $r$  đặc tính, biến ngẫu nhiên  $Y$  có  $s$  đặc tính. Có thể biểu diễn qua bảng  $(rs)$  phần tử
- Tần số đồng thời  $n_{ij} = (X = x_i, Y = y_j), i = 1, 2, \dots, r, j = 1, 2, \dots, s$



# Chú ý

- Biến ngẫu nhiên  $X$  có  $r$  đặc tính, biến ngẫu nhiên  $Y$  có  $s$  đặc tính. Có thể biểu diễn qua bảng  $(rs)$  phần tử
- Tần số đồng thời  $n_{ij} = (X = x_i, Y = y_j), i = 1, 2, \dots, r, j = 1, 2, \dots, s$
- Dễ thấy  $\sum_{j=1}^r n_{ij} = n_i; \sum_{i=1}^s n_{ij} = m_j; \sum_{j=1}^s m_j = n; \sum_{i=1}^r n_i = n$

# Cặp giả thuyết

Cặp giả thuyết được đưa ra

Cặp  $(H_0 \mid \overline{H_0})$

$(H_0 : X, Y \text{ độc lập} \mid \overline{H_0} : X, Y \text{ không độc lập})$

# Hàm kiểm định

- Thống kê

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - \frac{n_i m_j}{n})^2}{\frac{n_i m_j}{n}}$$

# Hàm kiểm định

- Thống kê

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - \frac{n_i m_j}{n})^2}{\frac{n_i m_j}{n}}$$

- Nếu giả thuyết đầu  $H_0$  đúng thì  $\chi^2$  có phân phối khi bình phương với  $(r-1)(s-1)$  bậc tự do

## Mức ý nghĩa

Với mức ý nghĩa  $\alpha \in (0, 1)$  cho trước và bậc tự do  $(r-1)(s-1)$  xác định, tra bảng khi bình phương, ta có

### phân vị

Phân vị khi bình phương  $\chi^2_{\alpha}(r-1)(s-1)$  cho kiểm định giả thuyết đầu  
( $H_0 : X, Y$  độc lập |  $\overline{H_0} : X, Y$  không độc lập)

# Miền bác bỏ

Với các phân vị khi bình phương  $\chi^2_\alpha(r-1)(s-1)$  đã xác định, miền bác bỏ có dạng sau:

Miền bác bỏ

$$\text{Miền bác bỏ } W_\alpha = \left( \chi^2_\alpha(r-1)(s-1), +\infty \right)$$

# Kiểm định

So sánh thống kê  $\chi^2$  với phân vị  $\chi_\alpha^2(r-1)(s-1)$ , có kết luận:

- 1 Nếu  $\chi^2 \in \left( \chi_\alpha^2(r-1)(s-1), +\infty \right)$ , thì bác bỏ giả thuyết  $H_0$ , khi đó X và Y không độc lập.

# Kiểm định

So sánh thống kê  $\chi^2$  với phân vị  $\chi_\alpha^2(r-1)(s-1)$ , có kết luận:

- 1 Nếu  $\chi^2 \in \left( \chi_\alpha^2(r-1)(s-1), +\infty \right)$ , thì bác bỏ giả thuyết  $H_0$ , khi đó X và Y không độc lập.
- 2 Nếu  $\chi^2 \notin \left( \chi_\alpha^2(r-1)(s-1), +\infty \right)$ , thì chấp nhận giả thuyết  $H_0$ , khi đó X và Y độc lập



# Kiểm định tính độc lập

## Ví dụ 2

Có 4 phân xưởng A, B, C và D với số công nhân tương ứng là 150, 170, 160 và 120. Sau một đợt cúm các phân xưởng có số công nhân bị ốm phải nghỉ làm lần lượt là 29, 39, 25 và 56. Số công nhân mệt nhưng vẫn làm việc lần lượt là 21, 11, 35, 24. Với  $\alpha = 0.05$  hãy xét xem tình trạng sức khỏe của công nhân có phụ thuộc vào việc họ làm việc ở phân xưởng nào hay không?

# Lời giải

- 1 Gọi  $X$  là phân xưởng.  $X$  có 4 đặc tính  $A$ ,  $B$ ,  $C$  và  $D$

# Lời giải

- 1 Gọi  $X$  là phân xưởng.  $X$  có 4 đặc tính  $A, B, C$  và  $D$
- 2 Gọi  $Y$  là tình trạng sức khỏe của công nhân.  $Y$  có ba đặc tính: khỏe, mệt (vẫn làm việc) và ốm (nghỉ việc)

# Lời giải

- 1 Gọi  $X$  là phân xưởng.  $X$  có 4 đặc tính A, B, C và D
- 2 Gọi  $Y$  là tình trạng sức khỏe của công nhân.  $Y$  có ba đặc tính: khỏe, mệt (vẫn làm việc) và ốm (nghỉ việc)
- 3 Giả thuyết cần kiểm định  
( $H_0 : X, Y$  độc lập |  $\overline{H_0} : X, Y$  không độc lập)

## Lời giải (tiếp tục)

① Hàm kiểm định  $\chi^2 = \sum_{i=1}^4 \sum_{j=1}^3 \frac{(n_{ij} - \frac{n_i m_j}{n})^2}{\frac{n_i m_j}{n}} = 64.45$

## Lời giải (tiếp tục)

- 1 Hàm kiểm định  $\chi^2 = \sum_{i=1}^4 \sum_{j=1}^3 \frac{(n_{ij} - \frac{n_i m_j}{n})^2}{\frac{n_i m_j}{n}} = 64.45$
- 2 Với mức ý nghĩa  $\alpha = 0.05$ , tra bảng khi bình phương, xác định  $\chi_{0.05}^2(3) = 12.59$

## Lời giải (tiếp tục)

- ① Hàm kiểm định  $\chi^2 = \sum_{i=1}^4 \sum_{j=1}^3 \frac{(n_{ij} - \frac{n_i m_j}{n})^2}{\frac{n_i m_j}{n}} = 64.45$
- ② Với mức ý nghĩa  $\alpha = 0.05$ , tra bảng khi bình phương, xác định  $\chi_{0.05}^2(3) = 12.59$
- ③ Miền bác bỏ  $W_\alpha = (12.59, +\infty)$

## Lời giải (tiếp tục)

- ① Hàm kiểm định  $\chi^2 = \sum_{i=1}^4 \sum_{j=1}^3 \frac{(n_{ij} - \frac{n_i m_j}{n})^2}{\frac{n_i m_j}{n}} = 64.45$
- ② Với mức ý nghĩa  $\alpha = 0.05$ , tra bảng khi bình phương, xác định  $\chi_{0.05}^2(3) = 12.59$
- ③ Miền bác bỏ  $W_\alpha = (12.59, +\infty)$
- ④ Do  $\chi^2 \in W_\alpha$ , nên bác bỏ giả thuyết ban đầu  $H_0$



## Lời giải (tiếp tục)

- 1 Hàm kiểm định  $\chi^2 = \sum_{i=1}^4 \sum_{j=1}^3 \frac{(n_{ij} - \frac{n_i m_j}{n})^2}{\frac{n_i m_j}{n}} = 64.45$
- 2 Với mức ý nghĩa  $\alpha = 0.05$ , tra bảng khi bình phương, xác định  $\chi_{0.05}^2(3) = 12.59$
- 3 Miền bác bỏ  $W_\alpha = (12.59, +\infty)$
- 4 Do  $\chi^2 \in W_\alpha$ , nên bác bỏ giả thuyết ban đầu  $H_0$
- 5 Kết luận: X và Y phụ thuộc (với mức ý nghĩa 0.05)

## 8.4 Kiểm định nhiều tỷ lệ

### Bài toán 3

Giả sử có  $s$  đối tượng ( $s$  tỷ lệ), mỗi đối tượng được chia theo 2 dấu hiệu đối lập (tốt-xấu, đúng-sai, sống chết, có tác dụng-không tác dụng...) Với mức ý nghĩa  $\alpha \in (0, 1)$ , cần kiểm định giả thuyết thống kê

$$H_0 : p_1 = p_2 = \dots = p_s \quad \text{với đối thuyết} \quad \overline{H_0} : p_1 \neq p_2 \neq \dots \neq p_s$$

# Chú ý

- Có thể coi như bài toán 2 với biến ngẫu nhiên  $X$  có  $s$  đặc tính, biến ngẫu nhiên  $Y$  có 2 đặc tính. Như vậy, số liệu có thể biểu diễn qua bảng  $(2s)$  phần tử

# Chú ý

- Có thể coi như bài toán 2 với biến ngẫu nhiên  $X$  có  $s$  đặc tính, biến ngẫu nhiên  $Y$  có 2 đặc tính. Như vậy, số liệu có thể biểu diễn qua bảng  $(2s)$  phần tử
- Tần số đồng thời  $n_{ij} = (X = x_i, Y = y_j), i = 1, 2, \dots, s, j = 1, 2$

# Chú ý

- Có thể coi như bài toán 2 với biến ngẫu nhiên  $X$  có  $s$  đặc tính, biến ngẫu nhiên  $Y$  có 2 đặc tính. Như vậy, số liệu có thể biểu diễn qua bảng  $(2s)$  phần tử
- Tần số đồng thời  $n_{ij} = (X = x_i, Y = y_j), i = 1, 2, \dots, s, j = 1, 2$
- Dễ thấy  $\sum_{j=1}^2 n_{ij} = n_i; \sum_{i=1}^s n_{ij} = m_j; \sum_{j=1}^2 m_j = n; \sum_{i=1}^s n_i = n$

# Cặp giả thuyết

Cặp giả thuyết được đưa ra

Cặp  $(H_0 \mid \overline{H_0})$

$H_0 : p_1 = p_2 = \dots = p_s$  với đối thuyết  $\overline{H_0} : p_1 \neq p_2 \neq \dots \neq p_s$ .

# Hàm kiểm định

- Thống kê

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^s \frac{(n_{ij} - \frac{n_i m_j}{n})^2}{\frac{n_i m_j}{n}}$$

- Thống kê

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^s \frac{(n_{ij} - \frac{n_i m_j}{n})^2}{\frac{n_i m_j}{n}}$$

- Nếu giả thuyết đầu  $H_0$  đúng thì  $\chi^2$  có phân phối khi bình phương với  $(s-1)$  bậc tự do



## Mức ý nghĩa

Với mức ý nghĩa  $\alpha \in (0, 1)$  cho trước và bậc tự do  $(s-1)$  xác định, tra bảng khi bình phương, ta có

### phân vị

Phân vị khi bình phương  $\chi^2_{\alpha}(s-1)$  cho kiểm định giả thuyết đầu  $H_0 : p_1 = p_2 = \dots = p_s$  với đối thuyết  $\overline{H_0} : p_1 \neq p_2 \neq \dots \neq p_s$ .

# Miền bác bỏ

Với các phân vị khi bình phương  $\chi^2_\alpha(s-1)$  đã xác định, miền bác bỏ có dạng sau:

Miền bác bỏ

$$\text{Miền bác bỏ } W_\alpha = \left( \chi^2_\alpha(s-1), +\infty \right)$$

# Kiểm định

So sánh thống kê  $\chi^2$  với phân vị  $\chi_\alpha^2(s-1)$ , có kết luận:

- 1 Nếu  $\chi^2 \in \left( \chi_\alpha^2(s-1), +\infty \right)$ , thì bác bỏ giả thuyết

$$H_0 : p_1 = p_2 = \dots = p_s \quad \text{với đối thuyết} \quad \overline{H_0} : p_1 \neq p_2 \neq \dots \neq p_s$$

# Kiểm định

So sánh thống kê  $\chi^2$  với phân vị  $\chi_\alpha^2(s-1)$ , có kết luận:

- ❶ Nếu  $\chi^2 \in \left( \chi_\alpha^2(s-1), +\infty \right)$ , thì bác bỏ giả thuyết

$$H_0 : p_1 = p_2 = \dots = p_s \quad \text{với đối thuyết} \quad \overline{H_0} : p_1 \neq p_2 \neq \dots \neq p_s$$

- ❷ Nếu  $\chi^2 \notin \left( \chi_\alpha^2(r-1)(s-1), +\infty \right)$ , thì chấp nhận giả thuyết

$$H_0 : p_1 = p_2 = \dots = p_s \quad \text{với đối thuyết} \quad \overline{H_0} : p_1 \neq p_2 \neq \dots \neq p_s.$$

# Kiểm định nhiều tỷ lệ

## Ví dụ 3

Có 4 giáo viên dạy cùng một giáo trình xác suất thống kê cho 4 lớp A, B, C và D với số lượng sinh viên lần lượt là 95, 160, 172 và 180. Thống kê số sinh viên bị thi lại của các lớp lần lượt là 18, 32, 34 và 36. Với mức ý nghĩa  $\alpha = 0.01$ , chất lượng học tập của sinh viên có phụ thuộc vào chất lượng giảng dạy của các giáo viên hay không?

# Lời giải

- 1 Gọi tỷ lệ sinh viên bị thi lại môn xác suất thống kê của các lớp A, B, C và D lần lượt là  $p_1, p_2, p_3, p_4$

# Lời giải

- 1 Gọi tỷ lệ sinh viên bị thi lại môn xác suất thống kê của các lớp A, B, C và D lần lượt là  $p_1, p_2, p_3, p_4$
- 2 Giả thuyết cần kiểm định  
( $H_0 : p_1 = p_2 = p_3 = p_4 \mid \overline{H_0} : p_1 \neq p_2 \neq p_3 \neq p_4$ )

## Lời giải (tiếp tục)

① Hàm kiểm định  $\chi^2 = \sum_{i=1}^4 \sum_{j=1}^2 \frac{(n_{ij} - \frac{n_i m_j}{n})^2}{\frac{n_i m_j}{n}} = 0.047$



## Lời giải (tiếp tục)

- 1 Hàm kiểm định  $\chi^2 = \sum_{i=1}^4 \sum_{j=1}^2 \frac{(n_{ij} - \frac{n_i m_j}{n})^2}{\frac{n_i m_j}{n}} = 0.047$
- 2 Với mức ý nghĩa  $\alpha = 0.01$ , tra bảng khi bình phương, xác định  $\chi_{0.01}^2(3) = 11.34$

## Lời giải (tiếp tục)

- 1 Hàm kiểm định  $\chi^2 = \sum_{i=1}^4 \sum_{j=1}^2 \frac{(n_{ij} - \frac{n_i m_j}{n})^2}{\frac{n_i m_j}{n}} = 0.047$
- 2 Với mức ý nghĩa  $\alpha = 0.01$ , tra bảng khi bình phương, xác định  $\chi_{0.01}^2(3) = 11.34$
- 3 Miền bác bỏ  $W_\alpha = (11.34, +\infty)$

## Lời giải (tiếp tục)

- 1 Hàm kiểm định  $\chi^2 = \sum_{i=1}^4 \sum_{j=1}^2 \frac{(n_{ij} - \frac{n_i m_j}{n})^2}{\frac{n_i m_j}{n}} = 0.047$
- 2 Với mức ý nghĩa  $\alpha = 0.01$ , tra bảng khi bình phương, xác định  $\chi_{0.02}^2(3) = 11.34$
- 3 Miền bác bỏ  $W_\alpha = (11.34, +\infty)$
- 4 Do  $\chi^2 \notin W_\alpha$ , nên chấp nhận giả thuyết ban đầu  
 $H_0 : p_1 = p_2 = p_3 = p_4$  và bác bỏ đối thuyết  $\overline{H_0} : p_1 \neq p_2 \neq p_3 \neq p_4$

## Lời giải (tiếp tục)

- 1 Hàm kiểm định  $\chi^2 = \sum_{i=1}^4 \sum_{j=1}^2 \frac{(n_{ij} - \frac{n_i m_j}{n})^2}{\frac{n_i m_j}{n}} = 0.047$
- 2 Với mức ý nghĩa  $\alpha = 0.01$ , tra bảng khi bình phương, xác định  $\chi_{0.02}^2(3) = 11.34$
- 3 Miền bác bỏ  $W_\alpha = (11.34, +\infty)$
- 4 Do  $\chi^2 \notin W_\alpha$ , nên chấp nhận giả thuyết ban đầu  
 $H_0 : p_1 = p_2 = p_3 = p_4$  và bác bỏ đối thuyết  $\overline{H_0} : p_1 \neq p_2 \neq p_3 \neq p_4$
- 5 Kết luận: tỷ lệ sinh viên thi lại như nhau (với mức ý nghĩa 0.05), nên rõ ràng chất lượng học tập của sinh viên có không phụ thuộc (độc lập) vào chất lượng giảng dạy của giáo viên

## 8.5 Tiêu chuẩn Mann-Whitney

### Bài toán 4

Giả sử hai mẫu ngẫu nhiên  $\omega_1 = (x_1, x_2, \dots, x_{n_1})$  và  $\omega_2 = (y_1, y_2, \dots, y_{n_2})$  sinh bởi các biến ngẫu nhiên  $X$  và  $Y$ , chưa xác định phân phối xác suất. Với mức ý nghĩa  $\alpha$ , hãy kiểm định giả thuyết cho rằng  $X$  và  $Y$  có cùng phân phối, tức là

$$H_0 : E(X) = E(Y) \mid \overline{H_0} : E(X) \neq E(Y)$$

# Chú ý

- ❶ Hai biến ngẫu nhiên  $X$  và  $Y$  không nhất thiết có phân phối chuẩn

# Chú ý

- ① Hai biến ngẫu nhiên  $X$  và  $Y$  không nhất thiết có phân phối chuẩn
- ② Phương pháp Mann-Whitney còn được gọi là phương pháp hạng

# Hạng của một phần tử trong dãy

## Định nghĩa

Giả sử có một dãy số được sắp thứ tự

$x_{[1]} < x_{[2]} < \dots < x_{[j]} < x_{[j+1]} < \dots < x_{[n]}$ . Khi đó, hạng của phần tử thứ  $j$  trong dãy chính là số thứ tự vị trí của phần tử đó. Ký hiệu,  $\text{rank}(x_{[j]}) = r_j = j$ .



# Hạng của dãy số

## Ví dụ

Dãy số 3, 1, 4, 6, 8, 5 sau khi xếp thứ tự có dạng 1, 3, 4, 5, 6, 8 nên  
 $r(1)=1, r(3)=2, r(4)=3, r(5)=4, r(6)=5, r(8)=6$

# Hạng của một phần tử trong dãy

## Định nghĩa

- Nếu một dãy số được sắp thứ tự

$x_{[1]} < x_{[2]} < \dots < x_{[j]} = x_{[j+1]} < \dots < x_{[n]}$ . Khi đó, hạng của phần tử thứ  $j$  trong dãy chính là trung bình cộng số thứ tự vị trí của hai phần tử  $x_{[j]}, x_{[j+1]}$ . Như vậy,

$$\text{rank}(x_{[j]}) = r_j = \frac{j + j + 1}{2} = j + 0.5$$

- Nếu một dãy số được sắp thứ tự

$x_{[1]} < x_{[2]} < \dots < x_{[j-1]} = x_{[j]} = x_{[j+1]} < \dots < x_{[n]}$ . Khi đó, hạng của phần tử thứ  $j$  trong dãy chính là trung bình cộng số thứ tự vị trí của ba phần tử  $x_{[j-1]}, x_{[j]}, x_{[j+1]}$ . Như vậy,

$$\text{rank}(x_{[j]}) = r_j = \frac{j - 1 + j + j + 1}{2} = j$$

# Hạng của dãy số

- ① Dãy số 3, 1, 4, 4, 8, 5 sau khi xếp thứ tự có dạng 1, 3, 4, 4, 5, 8 nên

$$r(1) = 1, r(3) = 2, r(4) = 3.5, r(5) = 5, r(8) = 6$$

# Hạng của dãy số

- ① Dãy số 3, 1, 4, 4, 8, 5 sau khi xếp thứ tự có dạng 1, 3, 4, 4, 5, 8 nên

$$r(1) = 1, r(3) = 2, r(4) = 3.5, r(5) = 5, r(8) = 6$$

- ② Dãy số 3, 1, 4, 4, 4, 8, 5 sau khi xếp thứ tự có dạng 1, 3, 4, 4, 4, 5, 8 nên

$$r(1) = 1, r(3) = 2, r(4) = 4, r(5) = 6, r(8) = 7$$

# Tiêu chuẩn Mann-Whitney

## Định nghĩa

- Gộp hai mẫu thành một mẫu có thứ tự

$$c_1 < c_2 < \dots < c_{n_1+n_2}$$

- Ký hiệu tổng các số hạng của  $x_i$  là  $R_1 = \sum_{i=1}^{n_1} \text{rank}(x_i)$  và tổng các số hạng của  $y_j$  là  $R_2 = \sum_{j=1}^{n_2} \text{rank}(y_j)$

# Tiêu chuẩn Mann-Whitney

## Hàm kiểm định K

- Đại lượng

$$V_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1; \quad V_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

- Hàm kiểm định:

$$K = \frac{(V - E(V))}{\sqrt{D(V)}}$$

trong đó,  $V = \min\{V_1; V_2\}$ ,

- Kỳ vọng và phương sai

$$E(V) = \frac{n_1 n_2}{2}; \quad D(V) = \frac{n_1 n_2 (n_1 + n_2 - 1)}{12}$$

## Phân vị chuẩn mức $\frac{\alpha}{2}$

- ① Nếu giả thuyết  $H_0$  đúng thì biến ngẫu nhiên  $K = \frac{(V - E(V))}{\sqrt{D(V)}} \sim \mathfrak{N}(0, 1)$

## Phân vị chuẩn mức $\frac{\alpha}{2}$

- 1 Nếu giả thuyết  $H_0$  đúng thì biến ngẫu nhiên  $K = \frac{(V-E(V))}{\sqrt{D(V)}} \sim \mathfrak{N}(0, 1)$
- 2 Với mức ý nghĩa  $\alpha \in (0, 1)$ , tra bảng Laplace xác định phân vị chuẩn  $x_{\frac{\alpha}{2}}$  sao cho

$$\Phi(x_{\frac{\alpha}{2}}) = \frac{1}{2} - \frac{\alpha}{2}$$



# Miền bác bỏ $W_{\frac{\alpha}{2}}$

Miền bác bỏ

Miền bác bỏ hai phía

$$W_{\frac{\alpha}{2}} = \left( -\infty, -x_{\frac{\alpha}{2}} \right) \cup \left( x_{\frac{\alpha}{2}}, +\infty \right)$$

# Kết luận

- 1 Nếu  $K \in W_{\frac{\alpha}{2}}$ , thì giả thuyết  $H_0$  bị bác bỏ ( $\overline{H_0}$  được chấp nhận), tức là X và Y không có cùng phân phối hay  $E(X) \neq E(Y)$ .

# Kết luận

- 1 Nếu  $K \in W_{\frac{\alpha}{2}}$ , thì giả thuyết  $H_0$  bị bác bỏ ( $\overline{H_0}$  được chấp nhận), tức là X và Y không có cùng phân phối hay  $E(X) \neq E(Y)$ .
- 2 Nếu  $K \notin W_{\frac{\alpha}{2}}$ , thì giả thuyết  $H_0$  được chấp nhận ( $\overline{H_0}$  bị bác bỏ), tức là X và Y có cùng phân phối hay  $E(X) = E(Y)$ .

# Quy tắc Mann-Whitney

## Ví dụ 4

Để nghiên cứu kết quả học tập của nam sinh viên và nữ sinh viên năm thứ nhất, người ta thống kê điểm của 11 nam sinh viên (X) và 12 nữ sinh viên (Y) với số liệu sau

$X_i$  : 28   30   35   35   40   40   40   45   45   48   54

$Y_i$  : 22   25   25   30   30   32   35   35   36   38   38   38

Bằng phương pháp hạng Mann-Whitney, với  $\alpha = 0,05$ , hãy xem tổng điểm trung bình của nam sinh viên và nữ sinh viên có như nhau hay không?

## 8.6 Tiêu chuẩn Wilcoxon

### Bài toán 5

Giả sử hai mẫu ngẫu nhiên  $\omega_1 = (x_1, x_2, \dots, x_n)$  và  $\omega_2 = (y_1, y_2, \dots, y_n)$  sinh bởi các biến ngẫu nhiên  $X$  và  $Y$ , chưa xác định phân phối xác suất. Với mức ý nghĩa  $\alpha$ , hãy kiểm định giả thuyết cho rằng  $X$  và  $Y$  có cùng phân phối, tức là

$$H_0 : E(X) = E(Y) \mid \overline{H_0} : E(X) \neq E(Y)$$

# Chú ý

- ① Hai biến ngẫu nhiên  $X$  và  $Y$  không nhất thiết có phân phối chuẩn

# Chú ý

- ① Hai biến ngẫu nhiên  $X$  và  $Y$  không nhất thiết có phân phối chuẩn
- ② Hai mẫu sinh bởi hai biến ngẫu nhiên  $X$  và  $Y$  có cùng cỡ  $n$

# Giải thích

- 1 Có thể coi  $X$  là hiệu quả của phương pháp thứ nhất



# Giải thích

- 1 Có thể coi  $X$  là hiệu quả của phương pháp thứ nhất
- 2  $Y$  là hiệu quả của phương pháp thứ hai tác động lên cùng một cá thể

# Giải thích

- 1 Có thể coi  $X$  là hiệu quả của phương pháp thứ nhất
- 2  $Y$  là hiệu quả của phương pháp thứ hai tác động lên cùng một cá thể
- 3 Với mức ý nghĩa  $\alpha$  kiểm định giả thuyết

$H_0 : X$  và  $Y$  có hiệu quả như nhau,

với đối thuyết

$\overline{H_0} : X$  và  $Y$  có hiệu quả không như nhau

# Tiêu chuẩn Wilcoxon

- 1 Xuất phát từ hai mẫu, tính hiệu  $d_i = x_i - y_i, i = 1, 2, \dots, n$ .

# Tiêu chuẩn Wilcoxon

- 1 Xuất phát từ hai mẫu, tính hiệu  $d_i = x_i - y_i, i = 1, 2, \dots, n$ .
- 2 Bỏ qua các giá trị  $d_i = 0$ .

# Tiêu chuẩn Wilcoxon

- 1 Xuất phát từ hai mẫu, tính hiệu  $d_i = x_i - y_i, i = 1, 2, \dots, n$ .
- 2 Bỏ qua các giá trị  $d_i = 0$ .
- 3 Tính hạng của  $|d_i| (d_i \neq 0)$ .

# Tiêu chuẩn Wilcoxon

- 1 Xuất phát từ hai mẫu, tính hiệu  $d_i = x_i - y_i, i = 1, 2, \dots, n$ .
- 2 Bỏ qua các giá trị  $d_i = 0$ .
- 3 Tính hạng của  $|d_i| (d_i \neq 0)$ .
- 4 Gọi  $\bar{n}$  số các giá trị  $d_i \neq 0$ .

# Tiêu chuẩn Wilcoxon

- 1 Xuất phát từ hai mẫu, tính hiệu  $d_i = x_i - y_i, i = 1, 2, \dots, n$ .
- 2 Bỏ qua các giá trị  $d_i = 0$ .
- 3 Tính hạng của  $|d_i| (d_i \neq 0)$ .
- 4 Gọi  $\bar{n}$  số các giá trị  $d_i \neq 0$ .
- 5 Ký hiệu  $R^+$  là tổng các hạng của  $|d_i|$  với  $d_i > 0$ ,  $R^-$  là tổng của các  $|d_i|$  với  $d_i < 0$ .

# Tiêu chuẩn Wilcoxon

- 1 Xuất phát từ hai mẫu, tính hiệu  $d_i = x_i - y_i, i = 1, 2, \dots, n$ .
- 2 Bỏ qua các giá trị  $d_i = 0$ .
- 3 Tính hạng của  $|d_i| (d_i \neq 0)$ .
- 4 Gọi  $\bar{n}$  số các giá trị  $d_i \neq 0$ .
- 5 Ký hiệu  $R^+$  là tổng các hạng của  $|d_i|$  với  $d_i > 0$ ,  $R^-$  là tổng của các  $|d_i|$  với  $d_i < 0$ .
- 6 Chọn  $R = \min(R^-; R^+)$ .



# Tiêu chuẩn Wilcoxon

## Hàm kiểm định

Tính giá trị của thống kê

$$K = \frac{(R - \frac{\bar{n}(\bar{n}+1)}{4})}{\sqrt{\frac{\bar{n}(\bar{n}+1)(2\bar{n}+1)}{24}}}$$

# Tiêu chuẩn Wilcoxon

## Phân vị chuẩn mức $\frac{\alpha}{2}$

Nếu giả thuyết  $H_0$  đúng, thì thống kê  $K$  có phân phối chuẩn chính tắc,  $K \sim \mathcal{N}(0, 1)$ , khi đó tra bảng Laplace xác định phân vị chuẩn mức  $\frac{\alpha}{2}$  là  $x_{\frac{\alpha}{2}}$  sao cho

$$\Phi_0(x_{\frac{\alpha}{2}}) = \frac{1}{2} - \frac{\alpha}{2}$$

# Tiêu chuẩn Wilcoxon

Miền bác bỏ  $W_{\frac{\alpha}{2}}$

$$W_{\frac{\alpha}{2}} = \left( -\infty, -x_{\frac{\alpha}{2}} \right) \cup \left( x_{\frac{\alpha}{2}}, +\infty \right)$$

# Kết luận

- 1 Nếu  $K \in W_{\frac{\alpha}{2}}$ , thì giả thuyết  $H_0$  bị bác bỏ ( $\overline{H_0}$  được chấp nhận), tức là  $E(X) \neq E(Y)$ .

# Kết luận

- 1 Nếu  $K \in W_{\frac{\alpha}{2}}$ , thì giả thuyết  $H_0$  bị bác bỏ ( $\overline{H_0}$  được chấp nhận), tức là  $E(X) \neq E(Y)$ .
- 2 Nếu  $K \notin W_{\frac{\alpha}{2}}$ , thì giả thuyết  $H_0$  được chấp nhận ( $\overline{H_0}$  bị bác bỏ), tức là  $E(X) = E(Y)$ .

# Bài tập chương 8

## Bài tập 1

Gieo một con xúc xắc 240 lần, trong đó mặt "lục" xuất hiện 70 lần. Với mức ý nghĩa  $\alpha = 0.01$  hỏi con xúc xắc có cân đối và đồng chất hay không?

# Bài tập chương 8

## Bài tập 2

Gieo một con xúc xắc 240 lần, trong đó mặt "nhất" xuất hiện 56 lần, mặt "nhị" xuất hiện 28 lần, mặt "tam" xuất hiện 52 lần, mặt "tứ" xuất hiện 36 lần, mặt "ngũ" xuất hiện 30 lần và mặt "lục" xuất hiện 38 lần. Với mức ý nghĩa  $\alpha = 0.05$  hỏi con xúc xắc có cân đối và đồng chất hay không?

## Bài tập chương 8

### Bài tập 3

Thống kê số vụ tai nạn ( $X_j$ ) và số ngày xảy ra tai nạn ( $n_j$ ) tại một thành phố, người ta có số liệu trong 2190 ngày như sau

$$X_j : 0 \quad 1 \quad 2 \quad 3 \quad 4$$

$$n_j : 1426 \quad 598 \quad 132 \quad 32 \quad 2$$

Với mức ý nghĩa 0.01 hãy kiểm định ý kiến cho rằng số vụ xảy ra tai nạn có phân phối Poisson



# Bài tập chương 8

## Bài tập 4

Giả sử tuổi đời ( $X$ ) được chia thành 3 giai đoạn: A (dưới 30), B (từ 30 tới 50) và C (trên 50) và sự đam mê luyện tập thể thao ( $Y$ ) có các mức độ D (tích cực), E (bình thường) và F (lười biếng). Theo dõi 500 người ở các lứa tuổi khác nhau về sự đam mê luyện tập thể thao, có thống kê sau: với độ tuổi A thì  $D=45$ ,  $E=35$ ,  $F=38$ ; với độ tuổi B thì  $D=62$ ,  $E=49$ ,  $F=95$ ; với độ tuổi C thì  $D=48$ ,  $E=49$ ,  $F=68$ . Với mức ý nghĩa 0.01 hãy kiểm định xem sự đam mê luyện tập thể thao có phụ thuộc vào độ tuổi hay không?

## Bài tập chương 8

### Bài tập 5

Một điều tra xã hội học cho thấy trong số 1076 sinh viên nam tốt nghiệp đại học có công việc ổn định trong năm đầu tiên là 526, còn trong số 1011 nữ sinh viên tốt nghiệp đại học có công việc ổn định trong năm đầu tiên là 313. Với mức ý nghĩa 0.05 hãy kiểm định giả thuyết cho rằng nữ sinh viên tốt nghiệp khó kiếm việc hơn nam sinh viên.