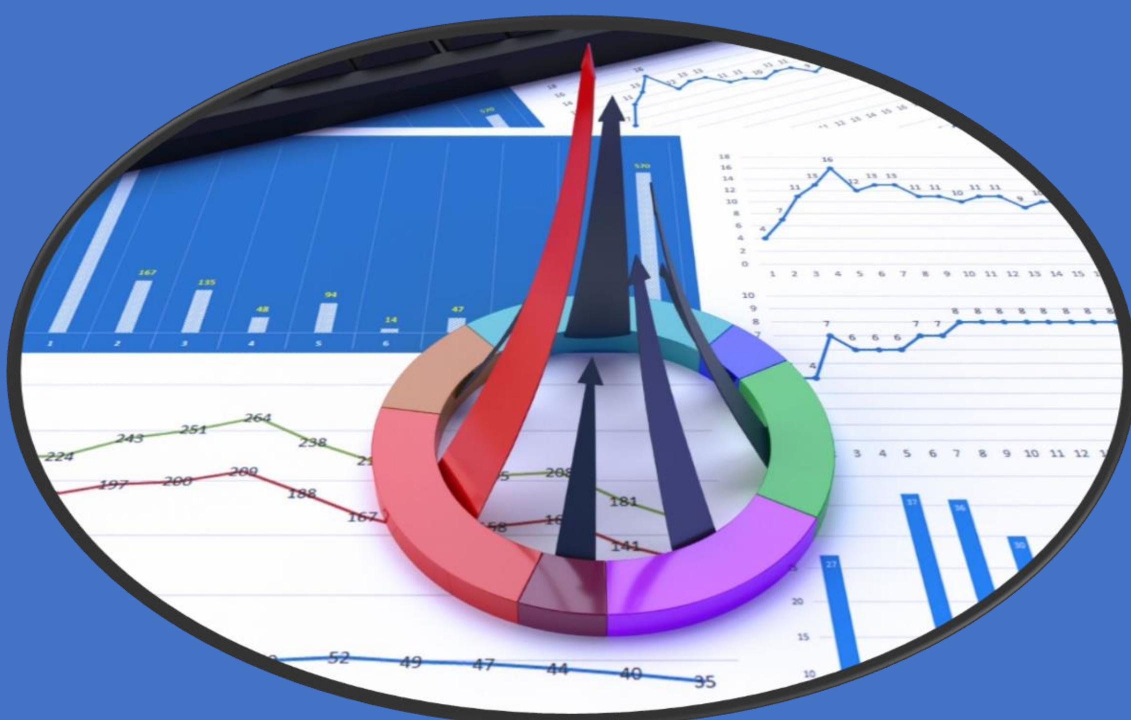
BỘ CÔNG THỨC TONY

THỐNG KÊ ỨNG DỤNG (APPLIED STATISTICS)



Pisces Kibo

MỤC LỤC

Bài 1: CÁC LOẠI PHÂN BỐ THƯỜNG GẶP

CÁC LOẠI PHÂN BỐ THÔNG THƯỜNG

1. **Phân bố nhị thức:**
   - Ký hiệu: X ~ B(n, p)
   - Công thức:

     $$\boxed{P(X = k) = C_n^k\, p^k\, (1-p)^k}$$

     - EX = np
     - DX = np (1 – p)
     - Mod X = [(n + 1) p]

2. **Phân bố Poisson:**
   - Ký hiệu: X ~ Poisson(λ)
   - Công thức:

     $$\boxed{P(X = k) = \frac{e^{-\lambda}}{k!}\lambda^k}$$

     - EX = DX = λ
     - Mod X = $[\lambda]$

3. **Phân bố chuẩn:**
   - Ký hiệu: X ~ N(μ, σ^2)
   - Công thức:

     $$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}$$

     - EX = μ
     - DX = σ^2

4. **Phân bố mũ:**
   - Ký hiệu: X ~ Exp(λ)
   - Công thức:

     $$\boxed{f(x) = \begin{cases} \lambda\, e^{-\lambda x}, & x > 0 \\ 0, & x \le 0 \end{cases}}$$

     - EX = $\frac{1}{\lambda}$
     - DX = $\frac{1}{\lambda^2}$

5. **Phân bố đều:**
   - Ký hiệu: X ~ U([a, b])
   - Công thức:

     $$\boxed{f(x) = \sqrt[n]{a^n} = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & \text{ngược lại} \end{cases}}$$

     - EX = $\frac{b+a}{12}$
     - DX = $\frac{(b-a)^2}{12}$

6. **Định lý giới hạn trung tâm:**
   - Ký hiệu: $S_n = \sum_{i=1}^{n} x_i$ ~ phân bố chuẩn
   - Công thức:

     $$\boxed{P\left(\frac{S_n - n\mu}{6\sqrt{n}} < X\right) = \phi(x)}$$

Bài 2: MẪU VÀ MỘT SỐ ĐỊNH LÝ PHÂN BỐ

BẢNG TẦN SỐ <=> $\begin{cases} \sum_{i=1}^{k} m_i = n \\ \sum_{i=1}^{n} f_i = 1 \end{cases}$

$\boldsymbol{x_i}$	$\boldsymbol{m_i}$	$\boldsymbol{f_i}$
x_1	m_1	f_1
x_2	m_2	f_2
……	……	……
x_k	m_k	f_k
X	n	1

1. **Trung bình mẫu:**
   - Ký hiệu: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
     - $E\bar{X} = \mu$
     - $D\bar{X} = \frac{\sigma^2}{n}$
2. **Phương sai mẫu:**
   - Ký hiệu: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
     - $(\sigma_{n-1})^2$ → $S^2 = \frac{1}{n-1}\hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
     - $(\sigma_n)^2$ → $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i{}^2 - (\bar{X})^2)$
3. **Một số định lý về phân phối mẫu:**
   - Định lý 1: Nếu $X_1, X_2, \dots, X_n$ là mẫu ngẫu nhiên có phân bố chuẩn N(μ, σ^2) thì $\bar{X}$ có phân bố chuẩn N(μ, $\frac{\sigma^2}{n}$)
     - ⇨ $\tilde{X}$ ~ N(μ, $\frac{\sigma^2}{n}$) ⇔ $\frac{\bar{X}-\mu}{\sigma\sqrt{n}}$ ~ N(0,1)
   - Định lý 2: Nếu $X_1, X_2, \dots, X_n$ là mẫu ngẫu nhiên có phân bố chuẩn N(μ, σ^2) thì $\bar{X}$ và S^2 độc lập với nhau
     - ⇨ $\frac{n-1}{\sigma^2}S^2 \sim \chi^2_{n-1} = Z_1^2 + Z_2^2 + \dots + (Z_{n-1})^2$
   - Định lý 3: Nếu $X_1, X_2, \dots, X_n$ là mẫu ngẫu nhiên quan sát phân bố chuẩn thì $t_{n-1} \sim T = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ cùng có phân bố t_{n-1} (với T là độ lệch tiêu chuẩn của mẫu)

- Định lý 4: Nếu $X_1, X_2, \dots, X_n$ là mẫu ngẫu nhiên quan sát X bất kỳ với kỳ vọng µ thì Z = $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ cũng có phân bố chuẩn N(0, 1) khi n --> ∞

- Định lý 5: Cho $\begin{cases} X_1, X_2, \dots, X_n \\ Y_1, Y_2, \dots, Y_n \end{cases}$, ta có $\bar{X} - \bar{Y}$ có phân bố chuẩn với $\begin{matrix} \mu = \mu_1 - \mu_2 \\ \sigma^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \end{matrix}$

Bài 3: BÀI TOÁN ƯỚC LƯỢNG THAM SỐ

1. **Ước lượng điểm:**

   $\boxed{\theta^* = T(X_1, X_2, \dots, X_n)}$

   - Ước lượng không chệch → ULKC => Eθ^* = θ
   - Ước lượng chệch → ULC => Eθ^* = $\theta + C$

- Ước lượng điểm của một số tham số quan trọng:
  - Nếu θ = EX thì $\theta^* = \bar{X}$ → Ước lượng không chệch của EX
  - Nếu θ = DX thì $\theta^* = S^2$ (hoặc $\hat{S}^2$)
    - $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\mathrm{X_i} - \bar{X})^2$ → ULKC của DX
    - $\hat{\mathrm{S}}^2 = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{X_i} - \bar{X})^2$ → ULC của DX

    Với đội chệch C = $\frac{-DX}{n}$

  - $\theta = \sigma$ => $\theta^* = S\ (hoặc\ \hat{S})$

    Với S là mẫu, $\hat{S}$ là tổng thể

  - $\theta = p(A)$ => θ^* = f = $\frac{k}{n}$

2. **Phương pháp momen (Momen cấp k):**

- Momen cấp k của X => $\mu_k = EX^k$ = $\frac{1}{n}\sum_{i=1}^{n}\mathrm{x_i}^2$

  → Đây là ước lương momen của μ_k

3. **Phương pháp ước lượng hợp lý cực đại:**

- Bước 1: phân bố X ~ ....

  $\boxed{\text{f(}x_i\text{ / tham số) = ...}}$

- Bước 2: Hàm hợp lý:

  $\boxed{\text{L(tham số)} = \prod_{i=1}^{n}\mathrm{f}(x_i\ /\ \text{tham số})}$

- Bước 3: Logarit 2 vế => l(tham số) = $\ln$(L(tham số))
- Bước 4: Đạo hàm l(tham số) theo tham số

  ⇨ $\frac{\partial l(\text{tham số})}{\partial(\text{tham số})}$ = 0 → tham số = θ^*

- *Chú thích:*

$\frac{1}{n}\sum_{i=1}^{n} x_i = \overline{X}$	$E(x - \overline{X})^2 = D\overline{X}$

4. Ước lượng hiệu quả:

- Ước lượng hợp lý cực đại:

$$\boxed{L_X(\theta) = \prod_{i=1}^{n} f(x_i, \theta)}$$

⇨ $\left[\begin{array}{l} \frac{\partial L_X(\theta)}{\partial \theta} = 0 \\ \frac{\partial \ln(L_X(\theta))}{\partial \theta} = 0 \end{array}\right.$ => θ^* = ?

- Ước lượng hiệu quả:

$$\boxed{I_X(\theta) = E\left(\left(\frac{\partial \ln(f_X(x,\theta))}{\partial \theta}\right)^2\right) = -E\left(\frac{\partial^2 \ln(f_X(x,\theta))}{\partial \theta^2}\right)}$$

  - Nếu $X_1, X_2, \dots, X_n$ độc lập → $I_X(\theta) = nI_{X1}(\theta)$

    ➔ $\begin{cases} E\theta^*(X_1, X_2, \dots, X_n) = \theta \\ D\theta^*(X_1, X_2, \dots, X_n) = \frac{1}{I_X(\theta)} \end{cases}$ $=> Hiệu\ quả$

  - Cách so sánh:
    - $D\theta^*$ = E($\theta^* -$ giá trị ước lượng$)^2$
    - $I(\lambda) = E\left(\left(\frac{\partial \ln(\lambda)}{\partial \lambda}\right)^2\right)$

⇨ Nếu $\left|\begin{array}{l} I(\lambda) = \frac{1}{D\theta^*} --> Có\ hiệu\ quả \\ I(\lambda) \neq \frac{1}{D\theta^*} --> Không\ hiệu\ quả \end{array}\right.$

- Chú thích:

$$\boxed{\begin{array}{l} \boldsymbol{Z_\alpha => \emptyset(Z_\alpha) = 1 - \alpha = \beta} \\ \boldsymbol{\rightarrow Z_\alpha = \emptyset^{-1}(1 - \alpha)} \end{array}}$$

Bài 4: ƯỚC LƯỢNG KHOẢNG TIN CẬY

1. **Ước lượng khoảng tin cậy hai phía:**
   a. Giá trị trung bình: $(\theta = EX)$
      - Nếu σ^2 đã biết với n ≥ 30:
        - Khoảng tin cậy → $(\bar{X} \pm z_{\frac{\alpha}{2}} . \frac{\sigma}{\sqrt{n}})$
        - Sai số → $\varepsilon = z_{\frac{\alpha}{2}} . \frac{\sigma}{\sqrt{n}}$
      - Nếu σ^2 chưa biết với n đủ lớn:
        - Khoảng tin cậy → $(\bar{X} \pm z_{\frac{\alpha}{2}} . \frac{S}{\sqrt{n}})$
        - Sai số → $\varepsilon = z_{\frac{\alpha}{2}} . \frac{S}{\sqrt{n}}$
      - Nếu σ^2 chưa biết với n nhỏ:
        - Khoảng tin cậy → $(\bar{X} \pm t_{\frac{\alpha}{2}; n-1} . \frac{S}{\sqrt{n}})$
        - Sai số → $\varepsilon = t_{\frac{\alpha}{2}; n-1} . \frac{S}{\sqrt{n}}$

   b. Tỷ lệ: $(\theta = p)$
      - Khoảng tin cậy → (f $\pm z_{\frac{\alpha}{2}} . \sqrt{\frac{f\,(1-f)}{n}}$)
      - Tỷ lệ → f = $\frac{k}{n}$

   c. Phương sai: $(\theta = \sigma^2)$
      - Khoảng tin cậy → $\left(\frac{(n-1)\,S^2}{\chi^2_{\frac{\alpha}{2};(n-1)}}; \frac{(n-1)\,S^2}{\chi^2_{\frac{\alpha}{2};(n-1)}}\right)$

2. **Cho $\boldsymbol{\varepsilon \leq a}$ → Tìm cỡ mẫu n:**
   a. Giá trị trung bình: $(\theta = EX)$
      - Nếu σ^2 đã biết với n ≥ 30:
        - $z_{\frac{\alpha}{2}} . \frac{\sigma}{\sqrt{n}}$ ≤ a ⇔ $\left(z_{\frac{\alpha}{2}} . \frac{\sigma}{a}\right)^2$ ≤ n
      - Nếu σ^2 chưa biết với n < 30:
        - $z_{\frac{\alpha}{2}} . \frac{S}{\sqrt{n}}$ ≤ a ⇔ $\left(z_{\frac{\alpha}{2}} . \frac{S}{a}\right)^2$ ≤ n
      - Nếu σ^2 chưa biết với n < 30:
        - $t_{\frac{\alpha}{2}; n-1} . \frac{S}{\sqrt{n}}$ ≤ a ⇔ $\left(t_{\frac{\alpha}{2}; n-1} . \frac{S}{a}\right)^2$ ≤ n

b. Tỷ lệ: ($\theta = p$)

- Nếu cho f:
  - Sai số $\varepsilon = z_{\frac{\propto}{2}} . \sqrt{\frac{f\,(1-f)}{n}} \leq$ a
- Nếu không cho f:
  - $\sqrt{f\,(1-f)} \leq \frac{f+(1-f)}{2} = \frac{1}{2}$

  → Sai số: $\varepsilon \leq z_{\frac{\propto}{2}} . \frac{1}{2\sqrt{n}} \leq a$ → Cỡ mẫu n = ?

3. <u>Ước lượng khoảng tin cậy 1 phía:</u>

a. Giá trị trung bình: ($\theta = EX$)

- Nếu σ^2 đã biết với n ≥ 30:
  - Khoảng tin cậy bên trái → $(-\infty;\ \bar{X} + z_{\frac{\propto}{2}} . \frac{\sigma}{\sqrt{n}})$
  - Khoảng tin cậy bên phải → $(\bar{X} - z_{\frac{\propto}{2}} . \frac{\sigma}{\sqrt{n}}\ ;\ +\infty)$
- Nếu σ^2 chưa biết với n ≥ 30:
  - Khoảng tin cậy bên trái → $(-\infty;\ \bar{X} + z_{\frac{\propto}{2}} . \frac{S}{\sqrt{n}})$
  - Khoảng tin cậy bên phải → $(\bar{X} - z_{\frac{\propto}{2}} . \frac{S}{\sqrt{n}}\ ;\ +\infty)$
- Nếu σ^2 chưa biết với n < 30:
  - Khoảng tin cậy bên trái → $(-\infty;\ \bar{X} + t_{\frac{\propto}{2};\,(n-1)} \cdot \frac{S}{\sqrt{n}})$
  - Khoảng tin cậy bên phải → $(\bar{X} - t_{\frac{\propto}{2};\,(n-1)} \cdot \frac{S}{\sqrt{n}}\ ;\ +\infty)$

b. Tỷ lệ: ($\theta = p$)

- Khoảng tin cậy bên trái → (0; f $+ z_{\frac{\propto}{2}} . \sqrt{\frac{f\,(1-f)}{n}}$)
- Khoảng tin cậy bên phải → (f $- z_{\frac{\propto}{2}} . \sqrt{\frac{f\,(1-f)}{n}}$; 1)

c. Phương sai: ($\theta = \sigma^2$)

- Khoảng tin cậy bên trái → $\left(0;\ \frac{(n-1)\,S^2}{\chi^2_{1-\propto;\,(n-1)}}\right)$
- Khoảng tin cậy bên phải → $\left(\frac{(n-1)\,S^2}{\chi^2_{\propto;\,(n-1)}};\ +\infty\right)$

Bài 5: BÀI TOÁN KIỂM ĐỊNH GIẢ THIẾT

1. **Tổng quan về kiểm định:**
   - Các bước làm cho bài toán kiểm định giả thiết:
     - B1: Xác định $H_0;\ H_1$
     - B2: Tìm Test thống kê
     - B3: Tìm miền bác bỏ Δ
     - B4: Nếu $\Delta\ xảy\ ra\ thì\ Bác\ bỏ\ H0\ (chấp\ nhận\ H1)$
     - B5: Kết luận

   - Bảng kiểm định giả thiết:

KẾT LUẬN	H0 đúng (H1 sai)	H0 sai (H1 đúng)
Bác bỏ H0 (chấp nhận H1)	SL1	Đúng
Không bác bỏ H0 (không chấp nhận H1)	Đúng	SL2

2. **Bài toán tìm tiêu chuẩn thỏa mãn khi biết tổng thể, mẫu, chọn:**
   - Cho f = $\frac{k}{n}$ biết k là chọn, n là mẫu
   - Độ tin cậy $\propto \rightarrow\ z_{\frac{\propto}{2}}$
   - Xác suất p ∈ (f $\pm z_{\frac{\propto}{2}} \cdot \sqrt{\frac{f\ (1-f)}{n}}$)

     => Số thỏa mãn = pN

     Với N là tổng thể

3. **Kiểm định cho giá trị trung bình GTTB:**
   - Xét bài toán so sánh GTTB ⇔ $\begin{cases} \mu = EX \\ \mu_0\ chưa\ biết \\ mức\ ý\ nghĩa\ \ \propto > \ 0 \end{cases}$
     - BT1 => giả thiết H0: $\mu = \ \mu_0$; đối thiết H1: $\mu\ \neq\ \mu_0$
     - BT2 => giả thiết H0: $\mu \leq\ \mu_0$; đối thiết H1: $\mu\ >\ \mu_0$
     - BT3 => giả thiết H0: $\mu \geq\ \mu_0$; đối thiết H1: $\mu\ <\ \mu_0$

   a) TH1: Nếu DX = σ^2 đã biết, X ~ N(μ, σ^2) với n ≥ 30:

   - Test thống kê → T = $\frac{\bar{X} - \mu_0}{\sigma\ /\ \sqrt{n}}$
   - Miền tiêu chuẩn bác bỏ H0:
     - S1 = { $|T|\ \geq\ z_{\frac{\propto}{2}}$ } → BT1
     - S2 = { $|T|\ \geq\ z_{\propto}$ } → BT2
     - S3 = { $|T|\ \leq\ -\ z_{\propto}$ } → BT3

b) TH2: Nếu phương sai DX chưa biết với n ≥ 30:

- Test thống kê → T = $\frac{\bar{X} - \mu_0}{S / \sqrt{n}}$
- Miền tiêu chuẩn bác bỏ H0:
  - S1 = { $|T| \geq z_{\frac{\propto}{2}}$ } → BT1
  - S2 = { $|T| \geq z_{\propto}$ } → BT2
  - S3 = { $|T| \leq -z_{\propto}$ } → BT3

c) TH3: Nếu phương sai DX chưa biết, X ~ N(μ, σ^2) với n nhỏ < 30:

- Test thống kê → T = $\frac{\bar{X} - \mu_0}{S / \sqrt{n}}$
- Miền tiêu chuẩn bác bỏ H0:
  - S1 = { $|T| \geq t_{\frac{\propto}{2};(n-1)}$ } → BT1
  - S2 = { $|T| \geq t_{\propto;(n-1)}$ } → BT2
  - S3 = { $|T| \leq -t_{\propto;(n-1)}$ } → BT3

4. **<u>Kiểm định cho tỷ lệ hay cho xác suất:</u>**

- Giả sử p = p(A) là tỷ lệ với p_0 $cho\ trước$: ($\propto > 0$)

  $\boxed{f = \frac{k}{n}\ với\ np_0 \geq 5 \text{ và } n(1 - p_0) \geq 5}$

- Test thống kê => T = $\frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$
- Bài toán:
  - BT1 => giả thiết H0: $p = p_0$; đối thiết H1: $p \neq p_0$
    → Miền $\Delta = \{ |T| \geq z_{\frac{\propto}{2}} \}$ = S1
  - BT2 => giả thiết H0: $p \leq p_0$; đối thiết H1: $p > p_0$
    → Miền $\Delta = \{ |T| \geq z_{\propto} \}$ = S2
  - BT3 => giả thiết H0: $p \geq p_0$; đối thiết H1: $p < p_0$
    → Miền $\Delta = \{ |T| \leq -z_{\propto} \}$ = S3

  => Nếu miền thỏa mãn => H0 bác bỏ, H1 chấp nhận

5. **<u>Kiểm định phương sai:</u>**

- Giả sử $\sigma^2 = DX\ chưa\ biết$, ${\sigma_0}^2$ cho trước mức ý nghĩa $\propto\ nhỏ > 0$:
  - Xét bài toán ⇔ $\begin{cases} H0: \sigma^2 = {\sigma_0}^2 \\ \begin{bmatrix} H1: \sigma^2 \neq {\sigma_0}^2 => BT1 \\ H1: \sigma^2 > {\sigma_0}^2 => BT2 \\ H1: \sigma^2 < {\sigma_0}^2 => BT3 \end{bmatrix} \end{cases}$
  - Test thống kê: T = $\frac{(n-1)S^2}{{\sigma_0}^2} \sim \chi^2_{(n-1)}$
- Miền bác bỏ giả thiết H0:

- $\Delta 1 = (0 ; \chi^2_{(n-1)\,;\,1-\frac{\propto}{2}}) \cup (\chi^2_{\frac{\propto}{2}\,;\,(n-1)} ; +\infty)$
  → $\Delta 1 = \{T < \chi^2_{(n-1)\,;\,1-\frac{\propto}{2}}\}$

- $\Delta 2 = \left(\chi^2_{\propto\,;\,(n-1)} ; +\infty\right)$
  → $\Delta 2 = \{T > \chi^2_{(n-1)\,;\,\propto}\}$

- $\Delta 3 = (0 ; \chi^2_{1-\propto\,;\,(n-1)})$
  → $\Delta 3 = \{T < \chi^2_{(n-1)\,;\,1-\propto}\}$

6. <u>Tiêu chuẩn phù hợp χ^2:</u>

- Bài toán kiểm định giả thiết khi cho sẵn số liệu và tỷ lệ xác suất:
  - H0: số liệu phù hợp với tỷ lệ đã cho
  - H1: số liệu đã cho không phù hợp với tỷ lệ đã cho

- Tần số quan sát m_i (ứng với tỷ lệ p_i) → n = $\sum_{i=1}^{k} m_i$
  => Tần số lý thuyết $\widehat{m_i} = np_i$ (được tính khi H0 đúng)

- Test thống kê T = $\sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i} = \frac{1}{n} \sum_{i=1}^{k} (\frac{{m_i}^2}{p_i}) - n$

- Miền bác bỏ H0 là: S = { $T \geq \chi^2_{k-1}(\propto)$ }
  => Tiêu chuẩn sử dụng tốt khi $m_i \geq 5$

Bài 6: BÀI TOÁN KIỂM ĐỊNH HAI TỔNG THỂ

SO SÁNH TRUNG BÌNH HAI TỔNG THỂ

$\mu_1 = EX$ và $\sigma_1 = DX$ $\mu_2 = EY$ và $\sigma_2 = DY$ **GIẢ THIẾT ĐÃ CHO**	+ BT1 => $\begin{cases} H0: \mu_1 = \mu_2 \\ H1: \mu_1 \neq \mu_2 \end{cases}$ + BT2 => $\begin{cases} H0: \mu_1 \leq \mu_2 \\ H1: \mu_1 > \mu_2 \end{cases}$ + BT3 => $\begin{cases} H0: \mu_1 \geq \mu_2 \\ H1: \mu_1 < \mu_2 \end{cases}$

1. **Với mẫu độc lập: (X và Y đều có phân bố chuẩn)**
   a) TH1: Nếu phương sai ${\sigma_1}^2; {\sigma_2}^2$ đã biết và $n_1; n_2 \geq 30$:
      - Test thống kê → T = $\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
      - Miền tiêu chuẩn:
        - S1 = $\left\{|T| \geq z_{\frac{\propto}{2}}\right\}$ → BT1
        - S2 = $\{T \geq z_{\propto}\}$ → BT2
        - S1 = $\{T \leq -z_{\propto}\}$ → BT3

   b) TH2: Nếu phương sai DX, DY chưa biết (${\sigma_1}^2 = {\sigma_2}^2$) và $n_i < 30$:
      - Test thống kê → T = $\frac{\bar{X} - \bar{Y} - \mu}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}} = \frac{\bar{X} - \bar{Y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$
        $với\ \mu\ là\ so\ sánh\ độ\ chênh\ lệch\ giữa\ X\ và\ Y$
      - Do ${\sigma_1}^2 = {\sigma_2}^2$ với $n_i\ nhỏ$ → thay σ^2 là phương sai chung:
        → $S^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}$
        Với bậc tự do df = $n_1 + n_2 - 2$
      - Miền tiêu chuẩn bác bỏ H0:
        - S1 = $\left\{|T| \geq t_{n_1 + n_2 - 2}(\frac{\propto}{2})\right\}$ → BT1
        - S2 = $\left\{T \geq t_{n_1 + n_2 - 2}(\propto)\right\}$ → BT2
        - S1 = $\left\{T \leq -t_{n_1 + n_2 - 2}(\propto)\right\}$ → BT3

c) TH3: Nếu phương sai DX, DY chưa biết (${\sigma_1}^2 \neq {\sigma_2}^2$) và n_i < 30:

- Test thống kê → T = $\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$
- Bậc tự do → df = $\frac{\left(\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}\right)^2}{\frac{\left(\frac{{s_1}^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{{s_2}^2}{n_2}\right)^2}{n_2 - 1}}$
- Miền bác bỏ giả thiết H0:
    - S1 = $\left\{|T| \geq t_{df}(\frac{\propto}{2})\right\}$ → BT1
    - S2 = $\{T \geq t_{df}(\propto)\}$ → BT2
    - S1 = $\{T \leq -t_{df}(\propto)\}$ → BT3

2. <u>Với mẫu phụ thuộc:</u>

- Ta xét D = X – Y sinh ra bài toán:
    - H0: ED = 0 và H1: ED ≠ 0
    - H0: ED ≤ 0 và H1: ED > 0
    - H0: ED ≥ 0 và H1: ED < 0
- Bài toán kiểm định so sánh hai giá trị trung bình sẽ được đưa về bài toán so sánh trung bình giữa D với 0 (1 biến → 1 chiều)

Bài 7: TIÊU CHUẨN PHI THAM SỐ

Không phương sai, không phân bố chuẩn, n → 0

1. **Tiêu chuẩn hạng (Tiêu chuẩn Mann – Whitney):**

- Cho hai mẫu độc lập X và Y biết: $\begin{cases} X = \{x_1, x_2, \ldots, x_n\} \text{ có cỡ mẫu } n \\ Y = \{y_1, y_2, \ldots, y_m\} \text{ có cỡ mẫu } m \end{cases}$
- Kiểm định giả thiết: $\begin{cases} H0: \text{"X và Y cùng phân bố"} \\ H1: \text{"X và Y khác phân bố"} \end{cases}$
- Áp dụng tiêu chuẩn Mann – Whitney:
  - B1: Gộp 2 mẫu m + n
  - B2: Sắp xếp m + n giá trị $(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m)$ theo thứ tự tăng dần:

    $\boxed{c_1 < c_2 < c_3 < \ldots < x_{n+m}}$

    Với $\begin{cases} \text{Nếu } x_i = c_k \Rightarrow \text{Hạng } x_i \text{ là } k \\ \text{Nếu } x_j = c_k \Rightarrow \text{Hạng } x_j \text{ là } k \end{cases}$

  - B3: Tìm tổng các hạng:
    - Giả sử x_i có hạng r_i → Tổng các hạng của x_i là $R_1 = r_1 + r_2 + \ldots + r_n$
    - Giả sử y_i có hạng s_i → Tổng các hạng của y_i là $R_2 = s_1 + s_2 + \ldots + s_m$

      => $R_1 + R_2 = r_1 + r_2 + \ldots + r_n + s_1 + s_2 + \ldots + s_m$

  - B4: Ta chọn R = min(R_1, R_2)
    - Nếu R = R_1 → $\begin{cases} m, n \geq 8 \text{ và } R_1 \sim N(\mu, \sigma^2) \\ \quad H0 \text{ đúng} \end{cases}$
      - ER = $\frac{n(n+m+1)}{2}$
      - DR = $\frac{nm(n+m+1)}{12}$
    - Nếu R = R_2 → $\begin{cases} m, n \geq 8 \text{ và } R_2 \sim N(\mu, \sigma^2) \\ \quad H0 \text{ đúng} \end{cases}$
      - ER = $\frac{m(n+m+1)}{2}$ → μ_R
      - DR = $\frac{nm(n+m+1)}{12}$ → ${\sigma_R}^2$
  - B5: Test thống kê: T = $\frac{R - ER}{\sqrt{DR}} = \frac{R - ER}{\sigma_R}$
  - B6: Xét miền bác bỏ:
    - $|T| > z_{\frac{\alpha}{2}}$ → BT1
    - $|T| > z_{\alpha}$ → BT2 + BT3

- **BẢNG TIÊU CHUẨN:**

Rank	Value	Biến
1	Val1	X
2	Val2	Y
3	Val3*	X, Y
…	…	…
n	Valn	X

➢ Chọn R = min(R_1, R_2)

Chú ý: Nếu (c_i) có nhiều giá trị trùng nhau thì ta quy ước hạng của các giá trị trùng nhau:
$c_{k-1} < c_k = c_{k+1} < c_{k+2}$

➢ ***Hạng c_k = hạng c_{k+1} = $\frac{k+(k+1)}{2}$***

➢ ***Hạng $c_{k+2} = k + 2$***

=> Tổng quát: Hạng $r_i = \frac{\sum \text{các phần tử trùng nhau}}{\text{số phần tử trùng nhau}}$

2. **Tiêu chuẩn dấu: (mẫu phụ thuộc)**

- Cho (X, Y) là cặp gồm 2 đại lượng ngẫu nhiên: ⇔ $\begin{cases} X: \text{"Hiệu quả phương pháp 1"} \\ Y: \text{"Hiệu quả phương pháp 2"} \end{cases}$

  → Tác động lên cùng một cá thể

- Xét bài toán:

$$\begin{cases} H0: \text{"Hiệu quả phương pháp 1 và phương pháp 2 như nhau"} \\ H1: \left[\begin{array}{l} \text{"Hiệu quả phương pháp 1 và phương pháp 2 khác nhau"} \rightarrow BT1 \\ \text{"Hiệu quả phương pháp 1} > \text{phương pháp 2"} \rightarrow BT2 \\ \text{"Hiệu quả phương pháp 1} < \text{phương pháp 2"} \rightarrow BT3 \end{array}\right. \end{cases}$$

- Có (X, Y) = (x_1, y_1); (x_1, y_1); …
  - B1: Đặt $d_i = x_i - y_i$ và loại bỏ các $d_i = 0$
    - Biết $\begin{cases} n^+ \text{ là các hạng } d_i \ (\text{với } d_i > 0) \\ n^- \text{ là các hạng } d_i \neq 0 \end{cases}$
    - Nếu H0 đúng thì số hạng dấu + có xu hướng mang –
  - B2: Độ lệch chuẩn $\sigma = \frac{1}{2\sqrt{\tilde{n}}}$
    - Nếu H0 đúng → n^+ có phân bố nhị thức $\begin{cases} p = 0{,}5 \\ \tilde{n} \end{cases}$ với f = $\frac{n^+}{\tilde{n}}$

- B3: Test thống kê: T = (f – 0,5) 2 $\sqrt{\tilde{n}}$ = $\frac{2n^{+} - \tilde{n}}{\sqrt{\tilde{n}}}$
- B4: Miền bác bỏ (Δ) ⇔ $\left[\begin{array}{l} |T| > z_{\frac{\alpha}{2}} \rightarrow BT1 \\ |T| > z_{\alpha} \rightarrow \left[\begin{array}{l} T > z_{\alpha} \rightarrow BT2 \\ T < -z_{\alpha} \rightarrow BT3 \end{array}\right. \end{array}\right.$

- **BẢNG TIÊU CHUẨN:**

X	Y	$d_i = x_i - y_i$	Dấu
x_1	y_1	$x_1 - y_1$	+
x_2	y_2	$x_2 - y_2$	-
x_3	y_3	$x_3 - y_3$	0
…	…	…	…
x_n	y_n	$x_n - y_n$	+

  - Với $\tilde{n} = count\ (dấu \neq 0)$
  - Với n^{+}= count (dấu > 0)

3. **Tiêu chuẩn hạng có dấu Wilcoxon: (mẫu phụ thuộc)**

- B1: Đặt $d_i = x_i - y_i$ và loại bỏ các $d_i = 0$

- B2: Tính hạng (d_i) với $d_i \neq 0$
  - $\tilde{n}$ là các $d_i \neq 0$
  - R^{+} là $Tổng\ các\ hạng$ $|d_i|$ với $d_i > 0$
  - R^{-} là $Tổng\ các\ hạng$ $|d_i|$ với $d_i < 0$

- B3: Nếu H0 đúng thì R^{+} và R^{-} có cùng phân bố với $\tilde{n} \geq 8$:
  - Kỳ vọng → $\mu = ER = \frac{\tilde{n}\ (\tilde{n}+1)}{4}$
  - Phương sai → S^2 = DR = $\frac{\tilde{n}\ (\tilde{n}+1)\ (2\tilde{n}+1)}{24}$

- B4: Test thống kê → T = $\frac{R - ER}{\sqrt{DR}}$
  - Chọn R = min(R^{+}, R^{-})

- B5: Kết luận: $\left[\begin{array}{l} |T| > z_{\frac{\alpha}{2}} \rightarrow BT1 \\ |T| > z_{\alpha} \rightarrow \left[\begin{array}{l} T > z_{\alpha} \rightarrow BT2 \\ T < -z_{\alpha} \rightarrow BT3 \end{array}\right. \end{array}\right.$

- **BẢNG TIÊU CHUẨN:**

X	Y	$d_i = x_i - y_i$	Rank($\lvert d_i \rvert$) ($d_i > 0$)	Rank($\lvert d_i \rvert$) ($d_i < 0$)
x_1	y_1	$x_1 - y_1$	rank($\lvert d_1 \rvert$)	
x_2	y_2	$x_2 - y_2$		rank($\lvert d_2 \rvert$)
x_3	y_3	$x_3 - y_3$		
…	…	…	…	…
x_n	y_n	$x_n - y_n$	rank($\lvert d_n \rvert$)	
			R^+	R^-

- Chọn $\begin{cases} \mathrm{R} = \min(R^+, R^-) \\ \tilde{n} \text{ là } count\ (d_i \neq 0) \end{cases}$

Bài 8: SO SÁNH HAI TỶ LỆ VÀ PHƯƠNG SAI

1. **So sánh hai tỷ lệ:**
   - Giả sử p_1, p_2 là hai tỷ lệ chưa biết → Cần so sánh:
     - H0: $p_1 = p_2$ với f = $\frac{k_1 + k_2}{n_1 + n_2}$
     - H1: $\begin{cases} p_1 \neq p_2 \rightarrow BT1 \\ p_1 > p_2 \rightarrow BT2 \\ p_1 < p_2 \rightarrow BT3 \end{cases}$
   - Test thống kê: T = $\frac{f_1 - f_2}{\sqrt{f\,(1-f)\,(\frac{1}{n_1} + \frac{1}{n_2})}}$
   - Miền bác bỏ:
     - S1 = { $|T| \geq z_{\frac{\propto}{2}}$ }
     - S2 = { $|T| \geq z_{\propto}$ }
     - S3 = { $|T| \leq -z_{\propto}$ }

2. **So sánh hai phương sai:**
   - Giả sử X, Y có phân bố chuẩn với ${\sigma_1}^2, {\sigma_2}^2$ là phương sai chưa biết:
     - H0: ${\sigma_1}^2 = {\sigma_2}^2$ → giả thiết
     - H1: $\begin{bmatrix} {\sigma_1}^2 \neq {\sigma_2}^2 \rightarrow BT1 \\ {\sigma_1}^2 > {\sigma_2}^2 \rightarrow BT2 \\ {\sigma_1}^2 < {\sigma_2}^2 \rightarrow BT3 \end{bmatrix}$
   - ➢ Tính chất phân phối mẫu T = $\frac{{S_1}^2/{\sigma_1}^2}{{S_2}^2/{\sigma_2}^2}$ theo phân bố Fisher ($n_1 - 1;\ n_2 - 1$)
   - Test thống kê → T = $\frac{{S_1}^2}{{S_2}^2}$
   - Miền bác bỏ:
     - S1 = $(0;\ f_{n_1 - 1, n_2 - 1}(1 - \frac{\propto}{2})) \cup (f_{n_1 - 1, n_2 - 1}(\frac{\propto}{2});\ inf)$
     - S2 = $(f_{n_1 - 1, n_2 - 1}(\propto);\ inf)$
     - S3 = $(0;\ f_{n_1 - 1, n_2 - 1}(1 - \propto))$

Bài 9: PHÂN TÍCH PHƯƠNG SAI

1. **Phân tích phương sai 1 nhân tố:**
   - Cho mẫu các mức nhân tố:

Các mức nhân tố					Tổng số
	1	**2**	**3**	**4**	
$\boldsymbol{x_1}$	T_{11}	T_{12}	T_{13}	T_{14}	
$\boldsymbol{x_2}$	T_{21}	T_{22}	T_{23}	T_{24}	
$\boldsymbol{x_3}$	T_{31}	T_{32}	T_{33}	T_{34}	
…	…	…	…	…	
$\boldsymbol{n_i}$	n_1	n_2	n_3	n_4	n
$\boldsymbol{T_i}$	$\sum T_1$	$\sum T_2$	$\sum T_3$	$\sum T_4$	T

   - Cho: $\begin{cases} H0: \mu_1 = \mu_2 = \ldots = \mu_n \\ H1: \exists\, i, j \text{ sao cho } y_i \neq y_j \end{cases}$
     - Trung bình mẫu thứ i:
       - $\overline{X_\iota} = \frac{T_i}{n_i} = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i}$
     - Trung bình mẫu:
       - $\bar{X} = \frac{T}{n} = \frac{\sum\sum x_{ij}}{n}$
     - Tổng bình phương chung:
       - SST = $\sum_i \sum_j \left(X_{ij} - \bar{X}\right)^2 = \sum_{i,j}\left(x_{ij}\right)^2 - \frac{T^2}{n} = \sum {T_i}^2 - \frac{T^2}{n}$
     - Tổng bình phương về sự khác nhau giữa nhóm:
       - SSF = $\sum \frac{{T_i}^2}{n_i} - \frac{T^2}{n} = \frac{{T_1}^2}{n_1} + \frac{{T_2}^2}{n_2} + \ldots + \frac{{T_k}^2}{n_k} - \frac{T^2}{n}$
     - Tổng bình phương sai số trong từng nhóm:
       - SSE = $\sum_i \sum_j \left(X_{ij} - \overline{X_J}\right)^2$ = SST – SSF

   ➔ SST = SSF + SSE

   - Cho bảng phân tích phương sai:

Bảng phân tích phương sai				
	Bậc tự do	**Tổng bình phương**	**Trung bình bình phương**	**Tỷ số F**
Nhân tố	k - 1	SSF	MSF = $\frac{SSF}{k-1}$	F = $\frac{MSF}{MSE}$
Sai số	n - k	SSE	MSE = $\frac{SSE}{n-k}$	
Tổng	n - 1	SST		

➔ F có phân bố Fisher(k – 1, n – k)

- Miền bác bỏ H0 → $\Delta = \{F \geq f_{k-1;\, n-k}\}$

2. **Phân tích phương sai 2 nhân tố:**

- $\begin{cases} A, B \text{ là 2 nhân tố} \\ \left| \begin{array}{l} A \text{ có } s \text{ mức tác động} \\ B \text{ có } r \text{ mức tác động} \end{array} \right. \end{cases}$

- Bảng các nhân tố:

		Nhân tố A				
		Mức 1	Mức 2	...	Mức s	Tổng
Nhân tố B	Mức 1	x_{11}	x_{12}	...	x_{1s}	T_{10}
	Mức 2	x_{21}	x_{22}	...	x_{2s}	T_{20}
	...	...	...	...	...	...
	Mức r	x_{r1}	x_{r2}	...	x_{rs}	T_{r0}
Tổng		T_{01}	T_{02}	...	T_{0s}	T

BẢNG PHÂN TÍCH PHƯƠNG SAI

	Bậc tự do	Tổng bình phương	TBBP	Tỷ số F
Nhân tố A	s - 1	SSF_A	MSF_A	$F_A = \frac{MSF_A}{MSE}$
Nhân tố B	r - 1	SSF_B	MSF_B	$F_B = \frac{MSF_B}{MSE}$
Tương tác	(s – 1)(r – 1)	SSI	MSI	$F_{AB} = \frac{MSI}{MSE}$
Sai số	n – rs	SSE	MSE	
Tổng	n - 1	SST		

- Miền bác bỏ giả thiết (3 bài toán):
  - $\Delta_1 = \{F_A \geq f_{s-1;\, n-rs;\, \propto}\}$
  - $\Delta_2 = \{F_B \geq f_{r-1;\, n-rs;\, \propto}\}$
  - $\Delta_1 = \{F_{AB} \geq f_{(s-1)(r-1);\, n-rs;\, \propto}\}$

- ***Một số lưu ý về công thức:***
  - $A = \sum X^2$
  - $SST = A - \frac{T^2}{n}$
  - $SSF_A = \frac{T_{01}{}^2}{n_{01}} + \frac{T_{02}{}^2}{n_{02}} + \dots + \frac{T_{0s}{}^2}{n_{0s}} - \frac{T^2}{n} = \frac{\sum T_{hàng}{}^2}{\sum n_{hàng}{}^2} - \frac{T^2}{n}$
  - $SSF_B = \frac{T_{10}{}^2}{n_{10}} + \frac{T_{20}{}^2}{n_{20}} + \dots + \frac{T_{r0}{}^2}{n_{r0}} - \frac{T^2}{n} = \frac{\sum T_{cột}{}^2}{\sum n_{cột}{}^2} - \frac{T^2}{n}$
  - $SSE = A - (\frac{T_{11}{}^2}{n_{11}} + \dots + \frac{T_{rs}{}^2}{n_{rs}})$
  - $SSI = SST - (SSF_A + SSF_B + SSE)$

Bài 10: KIỂM TRA TÍNH ĐỘC LẬP VÀ SO SÁNH

1. **Kiểm tra tính độc lập:**
   - BGKĐGT:
     - H_0: "X và Y độc lập với nhau"
     - H_1: "X và Y không độc lập với nhau"

   - Số liệu quan sát:

	y_1	…	y_c	$\sum hàng$
x_1	n_{11}	…	n_{1c}	n_{10}
…	…	…	…	…
x_r	n_{r1}	…	n_{rc}	n_{r0}
$\sum cột$	n_{01}	…	n_{0c}	n

   => với n_{ij} là số lần mà giá trị X = x_i và Y = y_j → Tần số quan sát

   - Tần số quan sát:
     - $\widetilde{n_{ij}} = \frac{n_{i0}\, n_{0j}}{n} = \frac{Tổng\ hàng\ i\ .\ Tổng\ cột\ j}{n}$

   - Test thống kê:
     - T = n $(\sum_{i,j} \frac{{n_{ij}}^2}{n_{i0}\, n_{0j}} - 1)$ = n $(\frac{{n_{11}}^2}{n_{01}\, n_{10}} + \ldots + \frac{{n_{r0}}^2}{n_{0c}\, n_{r0}} - 1)$

   - Miền bác bỏ:
     - $\Delta = \{T \geq \chi_{(r-1)(c-1);\, \propto}\}$

2. **So sánh nhiều tỷ lệ:**
   - BTKĐGT:
     - $H_0: p_1 = p_2 = \ldots = p_k$ -→ phân bố giống
     - H_1: k tỷ lệ $(p_i)_{i=1}^{k}$ khác nhau → phân bố khác

	Mẫu 1	…	Mẫu K	$\sum hàng$
Số lần xuất hiện A	m_1	…	m_k	m
Số lần xuất hiện $\bar{A}$	$n_1 - m_1$	…	$n_k - m_k$	n - m
$\sum cột$	n_1	…	n_k	n

- Test thống kê:
  - T = n ($\frac{{m_1}^2}{n_1\, m}$ + ... + $\frac{(n_k - m_k)}{(n - m)\, n_k}$ - 1)

- Miền bác bỏ: $\Delta\ =\ \{T\ \geq\ {\chi_{k\,-\,1;\,\propto}}^2\}$
  (gộp lại nếu < 5)

3. **<u>So sánh chính xác tỷ lệ:</u>**
- p_i tỷ lệ cần tính:
  - f_1 = $\frac{k_1}{m_1}$; f_2 = $\frac{k_2}{m_2}$; ... ; f_k = $\frac{k_k}{m_k}$

- Ta có $f_1\ <\ f_2\ <\ f_3\ <\ldots<\ f_k$ → kiểm tra lần lượt $f_1\ <\ f_2$
  => So sánh 2 tỷ lệ $\begin{cases} H_0: p_1\ >\ p_2 \\ H_1: p_1 <\ p_2 \end{cases}$

  + Nếu bác bỏ H_0=> H_1 đ$úng\ \ =>\ \ lấy\ kết\ quả$
  + Nếu không bác bỏ => xét tiếp

Bài 11: HỒI QUY TUYẾN TÍNH

1. **Hệ số tương quan:**
   - HSTQ lý thuyết:
     - $\rho(x,y) = \frac{cov(X,Y)}{\sqrt{DX\,DY}} = \frac{E[(X-EX)(Y-EY)]}{\sqrt{DX\,DY}}$
     - $0 \le |\rho| \le 1$
       - $Nếu\ |\rho| = 0 \Rightarrow phụ\ thuộc$
       - $Nếu\ |\rho| = 1 \Rightarrow độc\ lập$

   - HSTQ có mẫu quan sát $(x_i;\ y_i)_{i=1}^{n}$ (theo cặp):
     - r = $\frac{\sum (X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum(X_i-\bar{X})^2\sum(Y_i-\bar{Y})^2}} = \frac{n\sum XY - \sum X\sum Y}{\sqrt{[n\sum X^2-(\sum X)^2][n\sum Y^2-(\sum Y)^2]}}$
       - r < 0 => mối quan hệ nghịch biến
       - r > 0 => mối quan hệ đồng biến

2. **Các dạng bài tập hồi quy:**

   a) *Dạng 1: Kiểm tra xem có mối quan hệ tuyến tính tương quan không?*
   - $\begin{cases} \rho = 0 => X,Y\ không\ tương\ quan \\ \rho \ne 0 => X,Y\ có\ tương\ quan \end{cases}$
   - Test thống kê → T = $\frac{r\sqrt{n-2}}{1-r^2}$ với r là hệ số tương quan
   - Miền bác bỏ giả thiết:
     - $\Delta = \{|T| \ge t_{n-2;\frac{\alpha}{2}}\}$

   b) *Dạng 2: Kiểm tra mối quan hệ tương quan của nhân tố có giống nhau không? (tham số cho trước)?*
   - $\begin{cases} \rho = \rho_0 \\ \rho \ne \rho_0 \end{cases} (với\ \rho_0 \ne 0\ cho\ trước)$
   - Ta đặt: $\begin{cases} u = \frac{1}{2}ln(\frac{1+r}{1-r});\ \sigma = \frac{1}{\sqrt{n-3}} \\ m = \frac{1}{2}ln(\frac{1+\rho_0}{1-\rho_0}) \end{cases}$
   - $Test\ thống\ kê$: T = $\frac{u-m}{\sigma}$
   - Miền bác bỏ: $\Delta = \{|T| \ge Z_\alpha\}$

   c) *Dạng 3: Xác định KTC cho hệ số tương quan ρ_0?*
   - Ta có $\frac{u-m}{\sigma}$ ~ N(0, 1)
   - Ta xét:

$P(-Z_{\frac{\propto}{2}} < \frac{u-m}{\sigma} < Z_{\frac{\propto}{2}}) = 1 - \propto$

$P(-\sigma Z_{\frac{\propto}{2}} < u - m < \sigma Z_{\frac{\propto}{2}}) = 1 - \propto$

$P(u - \sigma Z_{\frac{\propto}{2}} < m < u + \sigma Z_{\frac{\propto}{2}}) = 1 - \propto$

Với $\theta_1 = u - \sigma Z_{\frac{\propto}{2}}$; $\theta = m$; $\theta_2 = u + \sigma Z_{\frac{\propto}{2}}$

- Khoảng tin cậy 1 - ∝ $của\ m\ là$: $(m_1;\ m_2)$ = $(u - \sigma Z_{\frac{\propto}{2}}$; $u + \sigma Z_{\frac{\propto}{2}})$
- Khoảng tin cậy 1 - ∝ $của\ \rho\ là$: $(\frac{e^{2m_1}-1}{e^{2m_1}+1};\ \frac{e^{2m_2}-1}{e^{2m_2}+1})$

d) *Dạng 4: Kiểm tra A, B có phụ thuộc tuyến tính không? (phi tuyến)*

- BTKĐGT:
  - H_0: $\vartheta^2 - \rho^2 = 0$
  - H_1: $\vartheta^2 - \rho^2 \neq 0$ → có tương quan phi tuyến

- Test thống kê:
  - F = $\frac{\frac{\hat{\vartheta}^2 - r^2}{k-2}}{\frac{1-\hat{\vartheta}^2}{n-k}} = \frac{(\hat{\vartheta}^2 - r^2)(n-k)}{(1-\hat{\vartheta}^2)(k-2)}$
  - Biết $\hat{\vartheta}^2 = \frac{SSE}{SST}$ => chỉ số tương quan lý thuyết

- Miền bác bỏ → $|F| \geq f_{(k-2;\, n-k;\, \frac{\propto}{2})}$

- Chú thích:
  - SST = $\sum {x_{ij}}^2 - \frac{T^2}{n}$
  - SSF = $\sum \frac{{T_i}^2}{n_i} - \frac{T^2}{n}$
  - Tỷ số tương quan → $\vartheta_{Y/X} = 1 - \frac{E[Y - E(Y/X)]^2}{DY}$

3. Hồi quy tuyến tính thực nghiệm:

Y = a + bX + ε => HQTT

$\boldsymbol{\hat{\varepsilon}_i = y_i - \hat{y}_i}$

- a = $\frac{\sum y - b\sum x}{n}$ = A
- b = $\frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$
- Sai số tiêu chuẩn → ${S_{y/x}}^2 = \frac{\sum y^2 - b\sum xy - a\sum y}{n-2} = \frac{\sum {\hat{\varepsilon}_i}^2}{n-2}$

+) Y theo X → Y = a + bX (X độc lập, Y phụ thuộc)

+) X theo Y → X = a + bY (Y độc lập, X phụ thuộc)

4. **<u>Kiểm tra sự phù hợp của mô hình (có hồi quy tuyến tính giữa X và Y không?)</u>**

Y = a + bX + ε

- BTKĐGT: $\begin{cases} b = 0 \\ b \neq 0 \end{cases}$ $với\ b\ là\ hệ\ số\ hồi\ quy$
- Test thống kê: T = $\frac{b}{S_B}$ với $S_B = \frac{S_{Y/X}}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{n}}}$
- Miền bác bỏ giả thiết: $\Delta = \{|T| \geq t_{n-2;\frac{\alpha}{2}}\}$

5. **<u>Hồi quy bội đưa về tuyến tính:</u>**

Y = a + bx^m

- Cách 1: đặt z = x^m $=> y = a + bz$ (x_i, y_i) -----> (z_i, y_i)
- Cách 2: đặt Z = $a' + bX$

6. **<u>Bài toán dự báo:</u>**
   - Mô hình: ***Y = a + bX + ε***
   - Mô hình thực nghiệm: ***Y = A + BX + ε***

+) Dự báo cho Y khi X = X_0 là → $\widehat{Y_0} = A + BX_0$

+) Dự báo cho EY khi X = X_0 là → E(Y/X = X_0)

➢ KTC 1 - ∝ cho Y khi x = x_0 là:

$(\widehat{Y_0} \pm t_{n-2;\frac{\alpha}{2}}; S_{Y/X}\sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum X^2 - \frac{1}{n}(\sum X)^2}})$

➢ KTC 1 - ∝ cho EY khi x = x_0 là:

$(\widehat{Y_0} \pm t_{n-2;\frac{\alpha}{2}}; S_{Y/X}\sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum X^2 - \frac{1}{n}(\sum X)^2}})$

7. **<u>Tỷ số tương quan với X và Y độc lập:</u>**

- ${\vartheta_{Y/X}}^2 = \frac{SSF}{SST}$ với $\begin{cases} \rho^2(X,Y) = \rho^2(Y,X) \\ {\vartheta_{Y/X}}^2 = {\vartheta_{X/Y}}^2 \end{cases}$
- Cho $0 \leq r^2 \leq {\vartheta_{Y/X}}^2 \leq 1$:
  - SST = $\sum {x_{ij}}^2 - \frac{T^2}{n}$
  - SSF = $\sum \frac{{T_i}^2}{n_i} - \frac{T^2}{n}$

Y đối với X	X1	X2	X3	...	Xk	
Y1	x_{11}	x_{12}	x_{13}	...	x_{1k}	
Y2	x_{21}	x_{22}	x_{23}	...	x_{2k}	
...	...	...	...	...	...	
Yk	x_{k1}	x_{k2}	x_{k3}	...	x_{kk}	
	T1	T2	T3	...	Tk	T = $\sum_{i=1}^{k} T_i$
	n1	n2	n3	...	nk	n = $\sum_{i=1}^{k} n_i$