PHẦN I: XÁC SUẤT

1. Biến cố ngẫu nhiên & xác suất của biến cố:

1.1. Công thức cộng xác suất:

1.1.1. p(A+B)=p(A)+p(B) (2 biến cố xung khắc)

1.1.2. p(A+B)=p(A)+p(B)-p(A.B) → p(A+B+C)=p(A)+p(B)+p(C)-[p(AB)+p(AC)+p(BC)]+p(ABC)

1.2. Công thức nhân xác suất:

1.2.1. p(A.B)=p(A).p(B) (2 biến cố độc lập)

1.2.2. p(A.B)=p(A).p(B/A) → $p(A_1A_2...A_n) = p(A_1).p(A_2/A_1)...p(A_n/A_1A_2..A_{n-1})$

1.3. Công thức Bernoulli: cho 2 biến cố A và $\overline{A}$

1.3.1. $p_n(x) = C_n^x p^x q^{n-x}$, p=p(A), q=1-p

1.4. Công thức xác suất đầy đủ:

$$p(F) = p(A_1).p(F/A_1) + p(A_2).p(F/A_2) + ... + p(A_n).p(F/A_n)$$

1.5. Công thức Bayes: $p(A_i/F) = \dfrac{p(A_i.F)}{p(F)} = \dfrac{p(A_i).p(F/A_i)}{p(F)}$

2. Biến ngẫu nhiên:

2.1. Bảng phân phối xác suất (biến ngẫu nhiên rời rạc)

2.2. Hàm mật độ xác suất ($f(x)$) (biễn ngẫu nhiên liên tục)

2.2.1. $f(x) \geq 0$

2.2.2. $\int_{-\infty}^{+\infty} f(x)dx = 1$

2.2.3. $p(a \leq x \leq b) = \int_a^b f(x)dx$

2.3. Hàm phân phối xác suất ($F(x)$) (dùng cho cả 2 loại biến-thường là biến ngẫu nhiên liên tục)

2.3.1. $F(x)$=p(F <x)

2.3.2. $F'(x) = f(x)$

2.3.3. $F(x) = \int_{-\infty}^{x} f(t)dt$

2.4. Kỳ vọng

2.4.1. $E(x) = x_1p_1 + x_2p_2 + ... + x_np_n$ (từ bảng phân phối xác suất)

2.4.2. $E(x) = \int_{-\infty}^{+\infty} xf(x)dx$

2.5. Phương sai:

2.5.1. $V(x) = E(x^2) - [E(x)]^2$

2.5.2. $V(x)=\int_{-\infty}^{+\infty} x^2 f(x)dx-[\int_{-\infty}^{+\infty} xf(x)dx]^2$

3. Một số phân phối xác suất thông dụng:

3.1.Phân phối chuẩn tổng quát: $X \sim N(\mu;\sigma^2)$

3.1.1. $f(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

3.1.2. $\int_{-\infty}^{+\infty} f(x)dx=1$

3.1.3. $ModX=MedX=\mu\,;E(x)=\mu,V(x)=\sigma^2$

3.1.4. $p(a\le x\le b)=\varphi(\dfrac{b-\mu}{\sigma})-\varphi(\dfrac{a-\varphi}{\sigma})$

3.1.5. Phân phối chuẩn tắc $\mu=0,\sigma^2=1$

3.1.5.1. $T\sim N(0,1)$

3.1.5.2. $f(t)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$

3.1.5.3. Đổi biến $T=\dfrac{X-\mu}{\sigma}$

3.1.5.4. $p(a\le x\le b)=\varphi(b)-\varphi(a)$

3.2.Phân phối Poisson: $X\sim P(\lambda)$, λ ***>0***

3.2.1. $p(\lambda=k)=e^{-\lambda}\dfrac{\lambda^k}{k!}$

3.2.2. $E(x)=V(x)=\lambda$

3.3.Phân phối nhị thức**:** $X\sim B(n,p)$

3.3.1. $p(X=k)=p_n(k)=C_n^k p^k q^{n-k}, p+q=1$

3.3.2. $\sum_{k=0}^{n} p(X=k)=1$

3.3.3. $E(x)=np\,,ModX=x_0,np-q\le x_0\le np+q$

3.3.4. Khi n=1: $X\sim B(1,p)$:phân phối không-một

3.3.4.1. $E(x)=p,E(x^2)=p,V(x)=pq$

3.3.5. Xấp xỉ phân phối nhị thức:

3.3.5.1. Bằng phân phối Poisson: $n>50$, $p<0.1$; $X\sim B(n,p)\approx X\sim P(\lambda),\lambda=np$.

$$p(x=k)=C_n^k p^k q^{n-k}=e^{-\lambda}\frac{\lambda^k}{k!}$$

3.3.5.2. Bằng phân phối chuẩn:

$$np \ge 0.5, nq \ge 0.5, \mu = np, \sigma = \sqrt{npq}\ .\ X \sim B(n,p) \approx X \sim N(np, npq)$$

$$p(x=k) = \frac{1}{\sigma} f(\frac{k-\mu}{\sigma})\ ;\ p(k_1 < X < k_2) = \varphi(\frac{k_2-\mu}{\sigma}) - \varphi(\frac{k_1-\mu}{\sigma})$$

3.4.*Phân phối siêu bội:* $X \sim H(N, N_A, n)$ [N:tổng số phần tử, N_A:Số phần tử có tính chất A trong N, n: số phần tử lấy ngẫu nhiên].Gọi X là số phần tử có tính chất A trong n.

$$p(X=k) = \frac{C_{N_A}^k . C_{N-N_A}^{n-k}}{C_N^n}$$

3.4.1. $E(X) = np, p = \frac{N_A}{N}\ ; V(X) = npq.\frac{N-n}{N-1}, q = 1-p$

3.4.2. Xấp xỉ phân phối siêu bội bằng phân phối nhị thức:

$$n \le 0.05N \Rightarrow X \sim B(n,p)\ ;\ p(X=k) = C_n^k p^k q^{n-k}, p = \frac{N_A}{N}$$

3.5.*Biến ngẫu nhiên 2 chiều: X* và Y độc lập $\Leftrightarrow P_{ij} = p(x_i).q(y_j)$ với mọi i,j

3.6.*Hiệp phương sai và hệ số tương quan:*

3.6.1. Hiệp phương sai(cov): $\text{cov}(X,Y) = E(XY) - E(X)E(Y)$

3.6.2. Hệ số tương quan $\rho_{X,Y}$: $\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma(X)\sigma(Y)}$

PHẦN 2: THỐNG KÊ

1. Tổng thể và mẫu

1.1.Thực hành tính toán trên mẫu:

1.1.1. Tính trung bình ($\overline{X_n}$): $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} x_i$

1.1.2. Tính tỷ lệ mẫu: (f_n); $f_n = \frac{m_A}{n}$ (m_A:số phần tử mang tính chất A; n: kích thước mẫu)

1.1.3. Tính phương sai mẫu: $S^2 = \frac{1}{n-1}[\sum_{1}^{k} n_i x_i^2 - n(\overline{X})^2]$

1.2.Ước lượng tham số của tổng thể:

1.2.1. Ước lượng điểm: $E(X_n) = \mu, E(f_n) = p, E(S^2) = \sigma^2$

1.2.2. Ước lượng khoảng:

1.2.2.1. Ước lượng khoảng cho trung bình: Với độ tin cậy 1-α cho trước, 1 mẫu kích thước n.

$n \ge 30, \sigma^2$ **biết**	$n \ge 30, \sigma^2$ **chưa biết**
$\overline{X}, \sigma$ $\mu_1 = \overline{X} - \varepsilon, \mu_2 = \overline{X} + \varepsilon$	$\overline{X}$,s $\mu_1 = \overline{X} - \varepsilon, \mu_2 = \overline{X} + \varepsilon$

$\varepsilon = u_{\frac{\alpha}{2}}.\frac{\sigma}{\sqrt{n}}$ $(1-\alpha \rightarrow 0.5\text{-}\frac{\alpha}{2} \rightarrow u_{\frac{\alpha}{2}})$	$\varepsilon = u_{\frac{\alpha}{2}}.\frac{s}{\sqrt{n}}$ $(1-\alpha \rightarrow 0.5\text{-}\frac{\alpha}{2} \rightarrow u_{\frac{\alpha}{2}})$
n <30, σ^2 biết	**n <30, σ^2 chưa biết**
Như TH1	$\overline{X}$,s $\mu_1 = \overline{X} - \varepsilon, \mu_2 = \overline{X} + \varepsilon$ $\varepsilon = t_{(n-1,\frac{\alpha}{2})}.\frac{s}{\sqrt{n}}$

1.2.2.2. Ước lượng khoảng cho tỷ lệ: tổng thể có tỷ lệ p chưa biết, với độ tin cậy $1-\alpha$ cho trước, với 1 mẫu kích thước n, tỷ lệ mẫu f_n. Tìm 2 số p_1, p_2 thoả:

$p(p_1 \le p \le p_2) = 1-\alpha$, $p_{1,2} = f_n \mp \varepsilon$ Công thức: $\varepsilon = u_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}$

1.2.2.3. Ước lượng khoảng cho phương sai:Giả sử tổng thể có σ^2 chưa biết. Dựa vào 1 mẫu kích thước n, với độ tin cậy 1-α cho trước.

TH1: μ chưa biết, biết S^2. Khi đó ta có $\sigma^2 \in [\frac{(n-1)S^2}{\chi_1^2}, \frac{(n-1)S^2}{\chi_2^2}]$ trong đó

$\chi_1^2 = \chi^2(n-1,\frac{\alpha}{2})$, $\chi_2^2 = \chi^2(n-1,1-\frac{\alpha}{2})$

TH2: μ biết. Khi đó $\sigma^2 \in [\frac{\sum n_i(x_i-\mu)}{\chi_1^2}, \frac{\sum n_i(x_i-\mu)}{\chi_2^2}]$, trong đó

$\chi_1^2 = \chi^2(n,\frac{\alpha}{2})$, $\chi_2^2 = \chi^2(n,1-\frac{\alpha}{2})$

1.2.3. Kiểm định giả thuyết thống kê:

1.2.3.1. Kiểm định giả thuyết thống kê cho μ

1.2.3.1.1.TH1: σ^2 **biết**

Giả thuyết thống kê	W_α **:** σ^2 **biết** (miền bác bỏ H_0)
$H_0: \mu = \mu_0$ $H_1: \mu \ne \mu_0$	$W_\alpha = \{u = \frac{\overline{X}-\mu_0}{\sigma}\sqrt{n}, \lvert u \rvert > u_{\frac{\alpha}{2}}\}$
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$W_\alpha = \{u = \frac{\overline{X}-\mu_0}{\sigma}\sqrt{n}$,u<-u_α }
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$W_\alpha = \{u = \frac{\overline{X}-\mu_0}{\sigma}\sqrt{n}$,u>u_α }

1.2.3.1.2.TH2: $n \ge 30, \sigma^2$ không biết

Giả thuyết thống kê	W_α (miền bác bỏ H_0)
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$W_\alpha = \{u = \frac{\overline{X} - \mu_0}{s}\sqrt{n}, \|u\| > u_{\frac{\alpha}{2}}\}$
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$W_\alpha = \{u = \frac{\overline{X} - \mu_0}{s}\sqrt{n}$,u<-u_α }
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$W_\alpha = \{u = \frac{\overline{X} - \mu_0}{s}\sqrt{n}$,u>u_α }

1.2.3.1.3.TH3: $n<30, \sigma^2$ không biết

Giả thuyết thống kê	W_α (miền bác bỏ H_0)
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$W_\alpha = \{t = \frac{\overline{X} - \mu_0}{s}\sqrt{n}, \|t\| > t_{(n-1,\frac{\alpha}{2})}\}$
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$W_\alpha = \{t = \frac{\overline{X} - \mu_0}{s}\sqrt{n}, t < -t_{(n-1,\frac{\alpha}{2})}\}$
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$W_\alpha = \{t = \frac{\overline{X} - \mu_0}{s}\sqrt{n}, t > t_{(n-1,\frac{\alpha}{2})}\}$

1.2.3.2. Kiểm định giả thuyết thống kê cho tỷ lệ:

Giả thuyết thống kê	W_α (miền bác bỏ H_0)
$H_{0:} p = p_0$ $H_{1:} p \neq p_0$	$W_\alpha = \{u = \frac{f - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}, \|u\| > u_{\frac{\alpha}{2}}\}$
$H_{0:} p = p_0$ $H_{1:} p < p_0$	$W_\alpha = \{u = \frac{f - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}, u < -u_\alpha\}$
$H_{0:} p = p_0$ $H_{1:} p > p_0$	$W_\alpha = \{u = \frac{f - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}, u > u_\alpha\}$

1.2.3.3. Kiểm định giả thuyết thống kê cho phương sai:

1.2.3.3.1.TH1: μ chưa biết

Giả thuyết thống kê	W_α (miền bác bỏ H_0)
$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$	$W_\alpha = \{\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}, \chi^2 < \chi_1^2$ hoặc $\chi^2 > \chi_2^2$ $\chi_1^2 = \chi^2_{(n-1,1-\frac{\alpha}{2})}, \chi_2^2 = \chi^2_{(n-1,\frac{\alpha}{2})}$

$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$	$W_\alpha = \{\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}, \chi^2 < \chi^2_{(n-1,1-\alpha)}$
$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$	$W_\alpha = \{\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}, \chi^2 > \chi^2_{(n-1,\alpha)}$

1.2.3.3.2.TH2: μ biết.

Giả thuyết thống kê	W_α (miền bác bỏ H_0)
$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$	$W_\alpha = \{\chi^2 = \frac{\sum n_i(x_i-\mu)^2}{\sigma_0^2}, \chi^2 < \chi_1^2$ hoặc $\chi^2 > \chi_2^2$ $\chi_1^2 = \chi^2_{(n,1-\frac{\alpha}{2})}, \chi_2^2 = \chi^2_{(n,\frac{\alpha}{2})}$
$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$	$W_\alpha = \{\chi^2 = \frac{\sum n_i(x_i-\mu)^2}{\sigma_0^2}, \chi^2 < \chi^2_{(n,1-\alpha)}$
$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$	$W_\alpha = \{\chi^2 = \frac{\sum n_i(x_i-\mu)^2}{\sigma_0^2}, \chi^2 > \chi^2_{(n,\alpha)}$

1.2.4. So sánh 2 tham số của tổng thể:

1.2.4.1. So sánh 2 số trung bình:

1.2.4.1.1.TH1: $m \geq 30, n \geq 30, \sigma_1^2, \sigma_2^2$ biết

GTTK	W_α
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	$W_\alpha = \left\{u = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}; \lvert u\rvert > u_{\frac{\alpha}{2}}\right\}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$	$W_\alpha = \left\{u = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}; u < -u_\alpha\right\}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$	$W_\alpha = \left\{u = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}; u > u_\alpha\right\}$

1.2.4.1.2.TH2: $m < 30, n < 30, \sigma_1^2, \sigma_2^2$ biết, X,Y có phân phối chuẩn

GTTK	W_α

$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	$W_\alpha = \left\{ u = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}}; \lvert u \rvert > u_{\frac{\alpha}{2}} \right\}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$	$W_\alpha = \left\{ u = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}}; u < -u_\alpha \right\}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$	$W_\alpha = \left\{ u = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}}; u > u_\alpha \right\}$

1.2.4.1.3.TH3: $m \geq 30, n \geq 30, \sigma_1^2, \sigma_2^2$ không biết

GTTK	W_α
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	$W_\alpha = \left\{ u = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}}}; \lvert u \rvert > u_{\frac{\alpha}{2}} \right\}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$	$W_\alpha = \left\{ u = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}}}; u < -u_\alpha \right\}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$	$W_\alpha = \left\{ u = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}}}; u > u_\alpha \right\}$

1.2.4.1.4.TH4: $m < 30, n < 30$, X,Y có phân phối chuẩn, $\sigma_1^2 = \sigma_2^2$ không biết

GTTK	W_α
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	$W_\alpha = \left\{ t = \dfrac{\overline{X} - \overline{Y}}{\sqrt{s^2\left(\dfrac{1}{m} + \dfrac{1}{n}\right)}}; \lvert t \rvert > t_{\left(m+n-2, \frac{\alpha}{2}\right)} \right\}$ $s^2 = \dfrac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2}$

$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$	$W_\alpha = \left\{ t = \dfrac{\overline{X} - \overline{Y}}{\sqrt{s^2\left(\dfrac{1}{m} + \dfrac{1}{n}\right)}}; t < -t_{(m+n-2,\alpha)} \right\}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$	$W_\alpha = \left\{ t = \dfrac{\overline{X} - \overline{Y}}{\sqrt{s^2\left(\dfrac{1}{m} + \dfrac{1}{n}\right)}}; t > t_{(m+n-2,\alpha)} \right\}$

1.2.4.1.5.TH5: $m < 30, n < 30$, X,Y có phân phối chuẩn, $\sigma_1^2 \neq \sigma_2^2$ chưa biết

GTTK	W_α
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	$W_\alpha = \left\{ g = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}}}; \lvert g \rvert > t; t_1 = t_{\left(m-1, \frac{\alpha}{2}\right)}, t_2 = t_{\left(n-1, \frac{\alpha}{2}\right)}; v_1 = \dfrac{s_1^2}{m}, v_2 = \dfrac{s_2^2}{n}; t = \dfrac{t_1 v_1 + t_2 v_2}{v_1 + v_2} \right\}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$	$W_\alpha = \left\{ g = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}}}; g < -t; t_1 = t_{(m-1,\alpha)}, t_2 = t_{(n-1,\alpha)} \right\}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$	$W_\alpha = \left\{ g = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}}}; g > t \right\}$

1.2.4.2. So sánh 2 tỷ lệ:

GTTK	W_α
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	$W_\alpha = \left\{ u = \dfrac{f_1 - f_2}{\sqrt{f(1-f)\left(\dfrac{1}{m} + \dfrac{1}{n}\right)}}; \lvert u \rvert > u_{\frac{\alpha}{2}}; f_1 = \dfrac{k_1}{m}, f_2 = \dfrac{k_2}{n} \right\}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$	$W_\alpha = \left\{ u = \dfrac{f_1 - f_2}{\sqrt{f(1-f)\left(\dfrac{1}{m} + \dfrac{1}{n}\right)}}; u < -u_\alpha \right\}$

$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$	$W_\alpha = \left\{ u = \dfrac{f_1 - f_2}{\sqrt{f(1-f)\left(\dfrac{1}{m} + \dfrac{1}{n}\right)}}; u > u_\alpha \right\}$

1.2.4.3. So sánh 2 phương sai:

GTTK	W_α
$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$	$W_\alpha = \left\{ g = \dfrac{s_1^2}{s_2^2}, g < \overline{f} hay g > f; f = f_{\frac{\alpha}{2}}(m-1, n-1), \overline{f} = \dfrac{1}{f_{\frac{\alpha}{2}}(n-1, m-1)} \right\}$
$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 > \sigma_2^2$	$W_\alpha = \left\{ g = \dfrac{s_1^2}{s_2^2}, \lvert g \rvert > f_\alpha(m-1, n-1) \right\}$