TÓM TẮT MỘT SỐ CÔNG THỨC XÁC SUẤT THỐNG KÊ

- Công thức cộng và nhân XS: P(A+B) = P(A) +P(B) – P(AB). P(AB) = P(A).P(B/A).
- CT Bernouli: $P_n(k,p) = C_n^k\ p^k q^{n-k}$;
- CT XSTP: $P(F)=P(A_1).P(F/A_1) + P(A_2).P(F/A_2) + \ldots + P(A_n).P(F/A_n)$. $\{A_1, A_2,\ldots,A_n\}$ là nhóm bcđđ
- CT Bayes: $P(A_k/F)= \dfrac{P(A_k)\times P(F/A_k)}{P(F)}$;
- Nếu Bnn X rời rạc: $E(X)=\sum_i x_i \times p_i$; $D(X)=\sum_i (x_i - E(X))^2 \times p_i$
- Nếu Bnn X liên tục: $E(X) = \int_{-\infty}^{+\infty} x.f(x)dx$ => $D(X) = \int_{-\infty}^{+\infty} (x-a)^2 f(x)dx$
- PP nhị thức: $P(X = k) = C_n^k\ p^k q^{n-k}$. k = 0;1;…;n. Kí hiệu X~ B(n,p). $E(X) = np$; $D(X) = npq$
  - $Mod(X) = [np - q]\ hoặc\ [np - q] + 1$
  - X ~ N(np; npq) khi n lớn và 0<<p<<1 ; X ~ P(np) khi n lớn và p hoặc q rất nhỏ.
- PP Siêu bội: $P(X=k) = \dfrac{C_M^k \times C_{N-M}^{n-k}}{C_N^n}$; Kí hiệu X~ H(N,M,n). p=M/N. E(X) = np; D(X) = $npq\dfrac{N-n}{N-1}$
- PP POISSON: $P(X=k) = \dfrac{e^{-a}\times a^k}{k!}$; $k = 0;1;2;\ldots$ $X\sim P(a)$ thì $E(X) = D(X) = a$
- PP đều: $f(x) = \begin{cases} \frac{1}{b-a} & nếu\ x \in [a,b] \\ 0 & nếu\ x \notin [a,b] \end{cases}$ kí hiệu $X\sim U(a,b)$
  - $E(X) = \frac{b+a}{2}, D(X) = \frac{(b-a)^2}{12}$
- PP mũ: $Hàm\ mật\ độ\ XS$: $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0\ , & x < 0 \end{cases}$ kí hiệu $X\sim E(\lambda)$
  - $E(X) = \frac{1}{\lambda}, D(X) = \frac{1}{\lambda^2}$
- PP chuẩn: $Hàm\ mật\ độ\ XS$: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$ kí hiệu $X\sim N(a,\sigma^2)$
  - $E(X) = a, D(X) = \sigma^2$

Cho VTNN (X,Y) liên tục với hàm mật độ xác suất đồng thời: f(x, y)

- $f_X(x)= \int_{-\infty}^{+\infty} f(x,y)dy$; $f_Y(y)= \int_{-\infty}^{+\infty} f(x,y)dx$
- $f_{X/y} = \frac{f(x,y)}{f_Y(y)}$ nếu $Y = y$; $f_{Y/x}(y) = \frac{f(x,y)}{f_X(x)}$ nếu $X = x$
- VTNN (X, Y) phân phối đều trên D: $f(x,y) = \begin{cases} \frac{1}{m(D)} & nếu\ (x,y) \in D \\ 0 & nếu\ (x,y) \notin D \end{cases}$
- Covarian: *cov(X, Y) = E(XY) – E(X).E(Y);* $E(XY) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j p_{ij}$
- Ma trận tương quan: $D(X, Y) = \begin{pmatrix} \operatorname{cov}(X, X) & \operatorname{cov}(X, Y) \\ \operatorname{cov}(Y, X) & \operatorname{cov}(Y, Y) \end{pmatrix} = \begin{pmatrix} D(X) & \operatorname{cov}(X,Y) \\ \operatorname{cov}(X, Y) & D(Y) \end{pmatrix}$
- Nếu Z = φ(x,y)thì $E(Z) = \sum_{i=1}^{h}\sum_{j=1}^{k} \varphi(x_i, y_i)P_{ij}$ nếu (X,Y) rời rạc
  - $= \iint_{R^2} \varphi(x,y)f(x,y)dxdy$ nếu (X,Y) liên tục

- PT hồi quy tuyến tính mẫu :
  - $y = A + Bx;\ (B = \frac{\overline{xy} - \bar{x}.\bar{y}}{\hat{s}_x{}^2}, A = \bar{y} - B\bar{x})$
  - $x = A + By;\ (B = \frac{\overline{xy} - \bar{x}.\bar{y}}{\hat{s}_y{}^2}, A = \bar{x} - B\bar{y})$
  - *Hệ số tương quan mẫu :* $r_{XY} = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\hat{s}_X \hat{s}_Y}$
- Khoảng tin cậy cho tỷ lệ tổng thể P:
  - $f_n = \frac{m}{n}, P = \frac{M}{N}\ (n \geq 100);\ f_n - \varepsilon < P < f_n + \varepsilon, \varepsilon = U_\alpha \sqrt{\frac{f_n(1-f_n)}{n}};$
- Khoảng tin cậy cho kỳ vọng:
  - (biết σ^2, X có phân phối chuẩn hoặc n đủ lớn) $\bar{x} - \varepsilon < a < \bar{x} + \varepsilon;\ \varepsilon = U_\alpha.\frac{\sigma}{\sqrt{n}};$
  - $(n \geq 30,\text{không biết } \sigma^2)$ $\bar{x} - \varepsilon < a < \bar{x} + \varepsilon;\ \varepsilon = U_\alpha.\frac{S}{\sqrt{n}};$
  - $(n < 30, \text{X có phân phối chuẩn, không biết } \sigma^2)$ $\bar{x} - \varepsilon < a < \bar{x} + \varepsilon;\ \varepsilon = T_{\alpha/2}^{n-1}.\frac{S}{\sqrt{n}}$
- Khoảng tin cậy cho phương sai: $\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(\mathrm{n}-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(\mathrm{n}-1)}$
- KĐGT về tỉ lệ tổng thể:
  - 1 mẫu
    - $H_0: P = P_0, \bar{H}: P \neq P_0,$ $TCKĐ: U_{qs} = \frac{F_n - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$
    - $|U_{qs}| \leq U_\alpha$: Chấp nhận H;. Bác bỏ $\overline{\mathrm{H}}$. $|U_{qs}| > U_\alpha$: Bác bỏ H;. Chấp nhận $\overline{\mathrm{H}}$
  - 2 mẫu
    - $H_0: P_1 = P_2, \bar{H}: P_1 \neq P_2,.$
    - $U_{qs} = \frac{F_{n1} - F_{n2}}{\sqrt{\frac{\bar{P}(1-\bar{P})}{\bar{n}}}};\ \bar{p} = \frac{m1+m2}{n1+n2};\ \bar{n} = \frac{n1.n2}{n1+n2}$
    - $|U_{qs}| \leq U_\alpha$: Chấp nhận H;. Bác bỏ $\overline{\mathrm{H}}$. $|U_{qs}| > U_\alpha$: Bác bỏ H;. Chấp nhận $\overline{\mathrm{H}}$
- KĐGT về kì vọng tổng thể. Bài toán 1 mẫu: $H: a = a_0, \bar{H}: a \neq a_0$
  - biết σ^2, X có phân phối chuẩn hoặc n đủ lớn
    - $U_{qs} = \frac{\bar{X} - a_0}{\sigma}\sqrt{n} \sim N(0,1)$
    - $|U_{qs}| \leq U_\alpha$: Chấp nhận H;. Bác bỏ $\overline{\mathrm{H}}$. $|U_{qs}| > U_\alpha$: Bác bỏ H;. Chấp nhận $\overline{\mathrm{H}}$
  - $n \geq 30$, không biết σ^2
    - $U_{qs} = \frac{\bar{X} - a_0}{S}\sqrt{n} \sim N(0,1)$
    - $|U_{qs}| \leq U_\alpha$: Chấp nhận H;. Bác bỏ $\overline{\mathrm{H}}$. $|U_{qs}| > U_\alpha$: Bác bỏ H;. Chấp nhận $\overline{\mathrm{H}}$
  - $n < 30$, X có phân phối chuẩn, không biết σ^2
    - $T_{qs} = \frac{\bar{X} - a_0}{S}\sqrt{n} \sim Student(n-1)$
    - $|T_{qs}| \leq T_{\alpha/2}^{n-1}$: chấp nhận H; $|T_{qs}| > T_{\alpha/2}^{n-1}$: chấp nhận $\overline{\mathrm{H}}$;
- KĐGT về kì vọng tổng thể. Bài toán 2 mẫu, xét trường hợp n_1; n_2 >30; chưa biết σ_1^2; σ_2^2:
  - $H_0: a_1 = a_2, \bar{H}: a_1 \neq a_2$
  - $U_{qs} = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0,1)$
  - $|U_{qs}| \leq U_\alpha$: Chấp nhận H;. Bác bỏ $\overline{\mathrm{H}}$. $|U_{qs}| > U_\alpha$: Bác bỏ H;. Chấp nhận $\overline{\mathrm{H}}$

- KĐGT về phương sai tổng thể. Bài toán 1 mẫu: H: $\sigma^2 = \sigma^2_0$; $\overline{H}$: $\sigma^2 \neq \sigma^2_0$

$$\chi^2_{qs} = \frac{(n-1)S^2}{\sigma_0^2}$$

Chấp nhận H khi $\chi^2_{qs} \in \left[\chi^2_{1-\frac{\alpha}{2}}(n-1);\ \chi^2_{\frac{\alpha}{2}}(n-1)\right]$ .Chấp nhận $\overline{H}$ khi $\chi^2_{qs} \notin \left[\chi^2_{1-\frac{\alpha}{2}}(n-1);\ \chi^2_{\frac{\alpha}{2}}(n-1)\right]$

- Mở rộng: Bài toán kiểm định 1 phía:

Khi $U_{qs} \sim N(0, 1)$:

Nếu gt đối H_1 là kiểm định 2 phía thì miền bác bỏ gtkđ H là $W = (-\infty; -z_\alpha) \cup (z_\alpha; +\infty)$

Nếu gt đối H_1 là kiểm định bên trái thì miền bác bỏ gtkđ H là $W = (-\infty; -z_{2\alpha})$

Nếu gt đối H_1 là kiểm định bên phải thì miền bác bỏ gtkđ H là $W = (z_{2\alpha}; +\infty)$

với $\Phi(z_\alpha) = (1-\alpha)/2$; $\Phi(z_{2\alpha}) = (1-2\alpha)/2$

Khi $U_{qs} \sim T(n-1)$:

Nếu gt đối H_1 là kđ 2 phía thì miền bác bỏ gtkđ H là $W = (-\infty; -t_{\alpha/2}(n-1)) \cup (t_{\alpha/2}(n-1); +\infty)$

Nếu gt đối H_1 là kiểm định bên trái thì miền bác bỏ gtkđ H là $W = (-\infty; -t_\alpha(n-1))$

Nếu gt đối H_1 là kiểm định bên phải thì miền bác bỏ gtkđ H là $W = (t_\alpha(n-1); +\infty)$

KIỂM ĐỊNH QUY LUẬT PHÂN PHỐI

Kiểm định giả thiết $H : X$ có phân phối $F(x)$ với mức ý nghĩa α.

- Tìm $\mathbf{X}^2_\alpha(k-r-1)$ từ bảng phân phối $\mathbf{X}^2$, ở đây r là số tham số của $F(x)$

- Tính thống kê $\mathbf{X}^2_o = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$

Nếu $\mathbf{X}^2_o \le \mathbf{X}^2_\alpha$ thì chấp nhận *H*.

Nếu $\mathbf{X}^2_o > \mathbf{X}^2_\alpha$ thì bác bỏ *H*.

KIỂM ĐỊNH TÍNH ĐỘC LẬP

Hãy kiểm định giả thiết *H*: *X* và *Y* độc lập với mức ý nghĩa α.

- Tìm $\mathbf{X}^2_\alpha = \mathbf{X}^2_\alpha[(k-1)(h-1)]$ từ bảng phân vị $\mathbf{X}^2$

- Tính thống kê

$$\mathbf{X}^2_o = \sum_{i=1}^{k}\sum_{j=1}^{h} \frac{(n_{ij} - \gamma_{ij})^2}{\gamma_{ij}}, \quad \gamma_{ij} = \frac{n_i m_j}{n}$$

Nếu $\mathbf{X}^2_o \le \mathbf{X}^2_\alpha$ thì chấp nhận *H*.

Nếu $\mathbf{X}^2_o > \mathbf{X}^2_\alpha$ thì bác bỏ *H*.