

QUELQUES NOTIONS DE PROBABILITE

$$P(AB|C) = P(A|BC) P(B|C) = P(B|AC) P(A|C) ; P(A + B|C) = P(A|C) + P(B|C) - P(AB|C)$$

$$\text{Si } \dots, P(A) = P(A|B_1)P(B_1) + \dots + P(A|B_n)P(B_n) ; P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$$

$$\text{Si } P(A) > 0, \text{ alors } P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

$$\text{Si } Y_1, \dots, Y_n \text{ sont des v.a. indépendantes, alors } V(Y_1 + \dots + Y_n) = V(Y_1) + \dots + V(Y_n)$$

$$\text{Si } Z = a + bY: E(Z) = a + b\mu_Y ; \sigma_Z^2 = b^2\sigma_Y^2$$

Variabes aléatoires discrètes:

$$\mu_Y = E(Y) = \sum_{y_i \in \mathcal{E}} p_i y_i ; \sigma_Y^2 = V(Y) = E(Y - \mu)^2 = \sum_{y_i \in \mathcal{E}} p_i (y_i - \mu)^2 = E(Y^2) - \mu^2$$

$$\text{Bernoulli: } \mathcal{E}_i = \{0, 1\} \text{ et } p_0 = 1 - \pi ; p_1 = \pi ; E(X) = \pi ; V(X) = \pi(1 - \pi)$$

$$\text{Binomiale: } X \sim \text{Bin}(n, \pi) ; p_x = \frac{n!}{x!(n-x)!} \pi^x (1 - \pi)^{n-x} ; E(X) = n\pi ; V(X) = n\pi(1 - \pi)$$

$$\text{Poisson: } X \sim \text{Pois}(\mu) ; p_x = \Pr(X = k) = e^{-\mu} \frac{\mu^k}{k!} ; p_k = p_{k-1} \times \frac{\mu}{k} ; E(X) = V(X) = \mu$$

$$\text{Si } n > 50 \text{ et } \pi < 0.10, \text{ Bin}(n, \pi) \approx \text{Pois}(n\pi)$$

Variabes aléatoires continues:

$$\mu_X = E(X) = \int_{\mathcal{E}} x f(x) dx ; \sigma_X^2 = V(X) = E(X - \mu)^2 = \int_{\mathcal{E}} (x - \mu)^2 f(x) dx = E(X^2) - \mu_X^2$$

$$\text{Distribution normale: } X \sim N(\mu, \sigma^2) ; E(X) = \mu ; V(X) = \sigma^2 ; Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

$$\text{Si } \mu \geq 5, \text{ Pois}(\mu) \approx N(\mu, \mu) ; \text{ Si } n\pi \geq 5 \text{ et } n\pi(1 - \pi) \geq 5, \text{ Bin}(n, \pi) \approx N(n\pi, n\pi(1 - \pi))$$

ESTIMATION ET COMPARAISON DE PROPORTIONS - McNemar - G^2

- Objectif: estimer $\pi = \sum_{i=1}^N X_i / N$.
- $\pi \sim \text{Beta}(a, b): E(\pi) = \frac{a}{a+b}; \text{Mode}(\pi) = \frac{a-1}{a+b-2}; V(\pi) = \frac{ab}{(a+b)^2(a+b+1)}$; Uniforme si $a = b = 1$.
- Si $y =$ nombre de succès, a priori $\pi \sim \text{Beta}(a, b)$, $n^* = (a + b - 2) + n$ et $\hat{\pi} = \frac{(a-1)+y}{n^*}$, alors $(\pi|Y = y) \sim \text{Beta}(a + y, b + n - y) \approx \mathcal{N}(\hat{\pi}, \hat{\pi}(1 - \hat{\pi})/n^*)$; $\hat{\pi} \pm 1.96\sqrt{\hat{\pi}(1 - \hat{\pi})/n^*}$
- Si $\delta = \pi_2 - \pi_1$, $(\delta = \pi_2 - \pi_1|Y_1 = y_1, Y_2 = y_2) \approx \mathcal{N}(\hat{\pi}_2 - \hat{\pi}_1, \hat{\pi}_2(1 - \hat{\pi}_2)/n_2 + \hat{\pi}_1(1 - \hat{\pi}_1)/n_1)$.

- Données paires: $(\pi_2 - \pi_1 | \text{données}) \stackrel{\text{approx.}}{\sim} \mathcal{N}\left(\frac{n_{21} - n_{12}}{n}, \frac{s_D^2}{n}\right)$; $\frac{n_{21} - n_{12}}{n} \pm 1.96 \sqrt{\frac{s_D^2}{n}}$
avec $s_D^2 = \frac{n_{21} + n_{12}}{n} - \frac{(n_{21} - n_{12})^2}{n^2}$. Condition d'utilisation: $n_{21} + n_{12} \geq 25$.
- McNemar: $\frac{(n_{21} - n_{12})^2}{n_{21} + n_{12}} < \chi_1^2(0.95) \dots$ si $n_{21} + n_{12} \geq 25$.
- $G^2 = 2 \sum_{i=1}^I \sum_{j=1}^J O_{ij} \log\left(\frac{O_{ij}}{A_{ij}}\right)$ avec O_{ij} et A_{ij} valeurs observée et attendue respvt.
- $\log(RR): \log\left(\frac{p_1}{p_2}\right) \pm 1.96 \times \sqrt{\frac{1}{n_{11}} - \frac{1}{n_{1+}} + \frac{1}{n_{21}} - \frac{1}{n_{2+}}}$

ESTIMATION ET COMPARAISON DE MOYENNES

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i ; \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \left(\frac{1}{n} \sum_{i=1}^n y_i^2\right) - \bar{y}^2 ; s^2 = \frac{n}{n-1} \hat{\sigma}^2$$

$$\bar{y} = \frac{1}{n} \sum_{k=1}^K n_k y_k = \sum_{k=1}^K w_k y_k ; \hat{\sigma}^2 = \sum_{k=1}^K w_k (y_k - \bar{y})^2 = \left(\sum_{k=1}^K w_k y_k^2\right) - \bar{y}^2 \text{ avec } w_k = n_k/n$$

- Moyenne isolée: $(\mu | \text{données}) \sim t_{n-1}(\bar{y}, s^2/n)$; $\bar{y} \pm t_{n-1}(0.975) \frac{s}{\sqrt{n}}$; $(\mu | \text{données}) \stackrel{n \rightarrow \infty}{\sim} \mathcal{N}\left(\bar{y}, \frac{s^2}{n}\right)$

- Comparaison de 2 moyennes: si $\theta = \mu_1 - \mu_2$,

$$\rightarrow \text{Approx: } (\theta | \text{données}) \stackrel{\text{approx.}}{\sim} \mathcal{N}\left(\bar{y} - \bar{x}, s_1^2/n_1 + s_2^2/n_2\right); (\bar{y} - \bar{x}) \pm 1.96 \sqrt{s_1^2/n_1 + s_2^2/n_2}$$

$$\rightarrow \text{Si } \sigma_1^2 = \sigma_2^2: (\theta | \text{données}) \sim t_{n_1+n_2-2}\left(\bar{y} - \bar{x}, s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right) \text{ avec } s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2};$$

$$(\bar{y} - \bar{x}) \pm t_{n_1+n_2-2}(0.975) \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) s_p^2}$$

$$\rightarrow \text{Si } \sigma_1^2 \neq \sigma_2^2: (\theta | \text{données}) \stackrel{\text{approx.}}{\sim} t_{df}(\bar{y} - \bar{x}, s_1^2/n_1 + s_2^2/n_2) \text{ avec}$$

$$df = (s_1^2/n_1 + s_2^2/n_2)^2 / \left\{ \frac{1}{n_1-1} (s_1^2/n_1)^2 + \frac{1}{n_2-1} (s_2^2/n_2)^2 \right\} \quad (\text{Satterthwaite}).$$

- Données paires: $(\mu_D | \text{données}) \sim t_{n_D-1}(\bar{d}, s_D^2/n_D)$; $\bar{d} \pm t_{n_D-1}(0.975) s_D/\sqrt{n_D}$

$$\bullet \text{ ANOVA I: } F_{obs} = \frac{\sum_{k=1}^K n_k (\bar{y}_k - \bar{y})^2 / (K-1)}{s_p^2} = \frac{\text{MC(Facteur)}}{\text{MC(Erreur)}} < F_{K-1, n-K}(0.95) \text{ avec } s_p^2 = \frac{(n_1-1)s_1^2 + \dots + (n_K-1)s_K^2}{n-K}$$

$$\text{avec } (\mu_k - \mu_j | \text{données}) \sim t_{n-K}\left(\bar{y}_k - \bar{y}_j, \left(\frac{1}{n_k} + \frac{1}{n_j}\right) s_p^2\right); (\bar{y}_k - \bar{y}_j) \pm t_{n-K}(0.975) \sqrt{\left(\frac{1}{n_k} + \frac{1}{n_j}\right) s_p^2}$$