QUELQUES NOTIONS DE PROBABILITE

$\mathrm{P}(AB|C) = \mathrm{P}(A|BC)\ \mathrm{P}(B|C) = \mathrm{P}(B|AC)\ \mathrm{P}(A|C)$; $\mathrm{P}(A+B|C) = \mathrm{P}(A|C) + \mathrm{P}(B|C) - \mathrm{P}(AB|C)$
Si ..., $\mathrm{P}(A) = \mathrm{P}(A|B_1)\mathrm{P}(B_1) + \ldots + \mathrm{P}(A|B_n)\mathrm{P}(B_n)$; $\mathrm{P}(A) = \mathrm{P}(A|B)\mathrm{P}(B) + \mathrm{P}(A|\bar{B})\mathrm{P}(\bar{B})$
Si $\mathrm{P}(A) > 0$, alors $\mathrm{P}(B|A) = \frac{\mathrm{P}(A|B)\mathrm{P}(B)}{\mathrm{P}(A)}$
Si $Y_1, \ldots, Y_n$ sont des v.a. indépendantes, alors $\mathrm{V}(Y_1 + \ldots + Y_n) = \mathrm{V}(Y_1) + \ldots + \mathrm{V}(Y_n)$
Si $Z = a + bY$: $\mathrm{E}(Z) = a + b\mu_Y$; $\sigma_Z^2 = b^2\sigma_Y^2$

Variables aléatoires discrètes:
$\mu_Y = \mathrm{E}(Y) = \sum_{y_i \in \mathcal{E}} p_i y_i$; $\sigma_Y^2 = \mathrm{V}(Y) = \mathrm{E}(Y-\mu)^2 = \sum_{y_i \in \mathcal{E}} p_i (y_i - \mu)^2 = \mathrm{E}(Y^2) - \mu^2$
Bernoulli: $\mathcal{E}_i = \{0, 1\}$ et $p_0 = 1 - \pi$; $p_1 = \pi$; $\mathrm{E}(X) = \pi$; $\mathrm{V}(X) = \pi(1-\pi)$
Binomiale: $X \sim \mathrm{Bin}(n, \pi)$; $p_x = \frac{n!}{x!\ (n-x)!}\ \pi^x\ (1-\pi)^{n-x}$; $E(X) = n\pi$; $\mathrm{V}(X) = n\pi(1-\pi)$
Poisson: $X \sim \mathrm{Pois}(\mu)$; $p_x = \Pr(X = k) = \mathrm{e}^{-\mu}\ \frac{\mu^k}{k!}$; $p_k = p_{k-1} \times \frac{\mu}{k}$; $E(X) = \mathrm{V}(X) = \mu$
Si $n > 50$ et $\pi < 0.10$, $\mathrm{Bin}(n, \pi) \approx \mathrm{Pois}(n\pi)$

Variables aléatoires continues:
$\mu_X = \mathrm{E}(X) = \int_{\mathcal{E}} x\ f(x)\ dx$; $\sigma_X^2 = \mathrm{V}(X) = \mathrm{E}(X-\mu)^2 = \int_{\mathcal{E}} (x - \mu_X)^2\ f(x)\ dx = \mathrm{E}(X^2) - \mu_X^2$
Distribution normale: $X \sim N(\mu, \sigma^2)$; $\mathrm{E}(X) = \mu$; $\mathrm{V}(X) = \sigma^2$; $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$
Si $\mu \geq 5$, $\mathrm{Pois}(\mu) \approx N(\mu, \mu)$; Si $n\pi \geq 5$ et $n\pi(1-\pi) \geq 5$, $\mathrm{Bin}(n, \pi) \approx N(n\pi, n\pi(1-\pi))$

ESTIMATION ET COMPARAISON DE PROPORTIONS - McNemar - G^2

- Objectif: estimer $\pi = \sum_{i=1}^N X_i/N$.
- $\pi \sim \mathrm{Beta}(a, b)$: $\mathrm{E}(\pi) = \frac{a}{a+b}$; $\mathrm{Mode}(\pi) = \frac{a-1}{a+b-2}$; $\mathrm{V}(\pi) = \frac{ab}{(a+b)^2\ (a+b+1)}$; Uniforme si $a = b = 1$.
- Si y =nbre de succès, a priori $\pi \sim \mathrm{Beta}(a, b)$, $n^* = (a + b - 2) + n$ et $\hat{\pi} = \frac{(a-1)+y}{n^*}$, alors $(\pi|Y = y) \sim \mathrm{Beta}(a + y, b + n - y) \approx \mathcal{N}(\hat{\pi}, \hat{\pi}(1-\hat{\pi})/n^*)$; $\hat{\pi} \pm 1.96\sqrt{\hat{\pi}(1-\hat{\pi})/n^*}$
- 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 pairées: $(\pi_2 - \pi_1|\text{données}) \overset{\text{approx.}}{\sim} \mathcal{N}(\frac{n_{21}-n_{12}}{n}, \frac{s_D^2}{n})$; $\frac{n_{21}-n_{12}}{n} \pm 1.96\sqrt{\frac{s_D^2}{n}}$
  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: $\frac{(n_{21}-n_{12})^2}{n_{21}+n_{12}} < \chi_1^2(0.95)$ ... 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

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

- Moyenne isolée: $(\mu|\text{donnees}) \sim t_{n-1}(\bar{y}, s^2/n)$; $\bar{y} \pm t_{n-1}(0.975)\frac{s}{\sqrt{n}}$; $(\mu|\text{donnees}) \overset{n\to\infty}{\sim} \mathcal{N}\left(\bar{y}, \frac{s^2}{n}\right)$
- Comparaison de 2 moyennes: si $\theta = \mu_1 - \mu_2$,

$\rightarrow$ Approx: $(\theta|\text{donnees}) \overset{\text{approx.}}{\sim} \mathcal{N}(\bar{y} - \bar{x}, s_1^2/n_1 + s_2^2/n_2)$; $(\bar{y} - \bar{x}) \pm 1.96\sqrt{s_1^2/n_1 + s_2^2/n_2}$
$\rightarrow$ Si $\sigma_1^2 = \sigma_2^2$: $(\theta|\text{donnees}) \sim t_{n_1+n_2-2}\left(\bar{y} - \bar{x}, s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$ avec $s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$;

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

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

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

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

$\rightarrow$ $F_{obs} = \frac{\sum_{k=1}^K n_k(\bar{y}_k - \bar{y})^2/(K-1)}{s_p^2} = \frac{\text{MC(Facteur)}}{\text{MC(Erreur)}} < F_{K-1,n-K}(0.95)$ avec $s_p^2 = \frac{(n_1-1)s_1^2 + \ldots + (n_K-1)s_K^2}{n-K}$
$\rightarrow$ $(\mu_k - \mu_j|\text{donnees}) \sim t_{n-K}\left(\bar{y}_k - \bar{y}_j, \left(\frac{1}{n_k} + \frac{1}{n_j}\right) s_p^2\right)$; $(\bar{y}_k - \bar{y}_j) \pm t_{n-K}(0.975)\sqrt{\left(\frac{1}{n_k} + \frac{1}{n_j}\right) s_p^2}$