Tests

M8 – Chapitre 4

Les tests permettent de mesurer si les variations entre données ont une probabilité raisonnable d’être dus au hasard ou non.

	2 qualitatives : test du χ^2	1 qualitative, 1 quantitative : test de Student	2 quantitatives : test de Student
Données	$\begin{array}{c\|c\|c\|c\|c} X \backslash Y & b_1 & \dots & b_J & \text{marge} \\ \hline a_1 & n_{11} & \dots & n_{1J} & n_{1\bullet} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_I & n_{I1} & \dots & n_{IJ} & n_{I\bullet} \\ \hline \text{marge} & n_{\bullet 1} & \dots & n_{\bullet J} & n_{\bullet\bullet} = n \end{array}$	$\begin{array}{c\|c} Y & X \\ \hline a & x_{a_1} \\ \vdots & \vdots \\ a & x_{a_{n_a}} \\ \hline b & = x_{b_1} \\ \vdots & \vdots \\ b & x_{b_{n_b}} \end{array}$	$\begin{array}{c\|c} Y & X \\ \hline y_1 & x_1 \\ \vdots & \vdots \\ y_n & x_n \end{array}$ $y_i = ax_i + b + \varepsilon_i$
Hypothèses	$\begin{cases} \mathcal{H}_0 : \text{référence} \Leftrightarrow \text{indépendance} \\ \mathcal{H}_1 : \text{alternative} \Leftrightarrow \text{dépendance} \end{cases}$	$\begin{cases} \mathcal{H}_0 : \mu_a = \mu_b & X \text{ et } Y \text{ indépendants} \\ \mathcal{H}_{1_1} : \mu_a < \mu_b & \\ \mathcal{H}_{1_2} : \mu_a > \mu_b & X \text{ et } Y \text{ dépendants} \\ \mathcal{H}_{1_3} : \mu_a \neq \mu_b & \end{cases}$	$\begin{cases} \mathcal{H}_0 : \text{indépendance} \Leftrightarrow a = 0 \\ \mathcal{H}_1 : \text{dépendance} \Leftrightarrow a \neq 0 \end{cases}$
Modèle	$\mathbb{P}\left((X = x_i) \cap (Y = y_j)\right)$ $= \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j)$	$\overline{X_a} \sim \mathcal{N}\left(\mu_a, \frac{\sigma^2}{n_a}\right)$ $\overline{X_b} \sim \mathcal{N}\left(\mu_b, \frac{\sigma^2}{n_b}\right)$ $\sigma = \sigma_a = \sigma_b$ même variance	$\varepsilon_i \sim \mathcal{N}(0, \sigma^2) \Rightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 \sim \chi^2_{n-2}$
Statistique	$D_{\chi^2} = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{\left(n_{ij} - \frac{n_{i\bullet}n_{\bullet j}}{n}\right)^2}{\frac{n_{i\bullet}n_{\bullet j}}{n}} \sim \chi^2_{(I-1)(J-1)}$ $D_{\chi^2} = n\sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(\overbrace{\hat{\mathbb{P}}_{ij}}^{\text{eff obs}} - \overbrace{\hat{\mathbb{P}}_i\hat{\mathbb{P}}_j}^{\text{eff th}})^2}{\hat{\mathbb{P}}_i\hat{\mathbb{P}}_j}$	$U = \frac{\overline{X_a} - \overline{X_b}}{\sqrt{\sigma^2\left(\frac{1}{n_a} + \frac{1}{n_b}\right)}} \sim \mathcal{N}(0,1)$ Variance connue σ^2 $T = \frac{\overline{X_a} - \overline{X_b}}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_a} + \frac{1}{n_b}\right)}} \sim \mathcal{T}_{n_a+n_b-2}$ Variance inconnue estimée $\hat{\sigma}^2$ $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n_a}(x_{a_i} - \overline{x_a})^2 + \sum_{i=1}^{n_b}(x_{b_i} - \overline{x_b})^2}{n_a + n_b - 2}$	$U = \frac{\hat{a} - a}{\sqrt{\frac{\sigma^2}{s_X^2}}} \sim \mathcal{N}(0,1)$ Variance connue σ^2 $T = \frac{\hat{a} - a}{\sqrt{\frac{\hat{\sigma}^2}{s_X^2}}} \sim \mathcal{T}_{n-2}$ Variance estimée $\hat{\sigma}^2$ $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{n-2} = \frac{\sum_{i=1}^{n}\left(y_i - (\hat{a}x_i + \hat{b})\right)^2}{n-2}$
Valeur	Calcul de $D_{\chi^2_{obs}}$, u, ou t ; valeur de D_{χ^2}, U, ou T à partir des données.		
p-valeur	$p\text{-}val = \mathbb{P}\left(D_{\chi^2_n} \geq D_{\chi^2_{obs}}\right)$	$p\text{-}val = \begin{cases} \mathbb{P}(U \leq u) & \text{si } \mathcal{H}_1 = \mathcal{H}_{1_1} \\ \mathbb{P}(U \geq u) & \text{si } \mathcal{H}_1 = \mathcal{H}_{1_2} \\ \mathbb{P}(U \leq -\|u\|) + \mathbb{P}(U \geq \|u\|) & \text{si } \mathcal{H}_1 = \mathcal{H}_{1_3} \end{cases}$	$p\text{-}val = \mathbb{P}(\|U\| \geq u) = \mathbb{P}(U \leq -\|u\|) + \mathbb{P}(U \geq \|u\|)$
Décision	**p-valeur :** probabilité d’obtenir un tableau encore plus « rare » au hasard. Se lit généralement dans la table de la loi utilisée. • $p\text{-}val < 5\%$: on garde $\mathcal{H}_1$ *(« peu de tableaux sont plus rares, ce n’est pas du hasard »)* • $p\text{-}val \geq 5\%$: on garde $\mathcal{H}_0$ *(« le tableau n’est pas si rare, ça peut être le hasard »)*		

Tests

M8 – Chapitre 4

I. Rappels sur les lois et les probabilités

1. Espérance et variance

$$\mathbb{E}(aX + Y + b) = a\mathbb{E}(X) + \mathbb{E}(Y) + b$$

$$V(aX + b) = a^2 V(X)$$

2. Loi normale

$$X \sim \mathcal{N}(0,1)$$

$$\mathbb{P}(X \leq x) = 1 - \mathbb{P}(X > x) \qquad \mathbb{P}(x \leq X \leq y) = \mathbb{P}(X \leq y) - \mathbb{P}(X \leq x)$$

II. Loi du χ^2 à n degré de liberté

1. Définition et propriétés

$$\boxed{Z_n = \sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2}} \qquad \begin{aligned} &Z_n \sim \chi_n^2 \\ &Y_i \sim \mathcal{N}(0,1) \\ &X_i \sim \mathcal{N}(\mu, \sigma^2) \end{aligned} \qquad \mathbb{E}(Z_n) = n \quad V(Z_n) = 2n \qquad \frac{Z_n - n}{\sqrt{2n}} \to \mathcal{N}(0,1)$$

$$Z_{n-1} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2$$

2. Théorème de Pearson

$$X_i = \frac{N_i - n\widehat{p_i}}{\sqrt{n\widehat{p_i}}} \qquad \sum_{i=1}^{I} X_i^2 \to \chi_{I-1}^2$$

$$X_{ij} = \frac{N_{ij} - n\widehat{p_{ij}}}{\sqrt{n\widehat{p_{ij}}}} \qquad \sum_{i=1}^{I}\sum_{j=1}^{J} X_{ij}^2 \to \chi_{(I-1)(J-1)}^2$$

III. Loi de Student à n degré de liberté

1. Définition

$$\boxed{T_n = \frac{N}{\sqrt{\frac{X_n}{n}}}} \qquad \begin{aligned} &T_n \sim \mathcal{T}_n \\ &N \sim \mathcal{N}(0,1) \\ &X_n \sim \chi_n^2 \end{aligned} \qquad T_n \to \mathcal{N}(0,1)$$

2. Construction du test de Student

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \quad Y = \bar{X} - \frac{\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim \mathcal{N}(0,1) \quad \text{et} \quad Z_{n-1} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2$$

$$T_{n-1} = \frac{Y}{\sqrt{\frac{Z_{n-1}}{n-1}}} = \frac{\bar{X} - \mu}{\frac{S_{n-1}}{\sqrt{n}}} \sim \mathcal{T}_{n-1}$$