

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	<table><tr><th>X \ Y</th><th>b_1</th><th>...</th><th>b_J</th><th>marge</th></tr><tr><th>a_1</th><td>n_{11}</td><td>...</td><td>n_{1J}</td><td>$n_{1\bullet}$</td></tr><tr><th>$\vdots$</th><td>$\vdots$</td><td>$\ddots$</td><td>$\vdots$</td><td>$\vdots$</td></tr><tr><th>a_I</th><td>n_{I1}</td><td>...</td><td>n_{IJ}</td><td>$n_{I\bullet}$</td></tr><tr><th>marge</th><td>$n_{\bullet 1}$</td><td>...</td><td>$n_{\bullet J}$</td><td>$n_{\bullet\bullet} = n$</td></tr></table>	X \ Y	b_1	...	b_J	marge	a_1	n_{11}	...	n_{1J}	$n_{1\bullet}$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	a_I	n_{I1}	...	n_{IJ}	$n_{I\bullet}$	marge	$n_{\bullet 1}$	...	$n_{\bullet J}$	$n_{\bullet\bullet} = n$	<table><tr><th>Y</th><th>X</th></tr><tr><td>a</td><td>x_{a_1}</td></tr><tr><td>$\vdots$</td><td>$\vdots$</td></tr><tr><td>a</td><td>$x_{a_{n_a}}$</td></tr><tr><td>b</td><td>$= x_{b_1}$</td></tr><tr><td>$\vdots$</td><td>$\vdots$</td></tr><tr><td>b</td><td>$x_{b_{n_b}}$</td></tr></table>	Y	X	a	x_{a_1}	$\vdots$	$\vdots$	a	$x_{a_{n_a}}$	b	$= x_{b_1}$	$\vdots$	$\vdots$	b	$x_{b_{n_b}}$	<table><tr><th>Y</th><th>X</th></tr><tr><td>y_1</td><td>x_1</td></tr><tr><td>$\vdots$</td><td>$\vdots$</td></tr><tr><td>y_n</td><td>x_n</td></tr></table> $y_i = ax_i + b + \varepsilon_i$	Y	X	y_1	x_1	$\vdots$	$\vdots$	y_n	x_n
X \ Y	b_1	...	b_J	marge																																														
a_1	n_{11}	...	n_{1J}	$n_{1\bullet}$																																														
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$																																														
a_I	n_{I1}	...	n_{IJ}	$n_{I\bullet}$																																														
marge	$n_{\bullet 1}$	...	$n_{\bullet J}$	$n_{\bullet\bullet} = n$																																														
Y	X																																																	
a	x_{a_1}																																																	
$\vdots$	$\vdots$																																																	
a	$x_{a_{n_a}}$																																																	
b	$= x_{b_1}$																																																	
$\vdots$	$\vdots$																																																	
b	$x_{b_{n_b}}$																																																	
Y	X																																																	
y_1	x_1																																																	
$\vdots$	$\vdots$																																																	
y_n	x_n																																																	
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	$\mathbb{P}\left((X = x_i) \cap (Y = y_j)\right)$ $= \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j)$	$\overline{X}_a \sim \mathcal{N}\left(\mu_a, \frac{\sigma^2}{n_a}\right) \quad \overline{X}_b \sim \mathcal{N}\left(\mu_b, \frac{\sigma^2}{n_b}\right) \quad \sigma = \sigma_a = \sigma_b$ même variance	$\varepsilon_i \sim \mathcal{N}(0, \sigma^2) \Rightarrow \frac{1}{\sigma^2} \sum_{i=1}^n \hat{\varepsilon}_i^2 \sim \chi_{n-2}^2$																																															
Statistique	<div>$D_{\chi^2} = \sum_{i=1}^I \sum_{j=1}^J \frac{\left(n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n}\right)^2}{\frac{n_{i\bullet} n_{\bullet j}}{n}} \sim \chi_{(I-1)(J-1)}^2$<p style="text-align: center;">eff obs eff th</p>$D_{\chi^2} = n \sum_{i=1}^I \sum_{j=1}^J \frac{(\widehat{\mathbb{P}}_{ij} - \widehat{\mathbb{P}}_i \widehat{\mathbb{P}}_j)^2}{\widehat{\mathbb{P}}_i \widehat{\mathbb{P}}_j}$</div>	<div>$U = \frac{\overline{X}_a - \overline{X}_b}{\sqrt{\sigma^2 \left(\frac{1}{n_a} + \frac{1}{n_b}\right)}} \sim \mathcal{N}(0,1)$<p style="text-align: center;">Variance connue σ^2</p></div> <div>$T = \frac{\overline{X}_a - \overline{X}_b}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n_a} + \frac{1}{n_b}\right)}} \sim \mathcal{T}_{n_a+n_b-2}$<p style="text-align: center;">Variance inconnue estimée $\hat{\sigma}^2$</p>$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n_a} (x_{a_i} - \overline{x}_a)^2 + \sum_{i=1}^{n_b} (x_{b_i} - \overline{x}_b)^2}{n_a + n_b - 2}$</div>	<div>$U = \frac{\hat{a} - a}{\sqrt{\frac{\sigma^2}{S_X^2}}} \sim \mathcal{N}(0,1)$<p style="text-align: center;">Variance connue σ^2</p></div> <div>$T = \frac{\hat{a} - a}{\sqrt{\frac{\hat{\sigma}^2}{S_X^2}}} \sim \mathcal{T}_{n-2}$<p style="text-align: center;">Variance estimée $\hat{\sigma}^2$</p>$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{n-2} = \frac{\sum_{i=1}^n \left(y_i - (\hat{a}x_i + \hat{b})\right)^2}{n-2}$</div>																																															
Valeur	Calcul de $D_{\chi^2_{obs}}$, u , ou t ; valeur de D_{χ^2} , U , ou T à partir des données.																																																	
p-valeur	$p\text{-val} = \mathbb{P}\left(D_{\chi_n^2} \geq D_{\chi^2_{obs}}\right)$	$p\text{-val} = \begin{cases} \mathbb{P}(U \leq u) & \text{si } \mathcal{H}_1 = \mathcal{H}_{1_1} \\ \mathbb{P}(U \geq u) & \text{si } \mathcal{H}_1 = \mathcal{H}_{1_2} \\ \mathbb{P}(U \leq - u) + \mathbb{P}(U \geq u) & \text{si } \mathcal{H}_1 = \mathcal{H}_{1_3} \end{cases}$	$p\text{-val} = \mathbb{P}(U \geq u) = \mathbb{P}(U \leq - u) + \mathbb{P}(U \geq u)$																																															
Décision	p-valeur : probabilité d’obtenir un tableau encore plus « rare » au hasard. Se lit généralement dans la table de la loi utilisée. <ul style="list-style-type: none">$p\text{-val} < 5\%$: 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 »)																																																	

I. Rappels sur les lois et les probabilités

1. Espérance et variance

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

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

2. Loi normale

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

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

II. Loi du χ^2 à n degré de liberté

1. Définition et propriétés

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

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

2. Théorème de Pearson

$$\begin{array}{l} X_i = \frac{N_i - n\hat{p}_i}{\sqrt{n\hat{p}_i}} \\ X_{ij} = \frac{N_{ij} - n\hat{p}_{ij}}{\sqrt{n\hat{p}_{ij}}} \end{array} \quad \begin{array}{l} \sum_{i=1}^I X_i^2 \rightarrow \chi_{I-1}^2 \\ \sum_{i=1}^I \sum_{j=1}^J X_{ij}^2 \rightarrow \chi_{(I-1)(J-1)}^2 \end{array}$$

III. Loi de Student à n degré de liberté

1. Définition

$$\boxed{T_n = \frac{N}{\sqrt{\frac{X_n}{n}}}} \quad \begin{array}{l} T_n \sim \mathcal{T}_n \\ N \sim \mathcal{N}(0,1) \\ X_n \sim \chi_n^2 \end{array} \quad T_n \rightarrow \mathcal{N}(0,1)$$

2. Construction du test de Student

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

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