Exercice 6 — Longueurs des mâch. inférieures.

(mâles) Groupe 1 $X_{1,1}, \dots, X_{n_1,1} \overset{iid}{\sim} \mathcal{N}(\mu_1, \sigma^2)$

(femelles) Groupe 2 $X_{1,2}, \dots, X_{n_2,1} \overset{iid}{\sim} \mathcal{N}(\mu_2, \sigma^2)$

1) • Les médianes ont l'air différentes entre les 2 groupes, la médiane pour les femelles est inférieur à celle des mâles (plus petite longueur de mâch. inf. pour les femelles).

• La dispersion chez les mâles a l'air légèrement plus importante que chez les femelles mais au vu du nbre de données, elles sont plutôt semblables.

Pos de valeurs ~~médiane~~ = 14,5 ...

Le 1er ~~quartile~~, le 3ème ~~quartile~~ ...

2) $m_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{i,1}$

$m_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} x_{i,2}$

AN $m_1 = 113,4$

$m_2 = 108,6$.

3) $s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_{i,1} - m_1)^2$ (non biaisé) → estim de σ^2

$s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (x_{i,2} - m_2)^2$ (non biaisé) → estim de σ^2 (à connaître)

$s^2 = \frac{1}{n_1 + n_2 - 2} \left((n_1 - 1) s_1^2 + (n_2 - 1) s_2^2 \right)$ → estim de σ^2 ("mixte" de s_1^2 et s_2^2)

$$= \frac{1}{n_1 + n_2 - 2} \left(\sum_{i=1}^{n_2} (x_{i,2} - m_2)^2 + \sum_{i=1}^{n_1} (x_{i,1} - m_1)^2 \right)$$

AN : $s^2 = 8,5$.

$X_{1,1}, \dots, X_{1,n_1} \sim \mathcal{N}(\mu, \sigma^2)$

$X_{1,2}, \dots, X_{1,n_2} \sim \mathcal{N}(\mu, \sigma^2)$

2 échantillons.

↳ On peut calculer la var. empir à partir des 2 éch.

S_1^2 estimateur de σ^2.

S_2^2 estimateur de σ^2.

On veut utiliser toutes les données pour estimer σ^2.
donc on choisit un estim qui "mixte" S_1^2 et S_2^2.

4) On teste au risque 5% : (égalité des moyennes)

à écrire

H_0 : « $\mu_1 = \mu_2$ » contre H_1 : « $\mu_1 \neq \mu_2$ ».

les moyennes des longueurs des mâch. des femelles et des mâles sont =

↳ cela voudrait dire le sexe n'a pas d'influence sur la long. moy. des mâch.

« p-value "petite" $\Rightarrow$ on rejette H_0 »

p-value < niveau de risque

On a la statistique de test

$$Z = \frac{\sqrt{n_1}\sqrt{n_2}\,(m_1 - m_2)}{\sqrt{n_1 + n_2}\; S} \sim T_{n_1+n_2-2} \text{ sous } H_0 .$$

On a la zone de rejet

$$ZR = \left\{ |Z| > \underbrace{t_{n_1+n_2-2,\,1-\alpha/2}} \right\}.$$

$\alpha = 0,05$

AN : $t_{18,\,0,975}$

où $t_{n_1+n_2-2,\,1-\alpha/2}$ est le quantile, d'ordre $1-\alpha/2$, de la loi de Student de degré n_1+n_2-2.

On a une p-value égale à $0,00336 < 0,05 \Rightarrow$ on rejette H_0.

5) On suppose $X_{1,1}, \dots, X_{1,n_1} \sim \mathcal{N}(\mu_1, \sigma_1^2)$

$X_{1,2}, \dots, X_{1,n_2} \sim \mathcal{N}(\mu_2, \sigma_2^2)$

(s_1^2 est un estim. de σ_1^2, s_2^2 est un estim. de σ_2^2)

On teste au risque 5% (égalité des variances)

H_0 : « $\sigma_1^2 = \sigma_2^2$ » contre H_1 : « $\sigma_1^2 \neq \sigma_2^2$ »

On a la statistique de test

$$Z = \frac{S_{X_1}^2}{S_{X_2}^2} \sim F_{n_1-1, n_2-1} \text{ sous } H_0.$$

(Fisher)

On a la zone de rejet :

$$ZR = \{Z < f_{n_1-1, n_2-1, \alpha/2}\} \cup \{Z > f_{n_1-1, n_2-1, 1-\alpha/2}\}$$

où $f_{n_1-1, n_2-1, 1-\alpha/2}$ est le quantile de la loi de Fisher de paramètres n_1-1, n_2-1, d'ordre $1-\alpha/2$

La p-value est ici $0,1578 > 0,05$ donc on ne rejette pas H_0.

Exo 7

2 échantillons. $X_{1,1}, \dots, X_{1,n_1} \sim \mathcal{N}(\mu_1, \sigma_1^2)$

$X_{1,2}, \dots, X_{1,n_2} \sim \mathcal{N}(\mu_2, \sigma_2^2)$

1) Test de Shapiro teste si l'échantillon est gaussien.

solution 1 test à 5%

Ici les p-valeurs pour l'échantillon 1 (resp. 2) sont de 0,3688 (resp. 0,8401). Or 0,3688 > 0,05 et 0,8401 > 0,05 donc on ne peut pas rejeter le caractère gaussien des données (ni dans l'échantillon 1, ni dans l'échantillon 2).

Solution 2

Pour tout niveau inférieur à 36%, on ne peut pas rejeter le caractère gaussien des données.

2) La médiane dans l'éch. 1 a l'air légèrement inférieure à celle dans le groupe 2. Les dispersions semblent similaires.

3) On teste au risque $\overbrace{5\%}^{=\alpha}$:

H_0 : " $\sigma_1^2 = \sigma_2^2$ " contre H_1 : " $\sigma_1^2 \neq \sigma_2^2$ ".

On a la statistique de test :

$$Z = \frac{S_1^2}{S_2^2} \sim F_{n_1-1, n_2-1} \text{ sous } H_0.$$

La zone de rejet est :

$$ZR = \{Z < f_{n_1-1, n_2-1, \alpha/2}\} \cup \{Z > f_{n_1-1, n_2-1, 1-\alpha/2}\},$$

$f_{n_1-1, n_2-1, \alpha/2}$ est le quantile de la loi de Fisher de paramètres n_1-1, n_2-1 d'ordre $\alpha/2$.

La p-value est ici égale à $0,7503 > \underbrace{0,05}_{=\alpha}$

$\Rightarrow$ on ne rejette pas H_0.

4) On suppose $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (c'est validé par le test de la p3)

On pose S^2 l'estimateur de σ^2.

On teste au risque 5%. (égalité des moyennes)

- H_0 : "$\mu_1 = \mu_2$" contre H_1 : "$\mu_1 \neq \mu_2$"

- On a la statistique de test

$$\sqrt{\frac{n_1 n_2}{n_1 + n_2}} \; \frac{\bar{X}_1 - \bar{X}_2}{S} \sim T_{n_1+n_2-2} \text{ sous } H_0$$

où $\bar{X}_1$ est la moyenne empirique de l'échantillon 1

$\bar{X}_2$ est ——— échantillon 2

- On a la zone de rejet : $\{|Z| > t_{n_1+n_2-2,\, 1-\alpha/2}\}$

où $t_{n_1+n_2-2}$ est le quantile de la loi de Student de degré n_1+n_2-2 d'ordre $1-\alpha/2$.

$\Rightarrow$ On a la p-value égale à 0,6048 > 0,05 donc on ne rejette pas H_0.