Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1:
2023/2024

Solution de l'exercice 1 de la série N°3

Exercice 1 *Une étude de sociologie porte sur le temps passé par des enfants, âgés de* 8 *à* 16 *ans, sur des jeux électroniques. La question est de savoir si le temps moyen par jour est de* 7 *heures. On a demandé à* 15 *enfants leurs nombres d'heures de jeu par jour et les réponses sont les suivantes:*

$$5 \quad 9 \quad 5 \quad 8 \quad 7 \quad 6 \quad 7 \quad 9 \quad 7 \quad 9 \quad 6 \quad 9 \quad 10 \quad 9 \quad 8$$

1. *En supposant que ce temps est normalement distribué, avec une variance égale à* 3*, que conclut-on au niveau de signification* 5% ?
2. *Répondre à la même question si la variance était inconnue.*

* * * * * * * * * * * * * * * * *

Solution. 1) Le cas variance connue $\sigma^2 = 3$. Il s'agit d'un test bilatéral:

$$\begin{cases} H_0 : & \mu = 7 \\ H_1 : & \mu \neq 7 \end{cases}$$

D'après le cours la région critique est de la forme

$$W := \left\{ (x_{1,\dots,} x_{15}) \in \mathbb{R}^+ : |\overline{x} - 7| \geq k \right\},$$

avec $\mathbf{P}_{\mu=7} \left(\left| \overline{X} - 7 \right| \geq k \right) = \alpha = 0.05$. En d'autres termes

$$\mathbf{P}_{\mu=7} \left(\left| \frac{\overline{X} - 7}{\sqrt{3/15}} \right| \geq \frac{k}{\sqrt{3/15}} \right) = 0.05,$$

ce qui est équivalent à $\mathbf{P} \left(|Z| \geq \sqrt{5}k \right) = 0.05$. Ceci implique que

$$\sqrt{5}k = \Phi^{-1} (1 - 0.05/2) = 1.96,$$

ainsi $k = 1.96/\sqrt{5} = 0.876$. La fonction test statistique est donc

$$\delta = \begin{cases} 1 & \text{si } |\overline{x} - 7| \geq 0.876 \\ 0 & \text{si } |\overline{x} - 7| < 0.876 \end{cases}.$$

Application: nous avons

$$\overline{x} = \frac{1}{15} (5 + 9 + 5 + 8 + 7 + 6 + 7 + 9 + 7 + 9 + 6 + 9 + 10 + 9 + 8) = 7.6.$$

Comme $|7.6 - 7| = 0.6 < 0.876$, alors en garde H_0, c'est à dire le temps passé par des enfants, âgés de 8 à 16 ans, sur des jeux électroniques est en effet de 7 heures.

2. **Le cas variance inconnue.** D'après le cours, la région critique est de la forme

$$W := \left\{ (x_{1,...,}x_{15}) \in \mathbb{R}^+ : \left| \frac{\overline{x} - 7}{\widetilde{s}/\sqrt{15-1}} \right| \geq t_{1-0.05/2} \right\},$$

où $\widetilde{s} = \sqrt{\frac{1}{15}\sum_{i=1}^{15}(x_i - \overline{x})^2}$ (la variance empirique observée) et $t_{1-0.05/2}$ est le quantile d'ordre $1 - 0.05/2 = 0.975$ de loi de student à $15 - 1 = 14$ degrè de liberté. De la table statistique en tire $t_{1-0.05/2} = 2.144$. La fonction test statistique est donc

$$\delta = \begin{cases} 1 & \text{si } \left| \frac{\overline{x}-7}{\widetilde{s}/\sqrt{15-1}} \right| \geq 2.144 \\ 0 & \text{si } \left| \frac{\overline{x}-7}{\widetilde{s}/\sqrt{15-1}} \right| < 2.144 \end{cases}.$$

Application: nous avons $\widetilde{s}^2 = \frac{1}{15}\sum_{i=1}^{15}(x_i - 7.6)^2 = 2.722$ qui donne $\widetilde{s} = 1.65$. Comme $\left| \frac{7.6-7}{1.65/\sqrt{15-1}} \right| = 1.360\,6 < 2.144$, alors on garde oncore une fois l'hypothèse nulle $H_{0.}$

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2023/2024
Module: tests statistiques

Solution de l'exercice 2 de la Série N°**3**

Exercice 2.

On suppose que le poids (en grammes) d'un paquet de café est une variable aléatoire normale de moyenne μ et de variance σ^2 (toutes les deux inconnues). Pour faire des tests, au niveau de signification 10%, sur ces deux paramètres, on dispose de l'échantillon suivant :

480 502 484 500 504 503 495 500 501 481

Que décidez-vous pour chacun des problèmes suivants :

$$\begin{cases} H_0 : & \mu = 500 \\ H_1 : & \mu \neq 500 \end{cases} \text{ et } \begin{cases} H_0 : & \sigma^2 = 25 \\ H_1 : & \sigma^2 > 25 \end{cases}$$

* * * * * * * * * * * * * * * * * *

Solution. Considérons tout d'abord le test bilatérale de la moyenne

$$\begin{cases} H_0 : & \mu = 500 \\ H_1 : & \mu \neq 500 \end{cases},$$

avec $n = 10$, $\alpha = 0.1$ et σ^2 est inconnue. D'après le cours, la région critique de ce test est donnée par

$$W = \left\{ (x_1, ..., x_{10}) \in \mathbb{R}_+^{10} \mid \frac{|\overline{x} - 500|}{\widetilde{s}/\sqrt{10-1}} \geq k \right\},$$

où k est la solution de l'équation $\mathbf{P}\left(\left|T_{(10-1)}\right| \geq k\right) = \alpha = 0.1$, avec $T_{(10-1)} = T_9$ est la variable aléatoire de Student à 9 degré de liberté. On rappelle que $\overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i$ et $\widetilde{s}^2 = \frac{1}{10}\sum_{i=1}^{10}(x_i - \overline{x})^2$ désignent la moyenne et la variance empiriques respectivement. De la table statistique de la loi de Student on obtient $k = 1.8331$, ainsi

$$W = \left\{ (x_1, ..., x_{10}) \in \mathbb{R}_+^{10} \mid \frac{|\overline{x} - 500|}{\widetilde{s}/3} \geq 1.8331 \right\}.$$

Le fonction statistique test correspondante est

$$\delta(x_1, ..., x_{10}) = \begin{cases} 1 & \text{si } \frac{|\overline{x}-500|}{\widetilde{s}/3} \geq 1.8331 \\ 0 & \text{si } \frac{|\overline{x}-500|}{\widetilde{s}/3} < 1.8331 \end{cases}.$$

La moyenne et l'écart-type des données observées sont $\overline{x}_{obs} = 495$ et $\widetilde{s}_{obs} \simeq 9$ respectivement, ainsi

$$\frac{|\overline{x}_{obs} - 500|}{\widetilde{s}_{obs}/3} = \frac{|495 - 500|}{9/3} = 1.6667 < 1.8331.$$

Donc on accepte l'hypothèse H_0 disant que le poids des paquets de café est de 500 grammes.

Traitons maintenant le test de la variance

$$\begin{cases} H_0 : & \sigma^2 = 25 \\ H_1 : & \sigma^2 > 25 \end{cases},$$

avec la moyenne μ inconnue. Nous avons aussi annoncé au cours la région critique de ce test est donnée par

$$W = \left\{ (x_1, ..., x_{10}) \in \mathbb{R}_+^{10} \mid \frac{10v^2}{25} \geq k \right\},$$

où $v^2 = \widetilde{s}^2 = \frac{1}{10}\sum_{i=1}^{10}(x_i - \overline{x})^2$, et k est la solution de l'équation $\mathbf{P}\left(\chi_9^2 \geq k\right) = 0.1$, où χ_9^2 désigne la variable de Chi-deux (ou de Pearson) à 9 dégres de liberté. De la table statistique on obtient $k = 14.683$, ainsi

$$\begin{aligned} W &= \left\{(x_1, ..., x_{10}) \in \mathbb{R}_+^{10} \mid \frac{10v^2}{25} \geq 14.683\right\} \\ &= \left\{(x_1, ..., x_{10}) \in \mathbb{R}_+^{10} \mid v^2 \geq 36.709\right\}. \end{aligned}$$

La fonction statistique test correspondante est

$$\delta(x_1, ..., x_{10}) = \begin{cases} 1 & \text{si } v^2 \geq 36.709 \\ 0 & \text{si } v^2 < 36.709 \end{cases}.$$

Comme $v_{obs}^2 = \widetilde{s}_{obs}^2 \simeq 9^2 = 81 > 36.709$, alors on rejette H_0; c'est à dire $\sigma^2 > 25$.

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2022/2023

Solution de l'exercice 3 de la Série N°3

Exercice 3. Le coiffeur du village affirme **qu'au moins** 90% de ses clients sont satisfaits de ses services. Les gens du village croient **qu'il exagère**. Alors ils décident de faire un test au niveau de signification 0.05. Sur 150 clients, 132 se disent satisfaits. Conclure.

* *

Solution. On note $X = 1$ (resp. $X = 0$) pour un client satisfait (resp. non satisfait) du service du client. Il donc s'agit d'une v.a. de Bernoulli de paramètre p :

$$\mathbf{P}(X = 1) = p \text{ et } \mathbf{P}(X = 0) = 1 - p.$$

Ceci peut être aussi réécrit sous la forme suivante

$$\mathbf{P}(X = x) = p^x (1 - p)^{1-x}, \; x = 0, 1.$$

Les gens du village veulent alors tester

$$\begin{cases} H_0 : & p \geq p_0 = 90\% \\ H_1 & p < 90\% \end{cases}$$

au niveau de signification 0.05, à partir d'un échantillon de taille $n = 150$. Pour cela nous allons utiliser le principe du rapport des vraisemblances monotone: pour $0 < p_1, p_2 < 1$, telles que $p_1 > p_2$, on écrit

$$\frac{L_1}{L_2} = \frac{\prod_{i=1}^{n} p_1^{x_i} (1 - p_1)^{1-x_i}}{\prod_{i=1}^{n} p_2^{x_i} (1 - p_2)^{1-x_i}} = \frac{p_1^t (1 - p_1)^{n-t}}{p_2^t (1 - p_2)^{n-t}}, \; t = \sum_{i=1}^{n} x_i.$$

En d'autres termes

$$\frac{L_1}{L_2} = \left(\frac{1 - p_1}{1 - p_2}\right)^n \left(\frac{p_1 (1 - p_2)}{p_2 (1 - p_1)}\right)^t.$$

Il est clair que $a := \frac{1-p_1}{1-p_2} > 0$, $b := \frac{p_1(1-p_2)}{p_2(1-p_1)} > 1$, donc la fonction $t \to a^n \times b^t$ est une fonction croissante en t. Donc la distribution de X, possède un rapport de vraisemblance croissant par rapport à t. est la statistique de test correspondante est $T_n = \sum_{i=1}^{n} X_i$. D'après le théorème central limite on a

$$Z = \frac{T_n - \mathbf{E}[T_n]}{\sqrt{\mathbf{Var}[T_n]}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1), \text{quand } n \to \infty.$$

Nous avons $\mathbf{E}[T_n] = n\mathbf{E}[X] = np$ et $\mathbf{Var}[T_n] = np(1 - p)$. En d'autres termes

$$T_n \rightsquigarrow \mathcal{N}(np, np(1 - p)), \text{ pour } n \text{ grand.}$$

Pour avoir une bonne approximation, on doit s'assurer que $n \geq 30$, $np \geq 5$ et $n(1 - p) \geq 5$. Nous avons $n = 150 > 30$, et pour $p = 0.9$,on a $150 \times 0.9 = 135 \geq 5$ et $150 \times (1 - 0.9) = 15 \geq 5$. Donc pour $p = 0.9$, on peut appliquer l'approximation gaussienne ci-dessus. Ainsi, notre région critique est

$$W = \left\{ (x_1, ..., x_{150}) \in \{0, 1\}^{150} \mid T_n = \sum_{i=1}^{150} X_i \leq c \right\},$$

ou

$$\mathbf{P}_{p=0.9} \left(\sum_{i=1}^{150} X_i \leq c \right) = 0.05.$$

Ce qui équivalent à

$$\mathbf{P}_{p=0.9} \left(\frac{T_{150} - 150 \times 0.9}{\sqrt{150 \times 0.9 (1 - 0.9)}} \leq \frac{c - 150 \times 0.9}{\sqrt{150 \times 0.9 (1 - 0.9)}} \right) = \mathbf{P}(Z \leq k) = 0.05.$$

De la table statistique, nous avons $k = -1.644854$, ainsi

$$\frac{c - 150 \times 0.9}{\sqrt{150 \times 0.9\,(1 - 0.9)}} = -1.644854,$$

ce qui donne $c = 128.96$. Alors

$$W = \left\{ (x_1, ..., x_{150}) \in \{0,1\}^{150} \mid \sum_{i=1}^{150} X_i \leq 128.96 \right\}.$$

Nous avons $T_{obs} = 132 > 128.96$, donc on ne rejette pas l'hypothèse H_0. En conclusion le coiffeur a raison d'affirmer qu'au moins 90% de ses clients sont satisfaits de ses services.

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2023/2024

Solution de l'exercice 4 de la Série N°3

Exercice 4. L'étude de l'exercice 1 était faite en 1999. La même étude est faite en 2009, avec l'échantillon suivant:

4 5 7 7 5 7 5 6 5 6 7 8 5 7 8 7

En supposant des distributions Gaussiennes de même variance, comparer les temps de jeu moyens par jour pour les deux années, au niveau signification 5%. Etudier le problème en supposant que les variances pas nécessairement égales.

* *

Solution. Il s'agit ici d' test de comparaison entre deux moyennes:

$$\begin{cases} H_0: & \mu_1 = \mu_2 \\ H_0: & \mu_1 \neq \mu_2 \end{cases}$$

Nous avons $\alpha = 0.05$, $n_1 = 15$, $n_2 = 16$, $\overline{x}_{1,obs} = 7.6$, $\overline{x}_{2,obs} = 6.18$, $\widetilde{s}^2_{1,obs} = 2.37$, $\widetilde{s}^2_{2,obs} = 1.40$. En supposant que les variances sont égales, la statistique de test est

$$\frac{\overline{X}_1 - \overline{X}_2}{\widetilde{S}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightsquigarrow t\left(n_1 + n_2 - 2\right),$$

où

$$\widetilde{S} = \sqrt{\frac{n_1 \widetilde{S}_1^2 + n_2 \widetilde{S}_2^2}{n_1 + n_2 - 2}},$$

et $t\left(n_1 + n_2 - 2\right)$ est la loi de Student à $n_1 + n_2 - 2$ degré de liberté. La région critique associée est

$$W = \left\{ \left(x_1^{(1)}, ..., x_1^{(15)}; x_2^{(1)}, ..., x_2^{(16)}\right) \in \mathbb{R}_+^{31} \mid \frac{\left|\overline{x}_1 - \overline{x}_2\right|}{\widetilde{s}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \geq t_{1-\alpha/2} \right\},$$

où

$$\widetilde{s} := \sqrt{\frac{n_1 \widetilde{s}_1^2 + n_2 \widetilde{s}_2^2}{n_1 + n_2 - 2}},$$

et $t_{1-\alpha/2}$ (la valeur critique) est le quantile d'ordre $1 - \alpha/2$ de $t\left(n_1 + n_2 - 2\right)$. En utilisons la table statistique de la loi de Student à $15 + 16 - 2 = 29$ degré de liberté on obtient $t_{1-0.05/2} = t_{0.975} = 2.04$. En outre nous avons

$$\widetilde{s}_{obs} = \sqrt{\frac{15 \times 2.37 + 16 \times 1.40}{15 + 16 - 2}} = 1.41,$$

et

$$\frac{\left|\overline{x}_{1,obs} - \overline{x}_{2,obs}\right|}{\widetilde{s}_{obs}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\left|7.6 - 6.18\right|}{1.4\sqrt{\frac{1}{15} + \frac{1}{16}}} = 2.82.$$

Comme $2.82 > 2.04$, alors on rejette l'hypothèse de l'égalité des moyennes de deux populations.

Supposons maintenant que les variances pas nécessairement égales, ce qui est une hypothèse réalistique. Dans ce cas on utilise **la statistique de Welch**:

$$\frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\widetilde{S}_1^2/n_1 + \widetilde{S}_2^2/n_2}} \simeq t\left(\nu\right),$$

où $t\left(\nu\right)$ est une v.a de Student à ν degré de liberté définit par

$$\nu := \frac{\left(\widetilde{s}_1^2/n_1 + \widetilde{s}_2^2/n_2\right)^2}{\widetilde{s}_1^4/\left(n_1^2\left(n_1 - 1\right)\right) + \widetilde{s}_2^4/\left(n_2^2\left(n_2 - 1\right)\right)}.$$

Pour notre cas:

$$\nu = \frac{(2.37/15 + 1.40/16)^2}{(2.37)^2 / \left((15)^2 (15-1)\right) + (1.40)^2 / \left((16)^2 (16-1)\right)} = 26.27.$$

Le degré de liberté est un nombre naturel non nulle, donc on doit effectuer une interpolation linéaire afin de calculer la valeur critique associée $t_{0.975}(\nu)$. En utilisant les deux quantiles les plus proches, $t_{0.975}(26)$ et $t_{0.975}(27)$, on obtient

$$\begin{aligned} t_{0.975}(26.27) &\simeq t_{0.975}(\lfloor 26.27 \rfloor) + (26.27 - \lfloor 26.27 \rfloor) \frac{t_{0.975}(\lceil 26.27 \rceil) - t_{0.975}(\lfloor 26.27 \rfloor)}{\lceil 26.27 \rceil - \lfloor 26.27 \rfloor}, \\ &= t_{0.975}(26) + (26.27 - 26) \frac{t_{0.975}(27) - t_{1-\alpha/2}(26)}{27 - 26} \\ &= 2.056 + 0.27 (2.052 - 2.056) \\ &\simeq 2.054. \end{aligned}$$

La région critique associée est donc

$$W = \left\{ \left(x_1^{(1)}, ..., x_1^{(15)}; x_2^{(1)}, ..., x_2^{(16)} \right) \in \mathbb{R}_+^{31} \mid \frac{|\overline{x}_1 - \overline{x}_2|}{\sqrt{\widetilde{s}_1^2/n_1 + \widetilde{s}_2^2/n_2}} \geq 2.054 \right\}.$$

La valeur observée (en valeur absolue) de la statistique du test est

$$\left| \frac{\overline{x}_{1,obs} - \overline{x}_{2,obs}}{\sqrt{\widetilde{s}_{1,obs}^2/n_1 + \widetilde{s}_{2,obs}^2/n_2}} \right| = \left| \frac{7.6 - 6.18}{\sqrt{2.37/15 + 1.4/16}} \right| = 2.86.$$

Nous avons $2.86 > 2.054$, encore une fois on rejette H_0. Passons maintenant au raisonnement par le biais de la p-value:

$$\begin{aligned} p - value &= \mathbf{P}(|T_{26.27}| \geq 2.948) = 2\mathbf{P}(T_{26.27} \geq 2.86) \\ &= 2(1 - \mathbf{P}(T_{26.27} \leq 2.86)). \end{aligned}$$

On pose $\mathbf{p}_\nu := \mathbf{P}(T_\nu \leq 2.86)$. En utilisant l'interpolation linéaire, on écrit:

$$\mathbf{p}_{26.27} \simeq 0.27 (\mathbf{p}_{27} - \mathbf{p}_{26}) + \mathbf{p}_{26}$$

:En utilisant le logiciel R, on trouve $\mathbf{p}_{27} = 0.9959$ et $\mathbf{p}_{26} = 0.9958$, ainsi

$$\mathbf{p}_{26.27} \simeq 0.27 (0.9959 - 0.9958) + 0.9958 = 0.9958$$

Finalement

$$\begin{aligned} p - value &= 2(1 - \mathbf{p}_{26.27}) \\ &= 2(1 - 0.9958) = 0.0084 < 0.05. \end{aligned}$$

Encore une fois on rejette H_0.

Remarque: Le tableau statistique de loi de Student (comme d'autres lois de probabilités) offre uniquement les quantiles, tandis que les valeurs de probabilités on les calculs numériquement à l'aide des langages Softwar, à savoir le R ou le Matlab. En utilisant un de ces deux derniers on obtient $\mathbf{p}_{27} = 0.9959$ et $\mathbf{p}_{26} = 0.9958$. Bien sur en examen on vous donne quelques valeurs de probabilités et l'étudiant choisi celle qui convient au problème.

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2021/2022

Solution de l'exercice 5 de la Série N°3

Exercice 5. On veut tester l'égalité des variances de deux populations $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ au niveau de signification $\alpha = 0.05$. Un échantillon de taille 16 de X_1 donne une variance empirique $\widetilde{s}^2_{1,obs} = 7.62$ et un échantillon de taille 12 de X_2 donne une variance empirique $\widetilde{s}^2_{2,obs} = 3.96$. Conclure.

****

Solution. Il s'agit ici d'un test (bilatéral) de comparaison entre deux variances:

$$\begin{cases} H_0: & \sigma_1^2 = \sigma_2^2 \\ H_0: & \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

Nous avons $\alpha = 0.05$, $n_1 = 16$, $n_2 = 12$, $\widetilde{s}^2_{1,obs} = 7.62$, $\widetilde{s}^2_{2,obs} = 3.96$. La statistique de test (variable de décision) à utiliser est:

$$\frac{\frac{16\widetilde{S}_1^2/\sigma_1^2}{15}}{\frac{12\widetilde{S}_2^2/\sigma_2^2}{11}} \rightsquigarrow F(15, 11), \quad \left(\text{sous } H_0 : \sigma_1^2 = \sigma_2^2\right), \tag{1}$$

où $F(15, 11)$ désigne la loi de Fisher à $(15, 11)$ degrés de liberté. A un seuil de signification $\alpha = 5\%$, la région critique associée est

$$W = \left\{ \left(x_1^{(1)}, ..., x_1^{(16)}; x_2^{(1)}, ..., x_2^{(12)}\right) \in \mathbb{R}^{28} \mid \frac{\widetilde{s}_1^2}{\widetilde{s}_2^2} \geq \frac{12 \times 15}{16 \times 11} c_1 \text{ ou } \frac{\widetilde{s}_1^2}{\widetilde{s}_2^2} \leq \frac{12 \times 15}{16 \times 11} c_2 \right\},$$

où c_1 et c_2 sont deux constantes telles que

$$\mathbf{P}(F(15, 11) \geq c_1) = \mathbf{P}(F(15, 11) \leq c_2) = \alpha/2 = 0.025.$$

Donc c_1 est le quantile d'ordre $1 - 0.025 = 0.975$ et c_2 est le quantile d'ordre 0.025 de la loi Fisher à $(15, 11)$ degrés de liberté. De la table statistique de Fisher on obtient $c_1 = 3.40$. Pour trouver la valeur de c_2, on utilise la formule suivante:

$$F(\nu_1, \nu_2) \rightsquigarrow \frac{1}{F(\nu_2, \nu_1)}.$$

Donc

$$\mathbf{P}(F(15, 11) \leq c_2) = 0.025 \Longleftrightarrow \mathbf{P}(F(11, 15) \geq 1/c_2) = 0.025.$$

De la table statistique on obtient $1/c_2 =$qui donne $c_2 = 0.33$. La région critique est donc

$$W = \left\{ \left(x_1^{(1)}, ..., x_1^{(16)}; x_2^{(1)}, ..., x_2^{(12)}\right) \in \mathbb{R}^{28} \mid \frac{\widetilde{s}_1^2}{\widetilde{s}_2^2} \geq 3.40 \text{ ou } \frac{\widetilde{s}_1^2}{\widetilde{s}_2^2} \leq 0.33 \right\},$$

Nous avons

$$\frac{\widetilde{s}^2_{1,obs}}{\widetilde{s}^2_{2,obs}} = \frac{7.62}{3.96} = 1.92.$$

Cette valeur est ni ≥ 3.40 ni ≤ 0.33, donc on accepte l'égalité des deux variances. La p-value dans notre cas est égale à

$$\begin{aligned} p - value &= 2\min\left\{ \mathbf{P}\left(\frac{\widetilde{S}_1^2}{\widetilde{S}_2^2} \geq 1.92\right), \mathbf{P}\left(\frac{\widetilde{S}_1^2}{\widetilde{S}_2^2} \leq 1.92\right) \right\} \\ &= 2\min\left\{ \mathbf{P}\left(\frac{\frac{16\widetilde{S}_1^2/\sigma_1^2}{15}}{\frac{12\widetilde{S}_2^2/\sigma_2^2}{11}} \geq \frac{\frac{16/\sigma_1^2}{15}}{\frac{12/\sigma_2^2}{11}} 1.92\right), \mathbf{P}\left(\frac{\frac{16\widetilde{S}_1^2/\sigma_1^2}{15}}{\frac{12\widetilde{S}_2^2/\sigma_2^2}{11}} \leq \frac{\frac{16/\sigma_1^2}{15}}{\frac{12/\sigma_2^2}{11}} 1.92\right) \right\} \\ &= 2\min\left\{ \mathbf{P}\left(F(15, 11) \geq \frac{\frac{16}{15}}{\frac{12}{11}} 1.92\right), \mathbf{P}\left(F(15, 11) \leq \frac{\frac{16}{15}}{\frac{12}{11}} 1.92\right) \right\}, \end{aligned}$$

car sous H_0 on a $\sigma_1^2 = \sigma_2^2$. Donc

$$p - value = 2\min\left\{\mathbf{P}\left(F\left(15, 11\right) \geq 1.87\right), \mathbf{P}\left(F\left(15, 11\right) \leq 1.87\right)\right\}.$$

En utilisant le langage R on obient $\mathbf{P}\left(F\left(15, 11\right) \leq 1.87\right) = 0.85$, ainsi

$$p - value = 2\min\left\{1 - 0.85, 0.85\right\} = 0.3 > 0.05,$$

donc, effet l'égalités des deux variances est encore une fois confirmée.

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2023/2024

Solution de l'exercice 6 de la Série N°3

Exercice 6. Selon un institut de sondage A, 510 sur 980 personnes interrogées sont favorables à une certaine mesure gouvernementale. Un autre institut de sondage B donne 505 personnes favorables (à la même mesure) sur 1030. Tester, au niveau de signification 5%, l'égalité des proportions de gens favorables proposées par les deux instituts.

****

Solution. Il s'agit ici d' test de comparaison entre deux proportions:

$$\begin{cases} H_0: & p_1 = p_2 \\ H_1: & p_1 \neq p_2 \end{cases}$$

Nous avons $\alpha = 0.05$, $n_1 = 980$, $n_2 = 1030$, $p_{1,obs} = 510/980$, $p_{2,obs} = 505/1030$. Sous l'hypothèse "$p_1 = p_2$", la statistique de test est

$$\frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}\,(1-\widehat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \rightsquigarrow \mathcal{N}(0,1),$$

où

$$\widehat{p} := \frac{n_1\widehat{p}_1 + n_2\widehat{p}_2}{n_1+n_2},$$

et l'estimateur commun de $p_1$et p_2, à condition que

$$n_1 + n_2 > 30,\ n_i p_i \geq 5 \text{ et } n_i(1-p_i) \geq 5,\ \text{pour } i = 1,2.$$

Pour notre problème nous avons:

$$n_1 + n_2 = 2010 > 30;\ n_1p_1 = 510 > 5;\ n_2p_2 = 505 > 5.$$

La région critique associée est :

$$W = \left\{ \left(x_1^{(1)},...,x_1^{(980)}; x_2^{(1)},...,x_2^{(1030)}\right) \in \mathbb{R}_+^{2010} \mid \frac{|\widehat{p}_1 - \widehat{p}_2|}{\sqrt{\widehat{p}\,(1-\widehat{p})\left(\frac{1}{980}+\frac{1}{1030}\right)}} \geq z_{1-0.05/2} \right\},$$

où $z_{1-0.05/2}$ (la valeur critique) est le quantile d'ordre $1-\alpha/2$ de $\mathcal{N}(0,1)$. De la table statistique de la loi normale centré réduite on obtient $z_{1-0.05/2} = z_{0.975} \simeq 1,88$, ainsi

$$W = \left\{ \left(x_1^{(1)},...,x_1^{(980)}; x_2^{(1)},...,x_2^{(1030)}\right) \in \mathbb{R}_+^{2010} \mid \frac{|\widehat{p}_1 - \widehat{p}_2|}{\sqrt{\widehat{p}\,(1-\widehat{p})\left(\frac{1}{980}+\frac{1}{1030}\right)}} \geq 1.88 \right\}$$

La valeur observée de $\widehat{p}$ donne:

$$p_{obs} = \frac{n_1p_{1,obs} + n_2p_{2,obs}}{n_1+n_2} = \frac{510+505}{980+1030} = 0.50498.$$

Ainsi la valeur observée de la statistique de test est

$$\frac{|p_{1,obs} - p_{2,obs}|}{\sqrt{p_{obs}(1-p_{obs})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{|510/980 - 505/1030|}{\sqrt{0.50498\,(1-0.50498)\left(\frac{1}{980}+\frac{1}{1030}\right)}} \simeq 1.35.$$

Comme $1.35 < 1.88$, alors on accepte l'hypothèse d'égalité de deux proportions.