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 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

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

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.