L3-SV, Biostatistiques
Université de Nice-Sophia Antipolis

Corrigés

Exercice 5

1. Testons la normalité de chacun des échantillons (on le le fait que pour le premier échantillon) avec le test de Kolmogorov Smirnov. On teste H_0=normalité contre H_1=non normalité.

x	z $(=\frac{x-\overline{x}}{s_x})$	F	$\widehat{F}$ (calculé avec excel)	$d_i^+ = \lvert F_i - \widehat{F}_i \rvert$	$d_i^- = \lvert F_{i-1} - \widehat{F}_i \rvert$
0,144	-1,783	1/12	0,037	0,046	0,037
0,171	-0,837	2/12	0,201	0,035	0,035
0,178	-0,592	3/12	0,277	0,027	0,027
0,184	-0,382	4/12	0,351	0,018	0,018
0,193	-0,067	5/12	0,473	0,057	0,057
0,197	0,073	6/12	0,529	0,029	0,029
0,198	0,108	7/12	0,543	0,040	0,040
0,199	0,143	9/12	0,557	0,193	0,193
0,199	0,143	9/12	0,557	0,193	0,193
0,206	0,388	10/12	0,651	0,182	0,182
0,216	0,739	11/12	0,77	0,147	0,147
0,258	2,210	12/12	0,986	0,014	0,014

$\overline{x} = 0,195$, $s_x = 0,029$ (calculé avec excel)

Remarques :

- Il est possible qu'excel ne donne pas le même résultat qu'un calcul « à la main ».
- $\widehat{F}$ est la colonne des cumulés théoriques. Pour calculer $\widehat{F}$ à l'aide de la table : si $t < 0$, $\widehat{F}(t) = \mathbb{P}(Z < t)$ (avec Z de loi $\mathcal{N}(0,1)$) est égal à $1 - \widehat{F}(-t)$ et on trouve $\widehat{F}(-t)$ dans la table.
- Pour remplir la première ligne des d_i^-, on reporte la première valeur de $\widehat{F}$.

On calcule D_{obs} l'élément max dans les d_i^+, d_i^- : $D_{obs} = 0,19$. Puis (pour $\alpha = 0,05$) :

$$S = \sqrt{n} - 0,01 + \frac{0,085}{\sqrt{n}} = 3,48\,,\; D_\alpha = 0,895/S = 0,26\,.$$

Comme $D_{obs} < D_\alpha$, on garde H_0.

2. Testons maintenant l'équivariance (H_0=équivariance contre H_1=pas équivariance) avec le mini-test F. On compare les variances :

$$F_{calc} = \frac{s_{x,2}}{s_{x,1}} = \frac{(0,038)^2}{(0,029)^2} = 1,77\,,$$

en mettant la plus grande variance en haut (0,038 est l'écart-type de la colonne « pesticides à 3 ans »). Puis on regarde dans la table (avec $n_1 = 12$ la taille du premier échantillon, $n_2 = 13$ la taille du deuxième, $\nu_1 = n_1$, $\nu_2 = n_2$)

$$F_{\frac{\alpha}{2};\nu_2;\nu_1} = F_{0,025;12;11} = 3,47\,.$$

Attention, si $s_{x,1}$ était plus grand que $s_{x,2}$, il faudrait regarder $F_{\frac{\alpha}{2};\nu_1;\nu_2}$ (qui est différent de $F_{\frac{\alpha}{2};\nu_2;\nu_1}$) (attention, c'est bien $\alpha/2$ et pas α dans le $F_{\dots}$). Ici, $F_{calc} < F_{\frac{\alpha}{2};12;11}$ donc on garde H_0.

3. Puisque nous avons deux échantillons de loi normale et de même variance qui ne sont indépendants, avec $n_1 < 30$ (il suffit d'avoir n_1 ou $n_2 < 30$), nous allons utiliser le test t de

Student pour échantillons indépendants. Nous testons $H_0 : \mu_1 = \mu_2$ contre $H_1 : \mu_1 \neq \mu_2$ (μ_1 et μ_2 sont les moyennes des deux échantillons). Il s'agit donc d'un test bilatéral. Calculons

$$s_{pd} = \sqrt{\frac{\nu_1 s_{x,1}^2 + \nu_2 s_{x,2}^2}{\nu_1 + \nu_2}} = 0,034\,;\; t_{calc} = \frac{\overline{x}_1 - \overline{x}_2}{s_{pd}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = -11,18\,.$$

Soit $\nu = n_1 + n_2 - 2 = 11 + 12 = 23$. On regarde dans la table (de la loi de Student) : $t_{\alpha/2;\nu} = t_{0,025;23} = 2,069$. Comme $|t_{calc}| > t_{\alpha/2;\nu}$, on rejette H_0 pour H_1 et on en conclut que les moyennes des deux échantillons diffèrent de façon significative.

Exercice 6

Les données sont appariées. Il faut commencer par tester la normalité des différences des données mais nous allons sauter cette étape (voir exercice précédent pour un exemple). Le nombre de données n est < 30. Soit $\alpha = 0,05$. On veut tester $H_0 : \mu_1 = \mu_2$ contre $H_1 : \mu_1 > \mu_2$ (μ_1 et μ_2 sont les moyennes des deux échantillons). Il s'agit donc d'un test unilatéral. Nous allons utiliser le test de Student pour données appariées. Notons $\overline{x}_1$, $\overline{x}_2$ les moyennes empiriques des deux échantillons.

Filtre A	Filtre B	différences
65	53	12
80	63	17
89	62	27
64	52	12
68	64	4
68	50	18
86	75	11
54	35	19
91	72	19
77	59	18
77	63	12
86	55	31

On calcule (s_d est l'écart-type empirique des différences)

$$\nu = n - 1 = 11\,;\; \overline{d} = \overline{x}_1 - \overline{x}_2 = 1,83\,;\; t_{\overline{d}} = \frac{\overline{d}\sqrt{n}}{s_d} = 8,75\,.$$

On regarde dans la table de la loi de Student : $t_{\alpha;\nu} = t_{0,05;11} = 21,796$. Nous avons $t_{\overline{d}} > t_{0,025;11}$ donc nous rejetons H_0 pour H_1. Nous concluons que le filtre A est meilleur que le filtre B (ce qui n'est pas très satisfaisant puisque le test nous dit que $\mu_1 > \mu_2$ est plus crédible que $\mu_1 = \mu_2$, pas que c'est plus crédible que $\mu_1 \leq \mu_2$).

Remarque. Si $\mu_1 = \mu_2$ alors $t_{\overline{d}}$ suit une loi de Student à ν degrés de liberté. Le $t_{\overline{d}}$ que nous observons est plus grand que 95% des variables de Student, c'est pourquoi nous en concluons que H_0 est moins crédible que H_1.
Si on veut tester H_0 contre $H_2 : \mu_1 \neq \mu_2$, il faut regarder $t_{0,025;\nu} = 2,201$. Puisque $|t_{\overline{d}}| > 2,201$, on rejette H_0 pour H_2.

Remarque. Comment lire la table de la loi de Student. Soit T_ν variable de loi de Student à ν degrés de liberté. Soit $\alpha \in [0;1/2]$. Le nombre $t_{\alpha;\nu}$ est le nombre (≥ 0) tel que $\mathbb{P}(T_\nu > t_{\alpha;\nu}) = \alpha$. On lit ce nombre en partant de la ligne « α unilatéral ». Comme la densité de T_ν est symétrique, nous avons alors aussi : $\mathbb{P}(T_\nu < t_{\alpha;\nu}) = \alpha$. En particulier, pour tout α,

$$\mathbb{P}(|T_\nu| > t_{\frac{\alpha}{2};\nu}) = \mathbb{P}(T_\nu > t_{\frac{\alpha}{2};\nu}) + \mathbb{P}(T_\nu < -t_{\frac{\alpha}{2};\nu}) = \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha\,.$$

Si je veux t tel que $\mathbb{P}(|T_\nu| > t) = \alpha$, je cherche dans la table en partant de la ligne « α bilatéral » et je trouve le même résultat qu'en partant de la line « α unilatéral » avec $\alpha/2$ à la place de α.