UFR de Sciences économiques et de Gestion

UNIVERSITÉ DE LYON

L3 - Statistique Inférentielle

Corrigé de l'examen de 2ème session - mai 2026

Durée 1h 30

Exercice 1 : **(Barème de notation : a) 3 pt b) 2.5 pts c) 2 pts = 7.5 pts)**

a) Test non paramétrique du khi-deux d'homogénéité :

Hypothèses statistiques :

$\begin{cases} H_0 : \text{la préférence en faveur d'un changement d'horaire est identique dans les trois départements} \\ H_1 : \text{la préférence en faveur dun changement d'horaire est différente dans les trois départements} \end{cases}$

Tableau des Effectifs Observés (Théoriques) :

Changement d'horaire	Rhône	Savoie	Cantal	Total
Favorable	304 (304,8)	238 (228,6)	220 (228,6)	762
Défavorable	96 (95,2)	62 (71,4)	80 (71,4)	238
Total	400	300	300	1000

Test non paramétrique du Khi-deux d'homogénéité au seuil de signification $\alpha = 5\%$.

Le nombre de degrés de liberté : $\nu = (p-1)(k-1) = 2$ d.d.l. avec $k = 3$ départements et $p = 2$ modalités de la variable changement d'horaires. Avec un risque $\alpha = 5\%$, le $\chi^2_{(5\%\,;\,2d.d.l.)} = 5.991$ cf. annexe table du khi-deux.

$$\chi^2_{calculé} = \sum_{i=1}^{6} \frac{(O_i - T_i)^2}{T_i} = \frac{(304-304,8)^2}{304,8} + \cdots + \frac{(80-71,4)^2}{71,4} = 2,99.$$

Règle de décision du test : A partir du tableau des répartitions observée et théorique, on détermine la valeur du khi-deux sous l'hypothèse nulle H_0 d'homogénéité : $\chi^2_{calculé} = 2,99$. On ne peut rejeter l'hypothèse nulle H_0 puisque $\chi^2_{calculé} = 2.99 < \chi^2_{(5\%;2)} = 5.991$. On en conclut, avec un risque d'erreur $\alpha = 5\%$, que les trois départements de la région Auvergne-Rhône-Alpes ont un comportement homogène en ce qui concerne le changement d'horaires.

b) Intervalle de confiance et ceci pour l'ensemble des trois départements, de la proportion de clients favorables au changement d'horaire :

$\widehat{p} = \frac{762}{1000} = 76.20\%$: Proportion estimée de clients favorables au changement d'horaires dans les trois départements de la région Auvergne-Rhône-Alpes.

La statistique de test : $\dfrac{\widehat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \rightarrow N(0,1)$

Risque d'erreur : $\alpha = 5\% \Rightarrow u_{\frac{\alpha}{2}=2.5\%} = \pm 1.96$ cf. annexe table N(0 , 1)

Marge d'erreur associée à l'estimation de la proportion:

$$E = u_{\frac{\alpha}{2}} \times \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.96 \times \sqrt{\frac{0.762 \times 0.238}{1000}} = 2.64\%$$

- I.C. de niveau 95% de p : $p \in [73.56\% \ ; \ 78.84\%]$

c) les clients du département de la Savoie sont plus favorables au changement d'horaire que les clients du département du Cantal :

Test de comparaison de 2 proportions :

Hypothèses statistiques : Test unilatéral risque à droite

$$\begin{cases} \text{H}_0 : \text{P}_S \leq P_C \\ \text{H}_1 : \text{P}_S > P_C \end{cases}$$

Conditions d'application du test : échantillons indépendants provenant de deux populations normales - Grands échantillons $n_S = n_C = 300$.

La statistique de test : $\dfrac{(\widehat{P}_S - \widehat{P}_C) - (p_S - p_C)}{\sqrt{\frac{p_S q_S}{n_S} + \frac{p_C q_C}{n_C}}} \rightarrow N(0\ ,\ 1)$

Proportions observées en Savoie : $\widehat{p}_S = \frac{238}{300} = 79.33\%$

Proportions observées dans le Cantal : $\widehat{p}_S = \frac{220}{300} = 73.33\%$

Ecart observé : $\widehat{p}_S - \widehat{p}_C = 79.33 - 73.33\% = 6\%$

Risque d'erreur : $\alpha = 5\% \ \Rightarrow \ u_{\alpha=5\%} = 1.645$ cf. table N(0 , 1)

Valeur de la statistique de test sous $H_0 : P_S = P_C$

$$u_0 = \frac{(\widehat{P}_S - \widehat{P}_C) - (p_S - P_C)}{\sqrt{\frac{\widehat{p}_S \widehat{q}_S}{n_S} + \frac{\widehat{p}_C \widehat{q}_C}{n_C}}} = \frac{0.06 - 0}{\sqrt{\frac{0.7933 \times 0.2067}{300} + \frac{0.7333 \times 0.2667}{300}}} = 1,7331$$

Règle de décision et conclusion : La valeur $u_0 = 1.7331 > u_{\alpha=5\%} = 1.645$ appartient à la zone de rejet de H_0, on peut conclure avec un risque d'erreur $\alpha = 5\%$, que la différence observée $(\widehat{p}_S - \widehat{p}_C) = 6\%$ est significative ; les clients du département de la Savoie sont plus favorables au changement d'horaire que les clients du département du Cantal.

********** ◦◦◦ **********

Exercice 2 : (Barème de notation : a) 2.5 pts b) 2.5 pts c) 2.5 pts = 7.5 pts)

a) Étude de la variabilité : comparaison de variances :

Type	Classique (C)	En ligne (L)
Nombre de banques	15	8
Rendement annuel moyen observé (%)????????????	2.45	2.60
Variance observée du rendement	0.04	0.09

Banques classiques : $n_C = 15$; $\bar{x}_C = 2.45$; $s_C^2 = 0.04$; $s_C^{*2} = \frac{n_C}{n_C - 1} s_C^2 = 0.0429$

Banques en ligne : $n_L = 8$; $\bar{x}_L = 3.10$; $s_L^2 = 0.09$; $s_L^{*2} = \frac{n_L}{n_L - 1} s_L^2 = 0.1029$

Conditions d'application du test : échantillons indépendants provenant de deux populations normales, les deux échantillon sont de petites tailles ($n_C = 15$; $n_L = 8 < 30$), les variances des rendements sont inconnues σ_C^2 et σ_L^2. La statistique de test : $\dfrac{\sigma_L^2 s_C^{*2}}{\sigma_C^2 s_L^{*2}} \rightarrow F_{(n_C - 1 = 14 \ ; \ n_L - 1 = 7)}$

Hypothèses statistiques : Test bilatéral symétrique

$$\begin{cases} \mathrm{H}_0 : \frac{\sigma_L^2}{\sigma_C^2} = 1 \;\Leftrightarrow\; \sigma_C^2 = \sigma_L^2 \;\text{ égalité des variances.} \\ \mathrm{H}_1 : \frac{\sigma_L^2}{\sigma_C^2} \neq 1 \;\Leftrightarrow\; \sigma_C^2 \neq \sigma_L^2 \;\text{ les variances sont différentes.} \end{cases}$$

Seuil de signification : $\alpha = 5\%$

Calcul de la statistique de test sous l'hypothèse nulle H_0 d'égalité des variances : $\frac{\sigma_L^2}{\sigma_C^2} = 1$

$$f_0 = \frac{s_C^{*2}}{s_L^{*2}} = \frac{0.0429}{0.1029} = 0.4167$$

Règle de décision - Les valeurs critiques (Table de Fisher) :

$$f_2 = f_{(2.5\%\,,\,14\,,\,7)} = 4.60 \quad ; \quad f_1 = f_{(97.5\%\,,\,14\,,\,7)} = \frac{1}{f_{(2.5\%\,,\,7\,,\,14)}} = \frac{1}{3.38} = 0.296$$

Conclusion : vu que $f_0 = 0.4167 \in [f_1 = 0.296\,;\, f_2 = 4.60]$, l'hypothèse nulle H_0 n'est pas rejetée. On peut donc conclure avec un seuil de signification $\alpha = 5\%$, qu'il n'y a pas de différence significative entre les variances des rendements de ces deux types de banques, on peut alors les considérer comme égales, $\sigma_C^2 \approx \sigma_L^2$.

b) Comparaison de deux moyennes :

Conditions d'application du test : petits échantillons $n_C = 18$, $n_L = 8\,(< 30)$ indépendants provenant de deux populations normales de variances inconnues, testées égales a) $\sigma_C^2 \approx \sigma_L^2$.

Les rendements moyens des banques en ligne sont-ils meilleurs ou plus importants que ceux des banques classiques ?

Hypothèses statistiques : Comparaison de 2 moyennes - Test Unilatéral risque à gauche

$$\begin{cases} \mathrm{H}_0 : m_C - m_L = 0 \;\Leftrightarrow\; m_C = m_L \\ \mathrm{H}_1 : m_C - m_L < 0 \;\Leftrightarrow\; m_C < m_L \;\text{ les rendements moyens des banques en ligne sont meilleurs} \end{cases}$$

La statistique de test : $\frac{(\overline{X}_C - \overline{X}_L) - (m_C - m_L)}{s^* \sqrt{\frac{1}{n_C} + \frac{1}{n_L}}} \;\rightarrow\; T_{(n_C + n_L - 2 = 21 d.d.l.)}$

La valeur critique (cf. table Student) : avec $\alpha = 5\%$; 21 degrés de liberté : $t_{(5\%\,,\,21)} = -1.721$.

$SCE_C = n_C \times s_C^2 = 0.6$ et $\mathrm{SCE}_L = n_L \times s_L^2 = 0.72$

Estimation de la variance commune $\sigma^2 = \sigma_C^2 \approx \sigma_L^2$:

$$s^{*2} = \frac{SCE_C + SCE_L}{n_C + n_L - 2} = \frac{0.6 + 0.72}{15 + 8 - 2} = 0.06286 = 0.2507^2$$

Valeur de la statistique de test sous H_0 :

$$t_0 = \frac{(\overline{X}_C - \overline{X}_L) - (m_C - m_L)}{s^* \sqrt{\frac{1}{n_C} + \frac{1}{n_L}}} = \frac{(2.45 - 2.60) - 0}{0.2507 \sqrt{\frac{1}{15} + \frac{1}{8}}} = -1.3667$$

Conclusion : vu que $t_0 = -1.3667 > t_{\alpha = 5\%} = -1.721$ alors non-rejet de l'hypothèse nulle H_0. On peut donc conclure avec un risque d'erreur $\alpha = 5\%$, que le rendement moyen de la banque an ligne n'est pas significativement supérieur à celui de la banque classique. L'écart observé de 0.15 n'est pas significatif, il est dû aux fluctuations d'échantillonnage.

c) Niveau de confiance $1 - \alpha$ associé à l'intervalle de confiance $(m_C - m_L) \in [-0.3682\,;\, 0.0682]$:

Différence observée : $(\bar{x}_C - \bar{x}_L) = -0.15$.
Marge de l'intervalle $\Rightarrow E = \frac{0.0682-(-0.3682)}{2} = 0.2182$

Marge d'erreur : $E = t_{\frac{\alpha}{2}} s^* \sqrt{(1/n_C)+(1/n_L)}$
$\Rightarrow t_{\frac{\alpha}{2}} = \frac{E}{s^*\sqrt{(1/n_C)+(1/n_L)}} = \frac{0.2182}{0.2507\sqrt{(1/15)+(1/8)}} = 1.988$
$F(1.988) = 0.97 \Rightarrow \frac{\alpha}{2} = 0.03 \Rightarrow \alpha = 6\% \Rightarrow 1-\alpha^* = 94\%$ le niveau de confiance.

********** °°° **********

Exercice 3 : (Barème de notation : a) 2.5 pts b) 2.5 pts = 5 pts)

- **a) Test non-paramétrique du khi-deux d'ajustement :**

Hypothèses statistiques :
Test d'adéquation ou d'ajustement à une loi de probabilité donnée - Test unilatéral risque à droite

$\begin{cases} H_0 : \text{les croisements suivent une distribution Mendélienne} \\ H_1 : \text{les croisements ne suivent pas une distribution Mendélienne.} \end{cases}$

Couleur	Rouge	Rose	Blanc	Total
Proportions mendéliennes théoriques	25%	50%	25%	100%
Nombre de plantes observées : O_i	130	325	145	600
Nombre de plantes mendéliennes théoriques T_i	150	300	150	600

Statistique de test du Khi-deux d'ajustement sous H_0 :

Seuil de signification : $\alpha = 5\%$; $k = 3$; $\nu = (k-1) = 2\ d.d.l.$
Calcul de la statistique de test sous H_0 : distribution de Mendel
$\chi^2_{calculé} = \sum_{i=1}^{k=3} \frac{(O_i-T_i)^2}{T_i} = \frac{(130-150)^2}{150} + \frac{(325-300)^2}{300} \frac{(145-150)^2}{150} = 4.92$

Règle de décision et conclusion :

Pour $\alpha = 5\%$, $Proba\ (\chi^2_{2\ d.d.l.} \leq k_{5\%}) = 0.95 \Rightarrow k_{5\%} = 5.991$
$\chi^2_{calculé} = 4.92 < \chi^2_{(5\%\ ,\ 2\ d.d.l.)} = 5.991$ cf. table du Khi-deux. On ne rejette pas l'hypothèse nulle H_0. On peut donc conclure avec un risque d'erreur $\alpha = 5\%$ que les croisements de ces plantes suivent bien une loi de Mendel.

********** °°° **********

- **b) Test non-paramétrique de Wilcoxon - Echantillons appariés :**

Entreprise	E1	E2	E3	E4	E5	E6	E7	E8
X : 2025	35.00	39.00	35.50	36.00	35.50	39.00	38.40	35.00
Y : 2024	36.00	38.00	38.20	36.50	35.00	39.50	36.00	32.50
Différence : $d_i = x_i - y_i$	-1	1	-2.70	-0.5	0.5	-0.5	2.40	2.5
Rang en valeur absolue $\mid d_i \mid$	4.5	**4.5**	7	8	**2**	2	**5**	**6**

Hypothèses statistiques : Test non-paramétrique

$$\begin{cases} H_0 : \widetilde{m}_x = \widetilde{m}_y & \text{les deux distributions du temps de travail sont identiques} \\ H_1 : \widetilde{m}_x \neq \widetilde{m}_y & \text{les deux distributions du temps de travail sont différentes} \end{cases}$$

Remarque : taille de l'échantillon $n = 8$ actions.

$T^+ = 17.5$: somme des rangs des différences positives $d_i > 0$,

$T^- = 21.5$: somme des rangs des différences négatives $d_i < 0$,

Valeur critique pour $n = 8$ et $\alpha = 5\%$, test bilatéral : $T_\alpha = 3$ (cf. table Wilcoxon échantillons appariés.)

$T = min(T^+, T^-) = min(17.5; 21.50) = 17.5 > T_\alpha = 3 \Rightarrow$ Non-Rejet de H_0. On peut donc considérer avec un risque d'erreur $\alpha = 5\%$, que les deux distributions du temps de travail par semaine ne sont pas différentes.

********** °°° **********