L3-S6 : Statistique Inférentielle

Corrigé de l'Examen Terminal 2ème Session - 2024-2025

Durée 1h 30

Exercice 1 : **(Barème de notation : a) 1.5 pt b) 1.5 pt c) 1.5 pt d) 2 pts e) 2 pts = 8.5 pts)**

Conditions d'application du test : grande taille d'échantillon $n_R = 400$.

La statistique de test : $\frac{\widehat{P}_R - p_R}{\sqrt{\frac{p_R q_R}{n_R}}} \rightarrow N(0\ ,\ 1)$

Echantillon Rhône : $n_R = 400$; $\widehat{p}_R = \frac{304}{300} = 76\%$

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

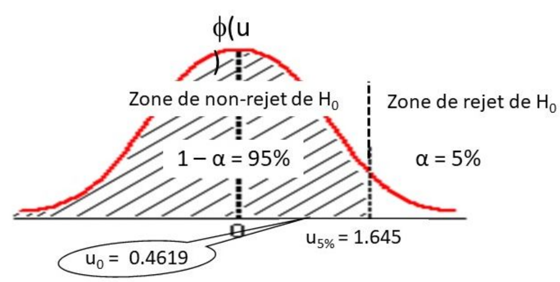
Hypothèses statistiques - Comparaison d'une proportion à une valeur donnée

Test Unilatéral - Risque à droite :

$$\begin{cases} \text{H}_0 : p_R \leq p_0 = 75\% \\ \text{H}_1 : p_R > p_0 = 75\% \text{ la proportion de clients favorables est supérieure à } 75\% \end{cases}$$

Statistique de test sous H_0 : $p_R = p_0 = 75\%$

$u_0 = \frac{\widehat{p}_R - p_0}{\sqrt{\frac{p_0 q_0}{n_R}}} = \frac{0.76-0.75}{\sqrt{\frac{0.75\times0.25}{400}}} = 0.4619$



Règle de décision et conclusion :

La valeur de la statistique de test $u_0 = 0.44619 < u_{5\%} = 1.645$ est dans la zone de non-rejet de H_0. On peut donc conclure avec un risque d'erreur $\alpha = 5\%$ que la proportion de clients favorables au changement dans le département du Rhône n'est pas significativement supérieure à 75%.

b) Les clients du département de la Savoie sont plus favorables au changement horaire que ceux du département du Rhône :

Conditions d'application du test : Échantillons indépendants provenant de deux populations normales - Grands échantillons $n_R = 400$ et $n_S = 300$.

La statistique de test : $\frac{(\widehat{P}_R - \widehat{P}_S) - (p_R - p_S)}{\sqrt{\frac{\widehat{p}_R \widehat{q}_R}{n_R} + \frac{\widehat{p}_S \widehat{q}_S}{n_S}}} \to N(0\ ,\ 1)$

Echantillon Rhône : $n_R = 400$; $p_R = \frac{304}{400} = 76\%$
Echantillon Savoie : $n_S = 300$; $p_S = \frac{237}{300} = 79\%$

Hypothèses statistiques : Test unilatéral risque à gauche

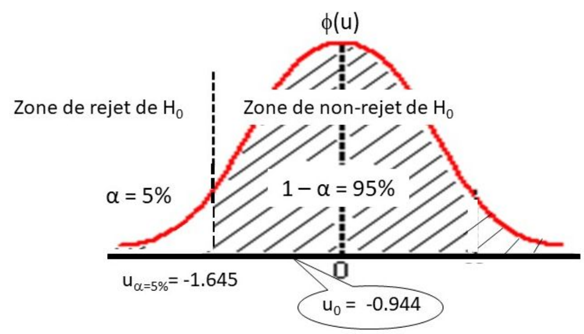
$$\begin{cases} \mathrm{H}_0 : p_R \geq p_S \\ \mathrm{H}_1 : p_R < p_S \Leftrightarrow p_R - p_S < 0 \end{cases}$$

Ecart observé : $\widehat{p}_R - \widehat{p}_S = 0.76 - 0.79 = -3\%$

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

Statistique de test sous H_0 : $p_R = p_S$

$$u_0 = \frac{(\widehat{p}_R - \widehat{p}_S) - 0}{\sqrt{\frac{\widehat{p}_R \widehat{q}_R}{n_R} + \frac{\widehat{p}_S \widehat{q}_S}{n_S}}} = \frac{-0.03}{\sqrt{\frac{0.76 \times 0.24}{400} + \frac{0.79 \times 0.21}{300}}} = -0.944$$



Règle de décision et conclusion : La valeur de la statistique de test $u_0 = -0.944 > -1.645$ est dans la zone de Non-Rejet de H_0. On peut donc conclure avec un risque d'erreur $\alpha = 5\%$ qu'il n'y a pas de différence significative, les clients du département de Savoie ne sont pas plus favorables au changement que ceux du département du Rhône.

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

$\widehat{p} = \frac{760}{1000} = 76\%$: Proportion estimée de clients favorables au changement d'horaires dans les trois départements de la région Auvergne-Rhône-Alpes.

- La statistique de test : $\frac{\widehat{P} - p}{\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}} \hookrightarrow N(0, 1)$
- Risque d'erreur : $\alpha = 5\% \Rightarrow u_{\frac{\alpha}{2} = 2,5\%} = 1.96$ cf. annexe table N(0 , 1)
- Marge d'erreur associée à l'estimation de la proportion:
  $E = u_{\frac{\alpha}{2}} \times \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$ =E = $1.96 \times \sqrt{\frac{0.76 \times 0.24}{1000}} = 2.65\%$
- I.C. de niveau 95% de p : $p \in [73.35\%\ ;\ 78.65\%]$.

d) Test non paramétrique du Khi-deux d'homogénéité :

Hypothèses statistiques :

$$\begin{cases} \text{H}_0 : \text{la préférence en faveur d'un changement horaire est identique dans les trois départements} \\ \text{H}_1 : \text{la préférence en faveur dun changement horaire est différente dans les trois départements} \end{cases}$$

Changement d'horaire	Rhône	Savoie	Cantal	Total
Favorable	304 (304)	237 (228)	219 (228)	760
Défavorable	96 (96)	63 (72)	81 (72)	238
Total	400	300	300	1000

Tableau des Effectifs Observés (Théoriques) :

Le nombre de degrés de liberté : $\nu = (k-1)(r-1) = 2$ d.d.l. avec k = 3 départements et r = 2 modalités de la variable changement d'horaires. Avec un risque $\alpha = 1\%$, le $\chi^2_{(5\%\,;\,2d.d.l.)} = 5.99$ cf. annexe table du khi-deux.

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

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 $\chi^2_{calculé} = 2{,}99 < \chi^2_{(5\%;2)} = 5.99$. 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.

e) Test d'ajustement du Khi-deux :

Changement d'horaire	Rhône	Savoie	Cantal	Total
Favorable Oi	304	237	219	760
Distribution Uniforme	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1
Ti	253.33	253.33	253.33	760

Hypothèses statistiques - Test d'Ajustement du khi-deux à une distribution donnée :

Hypothèses statistiques :

$$\begin{cases} \text{H}_0 : \text{les clients favorables au changement d'horaire se répartissent de façon uniforme sur les 3 départements} \\ \text{H}_1 : \text{les clients favorables au changement d'horaire ne se répartissent pas de façon uniforme sur les 3 départements} \end{cases}$$

Variable qualitative à $p = 3$ modalités. Le nombre de degrés de liberté pour le test d'ajustement du khi-deux : $\nu = p - 1 - r = 2$ degrés de liberté, avec $r = 0$, aucun paramètre estimé.

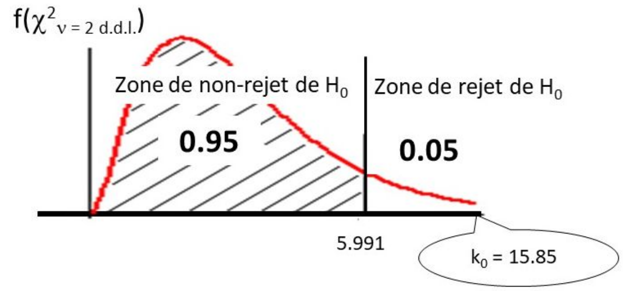
Statistique de test sous l'hypothèse nulle H_0 : $\Sigma_{i=1}^{p=3} \frac{(o_i - t_i)^2}{t_i} \rightarrow \chi^2_{2\,d.d.l.}$

Valeur de la statistique de test sous l'hypothèse nulle H_0 : Ajustement uniforme :

Pour $\alpha = 5\%$, $\chi^2_{5\%\,,\,2\,d.d.l.} = 5.991$ cf. table du khi-deux à 2 d.d.l.

Conclusion : puisque $k_0 = 15.84 > \chi^2_{5\%} = 5.991$ alors Rejet de l'hypothèse nulle H_0. Au seuil de signification $\alpha = 5\%$, on peut donc conclure que les clients favorables au changement d'horaire ne se répartissent pas de façon uniforme sur les 3 départements.

$k_0 = \frac{(304-253.33)^2}{253.33} + \frac{(237-253.33)^2}{253.33} + \frac{(219-253.33)^2}{253.33} = 15.84$



*************** °°° ***************

Exercice 2 : **(Barème de notation : a) 1.5 pt b) 1.5 pt c) 1.5 pt d) 1.5 pt e) 1.5 pt f) 2 pts g) 2 pts = 11.5 pts)**

Races bovines	Monbéliarde	Normande
Nombre de bêtes	$n_M = 75$	$n_N = 25$
Production journalière moyenne observée par vache	$\bar{x}_M = 21$	$\bar{x}_N = 25$
SCE : Somme des Carrés des Ecarts à la moyenne	$SCE_M = 1550$	$SCE_N = 900$

a) Production journalière moyenne des vache Normandes est inférieure à 28 litres :

Test paramétrique - Comparaison d'une moyenne à une valeur donnée :

Hypothèses statistiques : Test unilatéral risque à gauche .

$$\begin{cases} H_0 : m_N \geq 28 \text{ ppm} \\ H_1 : m_N < 28 \text{ ppm} \end{cases}$$

Seuil de signification et conditions d'application du test : $\alpha = 5\%$, petit échantillon $n_N = 25$, la variance σ^2 du rendement journalier est inconnue. $s_N^{*2} = \frac{SCE_N}{n_N-1} = \frac{900}{24} = 37.5 = 6.124^2$

Statistique de test : $\frac{\overline{X}_n - m_N}{\frac{s_N^*}{\sqrt{n_N}}} \to T_{n_N-1=24d.d.l.}$

Statistique de test sous H_0 : $m = 28$ $\quad t_0 = \frac{\overline{X}_n - m_N}{\frac{s_N^*}{\sqrt{n_N}}} = \frac{(25-28)}{\frac{6.124}{5}} = -2.45$

Conclusion : fractile de la loi Student $T_{24d.d.l.}$ (cf.table): $t_0 = -2.45 < t_{5\%} = -1.711$. t_0 appartient à la zone de Rejet de H_0, on peut donc conclure avec risque d'erreur $\alpha = 5\%$ qu'il y a une différence significative. Le rendement journalier moyen d'une vache Normande est significativement inférieur à 28 litres/jour.

b) La variance de la production journalière des vaches Normandes est supérieure à 35 litres^2 :

Comparaison d'une variance à une valeur donnée :

Test unilatéral risque à droite

Données : $n_N = 25$, $s_N^{*2} = \frac{SCE_N}{n_N-1} = \frac{900}{24} = 37.5 = 6.124^2$, $1 - \alpha = 0.95 \Rightarrow \alpha = 0.05$

Hypothèses statistiques

$$\begin{cases} H_0 : \sigma_N^2 \leq 35 \\ H_1 : \sigma_N^2 > 35 \end{cases}$$

La statistique de test : $\frac{(n_N-1)S_N^{*2}}{\sigma_N^2} \rightarrow \chi^2_{(n_N-1=24\, d.d.l.)}$

Valeurs tabulées du khi-deux à 24 degrés de liberté (cf. table du khi-deux) :

$$k_{\alpha=5\%} = \chi^2_{0.95\,;\,24} = 36.415$$

Statistique de test sous l'hypothèse nulle H_0 : $\sigma_N^2 = 35$

$$k_0 = \frac{(n_N-1)s_N^{*2}}{\sigma_N^2} = \frac{24\times 37.5}{35} = 25.71$$

Conclusion : vu que $k_0 = 25.77 < k_{\alpha=5\%} = 36.415$, la valeur k_0 appartient à la zone de Non-Rejet de H_0. On peut donc conclure avec un risque d'erreur $\alpha = 5\%$, que la variance du rendement journalier des vaches Normandes n'est pas significativement supérieure à 35.

c) Les variances des productions journalières moyennes de lait de ces deux races sont différentes

Test de comparaison de variances des productions de lait des deux races bovines :

$n_M = 75$; $n_N = 25$; $SCE_M = 1480$; $SCE_N = 970$

$\Rightarrow s_M^{*2} = \frac{SCE}{n_M-1} = \frac{1550}{74} = 20.95$; $s_N^{*2} = \frac{SCE}{n_N-1} = \frac{900}{24} = 37.50.$

Conditions d'application du test : $n_M = 75$ Grand échantillon, $n_N = 25$ petit échantillon provenant d'une population normale de variances inconnues σ_M^2 et σ_N^2.

La statistique de test : $\frac{\sigma_N^2}{\sigma_M^2} \times \frac{s_M^{*2}}{s_N^{*2}} \rightarrow F_{(n_M-1=74\,;\,n_N-1=24)}$

Hypothèses statistiques - Test bilatéral symétrique

$$\begin{cases} \mathrm{H}_0 : \frac{\sigma_N^2}{\sigma_M^2} = 1 \Leftrightarrow \sigma_M^2 = \sigma_N^2 \text{ égalité des variances.} \\ \mathrm{H}_1 : \frac{\sigma_N^2}{\sigma_M^2} \neq 1 \Leftrightarrow \sigma_M^2 \neq \sigma_N^2. \end{cases}$$

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

Les valeurs critiques - cf. Table de Fisher à $\nu_1 = 74$ et $\nu_2 = 24$ degrés de liberté :

$f_2 = f_{2,5\%\,,\,74\,,\,24} = 2.054$ et $f_1 = f_{97.5\%\,,\,74\,,\,24} = \frac{1}{f_{2.5\%\,,\,24\,,\,74}} = \frac{1}{1.835} = 0.545$

Statistique de test sous H_0 : $\sigma_M^2 = \sigma_N^2$

$$f_0 = \frac{\sigma_N^2 s_M^{*2}}{\sigma_M^2 s_N^{*2}} = \frac{s_M^{*2}}{s_N^{*2}} = \frac{24\times SCE_M}{74\times SCE_N} = \frac{24\times 1550}{74\times 900} = 0.559$$

<u>Conclusion</u> : La valeur $f_0 = 0.559 \in [0.545; 2.054]$ *appartient à la zone de non-rejet de* H_0*. On peut donc conclure, au seuil de signification* $\alpha = 5\%$*, qu'il n'y a pas de différence significative entre les variances des productions journalières de lait de ces deux races bovines. On peut alors les supposer comme égales* ($\sigma_M^2 \approx \sigma_N^2$).

d) La vache normande produit plus de lait par jour que la vache Monbéliarde :

Comparaison de moyennes :

Conditions d'application du test : $n_M = 75$ Grand échantillon, $n_N = 25$ petit échantillon provenant d'une population normale de variances inconnues σ_M^2 et σ_N^2.

Hypothèses statistiques - Test unilatéral risque à gauche

$$\begin{cases} \text{H}_0 : m_M - m_N \geq 0 \Leftrightarrow m_M \geq m_N. \\ \text{H}_1 : m_M - m_N < 0 \Leftrightarrow m_M < m_N. \end{cases}$$

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

La statistique de test : $\frac{(\overline{X}_M - \overline{X}_N) - (m_M - m_N)}{s^* \sqrt{\frac{1}{n_M} + \frac{1}{n_N}}} \rightarrow T_{\nu = n_M + n_N - 2 = 98 d.d.l.}$

Ecart observé : $(\bar{x}_M - \bar{x}_N) = 21 - 25 = -4$ litres.

$s^{*2} = \frac{SCE_M + SCE_N}{n_M + n_H - 2} = \frac{2450}{98} = 25 = 5^2$.

$\alpha = 5\% \Rightarrow t_\alpha = 1.661$ cf. table de Student.

Statistique de test sous H_0 : $m_M = m_N$

$t_0 = \frac{(m_M - m_N) - 0}{s^* \sqrt{\frac{1}{n_M} + \frac{1}{n_N}}} = \frac{-4}{5 \times \sqrt{\frac{1}{75} + \frac{1}{25}}} = -3.464$ litres.

<u>*Conclusion*</u> *: la valeur* $t_0 = -3.464 < t_\alpha = -1.661$*, appartient à la zone de rejet de* H_0*. On peut donc conclure, au seuil de signification* $\alpha = 5\%$*, qu'il y a une différence significative de rendement entre les deux races, la différence observée n'est pas attribuable au hasard de l'échantillonnage. La production journalière de lait des vaches Normandes sont significativement plus faible que celle des Monbéliardes.*

e) Risque d'erreur α^* associé à cet intervalle de confiance : $(m_M - m_N) \in [-6.197\ ;\ -1.803]$:

Différence observée : $(\bar{x}_M - \bar{x}_N) = -4$ l/j.

Marge d'erreur de l'intervalle $\Rightarrow E = \frac{(-1.803 + 6.197)}{2} = 2.197$.

Fractile de la loi de Student $t_{\frac{\alpha}{2}}$: $E = t_{\frac{\alpha}{2}} s^* \sqrt{\frac{1}{n_M} + \frac{1}{n_N}} \Rightarrow t_{\frac{\alpha}{2}} = \frac{E}{s^* \sqrt{\frac{1}{n_M} + \frac{1}{n_N}}} = \frac{2.197}{5\sqrt{\frac{1}{75} + \frac{1}{25}}} = 1.903$

$F(1.903) = 0.97 \Rightarrow \frac{\alpha^*}{2} = 0.03 \Rightarrow \alpha^* = 6\% \Rightarrow 1 - \alpha^* = 94\%$ le niveau de confiance.

f) Comparaison des distributions des productions journalières

Test non paramétrique de 2 échantillons indépendants : tests de Wilcoxon ou test U de Mann-Withney.

Hypothèses statistiques : test bilatéral

$$\begin{cases} \text{H}_0 : \text{Les distributions des productions journalières sont identiques} \\ \text{H}_1 : \text{Les distributions des productions journalières sont différentes} \end{cases}$$

Test de Wilcoxon :

Statistique de test de Wilcoxon sous H_0, $m = 10$: $W_M = \sum_{i=1}^{m=10} R_i = 73$

Rg M	15	5,5	2	8	11,5	4	8	2	15	2	$\sum_{i=1}^{m=10} RgM = 73$
Vache	1	2	3	4	5	6	7	8	9	10	
Monbéliarde	26	23	21	24	25	22	24	21	26	21	
Normande	28	25	24	25	23	26	27	25	28	28	
Rg N	18	11,5	8	11,5	5,5	15	17	11,5	18	18	$\sum_{j=1}^{n=10} RgM = 134$

Conclusion : vu que $W_M = 73 > w^* = 8$ cf. table de Wilcoxon pour des échantillons indépendants, on ne rejette pas l'hypothèse nulle H_0. On ne peut donc pas conclure avec un risque d'erreur $\alpha = 5\%$, que les distributions des productions journalières sont différentes.

g) Test non paramétrique de Wilcoxon - Echantillons dépendants :

Rg Eté	8	7	2,5	2,5	5	2,5	6	2,5
Vache	1	2	3	4	5	6	7	8
Eté	26	25	21	21	22	21	24	21
Hiver	28	25	24	25	23	26	27	25
Rg Hiver	8	4	2	4	1	6	7	4
Différence des Rgs	0	3	0,5	-1,5	4	-3,5	-1	-1,5
Rg en V.Absolue		5	1	3,5	7	6	2	3,5

Hypothèses statistiques :

$$\begin{cases} H_0 : & \text{les deux distributions Hiver et Eté sont identiques} \\ H_1 : & \text{les deux distributions Hiver et Eté sont différentes} \end{cases}$$

Remarque : La différence d_i des productions de la vache 1 est nulle, cette observation est supprimée, il ne reste donc qu'un échantillon de taille $n = 7 vaches$.

$T^+ = 13$: somme des rangs des différences positives $d_i > 0$,

$T^- = 15$: somme des rangs des différences négatives $d_i < 0$,

Valeur critique pour $n = 7$ et $\alpha = 5\%$, test bilatéral : $T_\alpha = 2$ (cf. table Wilcoxon échantillons appariés.)

$T = min(T^+, T^-) = min(13; 15) = 13 > T_\alpha = 2 \Rightarrow$ Non-Rejet de H_0. On peut donc considérer avec un risque d'erreur $\alpha = 5\%$, que les productions journalières sont identiques.

*************** °°° ***************