L3 - Economie & Gestion

Statistique Inférentielle - Corrigé RSE-CC3
Durée 1h30 - Année universitaire 2024 - 2025

Exercice 1 : **(Barème de notation : a) 2 pts b) 2 pts c) 2 pts d) 2 pts e) 2 pts = 10 pts)**

Statut vaccinal	Non-vacciné	Vacciné
Nombre de personnes testées positives	225	50
Nombre de personnes testées	300	200

a) La proportion de personnes non-vaccinées testées positives est supérieure à 70% :

Conditions d'application du test : grande taille d'échantillon $n_{NV} = 300$.

La statistique de test : $\frac{\widehat{P}_{NV} - p_{NV}}{\sqrt{\frac{p_{NV} q_{NV}}{n_{NV}}}} \rightarrow N(0\ ,\ 1)$

Echantillon Rhône : $n_{NV} = 300$; $\widehat{p}_{NV} = \frac{225}{300} = 75.00\%$

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 :

$$\begin{cases} H_0 : p_{NV} \leq p_0 = 70\% \\ H_1 : p_{NV} > p_0 = 70\% \text{ la proportion est supérieure à } 70\% \end{cases}$$

Statistique de test sous l'hypothèse nulle $H_0 : p_{NV} = p_0 = 70\%$:

$$u_0 = \frac{\widehat{p}_{NV} - p_0}{\sqrt{\frac{p_0 q_0}{n_{NV}}}} = \frac{0.75 - 0.70}{\sqrt{\frac{0.70 \times 0.30}{300}}} = 1.89$$

Conclusion : la valeur de la statistique de test $u_0 = 1.89 > u_{5\%} = 1.645$ est dans la zone de rejet de H_0. On peut donc conclure avec un risque d'erreur $\alpha = 5\%$ que la proportion de personnes non-vaccinées testées positives est supérieure à 70%.

b) Intervalle de confiance :

Conditions d'application du test : grand échantillon aléatoire de taille $n_{NV} = 300$.

Estimation de la proportion de personnes non-vaccinées testées positives : $\widehat{p}_{NV} = \frac{225}{300} = 75\%$.

La statistique de test : $\frac{\widehat{P}_{NV} - P}{\sqrt{\frac{p_{NV} q_{NV}}{n_{NV}}}} \rightarrow N(0, 1)$

Marge d'erreur : $E = u_{\frac{\alpha}{2}} \sqrt{\frac{\widehat{P}_{NV} \widehat{q}_{NV}}{n_{NV}}} = 1.96 \sqrt{\frac{0.75 \times 0.25}{300}} = 4.90\%$.

Fractile de la loi normale : $\alpha = 5\%$; $u_{\frac{\alpha}{2}=2.5\%} = \pm 1.96$, cf. table N(0 , 1).

Intervalle de confiance de niveau $1-\alpha = 95\%$ de $P_{NV} \in [70.10\% \; ; \; 79.90\%]$

c) Niveau de confiance : $p_{NV} \in [69.86\% \; ; \; 80.14\%]$:

Grand échantillon $n_{NV} = 300$.

Marge d'erreur : $E = u_{\frac{\alpha}{2}} \sqrt{\frac{p_{NV} q_{NV}}{n_{NV}}} = \frac{(0.8014-0.6986)}{2} = 0.0514$

$$\Rightarrow u_{\frac{\alpha}{2}} = \frac{E}{\sqrt{\frac{n_{NV}}{p_{NV} q_{NV}}}} = \frac{0.0514}{\sqrt{\frac{300}{0.75\times 0.25}}} = 2.055$$

cf. table N(0 , 1) : $P(U \leq 2.055) = \Phi(2.055) = 98\% = 1-\frac{\alpha}{2} \Rightarrow \frac{\alpha}{2} = 2\%$

$$\Rightarrow \alpha = 4\% \Rightarrow 1-\alpha = 96\%.$$

On attribue à cet intervalle un niveau de confiance de $1-\alpha = 96\%$.

d) Taille minimale de l'échantillon de personnes non-vaccinées à tester :

Niveau de confiance : $1-\alpha = 95\%$

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

Estimation ponctuelle de la proportion de personnes non-vaccinées testées positives : $\widehat{p} = 75\%$ avec $\widehat{q} = 25\%$,

Marge d'erreur : $E^* = u_{\frac{\alpha}{2}} \sqrt{\frac{\widehat{p}\widehat{q}}{n^*}} \leq 0.0245$

$$\Rightarrow n^* \geq \left(\frac{u_{\frac{\alpha}{2}}}{E}\right)^2 \widehat{p}\widehat{q} = \left(\frac{1.96}{0.0245}\right)^2 \times 0.75 \times 0.25 = 1200$$

Il faudrait tester $n^* \geq 1200$ personnes non-vaccinées.

Propriété : réduire de moitié la précision : $E^* = \frac{E}{2} = 0.0245 \Rightarrow n^* = 2^2 n = 4 \times 300 = 1200$

e) Comparaison de deux proportions :

Hypothèses statistiques : Comparaison de proportions - Test unilatéral risque à droite :

$$\begin{cases} H_0 : p_{NV} \leq p_V \Leftrightarrow p_{NV} - p_V \leq 0 \\ H_1 : p_{NV} > p_V \Leftrightarrow p_{NV} - p_V > 0 \end{cases}$$

Conditions d'application du test : échantillons indépendants provenant de deux populations normales - Grands échantillons $n_{NV} = 300$ et $n_V = 200$.

La statistique de test : $\frac{(\widehat{P}_{NV} - \widehat{P}_V) - (p_{NV} - p_A)}{\sqrt{\frac{\widehat{p}_{NV}\widehat{q}_{NV}}{n_{NV}} + \frac{\widehat{p}_V \widehat{q}_V}{n_V}}} \rightarrow N(0\,,\,1)$

Echantillon des non-vaccinés : $n_{NV} = 300$; $\widehat{p}_{NV} = \frac{225}{300} = 75\%$

Echantillon des vaccinés : $n_V = 200$; $\widehat{p}_V = \frac{50}{200} = 25\%$

Ecart observé : $\widehat{p}_{NV} - \widehat{p}_V = 75\% - 25\% = 50\%$

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

Statistique de test sous H_0 : $p_{NV} = p_V$

$$u_0 = \frac{(\widehat{p}_{NV} - \widehat{p}_V) - 0}{\sqrt{\frac{\widehat{p}_{NV}\widehat{q}_{NV}}{n_{NV}} + \frac{\widehat{p}_V\widehat{q}_V}{n_V}}} = \frac{0.50}{\sqrt{\frac{0.75\times 0.25}{300} + \frac{0.25\times 0.75}{200}}} = 12.65$$

Règle de décision et conclusion :

La valeur de la statistique de test u_0, $u_0 = 12.65 > 1.645$, est dans la zone de rejet de H_0. On peut donc considérer avec un risque d'erreur $\alpha = 5\%$ que la proportion de personnes non-vaccinées testées positives est significativement supérieure à celle des personnes vaccinés positives.

*************** °°° ***************

Exercice 2 : **(Barème de notation : a) 2 pts b) 2 pts c) 2 pts d) 2 pts e) 2 pts = 10 pts)**

Type	Delta	Omicron
Nombre de patients positifs	50	25
Durée moyenne de séjour observé (jours)	13.6	9.9
Variance observée de la durée de séjour	2.64	1.36

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

Hypothèses statistiques : Test unilatéral risque à gauche .

$$\begin{cases} H_0 : m_O \geq 10 \text{ jours} \\ H_1 : m_O < 10 \text{ jours} \end{cases}$$

Seuil de signification et conditions d'application du test : $\alpha = 5\%$, petit échantillon $n_O = 25$, la variance σ^2 de la durée de séjour en réanimation des personnes positives au variant Omicron est inconnue.

Statistique de test : $\frac{\overline{X}_{n_O} - m_O}{\frac{s^*_O}{\sqrt{n_O}}} \to T_{n_O - 1 = 24 d.d.l.}$

Statistique de test sous H_0 : $m = 10$ $\quad t_0 = \frac{\overline{X}_{n_O} - m_O}{\frac{s^*_O}{\sqrt{n_O}}} = \frac{\overline{X}_{n_O} - m_O}{\frac{s_O}{\sqrt{n_O - 1}}} = \frac{(9.9 - 10)}{\sqrt{\frac{1.36}{24}}} = -0.42$

Conclusion : fractile de la loi Student $T_{24d.d.l.}$ (cf.table): $t_{5\%} = -1.711, t_0$ appartient à la zone de non-rejet de H_0 ($t_0 = -0.420 > t_{5\%} = -1.711$), on peut donc conclure avec risque d'erreur $\alpha = 5\%$ qu'il n'y a pas de différence significative. La durée moyenne de séjour en réanimation des personnes positives au variant Omicron n'est pas significativement inférieur à 10 jours.

b) Comparaison d'une variance à une valeur donnée :

Test unilatéral risque à droite

Données : $n_O = 25$, $s^2_O = 1.36$, $1 - \alpha = 0.95 \Rightarrow \alpha = 0.05$

Hypothèses statistiques

$$\begin{cases} H_0 : \sigma^2_O \leq 1 \\ H_1 : \sigma^2_O > 1 \end{cases}$$

La statistique de test : $\frac{(n_O - 1)S^{*2}_O}{\sigma^2_O} \to \chi^2_{(n_O - 1 = 24\ d.d.l.)}$

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 : $\sigma_O^2 = 1$

$$k_0 = \frac{(n_O-1)s_O^{*2}}{\sigma_O^2} = \frac{n_O s_O^2}{\sigma_O^2} = \frac{25\times 1.36}{1} = 34$$

Conclusion : vu que $k_0 = 34 < k_{\alpha=5\%} = 36.415$, 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 de la durée de séjour en réanimation des personnes positives au variant Omicron n'est pas significativement supérieure à 1 jour.

c) Test paramétrique - Comparaison de 2 variances :

Seuil de signification et conditions d'application du test : $\alpha = 5\%$, les variances σ_D^2 et σ_O^2 sont inconnues, la taille de l'échantillon variant Delta est grande $(n_D = 50$ et la taille de l'échantillon variant Omicron est petite $n_O = 25)$.

Étude de la variabilité des durées de séjour en réanimation - Comparaison de variances :

Conditions d'application du test : petits échantillons indépendants provenant de deux populations normales de variances inconnues σ_D^2 et σ_O^2.

La statistique de test : $\frac{\sigma_O^2 s_D^{*2}}{\sigma_D^2 s_O^{*2}} \rightarrow F_{(n_D-1=49\,;\,n_O-1=24)}$

Solution 1 - Test d'hypothèses

Hypothèses statistiques - Test bilatéral symétrique

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

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

Statistique de test sous l'hypothèse nulle H_0 d'égalité des variances $\frac{\sigma_N^2}{\sigma_A^2} = 1$:

$$f_0 = \frac{s_D^{*2}}{s_O^{*2}} = \frac{50\times 24\times 2.64}{49\times 25\times 1.36} = \frac{48\times 2.64}{49\times 1.36} = 1.902$$

Les valeurs critiques (cf. table de Fisher) :

$f_2 = f_{(2.5\%\,,\,49\,,\,24)} = 1.94$; $f_1 = f_{(97.5\%\,,\,49\,,\,24)} = \frac{1}{f_{(2.5\%\,,\,24\,,\,49)}} = \frac{1}{2.11} = 0.4739.$

Conclusion : $f_1 = 0.4739 \leq f_0 = 1.902 \leq f_2 = 1.94$ on ne rejette pas l'hypothèse H_0. On peut donc conclure, avec un seuil de signification $\alpha = 5\%$, qu'il n'y a pas de différence significative entre les variances des durées de séjour en réanimation ; on peut les supposer comme égales $\sigma_D^2 \approx \sigma_O^2$.

Solution 2 - Intervalle de confiance

Les valeurs critiques (cf. table de Fisher) :

$f_2 = f_{(2.5\% \,, 49 \,, 24)} = 1.94 \quad ; \quad f_1 = f_{(97.5\% \,, 49 \,, 24)} = \frac{1}{f_{(2.5\% \,, 24 \,, 49)}} = \frac{1}{2.11} = 0.4739.$

$$f_1 \leq \frac{\sigma_O^2 s_D^{*2}}{\sigma_D^2 s_O^{*2}} \leq f_2 \;\Rightarrow f_1 \frac{s_O^{*2}}{s_D^{*2}} \leq \frac{\sigma_O^2}{\sigma_D^2} \leq f_2 \frac{s_O^{*2}}{s_D^{*2}} \;\Rightarrow 0.4739 \frac{1.36}{2.64}\frac{49}{48} \leq \frac{\sigma_O^2}{\sigma_D^2} \leq 1.94 \frac{1.36}{2.64}\frac{49}{48}$$

$$\Rightarrow 0.2492 \leq \frac{\sigma_O^2}{\sigma_D^2} \leq 1.0202$$

$I.C._{1-\alpha=95\%} \; \frac{\sigma_O^2}{\sigma_D^2} \in [0.2492 \,;\, 1.0202]$

Conclusion : $1 \in [0.2492 \,;\, 1.0202]$, non-rejet de l'hypothèse nulle H_0. On peut donc conclure, avec un seuil de signification $\alpha = 5\%$, qu'il n'y a pas de différence significative entre les variances des durées de séjour ; on peut les supposer comme égales $\sigma_D^2 \approx \sigma_O^2$.

d) Test paramétrique - Comparaison de 2 moyennes :

Hypothèses statistiques : unilatéral risque à droite

$$\begin{cases} H_0 : m_D \leq m_O \\ H_1 : m_D > m_O \quad \text{la durée moyenne de séjour pour le variant Delta est plus longue que celle du variant Omicron : } m_O < m_D \end{cases}$$

Seuil de signification et conditions d'application du test : $\alpha = 5\%$, les variances σ_D^2 et σ_O^2 sont inconnues, la taille de l'échantillon variant Delta est grande ($n_D = 50$ et la taille de l'échantillon variant Omicron est petite $n_O = 25$).

Selon la question c), les variances des taux des deux procédés étant égales $\sigma_D^2 \approx \sigma_O^2$, on choisit donc la statistique de test suivante :

Statistique de test : $\quad \frac{(\overline{X}_D - \overline{X}_O) - (m_D - m_O)}{s* \sqrt{\frac{1}{n_D} + \frac{2}{n_O}}} \rightarrow T_{(n_D + n_O - 2 = 73 \; d.d.l.)}$

Estimation ponctuelle de la variance commune de $\sigma^2 = \sigma_D^2 = \sigma_O^2$:

$$s^{*2} = \frac{n_D \times s_D^2 + n_O \times s_O^2}{n_D + n_O - 2} = \frac{(50 \times 2.64 + 25 \times 1.36)}{73} = 2.274 = 1.51^2$$

Calcul de la statistique de test sous $H_0 : m_D = m_O$

$$t_0 = \frac{(13.6 - 9.9) - 0}{1.51 \sqrt{\frac{1}{50} + \frac{1}{25}}} = \frac{3.70}{1.51 \sqrt{\frac{3}{50}}} = \frac{0.7}{0.3699} = 10.002$$

Fractile de la loi de Student : $t_{5\%} = 1.666$, cf. table de Student.

Conclusion du test : $t_0 = 10.002 > t_{5\%} = 1.666$ appartient à la zone de rejet de H_0. On peut donc conclure avec un risque d'erreur $\alpha = 5\%$, qu'il y a une différence significative ; les patients atteints du variant Omicron ont des durées de séjour à l'hôpital moins longues que celles des patients atteints du variant Delta, l'affirmation du chef de service est donc vraie.

e) Test d'indépendance du khi-deux :

Tableau des effectifs Observés & (**théoriques**)

Hypothèses statistiques : Test unilatéral risque à droite

	Statut vaccinal		
Genre	Delta	Omicron	Total
Femme	35 (**31.33**))	12 (**15.67**)	47
Homme	15 (**18.67**)	13 (**9.33**)	28
Total	50	25	75

$$\begin{cases} \text{H}_0 : \text{ indépendance entre le genre et le variant Covid-19} \\ \text{H}_1 : \text{ dépendance entre le genre et le le variant Covid-19} \end{cases}$$

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

Nombre de degrés de liberté pour le test d'indépendance khi-deux :

$$\nu = (L-1)(c-1) = (2-1)(2-1) = 1 \text{ d.d.l.}$$

Valeur critique : $\chi^2_{\alpha=5\%\ ,\ 1\ d.d.l.} = 3.841$ cf. table du khi-deux à 1 d.d.l.

Statistique de test sous l'hypothèse nulle H_0 d'indépendance :

$$k_0 = \frac{(35-31.33)^2}{31.33} + \frac{(12-15.67)^2}{15.67} + \frac{(15-18.67)^2}{18.67} + \frac{(13-9.33)^2}{9.33} = 3.45$$

Règle de décision et conclusion : Non-rejet de l'hypothèse nulle H_0 d'indépendance puisque $k_0 = 3.45 < \chi^2_{5\%} = 3.841$, il n'y pas de lien significatif au seuil $\alpha = 5\%$, entre le genre et le type de variant Covid-19.

*************** °°° ***************