UFR de Sciences Économiques et de Gestion — Université Lumière Lyon 2 | Université de Lyon

L3 - Economie & Gestion

Statistique Inférentielle - **Corrigé du Contrôle Continu n°2**

Durée 1h30 - Année universitaire 2024 - 2025

Exercice 1 : **(Barème de notation : a) 1.5 pt b) 1.5 pt c) 2 pts d) 2 pts e) 2 pts = 9 pts)**

	Procédé	
	Nouveau	Actuel
Nombre d'essais	$n_N = 50$	$n_A = 50$
Moyenne observée (ppm)	$\bar{x}_N = 6.9$	$\bar{x}_A = 7.6$
Variance observée	$s_N^2 = 1.36$	$s_A^2 = 2.64$

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

Hypothèses statistiques : Test unilatéral risque à gauche .

$$\begin{cases} H_0 : m_N \geq 7 \text{ ppm} \\ H_1 : m_N < 7 \text{ ppm} \end{cases}$$

Seuil de signification et conditions d'application du test : $\alpha = 5\%$, petit échantillon $n_N = 50$, la variance σ^2 du taux de CO est inconnue.

Statistique de test : $\frac{\overline{X}_n - m_N}{\frac{s_N^*}{\sqrt{n_N}}} \to T_{n_N - 1 = 49 d.d.l.}$

Statistique de test sous H_0 : $m = 7$ $\quad t_0 = \frac{\overline{X}_n - m_N}{\frac{s_N^*}{\sqrt{n_N}}} = \frac{\overline{X}_n - m_N}{\frac{s_N}{\sqrt{n_N - 1}}} = \frac{(6.9 - 7)}{\sqrt{\frac{1.36}{49}}} = -0.60$

Conclusion : fractile de la loi Student $T_{49 d.d.l.}$ (cf.table): $t_{5\%} = -1.677, t_0$ appartient à la zone de non-rejet de H_0 ($t_0 = -0.60 > t_{5\%} = -1.677$), on peut donc conclure avec risque d'erreur $\alpha = 5\%$ qu'il n'y a pas de différence significative. Le taux de CO du nouveau procédé n'est pas significativement inférieur à $7 ppm$.

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

Test unilatéral risque à droite

Données : $n_N = 50$, $s_N^2 = 1.36$, $1 - \alpha = 0.95 \Rightarrow \alpha = 0.05$

Hypothèses statistiques

$$\begin{cases} H_0 : \sigma_N^2 \leq 1.2 \\ H_1 : \sigma_N^2 > 1.2 \end{cases}$$

La statistique de test : $\frac{(n_N - 1) S_N^{*2}}{\sigma_N^2} \to \chi^2_{(n_N - 1 = 49 \; d.d.l.)}$

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

$$k_{\alpha = 5\%} = \chi^2_{0.95\,;\,49} = 66.34$$

Statistique de test sous l'hypothèse nulle H_0 : $\sigma_N^2 = 1.2$

$$k_0 = \frac{(n_N-1)s_N^{*2}}{\sigma_N^2} = \frac{n_N s_N^2}{\sigma_N^2} = \frac{50\times 1.36}{1.20} = 56.67$$

Conclusion : vu que $k_0 = 56.67 < k_{\alpha=5\%} = 66.34$, 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 taux de monoxyde de carbone du nouveau procédé n'est pas significativement supérieure à 1.20 ppm^2

c) Test paramétrique - Comparaison de 2 variances :

Étude de la variabilité des taux de CO - Comparaison de variances :

Grands échantillons indépendants de variances σ_A^2 et σ_N^2 inconnues.

La statistique de test : $\frac{\sigma_A^2 s_N^{*2}}{\sigma_N^2 s_A^{*2}} \rightarrow F_{(n_N-1=49\,;\,n_A-1=49)}$

Hypothèses statistiques - Test unilatéral risque à gauche

$$\begin{cases} \mathrm{H}_0 : \frac{\sigma_N^2}{\sigma_A^2} \geq 2 \Leftrightarrow \sigma_N^2 \geq 2\sigma_A^2 . \\ \mathrm{H}_1 : \frac{\sigma_N^2}{\sigma_A^2} < 2 \Leftrightarrow \sigma_N^2 < 2\sigma_A^2 . \end{cases}$$

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

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

$$f_0 = \frac{s_N^{*2}}{s_A^{*2}} = 2\frac{s_N^2}{s_A^2} = 2\frac{1.36}{2.64} = 1.03$$

Les valeurs critiques (cf. table de Fisher) :

$f_2 = f_{(5\%\,,\,49\,,\,49)} = 1.607$; $f_1 = f_{(95\%\,,\,49\,,\,49)} = \frac{1}{f_{(5\%\,,\,49\,,\,49)}} = \frac{1}{1.607} = 0.622.$

Conclusion : la valeur $f_0 = 1.03 > f_1 = 0.622$ appartient à la zone de Non-Rejet de l'hypothèse nulle H_0. On peut donc conclure, avec un seuil de signification $\alpha = 5\%$, que la variance du taux de CO du nouveau procédé n'est pas significativement 2 fois plus petite que celle du taux de CO de l'actuel procédé.

d) Test paramétrique - Comparaison de 2 moyennes :

Hypothèses statistiques : unilatéral risque à gauche

$$\begin{cases} H_0 : m_N \leq m_A \\ H_1 : m_N < m_A \text{ le taux moyen de CO du nouveau procédé est plus efficace} \end{cases}$$

Seuil de signification et conditions d'application du test : $\alpha = 5\%$, les variances σ_A^2 et σ_N^2 sont inconnues, grands échantillons ($n_A = 50$ et $n_N = 50$).

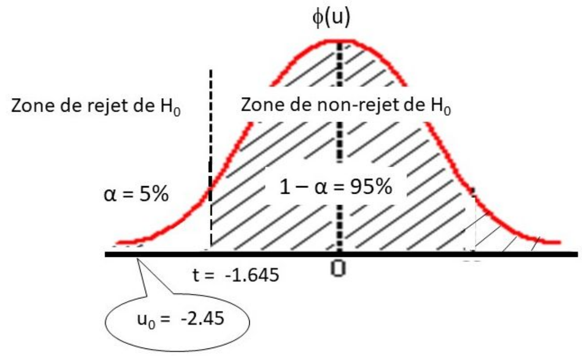
La statistique de test de la différence de deux moyennes :

$$\frac{(\overline{X}_N - \overline{X}_A) - (m_N - m_A)}{\sqrt{\frac{s_N^{*2}}{n_N} + \frac{s_A^{*2}}{n_A}}} \rightarrow N(0\,;\,1)$$

La valeur critique (cf. table N(0 ; 1) avec $\alpha = 5\%$: $u_{5\%} = -1.645$.

Ecart observé des taux de CO : $(\overline{x}_N - \overline{x}_A) = 6.9 - 7.6 = -0.7$.

Statistique de test sous H_0 : $m_A = m_N$ $u_0 = \frac{(\overline{X}_N - \overline{X}_A)-(m_N-m_A)}{\sqrt{\frac{s_N^{*2}}{n_N}+\frac{s_A^{*2}}{n_A}}} = \frac{(6.9-7.6)-0)}{\sqrt{\frac{s_N^2}{49}+\frac{s_A^2}{49}}} = -2.45$



Conclusion du test : $u_0 = -2.45 < t_{5\%} = -1.645$ 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 ; le taux moyen de réduction de monoxyde de carbone dans l'air du nouveau procédé est significativement plus faible que celui de l'actuel procédé ; le nouveau procédé est donc plus efficace.

e) Comparaison de 2 échantillons appariés

Test non paramétrique de Wilcoxon - Echantillons dépendants :

Usine	1	2	3	4	5	6	7	8	9	10
Ancien procédé : X	7.3	6.4	6.8	6.8	7.7	6.8	7.7	6.7	7.7	6.7
Nouveau procédé : Y	7.6	6.8	6.8	6.5	7.8	6.6	7.6	6.7	7.6	6.8
Différence : $d_i = x_i - y_i$	-0.3	-0.4	0	0.3	-0.1	0.2	0.1	0	0.1	-0.1
Rang en valeur absolue $\mid d_i \mid$	6.5	8		6.5	2.5	5	2.5		2.5	2.5

Hypothèses statistiques :

$$\begin{cases} H_0 \text{ :les deux distributions des taux de CO sont identiques} \\ H_1 \text{ :les deux distributions des taux de CO sont différentes} \end{cases}$$

Remarque : Les différences d_i des OC des usines n°3 et n°8 sont nulles, ces observations sont supprimées, il ne reste donc qu'un échantillon de taille $n = 8$ usines.

$T^+ = 16.5$: somme des rangs des différences positives $d_i > 0$,

$T^- = 19.5$: 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.)

$T = min(T^+, T^-) = min(16.5; 19.5) = 16.5 > T_\alpha = 3 \Rightarrow$ Non-Rejet de H_0. On peut donc considérer avec un risque d'erreur $\alpha = 5\%$, que les distributions des taux de CO sont identiques.

*************** °°° ***************

Exercice 2 : (**Barème de notation : a) 2 pts b) 2 pts c) 2 pts d) 2 pts e) 3 pts = 11 pts)**

a) La proportion d'aumobilistes verbalisés dans le Rhône est supérieure à 14.50% :

Organisme	Rhône	Ardèche
Nombre de véhicules verbalisés	54	25
Nombre de véhicules	300	200

Conditions d'application du test : grande taille d'échantillon $n_R = 300$.

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

Echantillon Rhône : $n_R = 300$; $\widehat{p}_R = \frac{54}{300} = 18.00\%$

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

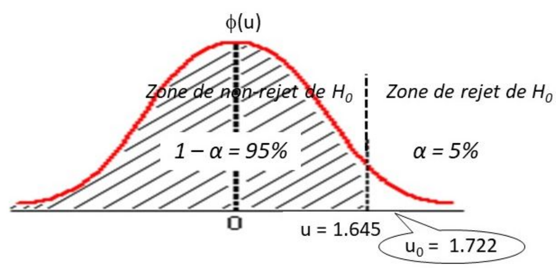
Hypothèses statistiques - Comparaison d'une proportion à une valeur donnée

Test Unilatéral - Risque à droite :

$$\begin{cases} \text{H}_0 : p_R \leq p_0 = 14.50\% \\ \text{H}_1 : p_R > p_0 = 14.50\% \text{ la proportion de véhicules verbalisés est supérieure à } 14.50\% \end{cases}$$

Statistique de test sous $H_0 : p_R = p_0 = 14.5\%$

$$u_0 = \frac{\widehat{p}_R - p_0}{\sqrt{\frac{p_0 q_0}{n_R}}} = \frac{0.18 - 0.145}{\sqrt{\frac{0.145 \times 0.855}{300}}} = 1.722$$



Règle de décision et conclusion :

La valeur de la statistique de test $u_0 = 1.722 > u_{5\%} = 1.645$ est dans la zone de rejet de H_0. On peut donc conclure avec un risque d'erreur $\alpha = 5\%$ que la proportion d'automobilistes verbalisés dans le département du Rhône est supérieure à 14.5%. L'affirmation du gendarme est correcte.

b) Hypothèse selon laquelle on a plus verbalisé d'automobilistes dans le département du Rhône que dans le département de l'Ardèche :

Hypothèses statistiques - Comparaison de moyennes - Test Unilatéral risque à droite

Conditions d'application du test : Échantillons indépendants provenant de deux populations normales - Grands échantillons $n_R = 300$ et $n_A = 200$.

La statistique de test : $\frac{(\widehat{P}_R - \widehat{P}_A) - (p_R - p_A)}{\sqrt{\frac{\widehat{p}_R \widehat{q}_R}{n_R} + \frac{\widehat{p}_A \widehat{q}_A}{n_A}}} \rightarrow N(0\ ,\ 1)$

Echantillon Rhône : $n_R = 300$; $\widehat{p}_R = \frac{54}{300} = 18.00\%$

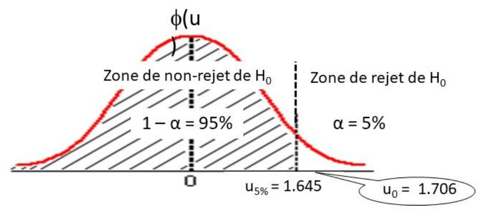
$$\begin{cases} \text{H}_0 : p_R \leq p_A \\ \text{H}_1 : p_R > p_A \;\Leftrightarrow\; p_R - p_A > 0 \end{cases}$$

Ecart observé : $\widehat{p}_R - \widehat{p}_A = 18\% - 12.5\% = 5.50\%$

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

Statistique de test sous H_0 : $p_R = p_A$

$$u_0 = \frac{(\widehat{p}_R - \widehat{p}_A) - 0}{\sqrt{\frac{\widehat{p}_R \widehat{q}_R}{n_R} + \frac{\widehat{p}_A \widehat{q}_A}{n_A}}} = \frac{0.055}{\sqrt{\frac{0.18 \times 0.82}{300} + \frac{0.125 \times 0.875}{200}}} = 1.706$$



Règle de décision et conclusion : La valeur de la statistique de test $u_0 = 1.706 > 1.645$ est dans la zone de rejet de H_0. On peut donc conclure avec un risque d'erreur $\alpha = 5\%$ qu'il y a pas bien eu plus d'automobilistes verbalisés dans le département du Rhône que dans le département de l'Ardèche. L'écart observé de 5.5% est significatif. On a plus verbalisé dans le Rhône que dans l'Ardèche pendant cette journée de pollution.

c)) Test non paramétrique - Echantillons indépendants :

Hypothèses statistiques : test unilatéral

$$\begin{cases} \text{H}_0 : \text{proportions de verbalisation identiques} \\ \text{H}_1 : \text{les proportions de verbalisation des camions sont plus faibles que celles des voitures} \end{cases}$$

Type de véhicule vérbalisé	1	2	3	4	5	6	7	8
Camions m = 8	2%	1,5%	4,3%	6,3%	9,6%	4,3%	10,2%	10,5%
Voitures n = 8	4%	6,3%	8%	5,4%	2,6%	6,3%	10%	9,3%
Rang C	2	1	5,5	9	13	5,5	15	16
Rang V	4	9	11	7	3	9	14	12

On réunit les deux échantillons et on les ordonne par ordre croissant en notant bien de quel groupe provient chaque valeur. On calcule les sommes des rangs des appréciations des diplômés et des non diplômés. On peut présenter les calculs de la manière suivante :

	Wilcoxon	Mann-Whitney
C : m = 8	$W_C = \sum_{i=1}^{m=8} R_i = 67$	$U_C = \sum_{i=1}^{m=8} (R_i - i) = W_C - \frac{m(m+1)}{2} = 31$
V : n = 8	$W_V = \sum_{i=1}^{n=8} R_i = 69$	$U_V = \sum_{i=1}^{n=8} (R_i - i) = W_V - \frac{n(n+1)}{2} = 33$

- Solution 1 - Test de Wilcoxon :

Statistique de test de Wilcoxon sous H_0 pour le plus petit échantillon : $W_C = \sum_{i=1}^{m=8} R_i = 67$

Conclusion : vu que $W_C = 67 > w^* = 3$ cf. table de Wilcoxon pour des échantillons indépendants, on ne rejette pas l'hypothèse nulle H_0. On peut donc conclure avec un risque d'erreur $\alpha = 5\%$, que les camions et les voitures n'ont pas été verbalisés de façon différente.

- Solution 2 - Test U de Mann-Whitney :

Les sommes U_C et U_V des valeurs obtenues sont respectivement 31 et 33. La variable de décision est $U = min(U_C, U_V) = U_C = 31$. Pour des groupes de tailles respectives 8 et 8, la valeur critique (cf. table de Mann-Whitney) est égale à 13.

Conclusion : non-rejet de l'hypothèse H_0 puisque $U_C = 31 > U_{5\%} = 13$. Il n'y a donc pas de différence significative. On peut conclure avec un risque d'erreur de $\alpha = 5\%$, que dans le Rhône, les camions et les voitures n'ont pas été verbalisés de façon différente.

d) Test d'ajustement du Khi-deux :

Type de véhicule	Rhône : Oi	Statistiques Nationales	Ti
Voiture de tourisme	150	45%	135
Utilitaire	60	30%	90
Camion	90	25%	75
Total	300	100%	300

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

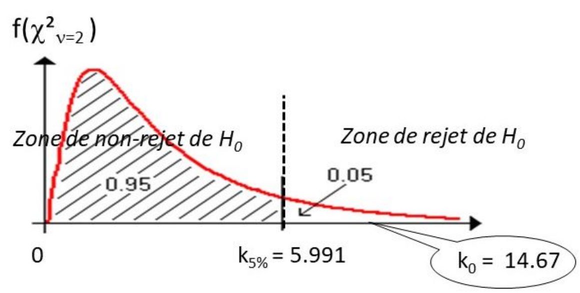
$$\begin{cases} \text{H}_0 : \text{ les verbalisations sont représentatives de celles du niveau national} \\ \text{H}_1 : \text{ les verbalisations ne pas représentatives de celles du niveau national} \end{cases}$$

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 au niveau national :

$$k_0 = \frac{(150-135)^2}{135} + \frac{(60-90)^2}{90} + \frac{(90-75)^2}{75} = 14.67$$



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

Conclusion : puisque $k_0 = 14.67 > \chi^2_{5\%} = 5.991$ alors Rejet de l'hypothèse nulle H_0. Au seuil de signification $\alpha = 5\%$, on peut donc conclure que les verbalisations ne représentent celles du niveau national.

e) Test d'homogénéité du Khi-deux :

Effectifs Observés & (Théoriques)

Type de véhicule	Rhône	Ardèche	Total
Voiture	150 (**156**)	110 (**104**)	260
Utilitaire	60 (**66**)	50 (**44**)	110
Camion	90 (**78**)	40 (**52**)	130
Total	300	200	500

Hypothèses statistiques - Test unilatéral risque à droite :

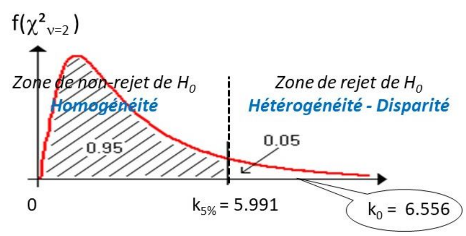
$$\begin{cases} \text{H}_0 : \text{ les verbalisations se répartissent de façon identique selon les 3 types de véhicules : homogénéité} \\ \text{H}_1 : \text{ les verbalisations ne se répartissent pas de façon identique selon les 3 types de véhicules : hétérogénéité} \end{cases}$$

Conditions d'application : $k = 2$ échantillons aléatoires indépendants observés selon un caractère qualitatif Votes à $p = 3$ modalités. Le nombre de degrés de liberté pour le test d'homogénéité du khi-deux : $\nu = (p-1)(k-1) = 2$ degrés de liberté.

Statistique de test sous l'hypothèse nulle H_0 : $\Sigma_{i=1}^{p=3}\Sigma_{j=1}^{k=2}\frac{(o_{ij}-t_{ij})^2}{t_{ij}} \to \chi^2_{2\, d.d.l.}$

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

$$k_0 = \frac{(150-144)^2}{144} + \cdots + \frac{(50-56)^2}{56} = 6.556$$



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

Conclusion : puisque $k_0 = 6.556 > \chi^2_{5\%} = 5.991$ alors Rejet de l'hypothèse nulle H_0 : d'homogénéité. Au seuil de signification $\alpha = 5\%$, on peut donc conclure que les verbalisations ne se répartissent pas de façon identique dans les 2 départements. Il y a donc disparité.

Localisation des disparités :

La proportion de camions verbalisés dans le Rhône est significativement différente (même supérieure $u_0 = 1.645$) que celle dans l'Ardèche.

*************** °°° ***************

	Proportions estimées		Statistique de test sous H_0	Décision
Type de véhicule	Rhône	Ardèche		
Voiture	$\widehat{p}_{VR} = 0.50$	$\widehat{p}_{VA} = 0.55$	$u_0 = -1.099 \in [-1.96; 1.96]$	Non-Rejet H_0
Utilitaire	$\widehat{p}_{UR} = 0.20$	$\widehat{p}_{UA} = 0.25$	$u_0 = -1.304 \in [-1.96; 1.96]$	Non-Rejet H_0
Camion	$\widehat{p}_{CR} = 0.30$	$\widehat{p}_{CA} = 0.20$	$u_0 = 2.582 \notin [-1.96; 1.96]$	Rejet H_0, Disparité