Université de Béjaïa | Département de Sociologie

EMD 2 : Statistiques Inférentielles – Année universitaire 2025/2026

CORRIGÉ TYPE

Statistiques Inférentielles

Durée : 1h30 | Note totale : 20 | Deux exercices de 10 points chacun

---

Exercice 1 – Test du Chi-deux (χ^2) [10 points]

Données : Enquête auprès de $N = 90$ étudiants de première année sciences sociales. Variable de ligne : lieu de résidence (Urbain / Rural) ; variable de colonne : pratique du sport (Oui / Non / Parfois).

Tableau des effectifs observés (O_{ij}) :

Résidence \ Pratique	Oui	Non	Parfois	Total
Urbain	20	12	13	**45**
Rural	16	18	11	**45**
Total	**36**	**30**	**24**	**90**

Question 1 – Calcul des effectifs théoriques E_{ij} [3 points]

La formule générale de l'effectif théorique est :

$$E_{ij} = \frac{L_i \times C_j}{N} = \frac{\texttt{Total ligne } i \times \texttt{Total colonne } j}{\texttt{Effectif total } N}$$

où L_i désigne le total marginal de la ligne i, C_j le total marginal de la colonne j, et $N = 90$ l'effectif total.

Tableau complet des effectifs théoriques :

Case (i, j)	Formule $E_{ij} = (L_i \times C_j)/N$	Calcul	E_{ij}	O_{ij}
Urbain / Oui	$(45 \times 36)/90$	$1620/90$	$18{,}00$	20
Urbain / Non	$(45 \times 30)/90$	$1350/90$	$15{,}00$	12
Urbain / Parfois	$(45 \times 24)/90$	$1080/90$	$12{,}00$	13
Rural / Oui	$(45 \times 36)/90$	$1620/90$	$18{,}00$	16
Rural / Non	$(45 \times 30)/90$	$1350/90$	$15{,}00$	18
Rural / Parfois	$(45 \times 24)/90$	$1080/90$	$12{,}00$	11

Remarque : On vérifie que la somme des effectifs théoriques par ligne et par colonne est identique à celle des effectifs observés – contrôle indispensable de la cohérence des calculs. De plus, chaque $E_{ij} > 5$, condition nécessaire à la validité du test du χ^2.

[3 pts – pour tous les effectifs théoriques correctement calculés]

Question 2 – Calcul du χ^2 calculé [3 points]

La statistique du test est calculée selon la formule :

$$\chi^2_{\texttt{calc}} = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Le calcul se déroule case par case. Pour chaque cellule, on mesure l'écart entre l'effectif observé et l'effectif théorique, on le met au carré (pour rendre tous les écarts positifs), puis on le divise par l'effectif théorique (pour obtenir un écart relatif, proportionnel à la taille attendue) :

Case (i, j)	O_{ij}	E_{ij}	$(O_{ij} - E_{ij})$	$(O_{ij} - E_{ij})^2$	$\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
Urbain / Oui	20	18	$+2$	4	$0{,}222$
Urbain / Non	12	15	-3	9	$0{,}600$
Urbain / Parfois	13	12	$+1$	1	$0{,}083$
Rural / Oui	16	18	-2	4	$0{,}222$
Rural / Non	18	15	$+3$	9	$0{,}600$
Rural / Parfois	11	12	-1	1	$0{,}083$
TOTAL	**90**	**90**	–	–	$1{,}810$

En additionnant les contributions de chaque case, on obtient :

$$\chi^2_{\texttt{calc}} \;=\; 0{,}222 + 0{,}600 + 0{,}083 + 0{,}222 + 0{,}600 + 0{,}083 \;=\; \mathbf{1{,}810}$$

[3 pts – pour le tableau de calcul complet et la valeur finale]

Question 3 – Conclusion : comparaison avec la valeur critique et interprétation [4 points]

Étape préliminaire : détermination du degré de liberté

$$\texttt{ddl} \;=\; (r-1)\times(c-1) \;=\; (2-1)\times(3-1) \;=\; 1\times 2 \;=\; 2$$

avec $r = 2$ lignes (Urbain, Rural) et $c = 3$ colonnes (Oui, Non, Parfois).

Paramètres du test :

- Degré de liberté : ddl $= 2$ (donnée du sujet)
- Seuil de signification retenu : $\alpha = 5\,\%$ (soit $0{,}05$) (donnée du sujet)
- Valeur critique : $\chi^2_{\text{théorique}}(\text{ddl} = 2\,;\ \alpha = 5\,\%) = 5{,}991$ (donnée du sujet)
- Valeur calculée sur l'échantillon : $\chi^2_{\text{calc}} = 1{,}810$

Règle de décision :

$$\begin{array}{l}
\texttt{Si}\ \ \chi^2_{\texttt{calc}} > \chi^2_{\texttt{théorique}} \ \longrightarrow\ \texttt{on rejette } H_0 \quad \texttt{(relation significative)} \\
\texttt{Si}\ \ \chi^2_{\texttt{calc}} \leq \chi^2_{\texttt{théorique}} \ \longrightarrow\ \texttt{on ne rejette pas } H_0 \quad \texttt{(pas de relation significative)}
\end{array}$$

Application numérique :

$$\underbrace{1{,}810}_{\chi^2_{\text{calc}}} \;<\; \underbrace{5{,}991}_{\chi^2_{\text{théorique}}}$$

Conclusion statistique :

$\chi^2_{\text{calc}}\,(1{,}810) < \chi^2_{\text{théorique}}\,(5{,}991)$: on ne rejette pas l'hypothèse nulle H_0. Au seuil de signification de $5\,\%$, le test du chi-deux ne permet pas de conclure à l'existence d'une relation statistiquement significative entre le lieu de résidence (Urbain / Rural) et la pratique du sport (Oui / Non / Parfois) au sein de cette population d'étudiants.

Commentaire :

Le test du χ^2 d'indépendance repose sur la confrontation des distributions observées aux distributions que l'on observerait si les deux variables étaient strictement indépendantes. Plus précisément, l'hypothèse nulle H_0 stipule qu'il n'existe aucune association entre le lieu de résidence et la pratique sportive : les différences constatées dans le tableau croisé seraient alors entièrement imputables aux fluctuations d'échantillonnage aléatoire, et non à un lien structurel entre les deux variables.

L'examen du tableau croisé révèle certes quelques disparités apparentes : les étudiants

> *ruraux tendent légèrement à déclarer « Non » plus fréquemment (18 contre 12 pour les urbains), tandis que les urbains déclarent « Oui » un peu plus souvent (20 contre 16). Cependant, ces écarts sont quantitativement modestes – les valeurs $(O_{ij} - E_{ij})$ ne dépassent pas 3 – ce qui indique que les fréquences observées s'éloignent très peu des fréquences attendues sous H_0.*

[4 pts – pour le commentaire]

Exercice 2 – Corrélation de Pearson [10 points]

Données : groupe de 8 étudiants. Variable X : heures de travail personnel par semaine ; variable Y : note obtenue à l'examen.

Question 1 – Tableau de calcul et moyennes $\bar{X}$ et $\bar{Y}$ [3 points]

Avant de calculer le coefficient de corrélation, on détermine les moyennes arithmétiques :

$$\bar{X} = \frac{\sum X_i}{n} = \frac{3+5+2+7+4+6+8+5}{8} = \frac{40}{8} = 5{,}00 \qquad \bar{Y} = 10{,}75$$

Tableau de calcul complet :

Ét.	X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i{-}\bar{X})(Y_i$-	$(X_i{-}\bar{X})^2$	$(Y_i{-}\bar{Y})^2$	X_i^2
1	3	8	-2	$-2{,}75$	$5{,}50$	4	$7{,}5625$	9
2	5	11	0	$+0{,}25$	$0{,}00$	0	$0{,}0625$	25
3	2	6	-3	$-4{,}75$	$14{,}25$	9	$22{,}5625$	4
4	7	14	$+2$	$+3{,}25$	$6{,}50$	4	$10{,}5625$	49
5	4	10	-1	$-0{,}75$	$0{,}75$	1	$0{,}5625$	16
6	6	13	$+1$	$+2{,}25$	$2{,}25$	1	$5{,}0625$	36
7	8	15	$+3$	$+4{,}25$	$12{,}75$	9	$18{,}0625$	64
8	5	9	0	$-1{,}75$	$0{,}00$	0	$3{,}0625$	25
Σ	**40**	**86**	0	0	$42{,}00$	28	$67{,}50$	**228**

On relève les trois sommes essentielles pour la suite :

$$\sum (X_i - \bar{X})(Y_i - \bar{Y}) = 42{,}00 \quad \Big| \quad \sum (X_i - \bar{X})^2 = 28{,}00 \quad \Big| \quad \sum (Y_i - \bar{Y})^2 = 67{,}50$$

[3 pts – $\bar{X}$ et $\bar{Y}$; les colonnes écarts ; et les sommes finales, indivisibles]

Question 2 – Calcul du coefficient de corrélation r de Pearson **[3 points]**

Le coefficient de corrélation linéaire de Pearson est défini par :

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \times \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Application numérique :

$$r = \frac{42}{\sqrt{28 \times 67{,}50}} = \frac{42}{\sqrt{1\,890}} = \frac{42}{43{,}475} \approx \mathbf{0{,}966}$$

[3 pts – pour la valeur finale]

Question 3 – Interprétation : force et sens de la relation **[4 points]**

Rappel de la grille d'interprétation standard du coefficient r :

$\lvert r \rvert$	Interprétation
$0{,}00 - 0{,}20$	Corrélation nulle à très faible
$0{,}20 - 0{,}40$	Corrélation faible
$0{,}40 - 0{,}60$	Corrélation modérée
$0{,}60 - 0{,}80$	Corrélation forte
$\mathbf{0{,}80 - 1{,}00}$	**Corrélation très forte**

Conclusion statistique :

$r = 0{,}966$ – Le coefficient de corrélation de Pearson est très proche de $+1$. Il existe donc une relation linéaire positive **très forte** entre le nombre d'heures de travail personnel hebdomadaire (X) et la note obtenue à l'examen (Y) : plus un étudiant consacre de temps au travail personnel, plus sa note tend à être élevée.

Commentaire :

Le coefficient de corrélation de Pearson (r) mesure le degré et le sens de la relation linéaire entre deux variables quantitatives continues. Sa valeur est toujours comprise entre -1 (relation linéaire négative parfaite) et $+1$ (relation linéaire positive parfaite) ; une valeur nulle indique l'absence de toute relation linéaire entre les variables.

La valeur obtenue, $r = 0{,}966$, est extrêmement proche de $+1$ et se situe très nettement au-delà du seuil communément admis de $0{,}80$ pour qualifier une corrélation de « très forte ». Deux éléments méritent d'être distingués : le signe positif du coefficient indique que les deux variables évoluent dans le même sens – un accroissement du temps de travail personnel

> *s'accompagne d'une augmentation de la note ; la valeur numérique élevée témoigne que cette relation linéaire est quasi-parfaite, avec très peu de dispersion autour de la droite de régression théorique.*

[4 pts – interprétation ; sens et force ; commentaire, indivisibles]

La séance de consultation des copies d'examen est prévue pour le lundi 25 mai 2026 à 12h00, au niveau du Bloc 9 salle 2