

Comprendre la significativité en analyse bivariable

Ahmed Fouad El Haddad

IEP de Fontainebleau UPEC

12 novembre 2025

Pourquoi ce cours ?

- En sciences sociales, on se demande souvent : *Est-ce que A est lié à B ?*
- Exemples :
 - Les femmes votent-elles différemment des hommes ?
 - Les pays plus riches sont-ils plus démocratiques ?
 - L'éducation favorise-t-elle la participation électorale ?
- Pour répondre, il ne suffit pas de comparer deux moyennes ou deux pourcentages.
- Il faut savoir si la relation est **significative**, c'est-à-dire *trop grande pour être due au hasard*.

Une histoire simple

Un chercheur analyse 200 bulletins de vote. Il remarque que 55 femmes ont voté à gauche, contre 40 hommes.

Question : Est-ce que les femmes votent vraiment plus à gauche ? Ou est-ce que c'est juste un petit hasard dans l'échantillon ?

C'est précisément le rôle des tests statistiques : *évaluer si ce qu'on observe dans un échantillon peut être généré par hasard.*

Deux types de variables à croiser

- **Variable qualitative :**

- Exemple : vote (gauche/droite), genre (homme/femme), statut (étudiant/non étudiant)

- **Variable quantitative :**

- Exemple : revenu, score de démocratie, taux de participation

- Le type de test dépend du type de variables :

- deux quantitatives : corrélation
- deux qualitatives : test du Chi²
- une de chaque : test t

Qu'est-ce qu'un test statistique ?

- C'est un outil pour **distinguer hasard et relation réelle**.
- Le test commence toujours par une hypothèse neutre :

H_0 : pas de lien, pas de différence

- Ensuite, on mesure l'écart observé dans les données.
- Enfin, on demande à la statistique : *Un tel écart est-il compatible avec le hasard ?*

Conclusion : si l'écart est trop grand pour être dû au hasard, alors on rejette l'hypothèse H_0 .

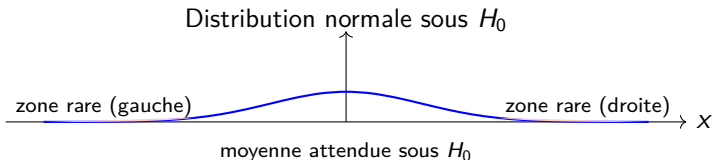
Comprendre une p-value

- Une p-value est une probabilité.
- Elle répond à la question :
Si le hasard seul gouvernait les résultats (H vrai), quelle est la probabilité d'observer un écart aussi extrême ou plus extrême que celui mesuré ?
- Si cette probabilité est très faible, on doute que ce soit le hasard.

Règle simple :

Si $p < 0,05 \Rightarrow$ l'écart est significatif

Illustration graphique de la p-value



p-value = surface rouge = rareté de l'écart observé sous H_0

Corrélation : exemple intuitif

Contexte : Un politiste étudie 40 pays.

- X = revenu moyen par habitant.
- Y = score de démocratie (0 à 10).

Observation : plus les pays sont riches, plus ils semblent démocratiques.

- Mais cette tendance est-elle due au hasard ?
- Ou reflète-t-elle un vrai lien ?

C'est exactement le but du test de corrélation de Pearson.

Corrélation : formule et test

Coefficient de corrélation de Pearson :

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} \quad \text{avec } -1 \leq r \leq 1$$

Test dhypothèse :

$$H_0 : \rho = 0 \quad (\text{pas de corrélation dans la population})$$

Statistique de test :

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$$

Si la p-value associée à t est < 0.05 , le lien est significatif.

Lecture politique

- Une corrélation de $r = 0.65$ avec $p < 0.01$ suggère un lien fort.
- Mais attention :
 - **Corrélation nest pas causalité** : d'autres facteurs peuvent expliquer la relation.
 - Une corrélation peut être influencée par une troisième variable (ex. éducation).
- Il faut toujours interpréter une corrélation à la lumière d'une théorie.

Chi-deux : exemple politique

Question : Le vote est-il lié au genre ?

	Gauche	Droite	Total
Hommes	40	60	100
Femmes	55	45	100
Total	95	105	200

Intuition : les femmes votent davantage à gauche. Mais cette différence est-elle significative ?

Test du Chi-deux : principe

- Le test compare les **valeurs observées** avec les **valeurs attendues** sous H_0 .
- Si les écarts sont trop grands, on rejette H_0 .

Formule :

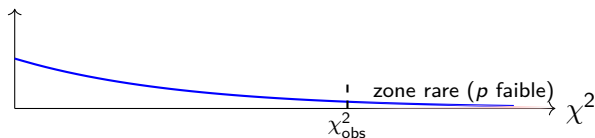
$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \quad E_{ij} = \frac{(\text{total ligne}) \times (\text{total colonne})}{n}$$

Hypothèse nulle :

H_0 : Vote et genre sont indépendants

Décision : Si $p < 0,05$, le lien est significatif.

Graphique simplifié du test Chi-deux



Si la valeur observée tombe dans la zone rouge, on rejette H_0 .

Et si l'échantillon est grand ?

- Plus un échantillon est grand, plus un petit écart devient significatif.
- Mais un **écart significatif** n'est pas toujours un **écart important**.
- Il faut donc mesurer la **taille de l'effet**, en plus de la p-value.

En résumé : *la significativité dit s'il y a un lien, la taille de l'effet dit si ce lien est fort ou faible.*

Taille deffet pour le Chi-deux : V de Cramer

Formule :

$$V = \sqrt{\frac{\chi^2}{n \cdot \min(r - 1, c - 1)}}$$

Interprétation :

- $V < 0.1$: effet très faible
- $0.1 \leq V < 0.3$: effet modéré
- $V \geq 0.5$: effet fort

Exemple : $\chi^2 = 4.6$, $n = 200 \Rightarrow V = 0.15$ *lien modéré entre genre et vote*

Résumé visuel des trois tests

Type de variables	Test utilisé	Loi statistique	Exemple
Quant – Quant	Corrélation de Pearson	Loi de Student (t)	Richesse /
Qual – Qual	Test du Chi-deux	Loi du Chiš	Genre / Vo
Quant – Qual	Test t de Student	Loi de Student (t)	Éducation ,

Chaque test :

- mesure un **écart à H** (absence de lien),
- convertit cet écart en **probabilité** (p-value),
- décide si le lien est significatif.

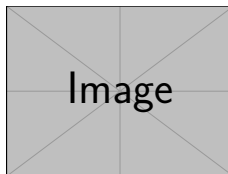
Conclusion générale

- Un test statistique évalue si une différence ou un lien observé est trop grand pour être dû au hasard.
- La p-value mesure la rareté d'un résultat sous l'hypothèse H_0 .
- Un résultat significatif ne signifie pas qu'il y a une causalité.
- Il faut articuler les résultats statistiques avec :
 - les données empiriques,
 - le contexte,
 - la théorie.

Un chiffre ne parle jamais seul. C'est l'analyse qui lui donne du sens.

Merci de votre attention

Questions ?



Prochaine séance : comparaison de moyennes (test t)