

Méthodes d'Analyse Statistique en Sciences Sociales

Chapitre 2 : Maîtrise des méthodes pour l'analyse statistique en sciences sociales

Introduction

Ce chapitre vous guidera à travers les concepts fondamentaux de l'analyse statistique spécifiquement adaptés aux sciences humaines et sociales. Ces outils vous permettront d'extraire du sens à partir de données quantitatives, de vérifier des hypothèses et de découvrir des relations entre différents phénomènes sociaux.

1. Les variables

1.1 Types de variables

En sciences sociales, nous manipulons différents types de données qui sont classées selon leur nature :

- **Variables qualitatives nominales** : catégories sans ordre (ex : genre, nationalité, profession)
- **Variables qualitatives ordinales** : catégories avec un ordre (ex : niveau d'éducation, échelle de Likert)
- **Variables quantitatives discrètes** : valeurs numériques entières (ex : nombre d'enfants)
- **Variables quantitatives continues** : valeurs numériques pouvant prendre n'importe quelle valeur dans un intervalle (ex : revenu, âge)

1.2 Niveau de mesure

- **Échelle nominale** : classification sans ordre naturel
- **Échelle ordinale** : classification avec un ordre
- **Échelle d'intervalle** : distances égales entre les valeurs, sans zéro absolu
- **Échelle de ratio** : distances égales entre les valeurs, avec un zéro absolu

1.3 Application en sciences sociales

En sociologie, anthropologie, psychologie ou science politique, le choix du type de variable est crucial pour déterminer les analyses statistiques appropriées. Par exemple, l'étude de l'influence du niveau d'éducation (variable ordinale) sur le revenu (variable continue) nécessite des méthodes spécifiques.

2. Mesures de tendance centrale et de dispersion

2.1 Moyenne

La moyenne arithmétique est la somme des valeurs divisée par le nombre d'observations :

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Exemple en sciences sociales : Calculer le revenu moyen d'un groupe démographique.

2.2 Autres mesures de tendance centrale

- **Médiane :** valeur qui divise une distribution en deux parties égales (50% au-dessus, 50% en-dessous)
- **Mode :** valeur la plus fréquente

2.3 Écart-type et variance

La variance mesure la dispersion des valeurs autour de la moyenne :

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

L'écart-type est la racine carrée de la variance :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Application : L'écart-type nous aide à comprendre l'homogénéité d'un groupe. Par exemple, deux communautés peuvent avoir le même revenu moyen mais des écarts-types très différents, indiquant des inégalités économiques différentes.

3. Analyse de la variance (ANOVA)

3.1 Principe de l'ANOVA

L'ANOVA permet de comparer les moyennes de plusieurs groupes pour déterminer s'il existe des différences significatives entre eux.

3.2 ANOVA à un facteur

Elle décompose la variance totale en :

- Variance inter-groupes (entre les différents groupes)
- Variance intra-groupes (à l'intérieur de chaque groupe)

3.3 Formulation et interprétation

- **Hypothèse nulle (H_0) :** Les moyennes des groupes sont égales
- **Hypothèse alternative (H_1) :** Au moins une moyenne diffère des autres

Le test F compare la variance inter-groupes à la variance intra-groupes :

$$F = \frac{\text{Variance, inter - groupes}}{\text{Variance, intra - groupes}}$$

Exemple : Comparer les attitudes politiques (variable dépendante) selon différentes catégories socio-professionnelles (variable indépendante).

4. Lois de probabilités usuelles

4.1 Loi normale

La distribution normale, ou gaussienne, est fondamentale en statistique :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Propriétés importantes :

- 68% des observations se situent à ± 1 écart-type de la moyenne
- 95% des observations se situent à ± 2 écarts-types de la moyenne
- 99,7% des observations se situent à ± 3 écarts-types de la moyenne

4.2 Loi du chi-carré (χ^2)

Utilisée notamment pour les tests d'indépendance et d'ajustement.

4.3 Loi de Student (t)

Utilisée pour comparer des moyennes, particulièrement avec de petits échantillons.

4.4 Application en sciences sociales

Ces distributions permettent de modéliser de nombreux phénomènes sociaux et sont à la base de nombreux tests statistiques utilisés pour valider des hypothèses.

5. Test du Khi-deux (χ^2)

5.1 Principe du test

Le test du χ^2 évalue l'indépendance entre deux variables qualitatives ou l'adéquation à une distribution théorique.

5.2 Test d'indépendance

Formule :

$$\chi^2 = \sum_{i,j} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

Où :

- $O_{i,j}$: fréquence observée
- $E_{i,j}$: fréquence théorique attendue

5.3 Interprétation et significativité

- **Hypothèse nulle (H_0)** : Les deux variables sont indépendantes
- **Hypothèse alternative (H_1)** : Il existe une relation entre les deux variables

Exemple concret : Étudier la relation entre le niveau d'éducation et les opinions politiques dans une population.

6. Coefficient de corrélation

6.1 Corrélation de Pearson

Le coefficient de corrélation de Pearson (r) mesure la force et la direction de la relation linéaire entre deux variables quantitatives :

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Valeurs possibles : entre -1 et 1

- $r = 1$: corrélation positive parfaite
- $r = -1$: corrélation négative parfaite
- $r = 0$: absence de corrélation linéaire

6.2 Autres mesures de corrélation

- **Coefficient de Spearman** : pour variables ordinales ou relation non linéaire
- **Tau de Kendall** : alternative robuste pour petits échantillons

6.3 Interprétation en sciences sociales

Important : Corrélation n'implique pas causalité !

Exemple : La corrélation entre le niveau d'éducation et le revenu ne signifie pas nécessairement que l'éducation cause directement un revenu plus élevé.

7. Régression linéaire simple

7.1 Principe de la régression

La régression linéaire simple modélise la relation entre une variable dépendante (Y) et une variable indépendante (X) :

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Où :

- β_0 : ordonnée à l'origine
- β_1 : pente
- ε : terme d'erreur

7.2 Estimation des paramètres

La méthode des moindres carrés estime les paramètres en minimisant la somme des carrés des résidus.

7.3 Interprétation des résultats

- **Coefficient de détermination (R^2)** : proportion de la variance expliquée par le modèle
- **Test de significativité** : déterminer si β_1 est significativement différent de zéro

Exemple : Étudier comment le temps passé sur les réseaux sociaux (X) influence le bien-être psychologique (Y).

8. Régression linéaire multiple

8.1 Extension à plusieurs variables

La régression multiple inclut plusieurs variables indépendantes :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

8.2 Interprétation des coefficients

Chaque coefficient β_1, β_2 , etc. représente l'effet de la variable correspondante lorsque les autres variables sont maintenues constantes.

8.3 Problèmes potentiels

- **Multicolinéarité** : forte corrélation entre variables indépendantes
- **Hétéroscédasticité** : variance non constante des résidus
- **Variables omises** : absence de variables importantes dans le modèle

Application : Analyser les facteurs (âge, éducation, région, etc.) qui influencent les attitudes envers l'immigration.

9. Introduction à l'analyse factorielle

9.1 Principe de l'analyse factorielle

L'analyse factorielle réduit un grand nombre de variables à un petit nombre de facteurs latents qui expliquent la majorité de la variance observée.

9.2 Analyse en composantes principales (ACP)

L'ACP transforme des variables corrélées en un ensemble de variables non corrélées appelées composantes principales.

9.3 Analyse factorielle exploratoire vs. confirmatoire

- **Exploratoire** : identifie la structure sous-jacente sans hypothèse préalable
- **Confirmatoire** : teste si les données correspondent à une structure hypothétique

9.4 Applications en sciences sociales

- Création d'indices composites (ex : indice de statut socio-économique)
- Réduction de nombreuses questions d'enquête à quelques dimensions clés
- Identification de construits latents (ex : traits de personnalité)

Exemple : Analyser un questionnaire de 25 items sur les attitudes politiques pour identifier les dimensions fondamentales sous-jacentes.

Conclusion

Ces méthodes statistiques constituent la boîte à outils essentielle du chercheur en sciences sociales. Elles permettent de passer de l'intuition à la démonstration, et des observations qualitatives aux généralisations quantitatives. Toutefois, il est crucial de se rappeler que les statistiques ne sont qu'un outil au service de la théorie et de l'interprétation.

Lectures recommandées

- Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics* (5e éd.). SAGE Publications.
- Fox, J. (2015). *Applied Regression Analysis and Generalized Linear Models* (3e éd.). SAGE Publications.
- Howell, D. C. (2014). *Méthodes statistiques en sciences humaines* (2e éd.). De Boeck.
- Tabachnick, B. G., & Fidell, L. S. (2019). *Using Multivariate Statistics* (7e éd.). Pearson.

Exercices pratiques

1. **Variables et mesures** : Identifiez le type de variable et le niveau de mesure approprié pour : l'âge, le niveau d'éducation, le genre, le revenu mensuel, et le code postal.
2. **Statistiques descriptives** : Calculez la moyenne, la médiane, l'écart-type et construisez un histogramme pour les données de revenu d'un échantillon de 30 personnes.

3. **Khi-deux** : Testez l'indépendance entre le genre et le choix politique à partir d'un tableau de contingence.
4. **Corrélation** : Calculez et interprétez le coefficient de corrélation entre le nombre d'années d'études et le revenu.
5. **Régression simple** : Construisez un modèle prédisant le score de bien-être (variable dépendante) à partir du temps de loisir hebdomadaire (variable indépendante).
6. **Régression multiple** : Ajoutez au modèle précédent les variables d'âge et de statut marital. Interprétez les coefficients et évaluez l'amélioration du modèle.
7. **ANOVA** : Comparez les scores moyens de satisfaction au travail entre trois groupes professionnels différents.
8. **Analyse factorielle** : Réduisez un ensemble de 10 questions sur les valeurs personnelles à 2-3 facteurs principaux et interprétez ces facteurs.