

Branches complémentaires - Bachelor - Semestre d'automne 2018

# MÉTHODES EMPIRIQUES DES SCIENCES SOCIALES

FLORENCE CAUHÉPÉ

Cours no 9 du 14 novembre 2018

# DÉPOUILLEMENT DES QUESTIONNAIRES ET PREMIÈRES ANALYSES QUANTITATIVES

- Introduction
- **I. Préparer les données**
- 1. Mesurer l'information
- 1.1. Codage des questions fermées
- 1.2. Codage des questions ouvertes
- 1.3. Création de variables synthétiques ou d'indices
- 2. Des questions aux variables: codes et échelles de mesures
- 3. La méthode des échelles
- **II. Traiter, analyser les données**
- 1. Décrire les données; le tri à plat ou tri marginal
- 1.1. La distribution sous forme de tableau descriptif
- 2. Comprendre les réponses; le tableau croisé
- 2.1. Le tableau croisé
- 2.2. Formulation des énoncés
- 2.3. Analyse des tableaux croisés
- 2.4. Le hasard aurait-il pu donner un tel tableau, le test du khi carré
- 3. Contrôler les relations; l'analyse multivariée
- **III. Saisir la diversité des situations**
- 1. Rôle des indicateurs de mesure
- 2. Les indicateurs de tendance centrale: mode, médiane, moyenne, exemples
- 3. Les indicateurs de dispersion; exemple
- 3.1. La dispersion autour de la moyenne: l'écart-type
- 4. La notion de variabilité
- 5. Comparer les résultats obtenus avec les résultats attendus
- 6. Ecriture du compte rendu
- 7. Les logiciels
- Bibliographie et Conclusion
- **IV. Partie pratique**

# Introduction

- Dans l'analyse qualitative ou quantitative, trois opérations sont toujours nécessaires :
  - la préparation des données ou des informations; réduction, codage
  - la mise en relation des données
  - la comparaison des résultats obtenus par rapport aux résultats attendus
- L'étape d'analyse sert essentiellement à corroborer ou à réfuter l'hypothèse de recherche
- Le développement informatique a profondément transformé l'analyse des données

# I. Préparer les données

- Les informations recueillies seront présentées, de façon économique, grâce au codage, sous une forme simple et systématique :
  - Le dictionnaire des codes
  - Le tableau/ fichier de présentation des données

Tableaux tirés de Nicole Bersier, 2011, Les Techniques d'enquête en sciences sociales, Cursus, Armand Colin, 4éd.

# Le dictionnaire des codes

Question (résumé)	Numéro de la variable	Nom de la variable (résumé)	Description du codage
numéro du questionnaire	1	numéro	nombres de 1 à 562
Q 1 – Sexe	2	sexe	1 = masculin 2 = féminin 9 = non-réponse
Q 2 – Âge	3	âge	nombre de 18 à 99
Q 3 – Venu avec qui ?	4	venu	1 = seul 2 = voisins, amis 3 = en famille 4 = collègues 9 = non-réponse
Q 4 – Intérêt de l'exposition	5	intérêt	1 = très intéressante 2 = assez intéressante 3 = peu intéressante 4 = pas intéressante du tout 9 = non-réponse

# Tableau, fichier, matrice des données

Numéro	Sexe	Âge	Venu	Intérêt	Etc.
1	1	35	2	1	
2	1	45	3	1	
3	2	28	3	2	
4	1	32	1	1	
etc.					

# 1.1. Codage des questions fermées

1. Un numéro de code est attribué à chaque modalité de réponses, c'est le cas le plus simple, le code est attribué à l'avance
2. Codification des questions fermées à choix multiple
3. Codification par décomposition logique
4. Codification des questions fermées à classement



# 1.2. Codage des questions ouvertes

1. Questions ouvertes quantitatives
  2. Questions ouvertes qualitatives
- Elaborer une catégorisation des réponses

# 1.3.La construction de variables synthétiques ou d'indices

- Grâce au codage, on peut créer **des variables dites synthétiques, aussi appelées indices**
- Ces variables synthétiques sont créées soit:
  - **par fusion (=combinaison) de réponses**
  - **par agrégation de plusieurs indicateurs** destinés à évaluer un même phénomène, une même notion
- C'est la construction de l'indice et l'observation de sa variation qui permettent de comprendre la notion étudiée

## 2. Des questions aux variables

- Lors de l'analyse statistique, chaque question devient une variable (cf cours no 6)

Trois types de variables

- Variable nominale ou catégorie
- Variable ordinale ou ordonnée
- Variable quantitative ou numérique

# 3. Codes et échelles de mesure

- Les codes attribués aux réponses n'auront pas la même signification suivant la nature des variables
- Lors d'une analyse quantitative des données, on compte, on quantifie
- Pour quantifier, il faut que les variables soient quantifiables
- Alors comment rendre quantifiables des faits qui ne sont pas quantitatifs ?
- **Par l'utilisation de la méthode des échelles**

# 4. La méthode des échelles

- L'échelle nominale
- L'échelle ordinale
- L'échelle d'intervalles (= échelle d'attitude)
  - L'échelle de Likert (degré d'accord)
  - Le différentiel ou différenciateur sémantique (échelle d'Osgood)
- Les échelles de rapport ou de proportion

# II. Traiter les données

- Trois types d'analyse:
  1. Décrire les variables; le tri à plat
  2. Comprendre les réponses; le tableau croisé
  3. Contrôler les relations; l'analyse multivariée

# 1. Décrire les variables; le tri à plat ou tri marginal

- 1<sup>ère</sup> étape de l'analyse : description simple de l'information
- Elle est réalisée à travers l'opération du tri à plat ou tri marginal, c'est-à-dire le calcul de la distribution des effectifs et des pourcentages des modalités de réponses pour chaque question

# Exemple de tri à plat

La culture, c'est une ouverture sur les autres :

	<i>Effectif</i>	<i>Pourcentage</i>
Tout à fait d'accord	527	65 %
Plutôt d'accord	218	27 %
Plutôt pas d'accord	35	4 %
Pas du tout d'accord	6	1 %
Non-réponse	24	3 %
Total	810	100 %



# Analyser les tris à plat; premières observations

- Qui a répondu ?
- L'échantillon est-il représentatif ?
- Examen des non-réponses
- Recodage (création de variables synthétiques ou d'indices)

## 2. Comprendre les réponses; le tableau croisé

- L'analyse de l'enquête commence véritablement avec l'examen de tableaux prenant en compte non plus une seule variable, **mais simultanément deux variables**
  - Un des supports les plus courants de cette analyse est un tableau croisant :
    - un indicateur d'un déterminant social (variable sociale ou indépendante)
- et
- un indicateur d'une pratique (variable dépendante ou à expliquer)

## 2.1. Le tableau croisé

- Dans la perspective explicative, un tableau croisé a pour objectif de mettre en lumière l'action éventuelle d'un facteur social, d'un déterminant social sur la pratique à étudier
- Pour construire le tableau croisé, il faut reprendre le fichier des données, afin d'observer, en même temps, les modalités de réponses à deux questions pour un répondant

# «La culture est une ouverture aux autres» avec effectifs

	Bac + 1	Bac + 2	Bac + 3	Bac + 4 et au-delà	Total
Tout à fait d'accord	137	123	131	136	527
Plutôt d'accord	82	65	49	22	218
Plutôt pas d'accord	14	8	8	5	35
Pas du tout d'accord	2	2	1	1	6
Total	235	198	189	164	786

# Avec les pourcentages calculés en colonnes

	Bac + 1	Bac + 2	Bac + 3	Bac + 4 et au-delà	Total
Tout à fait d'accord	58 %	62 %	69 %	83 %	67 %
Plutôt d'accord	35 %	33 %	26 %	13 %	28 %
Plutôt pas d'accord Pas du tout d'accord	7 %	5 %	5 %	4 %	5 %
Total	100 %	100 %	100 %	100 %	100 %
Effectif	(235)	(198)	(189)	(164)	(786)

## 2.2. Formulation des énoncés

- Une fois soulignés, colonne par colonne, les nombres les plus significatifs qui indiquent l'action de la variable indépendante, il convient d'écrire un ou plusieurs énoncés résumant la ou les variation-s observée-s
- Ex. d'énoncés

1<sup>er</sup> énoncé: pour la modalité **tout à fait d'accord**, le niveau Bac+1 a le score le plus faible (58%) et celui-ci augmente d'un groupe à l'autre, jusqu'au niveau le plus élevé (83%) pour la catégorie Bac+4

2<sup>ème</sup> énoncé: pour la modalité **plutôt d'accord**, **la variation est inverse**: le niveau Bac+1 a le plus fort score (35%) et le niveau d'étude le plus élevé Bac+4 a le score le plus faible (13%)

3<sup>ème</sup> énoncé: pour la modalité **pas d'accord**: les écarts sont très faibles. La variation est continue et décroissante entre les 4 groupes, mais le saut est plus marqué entre les Bac+1 et les Bac+4

4<sup>ème</sup> énoncé: dans les quatre groupes, on a toujours la même hiérarchie des modalités

**Résumé**: l'accord total avec l'opinion «la culture comme ouverture» est toujours majoritaire et augmente régulièrement avec le niveau d'étude au détriment de l'accord mitigé

## 2.3. Analyse d'un tableau croisé

- But:
  - Dans l'exemple présenté, le tableau croisé montre l'effet que peut avoir le niveau d'études sur l'accord avec la proposition: «La culture, c'est une ouverture sur les autres»
- Les étapes:
  - Calcul des pourcentages
  - Analyse comparative
  - Discussion et interprétation

## 2.4. Le hasard aurait-il pu donner un tel tableau ? Le Test du khi carré

- Le hasard aurait-il pu donner une telle répartition dans le tableau (hypothèse d'indépendance) ou au contraire, y a-t-il une liaison notable entre les deux variables ?
- Le tableau d'indépendance ou **tableau des effectifs théoriques (T)** permet de répondre
- Dans le cas d'une distribution aléatoire dans le tableau, les effectifs de chaque case devraient, théoriquement, être proportionnels aux effectifs totaux de la ligne et de la colonne, rapportés à l'ensemble des répondants (simple règle de trois)
- Pour évaluer si cet écart doit être considéré comme significatif, les logiciels d'analyse calculent une statistique: **le khi deux ou khi carré**
- $x^2 = \sum \frac{(O - T)^2}{T}$

T



# Règle du khi carré et règle de précaution

- Si le khi carré calculé est **inférieur** au khi carré théorique = il y a **indépendance**
- Si le khi carré calculé est **supérieur** au khi carré théorique = il y a **dépendance**
- **Ce test est valable pour les variables nominales et ordinales**
- Le lien statistique doit ensuite être appuyé par des éléments sociologiques explicatifs

# 3. Contrôler les relations; l'analyse multivariée

- Une relation constatée n'est pas une relation avérée
  
- Le contrôle de variables
  - la démarche consiste à introduire progressivement de nouvelles variables dans la relation à deux variables pour examiner leurs effets
  
- La nouvelle variable, dont on va contrôler l'effet, se nomme la variable-test

# Accord avec l'opinion

«La culture c'est une ouverture aux autres»,  
selon le niveau d'étude et l'université fréquentée

	Université I		Université II		Université III	
	Bac + 1 et 2	Bac + 3 et au-delà	Bac + 1 et 2	Bac + 3 et au-delà	Bac + 1 et 2	Bac + 3 et au-delà
Tout à fait d'accord						
Plutôt d'accord						
Plutôt pas d'accord Pas du tout d'accord						
Total	100 %	100 %	100 %	100 %	100 %	100 %

# III. Saisir la diversité des situations

- La notion de variabilité renvoie à l'idée de variation, de diversité, d'hétérogénéité
- Un phénomène de grande variabilité est un phénomène présentant de fortes différences entre les individus
- Inversement, un phénomène présentant une faible variabilité est un phénomène plutôt homogène, montrant des similitudes: les individus présentent des caractéristiques proches

# 1. Rôle des indicateurs de mesure

- Les indicateurs de mesure de variabilité doivent rendre compte :
  1. **de la fréquence** avec laquelle ces valeurs surviennent = les **indicateurs de tendance centrale**
  2. **de l'étendue des valeurs** (quelles sont les différentes valeurs prises par la variable) = les **indicateurs de dispersion**, ils expriment la dispersion des valeurs d'une variable (autour de la moyenne)

## 2. Les indicateurs de tendance centrale: mode, médiane, moyenne

- **Le mode** correspond à la valeur la plus fréquente dans la distribution d'une variable
- En mathématique; **la moyenne simple** est la somme des observations divisée par le nombre de ces observations
- Mais elle peut être **pondérée**
- **La médiane** notée  $Me$  correspond à la valeur qui sépare les unités considérées (la distribution), en deux groupes d'effectifs égaux

# Exemple d'une moyenne de notes pondérée par des coefficients

Sacha a obtenu les notes suivantes sur 20 en mathématiques au cours du premier trimestre :

Devoir	DM1	DS1	DM2	DS2	DM3	DS3
Note	14	14,5	9	15	20	16
Coefficient	1	2	1	3	2	3

Calculons M la moyenne de Sacha au premier trimestre :

$$M = \frac{14 \times 1 + 14,5 \times 2 + 9 \times 1 + 15 \times 3 + 20 \times 2 + 16 \times 3}{1 + 2 + 1 + 3 + 2 + 3} \approx 15,4$$

## Moyenne pondérée d'une série de valeurs statistiques

Reprenons l'exemple précédent :

Dans une classe de 25 élèves, les notes sur 20 obtenues lors d'un devoir surveillé sont :

16 ; 11 ; 8 ; 13 ; 9 ; 11 ; 9 ; 13 ; 15 ; 7 ; 7 ; 9 ; 11 ; 15 ; 16 ; 15 ; 11 ; 8 ; 9 ; 13 ; 14 ; 14 ; 11 ; 8 ; 13 ;

On range ces notes dans un tableau en suivant l'ordre croissant.

Note	7	8	9	11	13	14	15	16
Effectif	2	3	4	5	4	2	3	2

Calculer  $M_c$ , la moyenne de la classe, revient à faire le calcul suivant :

$$M_c = \frac{7 \times 2 + 8 \times 3 + 9 \times 4 + 11 \times 5 + 13 \times 4 + 14 \times 2 + 15 \times 3 + 16 \times 2}{2 + 3 + 4 + 5 + 4 + 2 + 3 + 2} = 11,44$$



# 3. Les indicateurs de dispersion

- On pourra aussi s'intéresser à la dispersion des valeurs autour de la moyenne (écart-type) et examiner en détail la distribution des réponses
- Les indicateurs de dispersion:
  - L'écart moyen arithmétique:  $em$
  - La variance:  $v$
  - La dispersion autour de la moyenne; **l'écart-type:  $\sigma$**

# Exemple: valeur moyenne des notes

Notes	1	2	3	4	5	6	7	8	9	10
Effectifs	1	0	5	8	12	7	9	2	1	1

La somme pondérée des écarts à la note moyenne:  
 $em =$

$$(1-5.4) + 5(3-5.4) + 8(4-5.4) + 12(5-5.4) + 7(6-5.4) + 9(7-5.4) + 2(8-5.4) + (9-5.4) + (10-5.4) / 46 = (4.4) + 5(2.4) + 8(1.4) + 12(0.4) + 7(0.6) + 9(1.6) + 2(2.6) + 3.6 + 4.6 / 46 = \mathbf{64.4/46 = 1.4}$$

Variance:  $v = (-4.4)^2 + 5(-2.4)^2 + 8(-1.4)^2 + 12(-0.4)^2 + 7(0.6)^2 + 9(1.6)^2 + 2(2.6)^2 + (3.6)^2 + (4.6)^2 / 46 =$   
 $\mathbf{138.96/ 46 = 3.02}$

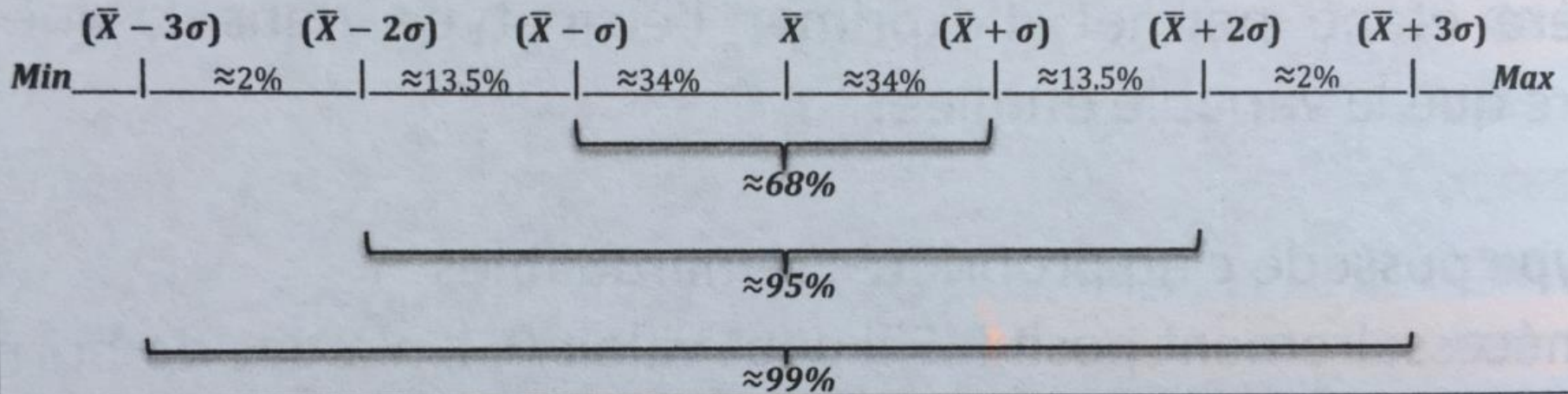
Ecart-type est la racine carrée de la variance:  $\mathbf{1.738}$

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

# 3.1. L'écart-type

## La dispersion autour de la moyenne ( $\bar{X}$ ) : l'écart-type ( $\sigma$ )

(en admettant que la variable suit une loi Normale)



IMG\_1614.JPG  
Type : Fichier JPG  
Taille : 1.91 Mo  
Dimension : 3264 x 2448 pixels

## 4. La notion de variabilité est centrale dans toute recherche empirique

- Le chercheur doit faire face à des situations présentant une certaine diversité, une certaine hétérogénéité
- L'essentiel est de retenir que la variance et l'écart-type sont toujours positifs et qu'ils **prennent des valeurs d'autant plus grande, que la variabilité, c'est-à-dire l'hétérogénéité, d'une variable est grande**

## 5. Comparer les résultats obtenus avec les résultats attendus

- Chaque hypothèse, élaborée lors du cadre théorique, exprime les relations supposées, les relations que l'on pensait correctes
- L'observation empirique et l'analyse viendront confirmer ou infirmer ces relations présumées
- C'est donc en comparant les résultats observés avec les résultats qui étaient attendus, selon l'hypothèse émise, que l'on pourra tirer des conclusions

# 6. Écriture du compte rendu

- **Dans un 1<sup>er</sup> chapitre: l'objet de l'enquête**, la manière dont il a été construit, le cadre théorique et les hypothèses de départ, l'échantillon, la logique du questionnaire, avec les indicateurs
- **Chapitres suivants:** les résultats, résumés de manière synthétique, selon la lecture systématique de tout le questionnaire par telle ou telle variable indépendante
- Montrer si une relation empirique renvoie à une relation hypothétique entre deux notions
- **Un dernier chapitre** donne une lecture transversale des chapitres précédents, grâce à l'analyse des données
- En annexe, on met le questionnaire et les tableaux croisés complémentaires

# 7. Les logiciels

- La mise au point de nouvelles procédures statistiques permet de visualiser et d'étudier les liaisons entre plusieurs dizaines de variables en même temps;
  - analyse factorielle des correspondances, AFC
  - analyse en composantes principales; ACP
  - analyse de la variance, ANOVA...
- L'outil statistique a un pouvoir d'élucidation limité aux postulats sur lesquels il repose, mais il ne dispose pas en lui-même d'un pouvoir explicatif
- Une initiation aux programmes informatiques de gestion et d'analyse est souvent nécessaire: **SPSS** est très utilisé, SPAD, SAS/STAT, STATVIEW

# Bibliographie

- Raymond Bourdon, 1993, *Les méthodes en sociologie*, Paris, PUF, coll. Que sais-je?
- Madeleine Grawitz, 2001, *Méthodes des sciences sociales*, Dalloz, 11<sup>ème</sup> éd.
- Philippe Cibois, 1984, *L'analyse des données en sociologie*, Paris, PUF, coll. Le sociologue
- Olivier Martin, 2012, *L'analyse quantitative des données*, Paris, Armand Colin
- Flora Chanvril-Ligneel, Viviane Le Hay, 2014, *Méthodes statistiques pour les sciences sociales*, Ellipses Edition



# Conclusion

Malgré les avancées de l'informatique

- c'est le chercheur qui donne un sens aux relations par le modèle théorique qu'il a construit au préalable
- les chercheurs ne renoncent pas pour autant à l'usage de techniques plus classiques comme celles des tableaux croisés notamment

## IV. Partie pratique

- Dans le tableau de la slide suivante, il s'agit du traitement de la réponse à la question: «Combien as-tu de frères et sœurs?»
- Question posée dans une enquête à 514 lycéens

# Combien as-tu de frères et sœurs ?

Nombre de frères et sœurs	Effectif	Pourcentage
0	37	7 %
1	95	18 %
2	123	24 %
3	79	15 %
4	54	11 %
5	39	8 %
6	23	4 %
7	27	5 %
8	9	2 %
9	10	2 %
10	8	2 %
11	0	0 %
12	5	1 %
13	2	0 %
14	3	1 %

# Exercice

- 1. Ce tableau est-il un tri à plat ou un tri croisé?
- 2. Donnez le mode et la médiane du nombre de frères et soeurs
- 3. Donnez la moyenne de la distribution
- 4. Donnez l'écart-type de la distribution
- 5. Décrivez en termes simples le contenu du tableau
- 6. Condensez la distribution en perdant le moins d'information possible, mais en ne conservant aucun groupe avec un effectif inférieur à 30

# INFORMATIONS IMPORTANTES pour les deux prochains cours

■ Journée des Gymnasiens le **21.11.2018**

➤ **Pas de cours**

■ **Cours du 28.11.2018**

➤ **de 13h15 à 15h Auditoire G 120**