

SOCIO532.2
(STATISTIQUES &) INFORMATIQUE APPLIQUÉE(S)
AUX SCIENCES SOCIALES

El Hadj Touré, Ph.D. Sociologie
Département de sociologie
Université Gaston Berger de St-Louis

Leçon 2

Comment synthétiser l'information statistique?

13.29 1

1

De la nécessité de synthétiser...

...l'information statistique

- ✓ Pourquoi synthétiser l'information statistique?
 - ✓ Une stratégie pour faire face à l'hétérogénéité des données dans une étude quanti et passer des indicateurs au concept
 - ✓ Une opération pour faciliter l'analyse sociologique, et donc la vérification des hypothèses de recherche
- ✓ Comment synthétiser l'information statistique?
 - ✓ Création de variables composites (indices)
 - ✓ Création de tableaux et de graphes synthétiques
- ✓ La synthèse de l'information statistique est plus qu'une simple opération technique: elle est une opération théorico-sociologique

13.29 2

2

Au programme

<ul style="list-style-type: none"> ■ De la nécessité du recodage de variables <ul style="list-style-type: none"> ❖ 5 stratégies de recodage de variables pour des raisons technico-pratiques et théoriques ■ Création de variables synthétiques <ul style="list-style-type: none"> ❖ 7 stratégies pour passer des indicateurs empiriques à l'indice statistique ou à une seule variable composite 	<ul style="list-style-type: none"> ■ Création de tableaux synthétiques <ul style="list-style-type: none"> ❖ Synthétiser les résultats d'une analyse statistique de plusieurs variables dans un seul tableau compact ■ Création de graphes synthétiques <ul style="list-style-type: none"> ❖ Synthétiser les résultats d'une analyse statistique de plusieurs variables dans un seul graphe compact
---	--

3

3

De la nécessité de recoder les variables...

...pour des raisons technico-pratiques et théoriques

- Nécessite de recoder les variables avant d'être en mesure de synthétiser l'information statistique
- Recoder une variable, c'est préparer les données
 - ✓ de façon à les rendre adéquates à certaines techniques statistiques (raison technico-pratique)
 - ✓ et à la problématique (raison théorique)
- Le recodage est une opération de traitement statistique des données
 - ✓ Or le traitement statistique précède l'analyse statistique des données

13.29

4

4

De la nécessité de recoder les variables...

...quali

1. Variable quali avec de multiples catégories à regrouper
 - Niveau de scolarité désiré
 - 1.L3 socio
 - 2.L3 autre
 - 3.M2 socio
 - 4.M2 autre
 - 5.Doctorat socio
 - 6.Doctorat autre
 - On peut regrouper les catégories selon les 3 niveaux
 - 1. L3
 - 2. M2
 - 3. Doctorat
 - Ou dichotomiser en termes de: 1.Doctorat; 0.Autrement
- ✓ Autres exemples: religion, région, etc.

13.29

5

5

De la nécessité de recoder les variables...

...quanti

2. Variable quanti avec des valeurs métriques à fusionner
 - Revenu mensuel en Fcfa
 - 0; 20000; 50000; 100000; 120000; 130000; 135000, etc.
 - On peut fusionner les valeurs en classes comme ceci:
 - 1. Moins de 50000 fcfa
 - 2. 50000-99999 fcfa
 - 3. 100000 fcfa et plus
 - Ou dichotomiser ainsi: 1.100000 et plus; 0.Autrement
 - On peut utiliser des seuils statistiques
 - médiane, terciles, un écart-type au-dessus de la moyenne...
- ✓ Autres exemples: âge, nombre d'heures sur le Net, etc.

13.29

6

6

De la nécessité de recoder les variables...

...quali à réponses multiples

3. Variable quali à réponses multiples à dichotomiser

- Quel(s) usage(s) faites-vous du téléphone?
 - 1.Appels; 2.Textos; 3.Internet; 4.Jeux; 5.Musique; 6.Photos..
 - ✓ Les répondants peuvent cocher plusieurs usages
- Respecter le principe suivant: les catégories d'une variable doivent être mutuellement exclusives
 - Pour chaque usage, on crée une variable dichotomique: 'a choisi/'n'a pas choisi' (1.Usage; 0.Non usage)
 - Donc autant de modalités (usages) que de variables dico.

Appels	Textos	Internet	Jeux	Musique	Photos
1. oui/usage	1. oui/usage	1. oui/usage	1. oui/usage	1. oui/usage	1. oui/usage
0. Non	0. Non	0. Non	0. Non	0. Non	0. Non

13.29 7

7

De la nécessité de recoder les variables...

...quali à réponses ouvertes

4. Données quali d'une question ouverte à coder

- Qu'est-ce qui vous motive à allier études et travail?
 - P1: Je dois travailler pour payer mes frais de scolarité...
 - P2: Ma famille n'a pas les moyens. Je dois l'aider...
 - P3: Je ne peux pas dépendre de mes parents...
 - P4: J'aime entreprendre, créer des affaires...
 - P4: Je travaille car la vie est une compétition même pour...
- Lire les données pour déterminer les catégories à retenir et les coder selon ces catégories, le nbre pouvant évoluer
- Au moins trois catégories à retenir
 - ✓ 1. Adversité familiale
 - ✓ 2. Autonomie financière/sociale
 - ✓ 3. Esprit d'entrepreneuriat/d'émulation

13.29 8

8

De la nécessité de recoder les variables...

...quali à données atypiques

5. Autres matériaux quali (texte, audio, vidéo...) à coder

- Annonces matrimoniales (De Singly, cité par Martin, 2009)
 - Coder les annonces: ressortir les manœuvres de séduction
 - ✓ Outre le sexe, l'âge, le statut socioéco. de l'annonceur,
 - ✓ il y a: nbre de mots, présence d'enfants, référence économique, morale, physique...
- Autre ex: L'affaire Sonko-Adjji: le traitement de l'information dans les journaux sénégalais
 - Ancienneté, région, idéologie, pro-pouvoir vs pro-opposition
- Autre ex: La recherche sociologique chez les étudiants. Méta-analyse des mémoires soutenus au dpt. de socio

13.29 9

9

Au programme

- De la nécessité du recodage de variables
 - ❖ 5 stratégies de recodage de variables pour des raisons technico-pratiques et théoriques
- Création de variables synthétiques
 - ❖ 6 stratégies pour passer des indicateurs empiriques à l'indice statistique ou à une seule variable composite
- Création de tableaux synthétiques
 - ❖ Synthétiser les résultats d'une analyse statistique de plusieurs variables dans un seul tableau compact
- Création de graphes synthétiques
 - ❖ Synthétiser les résultats d'une analyse statistique de plusieurs variables dans un seul graphe compact

10

10

Les variables synthétiques

1. Créer une variable synthétique par comptage

- L'intensité de la pratique de sport
 - Parmi les activités physiques et sportives suivantes, indiquez celles que vous pratiquez au moins une fois par an ? Natation ; Jogging ; Soccer ; Danse ; Autre
 - Recoder d'abord pour créer des variables dichotomiques (0/1)
 - On compte le nombre d'activités pratiquées. Pour cela, il suffit d'attribuer le score 1 à chacune des activités pratiquées, puis de compter au final le score total de chaque individu.
 - Indice avec valeurs métriques possibles : 0, 1, 2, 3, 4, 5.
 - Indice qu'on peut recoder sous forme de catégories pour mesurer la pratique de sport en termes d'intensité : 1.Faible (0-1); 2.Modérée (2-3); 3.Élevée (4-5)

13.29 11

11

Les variables synthétiques

2. Créer une variable synthétique par addition | soustraction

- Calcul du revenu familial total
 - Si on dispose du revenu du père et du revenu de la mère
 - Pour calculer le revenu familial, on applique la formule:

$$\text{RevenuTotal} = \text{RevenuPère} + \text{RevenuMère}$$
 - Pour chaque cas, on additionne les deux revenus
- Calcul de l'âge des cas sondés
 - Si on a la date de naissance et l'année de sondage
 - Pour calculer l'âge, on applique la formule:

$$\text{Age} = \text{AnneeSondage} - \text{DateNaissance}$$
 - Pour chaque cas, on soustrait la date de naissance de l'année de sondage

13.29 12

12

Les variables synthétiques

3. Créer une variable synthétique par addition

- Indice global d'individualisation du portable (Martin,2009)
 - « Arrive-t-il que le conjoint réponde à votre place avec votre portable ? » (0.oui, 1.non)
 - « Au cours de la dernière semaine, votre conjoint a-t-il emprunté votre portable ? » (0.oui, 1.non)
 - « Au cours de la dernière semaine, avez-vous reçu des appels sur votre portable pour votre conjoint? » (0.oui, 1.non)
 - « Votre conjoint connaît-il le code de votre portable ? » (0.oui, 1.non)
 - Pour chaque réponse négative, un individu se voit attribuer 1 point. La somme des scores mesure l'individualisation du portable. Les scores vont de 0 à 4, et correspondent à un très faible, faible, moyen, assez fort et très fort individualisme.

13.29 13

13

Les variables synthétiques

4. Créer une variable synthétique par addition

- Indice et intensité du risque de décrochage scolaire (Amadou Dia, étude de cas, 2024)
 - Absentez-vous fréquemment aux cours? (Très souvent, assez souvent, rarement, jamais)
 - À dichotomiser: 1.Très souvent/assez souvent, 0.Autre
 - Pensez-vous qu'abandonner ses études est une catastrophe? (0.oui, 1.non)
 - Mettriez-vous un terme à vos études si une opportunité de travail se présentait au cours de l'année? (1.oui, 0.non)
 - Pour chaque réponse de risque, un étudiant se voit attribuer 1 point. La somme des scores mesure le risque de décrochage. Les scores vont de 0 à 3, et correspondent à un faible (0), modéré (1-2), et fort (3) risque de décrochage universitaire

13.29 14

14

Les variables synthétiques

5. Créer une variable synthétique par division (rapport)

- Calcul de l'indice de masse corporelle
 - Si on a des données sur le Poids en kg et la taille au m²
 - On applique la formule de Quételet

$$IMC = \text{Poids kg} / \text{Taille m}^2$$
 - Pour chaque cas, on divise son poids kg par sa taille m²
 - Valeurs métriques possibles: 15, 17, 20, 40, etc.
 - Qu'on peut recoder sous forme de catégories pour mieux évaluer le risque d'obésité ou d'anorexie
 - Dénutrition: moins de 17
 - Maigreur: 17-18,49
 - Normal: 18,5-24,9
 - Surpoids: 25-29,9
 - Obésité: 30 et plus

13.29 15

15

Les variables synthétiques

6. Créer une variable synthétique par moyenne

- Calcul de l'indice de pauvreté vécue
 - Sur la base de 5 variables de nécessités mesurées selon une échelle ordinale de type Likert: 0.Jamais, 1.Juste une ou deux fois, 2.Quelques fois, 3.Plusieurs fois, 4.Toujours
 - Manque de nourriture
 - Manque d'eau potable
 - Manque de soins de santé
 - Manque de combustible pour la cuisson
 - Manque d'argent en espèces
 - La moyenne des scores est appropriée pour mesurer l'indice de pauvreté vécue. L'indice s'étend 0 à 4, avec recodage en termes d'intensité: 0-1 (pauvreté nulle), 1,20-2 (pauvreté modérée), 2,2-4 (pauvreté extrême)

13,29

16

16

Les variables synthétiques

7. Créer une variable synthétique par combinaison

- Deux variables quali à combiner (lorsqu'elles ont un effet multiplicateur sur un phénomène social)
 - Age et Sexe sur le temps passé sur Internet
 - On combine les catégories deux à deux
 - Sexe selon groupe d'âge (Sexe X Âge)
 - ✓ 1.Homme de -25ans
 - ✓ 2.Homme de 25ans+
 - ✓ 3.Femme de -25ans
 - ✓ 4.Femme de 25ans+
 - Autres ex: effet de l'alcool et de la drogue sur la présence de troubles mentaux

13,29

17

17

Au programme

- | | |
|--|---|
| <ul style="list-style-type: none">■ De la nécessité du recodage de variables<ul style="list-style-type: none">❖ 5 stratégies de recodage de variables pour des raisons technico-pratiques et théoriques■ Création de variables synthétiques<ul style="list-style-type: none">❖ 7 stratégies pour passer des indicateurs empiriques à l'indice statistique ou à une seule variable composite | <ul style="list-style-type: none">■ Création de tableaux synthétiques<ul style="list-style-type: none">❖ Synthétiser les résultats d'une analyse statistique de plusieurs variables dans un seul tableau compact■ Création de graphes synthétiques<ul style="list-style-type: none">❖ Synthétiser les résultats d'une analyse statistique de plusieurs variables dans un seul graphe compact |
|--|---|

18

18

Les tableaux synthétiques

Décrire les caractéristiques de l'échantillon

■ Caractéristiques des participants à l'étude (collégiens: n=60)

	Fréquence	%
Sexe		
Garçon	31	52
Fille	29	48
Age		
12-13 ans	7	18
14-16 ans	35	58
17-18 ans	18	30
Position dans la famille		
Cadet	19	32
Aîné	19	32
Autre	22	37
Vivre avec ses parents		
Oui	53	88
Non	7	12
Résidence		
Ndiawdoune	30	50
Makhana	30	50

Source. Étude de cas, Ndèye Faye (2022)

13:29

19

Les tableaux synthétiques

Décrire des variables mesurées de façon identique

■ Évaluation des incidences socioéconomiques de la SCL par les villageois (en %)

– 10 variables d'évaluation

Indices socio-économiques	D'accord	Plutôt d'accord	Plutôt en désaccord	Désaccord	Ne sait pas
Construction ou réfection de salles de classe	77,8	15,6	00	6,6	00
Construction ou réfection de case de santé	2,2	00	28,9	68,9	00
Construction ou réfection de lieux de culte	31,1	22,2	26,7	17,8	2,2
Promotion de l'emploi féminin	66,7	24,4	24,4	2,2	00
Relégation de l'agriculture au second plan par la SCL	4,4	4,4	29	62,2	00
Enclavement du village par la SCL	48,9	24,4	15,6	11,1	00
Accaparement des terres du village par la SCL	80	17,8	00	2,2	00
Relations conflictuelles SCL et agriculteurs	6,7	6,7	24,4	56,6	6,6
Relations conflictuelles entre la SCL et les éleveurs	66,7	17,8	6,7	6,7	2,1
Accès à l'eau potable	6,7	2,2	24,4	62,2	4,5

Source. Étude de cas, Papa Guitté Diop (2022)

13:29

20

Les tableaux synthétiques

Analyse bivariable de relations entre plusieurs variables (chi-2)

■ Changement de l'activité économique selon le changement dans les habitudes de vie et la configuration familiale (n=60)

Changement sur le plan social	Changement de l'activité économique		Chi ²	V de Cramer
	Oui	Non		
<i>Habitude de vie</i>			4,10*	0,26
Oui	52%	21%		
Non	48%	79%		
Cas(n)	(46)	(14)		

Source. Étude de cas, Paule Evelyne Tendeng (2022)

13:29

*p<0.05.

21

21

Les tableaux synthétiques

Analyse bivariée de relations entre plusieurs variables (test F)

■ Indice d'empowerment sociopolitique selon la scolarité et le statut des tâches chez les entrepreneures (n=40)

Groupes	Statistiques				
	Moyenne	Écart-Type	Cas (n)	Test F	Eta Carré
Niveau de scolarité de la femme				3,64*	0,28
Pas d'éducation formelle	2,05	0,63	09		
Niveau primaire/Second	5,60	1,91	21		
Bac ou plus	4	0,70	10		

Source: Étude de cas, Ramatoulaye Dabo (2022) **p<0.01; *p<0.05.

13:29 22

22

Au programme

■ De la nécessité du recodage de variables

- ❖ 5 stratégies de recodage de variables pour des raisons technico-pratiques et théoriques

■ Création de variables synthétiques

- ❖ 7 stratégies pour passer des indicateurs empiriques à l'indice statistique ou à une seule variable composite

■ Création de tableaux synthétiques

- ❖ Synthétiser les résultats d'une analyse statistique de plusieurs variables dans un seul tableau compact

■ Création de graphes synthétiques

- ❖ Synthétiser les résultats d'une analyse statistique de plusieurs variables dans un seul graphe compact

23

23

Les graphes synthétiques

Graphe compact portant sur une catégorie (%)

■ Changements suite au relogement des sinistrés (en %)

- Activité (oui,non)
- Habitude (oui, non)
- Famille (oui,non)

77

45

28

Activité économique

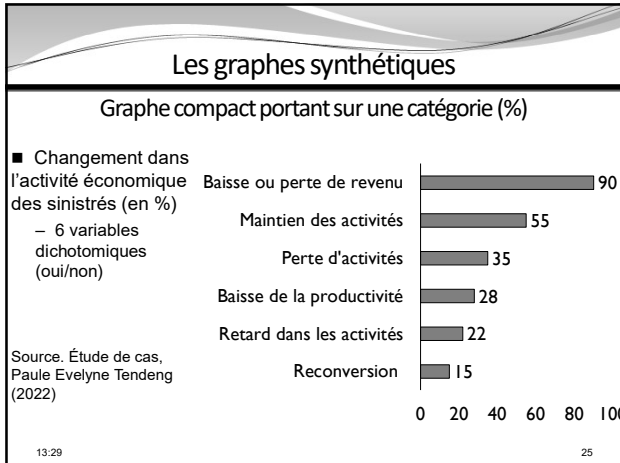
Habitude de vie

Configuration familiale

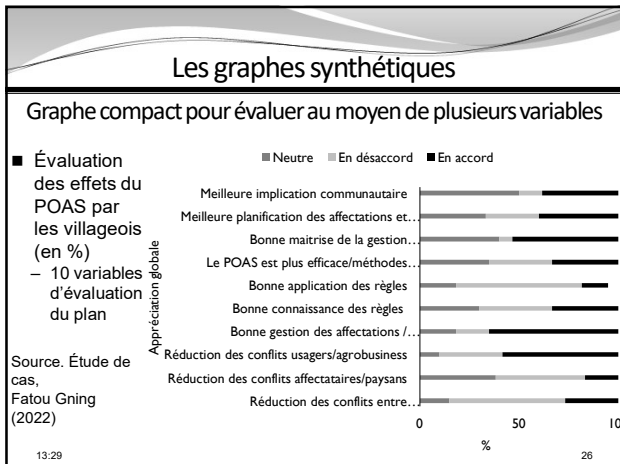
Source: Étude de cas, Paule Evelyne Tendeng (2022)

13:29 24

24



25



26

Tout prochainement

- Prochaine leçon
 - De l'analyse bivariée. Régression linéaire avec prédicteur qualitatif
- Au labo SPSS d'aujourd'hui
 - Recoder des variables
 - Créer des variables synthétiques
 - Créer des graphes synthétiques
 - Créer des tableaux synthétiques
 - Quelques analyses bivariées en guise de révision

13:29 27

27
