

MIASS 231.1
 MATHÉMATIQUES (APPLIQUÉES AUX SCIENCES
 SOCIALES) 3

© El Hadj Touré, Ph D. Sociologie
 Section de sociologie, UGB de St-Louis

Leçon 3
**Distributions de fréquences et de
 pourcentages**

5:35 1

1

Au programme

- Comment décrire les données d'une variable à l'aune des fréquences et % de façon à en faciliter l'analyse?
 - Reconnaître les conditions d'application et les situations de recherche où l'utilisation de ces techniques statistiques est requise
 - Transformer des données brutes en distributions de fréquences simples et regroupées
 - Transformer des fréquences en %, puis en % cumulatifs
 - % de variation comme mesure de rapport
 - Représenter les distributions de fréquences/% par des tableaux bien construits et des graphiques appropriés et les interpréter
- Remue-ménages & exercices « éclair »

5:35 2

2

Distribution de fréquences

Questions de recherche sociologique (univariées)

→ Questions relatives à l'ampleur d'un phénomène

- Quelle est l'opinion des citoyens américains concernant la désobéissance civile ? (Fox)
- Les travailleurs de la fonction publique sénégalaise sont-ils âgés?
- La religion est-elle aussi importante en FR qu'au CAN?
- La représentativité féminine est-elle en hausse dans les universités sénégalaises?
- Quel est l'intention de vote des Sénégalais en faveur du parti au pouvoir aux prochaines élections de 2024?

5:35 3

3

Distribution de fréquences

Définition & classification

→ Une façon d'y répondre consiste à compter le nombre de cas pour chaque valeur. « Ce résumé de la variation d'une variable est une distribution de fréquences » (Fox: 34)

- **Distribution de fréquences simples**
 - Fréquences avec les valeurs brutes | de départ
 - Ex: pour la variable «sexe», le nombre de cas correspondant aux valeurs catégorielles de départ «femme» ou «homme»
- **Distribution de fréquences regroupées**
 - Fréquences avec les valeurs regroupées en classes ou catégories
 - Ex: pour un échantillon n=100 cas, des classes d'âge peuvent être créées: 20-39ans; 40-59ans; 60 ans +

5:35 4

4

Distribution de fréquences simples

	Numéro du cas	Désobéissance	Numéro du cas	Désobéissance
n=30	01	Suivre sa conscience	16	Suivre sa conscience
	02	Suivre sa conscience	17	Suivre sa conscience
	03	Obéir aux lois	18	Suivre sa conscience
	04	Obéir aux lois	19	Obéir aux lois
	05	Suivre sa conscience	20	Obéir aux lois
	06	Obéir aux lois	21	Obéir aux lois
	07	Obéir aux lois	22	Suivre sa conscience
	08	Suivre sa conscience	23	Suivre sa conscience
	09	Suivre sa conscience	24	Suivre sa conscience
	10	Suivre sa conscience	25	Suivre sa conscience
	11	Suivre sa conscience	26	Suivre sa conscience
	12	Obéir aux lois	27	Suivre sa conscience
	13	Obéir aux lois	28	17 Suivre sa conscience
	14	Obéir aux lois	29	Obéir aux lois
	15	Obéir aux lois	30	13 Obéir aux lois

5:35 Source: Fox (p.35), Données tirées du Général Social Survey américain (1996) ⁵

5

Distribution de fréquences simples

Tableau de distribution univariée

Répartition des répondants selon l'opinion concernant la désobéissance civile (en fréquences)

Désobéissance civile	Fréquence (f)
Suivre sa conscience	17
Obéir aux lois	13
Total (n)	30

Source: Fox, Données tirées du Général Social Survey américain (1996)

→ L'analyse est maintenant plus facile à faire puisque la variable est agrégée à partir des microdonnées

5:35 6

6

Distribution de fréquences regroupées

Justification & procédure statistique

- Avec un petit nombre de scores, une mise en rang peut suffire pour visualiser la forme de la distribution
 - Ex: pour un échantillon de 5 personnes (n=5), l'âge peut donner la distribution : 20, 22, 25, 30, 50 ans
- Mais, avec un grand nombre de scores (n=100), la mise en forme apparaît si les fréquences sont regroupées :
 1. Calculer l'étendue de la distribution
 2. Déterminer le nombre de classes
 3. Définir l'intervalle et les limites des classes
 4. Calculer les fréquences pour les classes ainsi créées

5:35 7

7

Distribution de fréquences regroupées

1. Calculer l'étendue de la distribution (E)

$E = X_{\max} - X_{\min}$ Différence entre score maximal et score minimal

Distribution d'âge de 30 cas

25	33	40	43	46	
25	34	40	43	47	$E = 54 - 25 = 29$
26	34	40	44	47	
31	34	41	44	51	
31	36	41	44	52	
32	39	41	44	54	

Les microdonnées sont fictives et n'ont qu'une valeur illustrative!!!

5:35 8

8

Distribution de fréquences regroupées

2. Déterminer le nombre de classes (k)

- Décision arbitraire, mais raisonnée de préférence
- Idéalement, le choix dépend des orientations de l'étude
 - Ex.1: Étude sur les problèmes générationnels
 - ✦ 3 à 4 classes d'âges d'un intervalle de 20 ans suffisent
 - Ex.2: Étude sur les inégalités en termes de revenu
 - ✦ Jusqu'à 20 classes pour mieux observer les écarts possibles
- En pratique, le choix obéit à la règle de Sturges
 - Postulat: la taille de l'échantillon détermine le nombre de classes
 - Formule: $k = 1 + 3,3^{*}(\log_{10}(n))$

5:35 9

9

Distribution de fréquences regroupées

2. Déterminer le nombre de classes (k)

■ Représentation visuelle et formule mathématique de Sturges pour déterminer le nombre de classes

$k = 1 + \log_2(n)$

$k = 1 + 3.3 \log_{10}(n)$

→ $\log_{10}(100) = 2$

→ $\log_{10}(1000) = ?$

$k = 1 + 3.3 \log_{10}(30)$

$k = 1 + 3.3(1.477)$

$k = 5.9$

5:35 10

10

Distribution de fréquences regroupées

3. Définir l'intervalle (i_c) et les limites des classes

A. Déterminer l'intervalle (longueur) de classes

$$i_c = \frac{\text{Étendue } E}{\text{Nombre de classes } k}$$

B. Définir les limites des classes (inférieure et supérieure)

- Chaque classe commence idéalement par un multiple de la longueur de classe, le multiple le plus proche
- La 1^{ère} classe inclut le score min., la dernière le score max.

→ Ex: $i_c = 29/6 = 4.8 \rightarrow$ Il nous faut 6 classes d'intervalle 5
 [25-30]; [30-35]; [35-40]; [40-45]; [45-50]; [50-54]

5:35

11

11

Distribution de fréquences regroupées

4. Calculer les fréquences des classes (décompte)

Répartition des répondants selon les classes d'âge

Autres notations		Classes d'âge	Fréquences
(25-29)	$25 \leq X < 30$	[25;30[3
(30-34)	$30 \leq X < 35$	[30;35[7
-	$\leq X <$	[35;40[2
-	$\leq X <$	[40;45[12
-	$\leq X <$	[45;50[3
-	$\leq X <$	[50;55[3
Total (n)			30

Les microdonnées sont fictives et n'ont qu'une valeur illustrative!!!

5:35

12

12

Distribution de fréquences regroupées

Quelques remarques

- Plus souvent qu'autrement, on fusionne les valeurs numériques d'une variable quantitative
 - Pour pouvoir utiliser les distributions de fréquences et %
- On peut aussi fusionner les valeurs catégorielles d'une variable qualitative de façon à en réduire le nombre
 - Ex: fusionner les catégories de «religion» en 1.Chrétien et 0.Autre ou bien isoler 1.Protestant et 0.Catholique (Max Weber, 1904)
- Quoi qu'il en soit, lorsqu'on fusionne les valeurs d'une variable, il en résulte une perte d'informations
 - Mais on obtient une distribution plus informative

5:35 13

13

Distribution de pourcentages

Difficultés liées aux fréquences (1)

→ Lorsque n est grand, la compréhension d'une distribution de fréquences pose problème

L'importance de la religion au Canada (en fréquences)

Importance of religion	Frequencies
Very	459
Rather	456
Not very	371
Not at all	199
Total (n)	1485

Source: World Values Surveys and European Values Surveys, 1995–1997

5:35 14

14

Distribution de pourcentages

Difficultés liées aux fréquences (2)

→ Lorsque les n sont vraiment différents, la comparaison des distributions de fréquences pose problème

L'importance de la religion au Canada et en France (fréquences)

Importance of religion	Canada	France
Very	459	137
Rather	456	283
Not very	371	274
Not at all	199	289
Total (n)	1485	983

Source: World Values Surveys and European Values Surveys, 1995–1997

5:35 15

15

Distribution de pourcentages

Fréquences standardisées ou relatives

→ On peut surmonter ces difficultés si on rapporte les fréquences absolues à un nombre de cas égal à 100. Cette standardisation donne une distribution de pourcentages

L'importance de la religion au Canada (en fréquences et %)

Importance of religion	Frequencies	Percents
Very	459	30,9
Rather	456	30,7
Not very	371	25,0
Not at all	199	13,4
Total (n)	1485	100,0

5:35 16

16

Distribution de pourcentages

Rappel du calcul des %

1. Diviser chaque fréquence f par le nombre total de cas n
2. Multiplier ce résultat par la base de comparaison **100**

$$\text{Pourcentage} = \frac{f}{n} (100)$$

→ Exemple de la désobéissance civile (Fox)

- Des 30 cas sondés, 17 croient qu'il faut suivre sa conscience:
17/30 (100) = 56,7%
- Les 13 autres cas croient plutôt qu'il faut obéir aux lois:
13/30 (100) = 43,3%

5:35 17

17

Distribution de pourcentages

Exercice « éclair »

→ Comment est calculée la distribution de pourcentages selon l'importance de la religion en France?

Importance/religion	Canada		France	
	f	%	f	%
Very	459	31	137	14
Rather	456	31	283	29
Not very	371	25	274	28
Not at all	199	13	289	29
Total (n)	1485	100	983	100

5:35 18

18

Distribution de pourcentages

Proportions, %, taux & ratio: exercice éclair

- Une étude porte sur la représentativité féminine en socio à l'UGB: homme (79) ; femme (75); n = 154 (L2 2021)
 - Proportion de femmes = $\frac{75}{154} = 0,49$
 - Pourcentage de femmes = $0,49 \times 100 = 49\%$
 - Taux de féminité (100) = 49%
 - Ratio femmes: hommes = $\frac{75}{79} = 0.95$
- Selon la Banque mondiale, avec 15 420 000 habitants, le Sénégal compte 1 066 médecins (public) en 2017. Quel est le taux de médecins et le ratio au Sénégal?
 - Taux de médecins/100 000 = $(\frac{1066}{15\,420\,000}) \times 100\,000 = 6,9$
 - Ratio d'habitants par médecin = $(\frac{15\,420\,000}{1\,066}) = 14\,465$

19

Distribution de pourcentages

% de variation: mise en situation & exercice éclair

- Selon la Banque mondiale, le PIB/hbt du Sénégal est passé de 617\$ en 2000 à 1 318\$ en 2012. Quel est le % d'augmentation du PIB/hbt de 2000 à 2012?
- Formule du % de variation (augmentation ou diminution)

$$\% \text{ de variation} = \frac{\text{Valeur à un tps donné (T2)} - \text{Valeur au tps de référence (T1)}}{\text{Valeur au tps de référence (T1)}} \times 100$$
- Calcul $\% \text{ de variation} = \frac{1318 - 617}{617} \times 100 = 114\%$
- Quel est le % d'augmentation du PIB/hbt de 2012 à 2020, si en 2020 le PIB/hbt s'élève à 1 488\$? 13%

20

Distribution cumulative

Illustration

→ Pour la distribution ci-dessous, quel est le % de cas qui ont un diplôme de 1^{er} cycle ou moins?

Niveau d'instruction	Pourcentages	
Pas de secondaire	1%	79%
Secondaire	6%	
Collégial	20%	
1er cycle universitaire	52%	
2 ^e cycle universitaire	19%	
3 ^e cycle universitaire	2%	
Total	100%	

Source: Touré, sondage TÉLUQ, été 2011

21

Distribution cumulative

Calcul

→ Les distributions cumulatives F (fréquences ou % de tous les cas ayant un score donné ou inférieur) sont utiles pour les variables ordinales ou quantitatives

Niveau d'instruction	Pourcentages	Calcul	Pourcentages cumulés
Pas de secondaire	1%	1%	1%
Secondaire	6%	+6%	7%
Collégial	20%	+20%	27%
1er cycle univ.	52%	+52%	79%
2e cycle univ.	19%	+19%	98%
3e cycle univ.	2%	+2%	100%

5:35 22

22

Représentation tabulaire

Présentation d'un tableau univarié & interprétation

Tableau 1. Répartition des répondants américains selon l'opinion concernant la désobéissance civile (n=30)

Désobéissance	f	%
Suivre sa conscience	17	57
Obéir aux lois	13	43
Total	30	100

Source: Fox, Données tirées du Général Social Survey américain (1996)

- ✓ **Interprétation statistique (analyse):** Que disent les chiffres? (%/valeurs) Que suggèrent les chiffres (conclusion/variable)?
- ✓ **Interprétation théorique (sociologique):** Comment expliquer la conclusion? (éléments théoriques ou factuels, bon sens)

5:35 23

23

Représentation graphique

Aperçu des graphiques univariés

→ On retient plus facilement de l'information lorsqu'elle est présentée à l'aide de techniques graphiques. Pour ce faire, regardez le modèle d'ensemble et les variations autour...

Variables

Nominale

Diagramme en secteurs (peu de catégories)

Diag. en barres horizontales (8 catégories+)

Ordinale

Diagramme en barres verticales

Discrète

Diagramme en bâtons (linéaires)

Courbe de tendance (temps)

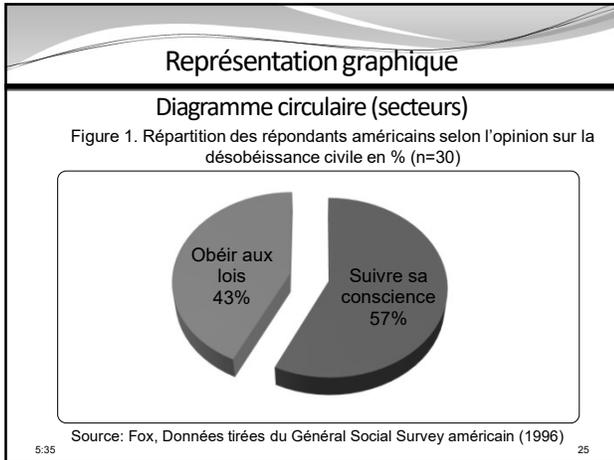
Continue

Histogramme (en classes)/polygone

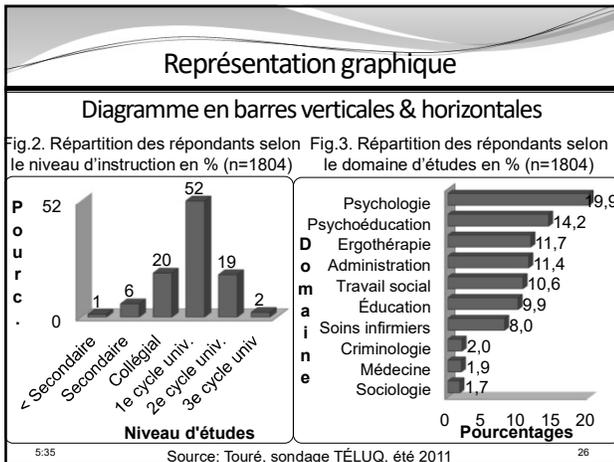
Ogive (limite sup. classes)

5:35 24

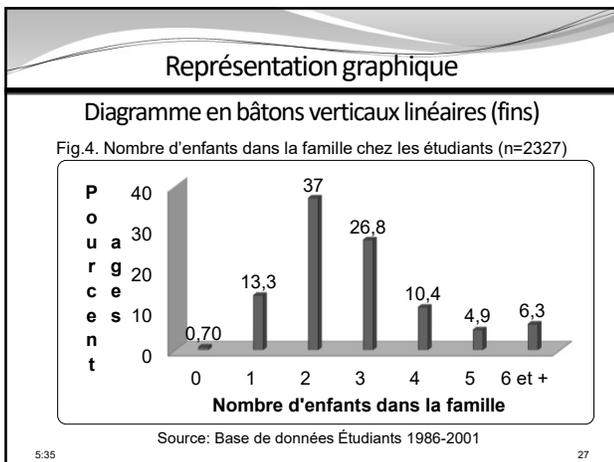
24



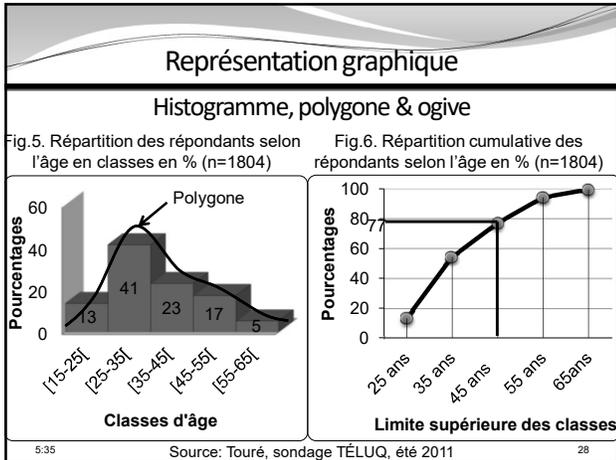
25



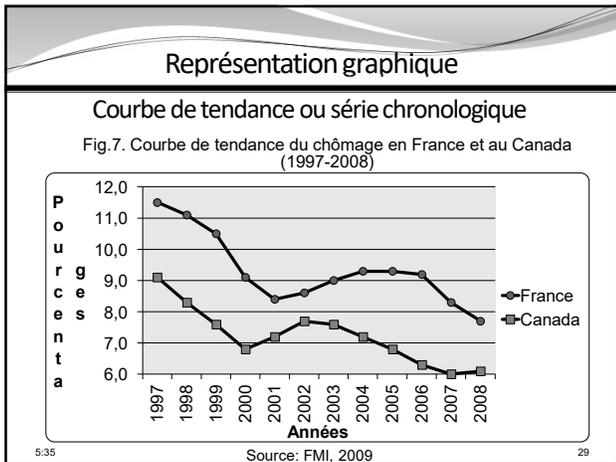
26



27



28



29

Tout prochainement

- À faire cette semaine
 - Compléter le quiz 3
 - Effectuer les exercices récapitulatifs (avec solutionnaire)

- Prochaine leçon
 - Mesures de tendance centrale

5:35 30

30
