

Seance 3.2: Distribution de fréquences et de pourcentages

Discussion en classe

Visseho Adjiwanou, PhD.

Département de Sociologie - UQAM

25 January 2023

Plan de présentation

- ➊ Revenir sur les notions de niveau de mesure des variables (30 minutes)
- ➋ Discussion sur les fréquences et les pourcentages (1 heure 10)
- ➌ Pause (10 minutes)
- ➍ Labo classe (1h)

Information: Organisation des fichiers sur Moodle

- Vous avez plusieurs fichiers sur Moodle qui peuvent vous mélanger.
- Tous les fichiers sont nommés: Il y a trois fichiers qui sont nommés:
 - Séance_chapitre.1: 1 à ma présentation de ce qui est dans le livre de Fox

Information: Organisation des fichiers sur Moodle

- Vous avez plusieurs fichiers sur Moodle qui peuvent vous mélanger.
- Tous les fichiers sont nommés: Il y a trois fichiers qui sont nommés:
 - Séance_chapitre.1: 1 à ma présentation de ce qui est dans le livre de Fox
 - Séance_chapitre.2: 2 fait référence à la discussion qu'on va avoir en classe. Tout ce qu'il y a dans le livre ne sera pas forcément couvert, mais l'essentiel et les plus importants. Cela viendra compléter vos lectures

Information: Organisation des fichiers sur Moodle

- Vous avez plusieurs fichiers sur Moodle qui peuvent vous mélanger.
- Tous les fichiers sont nommés: Il y a trois fichiers qui sont nommés:
 - Séance_chapitre.1: 1 à ma présentation de ce qui est dans le livre de Fox
 - Séance_chapitre.2: 2 fait référence à la discussion qu'on va avoir en classe. Tout ce qu'il y a dans le livre ne sera pas forcément couvert, mais l'essentiel et les plus importants. Cela viendra compléter vos lectures
 - Séance_livre.3: 3 fait référence à quelque chose qui est en développement. Il s'agit du chapitre du livre que je suis en train d'écrire

Résumé des notions

Symboles	Signification	Formules
N	Nombre total de cas / taille de l'échantillon	
f	Fréquence = nombre de cas pour un attribut	
F	Fréquence cumulative = somme des fréquences inférieures ou égales à une valeur donnée	$\sum f$
P	Proportion de l'échantillon	$\frac{f}{N}$
%	Pourcentage	$\frac{f}{N} * 100$
pc	Pourcentage cumulatif	$\frac{F}{N} * 100$
Ratio	La fréquence relative de cas dans différentes population	$Ratio \frac{f_1}{f_2}$

Résumé des notions

- Fréquences
 - Tous les niveaux de mesure (nominale, ordinale, intervalle/ratio)
- Fréquence cumulative
 - Normalement variable ordinale, intervalle/ratio

Résumé des notions

- Données manquantes
 - Non applicable, ne sait pas, Sans opinion, Incapable de choisir, Pas de réponse, Refus de répondre, Incertain

Résumé des notions

- Données manquantes
 - Non applicable, ne sait pas, Sans opinion, Incapable de choisir, Pas de réponse, Refus de répondre, Incertain
 - Le cas aussi dès fois de “autres”

Résumé des notions

- Données manquantes
 - Non applicable, ne sait pas, Sans opinion, Incapable de choisir, Pas de réponse, Refus de répondre, Incertain
 - Le cas aussi dès fois de “autres”
 - Exception: Quand “NSP” et “Sans opinion” ont une signification conceptuelle ou théorique, on ne devrait pas les exclure de l'analyse

Résumé des notions

- Données manquantes
 - Non applicable, ne sait pas, Sans opinion, Incapable de choisir, Pas de réponse, Refus de répondre, Incertain
 - Le cas aussi dès fois de “autres”
 - Exception: Quand “NSP” et “Sans opinion” ont une signification conceptuelle ou théorique, on ne devrait pas les exclure de l'analyse
 - Exemple:

Résumé des notions

- Données manquantes
 - Non applicable, ne sait pas, Sans opinion, Incapable de choisir, Pas de réponse, Refus de répondre, Incertain
 - Le cas aussi dès fois de “autres”
 - Exception: Quand “NSP” et “Sans opinion” ont une signification conceptuelle ou théorique, on ne devrait pas les exclure de l'analyse
 - Exemple:
 - Non applicable n'est pas forcément donnée manquante, on peut le recoder

Résumé des notions

- Données manquantes
 - Non applicable, ne sait pas, Sans opinion, Incapable de choisir, Pas de réponse, Refus de répondre, Incertain
 - Le cas aussi dès fois de “autres”
 - Exception: Quand “NSP” et “Sans opinion” ont une signification conceptuelle ou théorique, on ne devrait pas les exclure de l'analyse
 - Exemple:
 - Non applicable n'est pas forcément donnée manquante, on peut le recoder
 - Exemple

Résumé des notions

- Données manquantes
 - Non applicable, ne sait pas, Sans opinion, Incapable de choisir, Pas de réponse, Refus de répondre, Incertain
 - Le cas aussi dès fois de “autres”
 - Exception: Quand “NSP” et “Sans opinion” ont une signification conceptuelle ou théorique, on ne devrait pas les exclure de l'analyse
 - Exemple:
 - Non applicable n'est pas forcément donnée manquante, on peut le recoder
 - Exemple
- Cas déviants: le cas où les score peuvent biaiser l'analyse statistique

Résumé des notions

- Représentations graphiques
 - diagrammes circulaires (à privilégier quand le nombre de modalités est peu nombreux, cas des variables nominales)

Résumé des notions

- Représentations graphiques
 - diagrammes circulaires (à privilégier quand le nombre de modalités est peu nombreux, cas des variables nominales)
 - diagrammes en bâtons

Résumé des notions

- Représentations graphiques
 - diagrammes circulaires (à privilégier quand le nombre de modalités est peu nombreux, cas des variables nominales)
 - diagrammes en bâtons
 - Cartographie (variable écologiques)

Exemple numérique

Distribution de fréquences

Question 1: Est-ce que les gens devraient obéir aux lois sans exception ou suivre leur conscience même si cela demandait de violer la loi?

Question 2: Quel est ton sexe?

Question 3: Quel est ton niveau d'éducation?

Question 4: Quel est ton âge?

Distribution de fréquences

Sexe : Nominale

- Homme = 12 (f1)
- Femme = 13 (f2)
- Autre = 2 (f3)

$$N = 12 + 13 + 2 = 27$$

Distribution de fréquences

Niveau de scolarité: Ordinale

- Sans instruction = 6
- Secondaire = 7
- Collégial = 9
- Université ou plus = 5

Distribution de fréquences

Age: ratio

- 25 = 2
- 26 = 3
- 27 = 4
- 28 = 2
- 29 = 1
- 30 = 5
- 31 = 3
- 32 = 4
- 33 = 2
- 34 = 1

Distribution de fréquences

Désobéissance : Nominale

- Suivre sa conscience = 21
- Obéir aux lois = 3
- NSP = 3

Proportion

	proportion
Suivre sa conscience	$21/27 = 0.78$
Obéir aux lois	$3/27 = 0.11$
NSP	$3/27 = 0.11$

Pourcentage

	proportion	pourcentage
Suivre sa conscience	$21/27 = 0.78$	$0.78*100 = 77.8\%$
Obéir aux lois	$3/27 = 0.11$	$0.11*100 = 11.1\%$
NSP	$3/27 = 0.11$	$0.11*100 = 11.1\%$

Pourcentage valide

	pourcentage	Pourcentage valide
Suivre sa conscience	$0.78 \times 100 = 77.8\%$	$21/24 \times 100 = 87.5\%$
Obéir aux lois	$0.11 \times 100 = 11.1\%$	$3/24 \times 100 = 12.5\%$
NSP	$0.11 \times 100 = 11.1\%$	

$$\text{Échantillon valide} = 27 - 3 = 24$$

Distribution de la variable niveau de scolarité

	f	proportion	pourcentage
Université	5	$5/27 = 0.19$	$5/27*100 = 18.51$
Collège	9	$9/27 = 0.33$	$9/27*100 = 33.33$
Secondaire	7	$7/27 = 0.26$	$7/27*100 = 25.93$
Sans éducation	6	$6/27 = 0.22$	$6/27*100 = 22.22$

Distribution de la variable niveau de scolarité

	pourcentage	Pourcentage cumulatif
Université	$5/27 \cdot 100 = 18.51$	$18.51 + 33.33 + 25.93 + 22.22 =$
Collège	$9/27 \cdot 100 = 33.33$	$33.33 + 25.93 + 22.22 = 81.48$
Secondaire	$7/27 \cdot 100 = 25.93$	$25.93 + 22.22 = 48.15$
Sans éducation	$6/27 \cdot 100 = 22.22$	22.22

Distribution de la variable niveau de scolarité

- Il est fréquent que les modalités de la variables soient ordonnées en sens inverse: du plus petit au plus grand, du moins important au plus important. . .
- Cela ne change rien à l'interprétation des résultats.

	po	urcentage Po	urcentage cumulatif
Sans éducation	6/27*100 = 22.22	22.22	
Secondaire	7/27*100 = 25.93	22.22 + 25.93 = 48.15	
Collège	9/27*100 = 33.33	48.15 + 33.33 = 81.48	
Université	5/27*100 = 18.51	81.48 + 18.51 = 100	

Distribution de la variable niveau de scolarité

Interprétation

- 100% de l'échantillon ont le niveau universitaire ou moins (ou 100% de l'échantillon ont au plus le niveau universitaire)

Distribution de la variable niveau de scolarité

Interprétation

- 100% de l'échantillon ont le niveau universitaire ou moins (ou 100% de l'échantillon ont au plus le niveau universitaire)
- 81.48% de l'échantillon ont le niveau collège ou moins

Distribution de la variable niveau de scolarité

Interprétation

- 100% de l'échantillon ont le niveau universitaire ou moins (ou 100% de l'échantillon ont au plus le niveau universitaire)
- 81.48% de l'échantillon ont le niveau collège ou moins
- 48.15% de l'échantillon ont le niveau secondaire ou moins

Distribution de la variable niveau de scolarité

Interprétation

- 100% de l'échantillon ont le niveau universitaire ou moins (ou 100% de l'échantillon ont au plus le niveau universitaire)
- 81.48% de l'échantillon ont le niveau collège ou moins
- 48.15% de l'échantillon ont le niveau secondaire ou moins
- Quelle est la proportion des gens qui ont le niveau collège et plus?

Distribution de la variable niveau de scolarité

Interprétation

- 100% de l'échantillon ont le niveau universitaire ou moins (ou 100% de l'échantillon ont au plus le niveau universitaire)
- 81.48% de l'échantillon ont le niveau collège ou moins
- 48.15% de l'échantillon ont le niveau secondaire ou moins
- Quelle est la proportion des gens qui ont le niveau collège et plus?
- Peut-on calculer le pourcentage cumulatif pour les variables nominales?

Distribution de la variable Age

- Dresser le tableau de distribution de la variable âge.

De la mauvaise utilisation des statistiques (tiré de Krieg)

- **Pourcentage sur de petits échantillons:** Méfiez-vous des statistiques “époustouflantes” ! Les journaux, les magazines et autres médias font souvent état de changements rapides dans la société.

De la mauvaise utilisation des statistiques (tiré de Krieg)

- **Pourcentage sur de petits échantillons:** Méfiez-vous des statistiques “époustouflantes” ! Les journaux, les magazines et autres médias font souvent état de changements rapides dans la société.
- Par exemple, nous pouvons lire dans le journal local que le taux d'homicide de notre communauté a augmenté de 50 % par rapport à l'année précédente.

De la mauvaise utilisation des statistiques (tiré de Krieg)

- **Pourcentage sur de petits échantillons:** Méfiez-vous des statistiques “époustouflantes” ! Les journaux, les magazines et autres médias font souvent état de changements rapides dans la société.
- Par exemple, nous pouvons lire dans le journal local que le taux d'homicide de notre communauté a augmenté de 50 % par rapport à l'année précédente.
- Cela peut être une source d'inquiétude ou non.

De la mauvaise utilisation des statistiques (tiré de Krieg)

- **Pourcentage sur de petits échantillons**

De la mauvaise utilisation des statistiques (tiré de Krieg)

- **Pourcentage sur de petits échantillons**
- Si le nombre de meurtres à la même époque l'année dernière était de deux, une augmentation de 50 % signifie que notre communauté a connu trois meurtres cette année.

De la mauvaise utilisation des statistiques (tiré de Krieg)

- **Pourcentage sur de petits échantillons**
- Si le nombre de meurtres à la même époque l'année dernière était de deux, une augmentation de 50 % signifie que notre communauté a connu trois meurtres cette année.
- Passer de deux à trois n'est pas une grande augmentation (du moins dans la plupart des villes), et faire état d'une augmentation de 50 %, bien que statistiquement vrai, revient à tromper les consommateurs de ces statistiques en leur faisant croire que le problème est plus important qu'il ne l'est réellement.

De la mauvaise utilisation des statistiques (tiré de Krieg)

- **Spécificité sociale et géographique des statistiques.**
Souvent, les statistiques qui sont censées représenter une société entière ne représentent qu'une partie de la population.

De la mauvaise utilisation des statistiques (tiré de Krieg)

- **Spécificité sociale et géographique des statistiques.**
Souvent, les statistiques qui sont censées représenter une société entière ne représentent qu'une partie de la population.
- Par exemple, nous pouvons lire que le nombre de meurtres commis depuis le début de l'année est le double de celui de l'année dernière à la même époque.

De la mauvaise utilisation des statistiques (tiré de Krieg)

- **Spécificité sociale et géographique des statistiques.**
Souvent, les statistiques qui sont censées représenter une société entière ne représentent qu'une partie de la population.
- Par exemple, nous pouvons lire que le nombre de meurtres commis depuis le début de l'année est le double de celui de l'année dernière à la même époque.
- Cela ne signifie pas pour autant que tout le monde court deux fois plus de risques d'être assassiné.

De la mauvaise utilisation des statistiques (tiré de Krieg)

- **Spécificité sociale et géographique des statistiques.**
Souvent, les statistiques qui sont censées représenter une société entière ne représentent qu'une partie de la population.
- Par exemple, nous pouvons lire que le nombre de meurtres commis depuis le début de l'année est le double de celui de l'année dernière à la même époque.
- Cela ne signifie pas pour autant que tout le monde court deux fois plus de risques d'être assassiné.
- Il se peut que la majorité des meurtres aient lieu dans des villes de 250 000 habitants ou plus.

De la mauvaise utilisation des statistiques (tiré de Krieg)

- **Spécificité sociale et géographique des statistiques.**

De la mauvaise utilisation des statistiques (tiré de Krieg)

- **Spécificité sociale et géographique des statistiques.**
- Il est donc possible que l'augmentation globale du nombre de meurtres dans ces zones urbaines soit à l'origine de la hausse du taux de meurtres à l'échelle nationale, alors que d'autres types de communautés pourraient voir leur taux de meurtres diminuer.

De la mauvaise utilisation des statistiques (tiré de Krieg)

- **Autres exemple:** Il pourrait en être de même pour tout type de phénomène social - enlèvements d'enfants, fugues, viols, etc.

De la mauvaise utilisation des statistiques (tiré de Krieg)

- **Autres exemple:** Il pourrait en être de même pour tout type de phénomène social - enlèvements d'enfants, fugues, viols, etc.
- Voir statistiques sur els homicides au Canada: <https://www150.statcan.gc.ca/t1/tbl1/fr/tv.action?pid=3510007101>

De la mauvaise utilisation des statistiques (tiré de Krieg)

Attitudes à tenir devant les statistiques

- Soyez vigilant.es: Il est important de savoir quels types de questions poser lorsque nous entendons ces statistiques “époustouflantes”.

De la mauvaise utilisation des statistiques (tiré de Krieg)

Attitudes à tenir devant les statistiques

- Soyez vigilant.es: Il est important de savoir quels types de questions poser lorsque nous entendons ces statistiques “époustouflantes”.
- Quelles sont les fréquences sur lesquelles ces statistiques sont basées ?

De la mauvaise utilisation des statistiques (tiré de Krieg)

Attitudes à tenir devant les statistiques

- Soyez vigilant.es: Il est important de savoir quels types de questions poser lorsque nous entendons ces statistiques “époustouflantes”.
- Quelles sont les fréquences sur lesquelles ces statistiques sont basées ?
- À quelles populations ces statistiques peuvent-elles être généralisées ?

De la mauvaise utilisation des statistiques (tiré de Krieg)

Attitudes à tenir devant les statistiques

- Soyez vigilant.es: Il est important de savoir quels types de questions poser lorsque nous entendons ces statistiques “époustouflantes”.
- Quelles sont les fréquences sur lesquelles ces statistiques sont basées ?
- À quelles populations ces statistiques peuvent-elles être généralisées ?
- Et, ce qui est peut-être le plus important, qui a intérêt à communiquer ce type de statistiques ?

Étude de cas: Participation à la société québécoise

Exemple 1: Mesurer la participation des Québécoises et Québécois des minorités ethnoculturelles

Objectifs de l'étude:

- ❶ Décrire la participation des minorités ethnoculturelles dans 7 dimensions
 - Dimension 1: Économique
 - Dimension 2: Communautaire
 - Dimension 3: Culturelle
 - Dimension 4: Linguistique
 - Dimension 5: Citoyenne
 - Dimension 6: Identitaire
- ❷ Comparer la participation des minorités ethnoculturelles avec celle de la population majoritaire

Dimension économique - définition des concepts

- Personnes occupées sont celles qui exercent un emploi ou qui exploitent une entreprise

Dimension économique - définition des concepts

- Personnes occupées sont celles qui exercent un emploi ou qui exploitent une entreprise
- Chômeurs sont les personnes qui sont sans emploi, qui sont prêtes à travailler et qui cherchent activement du travail.

Dimension économique - définition des concepts

- Personnes occupées sont celles qui exercent un emploi ou qui exploitent une entreprise
- Chômeurs sont les personnes qui sont sans emploi, qui sont prêtes à travailler et qui cherchent activement du travail.
- Population active : Personnes occupées + Chômeurs

Dimension économique - définition des concepts

- Personnes occupées sont celles qui exercent un emploi ou qui exploitent une entreprise
- Chômeurs sont les personnes qui sont sans emploi, qui sont prêtes à travailler et qui cherchent activement du travail.
- Population active : Personnes occupées + Chômeurs
- Personnes inactives sont celles qui, considérant les conditions qui existaient sur le marché du travail dans leur région, ne désiraient ou ne pouvaient offrir ou fournir leurs services au cours de la semaine de référence (ceci comprend les personnes qui étaient actuellement aux études à temps plein).

Dimension économique - définition des concepts

- Personnes occupées sont celles qui exercent un emploi ou qui exploitent une entreprise
- Chômeurs sont les personnes qui sont sans emploi, qui sont prêtes à travailler et qui cherchent activement du travail.
- Population active : Personnes occupées + Chômeurs
- Personnes inactives sont celles qui, considérant les conditions qui existaient sur le marché du travail dans leur région, ne désiraient ou ne pouvaient offrir ou fournir leurs services au cours de la semaine de référence (ceci comprend les personnes qui étaient actuellement aux études à temps plein).
- La période de référence utilisée dans l'EIACA pour la situation d'emploi était la date de l'interview. La question suivante a été posée aux répondants : « Quelle est votre situation actuelle

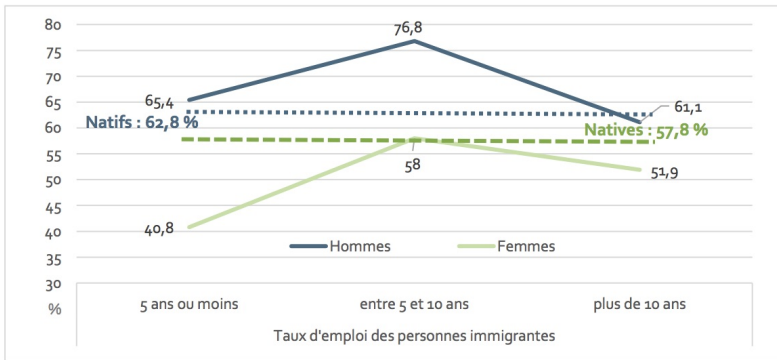
Dimension économique - définition des concepts

- Personnes occupées sont celles qui exercent un emploi ou qui exploitent une entreprise
- Chômeurs sont les personnes qui sont sans emploi, qui sont prêtes à travailler et qui cherchent activement du travail.
- Population active : Personnes occupées + Chômeurs
- Personnes inactives sont celles qui, considérant les conditions qui existaient sur le marché du travail dans leur région, ne désiraient ou ne pouvaient offrir ou fournir leurs services au cours de la semaine de référence (ceci comprend les personnes qui étaient actuellement aux études à temps plein).
- La période de référence utilisée dans l'EIACA pour la situation d'emploi était la date de l'interview. La question suivante a été posée aux répondants : « Quelle est votre situation actuelle

Exemple 1: Mesurer la participation des Québécoises et Québécois des minorités ethnoculturelles

- Décrivez ce graphique

GRAPHIQUE 4 : TAUX D'EMPLOI SELON LA DURÉE DE RÉSIDENCE PAR SEXE, 2015



Source : Enquête sur la population active, 2015

Visseho Adjwanou, PhD.

Department of Sociology - UQAM

Seance 3.2: Distribution de fréquences et de pourcentages

Exemple 1: Mesurer la participation des Québécoises et Québécois des minorités ethnoculturelles

“Les immigrants masculins participent au marché du travail avec un taux d’emploi dépassant celui des hommes natifs. Pour les femmes immigrantes, le taux d’emploi dépasse très légèrement celui des femmes natives chez celles résidant depuis 5 à 10 ans au Québec, mais demeure inférieur avant et après.” Laur, P. 19)

- http://www.midi.gouv.qc.ca/publications/fr/recherches-statistiques/RAP_Mesure_participation_2016.pdf

Pour la semaine prochaine

Séance 4 : Les mesures de tendance centrale

- ① Lectures
 - Fox: chapitre 3, pp.69-90
- ② Labo 4 - voir Moodle