MIASS 231
MATHÉMATIQUES
(APPLIQUÉES AUX SCIENCES SOCIALES) 3

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

Leçon 5

Mesures de variation

09:21 1

1

Mesures de variation

Illustration en lien avec la moyenne: Adapté de Fox (p. 92)

	Groupe A Relativement homogène	Groupe B Entre les deux	Groupe C Relativement hétérogène
	Écart faible	Écart moyen	Écart élevé
	64	44	34
	66	56	46
	68	63	58
	69	74	79
	70	80	82
	71	82	90
	82	91	101
Moyenne	70	70	70

2

Au programme

- Comment décrire les données d'une variable quantitative à l'aide des mesures de variation et les représenter?
  - Les mesures de variation: définition, calcul, interprétation
    - Étendue et écart interquartile
    - Variance et écart-type
    - Coefficient de variation
  - Les formes de distribution: asymétrie et aplatissement
  - Les scores-z: définition, calcul, interprétation
  - La distribution normale et ses applications pratiques
- Remue-méninges & exercices « éclair »

09:21 3

3

Mesures de variation

Justification & définition

- Les mesures de tendance centrale ne rendent pas compte de la dispersion des scores d'une distribution
- Les mesures de variation aident à en rendre compte en vue d'apprécier la variabilité d'un phénomène
  - Elles résument comment les scores sont dispersés les uns des autres et par rapport à la tendance centrale
  - Elles indiquent le degré d'homogénéité d'un groupe
- Les scores-z permettent de positionner des individus et représenter un phénomène à l'aide de la loi normale

09:21 4

4

Mesures de variation

Questions de recherche univariées

1. Quelle est la variation du nombre d'heures passées devant la TV chez les Sénégalais âgés de 18 ans et plus?
2. Les disparités dans les revenus des médecins sont-elles plus importantes au Québec qu'en Russie?
3. Les salaires sont-ils plus homogènes chez les femmes que chez les hommes dans la fonction publique sn.?
4. Du point de vue de leur performance, comment les élèves d'une école se situent les uns par rapport aux autres?
5. À l'échelle du continent africain, comment varie le taux de scolarisation d'un pays à l'autre?

09:21 5

5

Étendue

$$E = X_{max} - X_{min}$$

→Quelle est la variation des notes du groupe A? (n=7)

64 66 68 69 70 71 82

- E = 82 – 64 = 18
- Interprétation statistique: Dans le groupe A, la distance qui sépare les notes max. et min. est de 18 points en % (min=64; max=82)

- D'une part, malgré la facilité de son calcul, l'étendue n'est appropriée que pour des échantillons de petites tailles
- D'autre part, sa valeur n'est déterminée que par deux scores; aussi est-elle sensible aux valeurs extrêmes

09:21 6

6

Intervalle interquartile

- Étendue sur laquelle repose la moitié centrale (au moins 50%) autour de la médiane d'une distribution

IQ = Q3 – Q1

→ Dans l'exemple précédent : 64 66 $\underbrace{\overbrace{68\ 69\ 70}^{\text{Q1 Md Q3}}}$ 71 82

- Méthode des médianes partielles: Q1=67 ; Q3=70,5
- IQ = 70,5 – 67 = 3,5
- Interprétation statistique: Au moins 50% des notes du groupe A s'étendent de 67 à 70,5, soit un écart interquartile de 3,5 points

- L'intervalle interquartile est approprié en cas d'asymétrie
- Toutefois, il ne dépend que de deux mesures de position: il ne comporte pas d'information sur les autres scores

Variance & écart-type

Scores déviation

- Les mesures de variation les plus utilisées exploitent, informent sur l'ensemble des scores d'une distribution
- Ceci est rendu possible grâce aux scores déviation D_i

$$D_i = X_i - \bar{X}$$

- Un score déviation (D_i) renvoie à la différence ou l'écart entre un score (X_i) et la moyenne comme référence ($\bar{X}$)
  - Des cas éloignés de la moyenne auront des scores déviation élevés; des cas rapprochés auront des scores déviation bas
  - Des scores inférieurs ou supérieurs à la moyenne auront respectivement des scores déviation négatifs ou positifs

Variance & écart-type

Somme des carrés (des scores déviation)

- La somme des scores déviation est nulle: $\Sigma(X_i - \bar{X}) = 0$
  - Pour cette raison, elle ne peut servir comme mesure sommaire de la dispersion d'une distribution
- Pour rendre compte de la dispersion, on élève au carré les scores déviation avant de les sommariser:
  - On obtient ainsi la somme des carrés: $\Sigma(X_i - \bar{X})^2$
- La somme des carrés est un minimum: $\Sigma(X_i - \bar{X})^2$ Min
  - Mais elle est influencée par le nombre d'individus (et donc les différences entre les scores individuels et la moyenne)

Variance

Définition et formules

- La variance est la moyenne de la somme des carrés
- La variance d'une population est la somme des carrés divisée par la taille de la population ***N***

$$\sigma^2 = \frac{\Sigma (X_i - \mu)^2}{N}$$

- La variance d'un échantillon (estimé non biaisé de N) se calcule en divisant la somme des carrés par ***n*** - 1

$$s^2 = \frac{\Sigma (X_i - \bar{X})^2}{n - 1} \quad \begin{matrix}\text{Somme des carrés} \\ \text{Degrés de liberté}\end{matrix}$$

09:21 10

10

Variance

Variance d'un échantillon: Exemple (Adapté de Fox: 97)

$(X_i - \bar{X})$	$(X_i - \bar{X})^2$	*Groupe A* Relativement homogène	*Groupe B* Entre les deux	*Groupe C* Relativement hétérogène
64-70=-6	36	64	44	34
66-70=-4	16	66	56	46
68-70=-2	4	68	63	58
69-70=-1	1	69	74	79
70-70= 0	0	70	80	82
71-70= 1	1	71	82	90
82-70=12	144	82	91	101
Σ 0	202			
Moyenne		70	70	70
Variance = 202 / (7-1) =		33,7	270,3	600,3

→ Interprétation statistique: La variance du groupe A étant de 33,7 les scores (notes) divergent par rapport à la moyenne de 33,7 (% au carré)

09:21 11

11

Écart-type

Justification & formules

- La variance est exprimée dans une unité de mesure différente de celle des scores bruts (ex.: pour la taille, la variance s^2 est en cm^2 et les scores-X_i en cm)
- En statistique descriptive, on lui préfère l'écart-type, car il est ramené à la même échelle que les scores originaux
  - L'écart-type d'une population

$$\sigma = \sqrt{\sigma^2}$$

  - L'écart-type d'un échantillon

$$s = \sqrt{s^2}$$

09:21 12

12

Écart-type

Calcul, interprétation statistique & remarques

- Pour l'exemple précédent, voici les écarts-types :
  - $S_1 = \sqrt{33,7} =$ 5,8
  - $S_2 = \sqrt{270,3} =$ 16,4
  - $S_3 = \sqrt{600,3} =$ 24,5
- Interprétation statistique: L'écart-type du groupe A est de 5,8. Environ 2/3 des notes (68%) se situent à ± 5,8 de la moyenne 70, soit entre 64,2 (70-5,8) et 75,8 (70+5,8)
- Choisir ***s*** pour décrire la variabilité d'une variable quanti., ***s²*** étant réservée aux statistiques inférentielles
- Comme la moyenne, ils subissent l'effet des cas déviants

09:21 13

13

Coefficient de variation relative

Définition

- Pour évaluer la dispersion d'une distribution, de façon plus précise, il faut comparer l'écart-type à la moyenne

$$CV = (s / \overline{X}) *100$$

  - L'écart-type divisé par la moyenne est multiplié par 100
  - Si CV est ≤ à 15%, la distribution est homogène
- Le CV indique le degré d'homogénéité d'une distribution: plus il est faible, plus les données sont homogènes
- Approprié pour comparer des écarts-types/moyennes de distributions avec des échelles de mesure différentes

09:21 14

14

Coefficient de variation relative

Exercice « éclair »

- On désire comparer les revenus des médecins de la Russie à ceux du Québec (Simard: 97)

Nation	Revenu moyen	Écart-type du revenu	Coefficient de variation
Russie	1623 roubles	65 roubles	? **4,0%**
Québec	115 600 $	20 567 $	? **17,8%**

  - À quel endroit observe-t-on une plus grande disparité dans les revenus et quelle est la distribution la plus homogène?
- Interprétation statistique: Les CV sont de 17,8% (Québec) et 4% (Russie). Le Québec présente les revenus les plus disparates (CV≥15%) et c'est plus homogène en Russie

09:21 15

15

Mesures de variation

Scores standardisés ou scores-Z: Mise en situation

- Qui de ces deux étudiants d'une université a la meilleure note (sur 100) par rapport à sa classe ?

Étudiant	Note	Moyenne	Écart-type	Score-z
Mamad	87	81	6	**1**
Khady	83	76	4	**1,75**

- Note standardisée de Mamad: (87- 81) / 6 = 6/6 = 1,00
- Note standardisée de Khady: (83 – 76) / 4 = 7/4 = 1,75

- Interprétation stat.: Les scores-z respectifs de Khady et de Mamad sont de 1,75 et 1. La note de Khady est meilleure: elle se situe à 1,75 écart-type au-dessus de la moyenne de sa classe (précisément 7 pts en % au dessus de la moy.)

09:21 16

16

Scores standardisés ou scores-z

Définition, formule & propriétés

- Pour mesurer à combien d'écarts-types de la moyenne se situe un score donné, on utilise les scores-z

Groupe A Scores X_i	$\frac{(X_i - \overline{X})}{s}$	Scores Z_i
64	64-70/5,8	-1,03
66	66-70/5,8	-0,69
68	68-70/5,8	-0,34
69	69-70/5,8	-0,17
70	70-70/5,8	0
71	71-70/5,8	0,17
82	82-70/5,8	2,07
$\overline{X} = 70$ s = 5,8		$\overline{X} = 0$ s = 1

$$Z = \frac{(X_i - \overline{X})}{s}$$

Scores-z= Centrer chacun des scores-Xi à la moyenne et réduire la différence par l'écart-type s

Propriétés des scores-z

- Compris en général entre -3 et +3
- Leur moyenne est toujours égale à 0
- Leur écart-type est toujours égal à 1

09:21 17

17

Scores standardisés ou scores-z

Exercice « éclair »

- Un employeur veut engager un étudiant. Il examine les dossiers de 4 étudiants pour embaucher le meilleur. Doit-il se fier uniquement à la note absolue?

	Note	Moyenne du groupe	Écart type du groupe	$X_i - \overline{X}$	$(X_i-\overline{X})/s = Z_i$
Mamad	85	75	10	85-75	10/10= 1
Fatou	76	70	3	76-70	6/3 = 2
Khady	70	60	4	70-60	10/4= 2,5
Moussa	75	80	5	75-80	-5/5 = -1

- Il ne peut se fier à la note… Les scores-z aident à classer les candidats: Khady, Fatou, Mamad et Moussa

09:21 18

18

Scores standardisés ou scores-z

Avantage & inconvénient: illustration

- Avantage: Utilisation des scores-z dans le système éducatif québécois pour sélectionner les étudiants
  - Mettre à plat les différences artificielles liées à la notation
- Inconvénient: Lequel des deux génies aura le meilleur score-z? Celui du groupe A ou du groupe B?

	Groupe A	Groupe B
2 génies	100	100
2 bons étudiants	80	80
2 étudiants moyens	60	60
1 faible, 1 cancre	50	0

  - Ne considère pas les différences structurelles inhérentes aux groupes eux-mêmes

09:21 19

19

Distribution normale

Définition et lien avec les scores-z

- Ce qu'est une distribution normale:
  - Une distribution d'une variable aléatoire continue dont la courbe est symétrique, unimodale (sous forme de cloche)
  - La distribution normale est omniprésente en statistiques, notamment inférentielles, en raison de ses propriétés
  - Une distribution normale dépend de la moyenne μ et de l'écart-type σ et est notée N (μ, σ)
- Quel lien avec les scores-z?
  - Les scores-z permettent de standardiser les scores X_i d'une distribution normale et de les représenter
  - Une distribution normale centrée réduite (standardisée) à l'aide des scores-z est notée N (0, 1)

09:21 20

20

Distribution normale

Exemples de variables

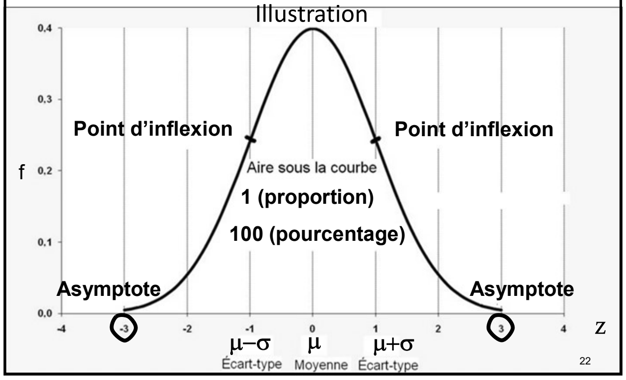
- Plusieurs variables se distribuent naturellement en adéquation avec la loi normale
  - Mesures physiques: poids, taille, pointure…
  - Mesures cognitives: résultats d'examens, cotes de rendement…
  - Mesures psychomotrices: temps de réaction, vitesse…
  - Mesures psychologiques: échelles de personnalité, stress…
  - Mesures psychométriques: QI, habiletés verbales…
- Toutefois, toutes les variables d'intérêt en sciences sociales ne suivent pas nécessairement la loi normale
  - Revenu (classe sociale)
  - Scolarité (niveau d'instruction)

09:21 21

21

Distribution normale

Illustration



22

22

Distribution normale

Caractéristiques

1. Il existe une infinité de courbes normales (μ, σ);
2. Elle est unimodale, symétrique et mésokurtique;
3. Ses points d'inflexion sont à $\mu \pm \sigma$ (concave à convexe)
4. Elle est asymptotique à ses extrémités;
5. Elle atteint son sommet à la valeur μ (μ=Mo=Md);
6. Plus on s'éloigne de X = μ, plus la hauteur de la courbe diminue, plus le pourcentage de données est faible;
7. L'aire totale sous la courbe est de 1 ou 100%
8. Les scores-z sont <u>en général</u> compris entre -3 et +3

09:21 23

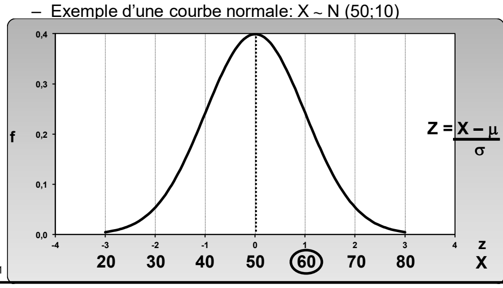
23

Distribution normale

Propriété 1: distribution des scores

→ Connaissant μ et σ, on peut décrire complètement la distribution des scores bruts (X) et standardisés (z).

– Exemple d'une courbe normale: X ~ N (50;10)



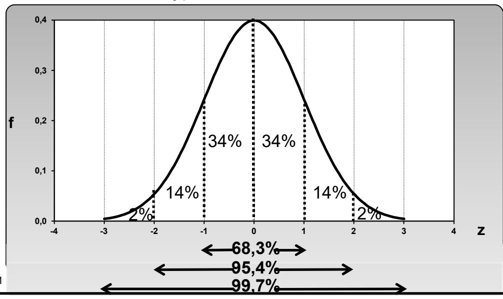
09:21 24

24

Distribution normale

Propriété 2: pourcentage de données

→ On peut trouver le pourcentage de données comprises entre deux écarts-types ou < ou > à une valeur donnée



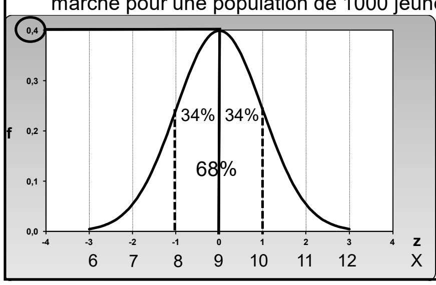
09:21 25

25

Distribution normale

Application pratique : solution à un problème

→ Un fabriquant veut connaître le nombre de chaussures de pointures (X) 7, 8, 9, 10, 11 qu'il lui faut mettre sur le marché pour une population de 1000 jeunes



Moyenne : 9
Écart-type : 1
X ~ N (9 ; 1)

- ✓ 40% auront la pointure 9
- ✓ 34% auront une pointure située entre 8 et 9
- ✓ 68% auront une pointure située entre 8 et 10

26

Tout prochainement

- À faire cette semaine
  - Complétez le quiz 5 (obligatoire)
  - Réalisez les 10 exercices récapitulatifs et vérifiez les solutions en vous référant au corrigé (optionnel)
- Prochaine séance (Labo Excel 1)
  - Initiation au tableur Excel : réaliser quelques opérations de calcul de base, manipuler les cellules de calcul, comprendre ce que sont les références absolues et relatives. Utilisation d'Excel pour créer et interpréter une table d'indice de masse corporelle, représenter visuellement les échelles de mesure des variables

09:21 27

27