

Leçon 4 Mesures de tendance centrale

06:39

1

1

Mesures de tendance centrale

Justification en lien avec les distributions de f et %

Répartition de 30 répondants selon les classes d'âge

Classe s d'âge	Fréquen- ces	Poucen- tages
[25;30[3	10
[30;35[7	23
[35;40[2	7
[40;45[12	40
[45;50[3	10
[50;55[3	10
Total (n)	30	100

06:39

2

Mesures de tendance centrale

Justification & définition

- Contrairement aux distribut. de fr. et de %, les mesures de tendance centrale donnent des résumés numériques précis
- Si les fréq. et % décrivent l'ampleur d'un phénomène, ces mesures décrivent la représentativité d'un phénomène
- Les trois mesures les plus rapportées sont :
 - le mode
 - la médiane
 - la moyenne
- Elles répondent à la question de recherche: quel est le score typique, le plus commun d'une distribution?

06:39

3

Au programme

- Comment décrire les données d'une variable à l'aide des mesures de tendance centrale ?
 - Justifier, reconnaître les situations de recherche et déterminer les conditions de leur utilisation adéquate
 - Définir et calculer les mesures de tendance centrale
 - Savoir les interpréter de façon appropriée
 - Saisir les effets des cas déviants sur ces mesures
 - Reconnaître les types de distribution d'une variable : unimodale, bimodale, symétrique, asymétrique
- Remue-méninges & exercices « éclair »

06:39

4

Mesures de tendance centrale

Questions de recherche univariées

1. Quel est le nombre « typique » d'heures passées devant la TV chez les Sénégalais?
2. Quelle est la famille « typique » en termes de nombre d'enfants chez les Sénégalais?
3. Quel est le salaire mensuel le plus « représentatif » chez les joueurs de l'équipe sénégalaise de football?
4. À l'échelle du continent africain, quel est le taux de scolarisation le plus « commun » entre les pays?

06:39

5

Le mode

Définition & détermination

- Valeur qui présente la fréquence la plus élevée dans une distribution donnée (Mo): «c'est la valeur à la mode»
- Seule mesure de tendance centrale qui convient aux variables nominales
 - Ex: le sexe modal au Sénégal est le féminin
- Quoiqu'il puisse être calculé pour toutes les variables, il ne convient pas aux variables aux nombreuses valeurs
 - Ex: nombre d'habitants des villes sénégalaises
- Il n'existe aucune formule pour son calcul, mais on peut se servir d'un arrangement, graphique ou tableau

06:39

6

Le mode

Types de distribution: caractérisation

→ Nombre d'heures passées devant la TV/semaine (n=10)

- Distribution **unimodale**
 - Echantillon A: 9, 11, 14, 15, 15, 15, 16, 17, 18, 19
- Distribution **bimodale**
 - Echantillon B: 9, 11, 14, 15, 15, 16, 17, 18, 18, 20
- Distribution **plurimodale**
 - Echantillon C: 9, 9, 11, 14, 15, 15, 16, 17, 18, 18
- Distribution **sans mode**
 - Echantillon D: 9, 11, 13, 14, 15, 16, 17, 18, 19, 20

06:39

7

Le mode

Types de distribution: illustration de la forme

<p>Sans mode</p> <p>Nbre d'habitants par ville</p>	<p>Unimodale</p> <p>Revenu</p>
<p>Bimodale</p> <p>Revenu</p>	<p>Plurimodale</p> <p>Revenu</p>

8

Le mode

Exercice «éclair» & interprétation statistique

→ Quel est le nombre d'enfants typique modal désiré... ?

Nombre d'enfants désiré chez les étudiants

→ Le mode est égal à 2. Une **pluralité** d'étudiants, soit 37%, désirent avoir une famille constituée de deux enfants

06:39

9

La médiane

Définition

- Mesure de tendance centrale qui divise en deux parties égales un ensemble ordonné de scores (**Md**)
- 50^e centile (C_{50}) d'une distribution
 - Les centiles C_i divisent le nombre de scores en 100 parties égales, chacun contenant 1% des scores
- 2^e quartile (Q_2) d'une distribution
 - Les quartiles Q_i divisent le nombre de scores en 4 parties égales, chacun contenant $\frac{1}{4}$ des scores

X_{\min} --- 25%|1/4 --- $(C_{25}; Q_1)$ --- 25%|1/4 --- **Md** --- 25%|1/4 --- $(C_{75}; Q_3)$ --- 25%|1/4 --- X_{\max}

06:39

10

La médiane

Calcul: données brutes

→ Si les données brutes sont disponibles et ordonnées:

- Avec **n impair** (n=5)
 - Score central d'une distribution $((n+1)/2^e$ score)
 - Salaire horaire de 5 personnes en \$: 31 35 **39** 45 50
Md= 39 soit le 3^e score en rang
- Avec **n pair** (n=6)
 - Point milieu des deux scores centraux $((n+1)/2^e$ score)
 - Salaire horaire de 6 personnes en \$: 30 34 **38 42** 49 55
Md=(38+42)/2 = 40 soit le 3,5^e score en rang

06:39

11

La médiane

Exercices «éclair» & interprétation statistique

- Quel est le revenu mensuel typique médian, en Fcfa, d'une micro-entreprise constituée de 7 travailleurs ?
82000 100000 35000 45500 55000 65000 72000
→ Commencez toujours par ordonner les scores!
35000 45500 55000 **65000** 72000 82000 100000
→ La médiane est 65000. Dans cette entreprise, au moins 50% des travailleurs, soit 4/7, gagnent 65000 Fcfa ou moins
- Que seraient le mode et la médiane de cette distribution?
35000 44500 **45000 55000** 72000 82000

06:39

12

La médiane

Données agrégées: estimation par l'ogive

→ Avec les fréq. regroupées, on perd les données brutes et leur ordre. L'ogive aide à estimer approximativement la Md.

Âge	f	F
[20-30[5	5
[30-40[15	20
[40-50[17	37
[50-60[36	73
[60-70[19	92
[70-80[8	100
Total (n)	100	2

06:39

13

La médiane

La médiane et les mesures de position: interprétation

- $Q2 = C50 = Md \approx 53$
 - Au moins 50% des cas de cet échantillon sont âgés approximativement de 53 ans ou moins
- $Q1 = C25 \approx 43$
 - Au moins 25% des cas de cet échantillon sont âgés approximativement de 43 ans ou moins
- $Q3 = C75 \approx 62$
 - Au moins 75% des cas de cet échantillon sont âgés approximativement de 62 ans ou moins

06:39

14

La médiane

Quelques remarques!

- Il est possible d'utiliser la formule de l'interpolation pour obtenir une valeur précise et raffinée de la médiane dans le cas de données agrégées (regroupées)
- Toutefois, on peut se contenter de l'estimation par l'ogive de Md, à défaut d'utiliser les données brutes (micro)
- La médiane s'applique surtout aux variables quantitatives mais aussi aux variables ordinales
- Contrairement au mode, la médiane tient compte certes de l'ordre inhérent aux scores d'une variable quantitative ou ordinale, mais elle ne tient pas compte du poids ou de la grandeur des scores d'une distribution

06:39

15

La moyenne

Formule de définition et de calcul

- La moyenne est la somme de tous les scores, relativisée (ou divisée) par le nombre d'individus d'une distribution

Données d'échantillon	Données de population
Statistique $\bar{X} = \frac{\sum X_i}{n}$	Paramètre $\mu = \frac{\sum X_i}{N}$

\bar{X} et μ = moyennes respectives d'un échantillon et d'une population
 X_i = score X du $i^{\text{ème}}$ individu
 Σ = lettre grecque « sigma » qui signifie sommation
n = nombre de cas de l'éch. / N = nombre d'individus de la pop.

- La moyenne s'applique aux variables quantitatives

06:39

16

La moyenne

Exemples

- Avec données brutes

→ Nombre d'enfants désiré chez 10 cas: 0 4 1 2 3 4 4 1 2 4

$\bar{X} = (0+4+1+2+3+4+4+1+2+4) / 10 = 2,5$ enfants par femme

- Avec données agrégées (fréquences simples)

Nbre d'enfants X_i	Fréquences f_i	$X_i * f_i$	
0	1	0	$\bar{X} = 25/10$ $= 2,5$
1	2	2	
2	2	4	$\bar{X} = \frac{\sum X_i f_i}{n}$
3	1	3	
4	4	16	
Total (n)	<u>10</u>	<u>25</u>	<u>n</u>

06:39

17

La moyenne

Moyenne pondérée: exercice

- Supposons qu'un étudiant de la classe ait obtenu les notes suivantes sur 20 en mathématiques 3 :
Quiz: 10; TP: 15; Examen: 18.

Trouver sa moyenne pour le cours si la pondération de chaque évaluation est: Quiz 20%; TP 20%; Examen 60%.

Items & pondération	Résultat
Quiz20%	
TP20%	
Examen60%	
Total (moyenne)	

06:39

18

La moyenne

Moyenne combinée: exercice

- Calculer le salaire moyen des 10 employés d'une micro-entreprise agricole sachant que les 2 cadres gagnent 200000 Fcfa chacun, les 3 techniciens 100000 Fcfa chacun, et les 5 ouvriers agricoles 60000 Fcfa chacun.

Employés & nombre Salaire

- 2 cadres
- 3 techniciens
- 5 ouvriers agricoles
- Total (moyenne)

06:39

19

La moyenne

Propriétés: inconvéient

→ Soit la distribution de revenu en Fcfa de 7 employés
20000, 22000, 30000, 34000, 35000, 36000, **500000**

- La moyenne est égale à 96714
- La médiane est égale à 34000

- La moyenne est peu représentative de cette distribution, du fait du cas déviant (cas ayant 500000 Fcfa)
- Conséquence: au contraire de la médiane, la moyenne est affectée par les cas déviants, leurs scores aberrants
- Face à cette situation :
 - Évaluer l'effet des cas déviants sur la moyenne et la médiane
 - ✦ Soit choisir la médiane comme mesure de tendance centrale
 - ✦ Soit exclure le cas déviant pour éviter qu'il ne biaise la moyenne

20

La moyenne

Illustration des cas déviants : Fox (p.58)

2.3. Écoute quotidienne de la télévision

06:3

21

La moyenne

Propriétés: avantage

→ Soit la distribution: note de 6 étudiants du SOL1020-A11

Cas	Score X_i	Écart à la moyenne (D)	Calcul $D = (X_i - \bar{X})$
1	79,3	-7,0	79,3 - 86,3
2	88,3	2,0	88,3 - 86,3
3	93,8	7,5	93,8 - 86,3
4	96,1	9,8	96,1 - 86,3
5	81,7	-4,6	81,7 - 86,3
6	78,7	-7,7	78,7 - 86,3
\bar{X}	86,3	Somme = 0,0	

- La somme des écarts à la moyenne est toujours nulle
 $\sum (X_i - \bar{X}) = 0$
- Conséquence: la moyenne équilibre une distribution

06:39

22

La moyenne

Inconvénient & avantage: Remue-méninges

- Illustrons comment l'usage de la moyenne et de la médiane peut faire l'objet d'une instrumentalisation selon les acteurs en jeu et les enjeux du moment...
- Supposons qu'on est dans un processus de négociation des salaires entre le gouvernement sénégalais et le syndicat des employés de la fonction publique. Pour défendre ses positions,
 - lequel des deux camps est plus susceptible d'invoquer le salaire moyen des employés,
 - ou lequel va avoir plutôt tendance à invoquer le salaire médian des employés?

06:39

23

Présentation des mesures

Interprétation statistique & tableau des moyennes

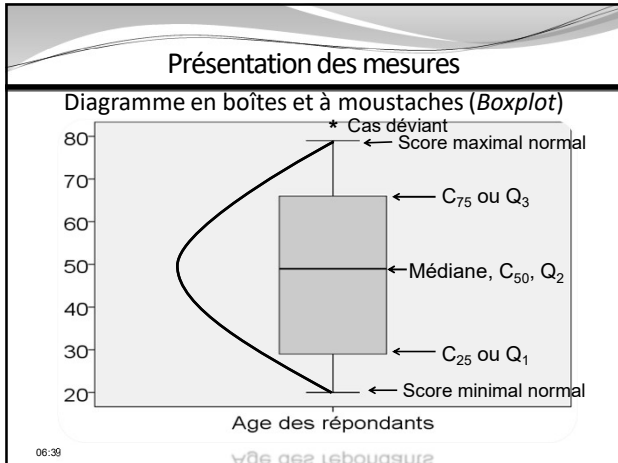
- Dans bien des cas, les mesures de tendance centrale sont présentées à l'intérieur du texte
 - Ex: La moyenne est égale à 2,5. Autrement dit, le nombre moyen d'enfants par femme est de 2,5 dans l'échantillon.
- Toutefois, il peut être nécessaire parfois de les présenter à l'aide d'un tableau des moyennes

Nombre d'enfants désiré chez les femmes et les hommes (n=10)

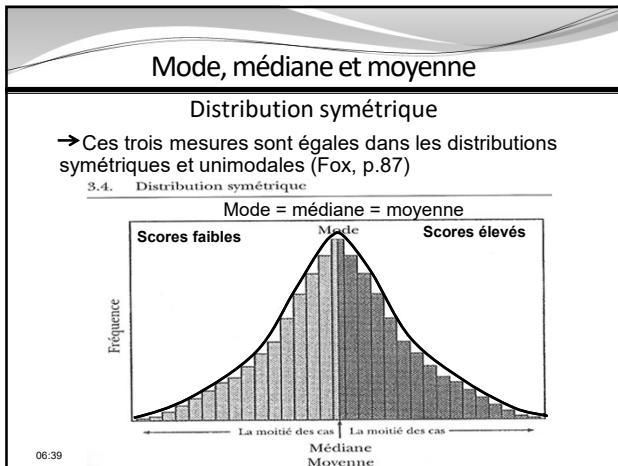
Nombre d'enfants	Sexe		Total
	Femme	Homme	
Moyenne	3	2	2,5
Nombre de cas (n)	5	5	10

06:39

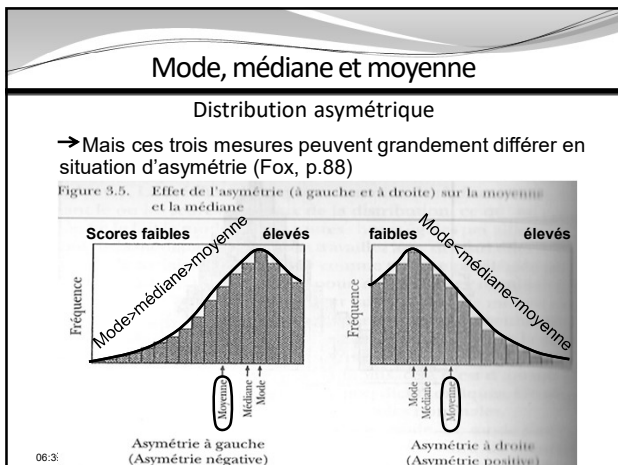
24



25



26



27

Mode, médiane et moyenne

Quelle mesure choisir? Comparaison

- **Mode**
 - Utilisé exclusivement à des fins descriptives (score réel)
 - Peut être utilisé pour toutes les variables, surtout nominales
 - « Avec la mode, il faut tenir compte du mode »
- **Médiane**
 - Utilisée principalement à des fins descriptives (seuil de coupure)
 - Indiquée pour les variables quantitatives surtout, ordinales aussi
 - Plus adaptée aux distributions asymétriques (cas déviants)
- **Moyenne**
 - Utilisée à la fois à des fins descriptives et inférentielles
 - Indiquée pour les variables quantitatives uniquement
 - Plusieurs techniques statistiques sont basées sur son calcul

06:39

28

Mode, médiane et moyenne

Quelle mesure choisir? Remue-méninges

- Un magasin qui vend des chaussures de femmes veut connaître les deux pointures d'une espadrille les plus représentatives dans la perspective de doubler les commandes pour ces deux pointures
- Un chercheur souhaite connaître le revenu le plus représentatif de la famille typique sénégalaise pour mieux rendre compte du niveau de vie ou du pouvoir d'achat standard des Sénégalais
- Un chef d'entreprise veut connaître le montant le plus représentatif qu'il a dépensé par employé dans le but d'avoir une meilleure idée sur la tendance centrale du coût salarial défrayé

06:39

29

Tout prochainement

- À faire cette semaine
 - Complétez le quiz 4 (obligatoire)
 - Réalisez les 10 exercices récapitulatifs et vérifiez les solutions en vous référant au corrigé (optionnel)
- Prochaine leçon
 - Mesures de variation

06:39

30
