Cours de Statistique Inférentielle

Jean Christophe Meunier

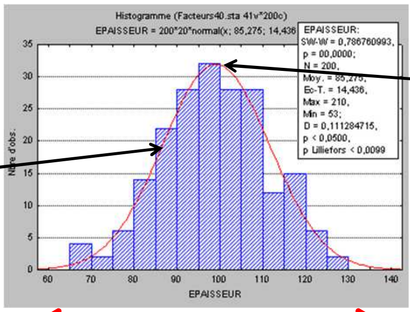
Rappel 1
Étude des séries statistiques : Tendance centrale et dispersion

2ème Bac, Commerce Extérieur

Année académique 2015-2016

EPFC

Indicateurs de la courbe de distribution



1. Position : Valeur qui occupe une position précise (ex. tendance centrale) ou qui découpe l'échantillon selon une proportion spécifiée

2. Dispersion : Dans l'absolu ou par rapport à la tendance centrale

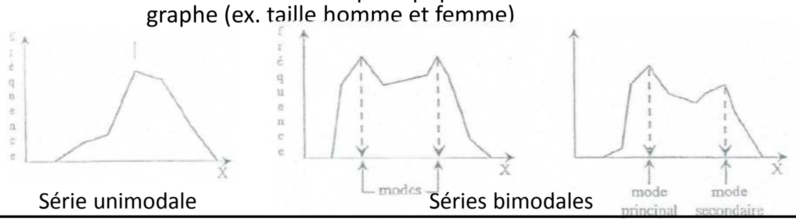
3. Forme : Asymétrie et aplatissement

4. Concentration : Ressource/argent : répartition des ressources au sein de population

A. Tendance centrale

I. Tendance centrale

1. Mode (x_0)
   - Valeur ou classe (x_i) de la série statistique dont l'effectif (n_i) est le plus élevé
   - Déterminé via table/graphe des effectifs
     - Repérer x_i dont n_i est le plus élevé
     - Données groupées : on parle de 'classe modale'
   - Un vs. Plusieurs modes
     - Unimodale : un seul 'pic' d'effectifs
     - Bimodale : deux 'pics' d'effectifs
       - Peut-être un indice que 2 populations ≠ sont considérées sur le même graphe (ex. taille homme et femme)



Série unimodale

Séries bimodales

4

I. Tendance centrale

2. Médiane ($\tilde{x}$)

– Valeur ou classe (x_i) qui 'coupe' l'échantillon en deux parties égales
  - Les effectifs des valeurs > et < à la valeur médiane sont égaux (à N/2 : 50%< et 50%>)
  - Si N est impair (2p + 1)
    – Une seule valeur se situe exactement à la moitié de l'échantillon : Médiane = (p+1)^{ème} valeur
    – <u>Ex</u> série statistique impaire (N=9) :
      » 1 1 2 2 **<u>2</u>** 3 4 5 5 → Médiane = 2 ((p+1)^{ème} valeur))
  - Si N est pair (2p)
    – Deux valeurs se situent 'à cheval' sur la moitié de l'échantillon : Médiane = moyenne de p^{ème} valeur et de (p+1)^{ème} valeur
    – <u>Ex</u> série statistique paire (N=10) :
      » 1 1 2 2 **<u>2 3</u>** 3 4 5 5 → Médiane = (2+3)/2 = 2,5

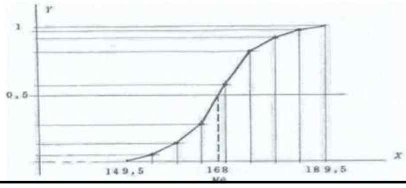
I. Tendance centrale

2. Médiane ($\tilde{x}$)

– Comment retrouver la médiane :
  - Par série statistique brute et ordonnée (cf. supra)
    – Ex série statistique impaire (N=9) : 1 1 2 2 **<u>2</u>** 3 4 5 5 → Médiane = 2 ((p+1)^{ème} valeur))
  - Par table des effectifs cumulés
    – Repérer valeur ou classe (x_i) qui comprend la (p+1)^{ème} valeur (N impair) ou la p^{ème} et (p+1)^{ème} valeur (N pair)

Réponses x_i de la variable x	45	55	60	75	80	85	90
Effectifs n_i	1	2	3	5	2	1	1
Effectifs cumulés n_i	1	3	6	11	13	14	15

  - Par graphe des effectifs cumulés



I. Tendance centrale

3. Moyenne (μ ou $\bar{x}$) *

- Somme de toutes les observations divisée par nombre d'observations
- Soit, $\boxed{\bar{x} = \frac{1}{N}\sum_{i=1}^{i=n} x_i = \frac{1}{N}\sum_{i=1}^{i=c} n_i x_i}$

Via données brutes

Via table effectif

Si données groupées par classes, le centre de classe peut être considéré comme estimation de x_i
- Sous l'h° d'équirépartition au sein des classes

- Ex série (N=13): 1 1 1 1 1 2 2 2 2 3 3 3 3

$$\bar{x} = \frac{1+1+1+1+\cdots+3+3+3}{13}$$ Via données brutes

$$\bar{x} = \frac{(5*1)+(4*2)+(4*3)}{13}$$ Via table effectif

* $\bar{x}$ quand échantillon ; μ quand population

I. Tendance centrale

3. Moyenne (μ ou $\bar{x}$) : propriétés

- Uniquement pour variables quantitatives
- Unique : une seule moyenne pour toute série statistique
- Somme des écarts entre x_i et la moyenne est nulle
  - Les différences positives et négatives s'annulent

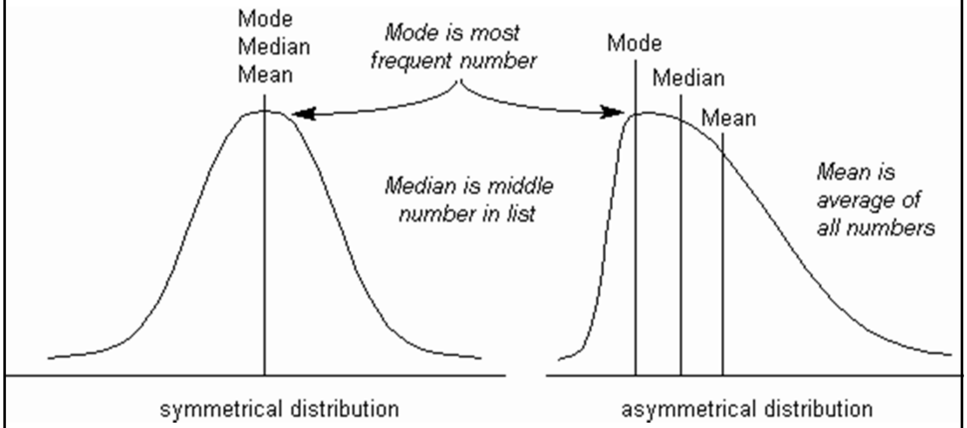
$$\sum_{i=1}^{i=n}(x_i - \bar{x}) = 0$$

- Moyenne de deux séries statistiques (pour une même variable)
  - Ex moyenne taille homme (série a) et taille femme (série b)

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{i=n} x_i(série\ a) + \sum_{i=1}^{i=n} x_i(série\ b)}{N_a + N_b} = \frac{\sum_{i=1}^{i=c} n_i x_i(série\ a) + \sum_{i=1}^{i=c} n_i x_i(série\ b)}{N_a + N_b}}$$

I. Tendance centrale

- Mode, médiane & moyenne

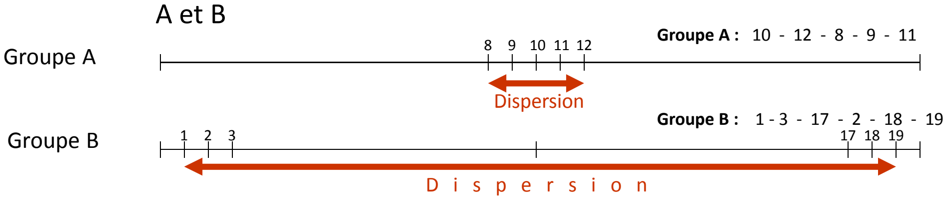


9

B. Dispersion

I. Variance et écart-type

- Compléments aux indices de position
  - Indices de position ne disent rien sur la dispersion
    - Ex : pour une même moyenne (10), scores sur 20 de deux groupes A et B



  - Notion de 'Moment' : écart moyen des x_i à la moyenne

$$\boxed{m = \frac{1}{N}\sum_{i=1}^{i=n}(x_i - \bar{x})}$$

- Donne une indication de la dispersion des valeurs autour de la moyenne mais 'pas utilisable' comme tel
  - Les différences positives et négatives s'annulent
  - Valeur absolue des différences ou les élever à la puissance 2 → moment d'ordre 2 = variance (cf. dia suivante)

I. Variance et écart-type

1. Variance (σ^2 ou s^2)*
   - Moyenne des carrés des écarts des valeurs x_i à la moyenne

Ecarts à la moyenne élevés au carré

$$\sigma^2 = \frac{\left(X_1 - \overline{X}\right)^2 + \left(X_2 - \overline{X}\right)^2 + \left(X_3 - \overline{X}\right)^2 + \ldots\ldots\ldots\ldots\left(X_n - \overline{X}\right)^2}{n}$$

   - ou, plus simplement

$$\boxed{\sigma^2 = \frac{1}{N}\sum_{i=1}^{i=n}(x_i - \bar{x})^2 = \frac{1}{N}\sum_{i=1}^{i=c} n_i(x_i - \bar{x})^2}$$

Via données brutes

Via table effectif

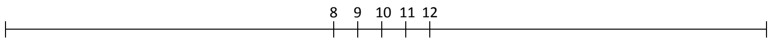
Si données groupées par classes, le centre de classe peut être considéré comme estimation de x_i
- Sous l'hypothèse d'équirépartition au sein des classes

* s^2 quand échantillon ; σ^2 quand population

I. Variance et écart-type

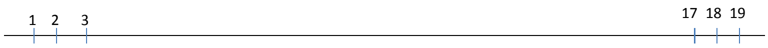
1. Variance (σ^2 ou s^2)

Groupe A



$$Variance = \frac{(8-10)^2+(9-10)^2+(10-10)^2+(11-10)^2+(12-10)^2}{5} = \frac{10}{5} = 2$$

Groupe B



$$Variance = \frac{(1-10)^2+(2-10)^2+(3-10)^2+(17-10)^2+(18-10)^2+(19-10)^2}{6} = \frac{388}{6} = 64{,}66$$

13

I. Variance et écart-type

2. Ecart-type (σ ou s)*
   - Racine carrée de la variance

$$\boxed{\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{i=n}(x_i-\bar{x})^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{i=c}n_i(x_i-\bar{x})^2}}$$

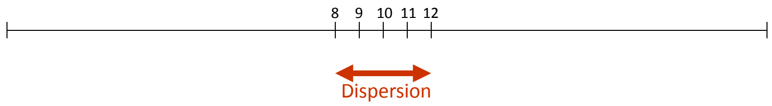
   - Indice similaire à la variance mais plus facilement interprétable
     - moyenne des écarts et non moyenne des carrés des écarts
     - La valeur d'écart-type peut s'exprimer selon la même métrique que la série statistique dont il est issu

14

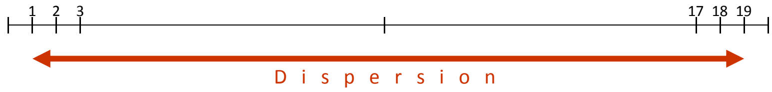
* s quand échantillon ; σ quand population

I. Variance et écart-type

2. Ecart-type (σ ou s)



Groupe A : Variance (σ^2)= 2 Ecart type (σ) = $\sqrt{2} = 1{,}41$



Groupe B : Variance (σ^2)= 64,66 Ecart type (σ) = $\sqrt{64{,}66} = 8{,}04$

15

C. Exercices

Exercices

1. Données

<u>Données sur le salaire net et le niveau d'études des employés d'une ASBL.</u>
- Salaires : euros
- Niveau d'études : 1 Primaire, 2 Secondaire inférieur, 3 Secondaire supérieur, 4 Supérieur de type court, 5 Supérieur de type universitaire

Sujets	x_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	x_{10}
Salaire net	1500	2250	1750	1500	2000	2500	2000	2000	1750	2000
Niveau d'étude	3	5	4	3	4	5	4	4	3	4

Exercices

2. Effectifs et fréquences

Variable : Salaire Net										
Réponses x_i de la variable	1500	1750	2000	2250	2500					
Effectifs (n_i)	2	2	4	1	1					
Effectifs cumulés	2	4	8	9	10					
Fréquences relatives (%)	20%	20%	40%	10%	10%					
Fréquences relatives cumulées (%)	20%	40%	80%	90%	100%					

Variable : Niveau d'études										
Réponses x_i de la variable	3	4	5							
Effectifs (n_i)	3	5	2							
Effectifs cumulés	3	8	10							
Fréquences relatives (%)	30%	50%	20%							
Fréquences relatives cumulées (%)	30%	80%	100%							

Exercices

3. Tendance centrale (1)

Salaire Net

– Moyenne :

$$\bar{X}\ (salaire\ net) = \frac{1500 + 2250 + 1750 + 1500 + 2000 + 2500 + 2000 + 2000 + 1750 + 2000}{10} = 1925$$

Ou, plus simplement

$$\bar{X}\ (salaire\ net) = \frac{(1500 * 2) + (1750 * 2) + (2000 * 4) + 2250 + 2500}{10} = 1925$$

– Médiane : 2000
– Mode : 2000

19

Exercices

3. Tendance centrale (2)

Niveau d’études

– Moyenne :

$$\bar{X}\ (niveau\ d'études) = \frac{3 + 5 + 4 + 3 + 4 + 5 + 4 + 4 + 3 + 4}{10} = 3{,}9$$

Ou, plus simplement

$$\bar{X}\ (niveau\ d'études) = \frac{(3 * 3) + (4 * 5) + (5 * 2)}{10} = 3{,}9$$

– Médiane : 4
– Mode : 4

20

Exercices

4. Dispersion (1)

Salaire Net

– Variance : σ²

$$= \frac{(1500-1925)^2+(2250-1925)^2+(1750-1925)^2+(1500-1925)^2+(2000-1925)^2+(2500-1925)^2+(2000-1925)^2+(2000-1925)^2+(1750-1925)^2+(2000-1925)^2}{10}$$

$$= 88125$$

– Ecart-type : $\sigma\,(salaire\ net) = \sqrt{88125}$

Exercices

4. Dispersion (2)

Niveau d'études

– Variance : σ²

$$= \frac{(3-3{,}9)^2+(5-3{,}9)^2+(4-3{,}9)^2+(3-3{,}9)^2+(4-3{,}9)^2+(5-3{,}9)^2+(4-3{,}9)^2+(4-3{,}9)^2+(3-3{,}9)^2+(4-3{,}9)^2}{10} = 0{,}49$$

– Ecart-type : $\sigma\,(niveau\ d'études) = \sqrt{0{,}49} = 0{,}7$