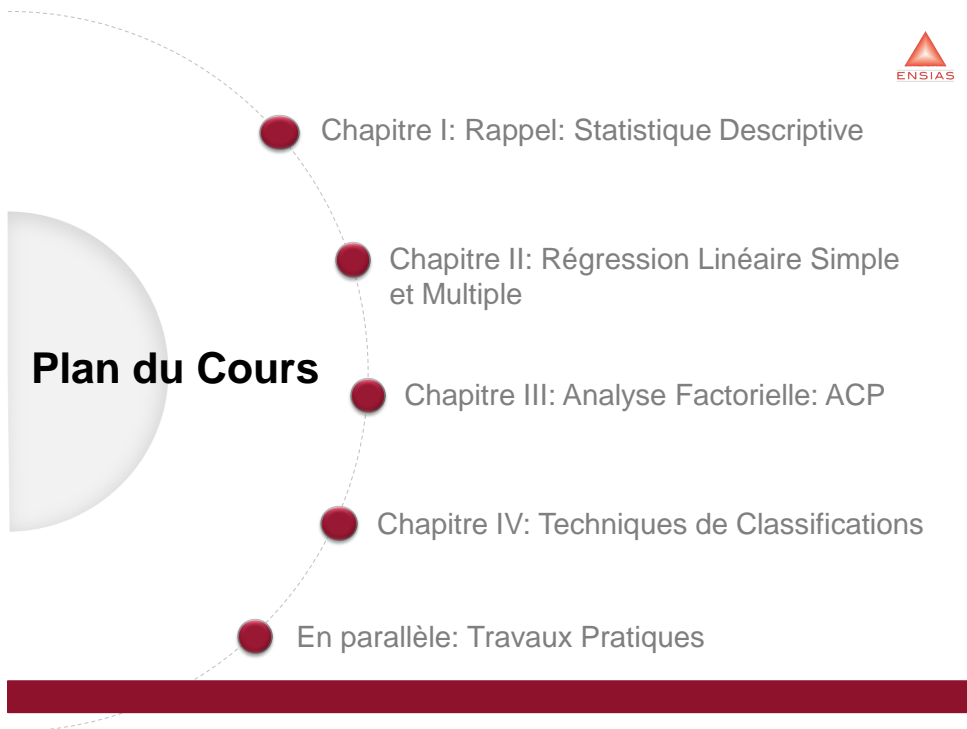




Analyse de Données

Par: Houda Benbrahim



Plan du Cours

Chapitre I: Rappel: Statistique Descriptive

Chapitre II: Régression Linéaire Simple et Multiple

Chapitre III: Analyse Factorielle: ACP

Chapitre IV: Techniques de Classifications

En parallèle: Travaux Pratiques



Plan du Cours

Chapitre I: Rappel: Statistique Descriptive

Chapitre II: Régression Linéaire Simple et Multiple

Chapitre III: Analyse Factorielle: ACP

Chapitre IV: Techniques de Classifications

En parallèle: Travaux Pratiques

3

H. Benbrahim



Section I: Introduction et Définition

Section II: Les Données

Section III: Représentation Graphique de Données

Section IV: Indicateurs Statistiques

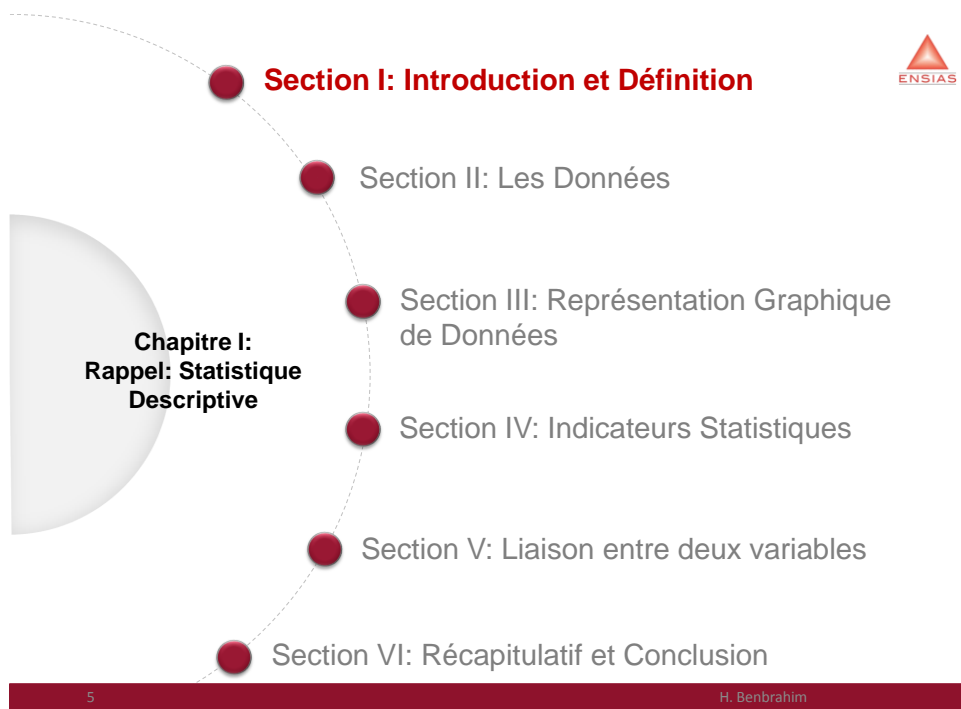
Section V: Liaison entre deux variables

Section VI: Récapitulatif et Conclusion

Chapitre I: Rappel: Statistique Descriptive

4

H. Benbrahim



[Définition]	Données	Graphique	Indicateurs	Liaison	Conclusion
Contexte général	Définition	Applications	Méthodologie		

Constatation, Exemple:

Le chiffre d'affaires a baissé en octobre. Pour y remédier, des décisions sont à prendre. Mais pour prendre la bonne décision, il faut savoir:

- Pourquoi le chiffre d'affaires a baissé?
- comment il a baissé?
- Dans quelle gamme de produits?
- Dans quels pays, quelles régions?
- Dans le portefeuille de clientèle de quels commerciaux?
- N'avait-on pas une baisse semblable en octobre chaque année?

➔ Les questions sont nombreuses, auxquelles il faut savoir répondre avant de prendre la bonne **décision**.

6 H. Benbrahim ENSIAS

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Contexte général	Définition	Applications	Méthodologie		


Exemple de Solution/Décision Classique:

L'entreprise doit améliorer:

- La qualité de ses produits & services
- La connaissance des marchés et leur dynamique
- Sa position face à la concurrence
- Les coûts de production
- La connaissance des technologies pour en tirer profit

7

H. Benbrahim



Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Contexte général	Définition	Applications	Méthodologie		

Nouvelle Tendance:

➔ Analyser les **données** de l'entreprise, afin d'en dégager des informations nouvelles qui vont fonder des décisions.

- Données (Explosion des données intra et extra entreprise)
 - Socio-démographique (Sexe, age, localisation, etc.)
 - Profil clients (loisirs, habitudes,...)
 - Comportement « d'achat » (volume, type,...)
 - Critères d'achat (études clients,...)

➔ Prendre des décisions:

- Anticipation des besoins des clients
- Adaptation de la stratégie de l'entreprise aux attentes du marché

8

H. Benbrahim



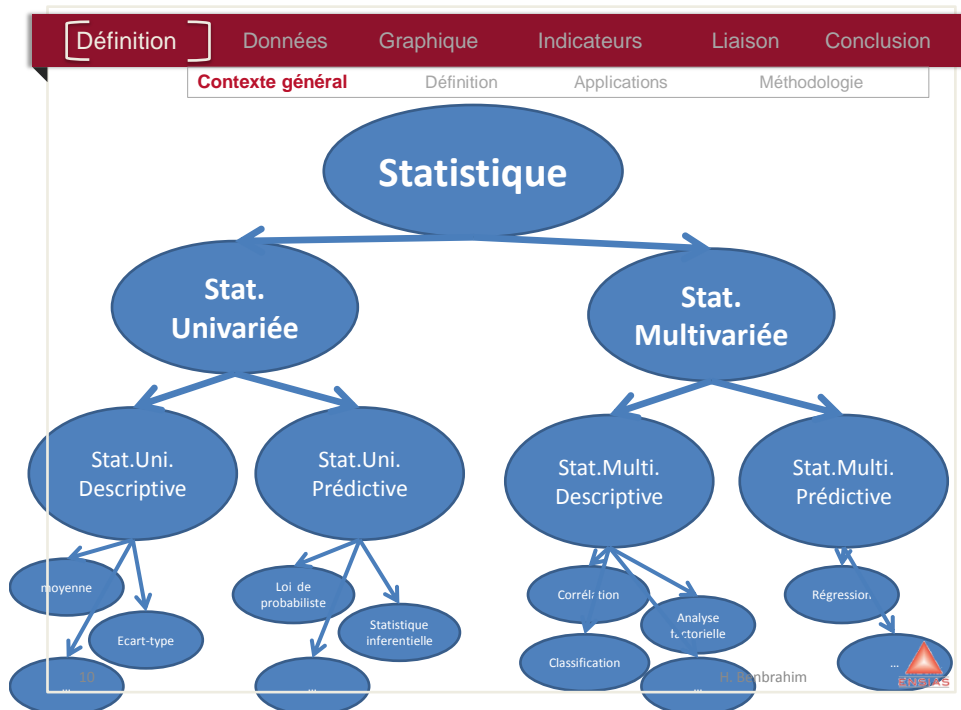
[Définition]	Données	Graphique	Indicateurs	Liaison	Conclusion
Contexte général	Définition	Applications	Méthodologie		

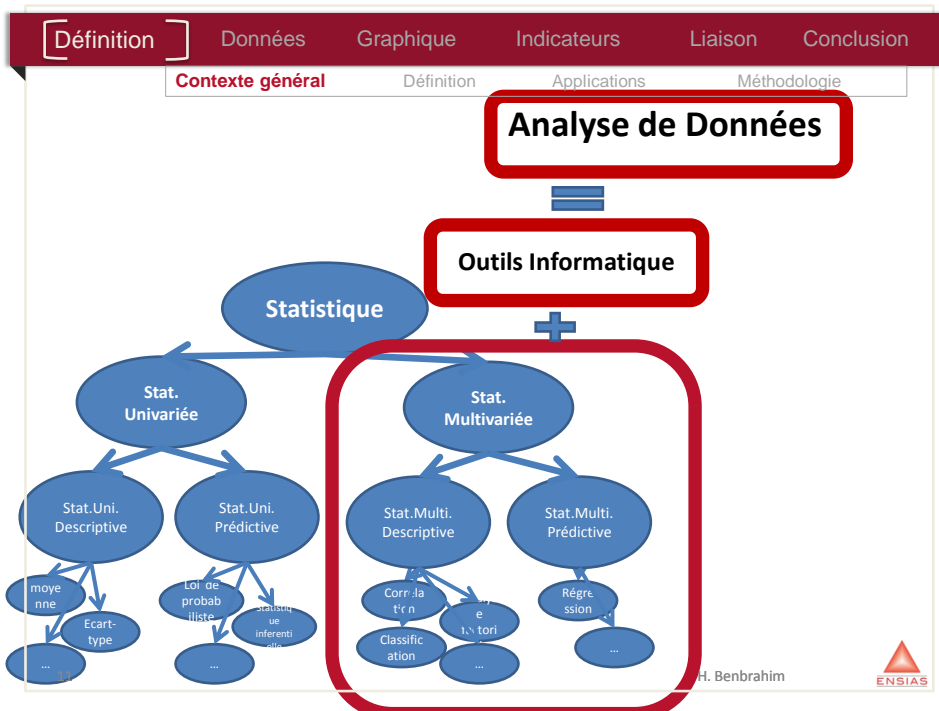
Analyser les données?

→ Statistique

- Le mot "**statistique**" regroupe à la fois un ensemble de données d'observation ainsi que la méthodologie de recueil, de traitement et d'interprétation de ces données
- Statistiques (latin « status » état):**
 - Ensemble cohérent de données numériques relatives à un groupe d'individus.
 - Statistiques démographiques
 - Statistiques annuelles des établissements financiers ou autres
 - Statistiques du chômage
- Statistique**
 - Ensemble des méthodes qui permettent de rassembler et d'analyser les données numériques

9 H. Benbrahim ENSIAS





[Définition] Données Graphique Indicateurs Liaison Conclusion

Contexte général **Définition** Applications Méthodologie

Définition

- « La statistique est l'ensemble des méthodes qui ont pour objet la collecte, le traitement et l'interprétation de données d'observation relatives à un groupe d'individus ou d'unités ».
- « La statistique est la science ayant pour objet l'étude quantitative des populations, à l'aide de données représentatives, le plus souvent incomplètes, et comportant généralement, de ce fait, un caractère d'incertitude ».

12

H. Benbrahim

ENSIAS

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Contexte général	Définition	Applications	Méthodologie		

Ex. Domaines d'applications

- Bio-médical: comparer de nouveaux traitements avec les solutions courantes (Contraintes: des données souvent coûteuses à obtenir).
- Environnement: localisation et abondance des plantes et des animaux, étendue et sources de pollution, effets possibles de changements dans l'activité humaine (Contraintes: des données souvent incomplètes, incertaines).
- Industrie: qualité de la production, analyse des données économiques de l'entreprise, enquêtes auprès de la clientèle.
- Transport: faire du « yield management »: tarifier au meilleur prix un billet d'avion ou de train.
- Statistique officielle: sondage auprès des entreprises, des ménages...
- Études de marché: données publiques + données d'enquêtes + tickets de caisse + fichiers de connexion à un site WEB... (Contraintes : données très abondantes).
- Banque, Assurance: segmentation de la clientèle, analyse de risque, calculs de primes.

13

H. Benbrahim



Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Contexte général	Définition	Applications	Méthodologie		

Les différentes étapes de toute étude statistique

1. Définition de la problématique de l'étude

2. Collecte des données :

- Simple observation
- Expérimentation : i.e en provoquant volontairement l'apparition de certains phénomènes contrôlés

3. Préparation des données

- Consolidation, fiabilité, calcul de nouveaux indicateurs

4. Analyse statistique

- Analyse "déductive" ou descriptive
 - a pour but de synthétiser et de présenter les données observées pour que l'on puisse en prendre connaissance facilement : tableaux, graphiques ...
- Analyse "inductive" ou inférence
 - permet d'étendre ou de généraliser dans certaines conditions les conclusions obtenues. Cette phase comporte certains risques d'erreur qui peuvent être mesurés en faisant appel à la théorie des probabilités.

5. Production et diffusion des résultats

14

H. Benbrahim



Définition

Données

Graphique

Indicateurs

Liaison

Conclusion

Contexte général

Définition

Applications


Méthodologie

Ce chapitre: Rappel de la statistique descriptive

- **But** : présenter les données pour que l'on puisse en prendre connaissance facilement (données vers l'information décisionnelle)
- **peut concerner** :
 - une variable à la fois : **statistique à une dimension**
 - deux variables à la fois : **statistique à deux dimensions**
 - plus de deux variables à la fois : **statistique multidimensionnelle**
- **comporte** :
 - les tableaux : distributions de fréquences
 - les diagrammes : graphiques
 - les paramètres statistiques : réduction des données à quelques valeurs numériques caractéristiques (tendances centrales et de dispersion)

15

H. Benbrahim
H. Benbrahim



Définition

Données

Graphique

Indicateurs

Liaison

Conclusion

Contexte général

Définition

Applications


Méthodologie

Ex. Données relatives aux employés d'une banque

	A	B	C	D	E	F	G	H	I
	ID	SEXE	DATENAI	EDUC	CATEMP	SALDEB	SALACT	TEMPS	EXP
1	1	m	3-févr-1952	15	3	27 000	57 000	98	144
2	2	m	23-mai-1958	16	1	18 750	40 200	98	36
3	3	f	26-juil-1929	12	1	12 000	21 450	98	381
4	4	f	15-avr-1947	8	1	13 200	21 900	98	190
5	5	m	9-févr-1955	15	1	21 000	45 000	98	138
6	6	m	22-août-1958	15	1	13 500	32 100	98	67
7	7	m	26-avr-1956	15	1	18 750	36 000	98	114
8	8	f	6-mai-1966	12	1	9 750	21 900	98	0
9	9	f	23-janv-1946	15	1	12 750	27 900	98	115
10	10	f	13-févr-1946	12	1	13 500	24 000	98	244
11	11	f	7-févr-1950	16	1	16 500	30 300	98	143
12	12	m	11-janv-1966	8	1	12 000	28 350	98	26
13	13	m	17-juil-1960	15	1	14 250	27 750	98	34
14	14	f	26-févr-1949	15	1	16 800	35 100	98	137
15	15	m	29-août-1962	12	1	13 500	27 300	97	66
16	16	m	17-nov-1964	12	1	15 000	40 800	97	24
17	17	m	18-juil-1962	15	1	14 250	46 000	97	48
18	18	m	20-mars-1956	16	3	27 510	103 750	97	70
19	19	m	19-août-1962	12	1	14 250	42 300	97	103
20	20	f	23-janv-1940	12	1	11 550	26 250	97	48
21	21	f	19-févr-1963	16	1	15 000	38 850	97	17
22	22	m	24-sept-1940	12	1	12 750	21 750	97	315
23	23	f	15-mars-1965	15	1	11 100	24 000	97	75
24	24	f	27-mars-1933	12	1	9 000	16 950	97	124
25	25	f	1-juil-1942	15	1	9 000	21 150	97	171
26	26	m	8-nov-1966	15	1	12 600	31 050	96	14
27	16 27	m	19-mars-1954	19	3	27 480	60 375	96	96

16

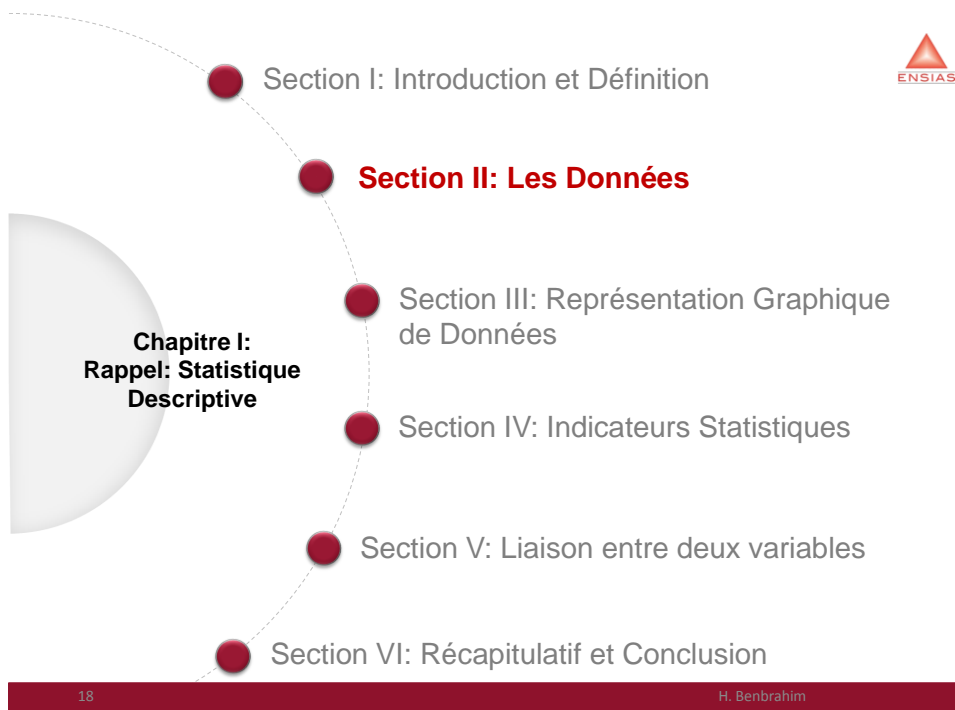
H. Benbrahim
H. Benbrahim



[Définition]	Données	Graphique	Indicateurs	Liaison	Conclusion
Contexte général	Définition	Applications	Méthodologie		

Ex. Données relatives aux employés d'une banque: Questions à se poser?

- Comment peut-on rendre plus intelligibles ces données ?
- Quelles sont les représentations graphiques requises pour mieux visualiser le comportement de ces variables et quelles interprétations peut-on en faire ?
- Existe-t-ils des valeurs typiques qui permettraient de résumer l'ensemble des données ?

17
17H. Benbrahim
H. Benbrahim

18

H. Benbrahim

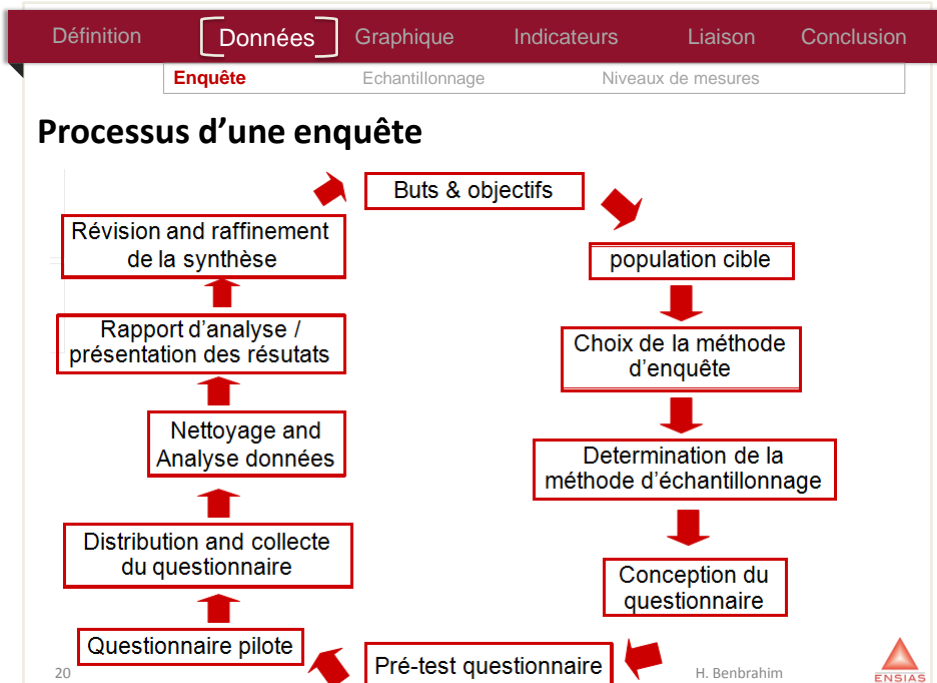
Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Enquête		Echantillonnage	Niveaux de mesures		

Collecte des données

Enquête

- Ensemble des opérations qui ont pour but de collecter de façon organisée des informations relatives à un groupe d'individus ou d'éléments observés dans leur milieu ou leur cadre habituel.
- Lorsque toutes les unités de la population sont observées l'enquête est exhaustive. Elle est encore **appelée recensement**.
- Lorsqu'au contraire, *une partie de la population* est observée, l'enquête est dite partielle ou par *échantillonnage*. Elle est encore appelée **sondage**. La partie de la population observée constitue l'échantillon.
- L'ensemble des unités auquel on s'intéresse est appelé **population** ou **univers** ou encore **ensemble statistique**.

19 H. Benbrahim ENSIAS



Définition **Données** Graphique Indicateurs Liaison Conclusion

Enquête Echantillonnage Niveaux de mesures

Enquêtes de satisfaction concernant l'usage du téléphone mobile

Dimensions mInterview

Overall, how do you rate the service you receive from us?

☐ Very Happy
☐ Happy
☐ Unhappy
☐ Very Unhappy

Next

Dimensions mInterview

How would you rate the service you've received from our staff?

	Very Good	Good	Average	Poor	Very Poor	Not Applicable
Telephone Support	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sales	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Accounts	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

Previous Next

Dimensions mInterview

How satisfied are you with your mobile phone?

☐ Very Happy
☐ Happy
☐ Unhappy
☐ Very Unhappy

Previous Next

Dimensions mInterview

Which features of your phone would you consider essential?

☒ Text Messaging (SMS)
☒ Voice mail
☒ Email / Internet
☒ Picture/Multimedia Messaging
☐ Camera
☐ Radio
☐ MP3 player
☐ Voice recorder
☒ Contacts / Address Book
☒ Calendar
☐ Alarm
☐ Games

Previous Next

Dimensions mInterview

How often do you use your mobile phone?

☒ Very frequently
☐ Frequently
☐ Occasionally
☐ In emergency

Previous Next

Dimensions mInterview

Do you have any further comments on your mobile phone service?

my phone sucks! I want a new one ASAP!

Previous Next H. Benbrahim

21

ENSIA

Définition **Données** Graphique Indicateurs Liaison Conclusion

Enquête Echantillonnage Niveaux de mesures

Collecte des données

- Les principaux problèmes qui se posent dans la préparation de l'enquête sont :
 - la définition de l'unité de base et de la population
 - la définition des observations à réaliser
 - le choix d'une méthode de collecte des données
 - le choix d'une méthode d'échantillonnage
 - la détermination de la taille de l'échantillon

22

H. Benbrahim

ENSIA

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
------------	----------------	-----------	-------------	---------	------------

Enquête	Echantillonnage	Niveaux de mesures
---------	------------------------	--------------------

Collecte des données: Quelques méthodes d'échantillonnage


- **Échantillonnage aléatoire simple** (*avec ou sans remise*)
le tirage sans remise est le plus fréquent et donne des estimations plus précises pour une même taille d'échantillon.
- **Échantillonnage stratifié ou par grappes**
 - A utiliser quand la population est très hétérogène et que l'on souhaite s'assurer que ses différentes composantes seront toutes bien représentées. La stratification peut apporter un gain de précision important par rapport à un échantillonnage aléatoire simple.
- **Échantillonnage à deux ou plusieurs niveaux**
 - Tirage au sort des familles
 - Puis tirage au sort dans chaque famille de la personne enquêtée.
- **Méthode des quotas** (quota) largement utilisée dans les sondages d'opinion.

➤ Toutes les méthodes nécessitent une base d'échantillonnage :

- on suppose que l'on dispose d'une liste de toutes les unités qui constituent la population

23

H. Benbrahim



Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
------------	----------------	-----------	-------------	---------	------------

Enquête	Echantillonnage	Niveaux de mesures
---------	------------------------	--------------------


Collecte des données: La taille de l'échantillon

- Fixée en valeur absolue ou en valeur relative : fraction de sondage
- La précision dans une enquête dépend :
 - de la taille de l'échantillon
 - du caractère plus ou moins homogène ou hétérogène de la population parent.
- La précision est d'autant meilleure que la taille de l'échantillon est importante et que la population est homogène.

➤ Pas de recette : pour fixer la taille d'un échantillon il est nécessaire d'avoir une idée suffisante de la précision souhaitée (risque accepté) et d'autre part du degré d'homogénéité (variabilité) de la population étudiée.

24

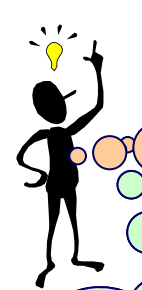
H. Benbrahim



DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

EnquêteEchantillonnageNiveaux de mesures

Niveaux de mesures



ORDINALE (Ordre)
Il est possible de classer du plus petit au plus grand

Mais, les intervalles entre les valeurs ne sont pas égaux ou inconnus

NOMINALE (Catégorielle)
Ne peuvent pas être ordonnées du plus petit au plus grand

CONTINUE (Echelle)
Valeurs sont continues peuvent être ordonnées les intervalles entre les valeurs sont égaux

26H. BenbrahimENSIAS

DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

EnquêteEchantillonnageNiveaux de mesures

La définition des observations

- Les observations à réaliser doivent être parfaitement définies.

ID	SEXE	DATENAI	EDUC	CATEMP	SALDEB	SALACT	TEMPS	EXP
1	m	3-févr-1952	15	3	27 000	57 000	98	144
2	m	23-mai-1958	16	1	18 750	40 200	98	36
3	f	26-juil-1929	12	1	12 000	21 450	98	381
4	f	15-avr-1947	8	1	13 200	21 900	98	190
5	m	9-févr-1955	15	1	21 000	45 000	98	138
6	m	22-août-1958	15	1	13 500	32 100	98	67
7	m	26-avr-1956	15	1	18 750	36 000	98	114
8	f	6-mai-1966	12	1	9 750	21 900	98	0
9	f	23-janv-1946	15	1	12 750	27 900	98	115
10	f	13-févr-1946	12	1	13 500	24 000	98	244
11	f	7-févr-1950	16	1	16 500	30 300	98	143
12	m	11-janv-1966	8	1	12 000	28 350	98	26
13	m	17-juil-1960	15	1	14 250	27 750	98	34
14	f	26-févr-1949	15	1	16 800	35 100	98	137
15	m	29-août-1962	12	1	13 500	27 300	97	66

27H. BenbrahimENSIAS

Définition

Données

Graphique

Indicateurs

Liaison

Conclusion

Enquête

Echantillonnage

Niveaux de mesures


La définition des observations: Nominal

ID	SEXE	DATENAI	EDUC	CATEMP	SALDEB	SALACT	TEMPS	EXP
1	m	3-févr-1952	15	3	27 000	57 000	98	144
2	m	23-mai-1958	16	1	18 750	40 200	98	36
3	f	26-juil-1929	12	1	12 000	21 450	98	381
4	f	15-avr-1947	8	1	13 200	21 900	98	190
5	m	9-févr-1955	15	1	21 000	45 000	98	138
6	m	22-août-1958	15	1	13 500	32 100	98	67
7	m	26-avr-1956	15	1	18 750	36 000	98	114
8	f	6-mai-1966	12	1	9 750	21 900	98	0
9	f	23-janv-1946	15	1	12 750	27 900	98	115
10	f	13-févr-1946	12	1	13 500	24 000	98	244
11	f	7-févr-1950	16	1	16 500	30 300	98	143
12	m	11-janv-1966	8	1	12 000	28 350	98	26
13	m	17-juil-1960	15	1	14 250	27 750	98	34
14	f	26-févr-1949	15	1	16 800	35 100	98	137
15	m	29-août-1962	12	1	13 500	27 300	97	66

- Observations qualitatives nominales
- Ex: Sexe : Féminin , Masculin
- Ex: Catemp : Cadre supérieur, Directeur, etc.
- Échelle nominale : les codes utilisés ne servent qu'à identifier la modalité à laquelle appartient l'individu.
- Dans cette échelle, il n'y a pas de relations d'ordre entre les codes.

28

H. Benbrahim



Définition

Données

Graphique

Indicateurs

Liaison

Conclusion

Enquête

Echantillonnage


Niveaux de mesures

La définition des observations: Ordinal

- Observations ordinales, les codes utilisés permettent :
 - d'identifier la modalité à laquelle appartient l'individu ;
 - d'établir une relation d'ordre entre les modalités observables et par le fait même entre les individus.
- Exemples:
 - Potentiel entrepreneurial : faible ; moyen; élevé
 - Groupe d'âge : moins de 18; de 18 à 24; 25 à 29; 30 à 34; etc
 - Niveau de scolarité : primaire, secondaire, collégial, universitaire
 - Niveau d'appréciation : très bonne qualité; bonne qualité; qualité moyenne

29

H. Benbrahim



Définition

Données

Graphique

Indicateurs

Liaison

Conclusion

Enquête

Echantillonnage

Niveaux de mesures

La définition des observations: quantitative

ID	SEXE	DATENAI	EDUC	CATEMP	SALDEB	SALACT	TEMPS	EXP
1	m	3-févr-1952	15	3	27 000	57 000	98	144
2	m	23-mai-1958	16	1	18 750	40 200	98	36
3	f	26-juil-1929	12	1	12 000	21 450	98	381
4	f	15-avr-1947	8	1	13 200	21 900	98	190
5	m	9-févr-1955	15	1	21 000	45 000	98	138
6	m	22-août-1958	15	1	13 500	32 100	98	67
7	m	26-avr-1956	15	1	18 750	36 000	98	114
8	f	6-mai-1966	12	1	9 750	21 900	98	0
9	f	23-janv-1946	15	1	12 750	27 900	98	115
10	f	13-févr-1946	12	1	13 500	24 000	98	244
11	f	7-févr-1950	16	1	16 500	30 300	98	143
12	m	11-janv-1966	8	1	12 000	28 350	98	26
13	m	17-juil-1960	15	1	14 250	27 750	98	34
14	f	26-févr-1949	15	1	16 800	35 100	98	137
15	m	29-août-1962	12	1	13 500	27 300	97	66

- observations quantitatives: résultat d'une mesure ou d'un comptage
- les opérations arithmétiques ont un sens
- Ex. Saldeb: Salaire du début
- Ex.Temps: Ancienneté de l'employé
- EDUC : Nb d'années d'étude
- Une variable quantitative peut être discrète ou continue
- Discrète : ne peut prendre qu'un Nb. limité de valeurs (souvent entières)
- Continue : peut prendre toutes les valeurs d'un intervalle fini ou infini

30

H. Benbrahim

ENSIAS

Définition

Données

Graphique

Indicateurs

Liaison

Conclusion

Enquête

Echantillonnage

Niveaux de mesures

En résumé : Niveaux de mesure

Niveau de mesure	Propriété		
Nominale	qualitative		
Ordinale	qualitative	Rang	
Continue/Echelle (Intervalle/Ratio)	quantitative	Rang	intervalle

31

H. Benbrahim

ENSIAS

Définition

Données

Graphique

Indicateurs

Liaison

Conclusion

Enquête

Echantillonnage

Niveaux de mesures

Exercice d'apprentissage

Précisez le type de caractère et le type d'échelle correspondant à chacune des situations suivantes :

Q1: Nb d'années d'existence de l'entreprise :

	code
Moins de 2 ans	1
2 mais moins de 5 ans	2
5 mais moins de 10 ans	3
10 ans et plus	4

Q2: Taille de l'entreprise :

	code
Petite	1
Moyenne	2
Grande	3
Très grande	4

Q3: Chiffre d'affaire de l'entreprise

	code
Moins de 50 000 Dh	1
50 000 Dh à 100 000 Dh	2
100 000 Dh à 500 000 Dh	3
500 000 Dh et plus	4

Q4: Nb de plaintes/j au service Client


2; 4; 0; 12; 14; 8;

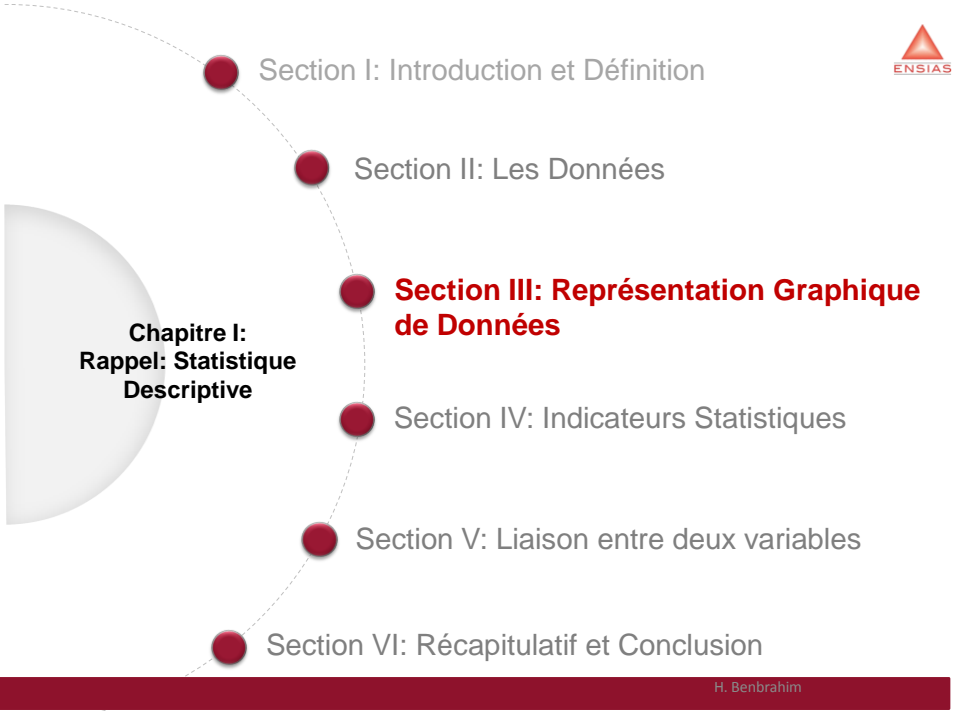
Q5: Opinion sur le DG choisi

	Code
Pour	1
Contre	2
Ne sais pas	3

32

H. Benbrahim





Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
	Généralités	Var. Qualitative	Var. Quantitative Discrète	Var. Quantitative Continue	

Comment peut-on résumer les différents types de mesures ?

Qualitative

Nominale → Exemple: Couleur des yeux, sexe, etc.
 Graphe: Diagramme sectoriel / Bâton
 Tendance centrale : Mode

Ordinale → Exemple: niveau de satisfaction, score, etc.
 Graphe: Diagramme sectoriel / Bâton
 Tendance centrale : Mode + Médiane, Rang

Quantitative (Echelle)

Ratio / Intervalle → Exemple: Revenu, Taille, Poids, âge, etc
 Graphe: Histogramme
 Tendance centrale : Mode + Médiane + Moyenne, Rang, Ecart type

34 H. Benbrahim ENSIAS

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
	Généralités	Var. Qualitative	Var. Quantitative Discrète	Var. Quantitative Continue	

Variables Qualitatives

Variables

- Catégorielles
 - Nominales: Sexe, CSP, Religion
 - Ordinales: Satisfaction, Niveau d'études...
- Quantitatives
 - Discrètes: Nombre d'enfants, Heures passées devant la télévision
 - Continues: Revenu, Age, Taille

Soit x une variable qualitative à k modalités. L'ensemble des n individus peut être subdivisé en k groupes sur lesquels x est constante.

Si la variable x est **ordinale**, les modalités sont écrites dans l'ordre :

$$\text{modalité } 1 < \text{modalité } 2 < \dots < \text{modalité } k.$$

Les graphiques les plus utilisés sont les suivants :

- diagrammes circulaires ou « en camembert »
- diagrammes en barres.

Dans les deux cas, les surfaces symbolisant les différentes modalités doivent être proportionnelles aux effectifs, ou aux fréquences associés.

35 H. Benbrahim ENSIAS

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Généralités	Var. Qualitative	Var. Quantitative Discrète	Var. Quantitative Continue		

Représentation graphique d'une variable qualitative

Exemple : la variable x désigne la situation familiale des individus

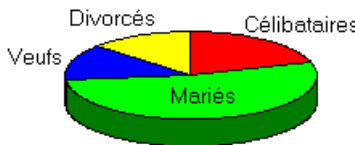


Diagramme sectoriel

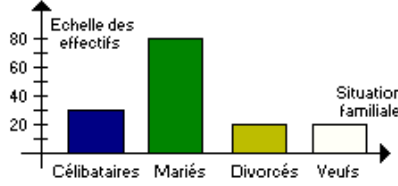


Diagramme en bâtons

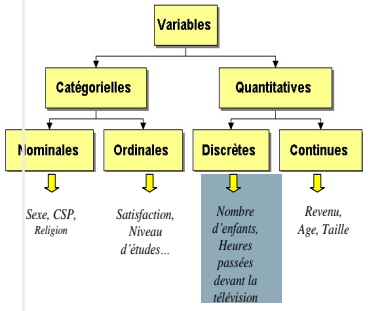
36

H. Benbrahim

ENSIA

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Généralités	Var. Qualitative	Var. Quantitative Discrète	Var. Quantitative Continue		

Variable quantitative discrète



On rappelle qu'une variable quantitative discrète est une variable ne prenant que des valeurs entières (plus rarement décimales). **Le nombre de valeurs distinctes d'une telle variable est habituellement assez faible.**

On notera dans la suite x la variable discrète et ses valeurs distinctes $\{x_1, x_2, \dots, x_p\}$. Chaque modalité x_i étant d'effectif (ou de **fréquence absolue**) n_i et de fréquence relative f_i , (n_i / n) $i=1, \dots, p$.

Exemple : On a noté l'âge des 48 salariés d'une entreprise, la série statistique brute est donnée ci-dessous

43 29 57 45 50 29 37 59 46 31 46 24 33 38 49 31
 62 60 52 38 38 26 41 52 60 49 52 41 38 26 37 59
 57 41 29 33 33 43 46 57 46 33 46 49 57 57 46 43

37

H. Benbrahim

ENSIA

Définition Données **Graphique** Indicateurs Liaison Conclusion

Généralités Var. Qualitative **Var. Quantitative Discrete** Var. Quantitative Continue

Présentation des données : Tableau statistique

Variables

Catégorielles

Nominales
↓
Sexe, CSP, Religion

Ordinales
↓
Satisfaction, Niveau d'études...

Quantitatives

Discrètes
↓
Nombre d'enfants, Heures passées devant la télévision

Continues
↓
Revenu, Age, Taille

xi

	ni	Ni	fi(%)	Fi(%)
24	1	1	2,08	2,08
26	2	3	4,17	6,25
29	3	6	6,25	12,50
31	2	8	4,17	16,67
33	4	12	8,33	25,00
37	2	14	4,17	29,17
38	4	18	8,33	37,50
41	3	21	6,25	43,75
43	3	24	6,25	50,00
45	1	25	2,08	52,08
46	6	31	12,50	64,58
49	3	34	6,25	70,83
50	1	35	2,08	72,92
52	3	38	6,25	79,17
57	5	43	10,42	89,58
59	2	45	4,17	93,75
60	2	47	4,17	97,92
62	1	48	2,08	100,0

↑

Effectifs cumulés

↑

Fréquences cumulées

H. Benbrahim

Colonne 1 : ensemble des observations distinctes de x rangées par ordre croissant et **non répétées**

Colonne 2 : les **effectifs** (nombre de réplifications)

Colonne 3 : Les effectifs cumulés sont définis par :
 $N_i = n_1 + \dots + n_i$, pour $i \geq 1$.

Colonne 4 : fréquences cumulées $F_i = f_i + \dots + f_1$,

Définition Données **Graphique** Indicateurs Liaison Conclusion

Généralités Var. Qualitative **Var. Quantitative Discrete** Var. Quantitative Continue

Représentation graphique (diagramme)

Variables

Catégorielles

Nominales
↓
Sexe, CSP, Religion

Ordinales
↓
Satisfaction, Niveau d'études...

Quantitatives

Discrètes
↓
Nombre d'enfants, Heures passées devant la télévision

Continues
↓
Revenu, Age, Taille

Diagramme en bâtons

Deux graphiques populaires :

- **Diagramme en bâtons**
- **Diagramme cumulatif**

Age des salariés H. Benbrahim

39

DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

GénéralitésVar. QualitativeVar. Quantitative DiscrèteVar. Quantitative Continue

Variables quantitatives continues

Variables

Catégorielles

Nominales

Sexe, CSP, Religion

Quantitatives

Ordinales

Satisfaction, Niveau d'études...

Discrètes

Nombre d'enfants, Heures passées devant la télévision

Continues

Revenu, Age, Taille

Rappelons qu'une variable quantitative est dite **continue** lorsque les observations qui lui sont associées ne sont pas des valeurs précises mais des **intervalles réels**.

Histogramme : remplace le diagramme en bâtons pour var discrètes.

A chaque classe $[x_i, x_{i+1}]$ est associé un rectangle dont la surface est proportionnelle à la base $e_i = x_{i+1} - x_i$.

Caractère continue x

DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

GénéralitésVar. QualitativeVar. Quantitative DiscrèteVar. Quantitative Continue

Cas d'une série chronologique ou temporelle

Variables

Catégorielles

Nominales

Sexe, CSP, Religion

Quantitatives

Ordinales

Satisfaction, Niveau d'études...

Discrètes

Nombre d'enfants, Heures passées devant la télévision

Continues

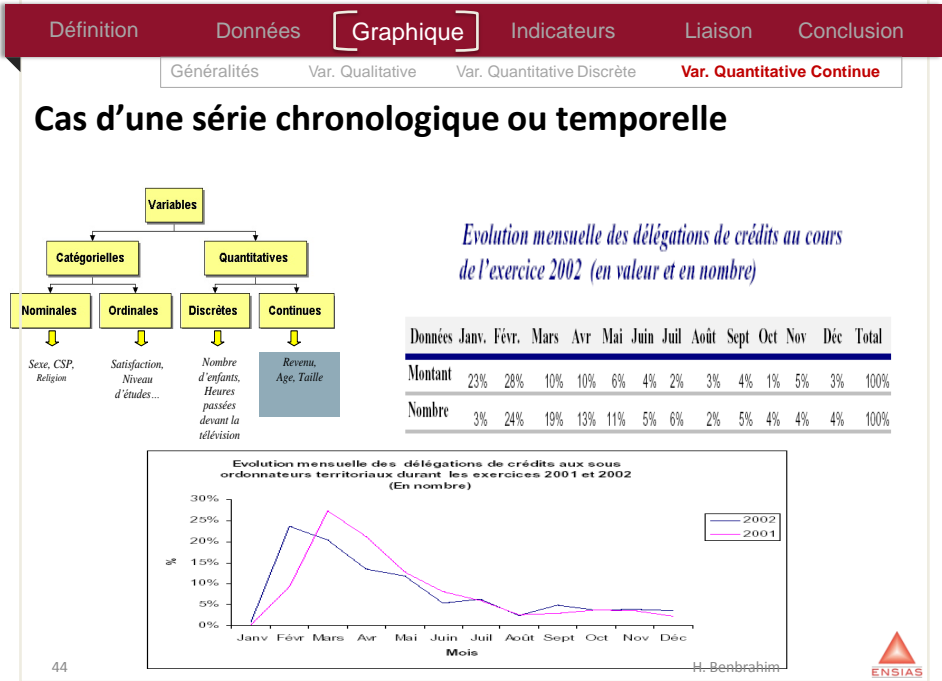
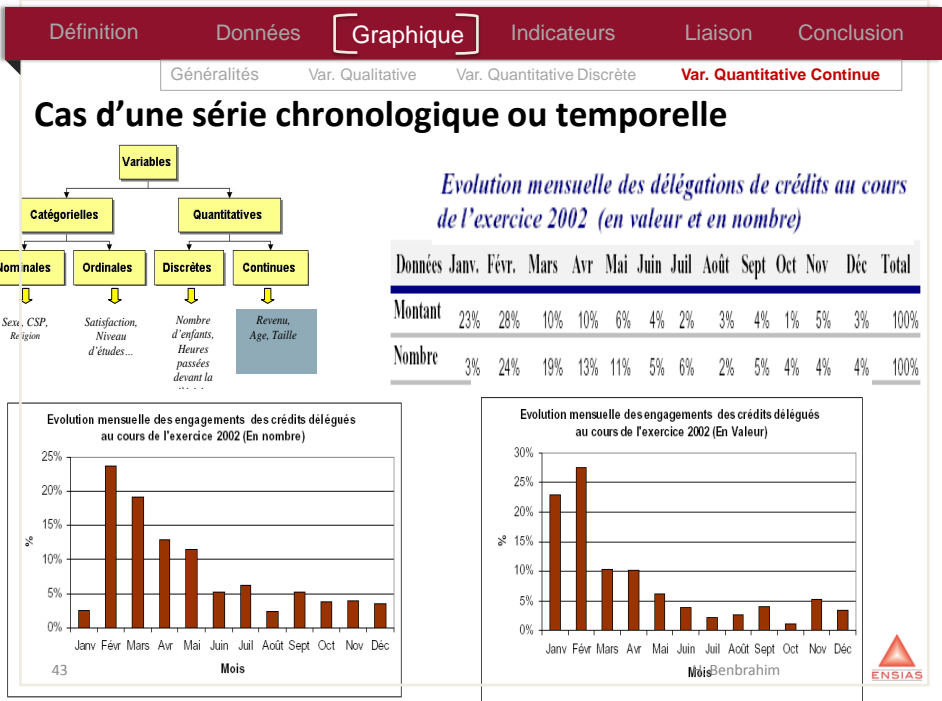
Revenu, Age, Taille

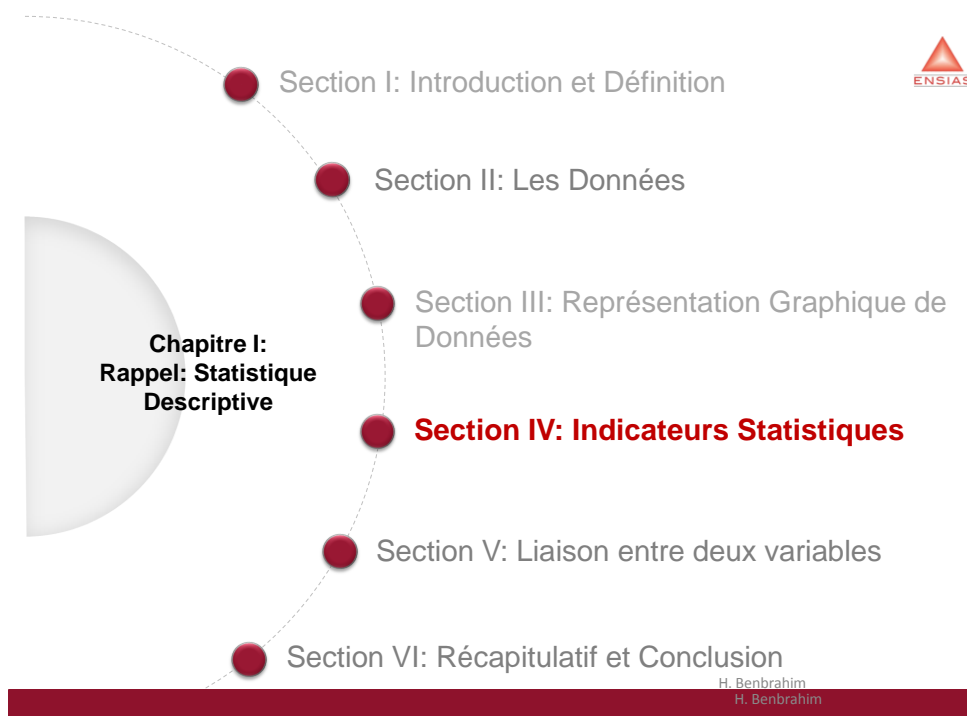
Une série chronologique est une série de données de nature quantitative qui ont été obtenues dans le temps à intervalles de temps réguliers.

Evolution mensuelle des délégations de crédits au cours de l'exercice 2002 (en valeur et en nombre)

Données	Janv.	Févr.	Mars	Avr	Mai	Juin	Juil	Août	Sept	Oct	Nov	Déc	Total
Montant	23%	28%	10%	10%	6%	4%	2%	3%	4%	1%	5%	3%	100%
Nombre	3%	24%	19%	13%	11%	5%	6%	2%	5%	4%	4%	4%	100%

42 H. Benbrahim





Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
	Généralités	Tendance Centrale	Dispersion	Mesure de Forme	

Quels Indicateurs Statistiques?

Caractéristiques de Tendances centrales	<ul style="list-style-type: none"> ▪ Moyenne arithmétique ▪ Médiane ▪ Mode
Dispersion	<ul style="list-style-type: none"> ▪ Etendu ▪ Ecart-type ▪ Indice de variabilité
Forme	<ul style="list-style-type: none"> ▪ Quantiles ▪ indice symétrie ▪ indice aplatissement

47

H. Benbrahim

ENSIAS

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Généralités	Tendance Centrale	Dispersion	Mesure de Forme		

Peut-on décrire les tendances centrales de la même manière pour tout type de variables ?


MOYENNE

MEDIANE

MODE

49

H. Benbrahim



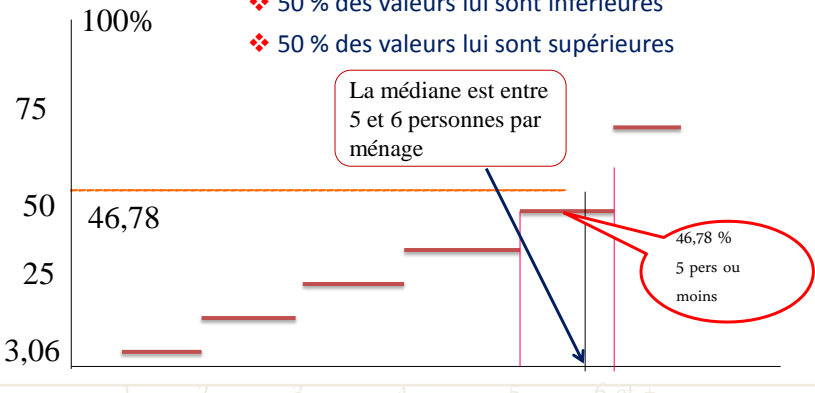
Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Généralités	Tendance Centrale	Dispersion	Mesure de Forme		
	Médiane	Mode	Moyenne		

La médiane

- La **médiane** Q_2 : valeur du caractère qui partage les données en deux sous ensemble de même fréquence :

- ❖ 50 % des valeurs lui sont inférieures
- ❖ 50 % des valeurs lui sont supérieures


La médiane est entre 5 et 6 personnes par ménage



Personnes par ménage	Fréquence cumulative (%)
1	3,06
2	25
3	50
4	75
5	90,78
6 et +	100

46,78

46,78 %
5 pers ou moins



DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

GénéralitésTendance CentraleDispersionMesure de Forme

MédianeModeMoyenne

La médiane d'une variable discrète

Cas de valeurs distinctes

Mois	Moy T° à Essaouira
janv	3
fév	3,6
mars	6,6
avril	9,6
mai	13
juin	16
juil	17,9
août	17,7
sept	15,3
oct	11,2
nov	6,4
déc	3,7

Tri

Mois	Moy T° à Essaouira
janv	3
fév	3,6
déc	3,7
nov	6,4
mars	6,6
avril	9,6
oct	11,2
mai	13
sept	15,3
juin	16
août	17,7
juil	17,9

Cas pair

$$Q_2 = (9,6 + 11,2)/2 = 10,4\text{ °C}$$

1. Classer les n données dans l'ordre croissant et numéroter les valeurs ordonnées de 1 à N. Soient $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ les valeurs distinctes ordonnées.

2. Si N est impair alors $Q_2 = x_{(k)}$ avec $k = (N+1)/2$
Si N est pair alors $Q_2 = (x_{(k)} + x_{(k+1)})/2$ avec $k = N/2$

51H. BenbrahimENSIAS

DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

GénéralitésTendance CentraleDispersionMesure de Forme

MédianeModeMoyenne

La Médiane d'une variable discrète

Cas de valeurs répétées

On utilise les effectifs cumulés croissants : pour les valeurs précédents Q_2 , ils sont inférieurs à $(N/2)$ et pour les valeurs suivants Q_2 , ils sont supérieurs.

Exemple de l'âge des employés

xi	ni	Ni	fi(%)	Fi(%)
24	1	1	2,08	2,08
26	2	3	4,17	6,25
29	3	6	6,25	12,50
31	2	8	4,17	16,67
33	4	12	8,33	25,00
37	2	14	4,17	29,17
38	4	18	8,33	37,50
41	3	21	6,25	43,75
43	3	24	6,25	50,00
45	1	25	2,08	52,08
46	6	31	12,50	64,58
49	3	34	6,25	70,83
50	1	35	2,08	72,92
52	3	38	6,25	79,17
57	5	43	10,42	89,58
59	2	45	4,17	93,75
60	2	47	4,17	97,92
62	1	48	2,08	100,0

$Q_2 = 43$

Fréquence relative cumulée

52H. BenbrahimENSIAS


Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Généralités	Tendance Centrale	Dispersion	Mesure de Forme		
	Médiane	Mode	Moyenne		

La Médiane

- La médiane tient compte du rang de tous les individus et non de leur valeur.
- Les valeurs exceptionnelles ne l'affectent pas.
- Elle est qualifiée d'estimateur robuste
- Valable sur caractères quantitatifs et qualitatifs ordinaux
- La médiane est la valeur centrale la plus proche de tous les individus

53

H. Benbrahim



Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Généralités	Tendance Centrale	Dispersion	Mesure de Forme		
	Médiane	Mode	Moyenne		

Médiane d'une variable continue

Cas des données groupées

Classes	Fréquence	Fi Cumulé
24 à 29,4	6	6
29,4 à 34,85	6	12
34,84 à 40,28	6	18
40,28 à 45,71	7	25
45,71 à 51,14	10	35
51,14 à 56,57	3	38
56,57 à 62	10	48

$B_{inf} = 40,28$
 $N/2 = 24$
 $F = 18$
 $F_{rne} = 7$
 $E = 45,71 - 40,28 = 5,43$

On effectue **une interpolation linéaire** à l'intérieur de la classe médiane afin de trouver la valeur de l'observation centrale. La formule requise pour déterminer Q_2 est la suivante :

$$Q_2 = B_{inf} + [(N/2 - F) / f_{Me}] * E$$

Q_2 : est la borne inférieure de la classe médiane ;
 F : la somme des fréquences absolues de toutes les classes précédant la classe médiane
 f_{Me} : la fréquence absolue de la classe médiane
 E : l'étendu de la classe médiane

L'âge médian est l'âge de la personne qui se trouve à 24^{ème} position dans le classement par ordre croissant


La classe médiane = [40,28 ; 45,14]

L'âge médian = $40,28 + ((24 - 18) / 7) * 5,43$
 $= 41,58 \text{ ans}$

La classe médiane correspond à la valeur 50% de la Fi cumulé

54

H. Benbrahim



DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

GénéralitésTendance CentraleDispersionMesure de Forme

MédianeModeMédianeMoyenne

Les quartiles Q_1 et Q_3

Q1

Q2

Q3

025%50%75%100%

Modalités
(âge en mois)

Effectif
ni

Effectif
cumulé

Fréquence
relative

Fréquence
cumulée

3

1

1

0,1

0,1

4

1

2

0,1

0,2

7

1

3

0,1

0,3

8

3

6

0,3

0,6

9

1

7

0,1

0,7

10

1

8

0,1

0,8

12

1

9

0,1

0,9

14

1

10

0,1

1

Total

10

Le rang de Q_1 est : $N/4 = 10/4 = 2,5 \rightarrow$ modalité 7

Le rang de Q_3 est $3 \cdot N/4 = 3 \cdot 10/4 = 7,5 \rightarrow$ modalité 10

ENSIAS

DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

GénéralitésTendance CentraleDispersionMesure de Forme

MédianeModeMoyenne

Le mode

Le **mode** = la valeur du caractère la plus fréquente

Le mode a l'avantage d'être utilisable avec les données qualitatives

Mode

Mode = 6 personnes et +

0102030405060

1,002,003,004,005,006,00

Nombre de personnes par ménage

Etat matrimonial	Effectifs	Pourcentage	$f_i \times 100$	$F_i \times 100$
Célibataire	158	3,1	3,1	3,1
Marié(monogame)	4200	81,9	81,9	84,9
Marié(polygame)	82	1,6	1,6	86,5
Divorcé	109	2,1	2,1	88,7
Veuf	582	11,3	11,3	100,0
Total	5131	100,0	100,0	

ENSIAS

26

Définition Données Graphique **Indicateurs** Liaison Conclusion

Généralités **Tendance Centrale** Dispersion Mesure de Forme

Médiane **Mode** Moyenne

Le Mode d'une variable quantitative discrète

x_i	n_i	N_i	$f_i(\%)$	$F_i(\%)$
24	1	1	2,08	2,08
26	2	3	4,17	6,25
29	3	6	6,25	12,50
31	2	8	4,17	16,67
33	4	12	8,33	25,00
37	2	14	4,17	29,17
38	4	18	8,33	37,50
41	3	21	6,25	43,75
43	3	24	6,25	50,00
45	1	25	2,08	52,08
46	6	31	12,50	64,58
49	3	34	6,25	70,83
50	1	35	2,08	72,92
52	3	38	6,25	79,17
57	5	43	10,42	89,58
59	2	45	4,17	93,75
60	2	47	4,17	97,92
62	1	48	2,08	100,0

46 est le mode de la variable âge

En se référant à la fréquence absolue ou effectif (colonne 2)

En se référant à la fréquence relative ou effectif (colonne 4)

Répartition des âges de 48 cadres

Définition Données Graphique **Indicateurs** Liaison Conclusion

Généralités **Tendance Centrale** Dispersion Mesure de Forme

Médiane Mode **Moyenne**

Moyenne arithmétique

C'est la valeur centrale la plus utilisée mais elle n'est calculable que sur des caractères quantitatifs.

caractères quantitatifs discrets :
c'est la somme des valeurs observées divisée par le nombre d'observations

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

caractères quantitatifs continus où les données sont groupées par classes.
On commet une légère erreur en remplaçant chacune des valeurs modalités par son centre de classe (CC)

$$\bar{x} = \frac{1}{n} \sum_{i=1}^p n_i x_i = \sum_{i=1}^p f_i x_i$$

- Les valeurs extrêmes décentrent la moyenne.
 Elles peuvent n'être que :
 - peu significatives
 - très exceptionnelles
 - voire aberrantes

Il faut donc contrôler leur pertinence

59

H. Benbrahim

DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

GénéralitésTendance CentraleDispersionMesure de Forme

MédianeModeMoyenne

(18+20)/2 = 19 ans

Moyenne

La moyenne d'une variable continue

On ne considère plus les valeurs des modalités, mais les centres des classes

$$\bar{X} = \frac{1}{N} \sum_{i=1}^k n_i c_i$$

k = Nombre de classes

Classes d'âge	ni	ci	fi	Fi	ci*fi
18 - 20	10	19	0,1	0,1	1,9
20 - 22	18	21	0,18	0,28	3,78
22 - 24	23	23	0,23	0,51	5,29
24 - 26	14	25	0,14	0,65	3,5
26 - 28	10	27	0,1	0,75	2,7
28 - 30	8	29	0,08	0,83	2,32
30 - 32	4	31	0,04	0,87	1,24
32 - 34	5	33	0,05	0,92	1,65
34 - 36	1	35	0,01	0,93	0,35
36 - 38	2	37	0,02	0,95	0,74

DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

GénéralitésTendance CentraleDispersionMesure de Forme

MédianeModeMoyenne

Autres moyennes

- La moyenne arithmétique n'est pas toujours la mieux adaptée. C'est le cas pour les phénomènes : multiplicatifs ; cumulatifs ; ou mettant en cause des fractions.
- Il faut utiliser une **moyenne** :
 - Géométrique : Quel est le taux d'accroissement annuel ?
 - Harmonique : permet de calculer des moyennes de pourcentage ou des moyennes de ratios
 - Quadratique : permet, de calculer des moyennes d'écart

DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

GénéralitésTendance CentraleDispersionMesure de Forme

MédianeModeMoyenne

Mesures appropriées des tendances centrales

	Mode	Médiane	Moyenne
Nominale	✓	✗	✗
Ordinale	✓	✓	?
Echelle	✓	✓	✓

62H. BenbrahimENSIAS

DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

GénéralitésTendance CentraleDispersionMesure de Forme

MédianeModeMoyenne

Comparaison des trois mesures de tendance centrale : la moyenne, la médiane et le mode

	Moyenne (\bar{x})	Médiane (Me_x)	Mode (Mo_x)
Calcul	<ul style="list-style-type: none">Facile	<ul style="list-style-type: none">Difficile (il faut trier les données)	<ul style="list-style-type: none">Difficile (il faut mettre les données en classes)
Valeurs except.	<ul style="list-style-type: none">Affectent beaucoup la valeur de \bar{x}	<ul style="list-style-type: none">Affectent peu la valeur de Me_x	<ul style="list-style-type: none">Affectent peu ou pas la valeur de Mo_x
Intérêt principal	<ul style="list-style-type: none">Bon estimateur de tendance centrale si distribution sym.\bar{x} est plus efficace que Me	<ul style="list-style-type: none">Plus précise que MoMoins affectée que \bar{x} par les valeurs extrêmes	<ul style="list-style-type: none">Pour décrire une distribution plurimodalePeut être calculé pour variables circulaireset pour var. qualitative

63H. BenbrahimENSIAS

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion	
	Généralités	Tendance Centrale	Dispersion	Mesure de Forme		
		Généralités	Etendu	Variance	IQ	Boite à Moustache

Mesures de dispersion

- Votre score est de 55% dans un test.
 - Quelle est votre performance si le score moyen est de 50%?
 - Mieux que la moyenne?
 - Oui, mais de combien?
- Pour le savoir, vous avez besoin de connaître
 - l'étendue
 - Variabilité
 - dispersion des données

64 H. Benbrahim ENSIAS

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion	
	Généralités	Tendance Centrale	Dispersion	Mesure de Forme		
		Généralités	Etendu	Variance	IQ	Boite à Moustache

Mesures de Dispersion

Minimum → 3 4 4 4 4 5 5 6 8 11 12 **Maximum**

IQR (Interquartile Range) is indicated by a bracket from the 4th figure (score 4) to the 9th figure (score 6).

Etendu (Max-Min) (Range) is indicated by a bracket from the first figure (score 3) to the last figure (score 12).

Ecart-type: en moyenne, de combien chaque score diffère de la moyenne ?

H. Benbrahim ENSIAS

DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

Généralités

Tendance Centrale

Dispersion

Mesure de Forme

Généralités

Etendu

Variance

IQ

Boite à Moustache

variation

Peut-on décrire les mesures de dispersion de la même manière pour tout type de données ?

Variance

ET

MIN /MAX

IQR

66H. BenbrahimENSIAS

DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

Généralités

Tendance Centrale

Dispersion

Mesure de Forme

Généralités

Etendu

Variance

IQ

Boite à Moustache

variation

Mesures de dispersion, pourquoi?

- Dans la plupart du temps, les mesures de tendance centrale ne peuvent à elles seules décrire et résumer convenablement un ensemble de données.
- Exemple 1** Dans le service d'urgence d'un hôpital on note à chaque intervalle de temps d'une heure le nombre d'arrivées de malades ou de blessés (l'observation a durée 12 heures). Les résultats sont données dans le tableau suivant
- Le service d'urgence traite en moyenne 6 patients par heure.
- La dispersion du nombre d'arrivés de cas urgents est en général très grande, il se peut très bien que, durant une certaine heure, il n'y ait qu'un seul arrivé ou aucun et que durant l'heure suivante il y en ait 12 ou 17. c'est le cas observé dans cet exemple.
- Pour éviter que le service soit trop souvent débordé. On doit l'organiser de telle sorte qu'il soit en mesure de traiter, par moments beaucoup plus que 6 patients par heure.
- La demande moyenne d'un service est un indice inadéquat des ressources nécessaires à sa prestation

Intervalles de temps	Nombre d'arrivées
1	5
2	2
3	4
4	3
5	0
6	12
7	10
8	17
9	11
10	1
11	2
12	5
Total	72

67H. BenbrahimENSIAS

DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

Généralités

Tendance Centrale

Dispersion

Mesure de Forme

Généralités

Etendu

Variance

IQ

Boite à Moustache

variation


Mesures de dispersion, pourquoi?

	Moyenne	Médiane	Ecart-type
Ensemble 1 : 20, 20, 20	20	20	0
Ensemble 2 : 10, 20, 30	20	20	8.16
Ensemble 3 : 1, 20, 39	20	20	15.51

- Exemple 2
- Dans les trois cas, la moyenne est égale à 20, ainsi que la valeur de la médiane.
- On ne saurait pour autant conclure que les trois ensemble sont identiques.
- la variabilité des données est plus grande dans l'ensemble 3 que dans les deux ensembles 2 et 1

68

H. Benbrahim



DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

Généralités

Tendance Centrale

Dispersion

Mesure de Forme

Généralités

Etendu

Variance

IQ

Boite à Moustache


variation

Etendu

- c'est la différence entre les valeurs extrêmes du caractère x observé :
 - $E = \max(x_i) - \min(x_i)$

69

H. Benbrahim



Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Généralités		Tendance Centrale	Dispersion	Mesure de Forme	
		Généralités	Etendu	Variance	IQ Boite à Moustache variation

Variance et Ecart-type

- La variance d'une distribution est la moyenne arithmétique des carrés des écarts à la moyenne.

$$S^2 = \frac{1}{N} \sum_{i=1}^k n_i (x_i - \bar{x})^2$$

$$S = \left(\frac{1}{N} \sum_{i=1}^k n_i (x_i - \bar{x})^2 \right)^{1/2}$$

- Sert à Caractériser de façon globale l'écart plus ou moins important de l'ensemble des valeurs de la distribution par rapport à la valeur moyenne.
- La variance permet de comparer la dispersion de deux ou plusieurs distributions d'une même variable.
- Dans le cas d'une variable continue groupée en classes on utilise les centres de classes à la place des x_i .

$$S^2 = \frac{1}{N} \sum_{i=1}^k n_i (c_i - \bar{x})^2$$

70 H. Benbrahim ENSIAS

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Généralités		Tendance Centrale	Dispersion	Mesure de Forme	
		Généralités	Etendu	Variance	IQ Boite à Moustache variation

Variance et Ecart-type

Exemple 1.

- Comparaison du nombre d'arrivés/heure de cas urgent dans deux hôpitaux différents

$$\bar{x}_{h1} = 5, \quad \bar{x}_{h2} = 5$$

$$s_{h1} = 1.6, \quad s_{h2} = 3.5$$

➤ La variabilité des arrivées dans h2 est plus grande que celle des arrivés dans h1.

71 H. Benbrahim ENSIAS

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion	
Généralités		Tendance Centrale	Dispersion	Mesure de Forme		
		Généralités	Etendu	Variance	IQ	Boîte à Moustache

L'écart interquartile

- **L'écart interquartile** = comprend 50% des observations, celles qui sont les plus centrales.

← Écart interquartile →

|-----|-----|-----|-----|

0 25% 50% 75% 100%

- **L'écart interquartile** = l'espace compris entre les quartiles 1 et 3

$$EQ = Q_3 - Q_1$$

- EQ est bien une mesure de dispersion, puisque plus les observations sont concentrées, plus Q_1 et Q_3 sont rapprochés et donc plus EQ est petite.
- EQ est moins utilisée que l'écart type.
- Est la mesure la plus appropriée pour des distributions fortement dissymétriques.

72 H. Benbrahim ENSIAS

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion	
Généralités		Tendance Centrale	Dispersion	Mesure de Forme		
		Généralités	Etendu	Variance	IQ	Boîte à Moustache

Boîte à moustaches

- La boîte à moustache ou box-plot est un résumé graphique d'une distribution.
- Le corps de la boîte est formé par le premier et troisième quartile et coupé par le deuxième quartile (médiane) plus deux autres valeurs qui sont:

$\text{Min}(1,5 * (Q_3 - Q_1) + Q_3, \text{Max}(x_i))$ et $\text{Max}(Q_1 - 1,5 * (Q_3 - Q_1), \text{Min}(x_i))$.

13	45	222	335	492	711
17	94	248	375	583	859
19	103	290	387	609	1693
31	104	295	444	618	1816
42	217	297	463	700	

revenu annuel en milliers de dirhams de 29 employés d'une société

73 H. Benbrahim ENSIAS

DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

Généralités

Tendance Centrale

Dispersion

Mesure de Forme

Généralités

Etendu

Variance

IQ

Boîte à Moustache


variation

Boîte à moustaches

- On repère sur la boîte à moustaches d’une variable:
 - L’échelle des valeurs de la variable, située sur l’axe vertical.
 - La valeur du 1er quartile Q1 (25% des effectifs), correspondant au trait inférieur de la boîte.
 - La valeur du 2ème quartile Q2 (50% des effectifs), représentée par un trait horizontal à l’intérieur de la boîte.
 - La valeur du 3ème quartile Q3 (75% des effectifs), correspondant au trait supérieur de la boîte.
 - Les 2 moustaches, délimitent les valeurs dites *adjacentes* qui sont déterminées à partir de l’écart interquartile (Q3-Q1).
 - Les valeurs dites extrêmes, atypiques, exceptionnelles, (*outliers*) situées au-delà des valeurs adjacentes sont individualisées. Elles sont représentées par des marqueurs (o, ou *, etc.).

74

H. Benbrahim



DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

Généralités

Tendance Centrale

Dispersion

Mesure de Forme

Généralités

Etendu

Variance

IQ

Boîte à Moustache

variation

Coefficient de variation

- Les mesures de dispersion considérées dans ce qui précède sont des mesures de dispersion absolue.
- Donc elles ne permettent pas de comparer la dispersion de deux ou plusieurs distributions d’une même variable mais de tendances centrales différentes
- Exemple 1** : Considérons les deux distributions suivantes:
 - L’écart-type est presque le même dans les deux distributions
- Moyenne (Médecins) ≠ Moyenne (infirmiers)**
- Nous nous pouvons pas comparer ces deux distributions en terme de dispersion.

Médecins	Fréquence	
xi	ni	ni*xi
1	4	4
2	4	8
3	5	15
4	2	8
Total	15	35
Moyenne	2,33	
Ecart-type	1,046	


Distribution du nombre de **médecins** dans 15 dispensaires d’une Wilaya

Infirmiers	Fréquence	
xi	ni	ni*xi
2	3	6
3	3	9
4	6	24
5	3	15
Total	15	54
Moyenne	3,6	
Ecart-type	1,055	

Distribution du nombre de **infirmiers** dans 15 dispensaires d’une Wilaya

- Rien à conclure en comparant seulement les mesures de dispersion absolue.
- Afin d’effectuer des comparaisons, nous avons besoin d’une mesure du degré de dispersion relative au sein de la distribution étudiée.

75



Définition	Données	Graphique	Indicateurs	Liaison	Conclusion	
Généralités		Tendance Centrale	Dispersion	Mesure de Forme		
		Généralités	Etendu	Variance	IQ	Boite à Moustache
		variation				

Coefficient de variation


- La dispersion relative la plus utilisée est le Coefficient de variation qui correspond à l'écart type exprimé en pourcentage de la moyenne.

$$CV = \frac{S}{\bar{X}} (100)$$

- C'est une mesure sans unité, donc plus pratique pour comparer deux distributions.
- Reprenons l'exemple précédent:

CV(Médecins)	=	1,046/2,33 = 0,45
CV(Infirmiers)	=	1,055/3,6 = 0,29

- La dispersion relative de la distribution des médecins est beaucoup plus grande que celle de la distribution des infirmiers.
- Le groupe des infirmiers est plus homogène que le groupe des médecins quand à leurs répartition dans les dispensaires.



Définition	Données	Graphique	Indicateurs	Liaison	Conclusion	
Généralités		Tendance Centrale	Dispersion	Mesure de Forme		
		Généralités	Etendu	Variance	IQ	Boite à Moustache
		variation				


Mesures de dispersion (résumé)

- Etendu** : c'est la différence entre les valeurs extrêmes du caractère x :

$$e = \max (x_i) - \min (x_i)$$
- Ecart interquartile** : $IQ = Q_3 - Q_1$
- Variance** : $\sigma^2 = 1/N \sum_{i=1}^p n_i (x_i - \bar{x})^2 = \sum_{i=1}^p f_i (x_i - \bar{x})^2$
- Ecart - type** : σ
- Le coefficient de variation** : $CV = \frac{\sigma}{\bar{x}}$

77

H. Benbrahim



Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
	Généralités	Tendance Centrale	Dispersion	Mesure de Forme	

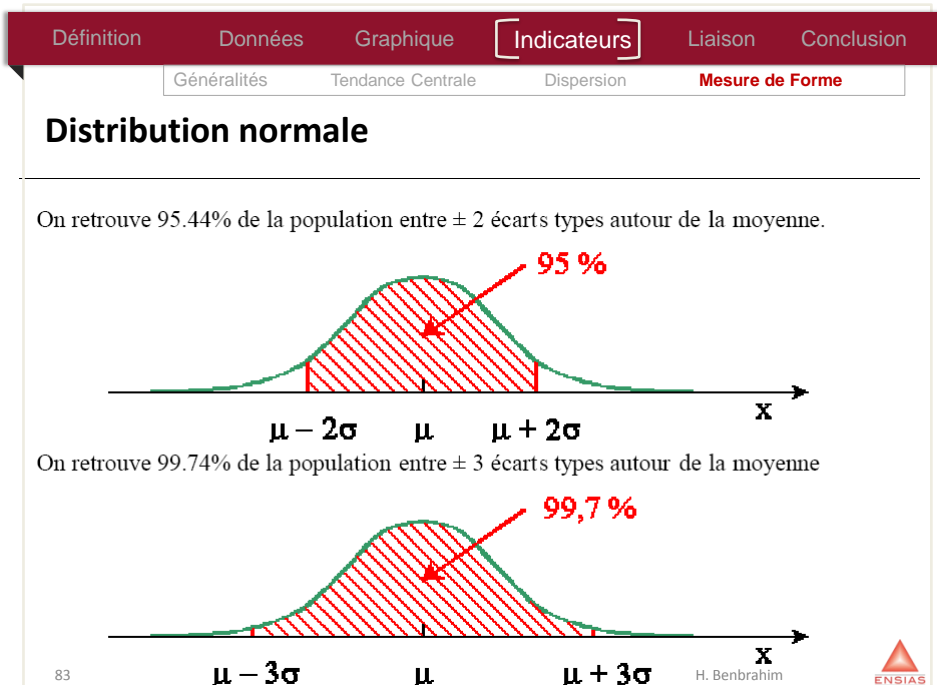
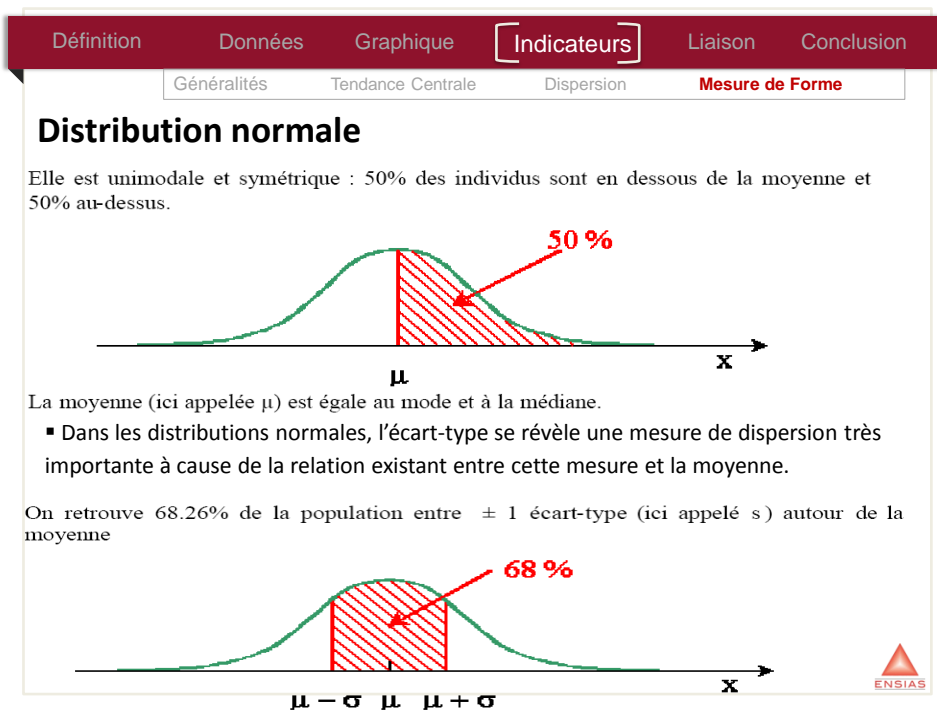
En guise de résumé

Type de variables	NOMINALE	ORDINALE	ECHELLE
Définition	Catégories Non ordonnées	Catégories ordonnées	Valeurs Numériques
Exemples	CSP, genre, statut marital	Niveau de Satisfaction, Tranches d'âge	Revenu, Poids, âge, Taille
Mesures de Tendance Centrale	Mode	Mode Médiane	Mode Médiane Moyenne
Mesures de Dispersion		Min/Max/	Min/Max Variance EQ
Graphes	Secteur (Bâton)	Bâton (Secteur)	Histogramme (Bâton)
Procédures	Fréquences	Fréquences	Fréquences, Descriptives

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
	Généralités	Tendance Centrale	Dispersion	Mesure de Forme	

Distribution normale


- On constate que les distributions de données continues présentent souvent une forme relativement régulière (en forme de cloche, atteignant progressivement le maximum avant de diminuer graduellement) qu'on appelle distribution normale ou distribution de Gauss.
- La distribution normale est une distribution très fréquente dans les phénomènes naturels → âge, hauteur, poids, erreurs aléatoires ...
- Une distribution est normale lorsque la majorité des sujets sont regroupés de façon symétrique autour de la moyenne.
- L'importance de cette distribution en statistique lui donne le nom de « distribution de référence ».
- Nous avons besoin de connaître certaines de ses propriétés.



Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Généralités		Tendance Centrale	Dispersion	Mesure de Forme	

Scores standardisés


- Le score Z (ou score standard) : est une technique statistique qui consiste à convertir un score individuel en un score standardisé, encore appelé score centré et réduit ou score Z.
- $Z = (X - \mu) / s$
- Le score Z permet de fournir une indication précise de la position du score de l'individu au sein de la distribution.
- Le score Z indique de combien en écart-type s'écarte une observation de sa moyenne.
- Pour comparer deux distributions obtenues sur des échelles d'intervalle d'un même échantillon, on transforme les données de chaque distribution en scores centrés réduits.
- Cette transformation consiste essentiellement à exprimer les données dans un système de mesure standard, correspondant à la courbe normale centrée réduite, symbolisé par Z
- Pour une distribution normale centrée réduite on a :**
 - Mode = médiane = moyenne = 0
 - l'écart-type vaut toujours 1 ($s = 1$)

84 H. Benbrahim 

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Généralités		Tendance Centrale	Dispersion	Mesure de Forme	

Vérification de la normalité d'une variable

- Vérifier si la distribution est en forme de cloche?
- Vérifier si la distribution est symétrique? (Histogramme)
- Vérifier si MOYENNE \approx MEDIANE \approx MODE
- Vérifier si Skewness AS \approx 0
- Vérifier si Kurtosis AP \approx 3
- Vérifier si il y a des Outliers (Explore - Box-plots)

85 H. Benbrahim 

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Généralités	Tendance Centrale	Dispersion	Mesure de Forme		

Mesures de forme

Exemple:

The figure shows three histograms. The top one is labeled 'Distribution symétrique' and shows a bell-shaped curve with the mean (M), mode (M), and median (X) all at the center. The bottom-left one is labeled 'Distribution étalée vers la gauche' and shows a long tail to the left, with the mean (M) and mode (M) to the right of the median (X). The bottom-right one is labeled 'Distribution étalée vers la droite' and shows a long tail to the right, with the mean (M) and mode (M) to the left of the median (X).

86

H. Benbrahim

ENSIA

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Généralités	Tendance Centrale	Dispersion	Mesure de Forme		

Coefficient d'asymétrie : SKWENESS

$$S = \frac{\mu_3}{\sigma^3}$$

Avec μ_3 correspondant au moment centré d'ordre 3
 σ , correspondant à l'écart type

$$\mu_3 = E[x - E(x)]^3$$

- Si $S = 0$, la distribution est symétrique comme la loi normale
- Si $S > 0$, la distribution penche à droite
- Si $S < 0$ la distribution penche à gauche

The figure shows two graphs. The left one is labeled 'S < 0' and shows a dashed blue curve that is skewed to the left. The right one is labeled 'S > 0' and shows a dashed red curve that is skewed to the right.

87

H. Benbrahim

ENSIA

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Généralités		Tendance Centrale	Dispersion	Mesure de Forme	

Coefficient d'aplatissement : KURTOSIS

$$K = \frac{\mu_4}{\sigma^4}$$

$$K = \frac{E[x - E(x)]^4}{(\sigma^2)^2}$$

Avec μ_4 correspondant au moment centré d'ordre 3
 σ , correspondant à l'écart type

- Si $K = 3$, la distribution a un coefficient d'aplatissement similaire à la distribution normale : **Distribution mésocurtique**
- Si $K > 3$, la distribution sera plus tassée que la distribution normale avec des queues épaisses: **Distribution Leptocurtique**
- Si $K < 3$, la distribution présente des queues plus fines que celle de la loi normale : **Distribution Platicurtique**

88

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Généralités		Tendance Centrale	Dispersion	Mesure de Forme	

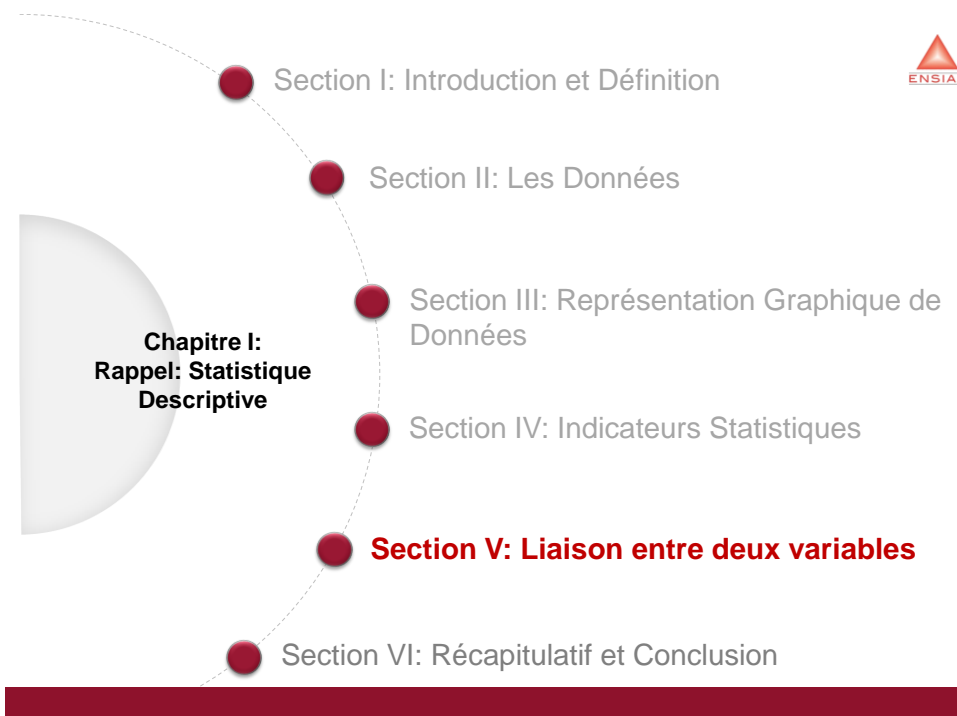
Les Indicateurs Statistiques

Caractéristiques de Tendances centrales	<ul style="list-style-type: none"> ▪ Moyenne arithmétique ▪ Médiane ▪ Mode
Dispersion	<ul style="list-style-type: none"> ▪ Etendu ▪ Ecart-type ▪ Indice de variabilité
Forme	<ul style="list-style-type: none"> ▪ Quantiles ▪ indice symétrie ▪ indice aplatissement

89

H. Benbrahim

ENSIAS



Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
	Var.. QL-QL	Var QT-QT		Var. QL-QT	
<ul style="list-style-type: none"> ▪ Liaisons entre deux variables qualitatives: <ul style="list-style-type: none"> • Coefficient χ^2 • Représentation graphique entre deux variables qualitatives ▪ Liaisons entre deux variables quantitatives <ul style="list-style-type: none"> • Représentation graphique • Coefficient de corrélation ▪ Liaisons entre une variables qualitative et quantitative <ul style="list-style-type: none"> • Représentation graphique quantitative contre qualitative 					

ENSIAS

Définition Données Graphique Indicateurs **Liaison** Conclusion

Var.. QL-QL Var QT-QT Var. QL-QT

Exploration de la liaison entre deux variables qualitatives (catégorielles)

92 H. Benbrahim ENSIAS

Définition Données Graphique Indicateurs **Liaison** Conclusion

Var.. QL-QL Var QT-QT Var. QL-QT

Liaisons entre deux variables qualitatives

- Soit x et y deux caractères qualitatifs.
- Les modalités de x sont notées $x_1, \dots, x_i, \dots, x_K$ et celles de y sont $y_1, \dots, y_j, \dots, y_L$.
- **Tableau de contingence**: pour chaque modalité x_i de x et y_j de y , n_{ij} est le nombre d'individus pour lesquels x vaut x_i et y vaut y_j .
- Les effectifs $n_{i.}$ et $n_{.j}$ désignent les totaux lignes et colonnes respectivement de ces individus:
 - ✓ $n_{i.} = \sum_j n_{ij}$
 - ✓ $n_{.j} = \sum_i n_{ij}$
 - ✓ avec $n = \sum_{ij} n_{ij} = \sum_i n_{i.} = \sum_j n_{.j}$

X \ Y	y_1	\dots	y_j	\dots	y_L	Total
x_1						
\vdots						
x_i			n_{ij}			$n_{i.}$
\vdots						
x_K						
Total			$n_{.j}$			n

93 H. Benbrahim ENSIAS

DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

Var.. QL-QLVar QT-QTVar. QL-QT

Coefficient χ^2 pour deux variables qualitatives

A partir d'un tableau de contingence, on peut calculer un **coefficient χ^2** (« chi-deux ») mesurant l'écart entre les n_{ij} et les « effectifs théoriques » que l'on aurait si x et y étaient indépendants, c'est-à-dire si les lignes, ou les colonnes du tableau, étaient proportionnelles.

$$\chi^2 = \sum_{i=1}^K \sum_{j=1}^L \frac{(n_{ij} - \frac{n_{i.} n_{.j}}{n})^2}{\frac{n_{i.} n_{.j}}{n}}$$

Valeur théorique
si indépendance

Le coefficient χ^2 est nul dans le cas de l'indépendance (profils identiques), et d'autant plus important que les profils sont différents entre eux.

94H. BenbrahimENSIAS

DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

Var.. QL-QLVar QT-QTVar. QL-QT

Coefficient χ^2 pour deux variables qualitatives: exemple

$$\chi^2 = \sum_{i=1}^K \sum_{j=1}^L \frac{(n_{ij} - \frac{n_{i.} n_{.j}}{n})^2}{\frac{n_{i.} n_{.j}}{n}}$$

Observée

théorique si indépendance

Différences au carrée

	Satisfait	Pas satisfait
Homme	70	30
Femme	30	70

Diviser par le théorique

	Satisfait	Pas satisfait
Homme	50	50
Femme	50	50

Différences au carrée

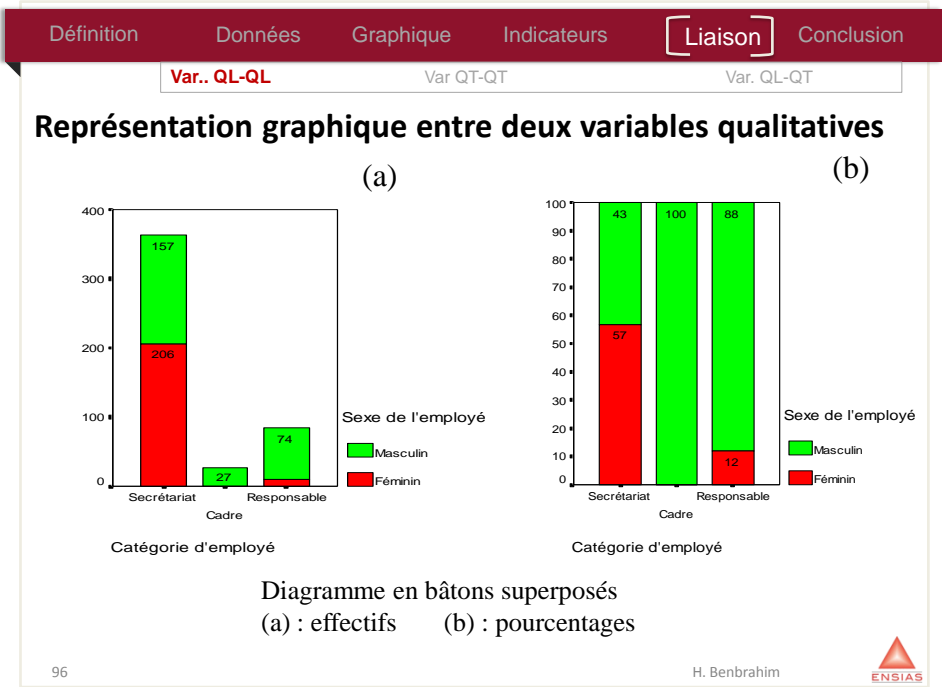
	Satisfait	Pas Satisfait
Homme	400	400
Femme	400	400

Diviser par le théorique

	Satisfait	Pas satisfait
Homme	8	8
Femme	8	8

Chi² = 32

H. BenbrahimENSIAS



Définition

Données

Graphique

Indicateurs

Liaison

Conclusion

Var.. QL-QL

Var QT-QT

Var. QL-QT

Liaisons entre deux variables quantitatives

- Si l'on dispose de l'observation de deux caractères x et y sur les mêmes n individus, on peut, en plus de l'étude séparée de chaque variable, décrire la liaison entre x et y au moyen d'un tableau de données brutes sous forme de n couples de valeurs (x_i, y_i) , $i = 1, \dots, n$.

ID	SALACT	SALDEB
1	\$57 000	\$27 000
2	\$40 200	\$18 750
3	\$21 450	\$12 000
4	\$21 900	\$13 200
5	\$45 000	\$21 000
6	\$32 100	\$13 500
7	\$36 000	\$18 750
8	\$21 900	\$9 750
9	\$27 900	\$12 750
10	\$24 000	\$13 500
11	\$30 300	\$16 500
12	\$28 350	\$12 000
13	\$27 750	\$14 250
14	\$35 100	\$16 800
15	\$27 300	\$13 500
16	\$40 800	\$15 000
17	\$46 000	\$14 250
18	\$103 750	\$27 510
19	\$42 300	\$14 250
20	\$26 250	\$11 550
21	\$38 850	\$15 000
.....		

97 ENSIAS

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
------------	---------	-----------	-------------	----------------	------------

Var.. QL-QL **Var QT-QT** Var. QL-QT

Liaisons entre deux variables quantitatives

Représentation graphique: Si x et y sont toutes deux variables quantitatives, la représentation graphique consiste en un **nuage de points** M_i de coordonnées (x_i, y_i) , $i = 1, \dots, n$.

98 H. Benbrahim ENSIAS

Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
------------	---------	-----------	-------------	----------------	------------

Var.. QL-QL **Var QT-QT** Var. QL-QT

Liaisons entre deux variables quantitatives

- **Notion de corrélation** : On dit qu'il y a corrélation entre deux variables observées sur des éléments d'une population lorsque les variations des deux variables quantitatives continues se produisent dans le même sens (corrélation positive) ou lorsque les variations sont de sens contraire (corrélation négative).
- **Coefficient de corrélation** : noté r est un indice qui rend compte numériquement de la manière dont deux variables quantitatives continues varient simultanément.

$$r_{xy} = \frac{\text{cov}(x,y)}{s_x s_y} \quad \text{où} \quad \text{cov}(x,y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

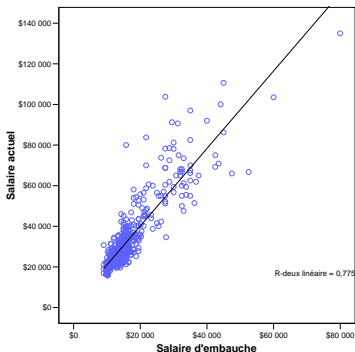
99 H. Benbrahim ENSIAS

Définition
Données
Graphique
Indicateurs
Liaison
Conclusion

Var.. QL-QL
Var QT-QT
Var. QL-QT

Exemple de calcul du coefficient de corrélation

Propriétés : r est un coefficient sans unité, indépendant de l'origine choisie, compris entre -1 et $+1$. Il est proche de -1 ou 1 s'il y a une relation presque affine entre x et y .



100

Matrice de corrélation		
	SALDEB	SALACT
SALDEB	1	
SALACT	0,88	1

Coefficient de corrélation dans l'exemple précédent = 0,88 révèle une forte liaison linéaire entre le salaire actuel et le salaire de début

H. Benbrahim

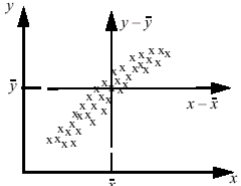
Définition
Données
Graphique
Indicateurs
Liaison
Conclusion

Var.. QL-QL
Var QT-QT
Var. QL-QT

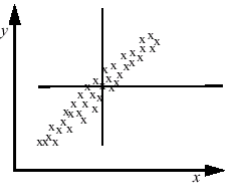
Interprétation du coefficient de corrélation

Coefficient de corrélation :

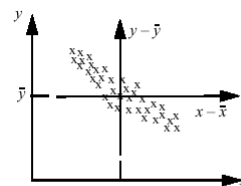
$$r_{xy} = \frac{\text{COV}(x,y)}{S_x S_y}$$



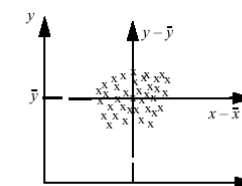
$r \approx 0,9$



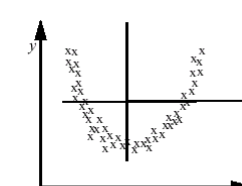
$r \approx 0,5$



$r < 0, |r| \text{ grand}$



$r \text{ voisin de zéro}$



$r \approx 0$

101


Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Var.. QL-QL	Var QT-QT			Var. QL-QT	

Liaisons entre une variable qualitative et une variable quantitative

- Si x est qualitative à k modalités, l'ensemble des n individus peut être subdivisé en k groupes sur lesquels x est constante.
- Si de plus y est quantitative, les k groupes peuvent être représentés par une boîte à moustache de façon à pouvoir les comparer.

102


H. Benbrahim



Définition	Données	Graphique	Indicateurs	Liaison	Conclusion
Var.. QL-QL	Var QT-QT			Var. QL-QT	

Exemples de Liaisons entre variable quantitative et variables qualitatives


- Analyse de la répartition des salaires
 - par sexe
 - par niveau d'étude
 - par catégories socioprofessionnelle



- Différents tableaux croisés 2 à 2 peuvent être générés
- Ou tableaux dynamiques multiples que l'on pourra pivoter selon les différentes dimensions.

103

H. Benbrahim



DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

Var.. QL-QLVar QT-QTVar. QL-QT

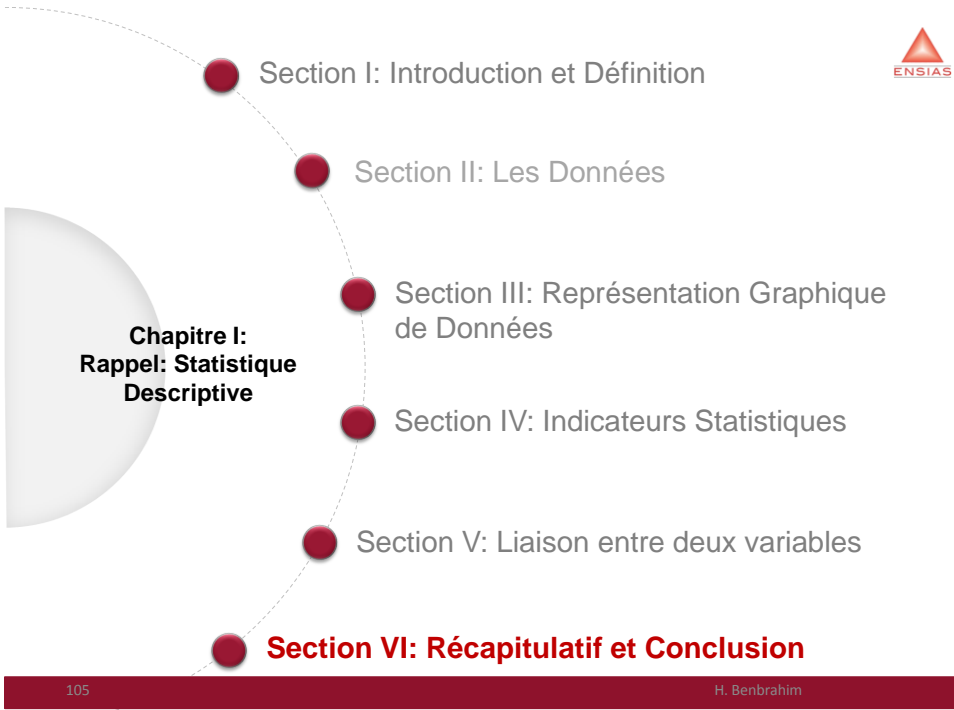
Représentation graphique quantitative contre qualitative

Exemple: Comparaison des poids des élèves en fonction du sexe (Féminin, Masculin).

		POIDS				
		Moyenne	Médiane	Q1	Q3	Ecart type
SEXE	Féminin	55	54	50	59	6
	Masculin	68	68	62	72	8

- Les mesures de position du poids des garçons sont supérieures à ceux des filles.
- De manière générale, la distribution des garçons est décalées vers le haut par rapport à celle des filles pour une dispersion quasi similaire.

104



Conclusion et récapitulatif?

106

H. Benbrahim



Exemple:

Il est usuel de modéliser la durée de bon fonctionnement ou durée de vie d'un système, ex., une ampoule électrique. Dans la pratique, l'utilisateur de ces ampoules est très intéressé par ces résultats. Il souhaite évidemment avoir une évaluation de leur durée de vie, de la probabilité qu'elles fonctionnent correctement pendant plus d'un mois, un an, etc...

C'est la statistique qui va permettre de résoudre ces problèmes. Pour cela, il faut faire une expérimentation, recueillir des données et les analyser.

On met donc en place ce qu'on appelle un essai ou une expérience. On fait fonctionner en parallèle et indépendamment les unes des autres $n = 10$ ampoules identiques, dans les mêmes conditions expérimentales, et on relève leurs durées de vie.

Admettons que l'on obtienne les durées de vie suivantes, exprimées en heures :

91.6 35.7 251.3 24.3 5.4 67.3 170.9 9.5 118.4 57.1

Moyenne = 83.15

Min = 5.4

Max = 251.3

(valeur aberrante due à une erreur de saisie ou bien d'expérimentation)

107

H. Benbrahim



Définition

Données

Graphique

Indicateurs

Liaison

Conclusion

Exemple:

Admettons que l'on obtienne les durées de vie suivantes, exprimées en heures :

91.6 35.7 251.3 24.3 5.4 67.3 170.9 9.5 118.4 57.1

Moyenne = 83.15

Min = 5.4

Max = 251.3

(valeur aberrante due à une erreur de saisie ou bien d'expérimentation)

➤ valeurs sensibles aux valeurs aberrantes

Médiane = 62.2 << 83.15

➤ une ampoule sur deux tombera en panne avant 62.2 h de fonctionnement.

➤ Cette propriété est caractéristique des distributions non symétriques dites « à queues lourdes » : un petit nombre d'ampoules auront une durée de vie nettement supérieure à la majeure partie des autres.

➤ On peut le remarquer sur l'histogramme aussi

➔ la moyenne et la médiane empiriques sont deux résumés de l'échantillon dont la connaissance simultanée peut être riche d'enseignements.

108 ➔ Quand la distribution est symétrique, moyenne et médiane sont proches



Définition

Données

Graphique

Indicateurs

Liaison

Conclusion

Exemple:

• On suppose que la variable étudiée est le salaire des marocains !

Pour un travail à temps plein, le salaire net mensuel moyen est de 10000 Dhs, alors que le salaire net mensuel médian est de 6500 Dhs.

➤ Un Marocain sur deux touche donc moins de 6500 Dhs par mois, mais un petit nombre de personnes ont un fort salaire, ce qui fait remonter la moyenne.

➤ Notons également que le seuil de pauvreté est défini comme la moitié du revenu médian, ce qui concerne ????? millions de personnes au Maroc.

109

H. Benbrahim




DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

Exemple:

•On donne les températures mensuelles moyennes, en degrés Celsius, à Meknes et à Rabat, calculées sur une période de 30 ans.

	J	F	M	A	M	J	J	A	S	O	N	D
Meknes	0	1	5	12	17	22	25	24	20	14	8	2
Rabat	9	11	12	13	14	16	17	17	18	16	13	9

- La température annuelle moyenne est de 12.5 à Meknes et de 13.7 à Rabat
- En se basant uniquement sur ces moyennes, on pourrait croire que les climats de ces deux villes sont similaires. Or il est clair que la différence de température entre l'hiver et l'été est beaucoup plus forte à Meknes qu'à Rabat.
- Pour le déceler, il suffit de calculer un indicateur qui exprime la variabilité des observations.

110H. Benbrahim

DéfinitionDonnéesGraphiqueIndicateursLiaisonConclusion

Exemple:

- L' écart-type des températures annuelles est de 8.8 à Meknes et de 3 à Rabat, ce qui exprime bien la différence de variabilité des températures entre les deux villes.

→ La variabilité doit toujours se comparer à la valeur moyenne.

- Une variabilité de 10 n'a pas le même sens si la température moyenne de référence est 12 ou 10000.
- Des données présentent une forte variabilité si l' écart-type est grand par rapport à la moyenne.

→ Le coefficient de variation

- Indicateur sans dimension.
- une variabilité significative si $cv_n > 0.15$.
- Si $cv_n < 0.15$, les données présentent peu de variabilité et on considère que la moyenne empirique à elle seule est un bon résumé de tout l' échantillon.

	\bar{x}_n	s_n^2	s_n	cv_n
ampoules	83.15	5540.2	74.4	0.89
Meknes	12.5	77.7	8.8	0.70
Rabat	13.7	8.9	3.0	0.22

111