

Cours d'Éléments de Statistique

Jean Christophe Meunier

MODULE 2

Étude des séries statistiques : Présentation des données

1^{ère} Bac, Commerce Extérieur
Année Académique 2015-2016



Introduction

- Statistique descriptive
 - Organisation, présentation et analyses des données d'une population/d'un échantillon
 - En vue d'explorer les caractéristiques de cette population/cet échantillon
- Etude des séries statistiques
 - Première étape de la statistique descriptive
 - Organise et présente les données afin d'en avoir un 'aperçu' intelligible et interprétable

A. Données brutes (non-groupées)

3

Illustration

- Lors d'une épreuve de sélection, 25 joueurs effectuent, chacun, 10 tirs au but
 - On note le nombre de buts marqués par chacun

	Prénom	Buts marqués		Prénom	Buts marqués
1	Kevin	5	14	Harry	2
2	Dieudonné	3	15	Dirk	5
3	Harry	4	16	Cédric	7
4	Dieumerici	2	17	Mbo	5
5	Geert	4	18	Sigfrig	2
6	Ronaldo	5	19	Mel	4
7	Regis	6	20	Carlos	6
8	Albert	3	21	Youssef	6
9	François	4	22	Rachid	9
10	Louis	5	23	Paolo	5
11	Merad	7	24	Roberto	2
12	Gad	6	25	Michel	5
13	Jo	4			

I. La variable

- Expression numérique du caractère observé
 - Peut être discrète ou continue
 - Dans l'exemple : nombre de buts marqués (variable discrète)
- Noté **X** ou **x**
- X_i ou x_i :
 - Chacune des valeurs individuelles que la variable peut prendre
 - Dans l'exemple : tous les nombres entiers allant de 2 à 9 (buts marqués)

5

II. Effectif

- Au sens général : nombre d'unités statistiques composant un groupe
- Effectif total (n ou N):
 - nombre d'unités statistiques observées pour la variable
 - Dans exemple : $N = 25 \rightarrow 25$ joueurs 'observés' qui ont un score/une valeur pour la variable 'nombre de buts'
 - Effectif total = taille de l'échantillon/de la population
- Effectif 'individuel' (n_i) :
 - nombre d'unités statistiques pour chaque valeur individuelle (x_i) que peut prendre la variable (x)
 - Dans exemple : pour $x_i = 2$ (buts marqués), on observe que 4 joueurs ont obtenu ce score (n_i)

6

II. Effectif

- Effectif vs. Effectif total

x_i	n_i (effectifs)
2	4
3	2
4	5
5	7
6	4
7	2
8	0
9	1
N (effectif total)	
	25

Effectif total (N)

$$n_1 + n_2 + \dots + n_c = N$$

Ou

$$\sum_{i=1}^{i=c} n_i = N$$

Où,

c = nombres de valeurs distinctes observées

7

II. Effectif

- Remarque
 - Le tableau des effectifs permet de visualiser la série statistique de manière groupée et intelligible

Sujets	Buts marqués (x_i)	Sujets	Buts marqués (x_i)
1	5	14	2
2	3	15	5
3	4	16	7
4	2	17	5
5	4	18	2
6	5	19	4
7	6	20	6
8	3	21	6
9	4	22	9
10	5	23	5
11	7	24	2
12	6	25	5
13	4		



x_i	n_i (effectifs)
2	4
3	2
4	5
5	7
6	4
7	2
8	0
9	1
N (effectif total)	
	25

8

II. Effectif

- Effectif cumulé (v_i)
 - Nombre de fois qu'on rencontre une valeur 'au plus' égale à x_i

x_i	n_i	v_i (effectifs cumulés)
2	4	4
3	2	6
4	5	11
5	7	18
6	4	22
7	2	24
8	0	24
9	1	25
	Effectif total	
	25	

Effectif cumulé (v_i)

$$v_i = \sum_{j=1}^{i=i} n_j$$

Logiquement, l'effectif cumulé pour la dernière valeur considérée est égal à l'effectif total

$$v_c = N$$

9

III. Fréquence

- On parle de fréquence ou de fréquence relative*
- Donne la même information que l'effectif mais en valeur relative par rapport à l'effectif total (N)
 - Ex : pour un échantillon dont effectif total $N = 100$ sujets
 - Un effectif de 10 sujets \rightarrow fréquence de 10%
- S'exprime soit en proportion (par rapport à l'unité) soit en pourcentage
 - Entre 0 = 0% et 1 = 100%
 - Ex : 0,25 \rightarrow 25 %

*Remarque : pour l'effectif, on parle parfois de 'fréquence absolue'

10

III. Fréquence

- Fréquence ou fréquence relative (f_i)

x_i	n_i	f_i
2	4	0,16
3	2	0,08
4	5	0,20
5	7	0,28
6	4	0,16
7	2	0,08
8	0	0,00
9	1	0,04
Effectif total		Somme des fréquences
25		1

Fréquence ou fréquence relative

$$f_i = \frac{n_i}{N} \text{ (en rapport à l'unité)}$$

Ou

$$f_i = \frac{n_i}{N} * 100 \text{ (en \%)}$$

Somme des fréquences = 1

$$\sum_{i=1}^{i=c} f_i = 1$$

11

III. Fréquence

- Fréquence cumulée (φ_i)
 - Idem qu'effectifs cumulés mais en valeurs relatives

Fréquence cumulée

$$\varphi_i = \frac{v_i}{N} = \frac{n_1 + n_2 + \dots + n_i}{N}$$

$$= \frac{n_1}{N} + \frac{n_2}{N} + \dots + \frac{n_i}{N}$$

$$= f_1 + f_2 + \dots + f_i$$

$$= \sum_{j=1}^{j=i} f_j$$

12

IV. Tableau effectifs/fréquences

	Effectifs	Effectifs cumulés	Fréquences	Fréquences cumulées
x_i	n_i	v_i	f_i	Φ_i
2	4	4	0,16	0,16
3	2	6	0,08	0,24
4	5	11	0,20	0,44
5	7	18	0,28	0,72
6	4	22	0,16	0,88
7	2	24	0,08	0,96
8	0	24	0,00	0,96
9	1	25	0,04	1,00
	Effectif total		Somme des fréquences	
	N			
	25		1	

13

B. Données groupées

14

Introduction

- Lorsque la variable ne prend qu'un nombre limité de valeurs, on peut travailler sur les données brutes
 - Ex : Illustration 'buts marqués', X ne peut prendre que quelques valeurs (entre 2 et 9 buts marqués)
- Lorsque la variable peut prendre un nombre important de valeurs, il faut grouper les données en classes (catégories) pour pouvoir calculer les effectifs/fréquences
 - Variables discrètes avec nombre important de valeurs distinctes
 - Variables continues

15

Illustration

- Points (sur 50) obtenus à l'examen théorique (x)

x_i	n_i	x_i	n_i	x_i	n_i	x_i	n_i	x_i	n_i
0	1	10	6	20	12	30	12	40	12
1	0	11	2	21	4	31	2	41	12
2	2	12	3	22	6	32	4	42	10
3	3	13	6	23	12	33	14	43	11
4	2	14	4	24	14	34	11	44	9
5	4	15	2	25	13	35	8	45	12
6	5	16	3	26	14	36	10	46	11
7	4	17	5	27	12	37	6	47	8
8	6	18	4	28	16	38	4	48	4
9	4	19	12	29	14	39	14	49	3
								50	2

Le nombre possible de valeurs (x_i) est trop important pour qu'on puisse se faire une représentation intelligible des données (effectifs/fréquences)

16

Illustration

Classes	Centres de classe	Effectifs	Effectifs cumulés	Fréquences	Fréquences cumulées
X_i		n_i	v_i	f_i	Φ_i
0-10	5	37	37	0,098 (9,8%)	0,098 (9,8%)
11-20	15	53	90	0,143 (14,3%)	0,241 (24,1%)
21-30	25	117	207	0,312 (31,2%)	0,553 (55,3%)
31-40	35	85	292	0,228 (22,8%)	0,781 (78,1%)
41-50	45	82	374	0,219 (21,9%)	1 (100%)
		Effectif total		Somme des fréquences	
		N			
		374		1	

On regroupe les valeurs en classe :

Ex : combien de personnes (effectif) ont obtenu entre 0 et 10 points

17

Regroupement en classes

- Quelles règles à suivre ?
 - Nombre de classe : entre 5 et 20
 - Assez grand : pour ne pas perdre trop d'information
 - Mais pas trop grand : pour faciliter la lecture et l'interprétation
 - Classes d'intervalles contigus et d'amplitude égale
 - Ex : 0-10, 11-20, 21-30, 31-40, 41-50 et non 0-2, 11-18, 22-30, 45-50
 - Chaque classe doit comporter au moins quelques éléments
- Pour obtenir un tableau aisément interprétable
- Remarque:
 - Inutile de regrouper par classe si les données sont d'emblée simples et facilement interprétables (peu de catégories → données brutes interprétables)

18

Présentation des classes

- Variables discrètes nombreuses (ex : points permis)

Classes	Centres de classe	Effectifs
X_i		n_i
0-10	5	37
11-20	15	53
...		

- Variables continues (ex : taille,...)

Classes	Centres de classe	Effectifs
X_i		n_i
[100;120cm[110	22
[120;130cm[130	44
...		

Utilisation d'accolade pour déterminer des intervalles contigus :

- 100-119,999cm
- 120-139,999cm
- ...

Par convention, la borne inférieure de la classe lui appartient, la borne supérieure ne lui appartient pas

19

C. Représentation graphique

20

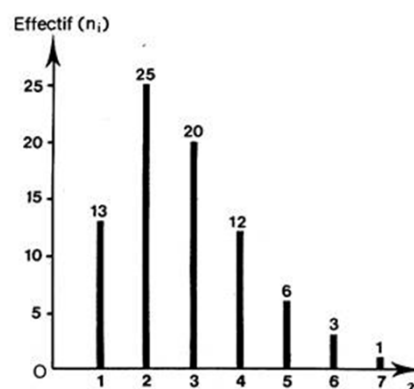
I. Graphe 'effectifs/fréquences'

- Objectif
 - Visualiser d'un coup d'œil le nombre de sujets (effectif/fréquence) pour chacune des modalités de réponses ou chacune des classes (x_i)
- Différentes options graphiques
 - Diagramme en bâtons
 - Histogramme
 - Polygone d'effectifs/de fréquences
 - Diagramme sectoriel ou 'camembert'
 - ...

21

I. Graphe 'effectifs/fréquences'

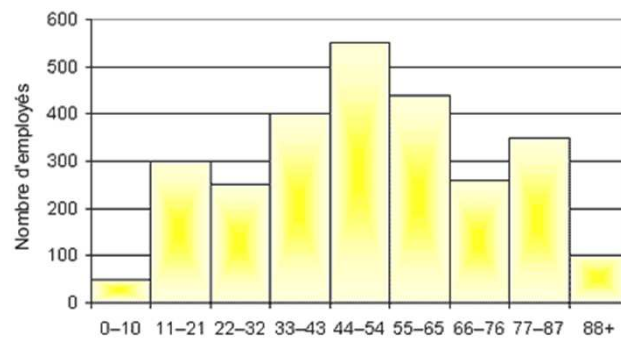
- Diagramme en bâtons



22

I. Graphe 'effectifs/fréquences'

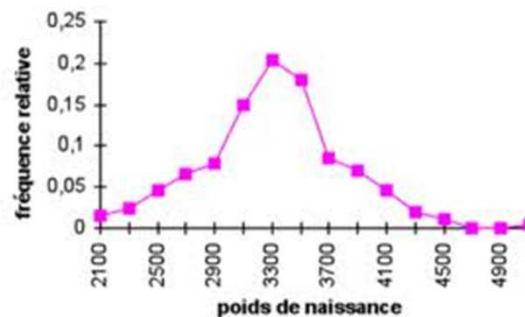
- Histogramme



23

I. Graphe 'effectifs/fréquences'

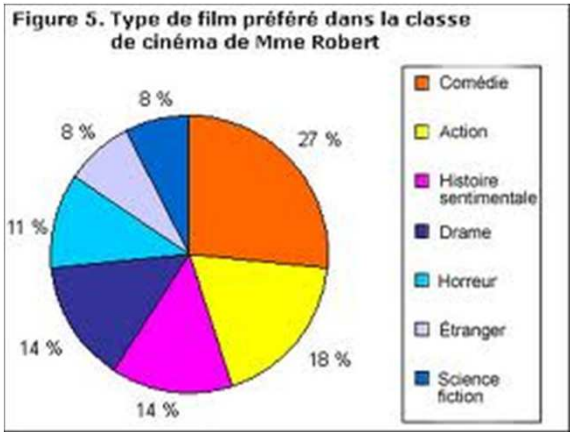
- Polygone d'effectifs/de fréquences



24

I. Graphe 'effectifs/fréquences'

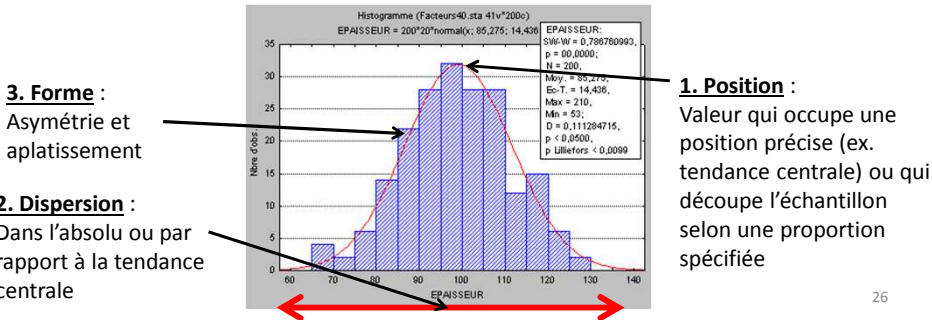
- Diagramme sectoriel ou 'camembert'



25

I. Graphe 'effectifs/fréquences'

- La ‘courbe de distribution’
 - Donnée par graphe de type diagramme en bâtons, histogramme, polygone,...
 - Comment l’échantillon se distribue-t-il en fonction des modalités de réponse ?
 - Donne différentes informations sur la distribution :



26

II. Graphe 'eff/fréq cumulé(e)s'

- Objectif

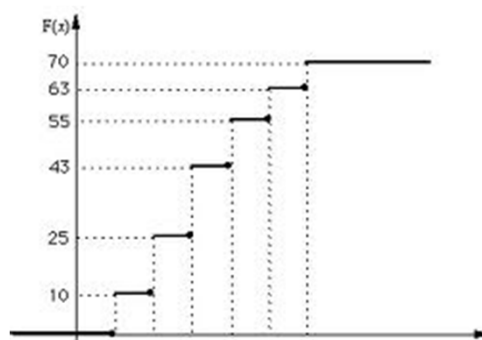
- Visualiser d'un coup d'œil le nombre de sujets ayant 'au plus' la valeur considérée (effectif/fréquence cumulé(e)) et ce, pour chacune des modalités de réponses (x_i) en partant successivement de la plus petite à la plus grande

27

II. Graphe 'eff/fréq cumulé(e)s'

1. Données brutes (non-groupées)

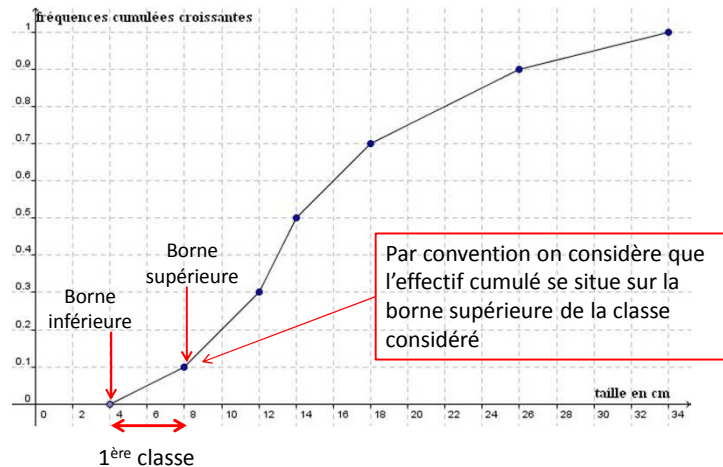
- Diagramme en escalier : chaque catégorie contient bien l'effectif cumulé indiqué



28

II. Graphe 'eff/fréq cumulé(e)s'

2. Données groupées



II. Graphe 'eff/fréq cumulé(e)s'

2. Données groupées

