

Statistique Descriptive

Plan du cours

- **Chap 1 : généralités sur la statistique**
- **Chap 2 : statistique descriptive univariée**
- **Chap 3 : statistique descriptive bivariée**

Statistique descriptive univariée

I. Tableaux statistiques

II. Représentations graphiques

- 1. Cas d'une variable qualitative**
- 2. Cas d'une variable quantitative discrète**
- 3. Cas d'une variable quantitative continue**

III. Indicateurs numériques

- 1. Indicateurs de tendance centrale**
- 2. Indicateurs de dispersion**
- 3. Forme d'une distribution**

Statistique descriptive univariée

I. Tableaux statistiques

II. Représentations graphiques

1. Cas d'une variable qualitative
2. Cas d'une variable quantitative discrète
3. Cas d'une variable quantitative continue

III. Indicateurs numériques

1. Indicateurs de tendance centrale
2. Indicateurs de dispersion
3. Forme d'une distribution



Cas des variables qualitatives ou quantitatives discrètes

Effectif partiel - effectif cumulé

On s'intéresse au nombre d'enfants à charge des 20 familles d'un village. Les observations recueillies sont présentées dans le tableau ci-dessous.

5	2	5	4	3	1	5	4	4	5
3	4	5	3	2	4	5	3	5	5

- n est le nombre total des observations : la taille de l'échantillon.
- k le nombre de modalités.
- x_i , $i \in [1, k]$ les modalités.
- n_i le nombre d'occurrences (l'effectif) de x_i dans l'échantillon.

Effectif partiel - effectif cumulé

Modalités	Effectifs	ECC	ECD
x_1	n_1	n_1	n
x_2	n_2	n_1+n_2	$n-n_1$
...
x_i	n_i	$n_1+n_2+\dots+n_i$	$n-n_1 - \dots - n_{i-1}$
...
x_k	n_k	n	n_k
Total	$n = \sum_i^K n_i$		

- Les Effectifs partiels sont également appelés fréquences absolues.

Effectif partiel - effectif cumulé

- D'abord on ordonne le tableau

1	2	2	3	3	3	3	4	4	4
4	4	5	5	5	5	5	5	5	5

Modalités	Effectifs	ECC	ECD
1	1	1	20
2	2	3	19
3	4	7	17
4	5	12	13
5	8	20	8
Total	20		

- Le nombre de familles ayant 4 enfants à charge est 5.
- Le nombre de familles ayant au plus 4 enfants à charge (au maximum 4 familles) est 12.
- Le nombre de familles ayant au moins 4 enfants à charge (au minimum 4 familles) est 13.

Fréquence partielle - Fréquence cumulée

Modalités	Fréquences	%	FCC	FCD
x_1	$f_1 = n_1/n$	$f_1 * 100$	f_1	1
x_2	$f_2 = n_2/n$	$f_2 * 100$	f_1+f_2	$1-f_1$
...
x_i	$f_i = n_i/n$	$f_i * 100$	$f_1+f_2+...+f_i$	$1-f_1 - -f_{i-1}$
...
x_k	$f_k = n_k/n$	$f_k * 100$	1	f_k
Total	$1 = \sum_i^K f_i$			

- Les fréquences partielles sont également appelées fréquences relatives.

Fréquence partielle - Fréquence cumulée

Modalités	Effectifs	Fréquences	%	FCC	FCD
1	1	0.05	5%	0.05	1
2	2	0.1	10%	0.15	0.95
3	4	0.2	20%	0.35	0.85
4	5	0.25	25%	0.6	0.65
5	8	0.4	40%	1	0.4
Total	20	1	100%		

- Le taux de familles ayant au plus 4 enfants à charge (au maximum 4 familles) est 12, soit 60 % des familles.
- Le taux de familles ayant au moins 4 enfants à charge (au minimum 4 familles) est 13 soit 65% des familles.



Cas des variables quantitatives

Fréquence absolue- Fréquence relative

- Une variable continue ne prend pas des valeurs isolées, mais des valeurs appartenant à des intervalles.
- On ne compte donc pas les modalités une à une, mais on définira des effectifs par intervalles, appelés **classes**.
- Une variable discrète comportant trop de valeurs est aussi traitée comme une variable continue.

Fréquence absolue- Fréquence relative

Les classes sont construites en respectant les règles suivantes :

1. le nombre de classes est compris entre 5 et 20 (de préférence entre 6 et 12) ;
2. chaque fois que cela est possible, les amplitudes des classes sont égales ;
3. chaque classe (sauf la dernière) contient sa borne inférieure, mais pas sa borne supérieure ;
4. lorsque plusieurs classes présentent des effectifs trop faibles ou nuls, il faut penser à les regrouper.

Fréquence absolue- Fréquence relative

- L'amplitude d'une classe est la distance entre la borne supérieure et la borne inférieure
- L'amplitude de chaque classe (dans le cas où elle est constante) est calculée de la manière suivante :
 - $a = \frac{L_{max} - L_{min}}{k}$
- L_{max} et L_{min} sont respectivement la valeur maximale et la valeur minimale de la série.

Fréquence absolue- Fréquence relative

- Il existe quelques règles empiriques pour le choix optimal du nombre de classes k :
 - « Première règle » : $k = \sqrt{N}$
 - Règle de Sturge : $k = 1 + (3,3 \times \log N)$
 - Règle de Yule : $k = 2,5\sqrt[4]{N}$
- N est la taille de l'échantillon.

Fréquence absolue- Fréquence relative

Exemple : nous avons recueilli le poids de 80 personnes présenté dans le tableau ci-dessous :

68	84	75	82	68	90	62	88	76	93
73	79	88	73	60	93	71	59	85	75
61	65	75	87	74	62	95	78	63	72
66	78	82	75	94	77	69	74	68	60
96	78	89	61	75	95	60	79	83	71
79	62	67	97	78	85	76	65	71	75
65	80	73	57	88	78	62	76	53	74
86	67	73	81	72	63	76	75	85	77

Fréquence absolue- Fréquence relative

- La plus petite valeur est égale à 53 et la plus grande valeur est égale à 97.
- L'étendue est de $97 - 53 = 44$.
- Règle de Sturge : $k = 1 + (3,3 \times \log 80) = 7,28$. On retient donc 7 classes.
- Règle de Yule : $k = 2,5 \sqrt[4]{80} = 7,47$. On retient donc 7 classes.
- L'amplitude constante des classes est $44/7 = 6,2$ ce qui équivaut à une amplitude de 7.

Fréquence absolue- Fréquence relative

Poids	centres	Effectifs	ECC	ECD	Fréquences	FCC	FCD
53-60	52	1	1	80	1.25%	1.25%	100%
60-67	57	2	3	79	2.5%	3.75%	98.75%
67-74	62	11	14	77	13.75%	17.5%	96.75%
74-81	67	10	24	66	12.5%	30%	82.5%
81-88	72	12	36	56	15%	45%	70%
88-95	77	21	57	44	26.25%	71.25%	55%
95-102	82	6	63	23	7.5%	78.75%	28.75%
85-89	87	9	72	17	11.25%	90%	21.25%
90-94	92	4	76	8	5%	95%	10%
95-97	96	4	80	4	5%	100%	5%
Total		80			100%		

- Le nombre de personnes pesant au moins 70 kg (au moins 70 kg ou plus de 69 kg) est 56 soit 70 % des personnes pesées.
- Le nombre de personnes pesant au plus 84 kg (au maximum 84 ou moins de 85 kg) est 63 soit 78,75 % des personnes pesées.

Application : énoncé

- On a recensé le nombre d'enfants à charge de 360 familles. Les résultats sont présentés dans le tableau ci-dessous :

Nb enfants à charge	0	1	2	3	4	5
Effectifs	103	115	95	35	10	2

- Quel est le nombre de familles possédant un nombre d'enfants inférieur ou égal à 2 ?
- Quel est le nombre de familles possédant un nombre d'enfants à charge supérieur ou égal à 3 ?
- Quelle est la proportion de familles possédant un nombre d'enfants à charge inférieur ou égal à 4 ?
- Quelle est la proportion de familles possédant un nombre d'enfants supérieur ou égal à 1 ?