Statistiques

Chapitre 1 . Séries statistique – Généralité

I) Définition

- Population : ensemble d'élément faisant l'objet d'une étude statistique.
- Echantillon : partie de la population.
- Unité statistique : Un élément de la population étudié dont on veut examiner un ou plusieurs caractères.
- Caractère : -qualitatif (nominale, ordinale) : non mesurable : (bon, mauvais; lieu d'habitation)

  -quantitatif (discrète, continue) : mesurable à l'aide d'une variable statistique.
- Variable statistique : -discrètes : prend des valeurs isolées.

  -continues : prend toutes les valeurs sur un intervalle.
- Séries statistique : ensemble des valeurs prises par une variable statistique sur l'ensemble de la population ou sur un échantillon.

II) Séries statistique d'un caractère quantitatif discret

Soit un échantillon de taille n (x1, x2, x3, …) sont les valeurs positives du caractère x, avec p le nombre de valeurs possible des x.

- Séries statistique : les valeurs prise par les n membres de l'échantillon.
- Effectif total : n
- Effectif partiel de xi (fréquence absolue) : c'est le nombre ni de fois qu'apparait la valeur xi (n1 +n2 +….np = n).
- Fréquence relative de xi : fi = $\frac{ni}{n}$ (f1 + f2 + … + fp = 1)
- Etude de la série : écart entre la plus grande et la plus faible des valeurs de xi
- Exemple :

On considère 100 familles de 4 enfants.

On étudie le caractère : nombre de garçons.

Série statistique : 3, 1, 3, 2, 4, 0, 1, 2, …

Effectif total : 100 étendu : 0 – 4 = 4

On compte les valeurs :

xi	0	1	2	3	4	Total
ni	7	20	43	25	5	100

Fréquence relative de famille avec 2 garçons : 43/ 100 : 43 %

III) Séries statistique d'un caractère quantitatif continu

Le nombre p de valeurs possible des x est infini (en réalité c'est fini à cause de la précision des mesures ➔ les fi sont presque nulles).

On va constituer les classes en divisant l'étendu de la série dans un certain nombre d'intervalles

Avec pour la classe i : un centre de classe xi et un effectif ui.

- Exemple : des nourrissons

Pesées effectuée à 10 g près. Donné le poids entre 2,24 et 4, 45.

	2,2-2,5	2,5-2,8	2,8 -3,1	3,1-3,4	3,4-3,7	3,7-4,0	4,0-4,3	4,3-4,6	total
classe	1	2	3	4	5	6	7	8	
centre	2,35	2,65	2,95	3,25	3,55	3,85	4,15	4,45	
effectif	5	11	24	40	42	20	13	6	161
fi	3,1	6,8	14,9	24,8	26,1	12,4	8,1	3,7	99 ,9 %

IV) Séries statistiques d'un caractère qualitatif

On groupe les résultats en autant de classes qu'il existe de modalités de caractères

- Variable cardinale : si les modalités peuvent être ordonnées (taille vestimentaire (+ ou – grande)).
- Variable nominale : si elles ne peuvent être ordonnées (couleurs)
- Variable dichotomiques : avec deux modalités.

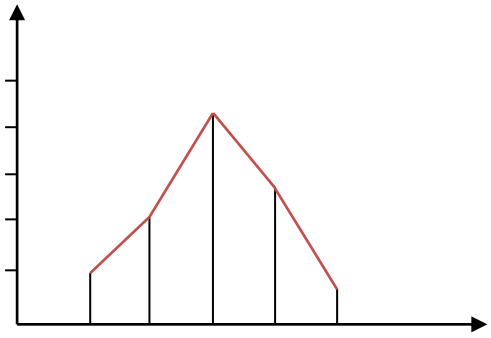
V) Représentation graphique des série statistiques

1) Cas des variables discrètes

- Diagramme en bâtons : (des effectifs et des fréquences absolues)

On trace ni = f(xi) ou fi = f(xi)

Exemple des familles de 4 enfants : (xi = nombre de garçons)



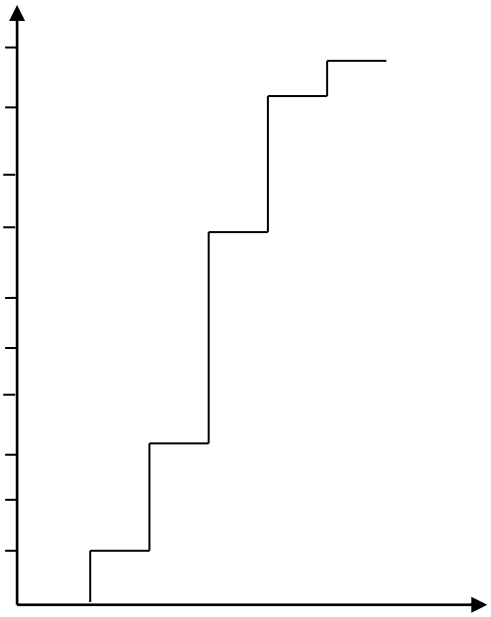
- Polygone : (des effectifs et de fréquences absolues)

On joint l'extrémité des bâtons par des segments (voir en rouge sur le graphique précédent)

- Diagramme cumulatif : (des effectifs et des fréquences)

Effectif cumulé jusqu'à la $i^{ème}$ valeur du caractère n1 + n2 + … + ni

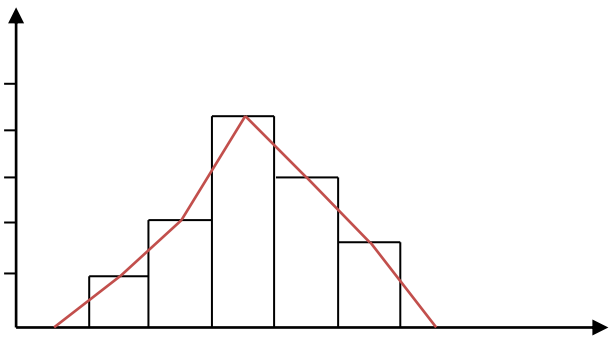
Fréquence relative cumulée jusqu'à la $i^{ème}$ valeur du caractère f1 + f2 + … + fi



2) Caractère continu

- Histogramme :

On va partager le domaine des valeurs de xi en classes : intervalles.



Si les classes sont inégales on les divise pour que toutes classes est la même ‘largeur’.

- Polygones : (des effectifs est des fréquences)

Il s’agit d’une ligne brisée joignant les milieux des sommets des rectangles. (en rouge)

Chapitre 2 . Parammètre de position et de dispersion

I) Paramètre de position

1) Moyenne arithmétique

Moyenne des x = $\frac{x1+x2+\cdots+xn}{n} = \frac{1}{n}\sum_{(i=n)}^{n} xi$

- Séries à caractère discret :

Moyenne des x = = $\frac{n1x1+n2x2+\cdots+npxp}{n} = \frac{1}{n}\sum_{(i=p)}^{p} nixi$

- Séries à caractère continu :

Moyenne des x = = $\frac{n1x1+n2x2+\cdots+npxp}{n} = \frac{1}{n}\sum_{(i=p)}^{p} nixi$ (xi est le centre de classe)

On peut faire une moyenne sur le ‘numéro de la classe’ : (zi : n° de classe ; M : minimum des xi; a : largeur de classe)

si xi = M + azi (zi : n° de classe) moy(x) = M + a*moy(z)

2) Médiane

On ordonne les valeurs de la série statistique et on prend la valeur du milieu.

Série discrète : série 1 : 3, 3, 4, **5**, 7, 7, 9 : médiane : Me = 5

Série 2 : 3, 3, 4, **5, 6**, 7, 7, 9 médianne : Me = (5 +6)/2 = 5,5

Série continu : On va considérer que la série est déjà ordonnée par classes et que la Me va tomber dans une classe qui correspond à la moitié de l’effectif total.

Me est, par approximation, dans la classe [l1, l2[avec F1 effectif cummulé en l1 et F2 en l2.

Me = l1 + $(\frac{n}{2}$ -F1$)\frac{l2-l1}{F2-F1}$

3) Quartiles

Les quartiles divisent les effectifs cumulés en 4 avec Q1, Q2 = Me, Q3.

Même principe que la médiane sauf que pour la formule des séries continues il faut changer le n/2 par n/4 et 3n/4.

4) Modes

Le mode est la valeur la plus fréquente (effectif le plus important)

Pour les <u>séries continues</u>, on a une classe dominante, mais on peut aussi trouver la valeur dominante.

Mode : $Mo = l1 + \frac{\Delta 1(l2-l1)}{\Delta 1+ \Delta 2}$

Où $\Delta 1$: excédant d'effectif de la classe modale par rapport à la classe inferieur et $\Delta 2$: excédant d'effectif de la classe modale par rapport à la classe supérieur

II) <u>Paramètre de dispersion</u>

1) Etendu

Ecart entre la plus grande et la plus petite valeur des x de la série.

2) Ecart moyen

Ecart de xi : $ei = |xi - moy(x)|$

Ecart moyen de la série : $moy(e) = \frac{1}{N}\sum_{i=1}^{n} ei$

<u>Cas discret</u> : $moy(e) = \frac{1}{n}\sum_{i=1}^{p} ni.ei = \sum_{i=1}^{p} fi.ei$

<u>Cas continu</u> : p classes de centre xi

3) Variance et écart-type

Variance : moyenne arithmétique des carrés des écarts des xi

Variance : $\sigma^2x = \frac{1}{N}\sum_{i=1}^{n}(xi - moy(x))^2$

Ecart-type : racine de la variance

Ecart-type : $\sigma x = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(xi - moy(x))^2}$

<u>Méthode rapide</u> :

(Cas continu xi : centre des classes)

$\sigma^2x = (\frac{1}{N}\sum_{i=1}^{n} xi^2) - moy(x)^2$ ➔ $\sigma^2x = (\frac{1}{N}\sum_{i=1}^{p} ni.\, xi^2) - moy(x)^2$

remarque : si xi = M + azi alors σx = a σz

4) Ecart-interquartile

L'écart-interquartile est représenté par Q3 – Q1 et il caractérise 50 % de la valeur de la série

5) Coefficient de variation

Coefficient de variation = $\frac{\sigma x}{moy(x)}$

Chapitre 3 . Ajustement de séries statistique

I) Méthode de lissage

1) Méthode des points médians

On trace l'enveloppe supérieur et inferieur avec les maximums/minimums locaux. Puis on trace un segment vertical pour chaque xi et on trace la nouvelle courbe qui passe par les milieux de chaque segments

2) Méthode de la moyenne mobile

On remplace yi par y'i = $\frac{y-+1+yi+ +yi+1}{3}$ (on effectue une moyenne sur trois point).

Sauf pour le premier et le dernier point (moyenne sur deux points.

Remarque : Possibilité de pondérer les moyennes.

II) Méthode d'ajustement

1) Méthode graphique

On trace une droite ayant autant de point au dessus qu'au dessous. Ensuite si on connait deux points de cette droite, on peut en déduire son équation.

2) Méthode des moindre carrés

On cherche yi' = f(xi) pour minimiser $\sum_{i=1}^{p}(yi - yi')^2$

P = nombre de points.

Pour mesurer la qualité de l'ajustement on utilise le coefficient de corrélation tel que :

$-1 \leq R \leq 1$

$$R = \frac{\sum_i (xi.yi) - \frac{1}{p}\sum_i xi \sum_i yi}{\sqrt{\left(\sum_i (xi^2) - \frac{1}{p}(\sum_i xi)^2\right)\left(\sum_i (yi^2) - \frac{1}{p}(\sum_i yi)^2\right)}}$$

a) Ajustement à l'aide d'une droite

Droite de la forme : y = ax + b où a et b sont donnés par :

$$a = \frac{p\sum xiyi - \sum xi \sum yi}{p\sum xi^2 - (\sum xi)^2} \qquad \text{et } b = moy(y) - a.moy(x)$$

b) Ajustement à l'aide d'une parabole

Parabole la forme : $y = ax^2 + bx + c$ où a, b et c sont donnés par :

$a\sum xi^2 + b\sum xi + cp = \sum yi$

$a\sum xi^3 + b\sum xi^2 + c\sum xi = \sum xiyi$

$a\sum xi^4 + b\sum xi^3 + c\sum xi^2 = \sum xi^2 yi$

c) Ajustement à l'aide d'une exponentielle

Exponentielle de la forme : $y = b.a^x$

Donc ➔ $\log y = \log b + \log a^x$

$Y = Ax + B$ où $y = \log y$, $B = \log b$ et $A = \log a$

$$A = \frac{p\sum xiYi - \sum xi \sum Yi}{p\sum xi^2 - (\sum xi)^2} \qquad \text{et } B = moy(Y) - A.moy(x)$$

d) Ajustement à l'aide d'une puissance

puissance de la forme : $y = b.x^a$

Donc ➔ $\log y = \log b + \log x^a$

$Y = aX + B$ où $y = \log y$, $X = \log x$ et $B = \log b$

$$a = \frac{p\sum XiYi - \sum Xi \sum Yi}{p\sum Xi^2 - (\sum Xi)^2} \qquad \text{et } B = moy(Y) - a.moy(X)$$