Les série statistiques doubles

Définition:

- Une série statistique double ou série statistique à deux variables est un ensembles des résultats issus de l'observation de deux caractères numériques sur une même population.

Exemple:

- La taille est le poids d'un groupe d'enfants.
- Le salaire et la qualification d'un ensemble de salariés.
- La température et la pression d'un milieu à différentes heures.

Notation et représentation des séries statistiques doubles:

- La sérié statistique des couples (x_i,y_j) sera notée (X,Y).
- Une série statistique double peut être donnée comme l'énumération d'un certain nombre de résultats, on distingue deux cas:
- 1) Les effectifs des valeurs égaux à 1.
- 2) Les effectifs des couples (x_i,y_j) égaux à n_{ij}.

Les effectifs des couples (x_i, y_i) égaux 1:

- Dans ce cas , une sérié statistique double se présente comme suit:

X_I	X_1	X_2	..	...	...	X_N
Y_I	Y_1	Y_2	...	...	...	Y_N

Exemple:

- On prélève un échantillon de 10 nouveau-nés pour lesquels, en plus de leur poids, on regarde leur taille à la naissance:

POIDS	2,1	2,5	2,65	2,8	3,1	3,35	3,65	3,75	4,245	4,4
TAILLE	40	45	43	50	48	51	49	51	53	55

La moyenne, la variance et l'écart type:

- Les variables statistiques X, Y ont chacune , une moyenne , une variance et un écart-type:

$$\bar{X} = \sum_{i=1}^{N} \frac{x_i}{N} \qquad \bar{Y} = \sum_{i=1}^{N} \frac{y_i}{N}$$

$$V(X) = \sum_{i=1}^{N} \frac{x_i^2}{N} - \bar{X}^2 \qquad V(Y) = \sum_{i=1}^{N} \frac{y_i^2}{N} - \bar{Y}^2$$

$$\sigma(X) = \sqrt{V(X)} \qquad \sigma(Y) = \sqrt{V(Y)}$$

Les effectifs des couples (x_i,y_j) égaux n_{ij}:

- Soit X et Y deux séries statistiques définie sur la même population.
- X prenant les valeurs $x_1,x_2,....,x_r$
- Y prenant les valeurs $y_1, y_2, ...,y_p$
- $(1 \leq i \leq r)$
- $(1 \leq j \leq p)$

Tableau d'effectifs:

X/Y	Y_1	…	Y_j	…	Y_p	TOTAL $N_{I\cdot}$
X_1	N_{11}		N_{1J}		N_{1P}	$N_{1.}$
…						
X_I	N_{I1}		N_{IJ}		N_{IP}	$N_{I.}$
…						
X_R	N_{R1}		N_{RJ}		N_{RP}	$N_{R.}$
TOTAL $N_{.J}$	$N_{.1}$		$N_{.J}$		$N_{.P}$	N

Définition:

- Si n_{ij} est l'effectif partiel d'un couple (x_i,y_j) et
- N l'effectif total.
- le rapport $f_{ij} = \frac{n_{ij}}{N}$ s'appelle la fréquence partiel du couple (x_i,y_j).
- <u>Remarque:</u>

$$\sum_i \sum_j n_{ij} = N$$

- N est l'effectif total

Remarque:

- La somme des effectifs partiels correspondant à la valeur x_i est égale à l'effectifs des individus pour lesquels X prend la valeur x_i, on représente par le symbole $n_{i.}$:

$$n_{i.} = n_{i1} + n_{i2} + \cdots + n_{ip}$$

Remarque:

- La somme des effectifs partiels correspondant à la valeur y_j est égale à l'effectifs des individus pour lesquels Y prend la valeur y_j, on représente par le symbole $n_{.j}$

$$n_{.j} = n_{1j} + n_{2j} + \cdots + n_{rj}$$

Remarque:

- Les nombres $n_{i.}$ et $n_{.j}$ sont appelés les effectifs partiels marginaux, donc:

$$\sum_i n_{i.} = \sum_j n_{.j} = N$$

- De la même manière on définit les fréquences partiels marginaux:
- On appelle fréquence marginale $f_{i.}$ de la valeur x_i de la variable X le rapport: $f_{i.} = \frac{n_{i.}}{N}$

Remarque:

- On appelle fréquence marginale $f_{.j}$ de la valeur y_j de la variable Y le rapport: $f_{.j} = \frac{n_{.j}}{N}$
- En fait les distributions $(x_i, n_{i.})$ $\{1 \leq i \leq r\}$ et

$(y_j, n_{.j})$ $\{1 \leq j \leq p\}$ sont les distributions des effectifs des séries X et Y respectivement.

Caractéristiques d'une série a deux variables quantitatives:

- Soit (X,Y) une série statistique double quantitative définie sur une même population.
- La moyenne de X : $\bar{X} = \sum_{i=1}^{r} \frac{n_{i.} x_i}{N}$
- La variance de X:

$$V(X) = \sum_{i=1}^{r} \frac{n_{i.} x_i^2}{N} - \bar{X}^2$$

- L'écart-type: $\sigma(X) = \sqrt{V(X)}$

Caractéristiques d'une série a deux variables quantitatives:

- La moyenne de Y : $\bar{Y} = \sum_{j=1}^{p} \frac{n_{.j} y_j}{N}$
- La variance de y:

$$V(Y) = \sum_{j=1}^{p} \frac{n_{.j} {y_j}^2}{N} - \bar{Y}^2$$

- L'écart-type de Y:

$$\sigma(Y) = \sqrt{V(Y)}$$

Définition:

- Etant donnée un couple (X,Y) de série statistiques quantitative définies sur une même population.
- les moyennes $\bar{X}, \bar{Y}$ s'appellent les moyennes marginales du série (X,Y).
- Les variances V(X),V(Y) s'appellent les variances marginales du série (X,Y).

Exemple:

- Soit la série double suivante:

X\Y	10	20	30	40	$n_{i.}$	$n_{i.}x_i$	$n_{i.}x_i^2$
2	2	0	1	0	3	6	12
3	3	1	3	0	7	21	63
4	7	4	9	2	22	88	352
5	0	2	1	1	4	20	100
$n_{.j}$	12	7	14	3	36=N	135	527
$n_{.j}y_j$	120	140	420	120	800		
$n_{.j}y_j^2$	1200	2800	12600	4800	21400		

Exemple:

- On calcul: $$\bar{X} = \sum_{i=1}^{r} \frac{n_{i.}x_i}{N} = \frac{137}{36} = 3{,}75$$

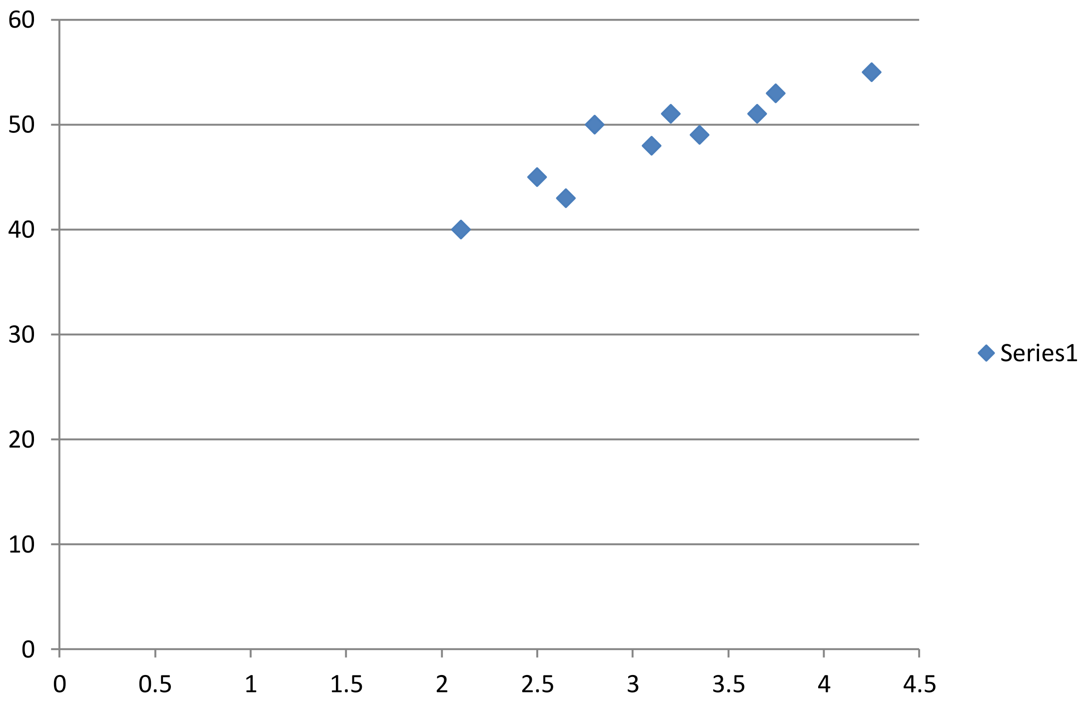
$$\bar{Y} = \sum_{j=1}^{p} \frac{n_{.j}y_j}{N} = \frac{800}{36} = 22{,}22$$

$$V(X) = \sum_{i=1}^{r} \frac{n_{i.}x_i^2}{N} - \bar{X}^2 = \frac{527}{36} - (3{,}75)^2 = 0{,}57 \Rightarrow \sigma(X) = 0{,}76$$

$$V(Y) = \sum_{j=1}^{p} \frac{n_{.j}{y_j}^2}{N} - \bar{Y}^2 = \frac{21400}{36} - (22{,}22)^2 = 100{,}72 \Rightarrow \sigma(Y) = 10{,}03$$

L'ajustement, La corrélation :

- Nuage de point:
- Dans un repère cartésien, on représente la série statistique à deux variables par les points M_{ij} de coordonnées (x_i,y_j).
- L'ensemble de ces points s'appelle le nuage de points de la série statistique.
- C'est une représentation des liens qu'il y a entre les x_i et les y_j



60
50
40
30
20
10
0
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
Series1

Le point moyen:

- Le point moyen du nuage de la série statistique à deux variables est le point M dont les coordonnées sont les moyennes $(\bar{X},\bar{Y})$ des X et Y respectivement.

La covariance:

• Définition:

- Si (X,Y) désigne un couple des séries statistiques quantitatives définies sur une même population prenant respectivement les valeurs $x_1, x_2, \ldots, x_r$ et $y_1, y_2, \ldots, y_p$
- n_{ij} désigne l'effectif partiel du couple (x_i, y_j)
- On appelle covariance du couple (X,Y) et on note Cov(X,Y)

$$Cov(X,Y) = \frac{1}{N}\left(\sum_{i}\sum_{j} n_{ij} x_i y_j\right) - \bar{X}\bar{Y} = \sigma_{XY}$$

Propriétés:

- $Cov(X,Y) = Cov(Y,X)$
- $Cov(X,X) = V(X)$
- $Cov(Y,Y) = V(Y)$
- Si X et Y sont indépendantes, alors: $Cov(X,Y)=0$

L’ajustement:

- Soit (X,Y) une série statistique double, une représentation graphique consiste à placer les points $M_i(x_i,Y_i)$ par rapport à un repère cartésien (OXY).
- Ajuster un ensemble de points consiste à déterminer une courbe simple " aussi proche que possible de points $M_i(x_i,Y_i)$