TARBES iut Occitanie UNIVERSITÉ TOULOUSE III

Fiche technique et formules du traitement des séries statistiques

BUT-Tech. de Co.
Semestre : 2
A.U. : 2021-2022
Prof. H. El-Otmany

NB : *cette fiche présente les techniques nécessaires* ***minimales*** *des sériés statistiques bivariées ; elle ne constitue donc pas un objectif mais un pré-requis pour le traitement des séries statistiques !*

1 Série statistique bivariée

Soit (X, Y) une série statistique quantitative observée sur une même population de taille n.

— Le point moyen ou le centre de gravité de la série (X, Y) est défini pour

a) les données simples par : $\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i, \quad \overline{Y} = \frac{1}{n}\sum\limits_{i=1}^{n} y_i.$

b) les données groupées par : $\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{p} n_i x_i, \quad \overline{Y} = \frac{1}{n}\sum\limits_{i=1}^{p} n_i y_i$ où $n = \sum\limits_{i=1}^{p} n_i.$

— La variance de Z (où Z joue le rôle de X ou de Y) est définie pour

a) les données simples par : $V(Z) = \overline{Z^2} - (\overline{Z})^2 = \frac{1}{n}\sum\limits_{i=1}^{n} z_i^2 - \left(\frac{1}{n}\sum\limits_{i=1}^{p} z_i\right)^2.$

b) les données groupées par : $V(Z) = \overline{Z^2} - (\overline{Z})^2 = \frac{1}{n}\sum\limits_{i=1}^{p} n_i z_i^2 - \left(\frac{1}{n}\sum\limits_{i=1}^{p} n_i z_i\right)^2$ où $n = \sum\limits_{i=1}^{p} n_i.$

— La covariance de la série (X, Y) est définie pour

a) les données simples par : $Cov(X, Y) = \overline{XY} - \overline{X} \times \overline{Y} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i y_i - \overline{X} \times \overline{Y}.$

b) les données groupées par : $Cov(X, Y) = \frac{1}{n}\sum\limits_{i=1}^{p} n_i x_i y_i - \overline{X} \times \overline{Y}$ où $n = \sum\limits_{i=1}^{p} n_i.$

2 Tableau de contingence

2.1 Distribution en effectif et en fréquence

— La distribution en effectif ou en fréquence de la série (X, Y) est donnée par le tableau suivant :

$X \backslash Y$	y_1	$\cdots$	y_j	$\cdots$	y_m	$\sum$
x_1	n_{11} ou f_{11}	$\cdots$	n_{1j} ou f_{1j}	$\cdots$	n_{1m} ou f_{1m}	$n_{1\cdot}$ ou $f_{1\cdot}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_i	$\vdots$	$\ddots$	n_{ij} ou f_{ij}	$\ddots$	$\vdots$	$n_{i\cdot}$ ou $f_{i\cdot}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_k	n_{k1} ou f_{k1}	$\cdots$	n_{kj} ou f_{kj}	$\cdots$	n_{km} ou f_{km}	$n_{k\cdot}$ ou $f_{k\cdot}$
$\sum$	$n_{\cdot 1}$ ou $f_{\cdot 1}$	$\cdots$	$n_{\cdot j}$ ou $f_{\cdot j}$	$\cdots$	$n_{\cdot m}$ ou $f_{\cdot m}$	$n = \sum\limits_{i=1}^{k} n_{k\cdot} = \sum\limits_{j=1}^{m} n_{\cdot j}$ ou 1

où

— n_{ij} est le nombre de fois où la modalité x_i de X et la modalité y_j de Y ont étés observées simultanément.

— $n_{i\cdot} = \sum_{j=1}^{m} n_{ij}$ est l'effectif marginal de la modalité x_i de X.

— $n_{\cdot j} = \sum_{i=1}^{k} n_{ij}$ est l'effectif marginal de la modalité y_i de Y.

— $f_{ij} = \frac{n_{ij}}{n}$ est la proportion $\frac{n_{ij}}{n}$ ou le pourcentage $\frac{n_{ij}}{n} \times 100\%$ des individus possédant à la fois la modalité x_i de X et la modalité y_j de Y.

— $f_{i\cdot} = \frac{n_{i\cdot}}{n}$ est la fréquence marginale de la modalité x_i de X indépendamment de Y.

— $f_{\cdot j} = \frac{n_{\cdot j}}{n}$ est la fréquence marginale de la modalité y_j de Y indépendamment de X.

2.2 Distribution marginale en effectif ou en fréquence

— La distribution marginale de X en effectif ou en fréquence est :

X	Effectif marginal
x_1	$n_{1\cdot}$
$\vdots$	$\vdots$
x_i	$n_{i\cdot}$
$\vdots$	$\vdots$
x_k	$n_{k\cdot}$
$\sum$	$n = \sum_{i=1}^{k} n_{i\cdot}$

X	fréquence marginale
x_1	$f_{1\cdot}$
$\vdots$	$\vdots$
x_i	$f_{i\cdot}$
$\vdots$	$\vdots$
x_k	$f_{k\cdot}$
$\sum$	1

— La distribution marginale de Y en effectif ou en fréquence est :

Y	Effectif marginal
y_1	$n_{\cdot 1}$
$\vdots$	$\vdots$
y_j	$n_{\cdot 1}$
$\vdots$	$\vdots$
y_m	$n_{\cdot m}$
$\sum$	$n = \sum_{j=1}^{m} n_{\cdot j}$

Y	fréquence marginale
y_1	$f_{\cdot 1}$
$\vdots$	$\vdots$
y_j	$f_{\cdot j}$
$\vdots$	$\vdots$
y_m	$f_{\cdot m}$
$\sum$	1

2.3 Moyenne marginale et variance marginale

— La moyenne marginale de X (respectivement de Y) est définie par :

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{k} n_{i\cdot} x_i \qquad \left(\text{resp. } \bar{Y} = \frac{1}{n}\sum_{j=1}^{m} n_{\cdot j} y_j\right)$$

— La variance marginale de X (respectivement de Y) est définie par :

$$S_X^2 = \frac{1}{n}\sum_{i=1}^{k} n_{i\cdot}(x_i - \bar{X})^2 \qquad \left(\text{resp. } S_Y^2 = \frac{1}{n}\sum_{j=1}^{m} n_{\cdot j}(y_j - \bar{Y})^2\right)$$

2.4 Dépendance entre deux variables X et Y

On considère la fréquence marginale $f_{i\cdot}$ du i^{eme} individus de la variable X et la fréquence marginale $f_{\cdot j}$ du j^{eme} individus de la variable Y. Pour $1 \leqslant i, j \leqslant n$, on note f_{ij} la fréquence du couple (x_i, y_j).

— Les variables X et Y sont **indépendantes** si pour tout $1 \leqslant i, j \leqslant n$, on a $f_{ij} = f_{i\cdot} \times f_{\cdot j}$.

— Les variables X et Y sont **dépendantes** s'il existe un $(i, j) \in \{1, \cdots, n\}$ telle que $f_{ij} \neq f_{i\cdot} \times f_{\cdot j}$.

3 Ajustement linéaire par la méthode des moindres carrés

— Le coefficient de corrélation linéaire est $r(X, Y) = \frac{Cov(X,Y)}{V(X)}$.

— La droite de régression ou des moindres carrés de Y en fonction de X est $y = ax + b$ où $a = \frac{Cov(X,Y)}{V(X)}$ et $b = \overline{Y} - a\overline{X}$.