SERIES STATISTIQUES DOUBLES

1 SERIES A DEUX VARIABLES

1.1 Généralités

Les séries statistiques doubles ont été développées dans le but d'étudier simultanément sur une même population de taille $n \geq 1$ deux caractères numérisables.
Soit $\Omega = \{\omega_1, \ldots, \omega_n\}$ une telle population et X,Y deux variables aléatoires réelles (v.a.r) définies sur Ω et prenant pour valeurs en $\omega_1, \ldots, \omega_n$ réspectivement $x_1, \ldots, x_n$ et $y_1, \ldots, y_n$.

DEFINITION 1 *On appelle série statistique double (s.s.d) associée aux caractères X et Y l'ensemble des couples de $\mathbb{R}^2$ donné par :*

$$\{(x_1, y_1), \ldots, (x_n, y_n)\}$$

On suppose de plus que $card\{X(\Omega)\} \geq 2$ et $card\{Y(\Omega)\} \geq 2$.

Remarque 1 *Classiquement, on a les formules donnant la moyenne et la variance de la v.a.r X définie sur Ω probabilisé par la* **loi uniforme discrète** *:*

$$\begin{cases} E(X) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i & \textit{notée} \quad \overline{x} \\ V(X) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} {x_i}^2 - \overline{x} & \textit{notée} \quad {\sigma_x}^2 \\ cov(X,Y) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) & \textit{notée} \quad \sigma_{X,Y} \end{cases}$$

Soit $(O, \vec{i}, \vec{j})$ un r.o.n du plan $\mathcal{P}$, on appelle **nuage de points** associé aux caractères X,Y le sous-ensemble de $\mathcal{P}$ constitué de l'ensemble suivant $\Gamma_n = \{M_1(x_1, y_1), \ldots, M_n(x_n, y_n)\}$.
On appelle **point moyen** du nuage, le point G isobarycentre deΓ_n :

$$\sum_{i=1}^{n} \overrightarrow{GM_i} = \vec{0} \Leftrightarrow G(\overline{x}, \overline{y})$$

Ci-dessous, deux exemples de nuages de points, sur le dessin de gauche ne semble se dégager aucune dépendance entre X et Y contrairement à celui de droite où semble se dégager une dépendance linéaire du type $y = ax + b$ entre X et Y.

But de la suite : Obtenir un ajustement affine du nuage, i.e approcher l'ensemble Γ_n par une droite permettant d'extrapôler une valeur de X ou de Y connaissant l'autre.

1.2 Covariance

THEOREME 1 *Pour tout couple (X, Y) de v.a.r définies sur Ω, on a la relation suivante :*

$$\boldsymbol{|\sigma_{X,Y}| \leq \sigma_X . \sigma_Y} \quad \textbf{(1)}$$

avec égalité ssi les points $\{M_i(x_i, y_i)\}_{1 \leq i \leq n}$ sont alignés (où $(x_i, y_i) = (X(\omega_i), Y(\omega_i))$.)

Preuve :
Considérons la fonction polynômiale P définie sur $\mathbb{R}$ par :
$P(\lambda) = V(\lambda X+Y) = \lambda^2 V(X)+2\lambda\sigma_{X,Y}+V(Y)$.Comme P est à valeurs positives où nulles, on a $\Delta_P \leq 0$, c'est à dire $\Delta_P = 4(\sigma_{X,Y})^2 - 4V(X)V(Y) \leq 0$
donc

$$|\sigma_{X,Y}| \leq \sigma_X.\sigma_Y.$$

Cas d'égalité :
.$\Leftarrow$:
Si tous les points de Γ_n sont alignés :

$$\exists (a,b) \in \mathbb{R}^2 \; t.q \; \forall i \in \{1,\ldots,n\} \; y_i = a.x_i + b$$

Tous les point sont sur la droite $(\mathcal{D})$ d'équation $y = a.x + b$, on a en particulier la relation $\overline{y} = a\overline{x} + b$.
On a d'une part $V(Y) = E((Y - E(Y))^2) = E((aX)^2) = a^2.V(x)$
et d'autre part $\sigma_{X,Y} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$ donc $\sigma_{X,Y} = \dfrac{a}{n}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2$.
Il en résulte que :

$$V(X).V(Y) = \sigma_{X,Y}^2 \text{ d'où } |\sigma_{X,Y}| = \sigma_X.\sigma_Y$$

$\Rightarrow$:
réciproquement si on a $|\sigma_{X,Y}| = \sigma_X.\sigma_Y$, d'après ce qui a été vu précédemment
on a $\Delta_P = {\sigma_{X,Y}}^2 - V(X)V(Y) = 0$, il existe donc $\lambda_0 \in \mathbb{R}$ t.q $P(\lambda_0) = 0$

$$\text{i.e} \qquad V(\lambda_0 X + Y) = 0$$

$$\text{i.e} \qquad \sum_{i=1}^{n}(\lambda_0 x_i + y_i - (\lambda_0\overline{x} + \overline{y}))^2 = 0$$

$$\text{i.e} \qquad \forall i \in \{1,\ldots,n\} \;\; \lambda_0 x_i + y_i - (\lambda_0\overline{x} + \overline{y}) = 0$$

$$\text{i.e} \quad \forall i \in \{1,\ldots,n\} \;\; M_i(x_i,y_i) \in (\mathcal{D}) : \lambda_0 x + y - (\lambda_0\overline{x} + \overline{y}) = 0)$$

DEFINITION 2 $\rho_{X,Y} = |\frac{\sigma_{X,Y}}{\sigma_X.\sigma_Y}|$ *est appelé le* **coefficient de corrélation** *des v.a.r X et Y et on a d'apres le résultat (1)* $|\rho_{X,Y}| \leq 1$.

2 Ajustement par la méthode des moindres carrés (M.M.C)

Description de la méthode :
Soit Γ_n un nuage de points associé à deux v.a.r X,Y et $(\mathcal{D})$ une droite du plan $\mathcal{P}$: pour tout point M_i du nuage, on note H_i le projeté de M_i sur la droite $(\mathcal{D})$ parallèlement à l'axe $(O,\vec{j})$.

On Cherche à déterminer **la** droite (Δ) du plan minimisant $\varepsilon(a,b) = \sum_{i=1}^{n} M_i H_i^2$, cette droite réalisant la meilleure approximation affine du nuage de points Γ_n.

THEOREME 2 *Il existe une unique droite* $(\Delta) : y = a.x + b$ *ajustant la s.s.d* Γ_n *par la m.m.c, ses coefficients sont donnés par :*

$$\begin{cases} a = \sigma_{X,Y}/\sigma_X^2 \\ b = \overline{y} - a.\overline{x} \end{cases}$$

Preuve :
. Analyse :

$$\begin{aligned} \varepsilon(a,b) &= \sum_{i=1} n(y_i - (ax_i + b))^2 \\ &= \sum_{i=1}^{n} y_i^2 + \sum_{i=1}^{n} a^2 x_i^2 + nb^2 + 2\sum_{i=1}^{n} ax_i b - 2\sum_{i=1}^{n} ax_i y_i - 2\sum_{i=1}^{n} y_i b \end{aligned}$$

Si $(D_{a,b})$ est une telle droite, on a nécessairement :

$$\begin{cases} \frac{\partial \varepsilon}{\partial a} = 0 \\ \frac{\partial \varepsilon}{\partial b} = 0 \end{cases} \quad \textbf{i.e} \quad \begin{cases} 2a\sum_{i=1}^{n} x_i^2 + 2b\sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} x_i y_i = 0 \\ 2nb + 2a\sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} y_i = 0 \end{cases}$$

$$i.e \begin{cases} a\sum_{i=1}^{n} x_i^2 + (\overline{y} - a\overline{x})\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i y_i = 0 \\ b = \overline{y} - a\overline{x} \end{cases}$$

$$i.e \begin{cases} a(\sum_{i=1}^{n} x_i^2 - \overline{x}\sum_{i=1}^{n} x_i) + \overline{y}\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i y_i = 0 \\ b = \overline{y} - a\overline{x} \end{cases}$$

$$i.e \begin{cases} aV(X) = \sigma_{X,Y} \\ b = \overline{y} - a\overline{x} \end{cases} \qquad i.e \begin{cases} a = \dfrac{\sigma_{X,Y}}{\sigma_X^2} \\ b = \overline{y} - a\overline{x} \end{cases}$$

Synthèse : Est ce que le couple (a_0, b_0) ainsi déterminé réalise pour autant un minimum de la fonction ε ? Etant donnée la nature du problème considéré, ce ne peut être qu'un minimum. Il reste cependant à vérifier que c'est un minimum (absolu) !

$$\text{Avec les notations de Monge une c.n.s est } \begin{cases} r > 0 \\ s^2 - r.t < 0 \end{cases}$$

où $r = \dfrac{\partial^2 \varepsilon}{\partial a^2}$, $s = \dfrac{\partial^2 \varepsilon}{\partial a \partial b}$, $t = \dfrac{\partial^2 \varepsilon}{\partial b^2}$. Une fois le calcul des dérivées partielles au point (a_0, b_0) effectuées , on trouve en tout et pour tout :

$$\begin{cases} s(a_0, b_0) = 2\sum_{i=1}^{n} x_i \\ r(a_0, b_0) = 2\sum_{i=1}^{n} x_i^2 \\ t(a_0, b_0) = 2n \end{cases}$$

De sorte que le calcul brut fournit : $\dfrac{s^2 - r.t}{4} = (\sum_{i=1}^{n} x_i)^2 - n.\sum_{i=1}^{n} x_i^2$.

Quel est son signe ? Considérons pour cela les vecteurs de $\mathbb{R}^n$ suivants :

$$\overrightarrow{X} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \qquad \overrightarrow{\mathbf{1_n}} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

D'après la relation de **Cauchy-Schwartz** on a : $|\overrightarrow{\mathbf{X}}.\overrightarrow{\mathbf{1_n}}| \leq \| \overrightarrow{\mathbf{X}} \| \| \overrightarrow{\mathbf{1_n}} \|$, ce qui se traduit analytiquement par $|\sum_{i=1}^{n} x_i| \leq \sqrt{\sum_{i=1}^{n} x_i^2} \times \sqrt{n}$, il en résulte donc :

$$(\sum_{i=1}^{n} x_i)^2 \leq n \times \sum_{i=1}^{n} x_i^2 \text{ i.e } \frac{s^2 - r.t}{4} \leq 0 \quad (\mathbf{2})$$

Comme les cas d'égalité de la relation de Cauchy-Schwartz ne sont réalisés que lorsque les vecteurs sont positivement dépendants donc dépendants, l'inégalité (2) est forcément stricte. Sinon, $\exists \lambda \in \mathbb{R}^+$ t.q $\overrightarrow{\mathbf{X}} = \lambda \overrightarrow{\mathbf{1_n}}$ et on a alors :

$$x_1 = \ldots = x_n = \lambda$$

Ce qui est contradictoire, car on a par définition d'une s.s.d $card\{X(\Omega)\} \geq 2$

Remarque 2 *Il existe une autre façon d'établir ce théorème sans passer par le calcul différentiel et les notations de Monge. On peut parvenir à ce résultat en factorisant de manière "adéquate" le développement de $\varepsilon(a, b)$, cependant cette méthode reste très lourde par son côté calculatoire et très astucieuse.*

DEFINITION 3 *:*

La droite $(\mathcal{D}) : y - \overline{y} = \frac{\sigma_{X,Y}}{\sigma_X^2}(x - \overline{x})$ *est la* **droite de régression de Y en X**.

Celle de X en Y est donnée par : $(\Delta) : x - \overline{x} = \frac{\sigma_{X,Y}}{\sigma_y^2}(y - \overline{y})$.*En particulier, on note que le point moyen du nuage,* $G(\overline{x}, \overline{y})$ *est sur cette droite.*

THEOREME 3 *:*
Les droites de régression $(\mathcal{D})$ *et* (Δ) *sont confondues s.s.i que* $\sigma^2_{X,Y} = \sigma^2_x.\sigma^2_y$*, s.s.i tous les points du nuage sont alignés.*

preuve : Comme G est à la fois sur $(\mathcal{D})$ et sur (Δ), les deux droites sont confondues s.s.i leurs coefficients directeurs sont égaux.

ssi $\frac{\sigma_{X,Y}}{\sigma^2_x} = \frac{\sigma^2_y}{\sigma_{X,Y}}$

ssi $\sigma^2_{X,Y} = \sigma^2_x.\sigma^2_y$

ssi tous les points du nuage sont alignés (d'après **(1)**)

Remarque 3 *:*
Comme $\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X.\sigma_Y}$*, on* $|\rho_{X,Y}| = 1$ *s.s.i* $|\sigma_{X,Y}| = \sigma_X.\sigma_Y$ *s.s.i les points du nuage sont alignés.*
Ainsi, $\rho_{X,Y}$ *va servir d'indice afin de savoir si l'ajustement affine par la m.m.c est valable ou non. En pratique, il est envisageable dès que* $|\rho_{X,Y}| \geq 0.8$*.*

3 Applications

3.1 Autres ajustements

3.1.1 Ajustements exponentiels

Si le nuage Γ_n a une forme exponentielle, le nuage serait "proche" d'être sur le graphe de la fonction :

$$f : \begin{cases} \mathbb{R}^+ \longrightarrow \mathbb{R}^+ \\ x \longmapsto \lambda.\exp(\mu.x) = \exp(a.x + b) \ [\ \lambda > 0\ ] \end{cases}$$

Dès lors, le nuage $\Gamma_n{}' = \{M_i'(x_i, \ln(y_i)\}_{1 \leq i \leq n}$ est ajustable sur la droite d'équation $y = a.x + b$ par la m.m.c (...), ce qui permet dès lors d'exprimer une relation fonctionnelle du type "$y = \exp(a.x + b)$" entre les v.a.r X et Y.

3.1.2 Ajustement logarithmique

Si le nuage Γ_n a une forme logarithmique, le nuage serait "proche" d'être sur le graphe de la fonction :

$$g : \begin{cases} \mathbb{R}^+ \longrightarrow \mathbb{R} \\ x \longmapsto b + a.\ln(x) \end{cases}$$

Dès lors, le nuage $\Gamma_n{}'' = \{M_i''(\ln(x_i), y_i\}_{1 \leq i \leq n}$ est ajustable sur la droite d'équation $y = a.x + b$ par la m.m.c (...), ce qui permet d'exprimer une relation fonctionnelle du type "$y = b + a.\ln(x)$" entre les v.a.r X et Y.

3.1.3 Ajustement puissance

Si le nuage Γ_n a une forme puissance, le nuage serait "proche" d'être sur le graphe de la fonction :

$$h : \begin{cases} \mathbb{R}^+ \longrightarrow \mathbb{R} \\ x \longmapsto a.x^b \end{cases}$$

Dès lors, le nuage $\Gamma_n{}'' = \{M_i''(\ln(x_i), \ln()y_i\}_{1 \leq i \leq n}$ est ajustable sur la droite d'équation $y = \alpha.x + \beta$ par la m.m.c (...), ce qui permet d'exprimer une relation fonctionnelle du type "$y = \alpha.x^{\beta}$" [1] entre les v.a.r X et Y.

Références

[1] Brigitte Bajou, Michaêl Ranguin, Xavier Sorbe. *Oral du CAPES : préparation à l'épreuve d'exposé*. Masson, 1996.

[2] Thomas Wonnacott, Ronald Wonacott.*Statistique*. Economica, 1995.

[3] *des extraits de la même leçon d'oral suivie à l'I.U.F.M de Saint-Etienne...il y'a quelques années de celà !*

[1] les coefficients α,β linéaires et puissances étant non forcément identiques.