VERSITÉ DE BÉJAIA — جامعة عبد الرحمن ميرة — TASDAWIT N'BGAYET

جامعة بجاية
Tasdawit n'Bgayet
Université de Béjaïa

Cours Math310

Première partie : statistique descriptive (Suite)

Chapitre 2: Séries statistiques à deux variables

M. BEZOUI

10 janvier 2012

Plan de travail

Un peu de silence S.V.P

Plan de travail

Introduction

Etudier le lien entre deux variables (caractères) sur une même population (taille et point d'un nouveau né, vitesse et consommation du véhicule,...). On appelera ces deux caractères X et Y, définis sur une population de taille n (n individus), on s'interesse à étudier le lien entre X et Y. Les données brutes sont les couples $(x_i, y_j), i = \overline{1, n}$.

Plan de travail

Tableau de contingence (Définition)

Le tableau de contingence est un moyen particulier de représenter simultanément deux caractères observés sur une même population, s'ils sont discrets ou bien continus et regroupés en classes. Les deux caractères sont X et Y, la taille de l'échantillon est n. Les modalités ou classes de x seront notées $x_1, x_2, \cdots, x_p$, celles de Y sont notées $y_1, y_2, \cdots, y_q$.

Notations

On note :

1. n_{ij} : l'effectif conjoint de x_i et y_j : c'est le nombre d'individus pour lesquels X prend la valeur x_k et Y la valeur y_j,
2. $n_{i\bullet} = \sum_{j=1}^{q} n_{ij}$: l'effectif marginal de x_i : c'est le nombre d'individus pour lesquels X prend la valeur x_i,
3. $n_{\bullet j} = \sum_{i=1}^{p} n_{ij}$: l'effectif marginal de y_j : c'est le nombre d'individus pour lesquels Y prend la valeur y_j.

Présentation

On représente ces valeurs dans un tableau à double entrée, dit tableau de contingence.

$X \setminus Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_q	Total
x_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1q}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2q}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\cdots$	$\cdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	$\vdots$	n_{ij}	$\vdots$	n_{iq}	$n_{i\bullet}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_p	n_{p1}	n_{p2}	$\vdots$	n_{pj}	$\vdots$	n_{pq}	$n_{p\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$	$\cdots$	$n_{\bullet j}$	$\cdots$	$n_{\bullet q}$	$n_{\bullet\bullet} = n$

Plan de travail

Distribution conjoint

fréquence jointe notée f_{ij}, c'est la proportion d'individus ayant pris simultanément la modalité x_i de X et la modalité y_j de Y.

$$f_{ij} = \frac{n_{ij}}{n}$$

Définition

On appelle distribution conjointe du couple (X, Y) ou encore série statistique à deux dimensions, les données des couples $(x_i, y_j), \ i = \overline{1, p}, \ \ j = \overline{1, q}$ et les fréquences jointes $f_{ij}, \ i = \overline{1, p}, \ \ j = \overline{1, q}$.

Plan de travail

fréquence marginale

Définition

On appelle fréquence marginale associée à la modalité x de X notée $f_{i\bullet}$ la proportion :

$$f_{i\bullet} = \frac{n_{i\bullet}}{n} = \sum_{i=1}^{q} f_{ij}$$

$$\sum_{i=1}^{p} f_{i\bullet} = 1$$

Distribution marginale

Définition

On appelle distribution marginale de la variable X (resp. Y), les données des couples $(x_i, f_{i\bullet})$, $i = \overline{1, p}$, (resp. $(y_j, f_{\bullet j}), j = \overline{1, q}$).

Caractéristiques numériques des distributions marginales

$$
\begin{aligned}
\overline{x} &= \tfrac{1}{n}\textstyle\sum_{i=1}^{p} n_{i\bullet}x_i = \sum_{i=1}^{p} f_{i\bullet}x_i \\
V(x) & \tfrac{1}{n}\textstyle\sum_{i=1}^{p} n_{i\bullet}x_i^2 - \overline{x}^2 = \sum_{i=1}^{p} f_{i\bullet}x_i^2 - \overline{x}^2 \\
\sigma(x) &= \sqrt{V(x)} \\
\overline{y} &= \tfrac{1}{n}\textstyle\sum_{j=1}^{q} n_{\bullet j}y_j = \sum_{j=1}^{q} f_{\bullet j}y_j \\
V(y) & \tfrac{1}{n}\textstyle\sum_{j=1}^{p} n_{\bullet j}y_j^2 - \overline{y}^2 = \sum_{j=1}^{q} f_{\bullet j}y_j^2 - \overline{y}^2 \\
\sigma(y) &= \sqrt{V(y)}
\end{aligned}
$$

Exemple illustratif

Soit le tableau de contingence suivant :

$X \setminus Y$	0	1	2	$n_{i\bullet}$
1	6	3	1	$n_{1\bullet} = 10$
2	5	5	5	$n_{2\bullet} = 15$
$n_{\bullet j}$	$n_{\bullet 1} = 11$	$n_{\bullet 2} = 8$	$n_{\bullet 3} = 6$	$n = 25$

Le tableau de fréquences correspondant est :

$X \setminus Y$	0	1	2	$f_{i\bullet}$
1	0.24	0.12	0.04	$f_{1\bullet} = 0.4$
2	0.44	0.32	0.24	$f_{2\bullet} = 0.6$
$f_{\bullet j}$	$f_{\bullet 1} = 0.44$	$f_{\bullet 2} = 0.32$	$f_{\bullet 3} = 0.24$	$f_{\bullet\bullet} = 1$

Distributions marginales

Distribution marginale de X

x_i	$n_{i\bullet}$	$f_{i\bullet}$
1	10	0.4
2	15	0.6
Total	25	1

Distribution marginale de X

y_i	$n_{i\bullet}$	$f_{i\bullet}$
0	11	0.44
1	8	0.32
2	6	0.24
Total	25	1

Plan de travail

Distribution conditionnelle

On peut Considérer la distribution de X (resp. Y) sur une sous population ayant pris la modalité y_j de Y (resp. x_i de X).
On les appelle distributions conditionnelles, notées : $X/Y = y_j$. (resp. $Y/X = x_i$).
$X/Y = y_j$ prend les modalitées $x_1, x_2, \cdots, x_p$.
$Y/X = x_i$ prend les modalitées $y_1, y_2, \cdots, y_q$.
L'effectif de la population ayant pris la modalité y_j de Y (resp. x_i de X) est $n_{\bullet j}$ (resp. $n_{i\bullet}$)

Fréquences conditionnelles

La fréquence conditionnelle de $X/Y = y_j$ (resp. $Y/X = x_i$) est définie par :

$$f_{i/j} = \frac{n_{ij}}{n\bullet j} = \frac{f_{ij}}{f_{\bullet j}}, \; i = \overline{1, p}$$

$$f_{j/i} = \frac{n_{ij}}{ni\bullet} = \frac{f_{ij}}{f_{i\bullet}}, \; j = \overline{1, q}$$

Exemple illustratif

Tableau de contingence :

X\Y	0	1	2
1	6	3	1
2	5	5	5

Distribution de $X/(Y = y_1 = 0)$

x_i	n_{i1}	$f_{i/j=1}$
1	6	0.55
2	5	0.45
Total	11	1

fréquences conditionnelle (exemple illustratif)

Distribution de $X/(Y=1)$

x_i	n_{i2}	$f_{i/j=2}$
1	3	0.38
2	5	0.62
Total	8	1

Distribution de $X/(Y=2)$

x_i	n_{i3}	$f_{i/j=3}$
1	1	0.17
2	5	0.62
Total	6	1

Caractéristiques numériques des distributions conditionnelles

Moyennes conditionnelles

$$\overline{x} = \frac{1}{n_{\bullet j}} \sum_{i=1}^{p} x_i n_{ij} = \sum_{i=1}^{p} x_i f_{i/j}, \; j = \overline{1,q}$$
$$\overline{y} = \frac{1}{n_{i\bullet}} \sum_{j=1}^{q} y_j n_{ij} = \sum_{j=1}^{p} y_j f_{j/i}, \; i = \overline{1,p}$$

Variances conditionnelles

$$V_j(x) = \frac{1}{n_{\bullet j}} \sum_{i=1}^{p} n_{ij}(x_i - \overline{x_j})^2, \; j = \overline{1,q}$$
$$V_i(y) = \frac{1}{n_{\bullet i}} \sum_{j=1}^{p} n_{ij}(y_j - \overline{y_i})^2, \; i = \overline{1,p}$$

Ecart type

$$\sigma_j(x) = \sqrt{V_j(x)}$$
$$\sigma_i(y) = \sqrt{V_i(y)}$$

Moyenne des moyenne

$$\overline{x} = \frac{1}{n} \sum_{j=1}^{q} n_{\bullet j} \overline{x_j}$$
$$\overline{y} = \frac{1}{n} \sum_{i=1}^{p} n_{\bullet i} \overline{y_i}$$

Plan de travail

Indépendance

On dit que deux variables X, Y sont indépendantes si les variations de l'une ne dépendent pas des variations de l'autre. Ou bien que les fréquences conditionnelles $f_{i/j}$ ne dépendent pas de j et $f_{j/i}$ ne dépendent pas de i.

Si $f_{i/j} = f_{i\bullet}, i = \overline{1, p}$ et $f_{j/i} = f_{\bullet j}, j = \overline{1, q}$
est équivalent à dire que ($\Leftrightarrow$) X et Y sont indépendantes.

$\Leftrightarrow f_{ij} = f_{i\bullet} \times f_{\bullet j}$

$\Leftrightarrow n_{ij} = \frac{n_{i\bullet} \times n_{\bullet j}}{n}$

Indépendance

On dit que deux variables X, Y sont indépendantes si les variations de l'une ne dépendent pas des variations de l'autre. Ou bien que les fréquences conditionnelles $f_{i/j}$ ne dépendent pas de j et $f_{j/i}$ ne dépendent pas de i.
Si $f_{i/j} = f_{i\bullet}, i = \overline{1, p}$ et $f_{j/i} = f_{\bullet j}, j = \overline{1, q}$
est équivalent à dire que ($\Leftrightarrow$**)** X et Y sont indépendantes.
$\Leftrightarrow f_{ij} = f_{i\bullet} \times f_{\bullet j}$
$\Leftrightarrow n_{ij} = \frac{n_{i\bullet} \times n_{\bullet j}}{n}$

Indépendance

On dit que deux variables X, Y sont indépendantes si les variations de l'une ne dépendent pas des variations de l'autre. Ou bien que les fréquences conditionnelles $f_{i/j}$ ne dépendent pas de j et $f_{j/i}$ ne dépendent pas de i.

Si $f_{i/j} = f_{i\bullet}, i = \overline{1, p}$ et $f_{j/i} = f_{\bullet j}, j = \overline{1, q}$

est équivalent à dire que (**$\Leftrightarrow$**) X et Y sont indépendantes.

$\Leftrightarrow f_{ij} = f_{i\bullet} \times f_{\bullet j}$

$\Leftrightarrow n_{ij} = \frac{n_{i\bullet} \times n_{\bullet j}}{n}$

Indépendance

On dit que deux variables X, Y sont indépendantes si les variations de l'une ne dépendent pas des variations de l'autre. Ou bien que les fréquences conditionnelles $f_{i/j}$ ne dépendent pas de j et $f_{j/i}$ ne dépendent pas de i.

Si $f_{i/j} = f_{i\bullet}, i = \overline{1, p}$ et $f_{j/i} = f_{\bullet j}, j = \overline{1, q}$

est équivalent à dire que ($\Leftrightarrow$**)** X et Y sont indépendantes.

$\Leftrightarrow f_{ij} = f_{i\bullet} \times f_{\bullet j}$

$\Leftrightarrow n_{ij} = \frac{n_{i\bullet} \times n_{\bullet j}}{n}$

Merci de votre attention