Document Chapitre 5

I Tableau de contingence

X \ Y	y_1	$\cdots$	y_j	$\cdots$	y_ℓ	totaux
x_1	n_{11}		n_{1j}		$n_{1\ell}$	$n_{1\bullet}$
$\cdots$					$\cdots$	$\cdots$
x_i	n_{i1}		n_{ij}		$n_{i\ell}$	$n_{i\bullet}$
$\cdots$					$\cdots$	$\cdots$
x_k	n_{k1}		n_{kj}		$n_{k\ell}$	$n_{k\bullet}$
totaux	$n_{\bullet 1}$		$n_{\bullet j}$		$n_{\bullet \ell}$	$N = n_{\bullet\bullet}$

II Exemple

Étude sur 5761 femmes de la survenue d'accouchement prématuré et de l'exposition à des événements stressants. On observe donc les deux variables : le type d'accouchement, noté X, c'est une *variable qualitative à 2 modalités*, et la variable score sur une échelle allant de 0 à 3, notée Y, qui est une *variable quantitative discrète à 4 valeurs*. Les effectifs observés sont rassemblés dans le tableau de contingence suivant

X \ Y	0	1	2	3	totaux
à terme	4698	413	250	197	5558
prématuré	165	16	12	10	203
totaux	4863	429	262	207	5761

▶ Distribution marginale de X

X	à terme	prématuré	effectif total
effectifs	5558	203	5761

▶ Distribution marginale de Y

Y	0	1	2	3	effectif total
effectifs	4863	429	262	207	5761

▶ Distribution conditionnelle de Y sachant que l'accouchement a été prématuré

$Y\|_{X=\text{prématuré}}$	0	1	2	3	total
effectifs	165	16	12	10	203

▶ Distribution conditionnelle de X sachant que la femme enceinte a subi un stress de niveau 2

$X\|_{Y=2}$	à terme	prématuré	total
effectifs	250	12	262

▶ Proportions des couples (x_i, y_j) :

X \ Y	0	1	2	3
à terme	0.815	0.072	0.043	0.034
prématuré	0.029	0.003	0.002	0.002

▶ Proportions marginales pour X :

X	à terme	prématuré	total
effectifs	5558	203	5761
proportions	0.964	0.036	1

▶ Proportions conditionnelles de X sachant $Y = 2$:

$X\|_{Y=2}$	à terme	prématuré	total
effectifs	250	12	262
proportions	0.954	0.046	1

▶ Proportions conditionnelles de Y sachant X =prématuré :

$Y\|_{X=\text{préma.}}$	0	1	2	3	tot.
effectifs	165	16	12	10	203
proportions	0.813	0.079	0.059	0.049	1

III Proportions

▶ lien entre les différentes notions de proportions :

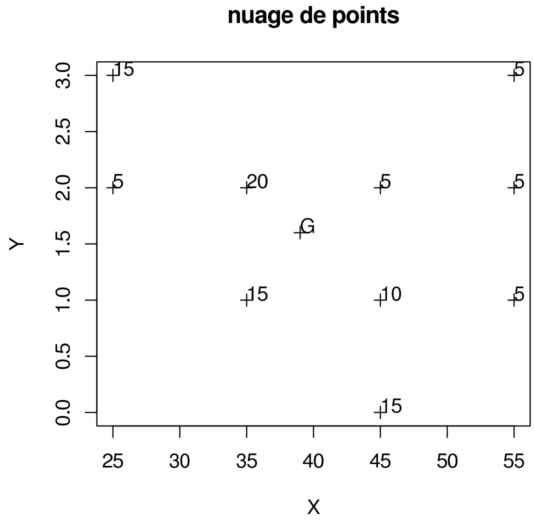
$$p_{ij} = p_{i|Y=y_j} \times p_{\bullet j} = p_{j|X=x_i} \times p_{i\bullet} \quad \text{ou encore} \quad p_{i|Y=y_j} = \frac{p_{ij}}{p_{\bullet j}} \quad \text{et} \quad p_{j|X=x_i} = \frac{p_{ij}}{p_{i\bullet}}$$

▶ propriété d'indépendance :
les variables dont indépendantes si et seulement si $p_{ij} = p_{i\bullet} \times p_{\bullet j}$.

IV Exemple 2

Une entreprise employant 100 femmes relève pour chaque femme son âge, noté X, et le nombre de journées d'absence durant le mois de janvier, noté Y. Les effectifs observés sont rassemblés dans le tableau de contingence suivant

X \ Y	0	1	2	3	totaux
$[20, 30[$	0	0	5	15	20
$[30, 40[$	0	15	20	0	35
$[40, 50[$	15	10	5	0	30
$[50, 60[$	0	5	5	5	15
totaux	15	30	35	20	100



V Principales caractéristiques

1) Caractéristiques de la distribution conditionnelle de X sachant $Y = y_j$

calcul à partir de	distribution en effectifs (x_i, n_ij)	distribution en proportions $(x_i, p_{i\|Y=y_j})$
moyenne, $\mu(X\|Y = y_j)$	$\dfrac{\sum_{i=1}^{k} (x_i \times n_{ij})}{n_{\bullet j}}$	$\sum_{i=1}^{k} \left(x_i \times p_{i\|Y=y_j}\right)$
variance $\sigma^2(X\|Y = y_j)$	$\dfrac{\sum_{i=1}^{k} \left[(x_i - \mu(X\|Y = y_j))^2 \times n_{ij}\right]}{n_{\bullet j}}$ $= \dfrac{\sum_{i=1}^{k} [{x_i}^2 \times n_{ij}]}{n_{\bullet j}} - \mu(X\|Y = y_j)^2$	$\sum_{i=1}^{k} \left[(x_i - \mu(X\|Y = y_j))^2 \times p_{i\|Y=y_j}\right]$ $= \sum_{i=1}^{k} \left[x_i^2 \times p_{i\|Y=y_j}\right] - \mu(X\|Y = y_j)^2$

2) Covariance et Corrélation

▶ Covariance

$$\text{cov}(X, Y) = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{\ell} n_{ij}(x_i - \mu(X))(y_j - \mu(Y)) = \left(\frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{\ell} n_{ij} x_i y_j \right) - \mu(X)\mu(Y)$$

▶ Corrélation

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}$$