TEST DE CHI2

Source : G. Saporta, "Probabilités, Analyse des données et Statistique", Editions Technip, 1990

Le test du Chi2

Le test

Les données sont structurées de la manière suivante :

	Modalité 1	Modalité 2	…	Modalité r	Total
Echantillon 1	n_{11}	n_{12}		n_{1r}	$n_{1.}$
Echantillon 2	n_{21}	n_{22}		n_{2r}	$n_{2.}$
Echantillon k	n_{k1}	n_{k2}		n_{kr}	$n_{k.}$
Total	$n_{.1}$	$n_{.2}$		$n_{.r}$	n

où $n_{i.}$ est l'effectif de l'échantillon i ($\sum_{j=1}^{r} n_{ij}$);

$n_{.j}$ est le nombre total des individus possédant j ($\sum_{i=1}^{k} n_{ij}$);

n est le nombre total des individus.

Le test du Chi2 permet de tester si les échantillons proviennent d'une même population. Si tel est le cas, la probabilité que les individus possèdent une caractéristique données (une modalité) devrait être la même quelque soit l'échantillon et devrait être égale à la probabilité moyenne observée dans la population. Comme on ne connaît en général pas la probabilité d'observer une modalité donnée dans la population, on estime celle-ci par $\hat{p}_j = \frac{n_{.j}}{n}$, ce qui représente r-1 estimations indépendantes (pour estimer r probabilités, on n'a besoin de r-1 relations puisque $\sum p_j = 1$. On obtient alors la statistique d^2 de la manière suivante : $d^2 = \sum_{i=1}^{k} \sum_{j=1}^{r} \frac{\left(n_{ij} - \frac{n_{i.} n_{.j}}{n} \right)^2}{\frac{n_{i.} n_{.j}}{n}}$. d^2 suit une loi du Chi^2. Il nous reste à déterminer le nombre de degrés de liberté :

- le nombre de termes sur lequel porte le test est kr;
- nous avons besoin de r-1 relations pour estimer les probabilités $\hat{p}_j$;
- les kr termes ne sont pas indépendants. Les sommes des lignes étant constantes, connaissant les r-1 effectifs pour un échantillon donné, on en déduire l'effectif pour la modalité r. On perd donc k degré de liberté (le nombre d'échantillons).

Le nombre de degrés de liberté est donc kr - (r-1) - k, soit (k-1) (r-1).