

Tests Statistiques sur un Echantillon (2)

Sous Hypothèse Nulle ($H_0 =$ il n'y a pas de différence), on détermine la limite de la probabilité, en général au risque $\alpha = 5\%$ ($z = 1.96$) :

$$|z| = \frac{|P - \pi|}{\sqrt{\frac{\pi(1 - \pi)}{n}}} \quad (\text{avec } P = \text{observé et } \pi = \text{théorique})$$

Si $|z| \sup.$ à 1.96 (α inf. à 5%), on rejette H_0 : il existe bien une différence significative.

Si $|z| \inf.$ à 1.96 (α sup. à 5%), on accepte H_0 : il n'y a pas de différence significative, ce qui ne veut pas dire qu'il n'y a pas de différence. On prend simplement le risque de considérer que cette différence est due au hasard de l'échantillonnage.

Le niveau de signification (p) est une évaluation de cette limite. Il est donné par la table.

Tableau de Contingence

	J=1	J=2	
I=1	n11	n12	n1.
I=2	n21	n22	n2.
	n..1	n..2	n..

Probabilité du tableau :

$$\frac{C_{n.1}^{n11} C_{n.2}^{n12}}{C_{n..}^{n1.}} = \frac{n.1! n.2! n1.! n2.!}{n1! n12! n21! n22! n..!}$$

= Produit des factorielles des effectifs marginaux / (Produit des factorielles des effectifs observés

X Produit des factorielles de l'effectif total)

Test du χ^2 de Pearson

(ν_{ij} sup. ou = à 5)

$$\text{ddl} = (i - 1) \cdot (j - 1)$$

$$\chi^2_{k-1, \text{ddl}} = \sum_{i=1}^k \frac{(ni - \nu_i)^2}{\nu_i} \quad (\text{avec } \nu_{ij} = \frac{ni \cdot nj}{n..} \text{ sup. ou = à 5})$$

Méthodologie :

Calculer ddl. Calculer les effectifs théoriques ν_{ij} . Calculer le χ^2 sous H_0 . Conclure.

Test Exact de Fisher

(= Test de Pearson avec ν_{ij} inf. à 5)

$$\text{ddl} = (i - 1) \cdot (j - 1)$$

$$\chi^2_{k-1, \text{ddl}} = \sum_{i=1}^k \frac{(ni - \nu_i)^2}{\nu_i} \quad (\text{avec } \nu_{ij} = \frac{ni \cdot nj}{n..} \text{ sup. ou = à 5})$$

Méthodologie :

On recherche la probabilité d'avoir une différence entre n et ν_{ij} sup. ou = à celle observée.

En conservant les mêmes effectifs marginaux, faire varier un des effectifs observés de 1 en 1 de façon à faire augmenter cette différence. En déduire les autres valeurs du tableau. Calculer la probabilité $P(n)$ de chaque tableau sous H_0 . Calculer $P = P(1) + P(2) + \dots$

Conclure (rejeter H_0 si P inf. ou = à 5%, et vice-versa).