Exercices : Martine Quinio

Tests et tests du khi deux

Exercice 1

On s'intéresse au problème des algues toxiques qui atteignent certaines plages de France; après étude on constate que 10% des plages sont atteintes par ce type d'algues et on veut tester l'influence de rejets chimiques nouveaux sur l'apparition de ces algues. Pour cela 50 plages proches de zones de rejet chimiques, sont observées; on compte alors le nombre de plages atteintes par l'algue nocive : on constate que 10 plages sont atteintes par l'algue. Pouvez-vous répondre à la question «Les rejets chimiques ont-t-il modifié, de façon significative, avec le risque $\alpha = 0.05$, le nombre de plages atteintes ?»

Correction ▼ [006032]

Exercice 2

On veut étudier la liaison entre les caractères : «être fumeur» (plus de 20 cigarettes par jour, pendant 10 ans) et «avoir un cancer de la gorge», sur une population de 1000 personnes, dont 500 sont atteintes d'un cancer de la gorge. Voici les résultats observés:
Tableau observé

Observé	cancer	non cancer	marge
fumeur	342	258	600
non fumeur	158	242	400
marge	500	500	1000

Faire un test d'indépendance pour établir la liaison entre ces caractères.

Correction ▼ [006033]

Pour plus de contenu théorique et pratique, et pour plus de détails sur les aspects historiques (par exemple études liens tabac et cancers, études statistiques sur énergie, travail et genre...), on pourra consulter l'ouvrage «Probabilités et statistique aujourd'hui», Martine Quinio Benamo, nouvelle édition 2009, L'Harmattan

Correction de l'exercice 1 ▲

Posons H_0 «les rejets chimiques ne modifient pas le nombre de plages atteintes par les algues».
Notons $p_0 = 0.1$ la proportion théorique de plages atteintes par l'algue verte avant les rejets chimiques; p la proportion théorique de plages atteintes par l'algue verte après les rejets chimiques et f la fréquence observée dans l'échantillon.
Considérons alors la variable aléatoire X_i, $i \leqslant 50$, qui a deux modalités: 1 si la plage est atteinte, 0 sinon. C'est une variable de Bernoulli, alors le nombre total de plages atteintes dans l'échantillon est une variable aléatoire qui, sous H_0, obéit à une loi binomiale de paramètres $n = 50$, $p_0 = 0.1$.
Sous H_0, «$p = p_0 = 0.1$» la variable «moyenne d'échantillon» :

$$\overline{X} = \frac{\sum_{i=1}^{i=50} X_i}{n}$$

dont une réalisation est la fréquence observée, soit $\frac{10}{50}$, obéit à une loi que l'on peut approcher par une loi normale de paramètres : moyenne p_0 et écart-type $\sqrt{\frac{p_0(1-p_0)}{50}}$.
A l'aide de la formule de cours, on détermine l'intervalle de confiance associé: $I \simeq [0.017; 0.183]$. On constate que la fréquence observée est dans la zone de rejet (non chimique) : 0.2 n'est pas dans l'intervalle de confiance au seuil 95%. On peut donc rejeter H_0 et conclure, au risque 0.05, que les rejets chimiques modifient de façon significative le nombre de plages atteintes par l'algue.

Correction de l'exercice 2 ▲

Mise en oeuvre du test:

1. On définit un risque: 5%. Pour étudier la dépendance de ces caractères faisons l'hypothèse H_0 : «les deux caractères sont indépendants » et voyons ce qui se passerait sous cette hypothèse. Notons les événements:

   - C : «avoir un cancer dans la population observée»
   - F: «être fumeur dans la population observée»

   Si les événements F et C sont indépendants, alors: $P(F \cap C) = P(F) \cdot P(C)$ et de même pour les trois autres possibilités: $P(\overline{C} \cap F)$, $P(\overline{C} \cap \overline{F})$,$P(C \cap \overline{F})$, quantités que l'on peut donc calculer sous H_0:
   $P(F) = \frac{600}{1000}$, $P(C) = \frac{500}{1000}$, $P(F) \cdot P(C) = \frac{3}{10}$, alors l'effectif théorique correspondant à la catégorie «fumeur et cancéreux» est de 300.

2. On en déduit le tableau théorique sous H_0 :

Théorique	cancer	non cancer	marge
fumeur	300	300	600
non fumeur	200	200	400
marge	500	500	1000

3. On calcule alors la valeur de $s = \sum_{i=1}^{i=4} \frac{(O_i - T_i)^2}{T_i}$: on obtient : $s = 34.73$. On a précisé le risque de %, mais pour $\alpha = 0,001$, on lit dans la table du khi-deux à un degré de liberté : $P[\chi^2 \geqslant 10.83] = 0.001$ et le χ^2 calculé est 34.73!

4. On décide de rejeter H_0. Ainsi, en rejetant l'hypothèse de l'indépendance des caractères «être fumeur» et «avoir un cancer de la gorge», on a moins de une chance sur 1000 de se tromper, puisque moins de un tableau possible sur mille conduit à un calcul de χ^2 plus grand que 10.83 ; beaucoup moins sans doute, conduiraient à un calcul de χ^2 plus grand que 34.73.
