Option Statistique :
Deux autres tests du Chi2

1 Test d'homogénéité

1.1 Un exemple

Dans une université où les initiatives pédagogiques différenciées sont vivement encouragées trois groupes de professeurs ont mis au point trois méthodes différentes d'apprentissage des Mathématiques qu'on a appliqué à trois échatillons d'étudiants ayant sensiblement le même niveau initial. A l'examen les résultats furent les suivants :

Observés	Admis	Ajournés	sommes
Méthode 1	**51**	**29**	80
Méthode 2	**38**	**12**	50
Méthode 3	**86**	**34**	120
sommes	175	75	250

On voudrait savoir si l'on peut afirmer que l'une des trois méthodes est plus efficace que les autres en termes de réussite à l'examen. Pour répondre à cette question, on teste l'hypothèse H_0 : *il n'y a pas de différence significative entre les pourcentages de réussite de ces trois échantillons, les différences observées n'étant dûes qu'aux fluctuations d'échantillonnage*, contre l'hypothèse H_1 : *une au moins de ces méthodes est significativement plus efficace que les autres.* Pour effectuer ce test on compare le **tableau de contingence observé** ci-dessus au **tableau de contingence théorique** ci-dessous correspondant à une population regroupant les trois échantillons et qui serait parfaitement homogène si H_0 est vrai, correspondant à un taux de réussite global de $175/250 = 0,7 = 70\%$:

Théoriques	Admis	Ajournés	sommes
Méthode 1	56	24	80
Méthode 2	35	15	50
Méthode 3	84	36	120
sommes	175	75	250
taux	70%	30%	100%

On calcule alors le D^2 de ces tableaux $l \times c$ de contingence, qui suit une loi du Chi2 à $(l-1)*(c-1)$ degrés de libertés (voyez-vous pourquoi ?) :

$$D^2 = \sum_{i=1}^{l}\sum_{j=1}^{c} \frac{(O_{ij}-T_{ij})^2}{T_{ij}} = \frac{(51-56)^2}{56} + \ldots + \frac{(34-36)^2}{36} = 3,26.$$

Au risque $\alpha = 5\%$, l'intervalle d'acceptation $I(\alpha) = [0\ ; x(\alpha)]$ est $I(\alpha) = [0\ ; 5,99]$. Donc

- Si $D^2 \geq 5,99$, on rejette l'hypothèse H_0
- Si $D^2 < 5,99$, on ne peut pas rejeter l'hypothèse H_0

Nous voyons donc qu'ici on ne peut pas rejeter H_0 : il est bien possible que malgré les différences de résultats à l'examen, les méthodes se valent...

1.2 Exercice 1

1. Retrouver les deux tableaux en ne saisissant que les 6 nombres imprimés en gras : tous les autres doivent être déduits par un calcul.
2. Confectionner un tableau 2×3 des $\frac{(O_{ij}-T_{ij})^2}{T_{ij}}$.
3. Calculer le D^2.
4. Retrouver la valeur de $x(\alpha) = x(5\%) = 5,99$.

1.3 Exercice 2

Deux parcelles identiques de vignes atteintes de phylloxera ont été traitées, la première avec un Traitement 1 et la seconde avec un Traitement 2. En vous inspirant de l'exercice précédent tester l'hypothèse H_0 : *les deux traitements ont le même effet.*

Observés	Eradication	Amélioration	Sans effet
Traitement 1	**280**	**210**	**110**
Traitement 2	**220**	**90**	**90**

2 Test d'indépendance

Rappelons que deux variables aléatoires X et Y prenant les valeurs $\{x_1, \ldots, x_i, \ldots, x_l\}$ et $\{y_1, \ldots, y_j, \ldots, y_c\}$ avec les probabilités $\mathbb{P}(X = x_i) = p_i$ et $\mathbb{P}(Y = y_j) = q_j$ sont dites indépendantes si et seulement si $\mathbb{P}(X = x_i, Y = y_j) = p_i q_j$.

2.1 Exemple

On veut tester si l'efficacité d'un vaccin contre la grippe est indépendante du fait qu'on ait moins de 55 ans ou strictement plus. Considérons un échantillon A de personnes vaccinées de 55 ans ou moins, et un échantillon B de personnes vaccinées de plus de 55 ans. Voici les résultats observés :

Observés	Grippés	Non-grippés	sommes	proportions
A	**38**	**82**	120	0,4
B	**72**	**108**	180	0,6
sommes	110	190	300	
proportions	0,37	0,63		

Soit p la proportion de A par rapport à l'effectif total. Soit q la proportion de personnes dans les deux échantillons réunis ayant malgré tout attrapé la grippe. Sous l'hypothèse H_0 : *les deux caractères sont indépendants* les proportions conjointes sont donc

Pr. théoriques	Grippés	Non-grippés	proportions
A	0,15	0,25	0,4
B	0,22	0,38	0,6
proportions	0,37	0,63	

Comme la population totale est de 300, on en déduit que les effectifs théoriques sous l'hypothèse H_0 sont

Effec. théoriques	Grippés	Non-grippés
A	44	76
B	66	114

On calcule alors le D^2 de ces tableaux $l \times c$ d'effectifs, qui suit une loi du Chi2 à $(l-1)*(c-1) = 1$ degrés de libertés : $D^2 = \frac{(38-44)^2}{44} + \ldots + \frac{(108-114)^2}{114} = 2,15$. Au risque $\alpha = 10\%$, l'intervalle d'acceptation $I(\alpha) = [0\ ; x(\alpha)]$ est $I(\alpha) = [0\ ; 2,71]$. Nous voyons donc qu'ici on ne peut pas rejeter H_0 : il est bien possible que malgré les différences d'âge, l'efficacité du vaccin soit la même.

2.2 Exercice 3

1. Retrouver les trois tableaux en ne saisissant que les 4 nombres imprimés en gras : tous les autres doivent être déduits par un calcul.

2. Confectionner un tableau 2×2 des $\frac{(O_{ij}-T_{ij})^2}{T_{ij}}$ et calculer le D^2.

3. Retrouver la valeur de $x(\alpha) = x(10\%) = 2,71$.