Université de Nice
Département de Mathématiques

Année 2003-2004
MP2-MI2

Option Statistique : **Deux autres tests du Khi Deux**

1 Test d'homogénéité

Dans une population formée d'individus répartis en différentes catégories (hommes/femmes, classes d'ages, niveaux socio-économiques, etc...), on observe une variable (effet d'un médicament, présence d'un comportement à risque, performances ...) et on se demande si ses variations selon les différentes catégories de la population sont simplement dues au fluctuations d'échantillonnage ou si au contraire elles révèlent des comportements différents de la variable dans chacune de ces catégories. Nous allons conclure grace à un test d'homogénéité dont nous expliquons tout d'abord le principe sur un exemple.

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 échantillons d'étudiants ayant sensiblement le même niveau initial. A l'examen les résultats furent les suivants :

Observés	Admis	Ajournés	sommes
Méthode 1	**51**	**29**	
Méthode 2	**38**	**12**	
Méthode 3	**86**	**34**	
sommes			

Peut-on affirmer 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 groupes d'étudiants,* 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 des effectifs observés** ci-dessus au **tableau des effectifs théoriques** ci-dessous correspondant à une population, regroupant les trois groupes, qui serait parfaitement homogène si H_0 était vrai. Le nombre théorique d'admis parmi les étudiants ayant suivi la méthode 1 sera égal à $56 = \frac{175}{250}80$ car il y a un taux de réussite global de $175/250 = 70\%$ et le nombre total d'étudiants ayant suivi la méthode 1 est de 80.

Théoriques	Admis	Ajournés	sommes
Méthode 1	56		80
Méthode 2			50
Méthode 3		36	120
sommes	175	75	250

On calcule alors le *Khi Deux* de ces tableaux $l \times c$ d'effectifs, noté X^2 de la façon suivante :

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

Sous des hypothèses généralement satisfaites dans la pratique (échantillons disjoints, sélectionnés de façon indépendante et de taille suiffisante), on peut considérer que la quantité X^2 qui est aléatoire (car elle dépend du choix des échantillons) suit une loi du $\chi^2(dl)$ ayant un nombre de degré de liberté dl égal à

$$dl = (c-1)(l-1).$$

Pour tout choix d'un seuil α, on peut construire alors un intervalle d'acceptation $I(\alpha) = [0\ ; x(\alpha)]$, où $x(\alpha)$ est défini par $P(X^2 > x(\alpha)) = \alpha$. On conclut, au risque α (c'est-à-dire avec une probabilité de rejet à tord de α), que :

- Si $X^2 \geq x(\alpha)$, l'hypothèse H_0 doit être rejetée, c'est-à-dire que les trois méthodes d'apprentissage ne peuvent être considérées comme équivalentes.
- Si $X^2 < x(\alpha)$, on ne peut pas rejeter l'hypothèse H_0, c'est-à-dire que, malgré les différences de résultats à l'examen, les méthodes se valent.

1.2 Exercice 1

1. Sur une feuille de calcul, saisir et compléter le tableau des effectifs observés en complétant les sommes. Calculer les taux de réussite et d'échec pour l'ensemble de la population.
2. Saisir et compléter le tableau des effectifs théoriques.
3. Confectionner le tableau 2×3 des $\frac{(O_{ij}-T_{ij})^2}{T_{ij}}$.
4. Calculer le X^2. Quel est le nombre de degré de liberté *dl* ?
5. Calculer $x(\alpha)$ pour $\alpha = 0,05$ et pour ce *dl*.
6. Conclure : indiquer le résultat de votre test.
7. Reprendre ce test avec un seuil α de 10% cette fois. Votre conclusion est-elle été modifiée ?
8. Reprendre ce test avec un seuil α de 1% cette fois. Votre conclusion est-elle été modifiée ?

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 $\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j) = p_i q_j$. Pour tester l'indépendance de deux caractères, on peut utiliser un test du Khi Deux de la façon suivante : on réunit dans un tableau les effectifs $O_{i,j}$ d'individus donnant à X et Y les valeurs (x_i) et (y_j) et on calcule les effectifs théoriques à partir de la proportion théorique produit $p_i q_j$. Expliquons le principe de ce test sur un exemple.

2.1 Exemple

On veut savoir si l'efficacité d'un vaccin contre la grippe est indépendante du fait qu'on l'administre à des patients de moins de 55 ans ou à des patients strictement plus agés. Considérons un échantillon A de personnes vaccinées de 55 ans ou moins, et un échantillon B de personnes vaccinées de plus de 55 ans.

Observés	Grippés	Non-grippés	sommes
A	**38**	**82**	120
B	**72**	**108**	180
sommes	110	190	300

Sous l'hypothèse H_0 : *les deux caractères sont indépendants*, les proportions conjointes sont :

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	1

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	.....
B	.....	114

On calcule alors le X^2 de ces échantillons, qui suit une loi du χ^2 à $(l-1)*(c-1) = 1$ degrés de libertés : $X^2 = \frac{(38-44)^2}{44} + \ldots + \frac{(108-114)^2}{114} = 2,15$. 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 le minimum de donées.
2. Confectionner le tableau 2×2 des $\frac{(O_{ij}-T_{ij})^2}{T_{ij}}$ et en déduire le X^2.
3. Retrouver la valeur de $x(\alpha) = x(10\%)$ indiquée. Conclure
4. Refaire le test au seuil de 5%. Que concluez-vous ?

2.3 Exercice 4

Dans une population de 500 personnes (300 hommes et 200 femmes), on a mesuré la tension artérielle de chaque individu, ce qui a donné les effectifs suivants :

Effec. observés	Hypertension	Tension normale	Hypotension
Hommes	72	192	36
Femmes	38	118	44

A-t-on, au risque de 5%, une liaison entre le sexe de l'individu et sa tension artérielle ?