EXERCICES CORRIGES : tests de χ^2

Exercice 1 : BTSA - Productions animales - Session 1993

(Troisième exercice sur 8 points)

Un étalon gris hétérozygote accouplé à des juments non grises produira des poulains gris avec une probabilité de 0,25 d'après les lois de Mendel. Des juments non grises accouplées à cet étalon et ayant produit 5 poulains ont donné les résultats suivants :

Nombre de poulains gris sur les 5 produits	0	1	2	3 et plus
Nombre de juments	10	18	16	6

1. Soit X la variable aléatoire prenant pour valeur le nombre de poulains gris par jument. Définir la loi de probabilité de X, sous l'hypothèse de Mendel.

2. A l'aide d'un test de χ^2 (au seuil de 5%), dire si les résultats observés permettent d'accepter l'hypothèse de Mendel.

Proposition de corrigé :

1. Nous sommes dans le cas d'un schéma de Bernoulli (On suppose qu'il n'y a qu'un poulain par portée):

- Pour chaque poulain il y a deux éventualités contraires :
  soit il est gris (avec une probabilité p = 0,25)
  soit il est non gris (avec une probabilité de 1 - p = 0,75)

- Les couleurs des 5 poulains d'une jument sont indépendantes et ont été obtenues dans les mêmes conditions.

Ces deux conditions nous permettent de dire que, sous l'hypothèse de Mendel, X est de loi binomiale B(5; 0,25).

2. Il s'agit ici d'un test d'ajustement :
Posons l'hypothèse nulle H_0 : Les résultats sont conformes à la théorie.
C'est-à-dire : H_0 : La variable aléatoire X suit la loi binomiale B(5; 0,25).

A l'aide de la formule $P(X=k)=C_5^k \times 0{,}25^k \times 0{,}75^{5-k}$ pour $0 \le k \le 5$, nous allons calculer les probabilités puis les effectifs théoriques correspondant aux diverses valeurs de X :

Nombre de poulains gris sur les 5 produits	0	1	2	3 ou plus	Totaux
effectifs observés (n_i)	10	18	16	6	50
Probabilités (p_i)	0,2373	0,3955	0,2637	0,1035 *	1
effectifs théoriques (np_i)	11,87	19,78	13,18	5,17 *	50
n_i - np_i	-1,87	-1,78	2,82	0,83	0

**valeurs calculées par différence*

La taille de l'échantillon est n = 50, c'est-à-dire l'effectif total.

Variable de décision :

Tous les effectifs théoriques étant supérieurs à 5, on peut dire que, sous l'hypothèse H_0, la variable $K_0 = \sum \frac{(N_i - np_i)^2}{np_i}$ suit approximativement la loi de χ² à $4-1=3$ degrés de liberté (ddl).

Remarque : le N_i *en majuscule signifie qu'il s'agit ici d'une variable aléatoire mais une minuscule est tolérée.*

Calcul de la valeur de K_0 pour l'échantillon prélevé :

$$k_0 = \sum \frac{(n_i - np_i)^2}{np_i} = \sum \frac{(n_{obs} - n_{th})^2}{n_{th}}$$

$$k_0 \approx \frac{1{,}87^2}{11{,}87} + \frac{1{,}78^2}{19{,}78} + \frac{2{,}82^2}{13{,}18} + \frac{0{,}83^2}{5{,}17} \text{ soit } k_0 \approx 1{,}19.$$

Décision : α = 5%

Pour 3 ddl on lit dans la table : $\chi^2_{0.95} = 7{,}81$ (ce type de test est unilatéral).

k_0 est inférieur à cette valeur donc on ne rejette pas H_0.

Nous pouvons donc en conclure, au seuil de 5 %, que les résultats observés ne contredisent pas la théorie.

Exercice 2 : BTSA - Industries agro-alimentaires - Remplacement 1994

(Quatrième exercice sur 5 points)

Afin de comparer l'action de deux levures sur une pâte à gâteaux, on prélève, pour chacune des levures, un échantillon aléatoire de gâteaux. L'aptitude des pâtes à lever est définie par les critères suivants : moyenne, bonne, très bonne.

Les résultats constatés sont rassemblés dans le tableau suivant :

levure \ aptitude à lever	moyenne	bonne	très bonne
A	41	16	63
B	22	27	51

A l'aide d'un test de χ^2, au risque de 5%, peut-on conclure à une différence d'activité des deux levures ?

Proposition de corrigé : *(Il y a de nombreuses notations possibles, celles choisies ci-dessous ne sont qu'un exemple.)*

Il s'agit ici d'un test de comparaison de deux distributions qui se ramène à un test d'indépendance.

H_0 : Il n'y a pas de différence d'activité entre ces deux levures.

qui se traduit par :

H_0 : L'aptitude à lever de cette pâte à gâteau est **in**dépendante du choix qui est fait entre les deux levures A et B.

Nous allons donc établir le tableau de contingences :

L'effectif théorique de la classe située à l'intersection de la i ième ligne et de la j ième colonne est donné par $n\hat{p}_{ij} = \dfrac{n_{i.} \times n_{.j}}{n}$ (où $n_{i.}$ est l'effectif total de la i ième ligne, $n_{.j}$ celui de la j ième colonne et n l'effectif total soit 220).

Les effectifs théoriques notés sont en italiques :

levure \ aptitude	moyenne	bonne	très bonne	total
A	41 *34.36*	16 *23.45*	63 *62.18*	120
B	22 *28.64*	27 *19.55*	51 *51.82*	100
total	63	43	114	220

Variable de décision :

Tous les effectifs théoriques sont supérieurs ou égaux à 5 donc, sous l'hypothèse H_0, la variable aléatoire

$$K_0 = \sum_i \sum_j \frac{(N_{ij} - n\hat{p}_{ij})^2}{n\hat{p}_{ij}}$$ suit (approximativement) la loi de χ^2 à (3-1)(2-1)=2 degrés de liberté.

Calcul de la valeur de K_0 pour l'échantillon prélevé :

$$k_0 = \frac{(41-34.36)^2}{34.36} + \frac{(16-23.45)^2}{23.45} + \frac{(63-62.18)^2}{62.18} + \frac{(22-28.64)^2}{28.64} + \frac{(27-19.55)^2}{19.55} + \frac{(51-51.82)^2}{51.82}$$

$$k_0 \approx 8.05$$

ou

$$k_0 = \frac{41^2}{34.36} + \frac{16^2}{23.45} + \frac{63^2}{62.18} + \frac{22^2}{28.64} + \frac{27^2}{19.55} + \frac{51^2}{51.32} - 220 \approx 8.05$$

Décision :

Pour α=5% et pour 2 ddl on lit dans la table de χ^2 : $\chi^2_{0.95} = 5.99$.

k_0 est supérieur à cette valeur donc on rejette H_0 au risque de 5%.

En conclusion on peut dire, au risque de 5% et au vu de ces observations, qu'il y a une différence d'activité entre les deux levures A et B sur cette pâte à gâteaux.

Remarque : la notation $n\hat{p}_{ij}$ vient du fait que, sous l'hypothèse d'indépendance des deux caractères, $\hat{p}_{ij} = \frac{n_{i.}}{n} \times \frac{n_{.j}}{n}$ est une estimation du pourcentage d'individus de la population appartenant à la $i^{ième}$ modalité du caractère "ligne" et à la $j^{ième}$ modalité du caractère "colonne".