

TP8 : Test du chi-deux

Objectif : savoir utiliser un test du chi-deux d'indépendance ou d'adéquation

Exercice 1

23 patients sont exposés à un certain facteur de risque et sont tombés malades, 56 étaient exposés mais ne sont pas tombés malades, 67 n'étaient pas exposés et sont tombés malades, 136 n'étaient pas exposés et ne sont pas tombés malades.

1. Créer la table de contingence :

```
ta <- rbind(c(23,56),c(67,136))  
rownames(ta) <- c("expose","non expose")  
colnames(ta) <- c("malades","sains")  
ta
```
2. Calculer les répartitions conditionnelles de la variable "Etat du patient" selon la variable "Exposition" (utiliser `prop.table`). Les représenter sur un même graphique et côte à côte (utiliser `barplot`). Interpréter.
3. On veut tester si les deux variables Exposition et Etat du malade sont indépendantes. Donner les hypothèses H_0 et H_1 . Appliquer le test du chi-deux avec la fonction `chisq.test` (on pourra consulter l'aide de cette fonction). Que peut-on en conclure ?

Exercice 2

1. Charger la base de données `titanic.csv`. Appeler les variables `pclass`, `survived`, `gender`, `age` respectivement `P`, `S`, `G`, `A`.
2. Calculer la table de contingence du genre et de la survie. Combien de passagers étaient des femmes ? Combien de femmes ont survécu ? Combien d'hommes ont survécu ? Calculer les distributions conditionnelles de la survie sachant le sexe. Interpréter.
3. En utilisant un test du chi-deux, les proportions de survivants sont-elles les mêmes chez les hommes et les femmes ? En d'autres termes la Survie dépend-elle du Sexe ?
4. Créer un vecteur valant 1 si les passagers avaient moins de 10 ans, 0 sinon. Appeler `A10` cette variable. Calculer la table de contingence de `A10` et de la survie. Combien de passagers avaient moins de 10 ? plus de 10 ans ? Combien parmi les moins de 10 ans ont survécu ? Appliquer un test du chi-deux pour décider si la survie était la même dans les deux populations (les moins de 10 ans et les plus de 10 ans). Conclure.
5. Dans la question précédente, remplacer 10 ans par 21 ans. La conclusion change-t-elle ?

Exercice 3

Exo 8.1 fait en TD On a effectué le croisement de balsamines blanches avec des balsamines pourpres. En première génération, les fleurs sont toutes pourpres. On obtient en deuxième génération quatre catégories avec les effectifs suivants :

Couleur	pourpre	rose	blanc lavande	blanc
Effectifs observés	1790	547	548	213

On veut savoir si la répartition observée est conforme à la loi de Mendel. Selon Mendel, en seconde génération on devrait obtenir la répartition théorique $(9/16; 3/16; 3/16; 1/16)$ pour les quatre modalités (pourpre, rose, blanc lavande et blanc) de la variable couleur.

Poser formellement les hypothèses du test \mathcal{H}_0 et \mathcal{H}_1 et le mettre en oeuvre en utilisant `chisq.test`.

Exercice 4

Dans le fichier `mtcars` de R sont données diverses caractéristiques d'un véhicule et de son moteur pour un échantillon de taille $n = 32$. On s'intéresse à la variable `qsec`. On voudrait savoir si cette variable a une répartition normale.

1. Représenter la répartition de cette variable avec un histogramme avec les classes $[14, 16[, [16, 17[, [17, 18[, [18, 19[, [19, 23[$.
2. Donner les effectifs de chaque classe à l'aide de la variable `counts` produite par la fonction `hist`.
3. Calculer l'estimation sans biais de la moyenne et de l'écart-type de la variable, puis les effectifs attendus dans chacune des classes si la répartition est normale (avec les deux paramètres précédemment estimés)
4. Appliquer la fonction `chisq.test`. Pour remédier au problème retourné par le logiciel et remplacer la probabilité théorique d'être dans $[14, 16[$ par celle d'être dans $] -\infty, 16]$ et la dernière probabilité par celle d'être dans $[19, \infty[$. Quelle valeur de statistique de test est retournée par la procédure `chisq.test`? Combien de degré de liberté? Quelle p-valeur?
5. Calculer la p-valeur en corrigeant le nombre de degrés de liberté.
6. Le test du chi-deux n'est pas idéal pour tester la normalité d'une variable. On lui préfère toujours le test de Shapiro-Wilks mis en oeuvre dans la procédure `shapiro.test`. Ce test conduit-il à la même conclusion?