

TD 4 - Analyse bivariée - Données qualitatives

But du TD : étude de la dépendance entre deux variables qualitatives. Tableaux de contingence, indices d'attraction-répulsion, test du χ^2 .

1 Etude de l'influence de la dose d'un traitement sur la guérison d'une maladie

Un traitement est administré à trois doses D_1 , D_2 et D_3 à un groupe de patients atteints d'une même maladie. On compte le nombre de guérisons pour chaque dose. Les résultats sont les suivants :

	Sujets guéris	Sujets non guéris
Dose D_1	30	30
Dose D_2	42	35
Dose D_3	58	31

La question posée est : l'efficacité du traitement est-elle liée à la dose utilisée ?
Référence utilisée :

http://nte-serveur.univ-lyon1.fr/nte/immediato/Math/Enseignement/07%20Statistiques/Statistiques_index.htm

1.1 Test d'indépendance

- Décrire les variables en jeu (nom, type).
- Formuler l'hypothèse nulle testée.
- Créer un tableau `tab` contenant les données à l'aide de la fonction `array`. Donner des noms aux colonnes et aux lignes à l'aide des fonctions `colnames` et `rownames`. Exécuter la commande `mosaicplot(tab)`. Expliquer le graphique obtenu.
- Décrire le test du χ^2 , la statistique de test, la distribution sous hypothèse nulle et les conditions d'application.
- Réaliser un test du χ^2 à l'aide de la fonction `chisq.test` (`c<-chisq.test(tab)`). Décrire les différents champs en sortie de la fonction. Taper `summary(c)` puis `c$expected`. A quoi cela correspond-il ?
- Les variables sont-elles indépendantes ? Donner votre décision en précisant le niveau d'erreur.
- Exécuter la commande `assocplot(tab)`. Que représente le graphique obtenu ? (utiliser l'aide). Quel est l'intérêt d'un tel graphique ?

1.2 Analyse détaillée

- Calculer et présenter la loi empirique du couple,
- présenter les lois empiriques des marginales,
- retrouver par le calcul le résultat de `c$expected`,
- calculer les lois empiriques conditionnelles,
- calculer les indices d'attraction-répulsion. Vérifier qu'ils sont cohérents avec le graphique obtenu avec `assocplot`.

2 Etude de données sur des nouveaux-nés

Charger la librairie MASS (`library(MASS)`). Récupérer les données `birthwt` et les nommer `B`. On s'intéresse à deux facteurs pouvant causer un poids trop faible des bébés à la naissance : l'âge de la mère et si elle est fumeuse ou non.

A l'aide de la fonction `table`, créer une table de contingence avec les données correspondant aux deux colonnes `B$low` et `B$smoke`. Tester l'indépendance des variables correspondantes en suivant la procédure décrite au paragraphe 1.

On veut maintenant étudier l'influence de l'âge de la mère. Exécuter la commande `boxplot(B$age~B$low)`. Que représente le graphique obtenu ? En déduit-on quelque chose ? Etudier la répartition de la variable `B$age` (à l'aide de `hist`, `summary`). Proposer une fonction que l'on nommera `classe`, qui permettra de répartir des données quantitatives en trois classes. Répartir ici les données correspondant à l'âge en trois classes : jeune, normal, moins jeune. Réaliser une table de contingence avec cette nouvelle variable catégorielle et la colonne `B$low`. Tester l'indépendance des deux variables comme dans le paragraphe 1.

3 Taux d'absentéisme à l'école chez les Australiens

Toujours dans la librairie `MASS`, récupérer les données `quine` et les nommer `x`. Que représentent les colonnes `x$Eth` et `x$Days` ? Exécuter `boxplot(x$Days~x$Eth)`. Commenter. Répartir les données `x$Days` en trois classes : faible, moyen et fort à l'aide de la fonction `classe`. Réaliser une table de contingence avec cette nouvelle variable et la variable catégorielle `x$Eth`. Etudier l'indépendance entre les deux variables comme dans le paragraphe 1.

4 Etude des données iris

Récupérer les données `iris` et les nommer `y`. On veut tester l'indépendance entre l'espèce (`y$Species`) et la longueur du sépale (`y$Sepal.Length`). Répartir comme dans les questions précédentes les données de longueur des sépales en trois catégories : petit, moyen, grand. Après avoir réalisé une table de contingence, tester l'indépendance entre les deux variables catégorielles comme dans le paragraphe 1.