

TP Analyse de données

Séance 3 – Tableaux de contingence

1. Exemple du cours

Considérons l'exemple fictif discuté au cours. Pour une population d'effectif $n = 1000$ on a mesuré les deux variables qualitatives "Couleur des yeux" et "Etat matrimonial". Les résultats sont résumés sous la forme d'un tableau de contingence:

	Bleu	Brun
Célibataire	290	410
Marié	110	190

- Créer une variable `tableau` à l'aide des commandes


```
tableau <- matrix(c(290,410,110,190), ncol=2, byrow=TRUE)
colnames(tableau) <- c("Bleu", "Brun")
rownames(tableau) <- c("Celib", "Marie")
tableau <- as.table(tableau)
```
- Afficher le contenu de la variable `tableau`. Afficher une représentation graphique à l'aide de la commande `barplot(tableau)`.
- Exécuter les commandes suivantes et comprendre leur signification:


```
n <- margin.table(tableau)
m1 <- margin.table(tableau,1)
m2 <- margin.table(tableau,2)
prop.table(tableau)
```

2. Le test du chi carré

- Créer un tableau `tab0` à l'aide des commandes


```
tab0 <- as.array(m1) %*% t(as.array(m2))/n
tab0 <- as.table(tab0)
```

 Quelle est sa signification?
- Exécuter les commandes


```
summary(tableau) et summary(tab0)
```

 Explication: `Chisq` donne la valeur de la distance χ^2 à l'indépendance, `df` est le nombre de degrés de liberté, et `p-value` est la probabilité que le χ^2 d'un échantillon indépendant dépasse la valeur calculée. En général, on considère les échantillons comme non indépendants si `p-value` est inférieur à 0.05.
- Créer un tableau dans lequel tous les individus aux yeux bleus sont mariés et tous les autres sont célibataires, et effectuer le test du chi carré.

3. Applications

- Appliquer le test du chi carré au tableau de contingence suivant,

	Blond	Châtain	Roux	Brun	Noir
Célibataire	71	64	12	58	90
Marié	152	122	28	120	178
Divorcé	27	14	10	22	32

- Appliquer le test du chi carré à quelques échantillons statistiques de R, par exemple `HairEyeColor`, `Titanic` et `UCBAdmissions`.