LSTAT2100 - Exercices - Série 1

Énoncés

Exercice 1

On s'intéresse à la couleur des yeux d'individus dans un pays. Pour $n = 300$ individus pris au hasard, on a obtenu :

bleu	marron	noir	vert
48	122	95	35

(a) Peut-on dire que toutes les couleurs sont équiprobables ?

(b) Peut-on dire que les yeux foncés (marron et noir) sont deux fois plus probables que les yeux clairs (bleu et vert) ?

(c) Peut-on dire que les yeux bleus et verts sont équiprobables ?

Exercice 2

Voici les fréquences (Freq) du nombre de passages de bus (nBus) par heure à un arrêt de bus. Ces données concernent une periode de 30 heures reparties sur 5 jours de semaine.

nBus	0	1	2	3	4	5
Freq	1	5	6	10	4	4

(a) Supposons que $nBus \sim Pois(\mu)$, $\mu > 0$. Tester l'hypothèse $H_0 : \mu = 3$ vs H_1: $\mu \neq 3$. Calculer un intervalle de confiance pour μ.

(b) En supposant que la fréquence moyenne de passage est de $\mu = 3$ par heure, testez l'ajustement d'une loi de Poisson aux données ?

(c) Refaire le même test, sans supposer que $\mu = 3$.

Exercice 3

Dans cet exercice, nous allons utiliser le jeu de données data.csv. Ce jeu de données comprend des observations liées à des clients qui ont contracté un crédit. Nous avons à notre disposition un certain nombre de variables, dont la variable Y qui nous dit si le client a pu rembourser dans les temps le crédit ($Y = 1$) ou pas ($Y = 2$).

Note: Bien qu'il y ait une multitude de variables disponibles, nous allons nous cantonner qu'à un sous ensemble de variables pour cet exercice.

Charger les données avec la commande suivante (cette commande suppose que votre répertoire de travail contient un répertoire `data` qui contient le fichier `data.csv`).

```r
mdata <- read.csv(file = "Data/data.csv")
mdata$Y <- factor(mdata$Y)
levels(mdata$Y) <- c("PasDefaut", "Defaut") # Pour plus de lisibilité.
```

(a) À l'aide de la fonction `xtabs`, calculer le nombre d'observations dans chaque niveau de la variable Y. Visualisez cette information dans un graphique adéquat.

(b) Construisez la table de contingence croisant les variables $X = check\_ac$ et Y et calculez les fréquences $\hat{P}(Y = j|X = i)$.

Note: La variable *check_ac* prend 4 valeurs et correspond aux flux moyens rentrants chaque mois sur le compte courant des clients:

- A11: <0
- A12: $[0, 200)$
- A13: ≥ 200
- A14: Pas de compte courant.

(c) Utiliser deux graphiques différents pour visualiser le lien entre ces deux variables.

(d) Tester l'indépendance entre ces deux variables.

(e) Etudier le lien entre les variables *tel* et Y, à l'aide de l'odds ratio. Interpréter.

Note: La variable **tel** nous renseigne si la personne a un numéro de téléphone (le jeu de données n'est pas tout récent…)

- A191 : Non
- A192 : Oui

(f) Sur base de l'oddsratio, que pouvez-vous dire sur l'indépendance entre *tel* et Y.

Exercice 4

Des statistiques sur la réussite ou non des "Lancer francs" en Basketball ont été récoltées. Il y a eu 5 lancers où aucun des deux tirs n'a été un succès, 34 où uniquement le premier tir a été un succès, 48 où uniquement le deuxième tir a été un succès et 251 tirs avec deux succès.

Notons X la variable aléatoire binaire du **premier tir** ($1 =$ Échec$, 2 =$ Succès), et Y la variable aléatoire binaire du **second tir** ($1 =$ Échec$, 2 =$ Succès).

(a) Construisez un tableau de contingence à partir de ces statistiques.

(b) Tester l'indépendance entre les deux tirs (i.e. entre les variables X et Y.).

(c) Soit $p_{ij} = P(X = i, Y = j)$, $i, j = 1, 2$. Considérons l'hypothèse suivante

$$H_0: \; p_{11} = \theta^2, p_{12} = p_{21} = \theta(1-\theta), \text{ et } p_{22} = (1-\theta)^2$$

Montrer que sous H_0, X et Y sont indépendants et identiquement distribués.

(d) En supposant un échantillonnage multinomial simple, donner l'estimateur du maximum de vraisemblance de θ et calculer le.

(e) Proposer une statistique de test pour tester H_0. Effectuez le test et concluez.