LSTAT2100 - Exercices - Série 2

Énoncés

Exercice 1

Dans cet exercice, nous allons reprendre le premier exercice de la Série 1, et le réaliser ***en utilisant uniquement des modèles log-linéaires***.

Nous nous intéressons à la couleur des yeux d'une certaine population. Pour $n = 300$ individus, pris au hasard, nous avons observé les chiffres suivants.

bleu	marron	noir	vert
48	122	95	35

(a) Ces couleurs sont-elles toutes réparties de manière uniforme (équiprobables) ? Répondez à l'aide d'un test de Pearson et un test LR (les deux à réaliser via une régression log-linéaire).

(b) Peut-on dire que les yeux foncés (marron et noir) sont deux fois plus probables que les yeux clairs (bleu et vert) ? Réaliser le test via une régression log-linéaire.

(c) Les proportions des yeux bleus et des yeux verts sont-elles les mêmes ? Réaliser le test via une régression log-linéaire.

Exercice 2

Dans cet exercice, nous utiliserons le jeu de données mdata.csv, que nous avions utilisé dans la Série 1. Vous devez charger/transformer les données comme expliqué dans le texte. Aussi, nous vous renvoyons à l' Exercice 3 de la Série 1 pour la description des variables.

(a) Ajustez un modèle log-linéaire *saturé* pour modéliser la table de contingence *tel* $\times$ *repay*. Construisez des intervalles de confiance à 98% pour tous les paramètres du modèle, à l'exception de l'intercept.

(b) Le tableau suivant donne le `summary()` du modèle saturé dont il est question en (a).

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	4.956	0.084	59.056	0.000
telYes	-0.434	0.134	-3.243	0.001
repayNotInDefault	0.764	0.102	7.525	0.000
telYes:repayNotInDefault	0.066	0.161	0.407	0.684

Utilisez cette sortie (*et uniquement celle-ci*) pour reconstruire le tableau de contingence *tel* $\times$ *repay*.

tel/repay	Default	NotInDefault
No	142	305

tel/repay	Default	NotInDefault
Yes	92	211

Que pouvez-vous dire des résidus du modèle étudié ? de sa déviance ? et du $pseudo - R^2$? Peut-on les obtenir sans faire de calculs ?

(c) Visualisez graphiquement des proportions conditionnelles $P(repay|tel)$. Que suggère votre graphique quand à l'association entre `repay` et `tel` ? Expliquez.

(d) Utilisez le modèle log-linéaire ajusté précédemment pour tester l'indépendance entre `repay` et `tel`. Proposer deux approches différentes (basées sur les GLM), dont une est le résultat direct des analyses précédentes (sans autres calculs).

(e) Créez la nouvelle variable `Acc` qui regroupe les modalités "No acc" et "<0" de la variable `account` en une seule modalité nommée "Noacc or <0" et les modalités "[0-200)" et ">=200" en une seule modalité nommée ">=0". Réalisez la table de contingence `employm`×`Acc`.

Construisez un modèle log-linéaire saturé sur ces données en choisissant "Noacc or <0" comme référence. Au regard des $p-$valeurs des termes d'interactions, que pouvez-vous conclure quand à l'association entre `employm` et `Acc` (ici, il n'est pas demandé d'effectuer de calculs supplémentaires) ?

Que signifie la $p-$valeur qui figure dans la ligne `Acc>=0` du `summary()` de votre modèle.

Utilisez le modèle ajusté pour tester l'indépendance entre `employm` et `Acc`.

(f) Indépendamment du résultat du test précédent, ajustez un modèle log-linéaire considérant les deux variables `emploi` et `Acc` comme indépendantes. Que signifie la $p-$valeur dans la ligne `Acc>=0` du `summary()` de ce nouveau modèle ? Calculez le $pseudo - R^2$ et comparez-le à celui du modèle saturé.

(g) Faites un tableau de contingence et un graphique (approprié) mettant en jeu les variables `repay`, `tel` et `employm`. Une association homogène entre ces trois variables est-elle envisageable ? À ce stade, il n'est pas demandé de réaliser un quelconque test.

Confirmez ou infirmez votre réponse antérieure à l'aide d'un test approprié.

(h) Simplifiez le modèle log-linéaire saturé `repay`×`employm`×`tel` autant que possible en utilisant des tests sur des modèles emboîtés. Répétez la même analyse en utilisant l'AIC et puis le BIC.

Exercice 3

Le but de cet exercice est de montrer que l'indépendance deux à deux n'implique pas l'indépendance mutuelle.

Pour cela, considérant l'expérience qui consistent à lancer deux pièces (indépendamment l'une de l'autre). On définit les variables suivantes:

- X la variable aléatoire binaire prenant la valeur 1 si la première pièce montre face et 0 sinon,
- Y la variable aléatoire binaire prenant la valeur 1 si la deuxième pièce montre face et 0 sinon.
- Z la variable aléatoire prenant la valeur 1 si les deux pièces montrent le même coté et 0 sinon; càd $Z = I(X = Y)$.

(a) Montrer que X, Y et Z sont indépendantes deux à deux.

(b) Montrer que X, Y et Z ne sont pas mutuellement indépendants.

Exercice 4

(a) Montrer que

$$\text{(a) } X \perp\!\!\!\perp Y|Z \text{ et } \text{(b) } X \perp\!\!\!\perp Z|Y \iff \text{(c) } X \perp\!\!\!\perp (Y,Z)$$

(b) Montrer que

$$\text{(a) } X \perp\!\!\!\perp (Y,Z) \text{ et } \text{(b) } Y \perp\!\!\!\perp (X,Z) \iff \text{(c) } X \perp\!\!\!\perp Y \perp\!\!\!\perp Z$$