

Statistique et Aide à la décision 2023

Contrôle des connaissances (durée : 1h)

1ère session

Nom Prénom :

Tous les documents sont autorisés. La calculatrice est autorisée. L'usage du téléphone portable est interdit, même pour sa fonction calculatrice.

Exercice - Lien entre personnalité et latéralité

Une équipe de chercheur.e.s en psychologie s'appuie sur un questionnaire normalisé construit pour décrire la personnalité dans le but de répondre à la question suivante : "existe-t'il un lien entre latéralité et personnalité ?". Le questionnaire de personnalité est ce que l'on appelle en psychologie une échelle : pour une série de phrases (on parle d'items), la.le répondant.e doit donner son degré d'accord sur une échelle quantitative entière allant de 0 (désaccord) à 5 (accord). Par exemple, l'item 40 est : "I really like dancing". L'échelle contient 44 items.

Question 1

Quel est votre degré d'accord avec l'item 40 ?

Réponse

Pour chacun.e des répondant.e.s, on dispose aussi du niveau d'études (1=Less than high school, 2=High school, 3=University degree, 4=Graduate degree) et du genre (1=Male, 2=Female, 3=Other).

La latéralité est mesurée par la réponse à la question suivante : "What hand do you use to write with?" (1=Right, 2=Left, 3=Both).

On donne ici un résumé des données (parmi les items, uniquement la réponse à l'item 40 est montrée) :

```
summary(dta[,c(40,45:47)])
```

	Q40	education	gender	hand
Min.	:0.00	1: 536	1:1569	1:3447
1st Qu.	:2.00	2:2016	2:2183	2: 435
Median	:3.00	3:1070	3: 301	3: 171
Mean	:2.98	4: 431		
3rd Qu.	:4.00			
Max.	:5.00			

Dans un premier temps, on s'intéresse à la question du lien éventuel entre le niveau d'études (`education`) et la latéralité (`hand`).

Question 2

Donnez le nom d'un test statistique permettant de conclure sur l'existence d'un tel lien à l'échelle de la population. Rappelez les hypothèses nulle et alternative de ce test ainsi que la loi de probabilité de la statistique de test sous l'hypothèse nulle.

Réponse

On met en oeuvre ce test par les commandes R ci-après :

```
tab <- table(dta$hand,dta$education)
RcmdrMisc::rowPercents(tab)
```

```
      1      2      3      4 Total Count
1 13.0 49.9 26.7 10.4   100   3447
2 12.0 51.5 23.2 13.3   100   435
3 21.1 42.1 29.2  7.6   100   171
```

```
test <- chisq.test(tab)
test
```

Pearson's Chi-squared test

```
data:  tab
X-squared = 17.73, df = 6, p-value = 0.00693
round(test$residuals,digits=2)
```

```
      1      2      3      4
1 -0.37  0.13  0.30 -0.34
2 -0.73  0.52 -1.29  1.73
3  2.81 -1.42  0.72 -1.22
```

Question 3

D'après les résultats ci-dessus, diriez-vous qu'il existe un lien à l'échelle de la population entre la latéralité et le niveau d'études (justifiez votre réponse) ? Si oui, quelle est la caractéristique la plus évidente de ce lien.

Réponse

L'équipe de chercheur.e.s en psychologie cherche à construire un modèle expliquant la latéralité par les degrés d'accord à tous les items de personnalité.

Question 4

Quelles sont les variables réponse et explicatives dans cette problématique ? Donnez la nature (quantitative ou catégorielle) de ces variables et, si catégorielle, leurs modalités ?

Réponse

Question 5

Donnez le nom usuel et l'expression mathématique d'un modèle statistique permettant d'aborder cette problématique. Combien ce modèle compte-t'il de paramètres ?

Réponse

Le modèle proposé dans la question 5 est estimé ci-dessous. On donne la déviance résiduelle et le BIC de ce modèle. Pour comparaison, le modèle nul (sans aucune variable explicative) est également estimé. Sa déviance résiduelle et son BIC sont aussi calculés :

```
mod <- nnet::multinom(hand~Q1+Q2+Q3+Q4+Q5+Q6+Q7+Q8+Q9+Q10+
                        Q11+Q12+Q13+Q14+Q15+Q16+Q17+Q18+Q19+Q20+
                        Q21+Q22+Q23+Q24+Q25+Q26+Q27+Q28+Q29+Q30+
                        Q31+Q32+Q33+Q34+Q35+Q36+Q37+Q38+Q39+Q40+
                        Q41+Q42+Q43+Q44,
                        data=dta,
                        trace=FALSE)
deviance(mod)
[1] 3908.76
```

```

BIC(mod)
[1] 4656.41
mod0 <- nnet::multinom(hand~1,
                        data=dta,
                        trace=FALSE)
deviance(mod0)
[1] 4140.85
BIC(mod0)
[1] 4157.46

```

Question 6

Lequel des deux modèles estimés ci-dessus réalise le meilleur compromis entre complexité du modèle et qualité d'ajustement ? Comment expliquez-vous ce résultat ?

Réponse

Pour réduire la complexité du modèle proposé dans la question 5, on propose de sélectionner un sous-ensemble d'items expliquant au mieux la latéralité. Les commandes suivantes mettent en oeuvre un algorithme visant à identifier le modèle réalisant le meilleur compromis entre complexité du modèle et qualité d'ajustement parmi ceux construits à partir d'un sous-ensemble d'items. Le BIC et la table d'analyse de la déviance de type II du modèle sélectionné sont également calculés :

```

select <- RcmdrMisc::stepwise(mod,direction="forward/backward",
                             criterion="BIC",
                             trace=0)

```

```

Direction: forward/backward
Criterion: BIC

```

```

BIC(select)
[1] 4095.74

```

```

car::Anova(select)

```

```

Analysis of Deviance Table (Type II tests)

```

```

Response: hand
      LR Chisq Df Pr(>Chisq)
Q35   66.10  2  4.44e-15 ***
Q4    26.23  2  2.01e-06 ***
---

```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Question 7

Dans la table d'analyse de la déviance ci-dessus, quelles sont les hypothèses nulle et alternative du test réalisé dans la 1ère ligne (Q35) ?

Réponse

Le profil de personnalité expliquant la latéralité est donc désormais réduit aux items 4 (I give people handmade gifts) et 35 (I have taken apart machines just to see how they work). On estime ci-après deux modèles construits à partir de ces deux items et du niveau d'études. On donne pour chacun de ces modèles la table d'analyse de la déviance de type II et les coefficients estimés :

```
mod1 <- nnet::multinom(hand~education*(Q35+Q4),data=dta,  
                        trace=FALSE)
```

```
car::Anova(mod1)
```

Analysis of Deviance Table (Type II tests)

Response: hand

	LR	Chisq	Df	Pr(>Chisq)
education	16.61	6		0.0108 *
Q35	65.16	2		7.10e-15 ***
Q4	26.51	2		1.75e-06 ***
education:Q35	4.91	6		0.5552
education:Q4	8.19	6		0.2244

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
mod2 <- nnet::multinom(hand~education+Q35+Q4,data=dta,  
                        trace=FALSE)
```

```
car::Anova(mod2)
```

Analysis of Deviance Table (Type II tests)

Response: hand

	LR	Chisq	Df	Pr(>Chisq)
education	16.61	6		0.0108 *
Q35	66.21	2		4.19e-15 ***
Q4	26.38	2		1.87e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
summary(mod2)$coefficients
```

	(Intercept)	education2	education3	education4	Q35	Q4
2	-1.91423	0.125086	-0.0411108	0.319947	-0.00234632	-0.0776473
3	-4.97870	-0.715928	-0.4634104	-0.752465	0.46151936	0.2511507

Question 8

D'après les résultats ci-dessus, le lien entre la latéralité et la personnalité est-il le même, quelque soit le niveau d'études ? (Justifiez votre réponse)

Réponse

Question 9

D'après les résultats ci-dessus, quelles sont les caractéristiques les plus évidentes du lien entre la latéralité et la personnalité ? (formulez vos réponses sous la forme suivante : "les gauchèr.e.s sont plus en accord que les droitèr.e.s avec l'item 'I really like dancing' ")

Réponse