

DEA Analyse et Modélisation des Systèmes Biologiques

Introduction au logiciel S-PLUS© - 1999/2000

Solutions

Question 11

```
> dbinom(15, 27, 0.45)
[1] 0.08369243
```

Question 12

```
> 1 - pchisq(5.26, 3)
[1] 0.1537191
```

Question 13

```
> qt(1e-005, 13)
[1] -6.501145
```

Question 14

```
> history("plot")
```

Question 15

La première fixe le dossier de travail et la seconde fixe le fichier de l'historique des commandes.

Question 16

Le bon ordre est le second car il y a des noms de variables, des séparateurs non « , » et pas de noms d'individus.

Question 17

La fonction doit s'appeler « .First ».

Question 18

Les fonctions servent à lire et écrire des fonctions S-Plus dans des fichiers textes.

Question 21 : La variable mesurée diffère t'elle entre les deux sexes chez cette espèce ?

Oui : le test de Wilcoxon donne une p-value de 5 pour mille

```
> males <- c(120, 107, 110, 116, 114, 111, 113, 117, 114, 112)
> femelles <- c(110, 111, 107, 108, 110, 105, 107, 106, 111, 111)
> males
[1] 120 107 110 116 114 111 113 117 114 112
> femelles
[1] 110 111 107 108 110 105 107 106 111 111
> wilcox.test(males, femelles)
Warning messages:
Warning in wil.rank.sum(x, y, alternative, exact, correct): cannot compute exact p-value
with ties
```

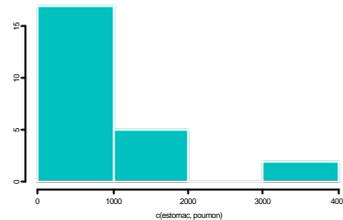
Wilcoxon rank-sum test

```
data: males and femelles
rank-sum normal statistic with correction Z = 2.8171, p-value = 0.0048
```

```
alternative hypothesis: true mu is not equal to 0
```

Question 22 : Pourquoi ? Lequel des deux résultats vous semble le plus solide ?

```
estomac_c(124,42,25,45,412,51,1112,46,103,876,146,340,396)
poumon_c(1235,24,1581,1166,40,727,,3808,791,1804,3460,719)
hist(c(estomac,poumon))
```



La normalité des données est grossièrement fautive. Le test paramétrique est invalide. Le test non paramétrique est largement plus solide.

Question 23

```
> z0 <- mean(cata)
> z0 <- z0 - 1461/2
> z0 <- z0/sqrt((1461 * 1461)/12/142)
> z0
[1] -3.579939
> 1 - pnorm(z0)
[1] 0.9998282
```

La date moyenne est très significativement inférieure à la moyenne attendue. Il y a plus de catastrophes au début. Je décide pour « La fréquence des catastrophes diminue pendant cette période ».

3 - Modèle linéaire

Question 31 : Quelle est l'erreur introduite ?

```
> y_5.75-2.365*x+rnorm(20, sd=0.2)
```

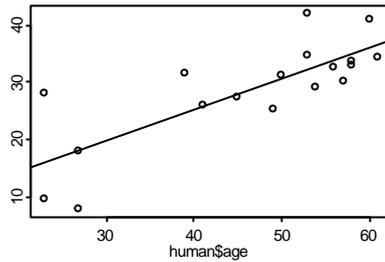
On fabrique un modèle linéaire dont l'écart-type résiduel vaut 0.2

Residual standard error: **20.68** on 18 degrees of freedom

Cet écart-type résiduel est estimé à 20.68. C'est cette valeur qui a été grossièrement modifiée.

Question 32 : Comment l'âge et le pourcentage de graisse sont-ils reliés ?

```
> plot(human$age, human$fat)
```



L'adiposité croît avec l'âge. Un modèle linéaire est acceptable.

```
> lm0_lm(fat~age*sex,data=human)
> anova(lm0)
Analysis of Variance Table
```

Response: fat

Terms added sequentially (first to last)

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
age	1	891.8737	891.8737	44.27361	0.00001089
sex	1	168.7867	168.7867	8.37876	0.01176673
age:sex	1	78.8532	78.8532	3.91436	0.06789632
Residuals	14	282.0243	20.1446		

Cette relation est elle différente chez les hommes (m) et chez les femmes (f) ?

Oui, on peut refuser au niveau 1% l'égalité des ordonnées à l'origine. Vu le nombre de mesures, on peut même refuser au risque de 6% l'égalité des pentes.

Question 33 :

```
> names(lm0)
[1] "coefficients" "residuals" "fitted.values" "effects"
[5] "R" "rank" "assign" "df.residual"
[9] "contrasts" "terms" "call"
```

Les noms qui manquent sont rank et call.

Question 41 : Qu'obtient-on avec l'ordre suivant ?

```
> tapply(fum$DEATHS, as.factor(fum$AGE), sum)
 40 45 50 55 60 65 70 75 80
293 335 402 1183 2009 1933 1436 914 625
```

On obtient le nombre de morts par classes d'âge.

Question 42 : Donner la fréquence des décès par classe d'âge, tout type de comportements confondus.

```
> a0 <- tapply(fum$DEATHS, as.factor(fum$AGE), sum)
> a1 <- tapply(fum$POP, as.factor(fum$AGE), sum)
> a1
 40 45 50 55 60 65 70 75 80
8742 5732 4465 8956 11756 8147 4720 2335 1269
> print(a0/a1, digits = 3)
 40 45 50 55 60 65 70 75 80
0.0335 0.0584 0.09 0.132 0.171 0.237 0.304 0.391 0.493
```

Question 43 : Quel glm redonne comme prédiction ce modèle simple ?

```
> glm0 <- glm(cbind(fum$DEATHS, fum$POP - fum$DEATHS) ~ as.factor(AGE), data = fum,
  family = binomial)
> predict0 <- predict(glm0, type = "response")
> print(predict0[1:9], digits = 3)
  1      2      3      4      5      6      7      8      9
0.0335 0.0584 0.09 0.132 0.171 0.237 0.304 0.391 0.493
```

La réponse est `glm(cbind(fum$DEATHS, fum$POP - fum$DEATHS) ~ as.factor(AGE), data = fum, family = binomial)`

Question 44 :

```
> glm1
Call:
glm(formula = cbind(DEATHS, POP - DEATHS) ~ AGE + SMOKE, family = binomial, data
  = fum)
```

```
Coefficients:
(Intercept)      AGE      SMOKE1      SMOKE2      SMOKE3
-6.74105  0.08421371  0.02368168  0.09153101  0.1100668
```

```
Degrees of Freedom: 36 Total; 31 Residual
Residual Deviance: 39.44169
```

```
> contrasts(fum$SMOKE)
[,1] [,2] [,3]
NO   -1  -1  -1
Y1    1  -1  -1
Y2    0   2  -1
Y3    0   0   3
```

On calcule la prédiction sur le lien.

```
> -6.74105 + 0.08421371*70 + 2*0.09153101 - 0.1100668
[1] -0.7730951
```

On inverse par $x = \log\left(\frac{p}{1-p}\right) \Leftrightarrow p = \frac{\exp(x)}{1+\exp(x)}$

```
> exp(-0.7730951)/(1+exp(-0.7730951))
[1] 0.31581
> predict(glm1)[fum$AGE==70&fum$SMOKE=="Y2"]
 25
-0.7730952
> predict(glm1, type="response")[fum$AGE==70&fum$SMOKE=="Y2"]
 25
0.3158099
```

Question 45 : Quel ordre utiliser pour obtenir ce qui suit :

`anova(glm1, test="Chisq")` *donne le résultat*

Question 46 : Quel ordre utiliser pour obtenir ce qui suit :

`summary(glm1)` *donne le résultat*

Question 47 : Résumer l'information acquise grâce au modèle.

La probabilité de décès est une fonction croissante de l'âge. Le comportement face au tabac dans cette étude est un facteur significatif. Le coefficient du premier contraste (test Y1 contre NO) n'est pas significatif. Par contre Y2 contre les deux premiers et Y3 contre les trois premiers est très significatif. On peut dire « Pipe+Cigare = Témoins, Cigarette + Autres > Témoins, Cigarette seule > Cigarette + Autres > Pipe + Cigare = Témoins ».

	Value	Std. Error	t value
(Intercept)	-6.74105029	0.086955256	-77.5232063
AGE	0.08421371	0.001339912	62.8501659
SMOKE1	0.02368168	0.027308065	0.8672047
SMOKE2	0.09153101	0.011190883	8.1790701
SMOKE3	0.11006680	0.007273717	15.1321260

Il n'y a pas d'interaction.

```
> glm2 <- glm(cbind(DEATHS, POP - DEATHS) ~ AGE * SMOKE, family = binomial, data =
fum)
> anova(glm2, test = "Chisq")
Terms added sequentially (first to last)
```

	Df	Deviance	Resid. Df	Resid. Dev	Pr(Chi)
NULL			35	4917.031	
AGE	1	4643.757	34	273.274	0.0000000
SMOKE	3	233.833	31	39.442	0.0000000
AGE:SMOKE	3	4.393	28	35.049	0.2220546