DEA-AMSB 2003-2004 - Logiciel R - Sur machine

Contrôle sur machine. Merci d'utiliser l'espace imparti pour vos réponses.

Il y a une grande indépendance entre plusieurs parties du questionnaire.

1. Catastrophes : lecture des données

Les 142 catastrophes aériennes [1] survenues entre le 01/01/1972 et le 31/12/1975 sont énumérées dans la chronique cata. La date de chaque accident est donnée en nombre de jours entre 0 et 1461. Ces données sont dans le fichier "cata.txt" à l'endroit habituel.

Utiliser `scan(file="http://pbil.univ-lyon1.fr/R/donnees/cata.txt")`.

```
cata
  [1]    7   17   21   21   26   34   36   42   63   74   79  104  107  109  111
 [16]  126  129  139  142  150  166  167  170  176  181  181  181  211  211  224
 [31]  227  229  240  245  254  257  268  276  276  287  295  301  304  309  323
...
[121] 1093 1094 1112 1126 1130 1154 1167 1209 1271 1308 1308 1311 1327 1338 1340
[136] 1363 1366 1369 1392 1399 1418 1422
```

Quel sont les jours de la chronique qui ont vu plusieurs avions détruits ? Quelle est la longueur en jours du plus grand intervalle de temps entre deux catastrophes ? Quand est-il observé ?

2. Catastrophes : étude de la date moyenne

On peut supposer que chaque accident survient à un quelconque moment dans l'intervalle $\left[a=0, b=1461\right]$ et que deux catastrophes sont indépendantes. D'après le théorème central limite, la moyenne des dates d'observations :

$$Y=\frac{1}{n}\sum_{i=1}^{n}X_i$$

suit alors une loi normale de moyenne $E\left(Y\right)=\frac{b-a}{2}$ et de variance $V\left(Y\right)=\frac{\left(b-a\right)^2}{12n}$. Quelle hypothèse a votre préférence entre :

1 - « La fréquence des catastrophes augmente pendant cette période »
2 - « La fréquence des catastrophes est constante pendant cette période »
3 - « La fréquence des catastrophes diminue pendant cette période » ?

3. Catastrophes : `as.numeric, table et cut`

Les données sont toujours celles de la question 1.

```
br0=seq(0,1461,le=25)
wcata=as.numeric(table(cut(cata,br=br0)))
wcata
[1]  8  7 12  6 10  9  5  4  9  8  6  7  7  6  3  4  5  6  4  2  1  5  6  2
```

Que représente `wcata` ?

4. Catastrophes : signification d'une simulation

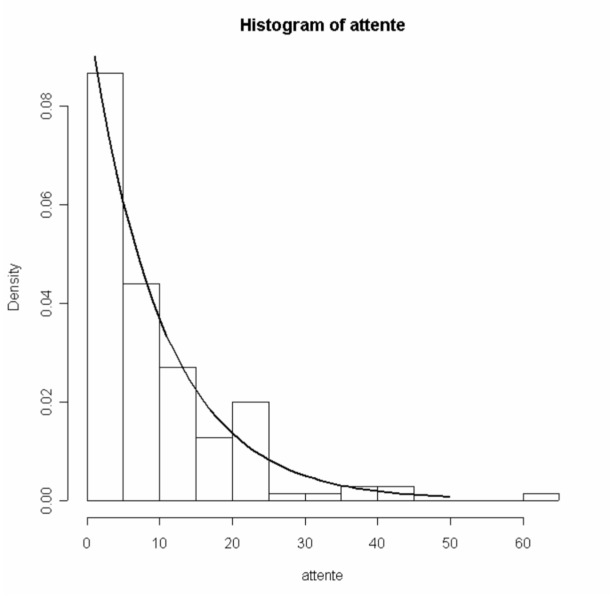
On considère généralement que le temps d'attente de la catastrophe suivante suit une loi exponentielle. Le paramètre d'une loi de ce type s'estime par l'inverse de la moyenne :

```
attente=diff(cata)
```

[1] Eddy P., Potter E. & Page B. (1976). *Destination désastre*. Grasset, Paris. pp. 330-332

```
attente
  [1] 10  4  0  5  8  2  6 21 11  5 25  3  2  2 15  3 10  3  8 16  1  3  6  5  0
 [26]  0 30  0 13  3  2 11  5  9  3 11  8  0 11  8  6  3  5 14 10  5  5  0 12  3
...
[101] 20 18  5  5 35 21 39  6  1  2 24  1  3  4 44 22 11  3 18  6  1 18 14  4 24
[126] 13 42 62 37  0  3 16 11  2 23  3  3 23  7 19  4
```

```
hist(attente,proba=T)
lines(1:50,dexp(1:50,1/mean(attente)),lwd=2)
```




```
f1<-function(k) {
 x=rexp(1,0.09965)
 y=rexp(1,0.09965)
 return(x/(x+y))
}
hist(unlist(lapply(1:10000,f1)))
```

Quelle information peut-on tirer de cette simulation ? A quoi peut-elle servir ?

5. Catastrophes : régression poissonienne

Tester la variation du nombre d'accidents par période avec une régression poissonienne. Comparer l'anova du modèle, le summary et le résultat de la question 2 et commenter le situation.

6. Khi2 et catastrophes

```
br1=seq(0,1461,by=10)
zcata=table(cut(cata,br=br1))
table(zcata)
zcata
 0  1  2  3  4
57 49 29  9  2
```

C'est la distribution du nombre de catastrophes par périodes de 10 jours. Tester par un khi2 d'ajustement si cette distribution suit une loi de Poisson.

7. Khi2 et Khi2 : une catastrophe ?

```
br1=seq(0,1470,by=210)
br1
[1]    0  210  420  630  840 1050 1260 1470
zcata=as.numeric(cut(cata,br=br1))
table(zcata)
zcata
```

```
 1  2  3  4  5  6  7
27 30 23 22 14 12 14
```

C'est la distribution du nombre de catastrophes par périodes de 210 jours. Tester par un Khi2 si la proportion d'accidents par période est constante. Comparer avec les questions précédentes.

8. Khi2 et tourterelles : loi de Mendel

Dans un élevage, on croise des tourterelles à plumage gris et pattes roses avec des tourterelles à plumage brun et pattes noires. La première génération donne uniquement des tourterelles à plumage gris et pattes noires. Le croisement de ces tourterelles entre elles donne les résultats suivants : 1300 tourterelles à plumage gris et pattes roses, 340 tourterelles à plumage brun et pattes roses, 1183 tourterelles à plumage brun et pattes noires et 3649 tourterelles à plumage gris et pattes noires. Cette distribution suit-elle une loi de Mendel ?

9. Khi2 et latéralité : `prop.test` et `chisq.test`

Une enquête sur la latéralité a été réalisée au sein de l'université Lyon 1 auprès de deux populations : les étudiants en STAPS et les étudiants en biologie. Les résultats ci-dessous ont été trouvé pour la main d'écriture :

	Droite	Gauche
STAPS	101	18
Biologie	81	7

Réaliser sur ces données, les deux tests suivants : `prop.test` et `chisq.test`. Commenter.

10. Khi2 et bruit : un piège

On fait passer un test de réactivité visuelle à un groupe de 118 sujets. Chaque sujet passe le test dans deux conditions différentes : avec ou sans bruit dans la salle. Le test se présente comme suit. Deux lampes sont placées à droite et à gauche du sujet. Chaque lampe s'allume de façon aléatoire, avec un temps d'attente variable (entre 0.2s et 0.5s). Le sujet est assis les mains sur les genoux. Dès qu'une lampe s'allume, il doit frapper une plaque située en dessous de la lampe correspondante. On considère qu'un sujet a réussi le test lorsqu'il a réalisé la bonne association « lumière, frappe » au moins 7 fois sur 10.

	avec bruit	
sans bruit	*Succès*	*Echec*
Succès	62	26
Echec	7	23

Quelle conclusion tirez vous de cette expérience ?

Seuls les couples de résultats incluant un changement ont un sens. Échec sans bruit et Succès avec bruit est observé 7 fois. Échec avec bruit et Succès sans bruit est observé 26 fois. Si le bruit n'a aucun rôle la proportion 1/2 est l'hypothèse nulle (l'indépendance est assurée par le changement de personnes).

11. Taille des enfants : lecture des données

Le site `http://www.inrialpes.fr/sel/donnees.html` propose un jeu de données sur la taille et poids d'enfants de 4 à 7 ans. Ces données ont été acquises par le Professeur M. Tauber au CHU de Toulouse. Après mise en forme, on ne garde que les enfants ayant plus de 60 mois et moins de 77 mois. On obtient un tableau à récupérer sous le nom tauber.txt à l'endroit habituel.

Utiliser `read.table(file="http://pbil.univ-lyon1.fr/R/donnees/tauber.txt",h=T)`

Les variables sont **sex**, le sexe de l'enfant M ou F, **age** son âge en mois, **tai** sa taille en cm et **poi** son poids en kg. Combien y a t-il d'enfants dans cette statistique ?

12. Taille des enfants : anova

Extraire la variable **tauber$age** en **age.num** (variable numérique). Extraire la variable **tauber$sex** en **sex** (variable qualitative). Utiliser **age.fac <- factor(age.num)** pour obtenir la variable âge comme variable qualitative. Que vous apprend la commande **anova(lm(y~age.num*sex*age.fac))** ?

13. Taille d'enfants : sous-espaces vectoriels

Expliquer l'origine de la différence entre les résultats de :

```
anova(lm(y~age.num+age.fac))
anova(lm(y~age.fac+age.num))
```

14. Taille d'enfants : predict.lm

Sachant que :

```
summary(lm0)

Call:
lm(formula = y ~ age.num + sex)

Residuals:
    Min      1Q  Median      3Q     Max
-17.135  -3.104  -0.132   3.136  22.009

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  81.0488     1.5857   51.11   <2e-16
age.num       0.4719     0.0231   20.42   <2e-16
sexM          0.5531     0.1750    3.16   0.0016

Residual standard error: 4.63 on 2803 degrees of freedom
Multiple R-Squared: 0.134,      Adjusted R-squared: 0.133
F-statistic:  217 on 2 and 2803 DF,  p-value: <2e-16
```

expliciter comment est calculée la valeur de :

```
predict(lm0,newdata=list(age.num=70.5,sex="M"))
```

15. Taille d'enfants : sélection de modèles

Résumer ce que vous pensez du rôle sur la taille de l'âge et du sexe.

16. Lois discrètes

Une urne contient 10 boules blanches et 990 boules noires. On en tire au hasard 10 sans remise. Qu'elle est la probabilité d'en avoir 2 blanches ? Quelle est l'approximation de cette valeur faite avec la loi binomiale ? Quelle est l'approximation de cette valeur faite avec la loi de Poisson ?

17. Vraisemblance d'une hypothèse

On extrait des données de l'exemple du logiciel WHICHRUN[2] les fréquences alléliques du locus microsatellites OTS-2 dans 5 populations[3] du Saumon quinnat à montaison hivernale, *Oncorhynchus tshawytscha* :

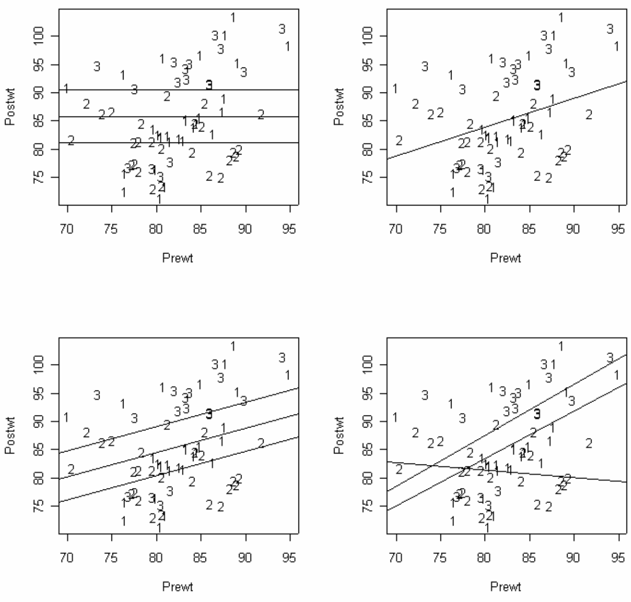
allèle	POP1	POP2	POP3	POP4	POP5
064	0.002	0.000	0.000	0.000	0.000
066	0.038	0.002	0.053	0.066	0.800
070	0.263	0.297	0.447	0.368	0.119
072	0.010	0.004	0.000	0.004	0.000
074	0.030	0.004	0.004	0.041	0.006
078	0.004	0.002	0.000	0.000	0.000
080	0.062	0.029	0.080	0.037	0.003
082	0.016	0.025	0.000	0.000	0.000
084	0.080	0.083	0.053	0.099	0.010
086	0.315	0.359	0.164	0.252	0.045
088	0.006	0.008	0.000	0.000	0.000
090	0.002	0.000	0.000	0.000	0.000
094	0.006	0.000	0.000	0.004	0.000
096	0.082	0.060	0.004	0.045	0.010
098	0.000	0.000	0.000	0.008	0.000
100	0.020	0.025	0.013	0.037	0.000
102	0.056	0.098	0.173	0.029	0.003
104	0.002	0.000	0.009	0.008	0.000
106	0.008	0.004	0.000	0.000	0.003

Sachant que les hypothèses du modèle de Hardy-Weinberg sont satisfaites, estimer au maximum de vraisemblance à laquelle des 5 populations appartient un individu hétérozygote typé 066-070.

18. Analyse de covariance

`?anorexia`

Dans la librairie `MASS`, on trouve le data.frame `anorexia` qui donne pour 72 jeunes filles le poids avant (Prewt) et le poids (Postwt) après un traitement qui appartient à une modalité parmi trois.



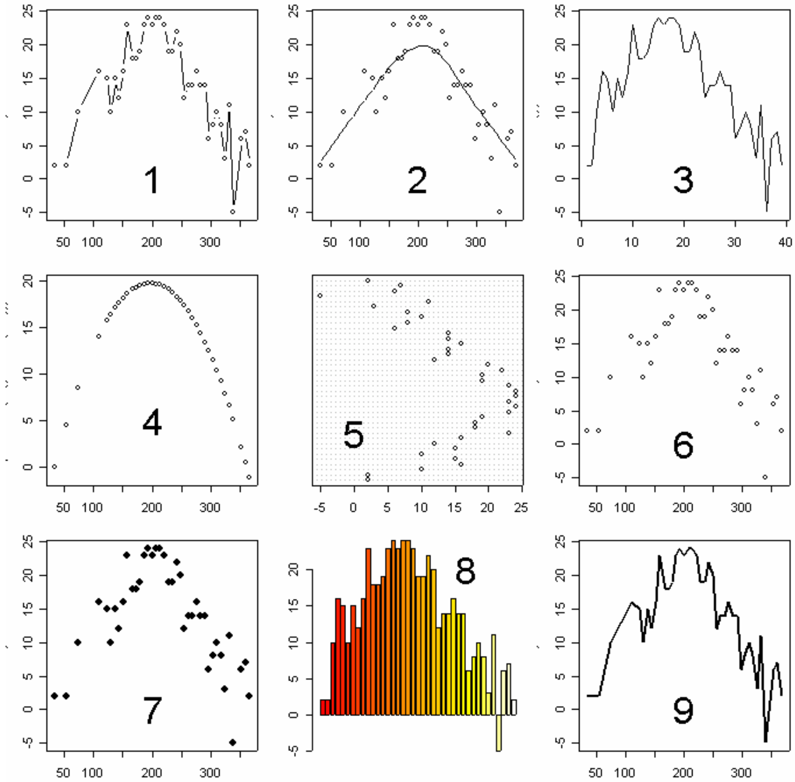
[2] Banks, M. A., and W. Eichert. 2000. Computer note. WHICHRUN (version 3.2): a computer program for population assignment of individuals based on multilocus genotype data. *Journal of Heredity* **91**:87-89.

[3] Greig, C., and M. A. Banks. 1999. Five multiplexed microsatellite loci for rapid response run identification of California's endangered winter chinook salmon. Animal Genetics 30:318-320.

Laquelle de ces 4 figures est-elle la plus appropriée ?

19. Graphique d'une chronique

```
 x
 [1]  33  53  73 109 123 129 138 144 151 158 167 173 179 186 192 200 207 213 221
[20] 228 235 242 249 255 262 269 276 283 291 297 304 311 318 325 332 339 353 360
[39] 367
 y
 [1]  2  2 10 16 15 10 15 12 16 23 18 18 19 23 24 23 24 24 23 19 19 22 20 12 14
[26] 14 16 14 14  6  8 10  8  3 11 -5  6  7  2
```



Donner pour chacune des figures l'ordre qui l'a générée. Les solutions sont dans la liste

- **A** `barplot(y)`
- **B** `plot(x,y,type="b")`
- **C** `plot.ts(ts(y))`
- **D** `dotchart(y)`
- **E** `plot(x,predict.lm(lm(y~x+I(x^2))))`
- **F** `scatter.smooth(x,y)`
- **G** `plot(x,y)`
- **H** `plot(x,y,pch=20,cex=2)`
- **I** `plot(x,y,type="l",lwd=2)`

20. `model.matrix`

```
w=gl(4,3)
```

Editer `w`, observer que c'est un facteur, analyser les résultats obtenus par `model.matrix(~w)` et `contrasts(w)`. Quelles opérations sont-elles nécessaires pour avoir le résultat ci-dessous :

```
model.matrix(~w)
   (Intercept) wA wB wC
1            1 -1 -1  0
2            1 -1 -1  0
3            1 -1 -1  0
4            1 -1  1  0
5            1 -1  1  0
6            1 -1  1  0
7            1  1  0 -1
8            1  1  0 -1
9            1  1  0 -1
10           1  1  0  1
11           1  1  0  1
12           1  1  0  1
```

DEA-AMSB 2003-2004 - Logiciel R - Solutions

1. Catastrophes : lecture des données

```
sort(table(cata))
...
1338 1340 1363 1366 1369 1392 1399 1418 1422   21  211  276  343  418  515  547
   1    1    1    1    1    1    1    1    1    2    2    2    2    2    2    2
 570 1308  181
   2    2    3
```

Aux dates 21, 211, 276, 343, 515, 547, 570 et 1308 il y a eu 2 catastrophes. Le jour 181, il y en a eu 3. C'est le jour le plus noir.

```
max(diff(cata))
[1] 62
which(diff(cata)==62)
[1] 128
cata[127:130]
[1] 1167 1209 1271 1308
```

Le plus grand intervalle sans accident est de 62 jours. Il est intervenu entre les jours 1209 et 1271.

2. Catastrophes : étude de la date moyenne

```
m0=1461/2 # 730.5
s0=sqrt(1461^2/12/142) # 35.39
mean(cata) # 603.8
(mean(cata)-m0)/s0
[1] -3.58
pnorm(-3.58)
[1] 0.0001718
```

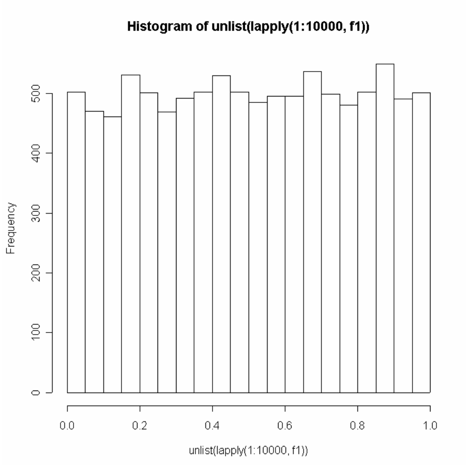
La date moyenne est très significativement inférieure à la valeur prévue. Donc en moyenne les dates sont trop petites, donc les accidents sont plus nombreux avant, donc la fréquence diminue.

3. Catastrophes : `as.numeric, table et cut`

`br0` définit 25 bornes de 0 à 1461 régulièrement réparties, donc 24 intervalles de temps égaux. `cut` fait un facteur qui attribue à chaque événement l'intervalle de temps auquel il appartient. `table` compte les valeurs du facteur par niveau. `wcata` est donc la suite des effectifs de catastrophes pendant 24 périodes de temps égales.

4. Catastrophes : signification d'une simulation

Quand on exécute ce qui précède, on trouve un histogramme caractéristique d'une distribution uniforme. Quand des événements se suivent avec un temps d'attente exponentiel, donc quand X et Y suivent une loi exponentielle et sont indépendants, la variable X/(X+Y) suit une loi uniforme sur [0,1]. On peut dire aussi que, pour trois événements, celui du milieu est tiré au hasard entre les deux autres. On pourrait s'en servir pour faire un test du caractère localement poissonien de la chronique :




```
wdif=diff(cata)
length(wdif)
[1] 141
wdif[-141]/(wdif[-141]+wdif[-1])
w2=wdif[-141]/(wdif[-141]+wdif[-1])
plot(ppoints(139),sort(w2))
abline(c(0,1))
```

5. Catastrophes : régression poissonienne

```
wcata
 [1]  8  7 12  6 10  9  5  4  9  8  6  7  7  6  3  4  5  6  4  2  1  5  6  2
tempo=1:24
glm0=glm(wcata~tempo,family=poisson)
summary(glm0)
...
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.2932     0.1543   14.86   <2e-16
tempo        -0.0451     0.0125   -3.61    3e-04
...
anova(glm0,test="Chisq")
Analysis of Deviance Table

      Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                     23       29.4
tempo  1     13.4        22       15.9   0.00025

Ces tests sont bilatéraux. On compare avec le test non paramétrique qui donne :
2*pnorm(-3.58)
[1] 0.0003436
```

On trouve 3 fois le même résultat avec des moyens très différents. C'est réconfortant.

6. Khi2 et catastrophes

```
 mean(zcata)
[1] 0.9726
 length(zcata)
[1] 146
 obs = c(57, 49, 29, 10)
 the = 146*dpois(0:3,0.9726)
 the[4]=146-sum(the[1:3])
 the
[1] 55.20 53.69 26.11 11.00
 sum(((obs-the)^2)/the)
[1] 0.8788
 pchisq(0.8788,2)
[1] 0.3556
 1-pchisq(0.8788,2)
[1] 0.6444
```

Vraiment rien d'anormal. Rien ne permet de rejeter la loi de Poisson.

7. Khi2 et Khi2 : une catastrophe ?

```
 as.numeric(table(zcata))
[1] 27 30 23 22 14 12 14
 w=as.numeric(table(zcata))
 chisq.test(w,p=rep(1/7,7))

        Chi-squared test for given probabilities

data:  w
X-squared = 14.66, df = 6, p-value = 0.02306
```

La densité d'accidents n'est pas constante. La question 2 et la question 5 l'ont déjà dit 3 fois. La question 6 dit que c'est poissonien. C'est contradictoire. En fait la puissance du test d'ajustement est faible contre l'alternative d'une évolution continue. La distribution est localement poissonienne, comme le dit la question 4.

8. Khi2 et tourterelles : loi de Mendel

Les caractères plumage gris (g) et pattes noires (N) sont dominants, les caractères plumage brun (b) et pattes roses (R) sont récessifs. en F2, on attend P(g) = P(N) = 3/4, P(b) = P(R) = 1/4. Les fréquences attendues sont P(gN) = 9/16, P(gR)=P(bN) = 3/16 et P(bR)=1/16. Les effectifs observés sont E(gN) = 3649, E(gR)=1300, E(bN) = 1183 et E(bR)=340. On fait un test du Khi2 :

```
chisq.test(c(3649,1300,1183,340),p=c(9,3,3,1)/16)

        Chi-squared test for given probabilities

data:  c(3649, 1300, 1183, 340)
X-squared = 17.24, df = 3, p-value = 0.0006316
On rejette l'hypothèse nulle. Cette distribution ne suit pas la loi de Mendel.
```

9. Khi2 et latéralité : `prop.test` et `chisq.test`

```
 chisq.test(matrix(c(101,81,18,7),nr=2))

        Pearson's Chi-squared test with Yates' continuity correction

data:  matrix(c(101, 81, 18, 7), nr = 2)
X-squared = 1.821, df = 1, p-value = 0.1771

 prop.test(matrix(c(101,81,18,7),nr=2))

        2-sample test for equality of proportions with continuity correction

data:  matrix(c(101, 81, 18, 7), nr = 2)
X-squared = 1.821, df = 1, p-value = 0.1771
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.16728  0.02384
sample estimates:
prop 1 prop 2
0.8487 0.9205

Il s'agit de la même chose.
```

10. Khi2 et bruit : un piège

```
 prop.test(7,26,alt="less")

        1-sample proportions test with continuity correction

data:  7 out of 26, null probability 0.5
X-squared = 4.654, df = 1, p-value = 0.01549
alternative hypothesis: true p is less than 0.5
95 percent confidence interval:
 0.0000 0.4489
sample estimates:
```

```
      p
0.2692
```

Le bruit a un effet significatif.

11. Taille des enfants : lecture des données

```
tauber=read.table("tauber.txt",h=T)
dim(tauber)
[1] 2806    4
names(tauber)
[1] "sex" "age" "tai" "poi"
```

La réponse est 2806.

12. Taille des enfants : anova

```
 age.num=tauber$age
 sex=tauber$sex
 age.fac=factor(age.num)
 summary(age.num)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   61.0    66.0    69.0    68.6    71.0    76.0
 summary(age.fac)
 61  62  63  64  65  66  67  68  69  70  71  72  73  74  75  76
 72 102 121 153 179 238 256 259 267 269 205 199 177 129  87  93

anova(lm(y~age.num*sex*age.fac))
Analysis of Variance Table

Response: y
              Df Sum Sq Mean Sq F value Pr(>F)
age.num        1   9075    9075  423.43 <2e-16
sex            1    214     214    9.99 0.0016
age.fac       14    354      25    1.18 0.2837
age.num:sex    1      3       3    0.13 0.7219
sex:age.fac   14    250      18    0.83 0.6326
Residuals   2774  59454      21
```

La croissance peut être considérée comme linéaire (l'âge comme facteur n'apporte rien à l'âge comme variable continue), avec une ordonnée à l'origine qui dépend du sexe mais une pente constante (pas d'interaction âge-sexe).

13. Taille d'enfants : sous-espaces vectoriels

```
anova(lm(y~age.num+age.fac))
Analysis of Variance Table

Response: y
            Df Sum Sq Mean Sq F value Pr(>F)
age.num      1   9075    9075  422.51 <2e-16
age.fac     14    347      25    1.16   0.30
Residuals 2790  59928      21

anova(lm(y~age.fac+age.num))
Analysis of Variance Table

Response: y
            Df Sum Sq Mean Sq F value Pr(>F)
age.fac     15   9423     628    29.3 <2e-16
Residuals 2790  59928      21
```

Il n'y a pas mention d'effet `age.num` après `age.fac` mais mention d'un effet `age.fac` après `age.num` (non significatif). Cela vient du fait que l'âge quantitatif est une combinaison linéaire des indicatrices de classes de l'âge qualitatif. Le sous-espace engendré par `age-num` est à l'intérieur du sous-espace engendré par `age.fac`.

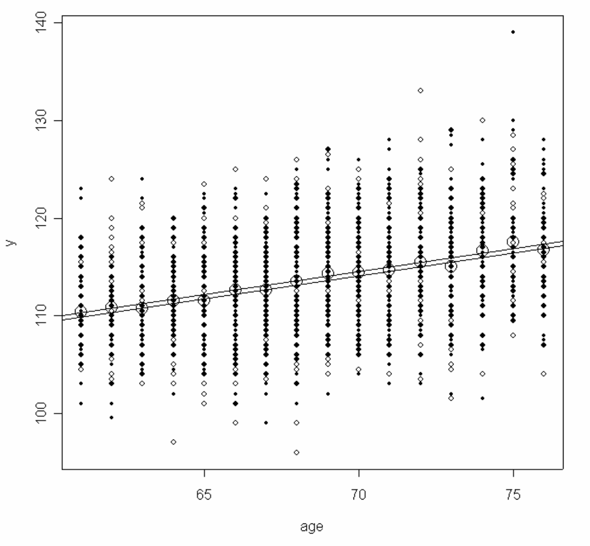
14. Taille d'enfants : predict.lm

```
 predict(lm0,newdata=list(age.num=70.5,sex=factor("M")))
[1] 114.9
à trouver avec :
 81.0488+0.4719*70.5+0.5531
[1] 114.9
```

15. Taille d'enfants : sélection de modèles

Le meilleur modèle est `lm(y ~ age.num + sex)`. On peut considérer, dans l'intervalle d'âge considéré, que la taille augmente linéairement avec l'âge (0.4 cm par mois) et est légèrement supérieure chez lez garçons (0.5 cm). Mais ceci n'explique que 13% de la variabilité : l'erreur standard résiduelle vaut 4.6 cm ce qui revient à un intervalle de variation de 16 cm. On le voit très bien sur un plot :

```
plot(y~age,pch=c(1,20)[as.numeric(sex)])
abline(c(81.05,0.4719))
abline(c(81.05+0.47,0.4719))
points(sort(unique(age.num)),unlist(lapply(split(y,age.fac),mean)),cex=2)
```



16. Lois discrètes

```
dhyper(2,10,990,10) (2 blanches pour 10 boules dont 10 blanches et 990 noires)
[1] 0.003800
dbinom(2,10,1/100) (2 succès pour 10 essais de probabilité 1/100)
[1] 0.004152
dpois (2,0.1) (2 pour une loi de Poisson de moyenne 0.1)
[1] 0.004524
```

17. Vraisemblance d'une hypothèse

La probabilité que l'individu porte un couple d'allèles ab est 2P(a)P(b), donc les vraisemblances (probabilité du résultat sous l'hypothèse que l'individu appartienne à une population donnée) est :

```
a = c(0.038, 0.002, 0.053, 0.066, 0.800)
b = c(0.263, 0.297, 0.447, 0.368, 0.119)
round(2*a*b, dig=3)
[1] 0.020 0.001 0.047 0.049 0.190
Au maximum de vraisemblance, on estime qu'il appartient à la POP5.
```

18. Analyse de covariance

```
Pour refaire la figure

coefficients(lm(Postwt~-1+Treat))
 TreatCBT TreatCont   TreatFT
```

```
     85.7      81.1      90.5
coefficients(lm(Postwt~Prewt))
(Intercept)       Prewt
    42.7006      0.5154
coefficients(lm(Postwt~Prewt+Treat))
(Intercept)       Prewt   TreatCont     TreatFT
    49.7711      0.4345     -4.0971      4.5631
coefficients(lm(Postwt~Prewt*Treat))
    (Intercept)          Prewt       TreatCont          TreatFT Prewt:TreatCont
       15.57724        0.84798        76.47423         -0.75749        -0.98217
  Prewt:TreatFT
        0.06124
```

```
abline(h=c(82.69,81558,83229))
abline(h=c(82.69,81.558,83.229))
```

```
par(mfrow=c(2,2))
plot(Prewt,Postwt,pch=c("1","2","3")[as.numeric(Treat)])
abline(h=c(85.7,81.1,90.5))
plot(Prewt,Postwt,pch=c("1","2","3")[as.numeric(Treat)])
abline(c(42.7,0.5154))
plot(Prewt,Postwt,pch=c("1","2","3")[as.numeric(Treat)])
abline(c(49.77,0.4345))
abline(c(49.77-4.0971,0.4345))
abline(c(49.77+4.5631,0.4345))
plot(Prewt,Postwt,pch=c("1","2","3")[as.numeric(Treat)])
abline(c(15.58,0.8480))
abline(c(15.58+76.474,0.8480-0.98217))
abline(c(15.58-0.75749,0.8480+0.06124))
```

Pour répondre à la question :

```
anova(lm(Postwt~Prewt*Treat, data=anorexia))
Analysis of Variance Table

Response: Postwt
            Df Sum Sq Mean Sq F value  Pr(>F)
Prewt        1    507     507   11.75 0.00105
Treat        2    766     383    8.89 0.00038
Prewt:Treat  2    466     233    5.41 0.00667
Residuals   66   2845      43
```

Les deux facteurs sont influents et il y a interaction. C'est la figure en bas à droite qui contient trois droites de pente différente qui résume correctement les données.

19. Graphique d'une chronique

Réponse : dans l'ordre B F C E D G H A I

Pour refaire la figure

```
par(mfrow=c(3,3))
par(mar=c(3,3,1,1))
plot(x,y,type="b"); text(200,0,"1",cex=3)
scatter.smooth(x,y) ; text(200,0,"2",cex=3)
plot.ts(ts(y)) ; text(20,0,"3",cex=3)
plot(x,predict.lm(lm(y~x+I(x^2)))) ; text(200,5,"4",cex=3)
dotchart(y) ; text(0,10,"5",cex=3)
plot(x,y) ; text(200,0,"6",cex=3)
plot(x,y,pch=20,cex=2) ; text(200,0,"7",cex=3)
barplot(y) ; text(35,20,"8",cex=3)
plot(x,y,type="l",lwd=2); text(200,0,"9",cex=3)
```

20. model.matrix

Il faut transformer les contrastes de w en :

```
contrasts(w)
  A  B  C
```

```
1 -1 -1  0
2 -1  1  0
3  1  0 -1
4  1  0  1
```

On peut le faire par :

```
contrasts(w)=matrix(c(-1,-1,1,1,-1,1,0,0,0,0,-1,1),4,3)
dimnames(contrasts(w))[[2]]=c("A","B","C")
```