

Contrôle continu numéro 3

Vendredi 09 décembre 2016
Durée : 1 heure et 30 minutes.

Le sujet comporte une question de cours et deux exercices indépendants.

- Les calculatrices sont autorisées.
- Le cours, les exercices de travaux dirigés, leurs corrigés ainsi que les notes de cours sont autorisés.
- Afin de pouvoir traiter les questions, plusieurs résultats numériques et graphique ont été intégrés au document.
- Vous prendrez un soin particulier à préciser quelles sont les hypothèses testées.
- Tous les tests seront effectués au seuil de signification $\alpha = 5\%$.

Question de cours : donner le nom de deux tests paramétriques présentés dans le cours et expliquer comment ils se déroulent : hypothèse nulle et hypothèse alternative, statistique de test, règle de décision utilisée. Un exemple dans chacun des deux tests serait le bienvenu.

Exercice 1. La survie des patients dépend-elle de leur traitement ?

Nous testons un médicament destiné à soigner une maladie en phase terminale. Nous traitons des patients avec ce médicament tandis que d'autres patients reçoivent un placebo (groupe contrôle). Nous notons dans la variable statut si les patients ont survécu plus de 48 jours. Voici le tableau obtenu :

traitement	statut	
	non	oui
contrôle	17	29
traité	13	41

Quelle est votre conclusion au seuil $\alpha = 5\%$?

La valeur critique du χ^2 à 1 degré de liberté au niveau 95% est égale à 3,841459.

Exercice 2. Satisfaction et enseignement. Cet exercice comporte six questions.

Une enquête sur 35 élèves ingénieurs de deuxième année, option SI, a porté sur le niveau de satisfaction (note de 0 à 20) des élèves pour cinq matières enseignées (réseaux IP, sécurité et virologie, programmation WEB, administration et programmation système, estimation-analyse de données) à l'ENSIIE (école d'ingénieurs à Strasbourg). Les indices de satisfaction, notés de 0 à 20, obtenus sont les suivants :

matière 1	matière 2	matière 3	matière 4	matière 5
14,5	9,0	8,5	12,0	16,5
12,5	8,0	16,5	17,0	15,5
6,5	12,0	12,5	14,5	16,0
13,0	11,5	14,0	12,5	14,5
8,5	16,0	10,5	15,0	14,0
15,0	10,5	14,0	13,5	17,0
9,5	18,0	10,5	16,5	15,5

1. Proposer un modèle statistique qui permet d'étudier une relation (préciser le type de relation) entre le niveau de satisfaction et la matière enseignée. Préciser la nature de chacune des variables présentes dans le modèle statistique proposé.
2. Les conditions d'application du modèle linéaire sont-elles vérifiées? Si oui, expliquer votre réponse.
3. Donner le tableau de l'analyse de la variance.
4. D'après les sorties statistiques réalisées avec le logiciel R qui se trouvent ci-dessous, pouvez-vous conclure à une éventuelle significativité de la matière enseignée sur le niveau de satisfaction? Pour répondre à cette question, utiliser un test. Vous citerez le nom du test, les hypothèses, la statistique du test et donnerez la conclusion du test (vous préciserez quelle règle vous utilisez).
5. Pouvez-vous séparer les matières enseignées en groupes ne présentant pas de différence significative au seuil de 5%? Si oui, expliquer comment vous procédez.
6. Dans le cas où vous avez répondu dans l'affirmative à la question précédente, faire cette répartition en groupes homogènes, en indiquant les matières enseignées et les moyennes correspondantes au niveau de satisfaction.

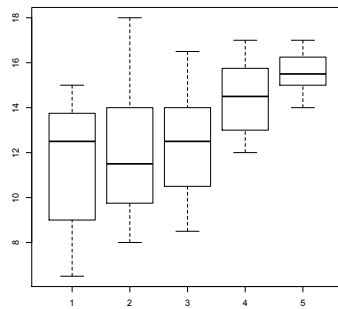


FIGURE 1. Les 5 boîtes à moustaches pour les 5 matières

```

> exo1<-matrix(c(17,29,13,41),
+ byrow=T,nrow=2,
+ dimnames=list(c("controle","soigne"),c("non","oui")))
> chisq.test(exo1)
Pearson's Chi-squared test with Yates' continuity correction
data:  exo1
X-squared = 1.3975, df = 1, p-value = 0.2371
> chisq.test(exo1)$expected
      non  oui
controle 13.8 32.2
soigne   16.2 37.8
> qchisq(0.95,1)
[1] 3.841459
> matière<-rep(1:5,c(7,7,7,7,7))
> matière<-factor(matière)
> note<-c(14.5,12.5,6.5,13.0,8.5,15.0,9.5,9.0,8.0,12.0,11.5,16.0,10.5,
18.0,8.5,16.5,12.5,14.0,10.5,14.0,10.5,12.0,17.0,14.5,12.5,15.0,13.5,
16.5,16.5,15.5,16.0,14.5,14.0,17.0,15.5)
> exo1<-data.frame(matière,note)
> str(exo1)
'data.frame': 35 obs. of 2 variables:
 $ matière: Factor w/ 5 levels "1","2","3","4",...: 1 1 1 1 1 1 1 2 2 2 ...
 $ note    : num  14.5 12.5 6.5 13 8.5 15 9.5 9 8 12 ...
> mean<-tapply(exo1$note,exo1$matière,mean)
> mean
      1      2      3      4      5
11.35714 12.14286 12.35714 14.42857 15.57143
> sd
      1      2      3      4      5
3.223795 3.636783 2.718981 1.902379 1.057850
> boxplot(note~matière)

> modele1<-aov(note~matière,data=exo1)
> modele1
Call:
  aov(formula = note ~ matière, data = exo1)
Terms:

```

```

          matière Residuals
Sum of Squares  86.47143 214.50000
Deg. of Freedom      4      30
Residual standard error: 2.673948
Estimated effects may be unbalanced
> options(contrasts=c("contr.sum","contr.poly"))
> modele2<-lm(note~matière,data=exo1)
> summary(modele2)
all:
lm(formula = note ~ matière, data = exo1)
Residuals:
    Min     1Q  Median     3Q     Max
-4.8571 -1.8571 -0.0714  1.6429  5.8571
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.1714     0.4520   29.142  <2e-16 ***
matière1     -1.8143     0.9040   -2.007   0.0538 .
matière2     -1.0286     0.9040   -1.138   0.2642
matière3     -0.8143     0.9040   -0.901   0.3749
matière4      1.2571     0.9040    1.391   0.1745
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 2.674 on 30 degrees of freedom
Multiple R-squared:  0.2873, Adjusted R-squared:  0.1923
F-statistic: 3.023 on 4 and 30 DF,  p-value: 0.03301
> TukeyHSD(modele1)
Tukey multiple comparisons of means
 95% family-wise confidence level
Fit: aov(formula = note ~ matière, data = exo1)
$matière
      diff      lwr      upr      p adj
2-1 0.7857143 -3.36008262 4.931511 0.9811369
3-1 1.0000000 -3.14579690 5.145797 0.9549702
4-1 3.0714286 -1.07436833 7.217225 0.2266633
5-1 4.2142857  0.06848881 8.360083 0.0448579
3-2 0.2142857 -3.93151119 4.360083 0.9998793
4-2 2.2857143 -1.86008262 6.431511 0.5093810
5-2 3.4285714 -0.71722548 7.574368 0.1433930
4-3 2.0714286 -2.07436833 6.217225 0.6019020
5-3 3.2142857 -0.93151119 7.360083 0.1898701
5-4 1.1428571 -3.00293976 5.288654 0.9286710
> residus1<-residuals(modele1)
> shapiro.test(residus1)
Shapiro-Wilk normality test
data:  residus1
W = 0.98887, p-value = 0.973
> bartlett.test(residus1~matière)
Bartlett test of homogeneity of variances
data:  residus1 by matière
Bartlett's K-squared = 8.5109, df = 4, p-value = 0.07456

```