Contrôle continu numéro 3

Université de Strasbourg - UFR de Mathématique et d'Informatique

Unité d'enseignement de Licence de mathématique : Étude de cas

Vendredi 04 décembre 2015

Durée : 1 heure et 30 minutes.

— Les calculatrices sont autorisées. Les imprimantes sont interdites.
— Les notes de cours, les exercices des travaux dirigés ainsi que leurs corrigés sont autorisés. Tout autre document est interdit
— Afin de pouvoir traiter les questions, plusieurs résultats numériques 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 significativité $\alpha = 5\%$.

Le sujet comporte quatre exercices indépendants.

Exercice 1. Contrôle de qualité et campagne marketing associée.

Laisser tomber son smartphone peut être un drame : on perd ses contacts, tout lien au monde, et en plus, cet accident n'est pas couvert par la garantie et il faudra débourser plusieurs centaines d'euros pour un smartphone de remplacement. Bref, voilà un créneau tout trouvé pour Lemon, un concurrent d'Apple et d'Orange : cette entreprise a construit un smartphone révolutionnaire qui peut encaisser jusqu'à 260 chocs en moyenne, selon l'argumentaire commercial développé. Sceptique, une association civique et féminine d'usagers de téléphonie mobile, « Téléphonées », mène un test sur 81 appareils de Lemon, dont les résultats sont reproduits ci-dessous :

Échantillon	$n = 81$ téléphones
Moyenne des nombres de chocs avant panne	$\widehat{\mu}_n = 264,9$
Écart-type des nombres de chocs	$s_{n,c} = 24,4$

Que faut-il en conclure ?

Vous pouvez répondre à la question posée par exemple avec le test adéquat qui sera effectué au seuil de significativité $\alpha = 5\%$.

```
> qt(0.975,81)
[1] 1.989686
> qt(0.975,80)
[1] 1.990063
> qt(0.95,81)
[1] 1.663884
> qt(0.95,80)
[1] 1.664125
> qnorm(0.95)
[1] 1.644854
> qnorm(0.975)
[1] 1.959964
```

Exercice 2. Les vers à soie.

Un sériciculteur a pesé 100 cocons provenant de son élevage de vers à soie. Il a obtenu les résultats suivants :

Poids en g	$0,64$	$0,65$	$0,66$	$0,67$	$0,68$	$0,69$	$0,70$	$0,71$
Effectif	3	5	2	6	6	10	12	10
Poids en g	$0,72$	$0,73$	$0,74$	$0,75$	$0,76$	$0,77$	$0,78$	$0,79$
Effectif	9	8	8	6	5	4	3	3

Nous appelons μ la moyenne des masses de tous les cocons de l'élevage et nous supposons que la variable « masse d'un cocon » notée X suit une loi normale de variance $0,0016g^2$.

a) Préciser les paramètres de cette loi normale.

b) Nous choisissons au hasard un échantillon de 100 cocons. Nous appelons X_i la masse du i-ème cocon de cet échantillon. Nous supposons que les variables $X_1, \cdots, X_{100}$ sont mutuellement indépendantes, de même loi que la variable X. Nous posons :

$$T_{100} = \frac{1}{100}\sum_{i=1}^{100} X_i.$$

Quelle loi suit la variable aléatoire T_{100} ? Préciser ses paramètres. Vous justifierez vos réponses.

c) Construire un intervalle de confiance pour μ au niveau de confiance de 95% et donner une estimation de cet intervalle.

```
> cocons<-c(rep(0.64,3),rep(0.65,5),rep(0.66,2),rep(0.67,6),rep(0.68,6),
rep(0.69,10),rep(0.70,12),rep(0.71,10),rep(0.72,9),rep(0.73,8),
rep(0.74,8),rep(0.75,6),rep(0.76,5),rep(0.77,4),rep(0.78,3),rep(0.79,3))
> mean(cocons)
[1] 0.7132
> sd(cocons)
[1] 0.03741333
> shapiro.test(cocons)
Shapiro-Wilk normality test
data:  cocons
W = 0.98003, p-value = 0.1336
> qt(0.95,99)
[1] 1.660391
> qt(0.975,99)
[1] 1.984217
> qt(0.95,100)
[1] 1.660234
> qt(0.975,100)
[1] 1.983972
> qnorm(0.95)
[1] 1.644854
> qnorm(0.975)
[1] 1.959964
```

Exercice 3. Mesures des arbres sur pieds.

Nous souhaitons exprimer la hauteur Y d'un arbre en fonction de son diamètre X à 1 mètre 30 du sol. Pour cela, nous avons mesuré 30 couples diamètre-hauteur et les résultats ci-dessous sont disponibles :

$$\widehat{\mu}_{30}(x) = 4,53 \quad s^2_{30}(x) = 10,97 \quad \widehat{\mu}_{30}(y) = 8,65 \quad s^2_{30}(y) = 2,24 \quad \widehat{\mathrm{Cov}}(x,y) = 3,77.$$

a) Nous notons $\widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 X$ l'estimation de la droite de régression des moindres carrés ordinaires. Donner l'expression de $\widehat{\beta}_0$ et de $\widehat{\beta}_1$ en fonction des statistiques élémentaires ci-dessus. Calculer-les.

b) Donner une mesure de la qualité de l'ajustement des données au modèle. Exprimer cette mesure à l'aide des statistiques élémentaires. Calculer et commenter le résultat obtenu.

c) Nous donnons les estimations de l'écart-type de $\widehat{\beta}_0$, $s_{B_0} = 1,62$, et de $\widehat{\beta}_1$, $s_{B_1} = 0,05$. Nous supposons les erreurs gaussiennes, centrées, de même variance et indépendantes.Tester au seuil de significativité $\alpha = 5\%$ les deux hypothèses suivantes :

$$\mathcal{H}_0 : \beta_j = 0$$

contre

$$\mathcal{H}_1 : \beta_j \neq 0$$

pour $j = 0,1$. Pourquoi ce test est-il intéressant dans notre contexte ? Que pensez-vous du résultat ?

```
> qt(0.95,28)
[1] 1.701131
> qt(0.975,28)
[1] 2.048407
> qnorm(0.95)
[1] 1.644854
> qnorm(0.975)
[1] 1.959964
```

Exercice 4. Traitement contre l'urée.

Cinq centres hospitaliers utilisent un traitement différent pour combattre le taux élevé d'urée dans le sang chez les malades atteints de lésions rénales. Le caractère étudié est le taux d'urée (en décigrammes par litre de sang) après traitement. Dans chaque centre hospitalier, nous l'avons mesuré chez sept patients. Les données sont présentées ci-dessous.

traitement 1	traitement 2	traitement 3	traitement 4	traitement 5
4,5	7,5	8,0	2,0	6,5
2,5	3,0	6,5	7,5	5,5
6,0	2,5	6,0	4,0	6,0
4,5	4,0	3,5	2,5	4,5
3,0	2,0	5,0	5,0	4,0
5,5	4,0	7,0	3,5	7,0
3,5	5,5	5,0	6,5	5,5

a) Proposer un modèle statistique qui permet d'étudier une relation (préciser le type de relation) entre le taux d'urée dans le sang et le traitement. Préciser la nature de chacune des variables présentes dans le modèle statistique proposé.

b) Les conditions d'application du modèle linéaire sont-elles vérifiées ? Si oui, expliquer votre réponse.

c) Donner le tableau de l'analyse de la variance.

d) D'après les sorties statistiques réalisées avec le logiciel `R` qui se trouvent ci-dessous, pouvez-vous conclure à une éventuelle significativité du traitement sur le taux d'urée dans le sang ? Pour répondre à cette question, utiliser un test. En particulier, citer le nom du test, les hypothèses, la statistique du test et donner la conclusion du test (préciser quelle règle est utilisée).

e) Pouvez-vous séparer les traitements en groupes ne présentant pas de différence significative au seuil de 5% ? Si oui, expliquer comment vous procédez.

f) Dans le cas où l'affirmative à la question précédente a été choisie, faire cette répartition en groupes homogènes, en indiquant les traitements et les moyennes correspondantes du taux d'urée dans le sang.

```
> traitement<-rep(1:5,c(7,7,7,7,7))
> traitement<-factor(traitement)
> taux<-c(4.5,2.5,6,4.5,3,5.5,3.5,7.5,3,2.5,4,2,4,5.5,8,6.5,6,3.5,5,
7,5,2,7.5,4,2.5,5,3.5,6.5,6.5,5.5,6,4.5,4,7,5.5)
> exo1<-data.frame(traitement,taux)
> str(exo1)
'data.frame':   35 obs. of  2 variables:
 $ traitement: Factor w/ 5 levels "1","2","3","4",
 ..: 1 1 1 1 1 1 1 2 2 2 ...
 $ taux      : num  4.5 2.5 6 4.5 3 5.5 3.5 7.5 3 2.5 ...
> mean<-tapply(exo1$taux,exo1$traitement,mean)
> mean
       1        2        3        4        5
4.214286 4.071429 5.857143 4.428571 5.571429
> var<-tapply(exo1$taux,exo1$traitement,var)
> var
       1        2        3        4        5
1.654762 3.619048 2.226190 4.119048 1.119048
> boxplot(taux~traitement)
```

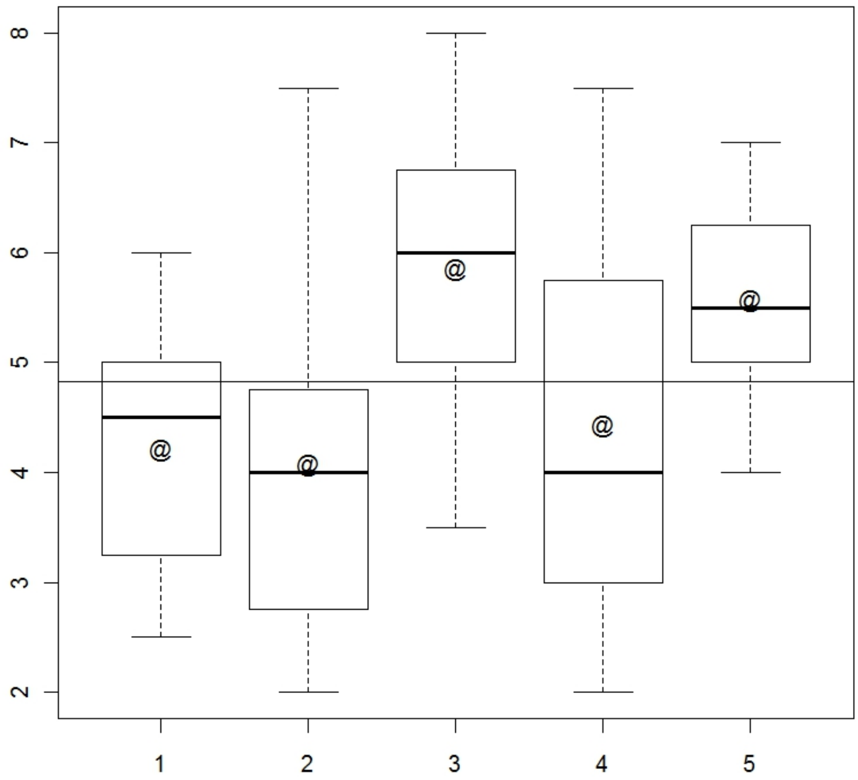

FIGURE 1. Les boîtes à moustaches pour les 5 traitements

```
> modele1<-aov(taux~traitement,data=exo1)
> modele1
Call:
   aov(formula = taux ~ traitement, data = exo1)
Terms:
                traitement Residuals
Sum of Squares    19.04286  76.42857
Deg. of Freedom          4        30
Residual standard error: 1.596126
Estimated effects may be unbalanced
> options(contrasts=c("contr.sum","contr.poly"))
> modele2<-lm(taux~traitement,data=exo1)
> summary(modele3)
Call:
lm(formula = taux ~ traitement, data = exo1)
Residuals:
     Min       1Q   Median       3Q      Max
-2.42857 -1.07143 -0.07143  1.03571  3.42857
Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.8286     0.2698  17.897   <2e-16 ***
traitement1  -0.6143     0.5396  -1.138   0.2639
traitement2  -0.7571     0.5396  -1.403   0.1708
traitement3   1.0286     0.5396   1.906   0.0662 .
traitement4  -0.4000     0.5396  -0.741   0.4643
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Residual standard error: 1.596 on 30 degrees of freedom
Multiple R-squared: 0.1995,     Adjusted R-squared: 0.09272
F-statistic: 1.869 on 4 and 30 DF,  p-value: 0.1419
> TukeyHSD(modele1)
  Tukey multiple comparisons of means
    95% family-wise confidence level
Fit: aov(formula = taux ~ traitement, data = exo1)
$traitement
          diff        lwr      upr     p adj
2-1 -0.1428571 -2.6175554 2.331841 0.9998128
3-1  1.6428571 -0.8318411 4.117555 0.3262525
4-1  0.2142857 -2.2604125 2.688984 0.9990697
5-1  1.3571429 -1.1175554 3.831841 0.5145534
3-2  1.7857143 -0.6889839 4.260413 0.2493343
4-2  0.3571429 -2.1175554 2.831841 0.9932296
5-2  1.5000000 -0.9746982 3.974698 0.4156363
4-3 -1.4285714 -3.9032696 1.046127 0.4641681
5-3 -0.2857143 -2.7604125 2.188984 0.9971344
5-4  1.1428571 -1.3318411 3.617555 0.6693955
> TukeyHSD(modele2)
Erreur dans UseMethod("TukeyHSD") :
  pas de méthode pour ’TukeyHSD’ applicable pour un objet de classe "lm"
> residus1<-residuals(modele1)
> shapiro.test(residus1)
Shapiro-Wilk normality test
data:  residus
W = 0.9744, p-value = 0.5734
> bartlett.test(residus1~traitement)
Bartlett test of homogeneity of variances
data:  residus by traitement
Bartlett’s K-squared = 3.1361, df = 4, p-value = 0.5353
> residus2<-residuals(modele2)
> bartlett.test(residus2~traitement)
Bartlett test of homogeneity of variances
data:  residus2 by traitement
Bartlett’s K-squared = 3.1361, df = 4, p-value = 0.5353
```