Contrôle continu 3

Le sujet comporte trois exercices indépendants.

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

Exercice 1. Prix d'un appartement.
Nous avons relevé quelques annonces d'appartements à vendre dans la presse parisienne journal. Les données sont reproduites dans le tableau suivant :

1.CENSIER, bas de R. Mouffetard, pied-à-terre, 28m2, tt confort. Visite vendredi, samedi, dim. 650.000 F à discuter. Facilités
2. CONTRESCARPE, imm. Ancien, pierre de taille, beau duplex caractère, 50m2, poutres, refait neuf, 1.400.000 F
3. RUE ST-SIMON, 4p., 106m2, en pleine verdure, calme, plein soleil, cuis. aménagée, s. de bains moderne, chff. cent. Parfait état. Px 3.250.000 F à discuter. Agence s'abstenir. Direct propriétaire.
4. RAPP 7P., 196m2 standing, 9 fenêtres plein soleil, 4.000.000 F
5. RUE ST ANDRE DES ARTS, beau liv + chbre, imm. XVIIIe siècle, 55m2, 1.350.000 F.
6. 5e PRES QUAIS, 7 pièces, 190m2 caractère, standing, 3.950.000 F
7. GOBELINS, Beau 5p., 110m2, gd cft, soleil, 250.000 F
8. GOBELINS, et. élevé, calme, asc., 2 pièces,60m2, 1.600.000 F
9. CENSIER, très grand studio + entrée 48m2, tt cft, ensoleillé, calme, bel imm., 1.250.000 F
10. PANTHEON, 7e étage, ascenseur, grand studio 35m2 + terrasse. Vue. 1.250.000 F
11. RUE MADAME, 3P. + Serv., 86m2, 1.750.000 F
12. RUE DE SEINE, 3P., tt cft, 65m2, calme, soleil, 1.500.000 F
13. PANTHEON, bel imm., verdure, magnifique studio 32m2, caractère, 775.000 F.
14. SEVRES BAB, 1er ét., 2P., grande cuis., bns 52m2, état neuf, 1.225.000 F
15. MONTPARNASSE, Part. vend atelier d'artiste 40m2, duplex, vue imprenable, tout confort, Prix 1.000.000 F
16. RUE D'ASSAS, imm. gd standing, bel appart 260m2, triple récept. + 5 ch., tt cft (travaux) 2 park., 2 ch. Serv., Prix 7.500.000 F à déb.
17. BD St-GERMAIN, 4P., 70m2, à amén., 4e ét., 1.625.000 F.
18. ILE St-LOUIS, Lux. appt., 117m2, en duplex, grande récept., grande chambre, 2 sdb, Terras., parf. et., décor tr. bon goût, 4.750.000 F
19. JUSSIEU, Charme, gd 3pces, 90m2, 1.890.000 F
20. QUARTIER LATIN, 30m2 à aménager, prix 390.000 F
21. MONTPARNASSE, Imm. p.d.t., 4-5 P., 105m2, bon état, 1.875.000 F
22. RUE MAZARINE, 52m2, 4e ét., sans ascens., à rénover. Prix 1.000.000 F
23. METRO CENSIER, Bel imm., 4P. 80m2, tt cft, petits travaux, 1.350.000 F
24. ASSAS LUXEMBOURG, 3P. 60m2 s/arbres, imm. caractère, 1.475.000 F
25. SUR JARDINS DE L'OBSERVATOIRE, 140m2, grand charme, 4.950.000 F
26. RUE DE SAVOIE, studio 20m2, 4e ét., dche, 425.000 F.
27. PRES LUXEMBOURG, Bel imm., pierre de taille, Appartement 100m2, salon, salle à manger, 2 chambres, office, cuis., bains, chf. cent., asc., Prix 2.475.000 F
28. METRO LES GOBELINS, studio 28m2, cuis., s. de bains, calme, 425.000 F
29. METRO NATION, studio 19m2, 1er ét., cuis., douche, 300.000 F
30. METRO PLACE D'ITALIE, 5P. 190m2, der. ét., salon, salle à manger, 3 chambres, cuis., s. de bains, PRIX 5.300.000 F

a) Un économiste souhaite utiliser le modèle linéaire simple pour modéliser le lien entre le prix d'un appartement et sa surface. Spécifier ce modèle (donner les coefficients) et bien identifier chacune des composantes du modèle (la variable explicative et la variable à expliquer) dans le contexte de ce problème.

b) Tester l'hypothèse nulle suivante avec un test approprié (vous donnerez le nom de ce test) :
$$\mathcal{H}_0 : \beta_1 = 0$$
contre
$$\mathcal{H}_0 : \beta_1 \neq 0.$$

c) Donner la valeur de la variation qui est expliquée par la droite des MCO et la variation qui est inexpliquée par la même droite. En déduire le pourcentage de variation qui est expliqué par la droite des MCO.

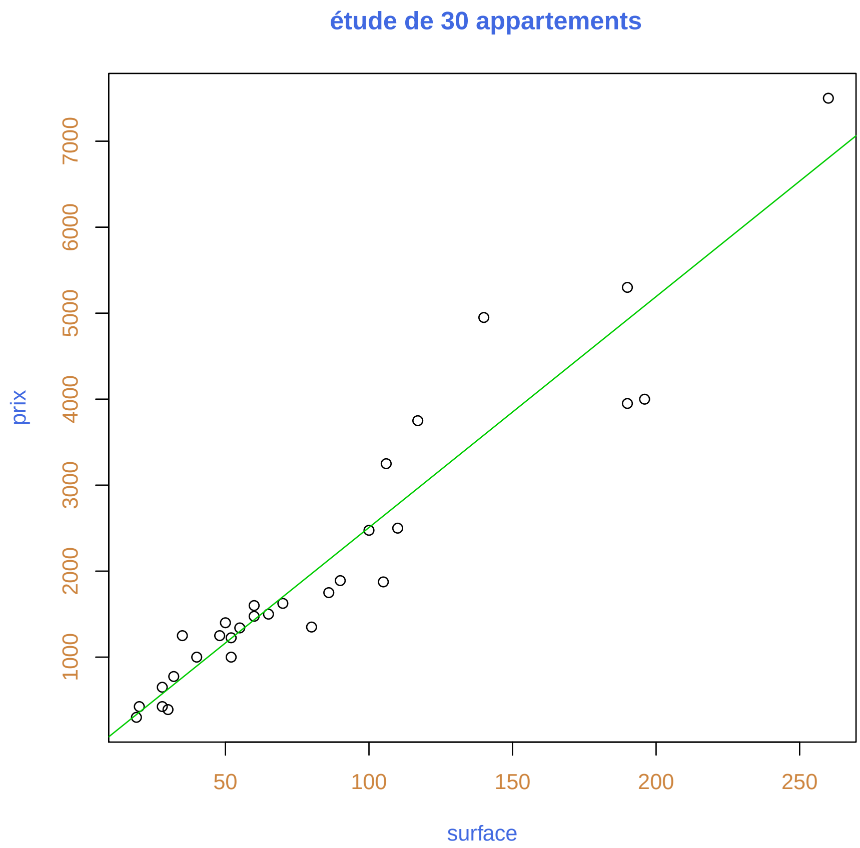
d) Calculer une estimation du prix d'un appartement pour les superficies suivantes : $X_1 = 50, X_2 = 100, X_3 = 200$. Pour quelle superficie l'estimation du prix d'un appartement serait-elle la plus précise parmi les trois valeurs précédentes ?

e) Entre quelles valeurs peut se situer la vraie valeur du prix d'un appartement pour les superficies dont le prix a été déterminé à la question précédente ? Utiliser un niveau de confiance de 95%. Quelle est la marge d'erreur dans l'estimation effectuée à la question précédente ?

```
> surface = c(28, 50, 106, 196, 55, 190, 110, 60, 48, 35, 86, 65, 32,
52, 40, 260, 70, 117, 90, 30, 105, 52, 80, 60, 140, 20, 100, 28, 19, 190)
> prix = c(650, 1400, 3250, 4000, 1340, 3950, 2500, 1600, 1250, 1250,
1750, 1500, 775, 1225, 1000, 7500, 1625, 3750, 1890, 390, 1875, 1000,
1350, 1475, 4950, 425, 2475, 425, 300, 5300)
> prix_au_m2 = c(23.21, 28, 30.66, 20.41, 24.36, 20.79, 22.73, 26.67,
26.04, 35.71, 20.35, 23.08, 24.22, 23.56, 25, 28.85, 23.21, 40.6,
21, 13, 17.86, 19.23, 16.88, 24.58, 35.36, 21.25, 24.75, 15.18, 15.79, 27.89)
> appart = data.frame(surface, prix, prix_au_m2)
> appart
   surface prix prix_au_m2
1       28  650      23.21
2       50 1400      28.00
3      106 3250      30.66
4      196 4000      20.41
5       55 1340      24.36
6      190 3950      20.79
7      110 2500      22.73
8       60 1600      26.67
9       48 1250      26.04
10      35 1250      35.71
11      86 1750      20.35
12      65 1500      23.08
13      32  775      24.22
```

```
14       52 1225      23.56
15       40 1000      25.00
16      260 7500      28.85
17       70 1625      23.21
18      117 3750      40.60
19       90 1890      21.00
20       30  390      13.00
21      105 1875      17.86
22       52 1000      19.23
23       80 1350      16.88
24       60 1475      24.58
25      140 4950      35.36
26       20  425      21.25
27      100 2475      24.75
28       28  425      15.18
29       19  300      15.79
30      190 5300      27.89

> appart.lm = lm(appart$prix~appart$surface)
> appart.lm

Call:
lm(formula = appart$prix ~ appart$surface)

Coefficients:
(Intercept) appart$surface
    -178.07          26.85

> plot(appart$surface,appart$prix,xlab = "surface", ylab = "prix",
col.axis = "peru",col.lab = "royalblue",col.main = "royalblue",
main = " étude de 28 appartements")
> abline(appart.lm, col = "green3")
> residus<-residuals(appart.lm)
> shapiro.test(residus)

        Shapiro-Wilk normality test

data:  residus
W = 0.9609, p-value = 0.327
> summary(appart.lm)

Call:
lm(formula = appart$prix ~ appart$surface)

Residuals:
     Min       1Q   Median       3Q      Max
-1085.40  -232.76    23.86   159.87  1368.45
```

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      -178.069    163.519  -1.089    0.285
appart$surface     26.854      1.597  16.812 3.65e-16 ***
---

Residual standard error: 514.4 on 28 degrees of freedom
Multiple R-squared: 0.9099,     Adjusted R-squared: 0.9066
F-statistic: 282.7 on 1 and 28 DF,  p-value: 3.646e-16
```




```
> anova(appart.lm)
Analysis of Variance Table

Response: appart$prix
               Df   Sum Sq  Mean Sq F value    Pr(>F)
appart$surface  1 74805169 74805169  282.66 3.646e-16 ***
Residuals      28  7410168   264649
---

> predict(appart.lm,interval="confidence")
        fit        lwr       upr
1   573.8555  308.62334  839.0876
```

```
2  1164.6532  942.73973 1386.5667
3  2668.5019 2462.85331 2874.1506
4  5085.4017 4670.93337 5499.8700
5  1298.9254 1084.69470 1513.1561
6  4924.2750 4527.09042 5321.4596
7  2775.9197 2565.29224 2986.5472
8  1433.1976 1225.64214 1640.7531
9  1110.9443  885.69827 1336.1904
10  761.8366  511.81740 1011.8558
11 2131.4131 1938.88487 2323.9413
12 1567.4698 1365.48215 1769.4575
13  681.2733  424.87418  937.6723
14 1218.3621  999.63614 1437.0880
15  896.1088  656.20675 1136.0108
16 6804.0859 6196.31943 7411.8524
17 1701.7420 1504.12108 1899.3630
18 2963.9008 2742.95909 3184.8425
19 2238.8309 2045.37074 2432.2910
20  627.5644  366.79347  888.3353
21 2641.6475 2437.13162 2846.1634
22 1218.3621  999.63614 1437.0880
23 1970.2865 1777.49151 2163.0814
24 1433.1976 1225.64214 1640.7531
25 3581.5530 3315.41822 3847.6877
26  359.0200   75.13437  642.9056
27 2507.3753 2307.81369 2706.9369
28  573.8555  308.62334  839.0876
29  332.1655   45.86543  618.4656
30 4924.2750 4527.09042 5321.4596
```

Exercice 2. Trouvez-vous votre évier trop bas ?

Avez-vous déjà fait la vaisselle dans un évier situé dans une cité U construite dans les années 70 ? Si oui, vous avez remarqué qu'il était très bas. Imaginez que vous soyez un constructeur de cuisines et que vous commenciez à entendre des plaintes sur la hauteur de vos éviers. Vous les construisiez jusqu'à présent pour une taille moyenne de 165 centimètres (vos données vous disaient que la taille moyenne des femmes était justement de 165 centimètres, et vos analyses marketing montrent que ce sont elles les prescriptrices d'achat dans les couples en ce qui concerne la cuisine). Vous conduisez donc une étude et mesurez toutes les femmes qui se présentent à votre magasin (en échange d'un petit cadeau). 64 femmes se prêtent au jeu, cela vous donne les hauteurs $x_1, \ldots, x_{64}$ dont la distribution parente suit une loi normale. La hauteur moyenne relevée est égale à $\overline{x}_{64} = 169,5$ centimètres et l'écart-type sur les données est égal à $s_{64,c} = 16,0$ centimètres.

Faut-il faire remonter cette information au syndicat des constructeurs de cuisines afin qu'il diligente une étude plus approfondie ?

Information dont il vous faudra tenir compte : vous voulez vous présenter à la présidence du syndicat l'an prochain.

Pour répondre à la question, vous allez suivre la démarche suivante :

a) allez-vous utiliser un test paramétrique ou non ? Justifier votre réponse.

b) Donner le nom du test que vous allez utiliser.

c) Donner les hypothèses du test. S'agit-il d'un test bilatéral ou d'un test unilatéral ? Justifier votre réponse.

d) Donner la statistique du test et calculer-la.

e) Conclure. Calculer le risque d'erreur associé à cette décision.

Nous rappelons également quelques valeurs de quantiles qui peuvent vous aider à répondre aux trois exercices :

```
> qnorm(0.975)
[1] 1.959964
> qnorm(0.95)
[1] 1.644854
> qt(0.975,64)
[1] 1.99773
> qt(0.95,64)
[1] 1.669013
> qt(0.975,63)
[1] 1.998341
> qt(0.95,63)
[1] 1.669402
```

Exercice 3. Le classement de logements dans une grande ville française.

Cette étude est réalisée à partir d'un échantillon aléatoire de trente logements situés dans une grande ville française. Pour chaque logement nous disposons de deux variables :

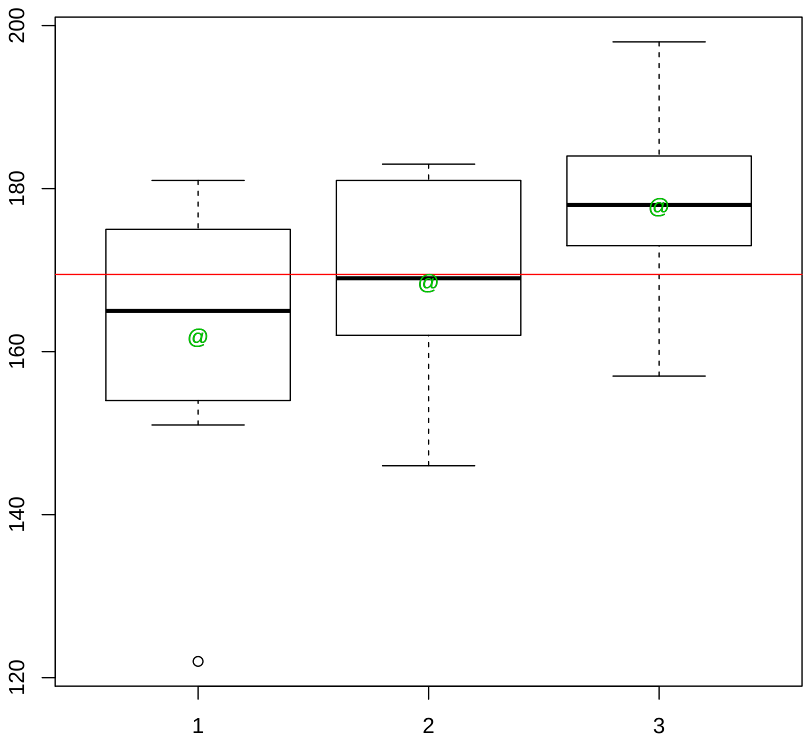
— Prix : prix de vente exprimé en milliers de dollars.

— Situation : variable à trois modalités :

1 : logements situés dans des quartiers du centre ville peu côtés.

2 : logements situés dans les faubourgs.

3 : logements situés dans les banlieues résidentielles.



1. Proposer un modèle statistique qui permet d'étudier une relation (préciser le type de relation) entre le prix de vente et la situation géographique. 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 situation géographique sur le prix de vente ? 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 situations géographiques 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 situations géographiques et les moyennes correspondantes au prix de vente.

```
> situation<-c(3,3,2,2,2,2,2,2,2,2,3,1,1,1,1,1,1,1,3,1,3,1,1,3,3,3,
3,3,2,2)
> prix<-c(198,185,165,170,170,183,158,146,168,162,184,154,170,122,
175,168,181,162,173,178,167,158,151,157,175,181,184,175,181,183)
> exo3<-data.frame(prix,situation)
> str(exo3)
'data.frame':   30 obs. of  2 variables:
 $ prix     : num  198 185 165 170 170 183 158 146 168 162 ...
 $ situation: Factor w/ 3 levels "1","2","3": 3 3 2 2 2 2 2 2 2 2 ...
> exo3
   prix situation
1   198         3
2   185         3
3   165         2
4   170         2
5   170         2
6   183         2
7   158         2
8   146         2
9   168         2
10  162         2
11  184         3
12  154         1
13  170         1
14  122         1
15  175         1
16  168         1
17  181         1
18  162         1
19  173         3
20  178         1
21  167         3
22  158         1
23  151         1
24  157         3
25  175         3
26  181         3
27  184         3
28  175         3
29  181         2
```

```
30  183         2
> modele1<-aov(prix~situation)
> residus<-residuals(modele1)
> shapiro.test(residus)

        Shapiro-Wilk normality test

data:  residus
W = 0.9444, p-value = 0.1198
> bartlett.test(residus~situation)

        Bartlett test of homogeneity of variances

data:  residus by situation
Bartlett's K-squared = 2.0298, df = 2, p-value = 0.3624
> anova(modele1)
Analysis of Variance Table
Response: prix
          Df Sum Sq Mean Sq F value  Pr(>F)
situation  2 1291.3  645.63  3.4354 0.04685 *
Residuals 27 5074.2  187.93
---
> mean<-tapply(exo3$prix,exo3$situation,mean)
> mean
    1     2     3
161.9 168.6 177.9
> TukeyHSD(modele1)
  Tukey multiple comparisons of means
    95% family-wise confidence level
Fit: aov(formula = prix ~ situation)
$situation
    diff        lwr      upr     p adj
2-1  6.7 -8.5008055 21.90081 0.5266048
3-1 16.0  0.7991945 31.20081 0.0376200
3-2  9.3 -5.9008055 24.50081 0.2990454
```