Examen de Statistiques applications avancées en biologie (NUV8SAA1)

****

• *Le cours, les exercices de travaux dirigés, leurs corrigés ainsi que les notes de cours sont autorisés. Tout autre document est interdit.*
• *Les téléphones portables sont formellement interdits.*
• *Les calculatrices sont autorisées.*
• *Tous les tests seront effectués au seuil de signification $\alpha = 5$ %.*
• *Les deux exercices sont indépendants.*
• *Afin de pouvoir traiter les questions, plusieurs résultats numériques et graphiques ont été intégrés au document.*
• *On prendra un soin particulier à préciser quelles sont les hypothèses testées.*

Durée de l'épreuve 2 heures

****

Exercice 1. Effets de la couleur ambiante et du son sur la vigilance.

Dans une expérience investiguant les effets de la couleur de la lumière ambiante sur la performance à une tâche de vigilance on demandait aux participants de presser sur un bouton quand ils pensaient pouvoir discerner un signal dans un fond sonore. L'expérimentateur s'attendait à ce que la couleur ambiante ait des effets variés sur la détection de différents types de sons. Trois types de signaux étaient utilisés : un klaxon, un sifflet et une cloche. Chaque signal était présenté 30 fois au cours d'une session d'auto-évaluation d'une heure au cours de laquelle chaque participant s'asseyait dans une cabine avec une lumière soit rouge soit bleue. La variable dépendante était le nombre de pressions correctes des boutons que l'on considère dans la suite comme une variable quantitative continue. Pour des objectifs théoriques, il était nécessaire d'utiliser différents participants pour différentes conditions de couleur ; en revanche, on considérait qu'il y aurait des avantages à tester chaque individu avec les trois types de signaux.

Participant	Couleur	Signal	Pressions	Participant	Couleur	Signal	Pressions
1	Rouge	Klaxon	25	7	Bleu	Klaxon	19
1	Rouge	Sifflet	18	7	Bleu	Sifflet	12
1	Rouge	Cloche	22	7	Bleu	Cloche	23
2	Rouge	Klaxon	22	8	Bleu	Klaxon	21
2	Rouge	Sifflet	16	8	Bleu	Sifflet	15
2	Rouge	Cloche	21	8	Bleu	Cloche	19
3	Rouge	Klaxon	26	9	Bleu	Klaxon	23
3	Rouge	Sifflet	19	9	Bleu	Sifflet	14
3	Rouge	Cloche	26	9	Bleu	Cloche	24
4	Rouge	Klaxon	23	10	Bleu	Klaxon	20
4	Rouge	Sifflet	21	10	Bleu	Sifflet	16
4	Rouge	Cloche	20	10	Bleu	Cloche	21
5	Rouge	Klaxon	19	11	Bleu	Klaxon	17
5	Rouge	Sifflet	18	11	Bleu	Sifflet	16
5	Rouge	Cloche	19	11	Bleu	Cloche	20
6	Rouge	Klaxon	27	12	Bleu	Klaxon	21
6	Rouge	Sifflet	23	12	Bleu	Sifflet	17
6	Rouge	Cloche	27	12	Bleu	Cloche	19

1. Écrire les deux modèles, dont les résultats sont reportés ci-dessous, d'analyse de la variance utilisés pour cette étude. Pour chacun, on précisera la nature des facteurs explicatifs ainsi que les hypothèses faites.

2. Les hypothèses des deux modèles sont-elles vérifiées ? On se contentera de vérifier les hypothèses pour lesquelles des graphiques sont fournis. On mentionnera explicitement celles qui restent et que l'on suppose désormais vérifiées.

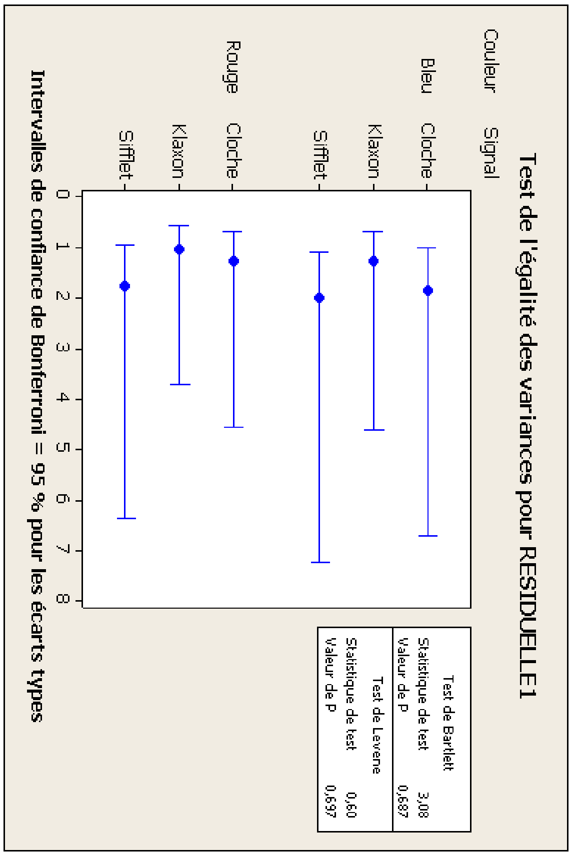
3. Procéder à l'étude du tableau de l'analyse de la variance du **Modèle 1**.

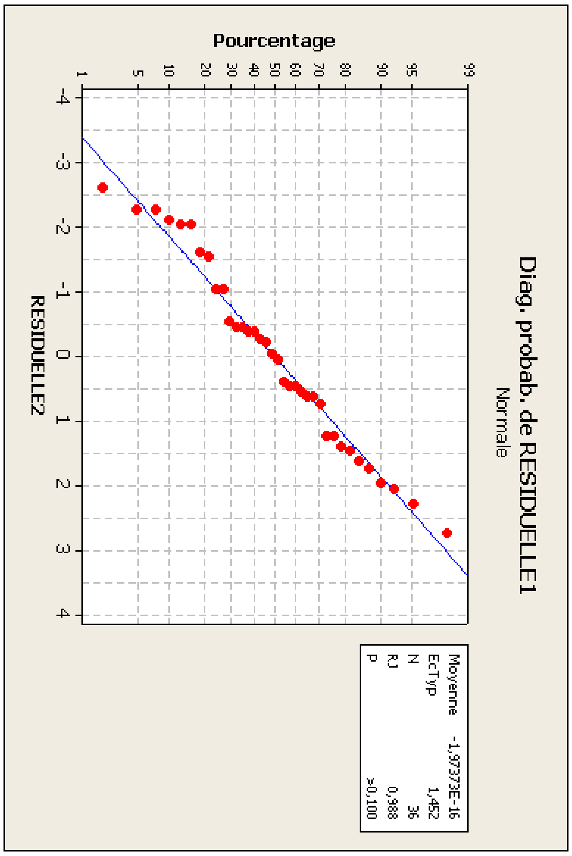
4. Compte tenu de la nature des facteurs peut-on procéder à des comparaisons multiples ? Il y a-t-il un facteur pour lequel cette procédure se justifie ? Le logiciel accepte-t-il de procéder à ces comparaisons ?

5. Expliquer pourquoi les résultats de la question **3.** permettent d'envisager l'utilisation du **Modèle 2**.

6. Compte tenu de la nature des facteurs peut-on procéder à des comparaisons multiples ? Il y a-t-il un facteur pour lequel cette procédure se justifie, si oui procéder aux tests correspondants et interpréter les résultats.

1. **Modèle 1**





Modèle linéaire général : Pressions en fonction de Couleur; Signal; ...

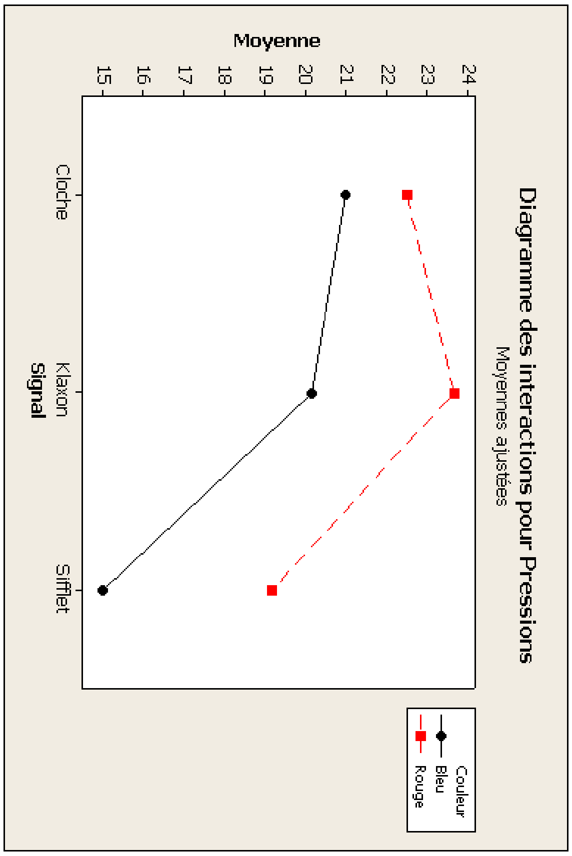
Facteur	Type	Niveaux	Valeurs
Couleur	fixe	2	Bleu; Rouge
Signal	fixe	3	Cloche; Klaxon; Sifflet
Participant(Couleur)	aléatoire	12	7; 8; 9; 10; 11; 12; 1; 2; 3; 4; 5; 6

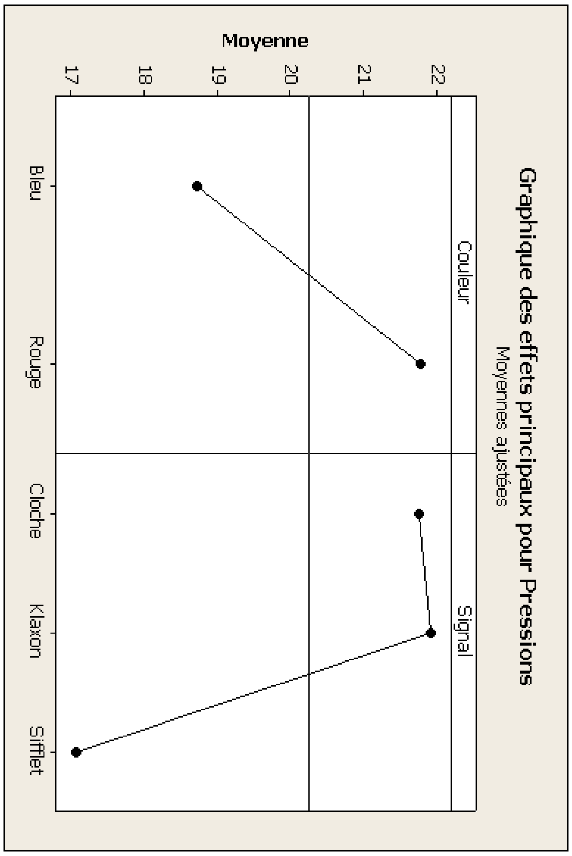
Analyse de la variance pour Pressions, avec utilisation de la somme des carrés ajustée pour les tests

Source	DL	SomCar séq	SomCar ajust	CM ajust	F	P
Couleur	1	84,028	84,028	84,028	7,45	0,021
Signal	2	180,667	180,667	90,333	24,49	0,000
Couleur*Signal	2	11,556	11,556	5,778	1,57	0,233
Participant(Couleur)	10	112,722	112,722	11,272	3,06	0,016
Erreur	20	73,778	73,778	3,689		
Total	35	462,750				

S = 1,92065 R carré = 84,06 % R carré (ajust) = 72,10 %

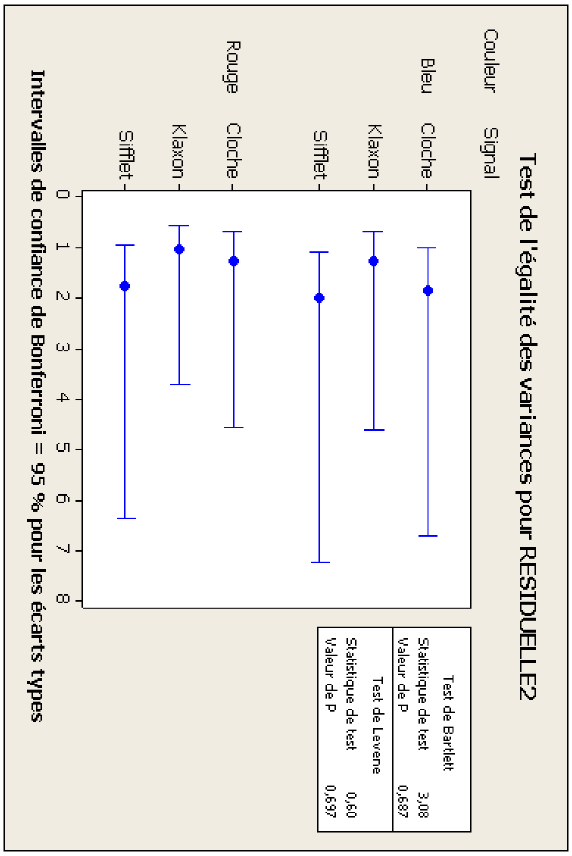
Terme	Coeff	Coef ErT	T	P
Constante	20,2500	0,3201	63,26	0,000
Couleur				
Bleu	-1,5278	0,3201	-4,77	0,000
Signal				
Cloche	1,5000	0,4527	3,31	0,003
Klaxon	1,6667	0,4527	3,68	0,001
Couleur*Signal				
Bleu Cloche	0,7778	0,4527	1,72	0,101
Bleu Klaxon	-0,2222	0,4527	-0,49	0,629
(Couleur)Participant				
Bleu 7	-0,722	1,012	-0,71	0,484
Bleu 8	-0,389	1,012	-0,38	0,705
Bleu 9	1,611	1,012	1,59	0,127
Bleu 10	0,278	1,012	0,27	0,787
Bleu 11	-1,056	1,012	-1,04	0,310
Rouge 1	-0,111	1,012	-0,11	0,914
Rouge 2	-2,111	1,012	-2,09	0,050
Rouge 3	1,889	1,012	1,87	0,077
Rouge 4	-0,444	1,012	-0,44	0,665
Rouge 5	-3,111	1,012	-3,07	0,006

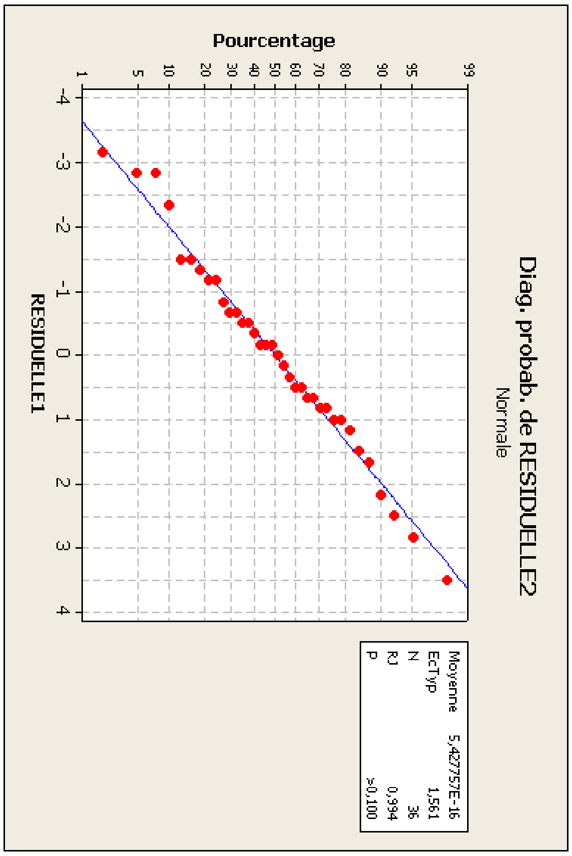




```
* ATTENTION * Aucune comparaison multiple n'a été calculée pour les termes
              suivants, qui contiennent (ou interagissant avec) des facteurs
              aléatoires.

Couleur; Signal
```

2. Modèle 2






```
Modèle linéaire général : Pressions en fonction de Couleur; Signal; ...

Facteur               Type       Niveaux  Valeurs
Couleur               fixe             2  Bleu; Rouge
Signal                fixe             3  Cloche; Klaxon; Sifflet
Participant(Couleur)  aléatoire       12  7; 8; 9; 10; 11; 12; 1;
                                          2; 3; 4; 5; 6

Analyse de la variance pour Pressions, avec utilisation de la
     somme des carrés ajustée pour les tests

                                       SomCar
Source                DL  SomCar séq    ajust  CM ajust      F      P
Couleur                1      84,028   84,028    84,028   7,45  0,021
Signal                 2     180,667  180,667    90,333  23,29  0,000
```

Participant(Couleur)	10	112,722	112,722	11,272	2,91	0,018
Erreur	22	85,333	85,333	3,879		
Total	35	462,750				

S = 1,96946 R carré = 81,56 % R carré (ajust) = 70,66 %

Terme	Coeff	Coef ErT	T	P
Constante	20,2500	0,3282	61,69	0,000
Couleur				
Bleu	-1,5278	0,3282	-4,65	0,000
Signal				
Cloche	1,5000	0,4642	3,23	0,004
Klaxon	1,6667	0,4642	3,59	0,002
(Couleur)Participant				
Bleu 7	-0,722	1,038	-0,70	0,494
Bleu 8	-0,389	1,038	-0,37	0,712
Bleu 9	1,611	1,038	1,55	0,135
Bleu 10	0,278	1,038	0,27	0,791
Bleu 11	-1,056	1,038	-1,02	0,320
Rouge 1	-0,111	1,038	-0,11	0,916
Rouge 2	-2,111	1,038	-2,03	0,054
Rouge 3	1,889	1,038	1,82	0,082
Rouge 4	-0,444	1,038	-0,43	0,673
Rouge 5	-3,111	1,038	-3,00	0,007

Observations aberrantes pour Pressions

Observation	Pressions	Valeur ajustée	Ajust ErT	Valeur résiduelle	Valeur résiduelle normalisée	
21	23,0000	19,5000	1,2282	3,5000	2,27	R
26	14,0000	17,1667	1,2282	-3,1667	-2,06	R

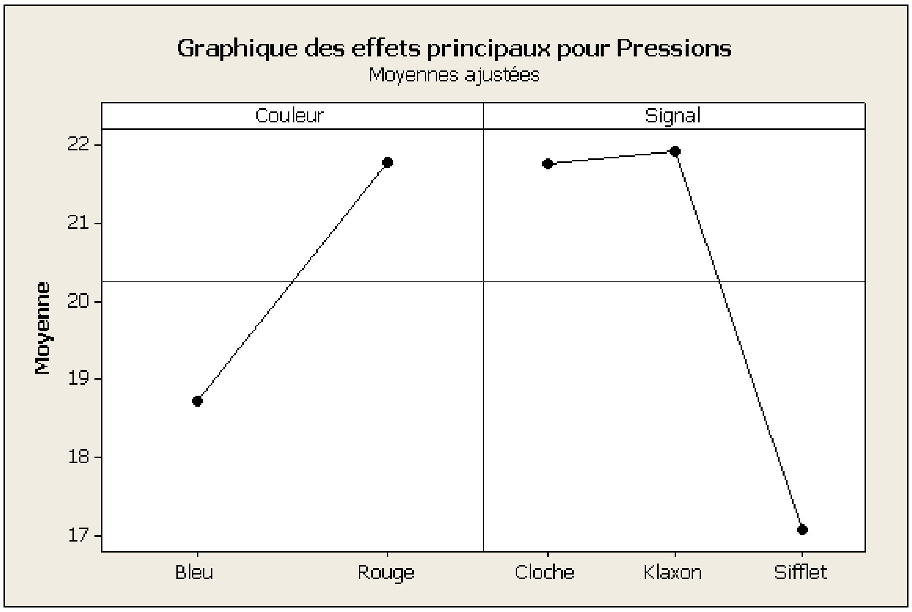
R indique une observation ayant une valeur résiduelle normalisée importante

Intervalles de confiance simultanés de Tukey = 95,0 %
Variable de réponse Pressions
Toutes les comparaisons deux à deux sur les niveaux de Signal
Signal = Cloche soustrait de :

```
Signal   Inférieur  Centré  Supérieur  -------+---------+---------+---------
Klaxon      -1,852   0,167      2,185                      (-------*-------)
Sifflet     -6,685  -4,667     -2,648  (-------*-------)
                                       -------+---------+---------+---------
                                           -5,0      -2,5       0,0
```

Signal = Klaxon soustrait de :

```
Signal   Inférieur  Centré  Supérieur  -------+---------+---------+---------
Sifflet     -6,852  -4,833     -2,815  (-------*-------)
                                       -------+---------+---------+---------
                                           -5,0      -2,5       0,0

Tests de simultanéité de Tukey
Variable de réponse Pressions
Toutes les comparaisons deux à deux sur les niveaux de Signal
Signal = Cloche  soustrait de :

                       Erreur type
         Différence         de la  Valeur  Valeur de
Signal   des moyennes  différence    de T  p ajustée
Klaxon          0,167      0,8040   0,207     0,9766
Sifflet        -4,667      0,8040  -5,804     0,0000

Signal = Klaxon  soustrait de :

                       Erreur type
         Différence         de la  Valeur  Valeur de
Signal   des moyennes  différence    de T  p ajustée
Sifflet        -4,833      0,8040  -6,011     0,0000
```



Exercice 2. Ventes de portraits.

Une entreprise spécialisée dans la vente de portraits de jeunes de moins de 16 ans souhaite étudier l'influence éventuelle sur le volume des ventes de chacun de ses studios, exprimées en *k*€, de deux variables explicatives, le nombre d'individus de moins de 16 ans recensé dans la ville où est installé le studio, exprimé en milliers d'habitants, et le revenu annuel moyen par habitant, exprimé en *k*€. Chaque studio est implanté dans une ville différente.
Le tableau suivant montre le relevé de ces variables pour l'année 2007 dans les 21 studios de l'entreprise.

Population	Revenus	Ventes	Population	Revenus	Ventes
68,5	16,7	174,4	72,8	17,1	191,1
45,2	16,8	164,4	88,4	17,4	232,0
91,3	18,2	244,2	42,9	15,8	145,3
47,8	16,3	254,6	52,5	17,8	161,1
46,9	17,3	282,6	85,7	18,4	109,7
66,1	18,3	207,5	41,3	16,5	146,4
49,5	15,9	152,8	51,7	16,3	144,0
52,0	17,2	163,2	89,6	18,1	232,6
48,9	16,6	145,4	82,7	19,1	224,1
38,4	16,0	137,2	52,3	16,0	266,5
87,9	18,3	241,9			

1. Écrire le modèle de régression permettant d'expliquer les ventes à l'aide de toutes les variables explicatives proposées. On précisera la nature des facteurs explicatifs ainsi que les hypothèses faites.

2. Les hypothèses du modèle sont-elles vérifiées ? Calculer les estimations de tous les paramètres du modèle.

4. Ce modèle est-il intéressant ?

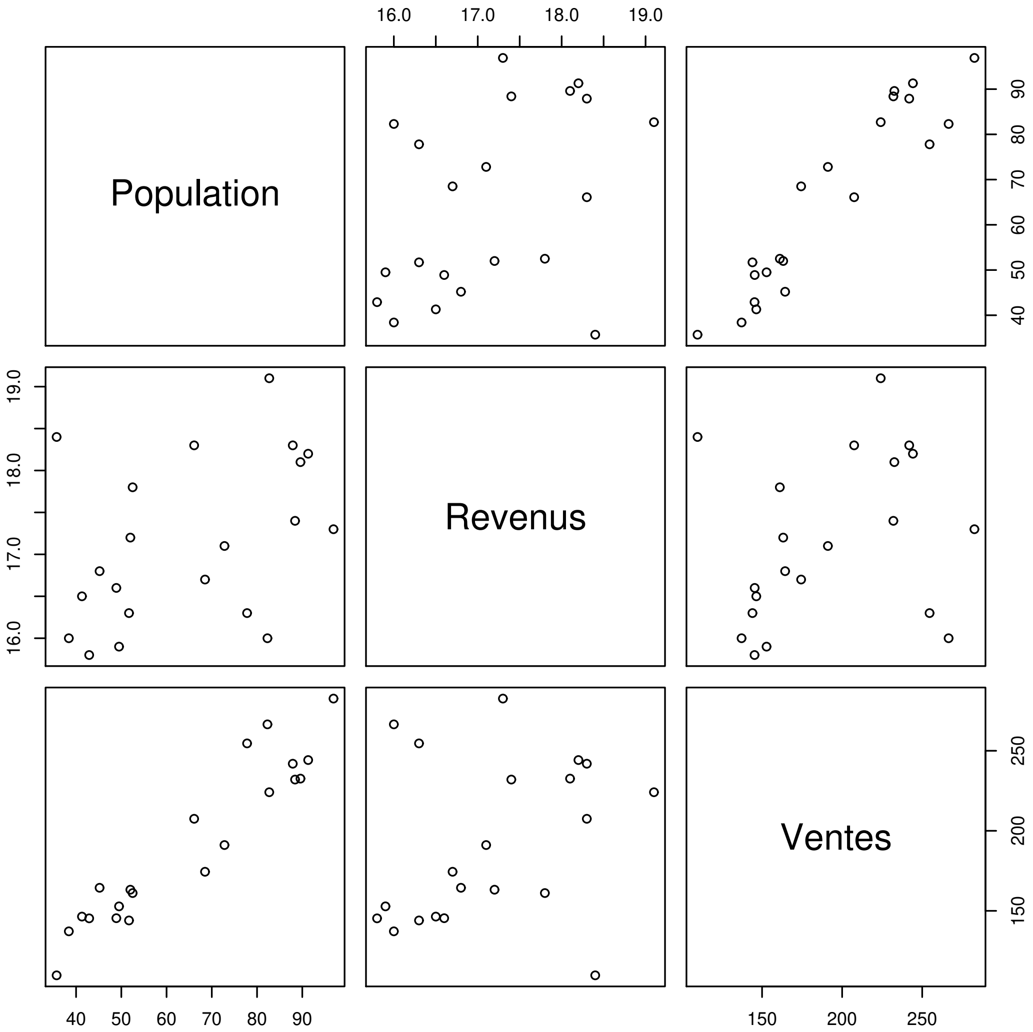
5. Quelles sont les variables explicatives qui influent significativement sur le volume des ventes ? Est-il pertinent de simplifier le modèle introduit en **1.** ?

6. À l'aide des résultats ci-dessous, donner un intervalle de confiance à 95% pour les paramètres du modèle.

7. L'entreprise souhaite implanter un nouveau studio dans une nouvelle ville. Celle-ci compte 50500 jeunes de moins de 16 ans et le revenu annuel moyen des habitants est de 17100 €. Quel est le volume des ventes prédit par le modèle ? À l'aide des résultats ci-dessous, donner un intervalle de confiance à 95% pour ce volume.

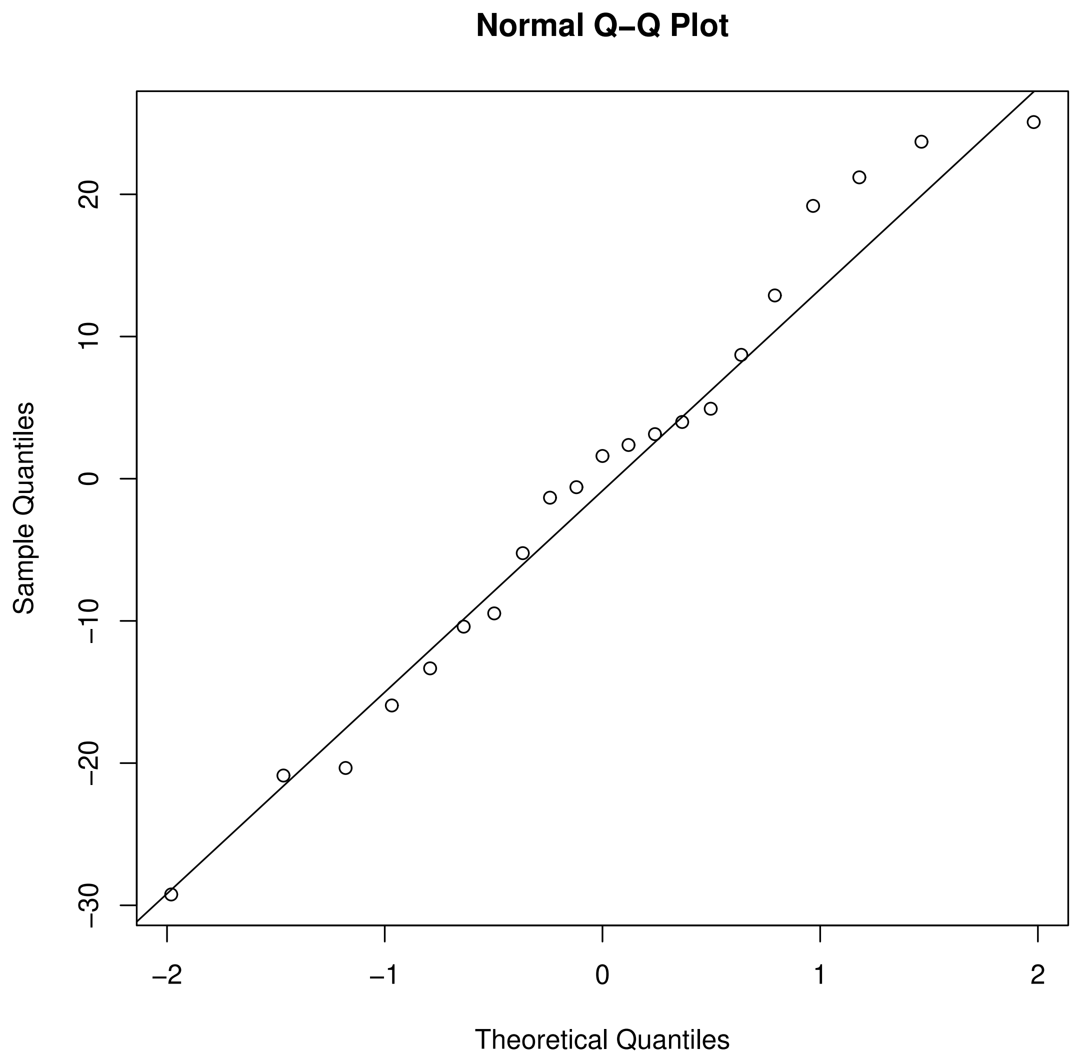
```
> donnees <- read.csv(file.choose())
> attach(donnees)

> str(donnees)
```

```
'data.frame':   21 obs. of  3 variables:
 $ Population: num  68.5 45.2 91.3 77.8 96.9 66.1 49.5 52 48.9 38.4 ...
 $ Revenus   : num  16.7 16.8 18.2 16.3 17.3 18.3 15.9 17.2 16.6 16 ...
 $ Ventes    : num  174 164 244 255 283 ...

> plot(donnees)
```




```
> model <- lm(Ventes~Population+Revenus)
> print(model)

Call:
lm(formula = Ventes ~ Population + Revenus)

Coefficients:
(Intercept)   Population      Revenus
```

```
  152.834        2.477       -7.118

> summary(model)

Call:
lm(formula = Ventes ~ Population + Revenus)

Residuals:
    Min      1Q  Median      3Q     Max
-29.238 -10.403   1.598   8.710  25.080

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 152.8342    65.5992   2.330   0.0317 *
Population    2.4769     0.1958  12.647 2.16e-10 ***
Revenus      -7.1176     4.0745  -1.747   0.0977 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.11 on 18 degrees of freedom
Multiple R-Squared: 0.9071,     Adjusted R-squared: 0.8968
F-statistic: 87.92 on 2 and 18 DF,  p-value: 5.135e-10

> shapiro.test(residuals(model))

        Shapiro-Wilk normality test

data:  residuals(model)
W = 0.9718, p-value = 0.7735
```




```
> confint(model)
                 2.5 %     97.5 %
(Intercept)  15.015441 290.653031
Population    2.065447   2.888345
Revenus     -15.677689   1.442521

> predict(model,list(Population=50.5,Revenus=17.1),
+ interval = "prediction")
          fit      lwr      upr
[1,] 156.2068 121.0884 191.3251
```