Frédéric Bertrand Session Janvier 2011

Université de Strasbourg Master Éthologie-Écophysiologie 1ère année
1er semestre 2010/2011

Examen de « Outils Élémentaires de statistiques appliquées » (VI22GUST)

****

• *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

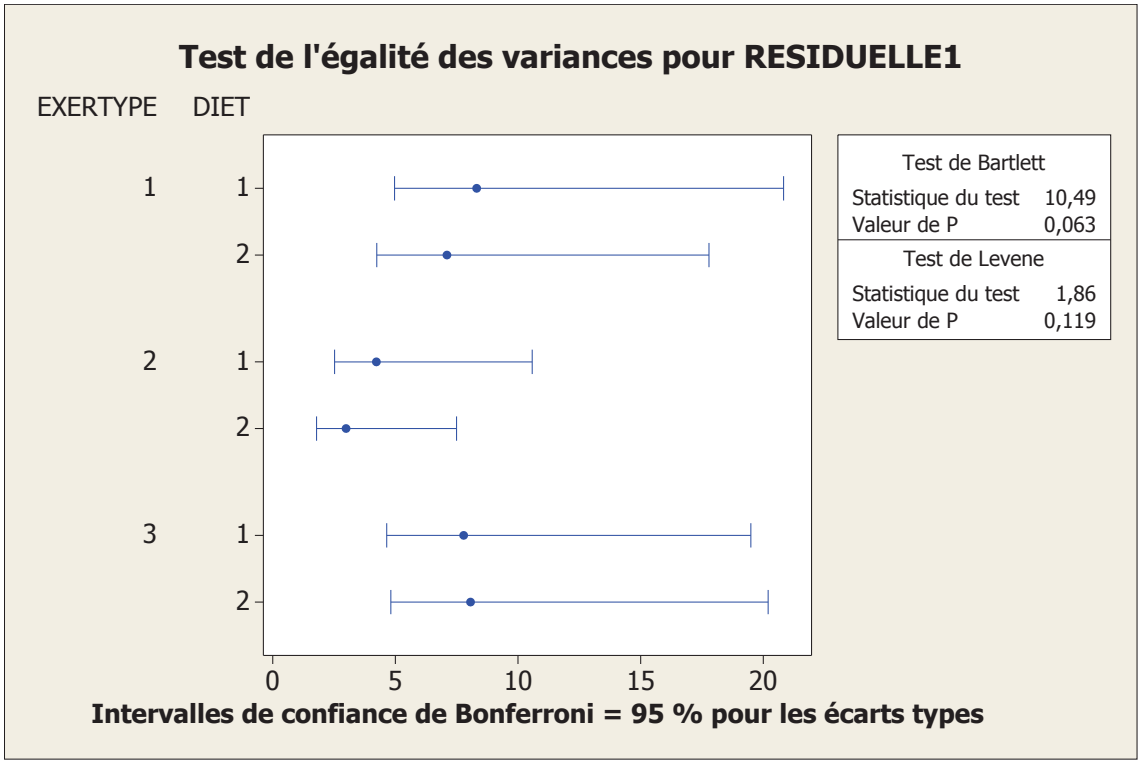
****

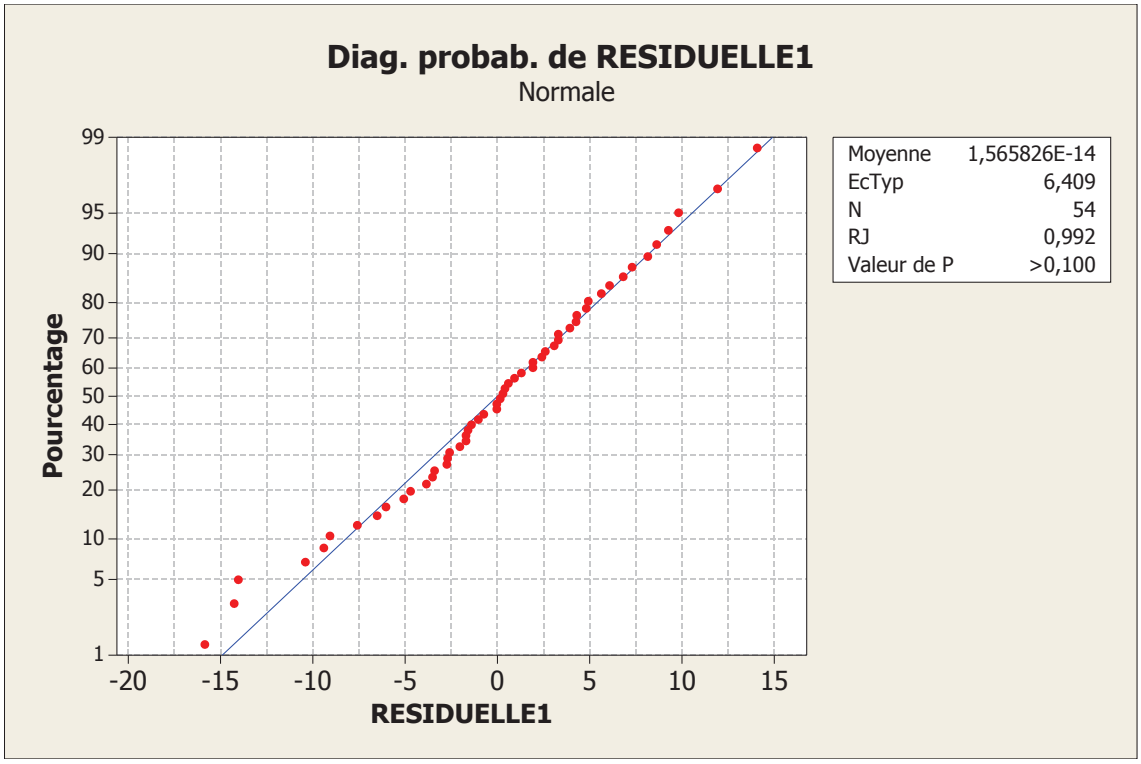
Exercice 1. Vous avez mesuré la pulsation cardiaque de dix-huit sujets dans trois conditions expérimentales différentes. Les premières mesures (EXERTYPE=1) ont été prises lors de l'échauffement des sujets avant l'effort. Les secondes mesures (EXERTYPE=2) ont été prises après 10 minutes de course. Les dernières mesures (EXERTYPE=3) ont été prises lors de la phase de récupération après l'effort. La variable régime alimentaire (DIET) indique les préférences alimentaires des sujets : (DIET=1) pour les carnivores et (DIET=2) pour les végétariens.

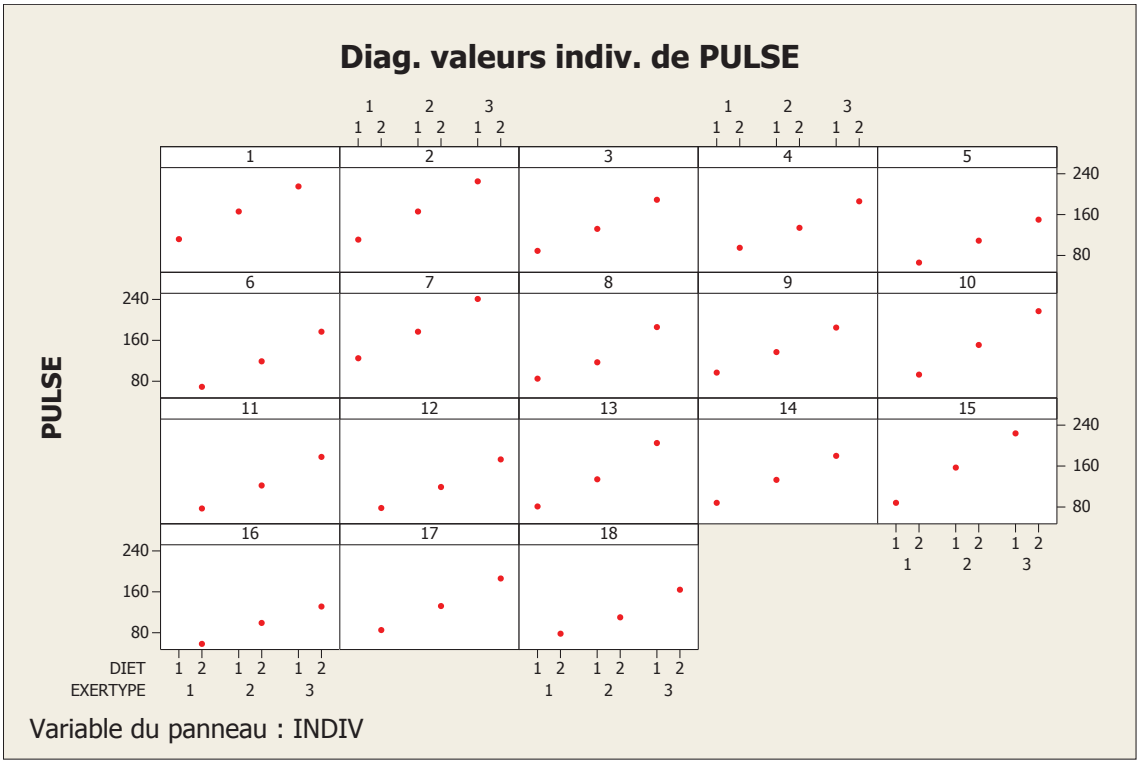
Les résultats sont présentés dans le tableau ci-après.

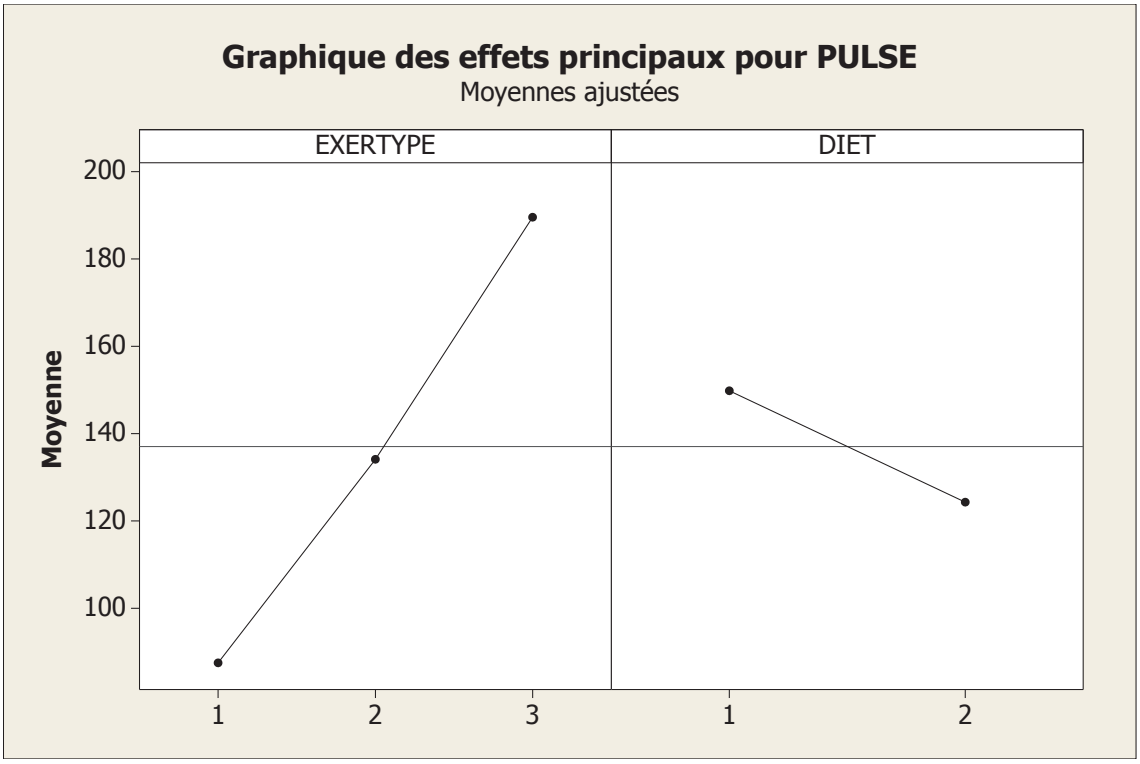
1. Écrire le modèle d'analyse de la variance relatif au modèle dont les résultats sont reproduits dans la suite. On précisera la nature des facteurs explicatifs ainsi que les hypothèses faites.
2. Indiquer pourquoi il faut considérer dans le modèle un troisième facteur « Indiv », de nature aléatoire.
3. Les hypothèses du modèle sont-elles vérifiées ? Calculer les estimations de tous les paramètres du modèle.
4. Procéder à l'étude du tableau de l'analyse de la variance.
5. 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.

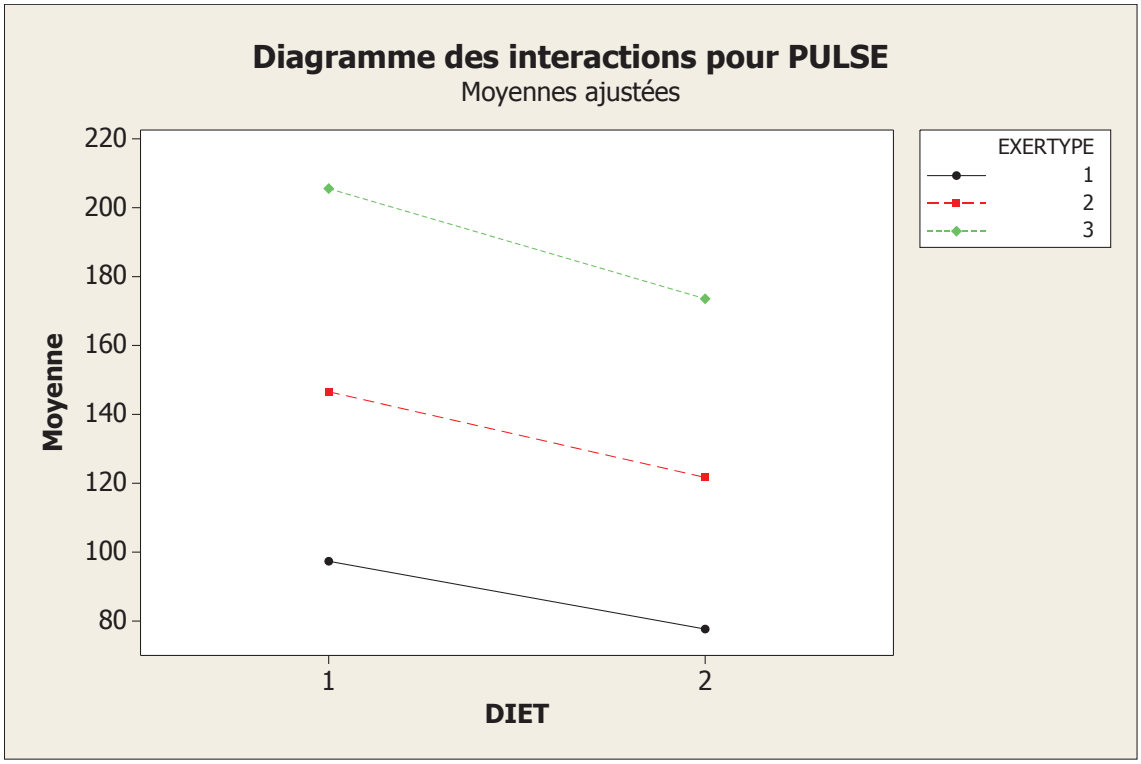
EXERTYPE	DIET	INDIV	PULSE	EXERTYPE	DIET	INDIV	PULSE	EXERTYPE	DIET	INDIV	PULSE
1	1	1	112	2	1	1	166	3	1	1	215
1	1	2	111	2	1	2	166	3	1	2	225
1	1	3	89	2	1	3	132	3	1	3	189
1	2	4	95	2	2	4	134	3	2	4	186
1	2	5	66	2	2	5	109	3	2	5	150
1	2	6	69	2	2	6	119	3	2	6	177
1	1	7	125	2	1	7	177	3	1	7	241
1	1	8	85	2	1	8	117	3	1	8	186
1	1	9	97	2	1	9	137	3	1	9	185
1	2	10	93	2	2	10	151	3	2	10	217
1	2	11	77	2	2	11	122	3	2	11	178
1	2	12	78	2	2	12	119	3	2	12	173
1	1	13	81	2	1	13	134	3	1	13	205
1	1	14	88	2	1	14	133	3	1	14	180
1	1	15	88	2	1	15	157	3	1	15	224
1	2	16	58	2	2	16	99	3	2	16	131
1	2	17	85	2	2	17	132	3	2	17	186
1	2	18	78	2	2	18	110	3	2	18	164


Test de l'égalité des variances pour RESIDUELLE1
EXERTYPE DIET
1 1
2
2 1
2
3 1
2
0 5 10 15 20
Intervalles de confiance de Bonferroni = 95 % pour les écarts types
Test de Bartlett
Statistique du test 10,49
Valeur de P 0,063
Test de Levene
Statistique du test 1,86
Valeur de P 0,119


Diag. probab. de RESIDUELLE1
Normale
Pourcentage
99 95 90 80 70 60 50 40 30 20 10 5 1
-20 -15 -10 -5 0 5 10 15
RESIDUELLE1
Moyenne 1,565826E-14
EcTyp 6,409
N 54
RJ 0,992
Valeur de P >0,100


Diag. valeurs indiv. de PULSE
PULSE
DIET
EXERTYPE
Variable du panneau : INDIV


Graphique des effets principaux pour PULSE
Moyennes ajustées
EXERTYPE
DIET
Moyenne



Modèle linéaire général : PULSE en fonction de EXERTYPE; DIET; INDIV

Facteur	Type	Niveaux	Valeurs
EXERTYPE	fixe	3	1; 2; 3
DIET	fixe	2	1; 2
INDIV(DIET)	aléatoire	18	1; 2; 3; 7; 8; 9; 13; 14; 15; 4; 5; 6; 10; 11; 12; 16; 17; 18

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

Source	DL	SomCar séq	SomCar ajust	CM ajust	F	P
EXERTYPE	2	93972,1	93972,1	46986,1	690,67	0,000
DIET	1	8791,1	8791,1	8791,1	9,65	0,007
EXERTYPE*DIET	2	344,9	344,9	172,5	2,54	0,095
INDIV(DIET)	16	14577,7	14577,7	911,1	13,39	0,000
Erreur	32	2177,0	2177,0	68,0		
Total	53	119862,8				

S = 8,24804 R carré = 98,18 % R carré (ajust) = 96,99 %

Terme		Coeff	Coef ErT	T	P
Constante		137,056	1,122	122,11	0,000
EXERTYPE					
1		-49,556	1,587	-31,22	0,000
2		-2,944	1,587	-1,85	0,073
DIET					
1		12,759	1,122	11,37	0,000
EXERTYPE*DIET					
1	1	-2,926	1,587	-1,84	0,075
2	1	-0,315	1,587	-0,20	0,844
(DIET)INDIV					
1	1	14,519	4,490	3,23	0,003
1	2	17,519	4,490	3,90	0,000
1	3	-13,148	4,490	-2,93	0,006
1	7	31,185	4,490	6,95	0,000
1	8	-20,481	4,490	-4,56	0,000
1	9	-10,148	4,490	-2,26	0,031
1	13	-9,815	4,490	-2,19	0,036
1	14	-16,148	4,490	-3,60	0,001
2	4	14,037	4,490	3,13	0,004
2	5	-15,963	4,490	-3,56	0,001
2	6	-2,630	4,490	-0,59	0,562
2	10	29,370	4,490	6,54	0,000
2	11	1,370	4,490	0,31	0,762
2	12	-0,963	4,490	-0,21	0,832
2	16	-28,296	4,490	-6,30	0,000
2	17	10,037	4,490	2,24	0,032

Observations aberrantes pour PULSE

Observation	PULSE	Valeur ajustée	ErT ajust	Valeur résiduelle	Valeurs résiduelles normalisées	
10	93,000	107,037	5,265	-14,037	-2,21	R
15	88,000	103,852	5,265	-15,852	-2,50	R
46	217,000	202,926	5,265	14,074	2,22	R
52	131,000	145,259	5,265	-14,259	-2,25	R

R indique une observation ayant des valeurs résiduelles normalisées importantes

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

```
EXERTYPE; DIET; EXERTYPE*DIET


Test de l'égalité des variances : RESIDUELLE1 en fonction de EXERTYPE; DIET

Intervalles de confiance de Bonferroni = 95 % pour les écarts types

EXERTYPE  DIET  N  Inférieur    EcTyp  Supérieur
       1     1  9    4,96610  8,31684    20,8251
       1     2  9    4,24148  7,10329    17,7864
       2     1  9    2,52311  4,22551    10,5805
       2     2  9    1,78704  2,99279     7,4938
       3     1  9    4,64848  7,78492    19,4932
       3     2  9    4,81695  8,06704    20,1996


Test de Bartlett (loi normale)
Résultat du test = 10,49 ; valeur p = 0,063


Test de Levene (toute loi de distribution continue)
Résultat du test = 1,86 ; valeur p = 0,119
```

.....

Exercice 2. Des chercheurs ont souhaité evaluer l'effet de l'alcool sur les erreurs de conduite. Ils ont fait conduire cinq sujets, à plusieurs reprises et pour trois quantités différentes d'alcool présentes dans le sang (0,04 ; 0,06 et 0,08 en mg/l). Ces périodes d'essai duraient 15 minutes et se déroulaient dans un simulateur.

Les données obtenues ont été reproduites dans le tableau ci-dessous.

Sujet	Qte_alc_sang	Erreurs
1	0.04	14
2	0.04	16
3	0.04	19
4	0.04	13
5	0.04	11
1	0.06	17
2	0.06	19
3	0.06	21
4	0.06	23
5	0.06	21
1	0.08	25
2	0.08	29
3	0.08	24
4	0.08	21
5	0.08	22

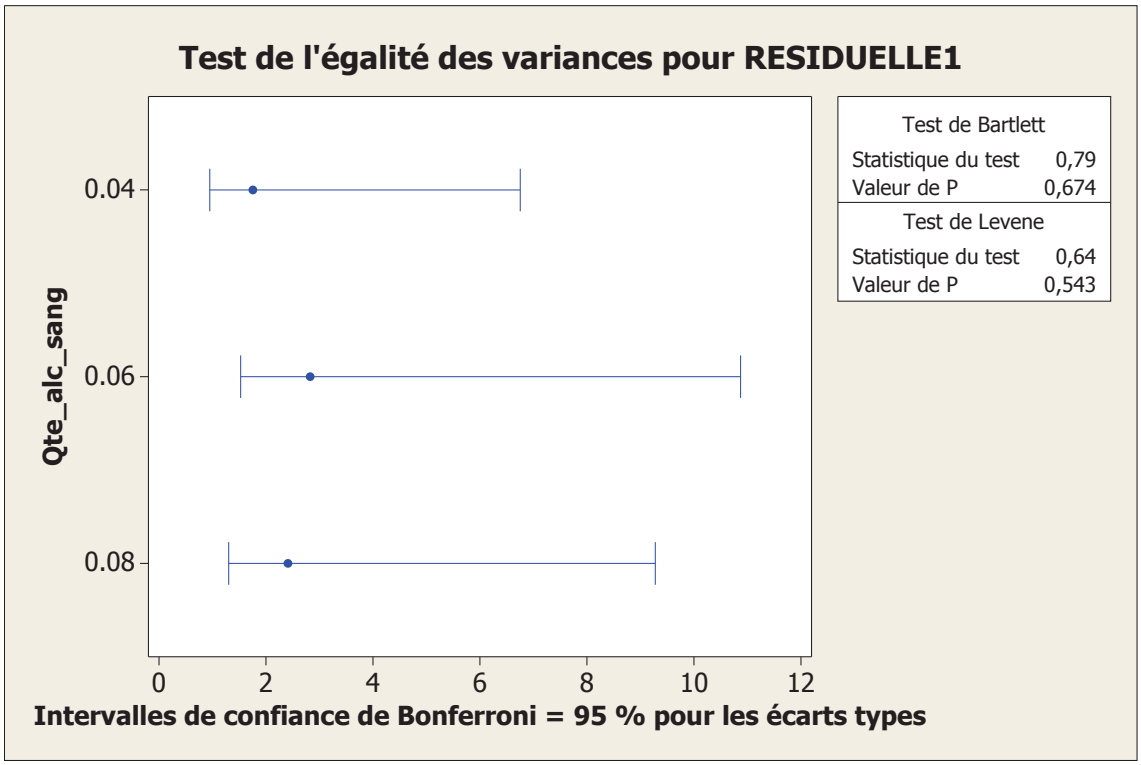
1. Donner le modèle complet de l'analyse de la variance. Donner les effets principaux et interactions qu'il est possible de tester. Les conditions d'utilisation du modèle sont-elles remplies ?
2. Effectuer l'analyse de la variance en indiquant comment sont obtenus les degrés de liberté.
3. Donner, le cas échéant, les effets et les interactions qui ont un effet significatif avec le risque de mauvaise décision associé. Analyser ces effets à l'aide de graphiques et / ou, si cela est possible, des comparaisons multiples.
4. Évaluer, le cas échéant, le risque de mauvaise décision pour les effets principaux et les interactions qui ne sont pas significatifs.
5. Donner un bref commentaire de l'ensemble des résultats.

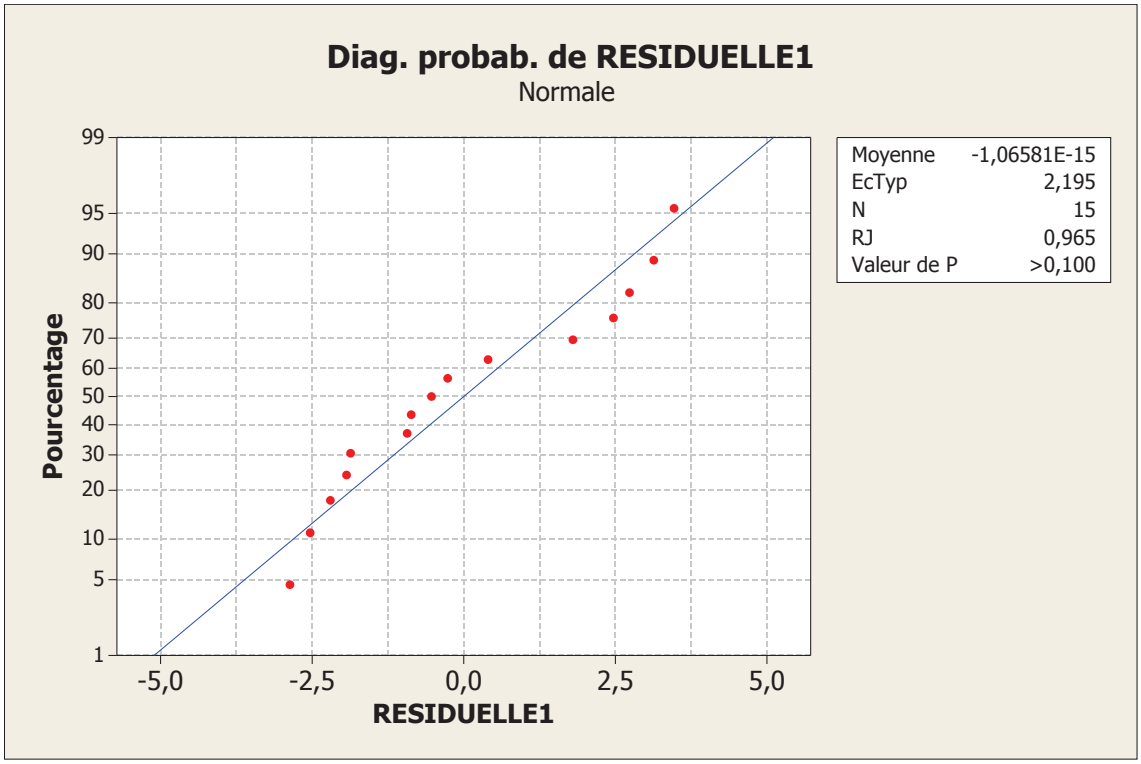
```
Modèle linéaire général : Erreurs en fonction de Qte_alc_sang; Sujet

Facteur       Type       Niveaux  Valeurs
Qte_alc_sang  fixe             3  0.04; 0.06; 0.08
Sujet         aléatoire        5  1; 2; 3; 4; 5


Analyse de la variance pour Erreurs, avec utilisation de la somme des carrés
```

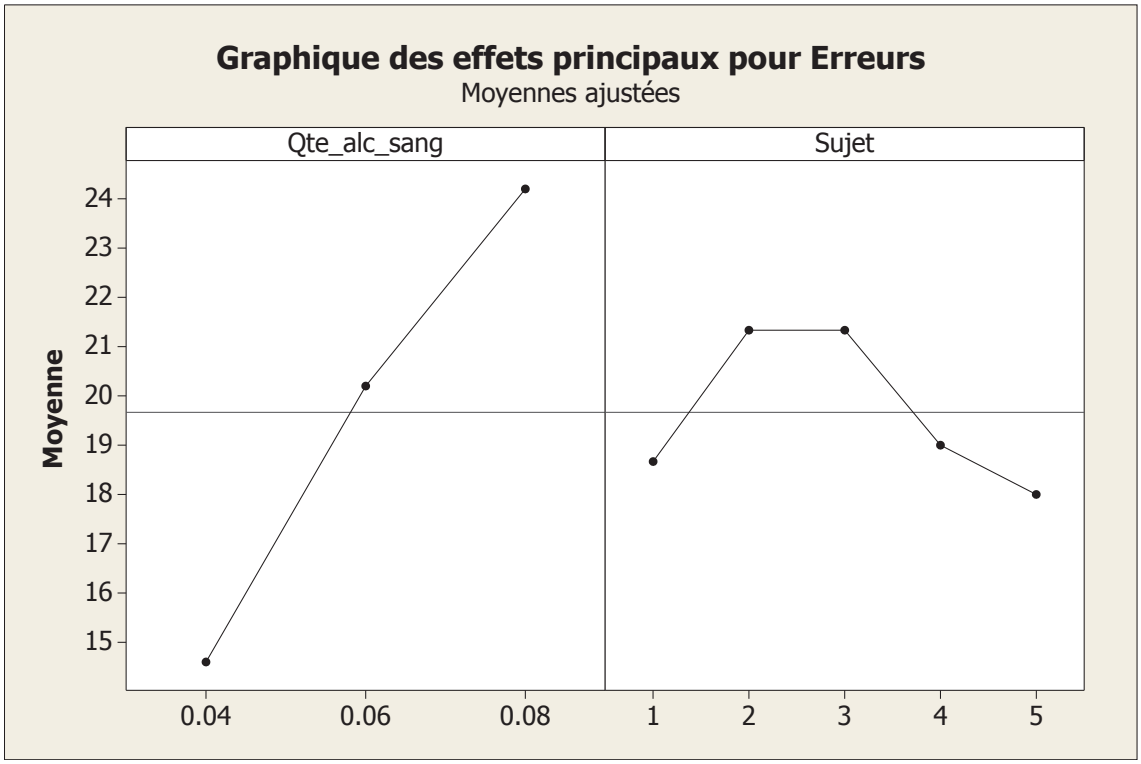
```
     ajustée pour les tests

                                SomCar
Source         DL  SomCar séq    ajust  CM ajust      F      P
Qte_alc_sang    2     232,533  232,533   116,267  13,79  0,003
Sujet           4      29,333   29,333     7,333   0,87  0,522
Erreur          8      67,467   67,467     8,433
Total          14     329,333


S = 2,90402   R carré = 79,51 %   R carré (ajust) = 64,15 %


Terme            Coeff  Coef ErT      T      P
Constante     19,6667    0,7498  26,23  0,000
Qte_alc_sang
0.04           -5,067     1,060  -4,78  0,001
0.06            0,533     1,060   0,50  0,629
Sujet
1              -1,000     1,500  -0,67  0,524
2               1,667     1,500   1,11  0,299
3               1,667     1,500   1,11  0,299
4              -0,667     1,500  -0,44  0,668


Informations de groupement avec la méthode de Tukey et un niveau de confiance
     de 95,0%

Qte_alc_sang  N  Moyenne  Groupement
0.08          5     24,2  A
0.06          5     20,2  A
0.04          5     14,6    B

Les moyennes ne partageant aucune lettre sont sensiblement différentes.


Intervalles de confiance simultanés de Tukey = 95,0 %
Variable de réponse Erreurs
Toutes les comparaisons deux à deux sur les niveaux de Qte_alc_sang
Qte_alc_sang = 0.04  soustrait de :

Qte_alc_sang  Inférieur  Centré  Supérieur
0.06             0,3532   5,600      10,85
0.08             4,3532   9,600      14,85

Qte_alc_sang  --+---------+---------+---------+----
0.06             (---------*----------)
```

```
0.08                    (---------*----------)
              --+---------+---------+---------+----
              0,0       5,0      10,0      15,0


Qte_alc_sang = 0.06  soustrait de :

Qte_alc_sang  Inférieur  Centré  Supérieur
0.08             -1,247   4,000      9,247

Qte_alc_sang  --+---------+---------+---------+----
0.08          (---------*---------)
              --+---------+---------+---------+----
              0,0       5,0      10,0      15,0


Tests de simultanéité de Tukey
Variable de réponse Erreurs
Toutes les comparaisons deux à deux sur les niveaux de Qte_alc_sang
Qte_alc_sang = 0.04  soustrait de :

                            Erreur type
                Différence        de la  Valeur  Valeur de
Qte_alc_sang  des moyennes   différence    de T  p ajustée
0.06                5,600        1,837   3,049     0,0378
0.08                9,600        1,837   5,227     0,0020


Qte_alc_sang = 0.06  soustrait de :

                            Erreur type
                Différence        de la  Valeur  Valeur de
Qte_alc_sang  des moyennes   différence    de T  p ajustée
0.08                4,000        1,837   2,178     0,1352


Test de l'égalité des variances : RESIDUELLE1 en fonction de Qte_alc_sang

Intervalles de confiance de Bonferroni = 95 % pour les écarts types

Qte_alc_sang  N  Inférieur    EcTyp  Supérieur
        0.04  5    0,94811  1,75436     6,7523
        0.06  5    1,52645  2,82450    10,8712
        0.08  5    1,30278  2,41062     9,2782


Test de Bartlett (loi normale)
```

```
Résultat du test = 0,79 ; valeur p = 0,674


Test de Levene (toute loi de distribution continue)
Résultat du test = 0,64 ; valeur p = 0,543
```




Diag. probab. de RESIDUELLE1
Normale
Pourcentage
99
95
90
80
70
60
50
40
30
20
10
5
1
-5,0
-2,5
0,0
2,5
5,0
RESIDUELLE1
Moyenne -1,06581E-15
EcTyp 2,195
N 15
RJ 0,965
Valeur de P >0,100


Graphique des effets principaux pour Erreurs
Moyennes ajustées
Qte_alc_sang
Sujet
Moyenne
24
23
22
21
20
19
18
17
16
15
0.04
0.06
0.08
1
2
3
4
5