# Examen de Statistique Approfondie I

**********************************************************

• *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.** Comparaison de l'évaluation de la résistance d'un même type de ciment

Davies et Goldsmith[1] ont récolté les données d'une expérience dont le but était d'étudier les différentes sources de variabilité possibles de la résistance d'un ciment fabriqué à Portland. On note $Y$ la variable associée à la résistance du ciment.

L'expérience s'est déroulée ainsi : plusieurs petits prélèvements d'un même type de ciment ont été mélangés à de l'eau et travaillés par trois personnes différentes, les « mélangeurs ». On a alors formé douze cubes à l'aide de chacune des préparations des « mélangeurs ». Puis on a donné ces 36 cubes à trois personnes chargées d'évaluer leur résistance, les « casseurs ». La répartition des 36 cubes entre ces « casseurs » a été faite de telle sorte que chaque « casseur » reçoive quatre cubes provenant de chacune des préparations des « mélangeurs » soit douze cubes au total.

[1] **Davies, O.L. et Goldsmith, P.L. (Eds.)**, *Statistical Methods in Research and Production*, $4^{\text{th}}$ edition, Oliver and Boyd, Edinburgh, 1972.

Tous les tests de résistance ont été faits sur la même machine. L'objectif principal de cette expérience était d'étudier et de quantifier l'importance de la variabilité dans les tests de résistance qui pouvait provenir des différences individuelles entre les « mélangeurs » et les « casseurs ». Les données ci-dessous, exprimées dans les unités d'origine c'est-à-dire en livres par pouces carrés, ont été recopiées dans le tableau ci dessous.

| | « Casseur » 1 | | « Casseur » 2 | | « Casseur » 3 | |
|---|---|---|---|---|---|---|
| « Mélangeur » 1 | 5280 | 5520 | 4340 | 4400 | 4160 | 5180 |
| | 4760 | 5800 | 5020 | 6200 | 5320 | 4600 |
| « Mélangeur » 2 | 4420 | 5280 | 5340 | 4880 | 4180 | 4800 |
| | 5580 | 4900 | 4960 | 6200 | 4600 | 4480 |
| « Mélangeur » 3 | 5360 | 6160 | 5720 | 4760 | 4460 | 4930 |
| | 5680 | 5500 | 5620 | 5560 | 4680 | 5600 |

**Partie I :**

Dans l'expérience d'origine, on ne s'intéressait qu'aux différences ne pouvant être dues qu'à ces trois « mélangeurs » et à ces trois « casseurs ».

**1.** Écrire le modèle d'analyse de la variance relatif à cette étude. On précisera la nature des facteurs explicatifs ainsi que les hypothèses faites.

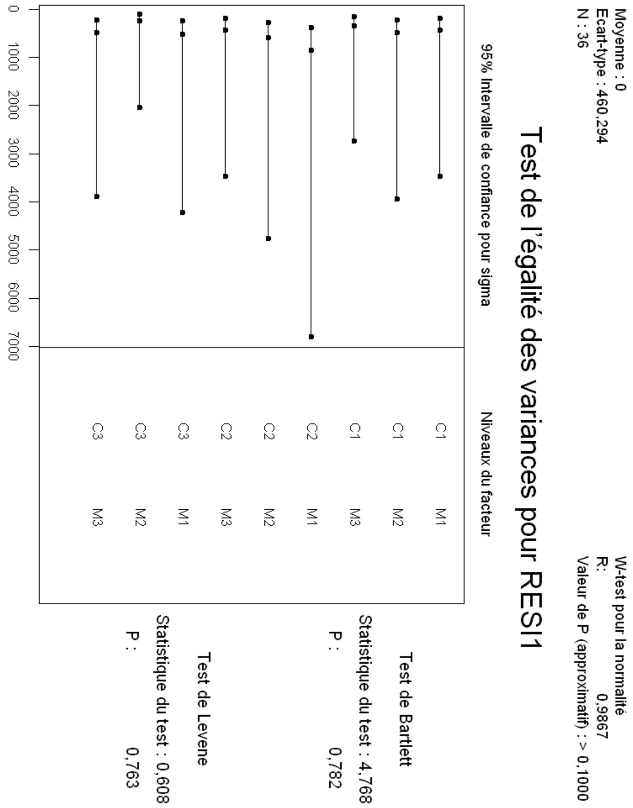
**2.** Les hypothèses du modèle sont-elles vérifiées ? Calculer les estimations de tous les paramètres du modèle.

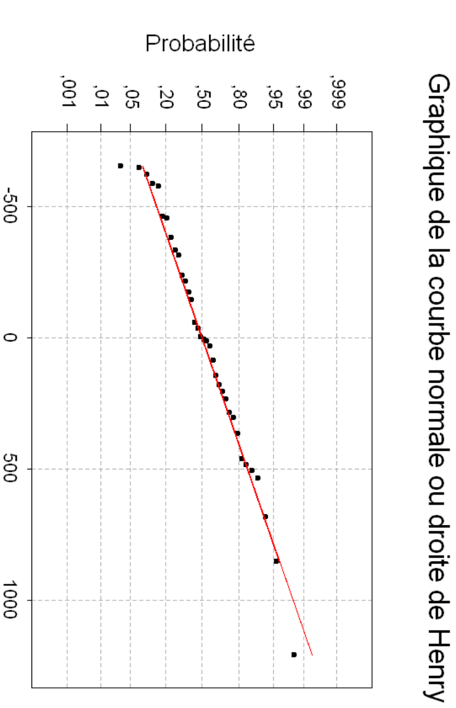
**3.** Existe-t-il une interaction dans l'évaluation de la résistance du ciment entre les « mélangeurs » et les « casseurs » ?

**4.** Existe-t-il des différences dans l'évaluation de la résistance dues aux « mélangeurs » ?

**5.** Existe-t-il des différences dans l'évaluation de la résistance dues aux « casseurs » ?

**6.** Compte tenu de la nature des deux 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.



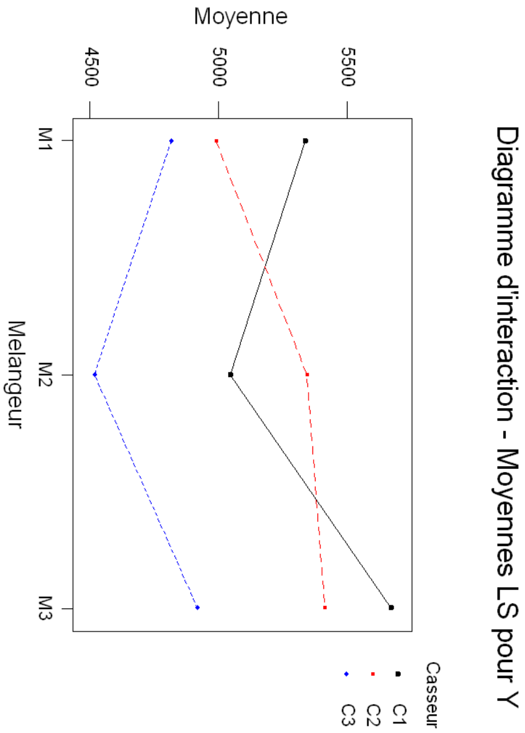

```
Modèle linéaire généralisé : Y en fonction de Casseur; Melangeur


Facteur      Type Niveaux  Valeurs
Casseur      fixe       3  C1 C2 C3
Melangeu     fixe       3  M1 M2 M3

Analyse de la variance pour Y, en utilisant la SC ajustée pour les tests

Source             DL     SC séq    SC ajust    CM ajust       F       P
Casseur             2    2506117     2506117     1253058    4,56   0,020
Melangeu            2     896450      896450      448225    1,63   0,214
Casseur*Melangeu    4     663833      663833      165958    0,60   0,663
Erreur             27    7415475     7415475      274647
Total              35   11481875

Terme                  Coef      Er-T coef        T       P
Constante           5117,50          87,34    58,59   0,000
Casseur
C1                    235,8          123,5     1,91   0,067
C2                    132,5          123,5     1,07   0,293
Melangeu
```

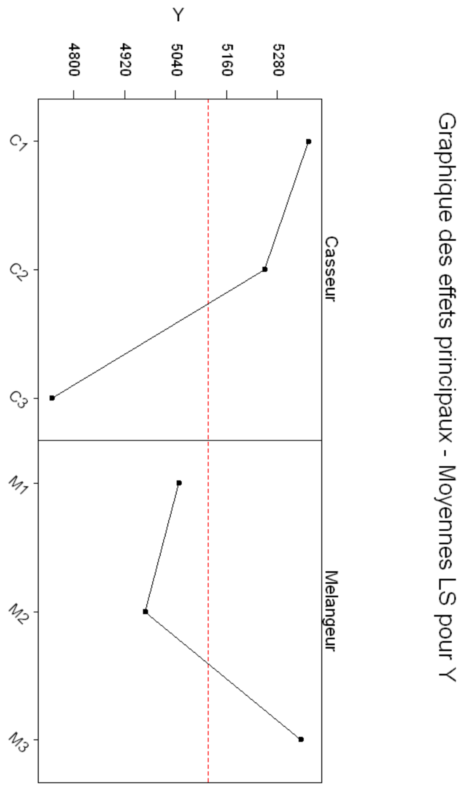
| | | | | | |
|---|---|---|---|---|---|
| M1 | | -69,2 | 123,5 | -0,56 | 0,580 |
| M2 | | -149,2 | 123,5 | -1,21 | 0,238 |
| Casseur*Melangeu | | | | | |
| C1 | M1 | 55,8 | 174,7 | 0,32 | 0,752 |
| C1 | M2 | -159,2 | 174,7 | -0,91 | 0,370 |
| C2 | M1 | -190,8 | 174,7 | -1,09 | 0,284 |
| C2 | M2 | 244,2 | 174,7 | 1,40 | 0,174 |

Observations aberrantes pour Y

| Obs | Y | Ajust | Er-T ajust | Val résid | Val résid norm |
|---|---|---|---|---|---|
| 20 | 6200,00 | 4990,00 | 262,03 | 1210,00 | 2,67R |

R indique une observation avec une valeur résiduelle normalisée importante.




Tests de simultanéité de Tukey
Variable de réponse Y
Toutes comparaisons deux à deux entre niveaux de Casseur

Casseur = C1 soustraites de :

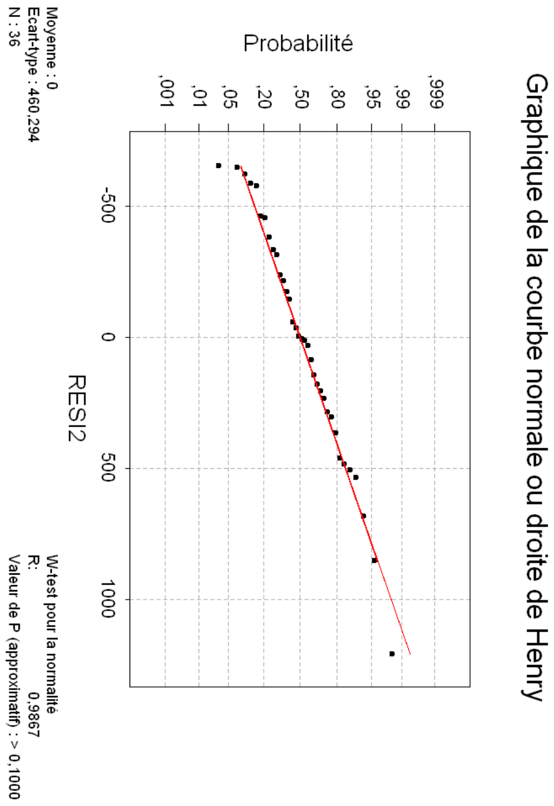
| Niveau Casseur | Différence des moyennes | Er-T de la différence | Valeur de T | Valeur ajustée de P |
|---|---|---|---|---|

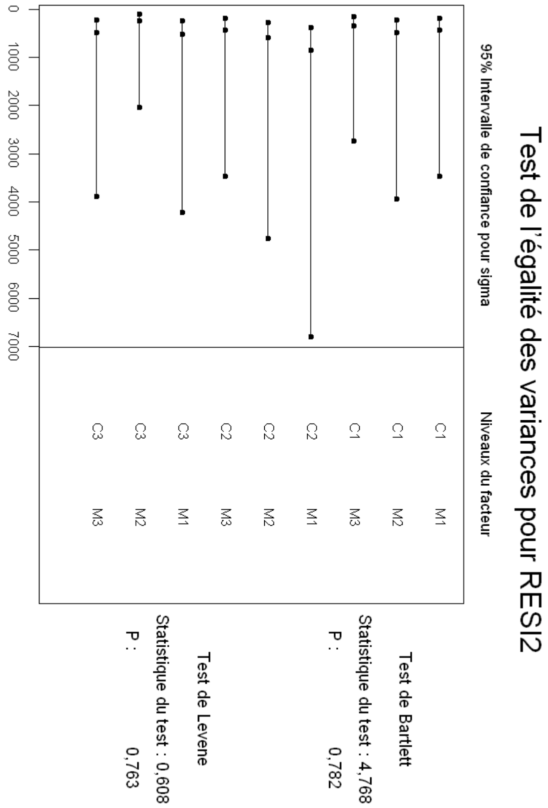
```
C2                  -103,3            213,9     -0,483             0,8799
C3                  -604,2            213,9     -2,824             0,0232

Casseur = C2 soustraites de :

Niveau          Différence          Er-T de     Valeur   Valeur ajustée
Casseur        des moyennes   la différence       de T             de P
C3                  -500,8            213,9     -2,341             0,0670

Tests de simultanéité de Tukey
Variable de réponse Y
Toutes comparaisons deux à deux entre niveaux de Melangeu

Melangeu = M1 soustraites de :

Niveau          Différence          Er-T de     Valeur   Valeur ajustée
Melangeu       des moyennes   la différence       de T             de P
M2                  -80,00            213,9    -0,3739             0,9260
M3                  287,50            213,9     1,3438             0,3840

Melangeu = M2 soustraites de :

Niveau          Différence          Er-T de     Valeur   Valeur ajustée
Melangeu       des moyennes   la différence       de T             de P
M3                   367,5            213,9      1,718             0,2171
```

**Partie II :**

On souhaite désormais que les résultats de cette expérience ait une portée plus générale que celle de la première partie : on ne restreint plus l'étude à cette population de « casseurs » et de « mélangeurs ».

6. Écrire le modèle d'analyse de la variance relatif à cette étude. On précisera la nature des facteurs explicatifs ainsi que les hypothèses faites.
7. Les hypothèses du modèle sont-elles vérifiées ? Calculer les estimations de tous les paramètres du modèle.
8. Existe-t-il une interaction dans l'évaluation de la résistance du ciment entre les « mélangeurs » et les « casseurs » ?
9. Existe-t-il des différences dans l'évaluation de la résistance dues aux « mélangeurs » ?
10. Existe-t-il des différences dans l'évaluation de la résistance dues aux « casseurs » ?




Modèle linéaire généralisé : Y en fonction de Casseur; Melangeur

| Facteur | Type | Niveaux | Valeurs |
|---|---|---|---|
| Casseur | aléatoire | 3 | C1 C2 C3 |
| Melangeu | aléatoire | 3 | M1 M2 M3 |

Analyse de la variance pour Y, en utilisant la SC ajustée pour les tests

| Source | DL | SC séq | SC ajust | CM ajust | F | P |
|---|---|---|---|---|---|---|
| Casseur | 2 | 2506117 | 2506117 | 1253058 | 7,55 | 0,044 |
| Melangeu | 2 | 896450 | 896450 | 448225 | 2,70 | 0,181 |
| Casseur*Melangeu | 4 | 663833 | 663833 | 165958 | 0,60 | 0,663 |
| Erreur | 27 | 7415475 | 7415475 | 274647 | | |
| Total | 35 | 11481875 | | | | |

Observations aberrantes pour Y

| Obs | Y | Ajust | Er-T ajust | Val résid | Val résid norm |
|---|---|---|---|---|---|
| 20 | 6200,00 | 4990,00 | 262,03 | 1210,00 | 2,67R |

R indique une observation avec une valeur résiduelle normalisée importante.

. . . . . . . . . . . . . . .

**Exercice 2.** Insertion professionnelle et niveau d'étude[2]

Dans le cadre d'une étude régionale sur le chômage, on a sélectionné 12 diplômés issus d'une filière universitaire professionnalisée $U$ et 12 titulaires d'un baccalauréat professionnel $P$.

On a recencé ci-dessous le temps, exprimé en semaines, qui leur a été nécessaire pour avoir un emploi après l'obtention de leur diplôme et éventuellement déduction faite de la période de temps correspondant à la durée du service militaire.

On se demande si l'on peut considérer que le temps $Y$ mis par le titulaire d'un bac professionnel est significativement supérieur à celui $X$ mis par un diplômé de la filière universitaire.

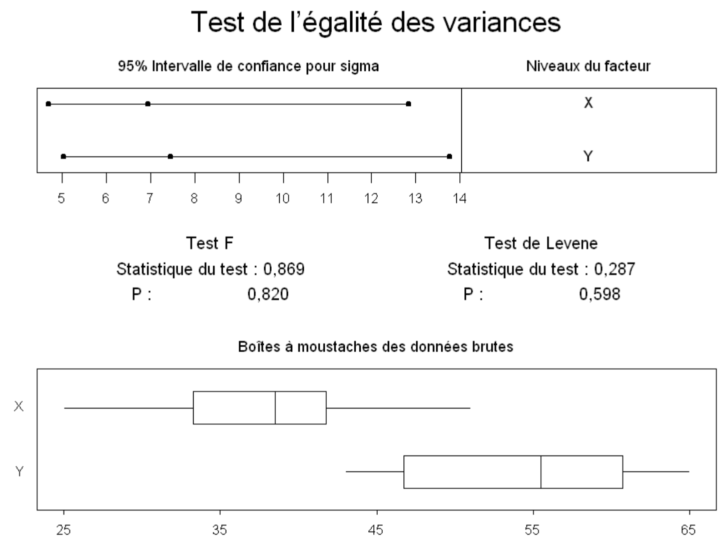
| Diplômé $U$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Durée $X_i$ | 38 | 35 | 40 | 45 | 51 | 41 | 42 | 39 | 34 | 33 | 30 | 25 |
| Rang $r_i$ | 6 | 5 | 8 | ? | 17 | 9 | ? | 7 | 4 | 3 | 2 | 1 |
| Diplômé $P$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Durée $Y_j$ | 49 | 50 | 57 | 61 | 44 | 55 | 60 | 43 | 46 | 56 | 62 | 65 |
| Rang $s_j$ | 15 | 16 | 20 | 22 | ? | 18 | 21 | ? | 14 | 19 | 23 | 24 |

**1.** Quel est le test, ci-dessous, qui a été employé pour étudier ces données ? Rappeler brièvement les hypothèses et la procédure de ce test.

**2.** Expliquer pourquoi il est intéressant dans cette situation de réaliser soit-même ce test. Le faire et conclure quant à la question posée.

**3.** Conclure à partir des résultats précédents et de ceux reproduits ci-dessous.

[2] Cet exercice est tiré du livre de G. Pupion et P.-C. Pupion, *Méthodes statistiques non paramétriques.*




```
Test de Mann-Whitney et IC : X; Y

X           N =  12     Médiane =       38,50
Y           N =  12     Médiane =       55,50
L'estimation ponctuelle pour ETA1-ETA2 est      -16,00
95,4 Le pourcentage IC pour ETA1-ETA2 est (-23,00;-10,00)
W = 85,0
Test de ETA1 = ETA2  fn de  ETA1 non = ETA2 est significatif à 0,0002
```

...............