Frédéric Bertrand
Université de Strasbourg

Examen de Statistique Appliquée 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*, 4th 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.

	« Mélangeur » 1		« Mélangeur » 2		« Mélangeur » 3	
« Casseur » 1	4420	5520	4340	4400	4160	5180
	5280	5800	5020	6200	5320	4600
« Casseur » 2	4760	5280	5340	4880	4180	4800
	5580	4900	4960	6200	4600	4480
« Casseur » 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.

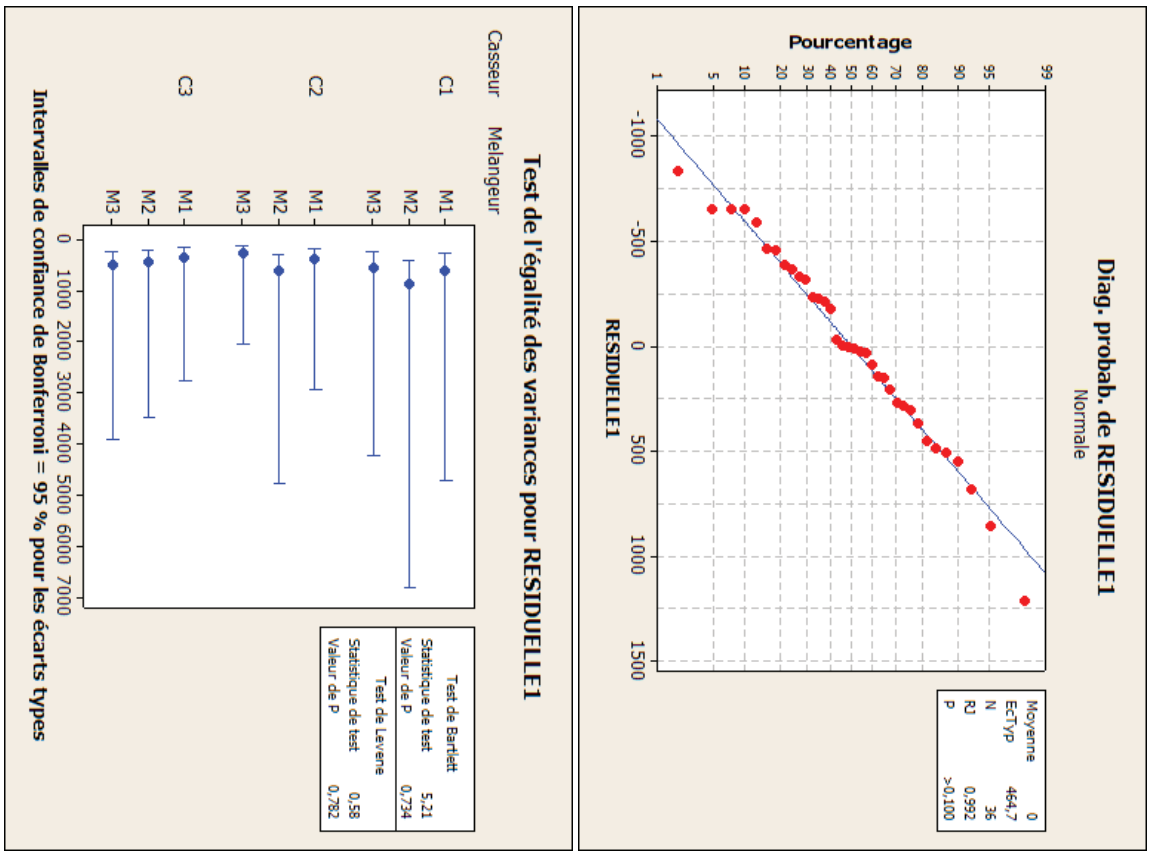
2. Les hypothèses du modèle sont-elles vérifiées ? Calculer les estimations de tous les paramètres du modèle.

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 général : Y en fonction de Casseur; Melangeur

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

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

Source	DL	SomCar séq	SomCar ajust	CM ajust	F	P
Casseur	2	861317	861317	430658	1,54	0,233
Melangeur	2	2506117	2506117	1253058	4,48	0,021
Casseur*Melangeur	4	556167	556167	139042	0,50	0,738
Erreur	27	7558275	7558275	279936		
Total	35	11481875				

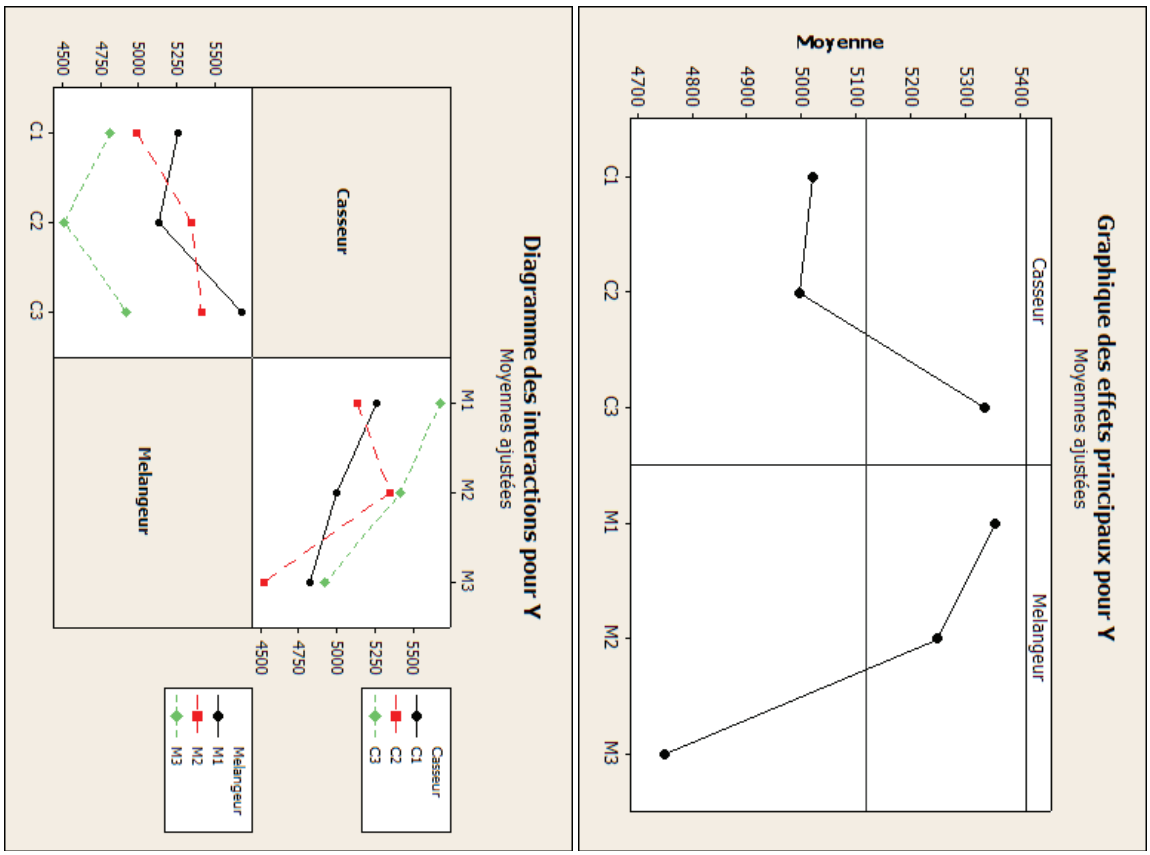
S = 529,090 R carré = 34,17 % R carré (ajust) = 14,67 %

Terme		Coeff	Coef ErT	T	P
Constante		5117,50	88,18	58,03	0,000
Casseur					
C1		-97,5	124,7	-0,78	0,441
C2		-120,8	124,7	-0,97	0,341
Melangeur					
M1		235,8	124,7	1,89	0,069
M2		132,5	124,7	1,06	0,297
Casseur*Melangeur					
C1	M1	-0,8	176,4	-0,00	0,996
C1	M2	-162,5	176,4	-0,92	0,365
C2	M1	-102,5	176,4	-0,58	0,566
C2	M2	215,8	176,4	1,22	0,232

Observations aberrantes pour Y

Observation	Y	Valeur ajustée	Ajust ErT	Valeur résiduelle	Valeur résiduelle normalisée	
20	6200,00	4990,00	264,54	1210,00	2,64	R

R indique une observation ayant une valeur résiduelle normalisée importante

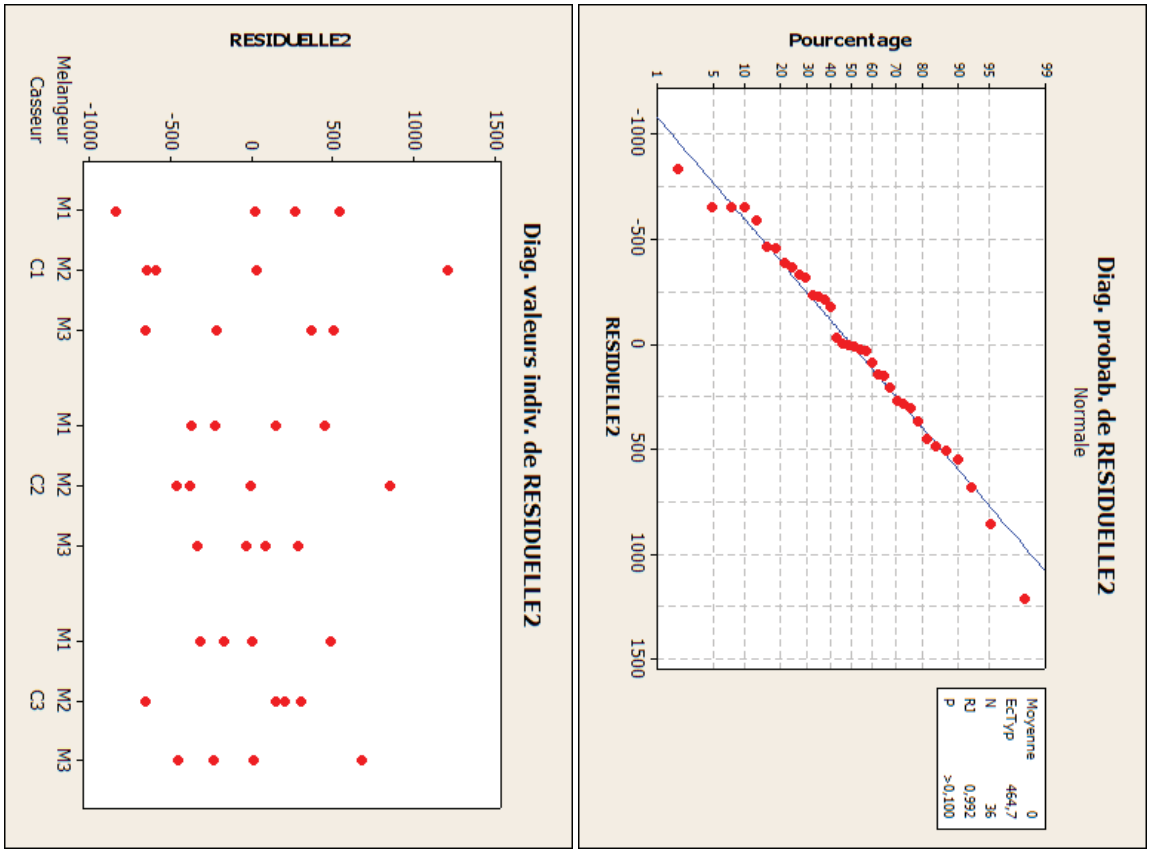


```
Tests de simultanéité de Tukey
Variable de réponse Y
Toutes les comparaisons deux à deux sur les niveaux de Casseur
Casseur = C1  soustrait de :

                      Erreur type
          Différence        de la                Valeur de
Casseur  des moyennes   différence  Valeur de T  p ajustée
C2            -23,33        216,0      -0,1080     0,9936
C3            315,83        216,0       1,4622     0,3245


Casseur = C2  soustrait de :

                      Erreur type
          Différence        de la  Valeur  Valeur de
Casseur  des moyennes   différence    de T  p ajustée
C3             339,2        216,0   1,570     0,2755


Intervalles de confiance simultanés de Tukey = 95,0 %
Variable de réponse Y
Toutes les comparaisons deux à deux sur les niveaux de Melangeur
Melangeur = M1  soustrait de :


Tests de simultanéité de Tukey
Variable de réponse Y
Toutes les comparaisons deux à deux sur les niveaux de Melangeur
Melangeur = M1  soustrait de :

                        Erreur type
            Différence        de la  Valeur  Valeur de
Melangeur  des moyennes   différence    de T  p ajustée
M2              -103,3        216,0  -0,478     0,8820
M3              -604,2        216,0  -2,797     0,0247


Melangeur = M2  soustrait de :

                        Erreur type
            Différence        de la  Valeur  Valeur de
Melangeur  des moyennes   différence    de T  p ajustée
M3              -500,8        216,0  -2,319     0,0702
```

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 général : Y en fonction de Casseur; Melangeur

Facteur	Type	Niveaux	Valeurs
Casseur	aléatoire	3	C1; C2; C3
Melangeur	aléatoire	3	M1; M2; M3

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

Source	DL	SomCar séq	SomCar ajust	CM ajust	F	P
Casseur	2	861317	861317	430658	3,10	0,154
Melangeur	2	2506117	2506117	1253058	9,01	0,033
Casseur*Melangeur	4	556167	556167	139042	0,50	0,738
Erreur	27	7558275	7558275	279936		
Total	35	11481875				

S = 529,090 R carré = 34,17 % R carré (ajust) = 14,67 %

Terme		Coeff	Coef ErT	T	P
Constante		5117,50	88,18	58,03	0,000
Casseur					
C1		-97,5	124,7	-0,78	0,441
C2		-120,8	124,7	-0,97	0,341
Melangeur					
M1		235,8	124,7	1,89	0,069
M2		132,5	124,7	1,06	0,297
Casseur*Melangeur					
C1	M1	-0,8	176,4	-0,00	0,996
C1	M2	-162,5	176,4	-0,92	0,365
C2	M1	-102,5	176,4	-0,58	0,566
C2	M2	215,8	176,4	1,22	0,232

Observations aberrantes pour Y

Observation	Y	Valeur ajustée	Ajust ErT	Valeur résiduelle	Valeur résiduelle normalisée	
20	6200,00	4990,00	264,54	1210,00	2,64	R

R indique une observation ayant une valeur résiduelle normalisée importante

Espérance mathématique des carrés moyens, en utilisant la somme des carrés ajustée

	Source	Espérance mathématique des carrés moyens pour chaque terme
1	Casseur	(4) + 4,0000 (3) + 12,0000 (1)
2	Melangeur	(4) + 4,0000 (3) + 12,0000 (2)
3	Casseur*Melangeur	(4) + 4,0000 (3)
4	Erreur	(4)

Termes d'erreur pour les tests, en utilisant la somme des carrés ajustée

	Source	Erreur - DL	Erreur - CM	Synthèse du carré moyen de l'erreur
1	Casseur	4,00	139042	(3)
2	Melangeur	4,00	139042	(3)
3	Casseur*Melangeur	27,00	279936	(4)

Composantes de la variance, avec la somme des carrés ajustée

Source	Valeur estimée
Casseur	24301
Melangeur	92835
Casseur*Melangeur	-35224
Erreur	279936

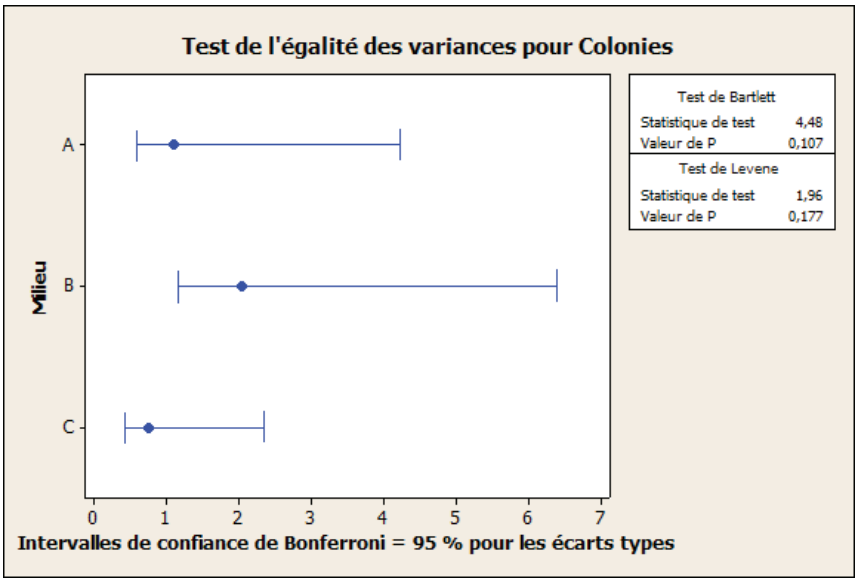
.....

Exercice 2. Milieux de culture et colonies.

Des chercheurs souhaitent comparer trois milieux de culture différents A B et C. Pour cela ils ont compté le nombre de colonies bactériennes dans chaque milieu après deux jours. Les résultats ont été reportés dans le tableau suivant.

Milieu	Colonies	Milieu	Colonies	Milieu	Colonies
A	7	B	5	C	6
A	4	B	4	C	7
A	3	B	4	C	6
A	2	B	1	C	5
A	4	B	3	C	7
		B	5	C	6

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. Justifier l'utilisation de ce test plutôt que celle d'un modèle d'analyse de la variance.
2. Conclure à partir des résultats reproduits ci-dessous.
3. Retrouver la valeur de la statistique du test en la calculant vous-même.




```
Test de Kruskal-Wallis : Colonies en fonction de Milieu

Test de Kruskal-Wallis sur Colonies

                        Rang
Milieu   N  Médiane    moyen       Z
A        5    4,000      6,3   -1,42
B        6    3,500      6,8   -1,36
C        6    6,000     13,5    2,71
```

```
Global  17            9,0

B H = 7,39  DL = 2  P = 0,025
H = 7,60  DL = 2  P = 0,022  (ajusté pour les nombres de même grandeur)
```

.....