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. Activité d'une enzyme

L'étude de l'activité d'une enzyme a été réalisée chez des sujets jeunes en fonction de l'âge et du sexe.

48 essais ont été réalisés dans des conditions similaires sur 48 individus distincts. Leurs résultats, exprimés en unités arbitraires, ont été reproduits dans le tableau ci-dessous.

Age	Sexe	Enzyme	Age	Sexe	Enzyme
moins de 12 ans	garçon	4,9	plus de 12 ans	garçon	2,1
moins de 12 ans	garçon	2,9	plus de 12 ans	garçon	2,2
moins de 12 ans	garçon	2,7	plus de 12 ans	garçon	1,1
moins de 12 ans	garçon	3,9	plus de 12 ans	garçon	2,9
moins de 12 ans	garçon	4,6	plus de 12 ans	garçon	5,0
moins de 12 ans	garçon	3,3	plus de 12 ans	garçon	3,5
moins de 12 ans	garçon	5,9	plus de 12 ans	garçon	2,4
moins de 12 ans	garçon	4,8	plus de 12 ans	garçon	4,4
moins de 12 ans	garçon	4,1	plus de 12 ans	garçon	2,1
moins de 12 ans	garçon	3,5	plus de 12 ans	garçon	3,0
moins de 12 ans	garçon	7,2	plus de 12 ans	garçon	3,9
moins de 12 ans	garçon	6,1	plus de 12 ans	garçon	5,6
moins de 12 ans	fille	4,5	plus de 12 ans	fille	2,4
moins de 12 ans	fille	6,9	plus de 12 ans	fille	3,6
moins de 12 ans	fille	4,0	plus de 12 ans	fille	4,8
moins de 12 ans	fille	5,4	plus de 12 ans	fille	3,9
moins de 12 ans	fille	1,9	plus de 12 ans	fille	5,5
moins de 12 ans	fille	3,6	plus de 12 ans	fille	5,0
moins de 12 ans	fille	4,8	plus de 12 ans	fille	6,8
moins de 12 ans	fille	3,3	plus de 12 ans	fille	2,2
moins de 12 ans	fille	7,5	plus de 12 ans	fille	3,1
moins de 12 ans	fille	5,8	plus de 12 ans	fille	5,0
moins de 12 ans	fille	4,4	plus de 12 ans	fille	4,1
moins de 12 ans	fille	6,0	plus de 12 ans	fille	4,7

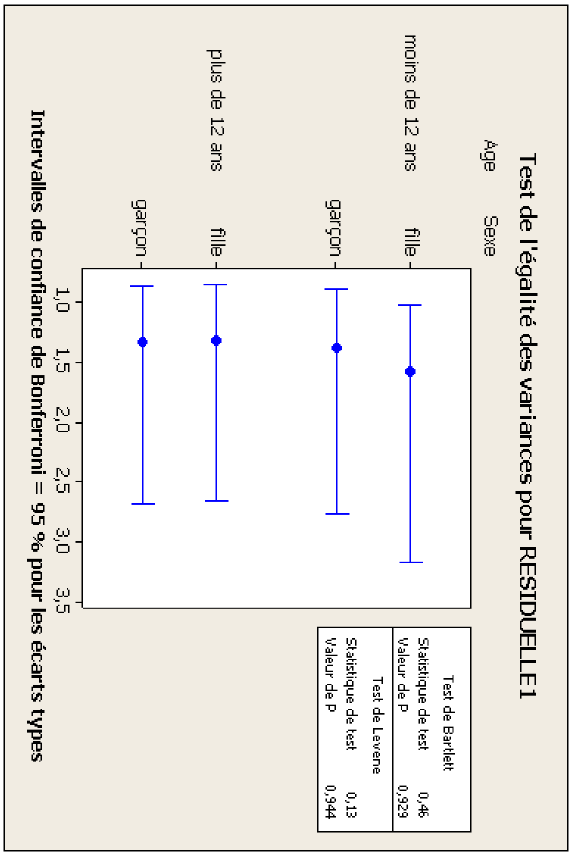
Partie I

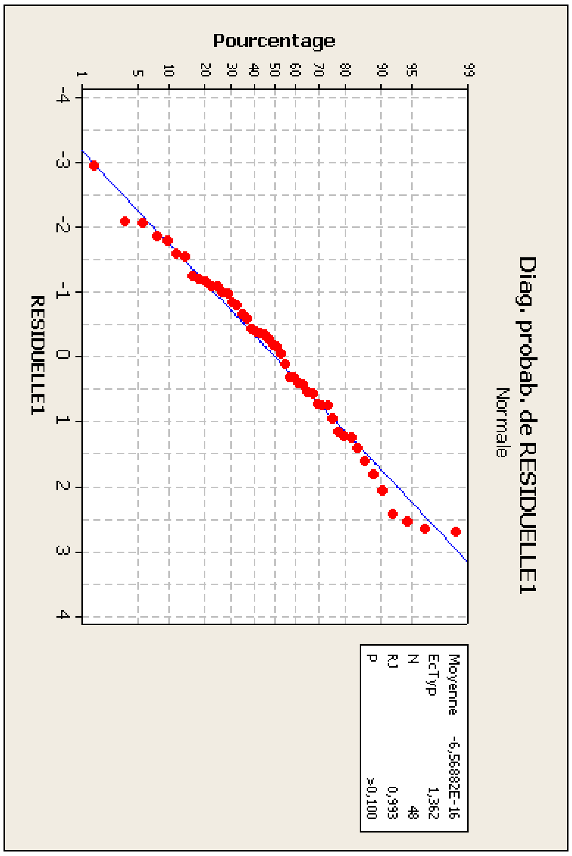
I.1. Écrire le modèle d'analyse de la variance relatif au modèle dont les résultats sont reproduits dans la suite de la **partie I**. On précisera la nature des facteurs explicatifs ainsi que les hypothèses faites.

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

I.3. Procéder à l'étude du tableau de l'analyse de la variance.

I.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, si oui procéder aux tests correspondants et interpréter les résultats.





Modèle linéaire général : Enzyme en fonction de Age; Sexe

Facteur	Type	Niveaux	Valeurs
Age	fixe	2	moins de 12 ans; plus de 12 ans
Sexe	fixe	2	fille; garçon

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

Source	DL	SomCar séq	SomCar ajust	CM ajust	F	P
Age	1	10,735	10,735	10,735	5,42	0,025
Sexe	1	6,092	6,092	6,092	3,08	0,086
Age*Sexe	1	1,577	1,577	1,577	0,80	0,377
Erreur	44	87,124	87,124	1,980		
Total	47	105,528				

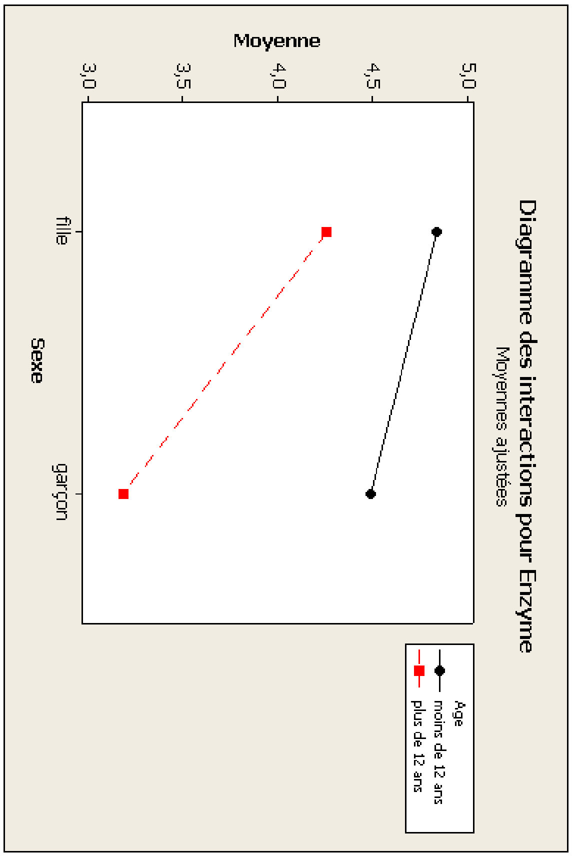
S = 1,40716   R carré = 17,44 %   R carré (ajust) = 11,81 %

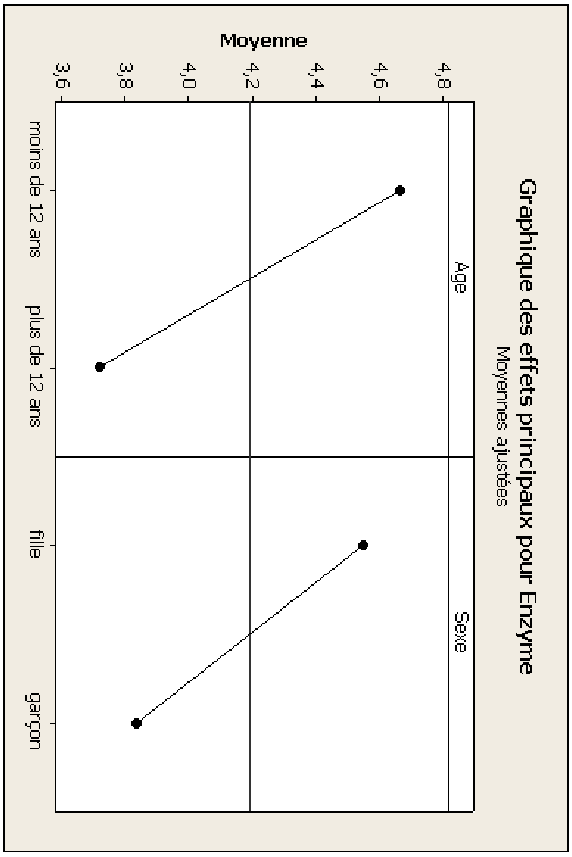
Terme	Coeff	Coef ErT	T	P
Constante	4,1938	0,2031	20,65	0,000
Age				
moins de 12 ans	0,4729	0,2031	2,33	0,025
Sexe				
fille	0,3562	0,2031	1,75	0,086
Age*Sexe				
moins de 12 ans fille	-0,1812	0,2031	-0,89	0,377

Observations aberrantes pour Enzyme

Observation	Enzyme	Valeur ajustée	Ajust ErT	Valeur résiduelle	Valeur résiduelle normalisée	
11	7,20000	4,49167	0,40621	2,70833	2,01	R
17	1,90000	4,84167	0,40621	-2,94167	-2,18	R

R indique une observation ayant une valeur résiduelle normalisée importante





Tests de simultanéité de Tukey
Variable de réponse Enzyme

Toutes les comparaisons deux à deux sur les niveaux de Age
Age = moins de 12 ans soustrait de :

Age	Différence des moyennes	Erreur type de la différence	Valeur de T	Valeur de p ajustée
plus de 12 ans	-0,9458	0,4062	-2,328	0,0246

Tests de simultanéité de Tukey
Variable de réponse Enzyme
Toutes les comparaisons deux à deux sur les niveaux de Sexe
Sexe = fille soustrait de :

Sexe	Différence des moyennes	Erreur type de la différence	Valeur de T	Valeur de p ajustée
garçon	-0,7125	0,4062	-1,754	0,0864

Tests de simultanéité de Tukey
Variable de réponse Enzyme
Toutes les comparaisons deux à deux sur les niveaux de Age*Sexe
Age = moins de 12 ans
Sexe = fille soustrait de :

Age	Sexe	Différence des moyennes	Erreur type de la différence	Valeur de T	Valeur de p ajustée
moins de 12 ans	garçon	-0,350	0,5745	-0,609	0,9286
plus de 12 ans	fille	-0,583	0,5745	-1,015	0,7413
plus de 12 ans	garçon	-1,658	0,5745	-2,887	0,0296

Age = moins de 12 ans
Sexe = garçon soustrait de :

Age	Sexe	Différence des moyennes	Erreur type de la différence	Valeur de T	Valeur de p ajustée
plus de 12 ans	fille	-0,233	0,5745	-0,406	0,9771
plus de 12 ans	garçon	-1,308	0,5745	-2,277	0,1190

Age = plus de 12 ans
Sexe = fille soustrait de :

Age	Sexe	Différence des moyennes	Erreur type de la différence	Valeur de T	Valeur de p ajustée
plus de 12 ans	garçon	-1,075	0,5745	-1,871	0,2551

Partie II

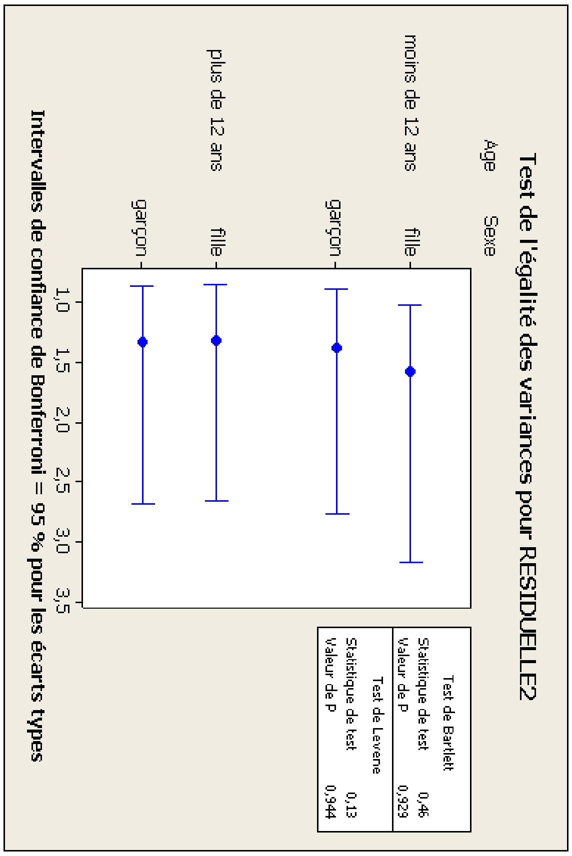
II.1. Écrire le modèle d'analyse de la variance relatif au modèle dont les résultats sont reproduits ci-dessous. On précisera la nature des facteurs explicatifs ainsi que les hypothèses faites.

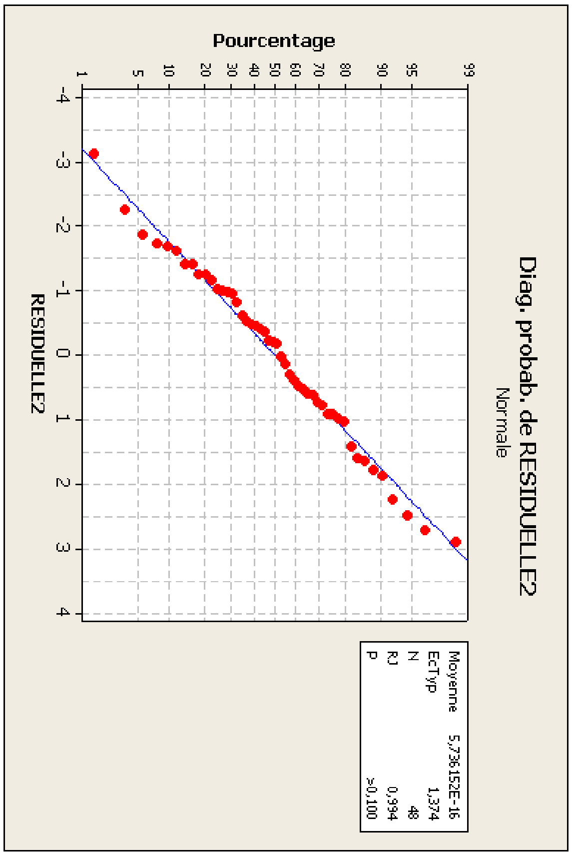
II.2. Les hypothèses du modèle sont-elles vérifiées ? Calculer les estimations de tous les paramètres du modèle.

II.3. Procéder à l'étude du tableau de l'analyse de la variance.

II.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, si oui procéder aux tests correspondants et interpréter les résultats.

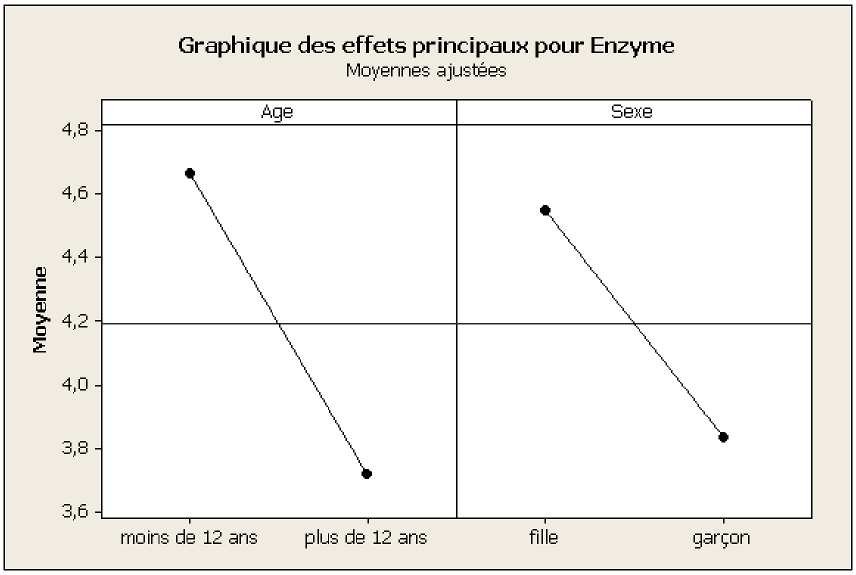
II.5. En comparant les résultats obtenus dans cette seconde partie avec ceux de la première partie, que constatez-vous ?






```
Modèle linéaire général : Enzyme en fonction de Age; Sexe

Facteur  Type  Niveaux  Valeurs
Age      fixe        2  moins de 12 ans; plus de 12 ans
Sexe     fixe        2  fille; garçon
```

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

                         SomCar
Source  DL  SomCar séq    ajust  CM ajust     F      P
Age      1      10,735   10,735    10,735  5,45  0,024
Sexe     1       6,092    6,092     6,092  3,09  0,086
Erreur  45      88,701   88,701     1,971
Total   47     105,528


S = 1,40397   R carré = 15,95 %   R carré (ajust) = 12,21 %


Terme             Coeff  Coef ErT      T      P
Constante        4,1937    0,2026  20,69  0,000
Age
moins de 12 ans  0,4729    0,2026   2,33  0,024
Sexe
fille            0,3562    0,2026   1,76  0,086


Observations aberrantes pour Enzyme

                                                           Valeur
                        Valeur                   Valeur  résiduelle
Observation   Enzyme  ajustée  Ajust ErT  résiduelle  normalisée
         11  7,20000  4,31042    0,35099     2,88958        2,13 R
         17  1,90000  5,02292    0,35099    -3,12292       -2,30 R
         43  6,80000  4,07708    0,35099     2,72292        2,00 R

R indique une observation ayant une valeur résiduelle normalisée
importante
```




```
Tests de simultanéité de Tukey
Variable de réponse Enzyme
Toutes les comparaisons deux à deux sur les niveaux de Age
Age = moins de 12 ans  soustrait de :

                              Erreur type
                Différence          de la  Valeur  Valeur de
Age             des moyennes   différence    de T  p ajustée
plus de 12 ans       -0,9458       0,4053  -2,334     0,0241

Tests de simultanéité de Tukey
Variable de réponse Enzyme
Toutes les comparaisons deux à deux sur les niveaux de Sexe
Sexe = fille  soustrait de :

                      Erreur type
        Différence          de la  Valeur  Valeur de
Sexe    des moyennes   différence    de T  p ajustée
garçon       -0,7125       0,4053  -1,758     0,0856
```

....

Exercice 2. Dosage de calcium.

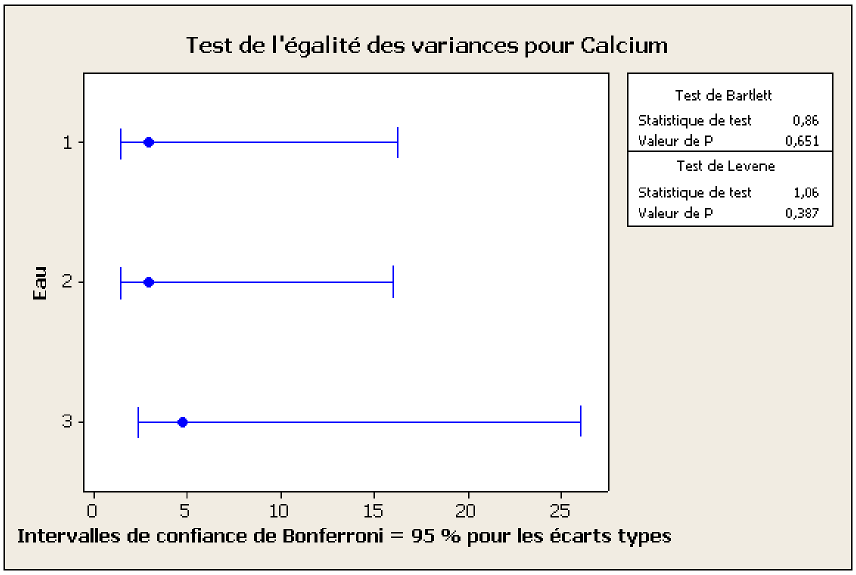
On a dosé la teneur en calcium de trois types d'eaux issues d'origines géographiques différentes. Chaque type d'eau a fait l'objet de quatre prélèvements. Les résultats des dosages (en milligramme de calcium par litre d'eau) ont été reproduits dans le tableau ci-dessous.

Eau 1	18	20	22	25
Eau 2	15	16	17	21
Eau 3	15	20	21	25

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 : Calcium en fonction de Eau

                          Rang
Eau        N  Médiane  moyen      Z
1          4    21,00    8,0   1,02
2          4    16,50    4,0  -1,70
3          4    21,00    7,5   0,68
Global    12             6,5

B H = 2,92  DL = 2  P = 0,232

* REMARQUE * Un ou plusieurs petits échantillons
```

....

Valeurs critiques de la statistique de Kruskal-Wallis pour 3 échantillons.

Effectifs			Risque de première espèce α				
n_1	n_2	n_3	.10	.05	.01	.005	.001
2	2	2	4.25				
3	2	1	4.29				
3	2	2	4.71	4.71			
3	3	1	4.57	5.14			
3	3	2	4.56	5.36			
3	3	3	4.62	5.60	7.20	7.20	
4	2	1	4.50				
4	2	2	4.46	5.33			
4	3	1	4.06	5.22			
4	3	2	4.51	5.44	6.44	7.00	
4	3	3	4.71	5.73	6.75	7.32	8.02
4	4	1	4.17	4.97	6.67		
4	4	2	4.55	5.60	7.14	7.59	8.32
4	4	3	4.65	5.69	7.66	8.00	8.65
4	4	4	4.20	5.00			
5	2	1	4.20	5.00			
5	2	2	4.36	5.16	6.53		
5	3	1	4.02	4.96			
5	3	2	4.65	5.25	6.82	7.18	
5	3	3	4.53	5.65	7.08	7.51	8.24
5	4	1	3.99	4.99	6.95	7.36	
5	4	2	4.54	5.27	7.12	7.57	8.11
5	4	3	4.55	5.63	7.44	7.91	8.50
5	4	4	4.62	5.62	7.76	8.14	9.00
5	5	1	4.11	5.13	7.31	7.75	
5	5	2	4.62	5.34	7.27	8.13	8.68
5	5	3	4.54	5.71	7.54	8.24	9.06
5	5	4	4.53	5.64	7.77	8.37	9.32
5	5	5	4.56	5.78	7.98	8.72	9.68