

Régression linéaire simple

Exercice 1

1. Relation entre le rendement (x) et la dose d'irrigation (y)

$$\sum_{i=1}^n x_i = 315 ; \sum_{i=1}^n y_i = 134 ; \sum_{i=1}^n x_i^2 = 16975 ; \sum_{i=1}^n y_i^2 = 3638 ; \sum_{i=1}^n x_i y_i = 7545 ; n = 6 ; \bar{x} = 52.5 ; \bar{y} = 22.33$$

- Calcul du coefficient de corrélation :

$$r = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}\right) \left(\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n}\right)}} = 0.95$$

- Test de signification du coefficient de corrélation :

$$t_{cal} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = 6.08$$

$$\begin{cases} ddl = n - 2 = 4 \\ \alpha = 5\% \end{cases} \Rightarrow t_{théo} = 2.776$$

$t_{cal} > t_{théo} \Rightarrow$ il existe une corrélation positive et significative entre le rendement et la dose d'irrigation au seuil de 5%.

2. Droite de régression

$$y = a + bx$$

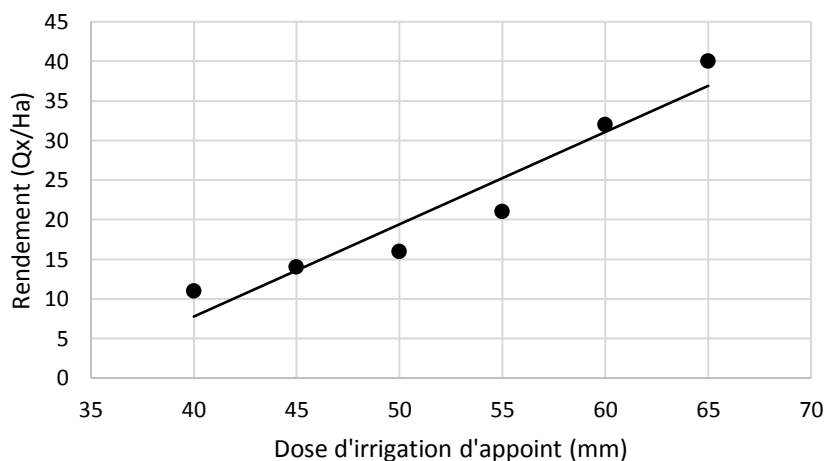
$$b = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}} = 1.16$$

$$a = \bar{y} - b\bar{x} = -38.86$$

$$\text{Donc : } y = 1,16x - 38.86$$

$R^2(\%) = r^2 \times 100 = 90.25\%$. Donc 90.25% de la variation observée pour le rendement est expliquée par le modèle linéaire

3. Courbe de tendance



Exercice 2

1. Il y a 7 points sur le nuage de points..... **Vrai**
2. Il y a 20 points sur le nuage de points..... **Faux**
3. La taille de l'échantillon est $n = 7$ **Vrai**
4. La taille de l'échantillon est $n = 20$ **Faux**
5. Le graphique sur lequel le nuage de points est représenté s'appelle un diagramme de dispersion..... **Vrai**
6. X est une variable à expliquer et Y est une variable explicative..... **Faux**
7. Aucune des propositions précédentes n'est exacte..... **Faux**

Exercice 3

1. Le coefficient de corrélation linéaire r entre les variables X et Y est négatif..... **Vrai**
2. La valeur absolue de r vaut 0,924..... **Vrai**
3. La statistique de test (t) vaut : - 5,413..... **Vrai**
4. 85% de la variabilité totale du temps d'attente aux caisses est expliqué par le modèle linéaire..... **Vrai**
5. La pente (b) de la droite de régression de Y en X vaut - 1,209..... **Vrai**
6. 16,889 est la valeur du 2^{ème} paramètre qui définit la droite de régression (a)..... **Vrai**
7. Aucune des propositions précédentes n'est exacte..... **Faux**

NB : les calculs se font de la même façon que l'exercice 1

ANOVA à 1 critère de classification

Exercice 1

- Analyse de la variance

Engrais	Répétitions			Ti	Moyennes
	Rep 1	Rep 2	Rep 3		
E1	12.2	12.5	13.5	38.2	12.7
E2	14.0	13.0	15.6	42.6	14.2
E3	18.1	21.3	17.8	57.2	19.1
E4	31.4	29.4	30.9	91.7	30.6
E5	41.5	39.6	38.7	119.8	39.9
G				349.5	

$$CF = \frac{G^2}{n} = \frac{(349.5)^2}{15} = 8143.35$$

$$SSTO = \sum_{i=1}^t \sum_{j=1}^{r_i} y_{ij}^2 - CF = [(12.2)^2 + \dots + (38.7)^2] - 8143.35 = 1643.72$$

$$SST = \sum_{j=1}^t \frac{T_j^2}{r_j} - CF = [(38.2)^2 + \dots + (119.8)^2] - 8143.35 = 1625.57$$

$$SSE = SSTO - SST = 1643.72 - 1625.57 = 18.14$$

$$MST = SST/4 = 406.39$$

$$MSE = SSE/10 = 1.81$$

$$F_{cal} = MST/MSE = 223.94$$

SV	df	SS	MS	F _{cal}	F _{théo (5%)}
Traitement	4	1625.57	406.39	223.94*	3.48
Erreur	10	18.14	1.81		
Total	14	1643.72			

$$\begin{cases} ddl_T = 4 \\ ddl_E = 10 \\ \alpha = 5\% \end{cases} \Rightarrow F_{théo(5\%)} = 3.48$$

$F_{cal} > F_{théo} \Rightarrow$ Il existe une différence significative entre les traitements étudiés. Donc les quatre engrais agissent différemment sur le rendement de la culture de blé.

$$CV (\%) = \frac{\sqrt{MSE}}{Moyenne} \times 100 \quad \text{Moyenne générale} = \frac{G}{n} = \frac{349.5}{15} = 23.3$$

$$CV (\%) = \frac{\sqrt{1.81}}{23.3} \times 100 = 5.77 \%$$

- Comparaison des moyennes :

$$\Delta_{NK} = Q \sqrt{\frac{MSE}{r}} \quad \begin{cases} k = 5 \\ ddl_E = 10 \\ \alpha = 5\% \end{cases} \Rightarrow Q = 4.65 \quad \begin{cases} r = 3 \\ MSE = 1.81 \end{cases}$$

$$\Delta_{NK} = 4.65 \sqrt{\frac{1.81}{3}} \Rightarrow \Delta_{NK} = 3.61$$

	E1	E2	E3	E4	E5
E1	-	NS	S	S	S
E2		-	S	S	S
E3			-	S	S
E4				-	S
E5					-

Exercice 2

SV	df	SS	MS	F _{cal}	F _{théo (5%)}
Traitement	1	1040.4	1040.4	76.21*	5.32
Erreur	8	109.2	13.65		
Total	9	1149.6			

$$\begin{cases} ddl_T = 1 \\ ddl_E = 8 \\ \alpha = 5\% \end{cases} \Rightarrow F_{théo(5\%)} = 5.32$$

$F_{cal} > F_{théo} \Rightarrow$ Il existe une différence significative entre les traitements étudiés.

$$CV (\%) = \frac{\sqrt{MSE}}{Moyenne} \times 100 \quad \text{Moyenne générale} = \frac{G}{n} = \frac{262}{10} = 26.2$$

$$CV (\%) = \frac{\sqrt{13.65}}{26.2} \times 100 = 14.10 \%$$

NB : les calculs se font de la même façon que l'exercice 1

Exercice 3

- Analyse de la variance

Traitements	Répétitions			T _i	Moyennes
	Rep 1	Rep 2	Rep 3		
Tallage	12	12	13	37	12.33
Montaison	14	13	15	42	14
Floraison	17	21	18	56	18.67
Remplissage	29	31	30	90	30
R_j	72	77	76		
G				225	18.75

$$CF = \frac{G^2}{n} = \frac{(225)^2}{12} = 4218.75$$

$$SSTO = \sum_{i=1}^t \sum_{j=1}^{r_i} y_{ij}^2 - CF = [(12)^2 + \dots + (30)^2] - 4218.75 = 584.25$$

$$SSR = \sum_{j=1}^r \frac{R_j^2}{t} - CF = [(72)^2 + \dots + (76)^2] - 4218.75 = 3.5$$

$$SST = \sum_{i=1}^t \frac{T_i^2}{r_i} - CF = [(37)^2 + \dots + (90)^2] - 4218.75 = 570.91$$

$$SSE = SSTO - SST - SSR = 584.25 - 570.91 - 3.5 = 18.14$$

$$MSR = SSR/2 = 1.75$$

$$MST = SST/4 = 406.39$$

$$MSE = SSE/10 = 1.81$$

$$F_{cal} = MST/MSE = 223.94$$

SV	df	SS	MS	F _{cal}	F _{théo (5%)}
Bloc	2	3.5	1.75	1.06	
Traitement	3	570.91	190.3	116.11*	4.76
Erreur	6	9.83	1.63		
Total	11	584.25			

$$\begin{cases} ddl_T = 3 \\ ddl_E = 6 \\ \alpha = 5\% \end{cases} \Rightarrow F_{théo(5\%)} = 4.76$$

$F_{cal} > F_{théo} \Rightarrow$ Il existe une différence significative entre les traitements étudiés. Donc les quatre engrais agissent différemment sur le rendement de la culture de blé.

$$CV (\%) = \frac{\sqrt{MSE}}{Moyenne} \times 100 \quad \text{Moyenne générale} = \frac{G}{n} = \frac{225}{12} = 18.75$$

$$CV (\%) = \frac{\sqrt{1.63}}{18.75} \times 100 = 6.8 \%$$

- Comparaison des moyennes :

$$\Delta_{NK} = Q \sqrt{\frac{MSE}{r}} \quad \begin{cases} k = 4 \\ ddl_E = 6 \\ \alpha = 5\% \end{cases} \Rightarrow Q = 4.90 \quad \begin{cases} r = 3 \\ MSE = 1.63 \end{cases}$$

$$\Delta_{NK} = 4.90 \sqrt{\frac{1.63}{3}} \Rightarrow \Delta_{NK} = 3.61$$

	Tallage	Montaison	Floraison	Remplissage
Tallage	-	NS	S	S
Montaison		-	S	S
Floraison			-	S
Remplissage				-

ANOVA à 2 critères de classification

Exercice 1

- But : étudier l'influence de la date de semis et l'espacement entre les rangs sur le rendement d'une variété de blé
 - Dispositif : factoriel en blocs complètement randomisé (DFBCR)
 - Facteur A : date de semis (a = 2)
 - Facteur B : espacement entre les rangs (b = 4)
 - Trois blocs (r=3)
 - Huit traitements (t=ab=8)
 - Vingt-quatre unités expérimentales (n=rab=24)
2. Analyse de la variance du DFBCR

Traitements	Rep1	Rep2	Rep3	T _{ij}	Moyennes
D ₁ S ₁	2.41	2.31	1.45	6.17	2.06
D ₁ S ₂	2.97	2.61	2.36	7.94	2.65
D ₁ S ₃	3.14	2.61	2.58	8.33	2.78
D ₁ S ₄	3.88	2.79	2.58	9.25	3.08
D ₂ S ₁	2.53	2.16	2.86	7.55	2.52
D ₂ S ₂	3.88	2.47	2.68	9.03	3.01
D ₂ S ₃	3.64	2.75	2.48	8.87	2.96
D ₂ S ₄	4.05	3.54	2.66	10.25	3.42
R _k	26.5	21.24	19.65	G=67.39	

$$CF = \frac{G^2}{n} = \frac{(67.39)^2}{24} = 189.22$$

$$SSTO = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r y_{ijk}^2 - CF = [(2.41)^2 + \dots + (2.66)^2] - 189.22 = 8.60$$

$$SSR = \sum_{k=1}^r \frac{R_k^2}{ab} - CF = [(26.5)^2 + \dots + (19.65)^2] - 189.22 = 3.21$$

$$SST = \frac{\sum_{i=1}^a \sum_{j=1}^b T_{ij}^2}{r} - CF = [(6.17)^2 + \dots + (10.25)^2] - 189.22 = 3.55$$

$$SSE = SSTO - SST - SSR = 1.83$$

Il existe une différence significative entre les traitements étudiés au seuil de 5% (3.84 > 2.76).

Décomposition de l'effet 'traitement' sur facteur A (date de semis), facteur B (espacement) et leur interaction :

Exercice 2

1. Tableau d'analyse de la variance

Source de variation	ddl	SCE	CM	F_{cal}	F_{théo (5%)}
Répétition	3	1002	334		///
Variété	7	34741	4963	24.82*	2.14
Régime	2	46518	23259	116.30*	3.13
Variété*Régime	14	2011	143.64	0.72 ^{ns}	1.84
Erreur	69	13800	200	///	///
Totale	95	98072	///	///	///

2. Il existe une différence significative entre les variétés de tomate étudiées au seuil de 5% ($24.82 > 2.14$) ;
Il existe une différence significative entre les régimes d'irrigation appliqués au seuil de 5% ($116.30 > 3.13$) ;
L'interaction des deux facteurs étudiés (variété et régime) est statistiquement non significative au seuil de 5% ($0.72 < 1.84$).

3. Coefficient de variation

$$CV (\%) = \frac{\sqrt{MSE}}{Moyenne} \times 100$$

$$CV (\%) = \frac{\sqrt{200}}{127} \times 100 = 11.13 \%$$