

TP 2

0.1 Objectifs

Le but de ce tp est :

- de saisir des données sous SPSS.
- de faire une analyse de variance sous SPSS.
- de résoudre des exercices d'ANOVA.
- de préparer à analyser les données aux prochains exercices.

0.2 Exercice 1 : (ANOVA 1 à un plan équilibré)

Dans cet exercice, on veut étudier l'effet de traitement (il y' a quatre types de traitement A, B, C et D) sur la durée séparant deux crises d'asthme.

Donc les hypothèses à tester sont

$$\begin{cases} H_0 : \text{pas d'effet de traitement sur la durée séparant deux crises d'asthme,} \\ H_1 : \text{Il y'a un effet de traitement sur la durée séparant deux crises d'asthme,} \end{cases}$$

ou d'une autre façon

$$\begin{cases} H_0 : \mu_A = \mu_B = \mu_C = \mu_D, \\ H_1 : \exists i, j \in \{A, B, C, D\} \text{ tel que } \mu_i \neq \mu_j, \quad i \neq j. \end{cases}$$

L'hypothèse H_0 est-elle vraie? Pour le savoir, il faut appliquer un test d'ANOVA à un facteur avec SPSS.

Cet exercice se divise en deux étapes :

- Définir les variables et ses valeurs dans SPSS.
- Répondre aux questions d'exercice en analysant les données avec SPSS.

Étape 1 :

a) DÉFINIR LES VARIABLES DANS SPSS

Dans cet exercice il y 'a deux variables :

Variable quantitative : la durée séparant deux crises d'asthme

Variable qualitative (facteur) : Traitement (a 4 modalités)

Pour définir ces deux variables dans SPSS, cliquez sur **Affichages des variables** puis saisissez tous ces caractéristiques : Nom, Type, Largeur, Décimales, Etiquette, Valeurs et Mesure (voir l'image ci dessous).

Les autres caractéristiques seront définies par défaut (automatiquement par SPSS).

1- Nom :

Ecrivez le nom de la première variable (durée), puis ajoutez le nom de la deuxième variable (Traitement). (SPSS définit automatiquement les autres champs, par défaut).

(Le nom d'une variable ne doit jamais contenir d'espace ou d'accent).

2- Type :

Cliquez sur le carré bleu sous Type pour faire apparaître les différents types disponibles.

Une fenêtre apparaît ; toujours choisir le type [Numérique].

Remarque : Dans la même fenêtre, vous pouvez déterminez la **Largeur** de vos données, ainsi que le nombre de **Décimales**. Puis cliquez sur OK ou vous pouvez les déterminés séparément.

3- Etiquette : pour précisez les noms de vos variables.

Exemple : pour la variable Durée : vous écrivez Durée séparant deux crises d'asthme.

4- Valeur : pour définir la Valeur de vos variables qualitatives.

Allez à la variable Traitement→Valeur, cliquez sur le petit carré bleu

Une petite fenêtre s'ouvre : attribuez une [Valeur] à chacun des niveaux de votre variable

(1= "A" : 2 = "B", 3="C", 4="D").

Commencez par Valeur 1 puis Etiquette A, puis cliquez sur le bouton Ajouter.

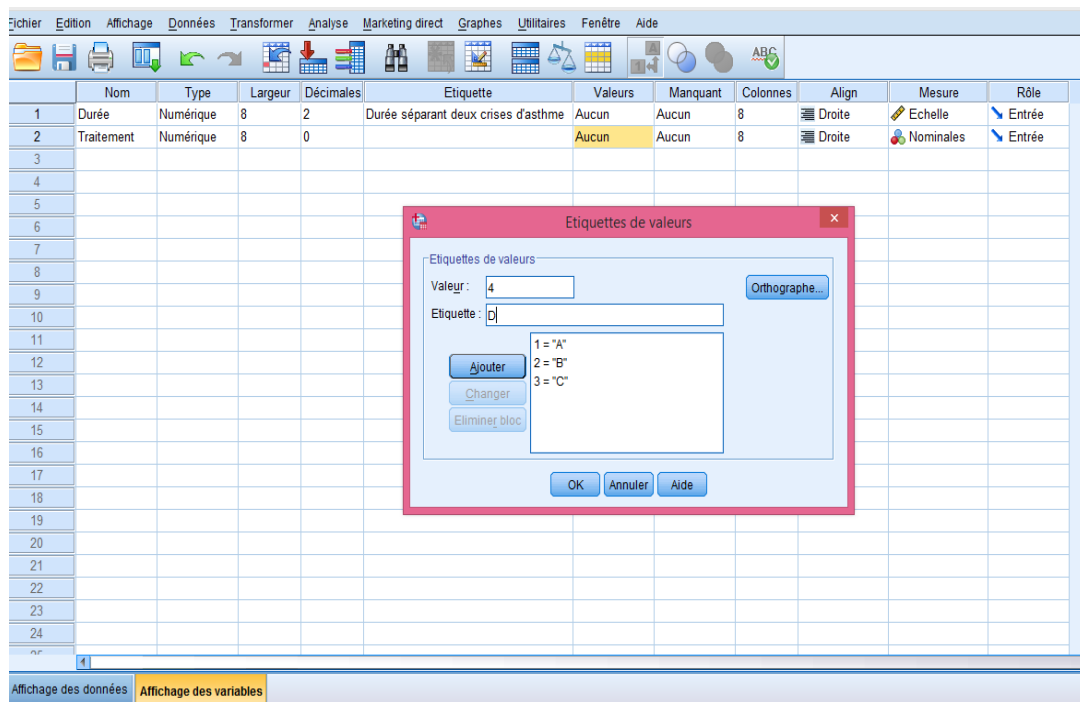
Ensuite Valeur 2 puis Etiquette B, puis cliquez sur le bouton Ajouter, faites la même chose pour les deux autres niveaux. Puis Cliquez sur OK.

Remarque : L'ajout des caractéristiques Etiquette et Valeur permettra de nommer avec plus de précision les variables dans vos tableaux et les axes de vos graphiques.

5- Mesure :

Vous choisissez **ÉCHELLE** pour les variables quantitatives et **NOMINALES** et **ORDINALES** pour les variables qualitatives.

Remarque : Enregistrez ensuite votre fichier SPSS dans votre espace personnel



b) Saisir les valeurs des variables sous SPSS

Cliquez sur **Affichage des données** pour entrer les valeurs des variables où chaque colonne représente une variable.

	Durée	Traitement	var	var	var	var	var	var	var	var	var	var	var	var	var
1	36,00	A													
2	37,00	A													
3	35,00	A													
4	38,00	A													
5	41,00	A													
6	42,00	B													
7	38,00	B													
8	39,00	B													
9	42,00	B													
10	44,00	B													
11	26,00	C													
12	26,00	C													
13	30,00	C													
14	38,00	C													
15	34,00	C													
16	42,00	D													
17	45,00	D													
18	50,00	D													
19	56,00	D													
20	58,00	D													
21															
22															
23															

Remarque : Il faut sauvegarder votre fichier.

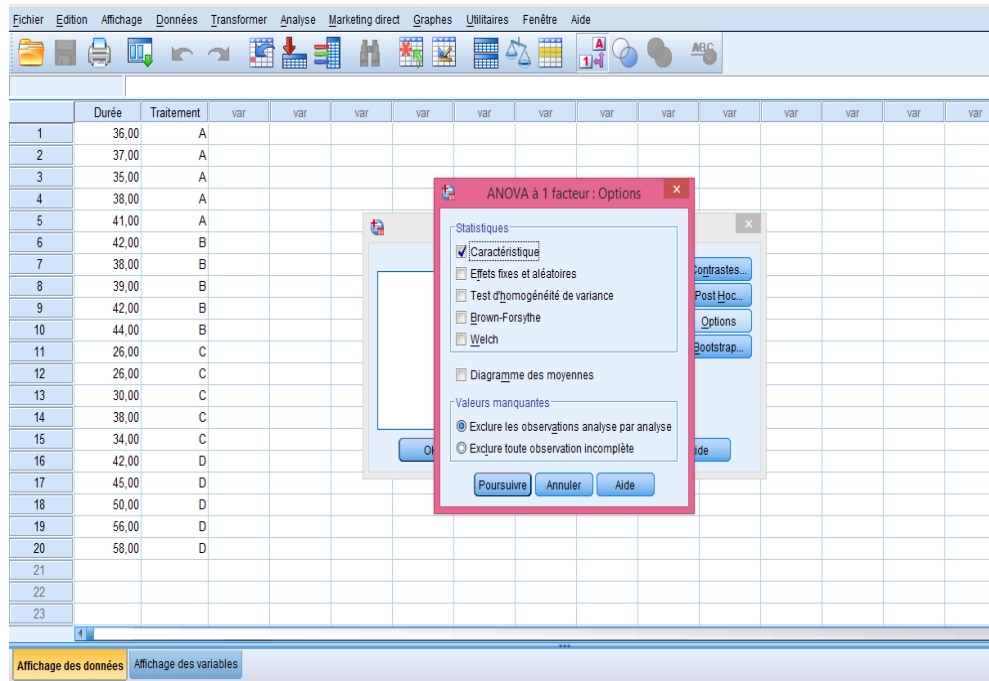
Etape 2 : Maintenant, on passe à la réponse à notre question.

Pour répondre à notre objectif

1- Allez à la barre des menus cliquez sur **Analyse**→**Comparer les moyennes**→**Anova à un facteur**.

2- Au moyen des flèches mettez dans la liste des variables dépendantes la variable **durée**, et dans la case critère la variable **Traitement**.

3- Cliquez ensuite sur Options, une fenêtre s'ouvre cochez **Caractéristiques** pour obtenir les moyennes et les variances des quatre groupes.



4- Cliquez ensuite sur Poursuivre, puis cliquez sur OK.

Le résultat final d'application de l'ANOVA1 sur les données nous fournies ce qui suit :

(a) Les caractéristiques descriptives des données qui sont résumées dans la table suivante :

Descriptives

Durée séparant deux crises d'asthme

	N	Moyenne	Ecart-type	Erreur standard	Intervalle de confiance à 95% pour la moyenne		Minimum	Maximum
					Borne inférieure	Borne supérieure		
A	5	37,4000	2,30217	1,02956	34,5415	40,2585	35,00	41,00
B	5	41,0000	2,44949	1,09545	37,9586	44,0414	38,00	44,00
C	5	30,8000	5,21536	2,33238	24,3243	37,2757	26,00	38,00
D	5	50,2000	6,87023	3,07246	41,6695	58,7305	42,00	58,00
Total	20	39,8500	8,34944	1,86699	35,9423	43,7577	26,00	58,00

b) La table de l'ANOVA 1 suivante

ANOVA à 1 facteur					
Durée séparant deux crises d'asthme					
	Somme des carrés	ddl	Moyenne des carrés	F	Signification
Inter-groupes	981,750	3	327,250	15,274	,000
Intra-groupes	342,800	16	21,425		
Total	1324,550	19			

De la table d'ANOVA 1 et plus précisément de la colonne Signification (Sig), on constate pour un risque 1% qu'on peut rejeter H_0 , c'est à dire, il y'a un effet significatif du traitement sur la durée séparant deux crises d'asthme (car $Sig < 1\%$).

Remarques

1- Le résultat significatif de ce test indique que certains groupes pris deux à deux sont significativement différents; et pour connaitre quelle paire de groupes présente une différence significative, il faut faire un test supplémentaire dit test post-hoc.

Ce test donc compare les quatre groupes. Pour accéder à ce test répéter les mêmes étapes précédentes puis cliquez sur **Post Hoc** comme vous pouvez voir qu'il y 'a plusieurs tests post Hoc répartis en deux ensembles selon la vérification de la condition d'homogénéité des variances (le premier ensemble de tests appliqué si les variances des groupes sont égaux et le deuxième si non) puis choisissez le test de Tukey (par exemple) et fixez le niveau de signification (1%), puis cliquez sur Poursuivre→OK.

La table de résultats suivante apparaît

Tests post hoc

Comparaisons multiples

Variable dépendante: Durée séparant deux crises d'asthme
Test de Tukey

(I) Traitement	(J) Traitement	Différence de moyennes (I-J)	Erreur standard	Signification	Intervalle de confiance à 99%	
					Borne inférieure	Borne supérieure
A	B	-3,60000	2,92746	,618	-14,3474	7,1474
	C	6,60000	2,92746	,151	-4,1474	17,3474
	D	-12,80000*	2,92746	,002	-23,5474	-2,0526
B	A	3,60000	2,92746	,618	-7,1474	14,3474
	C	10,20000	2,92746	,015	-,5474	20,9474
	D	-9,20000	2,92746	,029	-19,9474	1,5474
C	A	-6,60000	2,92746	,151	-17,3474	4,1474
	B	-10,20000	2,92746	,015	-20,9474	,5474
	D	-19,40000*	2,92746	,000	-30,1474	-8,6526
D	A	12,80000*	2,92746	,002	2,0526	23,5474
	B	9,20000	2,92746	,029	-1,5474	19,9474
	C	19,40000*	2,92746	,000	8,6526	30,1474

*. La différence moyenne est significative au niveau 0.01.

De cette table, on constate que pour un risque $\alpha = 1\%$, les paires des durées moyennes séparant deux crises d'asthme qui sont significativement différentes sont $(\mu_A \text{ et } \mu_D)$, $(\mu_C \text{ et } \mu_D)$.

Le logiciel SPSS utilise la notation * pour indiquer laquelle des paires de groupes sont significativement différents.

2- Pour tester l'homogénéité des variances cliquer sur Analyse → Comparer les moyennes → Anova à un facteur → Option → Cochez test d'homogénéité des variances → Poursuivre → Ok.

En suivant ces étapes vous obtenez une table de résultats contient la valeur de la statistique de Levene, les d.d.l associés et la signification (Le logiciel SPSS applique le test de Levene pour tester l'homogénéité des variances). Si signification $< \alpha$, vous pouvez rejeter l'hypothèse d'égalité des variances.

Pour cet exercice, l'application de test de Levene donne la table suivante :

Test d'homogénéité des variances			
Durée séparant deux crises d'asthme			
Statistique de Levene	ddl1	ddl2	Signification
3,462	3	16	,041

0.3 Exercice 3 : (ANOVA 1 à un plan non équilibré)

Suivez les mêmes étapes de l'exercice précédent vous pouvez obtenir les résultats suivants :

1- Les estimations des différentes moyennes 1, 2, 3, 4 et m sont données respectivement dans la partie sélectionnée de la table suivante :

Descriptives

Variable dépendante Rendement d'orge

	N	Moyenne	Ecart-type	Erreur standard	Intervalle de confiance à 95% pour la moyenne		Minimum	Maximum
					Borne inférieure	Borne supérieure		
Variété 1	3	45,6667	2,51661	1,45297	39,4151	51,9183	43,00	48,00
Variété 2	5	51,0000	5,52268	2,46982	44,1427	57,8573	43,00	57,00
Variété 3	4	45,0000	4,24264	2,12132	38,2490	51,7510	41,00	50,00
Variété 4	4	44,5000	5,00000	2,50000	36,5439	52,4561	39,00	51,00
Total	16	46,8750	5,09738	1,27435	44,1588	49,5912	39,00	57,00

2- Considérons l'hypothèse (H_0) : les rendements moyens de chaque variété sont égaux.

a) Pour obtenir la table d'ANOVA, on réalise une ANOVA à un facteur (Variété) avec des échantillons de tailles différentes.

Les étapes de raisonnement de l'ANOVA sous SPSS ne changeront pas et elles donnent la table suivante :

ANOVA à 1 facteur

Rendement d'orge

	Somme des carrés	ddl	Moyenne des carrés	F	Signification
Inter-groupes	126,083	3	42,028	1,913	,181
Intra-groupes	263,667	12	21,972		
Total	389,750	15			

b) On a $Sig > \alpha$ ($0.181 > 0.01$), donc on ne peut pas rejeter H_0 , cela signifie qu'il n'y a pas une différence significative entre les rendements des différentes variétés d'orge.

0.4 Exercice 4 : (ANOVA 2 à un plan équilibré)

Dans cet exercice le problème posé est l'analyse de l'influence d'un traitement hormonal, du sexe et leurs interactions sur la concentration de calcium dans le plasma.

1. Sachant que le modèle correspondant au problème est :

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, i = \overline{1, I}, j = \overline{1, J}, k = \overline{1, K};$$

alors les hypothèses à tester sont :

Effet du premier facteur (le traitement hormonal) :

$$\begin{cases} H_0 : \alpha_i = 0, \forall i = \overline{1, I}; \\ H_1 : \exists i(i = \overline{1, I}) / \alpha_i \neq 0. \end{cases}$$

Effet du second facteur (le sexe) :

$$\begin{cases} H_0 : \beta_j = 0, \forall j = \overline{1, J}; \\ H_1 : \exists j(j = \overline{1, J}) / \beta_j \neq 0. \end{cases}$$

Effet de l'interaction des deux facteurs (le traitement et le sexe simultanément) :

$$\begin{cases} H_0 : \gamma_{ij} = 0, \forall (i, j)(i = \overline{1, I}, j = \overline{1, J}); \\ H_1 : \exists (i, j)(i = \overline{1, I}, j = \overline{1, J}) / \gamma_{ij} \neq 0. \end{cases}$$

2. La technique adéquate pour l'analyse du problème est bien que l'ANOVA à deux facteurs.

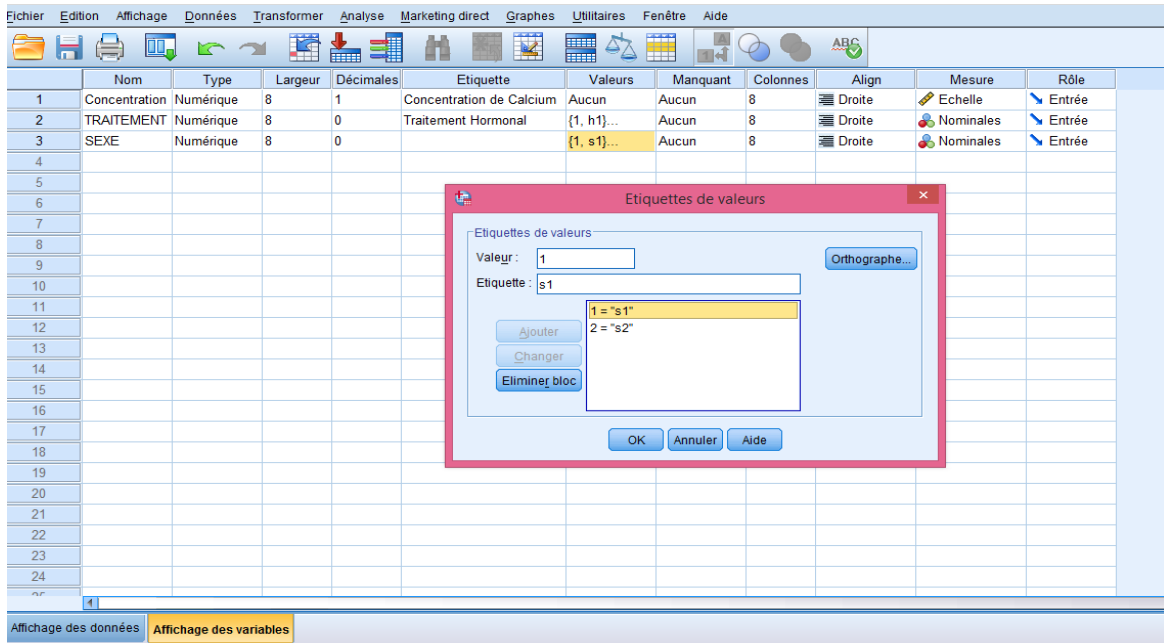
3. L'application de l'ANOVA 2 sur les données sous SPSS se fait de la manière suivante

a- Définissez les variables et entrez les données comme vous avez vu dans l'ANOVA à un facteur.

Remarque : Les variables étudiées dans cet exercice sont :

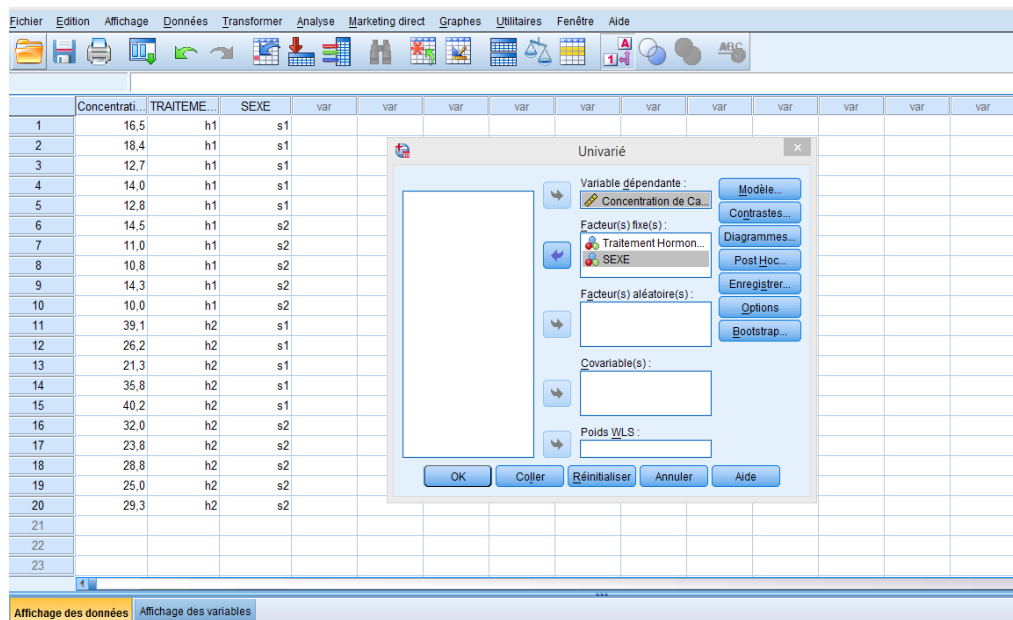
Une variable quantitative : concentration de calcium dans le plasma.

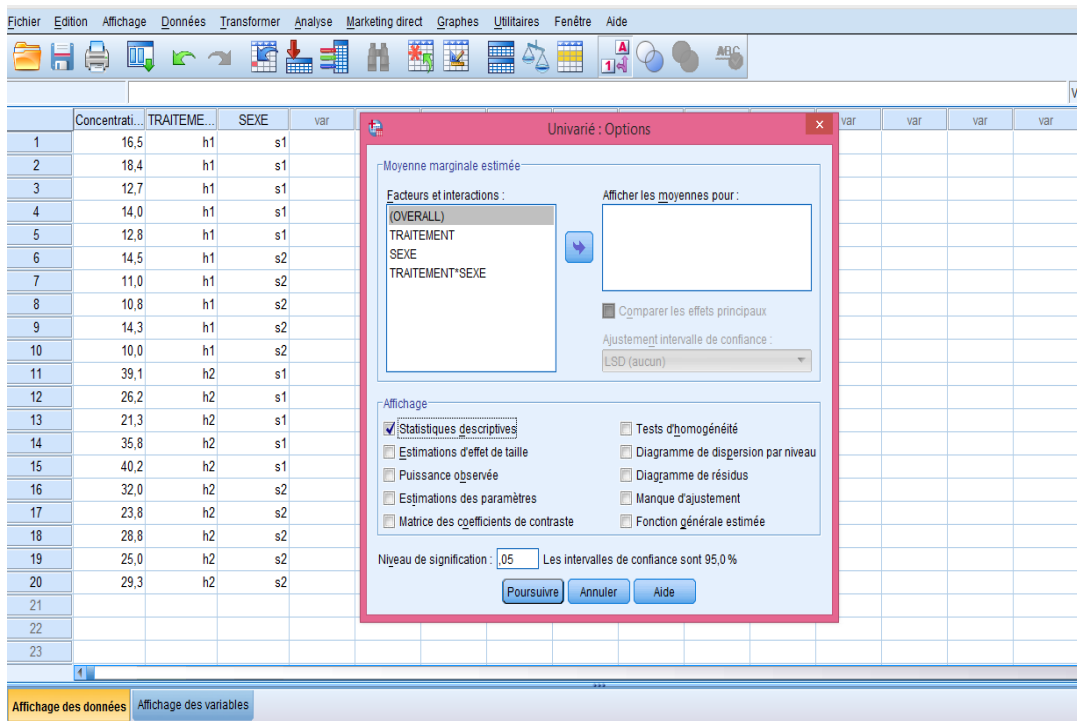
Deux variables qualitatives : traitement hormonal et sexe.



b- Allez à la barre de menus choisissez Analyse→Modèle linéaire général→Univarié

c- Dans la fenêtre qui apparaît, mettez la variable quantitative Concentration dans la case de variable dépendante et les deux variables qualitatives (traitement et sexe) dans la case facteur(s) fixe(s).





d- Cliquez sur le bouton Options puis cochez statistiques descriptives, cliquez ensuite sur Poursuivre, puis sur Ok.

Alors, le logiciel SPSS nous fournies ce qui suit :

- Tableaux des statistiques descriptives : SPSS donne deux tableaux :

Le premier tableau fournit par SPSS nous indique les tailles des nos échantillons selon les deux facteurs.(Traitement et sexe)

Le deuxième tableau résume les moyennes, les écarts-types obtenues pour chacun des échantillons.

Facteurs inter-sujets

		Etiquette de valeur	N
Traitement Hormonal	1	h1	10
	2	h2	10
SEXE	1	s1	10
	2	s2	10

Statistiques descriptives

Variable dépendante: Concentration de Calcium

Traitement Hormonal	SEXE	Moyenne	Ecart-type	N
h1	s1	14,880	2,4934	5
	s2	12,120	2,1159	5
	Total	13,500	2,6209	10
h2	s1	32,520	8,3497	5
	s2	27,780	3,3439	5
	Total	30,150	6,4959	10
Total	s1	23,700	10,9629	10
	s2	19,950	8,6649	10
	Total	21,825	9,8079	20

- Tableau de l'analyse principale

Les éléments importants à observer dans le tableau sont dans la partie sélectionnée. Ces dernières nous permettront de répondre à nos questions.

Tests des effets inter-sujets

Variable dépendante Concentration de Calcium

Source	Somme des carrés de type III	ddl	Moyenne des carrés	D	Sig.
Modèle corrigé	1461,326 ^a	3	487,109	21,273	,000
Ordonnée à l'origine	9526,613	1	9526,613	416,041	,000
TRAITEMENT	1386,113	1	1386,113	60,534	,000
SEXE	70,313	1	70,313	3,071	,099
TRAITEMENT * SEXE	4,901	1	4,901	,214	,650
Erreur	366,372	16	22,898		
Total	11354,310	20			
Total corrigé	1827,698	19			

a. R deux = ,800 (R deux ajusté = ,762)

Supposons qu'on désire prendre la décision avec un risque $\alpha = 5\%$, de se tromper. Alors d'après les résultats obtenus (Voir le tableau précédent), on constate que :

Effet du facteur traitement hormonal :

On a $Sig < \alpha$, donc le traitement hormonal a un effet significatif sur la concentration de calcium dans le plasma.

Effet du facteur sexe :

On a $Sig > \alpha$, donc nous ne pouvons pas rejeter H_0 , c'est à dire, il n'y a aucune différence significative de concentration du calcium chez les deux sexes.

Effet de l'interaction des deux facteurs (le traitement hormonal et le sexe simultanément) :

Pas d'effet significatif de l'interaction traitement hormonal et sexe (le fait que $Sig > \alpha$).

Remarque :

Pour le bouton **Modèle** vous pouvez laisser l'option par défaut, soit le modèle **factoriel complet**, qui inclut l'effet de toutes les variables qualitatives et l'effet d'interaction entre ces dernières. Si vous voulez tester un élément plus précis, vous pouvez sélectionner le modèle **personnalisé** et transférer les variables pour lesquelles vous désirez mesurer l'effet dans la boîte de droite.

0.5 Exercice 6 : (ANOVA 2 à un plan sans répétitions)

Deux techniques statistiques sont possibles pour résoudre cet exercice, et cela selon l'hypothèse posée.

1- Application d'ANOVA 2 sans répétitions : En effet, si nous considérons les Quadrats comme facteur donc on réalise ANOVA 2 sans répétitions pour répondre au problème.

Sachant que le modèle correspondant au problème est :

$$X_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}, i = \overline{1, I}, j = \overline{1, J};$$

alors les hypothèses à tester sont :

Effet du facteur Cages (Présence d'insectes) :

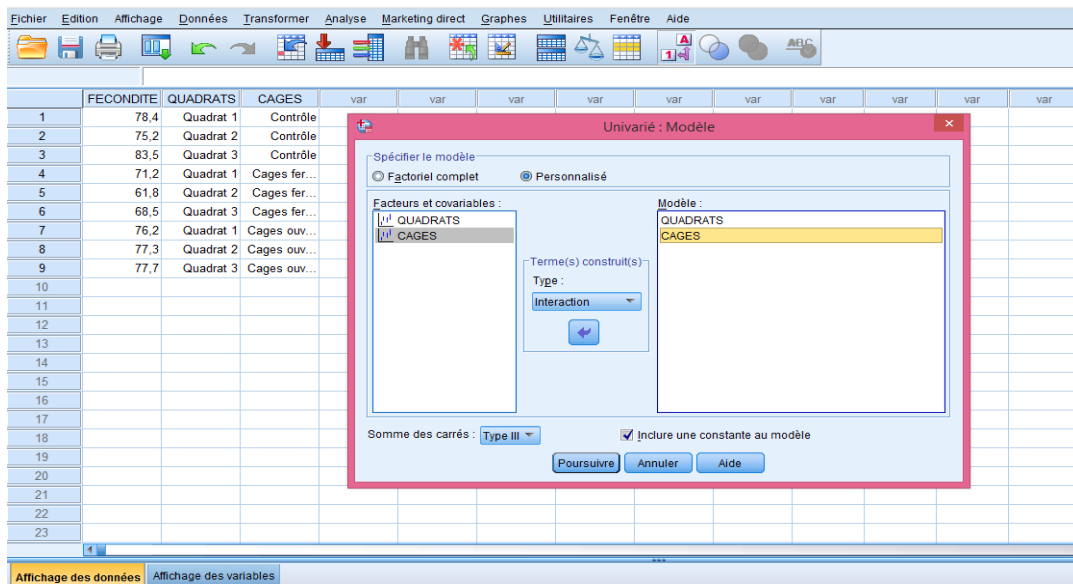
$$\begin{cases} H_0 : \alpha_i = 0, \forall i = \overline{1, I}; \\ H_1 : \exists i(i = \overline{1, I}) / \alpha_i \neq 0. \end{cases}$$

Effet du facteur Quadrats :

$$\begin{cases} H_0 : \beta_j = 0, \forall j = \overline{1, J}; \\ H_1 : \exists j(j = \overline{1, J}) / \beta_j \neq 0. \end{cases}$$

L'application de l'ANOVA 2 sur les données sous SPSS se fait de la manière suivante

- a- Définissez les variables et entrez les données comme vous avez vu précédemment.
- b- Répétez les mêmes étapes d'ANOVA 2 (Plan équilibré), c'est à dire allez à la barre de menus cliquez sur Analyse→Modèle linéaire général→Univarié.
- c- Dans la fenêtre apparaît, insérez la variable quantitative dans la boîte de variable dépendante et les deux variables qualitatives dans la boîte facteur(s) fixe(s).
- d- Cliquez sur le bouton **Modèle**, puis cochez **Personnalisé** et ensuite transférez les variables **Quadrats** et **Cages** dans la boîte de droite.



- e- Cliquez sur Poursuivre.

f- Cliquez sur le bouton Options puis cochez statistiques descriptives, cliquez ensuite sur Poursuivre, puis sur Ok.

Vous obtenez les résultats suivants :

- Tableaux des statistiques descriptives

		Etiquette de valeur	N
QUADRATS	1	Quadrat 1	3
	2	Quadrat 2	3
	3	Quadrat 3	3
CAGES	1	Contrôle	3
	2	Cages fermées	3
	3	Cages ouvertes	3

QUADRATS	CAGES	Moyenne	Ecart-type	N
Quadrat 1	Contrôle	78,400	.	1
	Cages fermées	71,200	.	1
	Cages ouvertes	76,200	.	1
	Total	75,267	3,6896	3
Quadrat 2	Contrôle	75,200	.	1
	Cages fermées	61,800	.	1
	Cages ouvertes	77,300	.	1
	Total	71,433	8,4085	3
Quadrat 3	Contrôle	83,500	.	1
	Cages fermées	68,500	.	1
	Cages ouvertes	77,700	.	1
	Total	76,567	7,5639	3
Total	Contrôle	79,033	4,1861	3
	Cages fermées	67,167	4,8398	3
	Cages ouvertes	77,067	,7767	3
	Total	74,422	6,3816	9

- Tableau d'analyse principale

Source	Somme des carrés de type III	ddl	Moyenne des carrés	D	Sig.
Modèle corrigé	285,431 ^a	4	71,358	7,071	,042
Ordonnée à l'origine	49848,004	1	49848,004	4939,793	,000
QUADRATS	42,736	2	21,368	2,117	,236
CAGES	242,696	2	121,348	12,025	,020
Erreur	40,364	4	10,091		
Total	50173,800	9			
Total corrigé	325,796	8			

a. R deux = ,876 (R deux ajusté = ,752)

A partir les résultats de ce dernier tableau, on constate que pour un risque de 5%, les Quadrats n'ont pas d'influence significative sur la fécondité moyenne tandis que le facteur

cages a une influence significative sur la fécondité. On conclue que la présence des insectes a une influence significative sur la fécondité moyenne des plantes en question.

2- Application d'ANOVA 1 : Si nous considérons les Quadrats comme répliation donc on réalise une ANOVA à un facteur (Cages) pour répondre au problème.

Exercice : Réalisez cette technique (ANOVA à un facteur), et que peut-on conclure ?