

Protocole de TP 02 :

Exécution du test ANOVA à deux facteurs

Pour validé test d'ANOVA à deux facteurs, il faut que les conditions suivantes soient valides :

- a) Une seule variable quantitative mesurée X , et deux variables qualitatives A avec k modalités et B avec l modalités.
- b) La distribution soit gaussienne (suit la loi Normale).
- c) L'échantillon est homogène pour la variance pour la variable quantitative, on ne peut pas trouver des valeurs plus loin que la moyenne, (Il n y a pas des valeurs extrêmes dans la série statistique).

Il faut comprendre que la réponse d'une question posée s'appelle une variable.

Tout réponse qui termine par une unité des mesures, est appelle une variable quantitative mesurable.

Tout réponse qui termine par des lettres, est appelle une variable qualitative.

Dans ANOVA à deux facteurs, on a deux variables qualitatives, et une seule variable quantitative.

I) Test ANOVA

On peut poser les questions suivantes :

Existe-il une influence de deux variables qualitatives « nominale ou bien ordinale » (qui s'appelle aussi le Facteur A et B , avec nombre k et l des modalités respectivement) sur la variable quantitative mesuré X ?

D'autre manière, est ce que les k échantillons sont homogènes ou bien non (par rapport la variable mesurable).

Est-ce que les l échantillons sont homogènes ou bien non (par rapport la variable mesurable).

Tableau des données.

Voici la table des données, avec les moyennes marginales pour les lignes et pour les colonnes, ainsi que la moyenne totale.

		Facteur A					
		Modalité 1	Modalité 2	Modalité 3	...	Modalité k	Moyennes marginales
Facteur B	Modalité 1	X_{11}	X_{12}	X_{13}		X_{1k}	
		X_{21}	X_{22}	X_{23}		X_{2k}	
		
		\bar{x}_1	\bar{x}_2	\bar{x}_3		\bar{x}_k	\bar{y} la moyenne des moyennes pour la modalité 1
	Modalité 2	X_{11}	X_{12}	X_{13}		X_{1k}	
		X_{21}	X_{22}	X_{23}		X_{2k}	
		
		\bar{x}_1	\bar{x}_2	\bar{x}_3		\bar{x}_k	\bar{y} la moyenne des moyennes pour la modalité 2
	X_{11}	X_{12}	X_{13}		X_{1k}	
		X_{21}	X_{22}	X_{23}		X_{2k}	
		
		\bar{x}_1	\bar{x}_2	\bar{x}_3		\bar{x}_k	\bar{y} la moyenne des moyennes pour la modalité I
Modalité I	X_{11}	X_{12}	X_{13}		X_{1k}		
	X_{21}	X_{22}	X_{23}		X_{2k}		
		
	\bar{x}_1	\bar{x}_2	\bar{x}_3		\bar{x}_k	\bar{y} la moyenne des moyennes pour la modalité I	
Moyennes marginales	\bar{x} la moyenne des moyennes pour la modalité 1	\bar{x} la moyenne des moyennes pour la modalité 2	\bar{x} la moyenne des moyennes pour la modalité 3	\bar{x}	\bar{x} la moyenne des moyennes pour la modalité k	$\bar{X}_{..}$ la moyenne totale pour les moyennes marginales	

Pour cela on doit ordonner les réponses par les étapes suivantes

Etape 1 : Proposition d'hypothèses

Pour le facteur A Les colonnes:

Hypothèse nulle H_0 : (Il existe homogénéité sur les k échantillons)=(tous les moyennes marginales sont significativement égales), (Il n'a pas une influence du facteur A sur la variable X).

$$(\bar{x}_1 = \bar{x}_2 = \bar{x}_3 =, \dots \bar{x}_k)$$

Hypothèse alternative H_1 : (Il n'y a pas homogénéité sur les k échantillons)=(tous les moyennes sont significativement différentes)=(Il y a une influence du facteur A sur la variable mesuré X).

Pour le facteur B Les lignes:

Hypothèse nulle H_0 : (Il existe homogénéité sur les l échantillons)=(tous les moyennes marginales sont significativement égales), (Il n'a pas une influence du facteur B sur la variable X).

$$(\bar{y}_1 = \bar{y}_2 = \bar{y}_3 =, \dots, \bar{y}_l)$$

Hypothèse alternative H_1 : (Il n'y a pas homogénéité sur les k échantillons)=(tous les moyennes sont significativement différentes)=(Il y a une influence du facteur B sur la variable mesuré X).

Pour le facteur d'interaction entre A et B :

Hypothèse nulle H_0 : (Il n'existe pas effet d'interaction entre A et B)

Hypothèse alternative H_1 : (Il existe effet d'interaction entre A et B).

Etape 2 : Les calculs (tableau d'ANOVA)

Source des variations	Somme des carrés	ddl	Moyenne des carrés	F
Facteur A	SCE_A	k-1	$CM_A = SCE_A / k - 1$	$F_A = \frac{CMA}{CMR}$
Facteur B	SCE_B	l-1	$CM_B = SCE_B / l - 1$	$F_B = \frac{CMB}{CMR}$
Facteur d'interaction A sur B	SCE_{AB}	$(k-1)(l-1)$	$CM_{AB} = SCE_{AB} / (k-1)(l-1)$	$F_{AB} = \frac{CMAB}{CMR}$
Résiduelle	SCE_R	$kl(n-1)$	$CM_R = SCE_R / kl(n-1)$	
Total	nkl		

Etape 3 : Conclusion (la décision)

Pour la décision, on utilise souvent la règle suivante :

Si Signification inférieure à $\alpha\%$. Alors on rejette H_0 .

Si Signification supérieure à $\alpha\%$. Alors on accepte H_0 .

Exemple:

Dans le cadre d'une expérience sur la durée de vie des bactéries, le but est de déterminer la durée de vie en fonction du type solution Hydro-Alcoolique. Comme on sait que les bactéries ont une durée de vie qui dépend de la température d'utilisation et aussi avec le type de solution

Hydro-Alcoolique qu'il été creusé sur la surface, un plan à deux facteurs (type de solution et température d'utilisation). Les résultats dans le tableau suivant

	15°C				25°C				35°C			
Type I	130	155	74	180	34	40	80	75	20	70	82	58
Type II	150	188	159	126	136	122	106	115	25	70	58	45
Type III	138	110	168	160	174	120	150	139	96	104	82	60

- 1) Déterminer l'objectif pour cette expérience.
- 2) Déterminer la variable qualitative qui exprime les échantillons (Facteur A et B), et la variable quantitative a mesuré X.
- 3) Déterminer l'hypothèse nulle et alternative pour le test d'ANOVA à deux facteurs.
- 4) Avec un risque de signification de 5%, que peut-on dire sur l'influence du facteur A et B sur la variable X?

Il est nécessaire de réécrire le tableau des données sous la forme suivante, pour simplifier l'introduction des données sous SPSS.

Température	Type de solution	Durée de Vie
15°C	Type I	130
15°C	Type I	150
15°C	Type I	138
15°C	Type II	155
15°C	Type II	188
15°C	Type II	110
15°C	Type II	74
15°C	Type III	159
15°C	Type III	168
15°C	Type III	...
15°C

Reponse

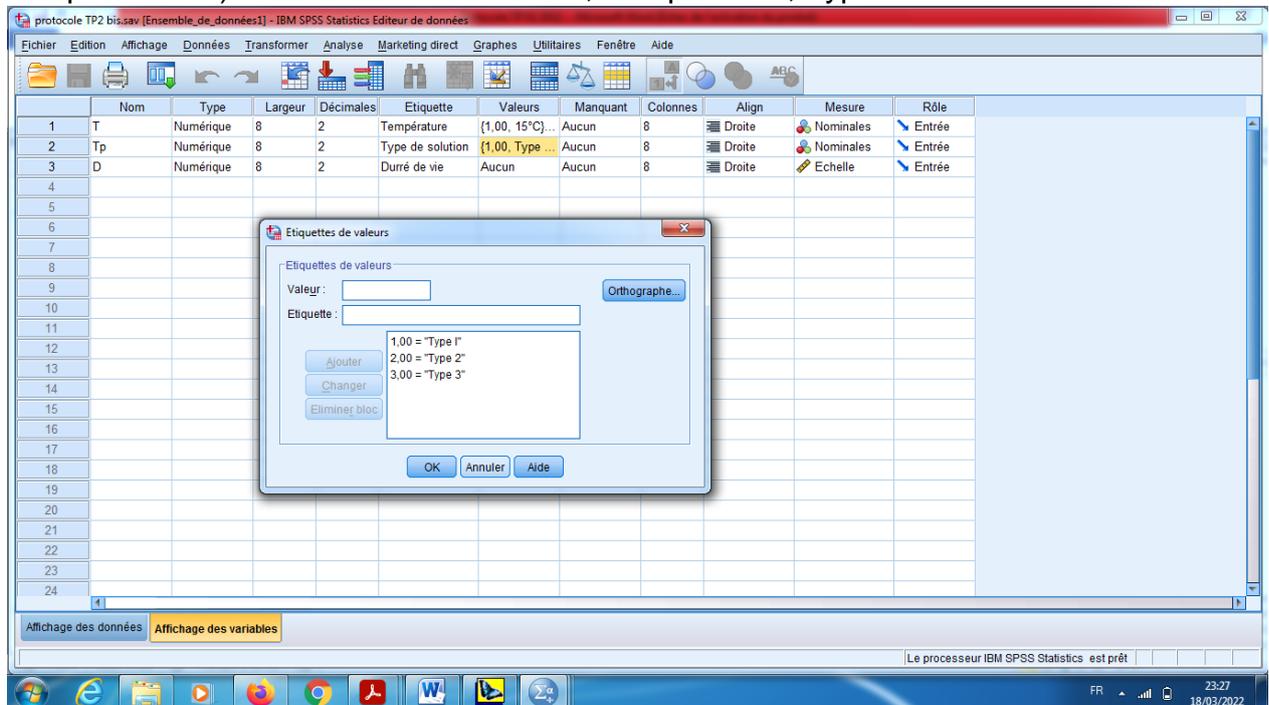
- 1) Notre objectif est de savoir l'existence d'effet éventuel de la température sur la durée de vie des bactéries (la variable X), et de plus de savoir l'existence d'effet éventuel de la type de solution sur la durée de vie des bactéries (la variable X), et de plus pour déterminer les groupes qui sont homogènes, ainsi que d'établir qu'elle est la meilleur modalité pour chaque facteur.
- 2) Facteur A « variable qualitative » : Température, dont les modalités sont groupe (15°C, 25°C, et 35°C), et Facteur B « variable qualitative » : Type de solution, dont les modalités

sont groupe (Type I, Type II, et Type III). La variable quantitative mesuré X c'est la durée de vie.

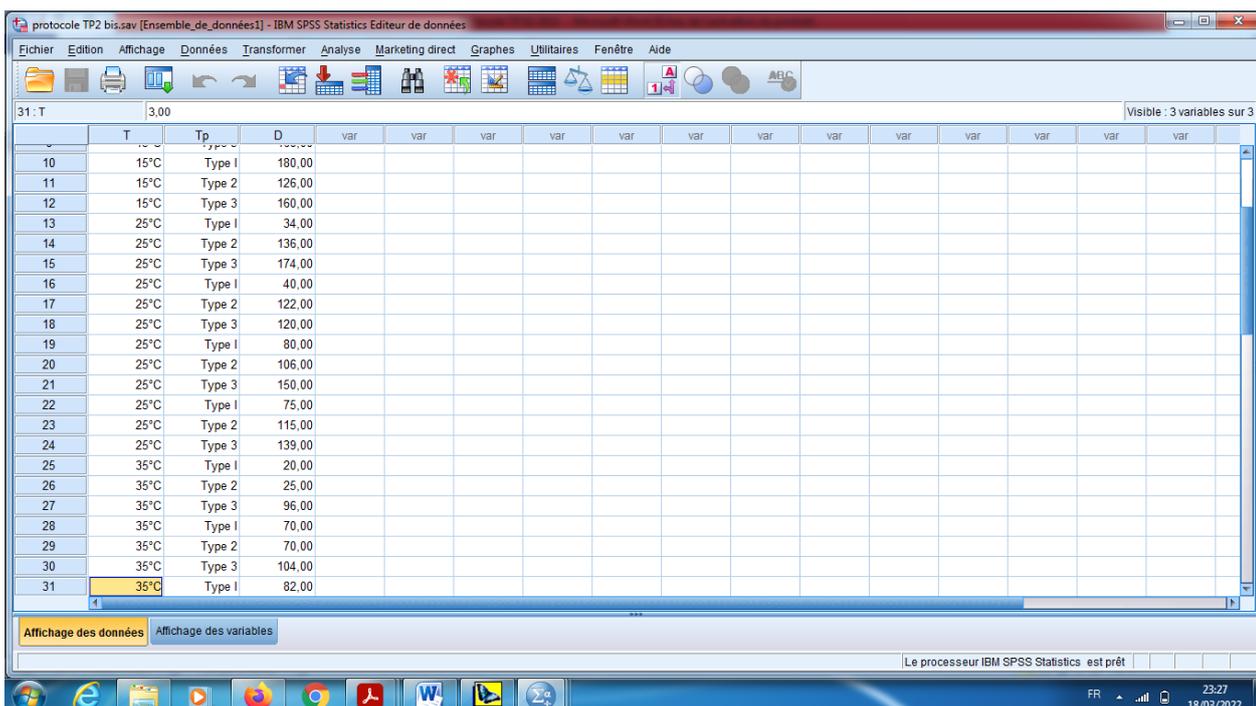
Pour vérifier les propositions d'hypothèses, tout d'abord il faut entrer ces données dans SPSS.

Nous suivons les étapes suivantes :

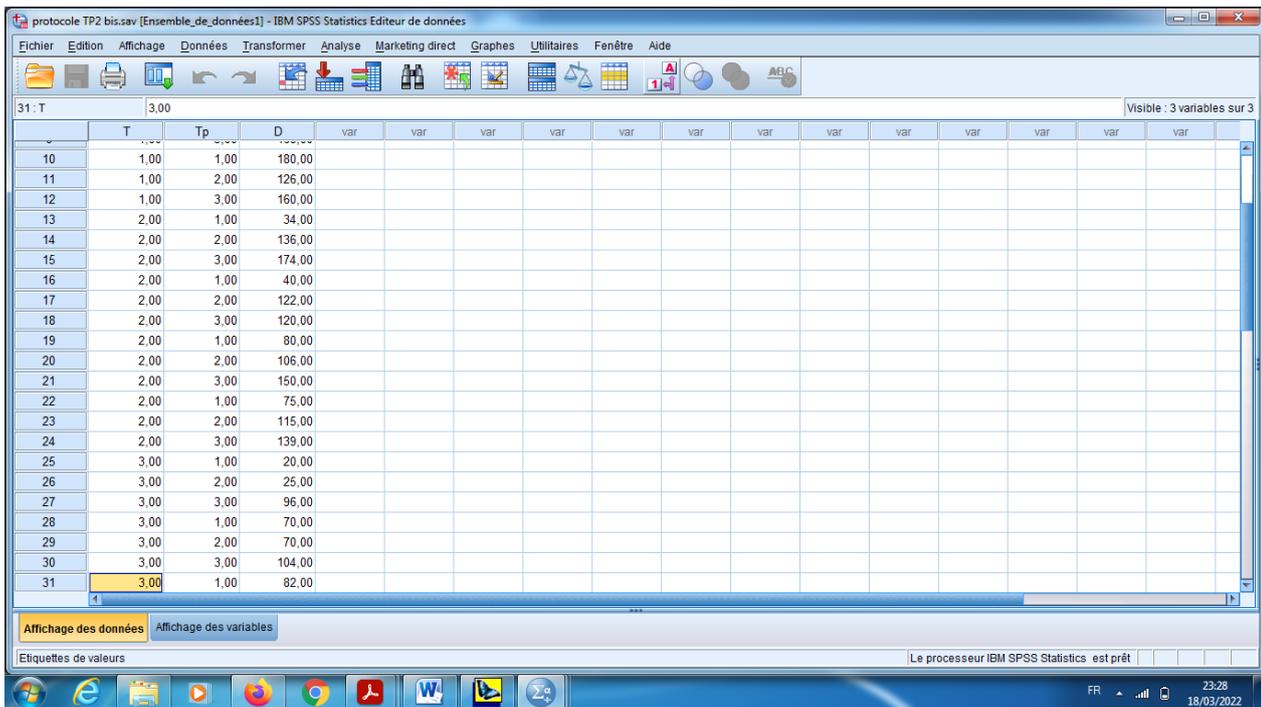
- a) Il faut définir dans la barre en bas « Affichage des variables » : les variables (qualitative et deux quantitative) suivantes : observations, Température, Type de solution et la durée de vie.



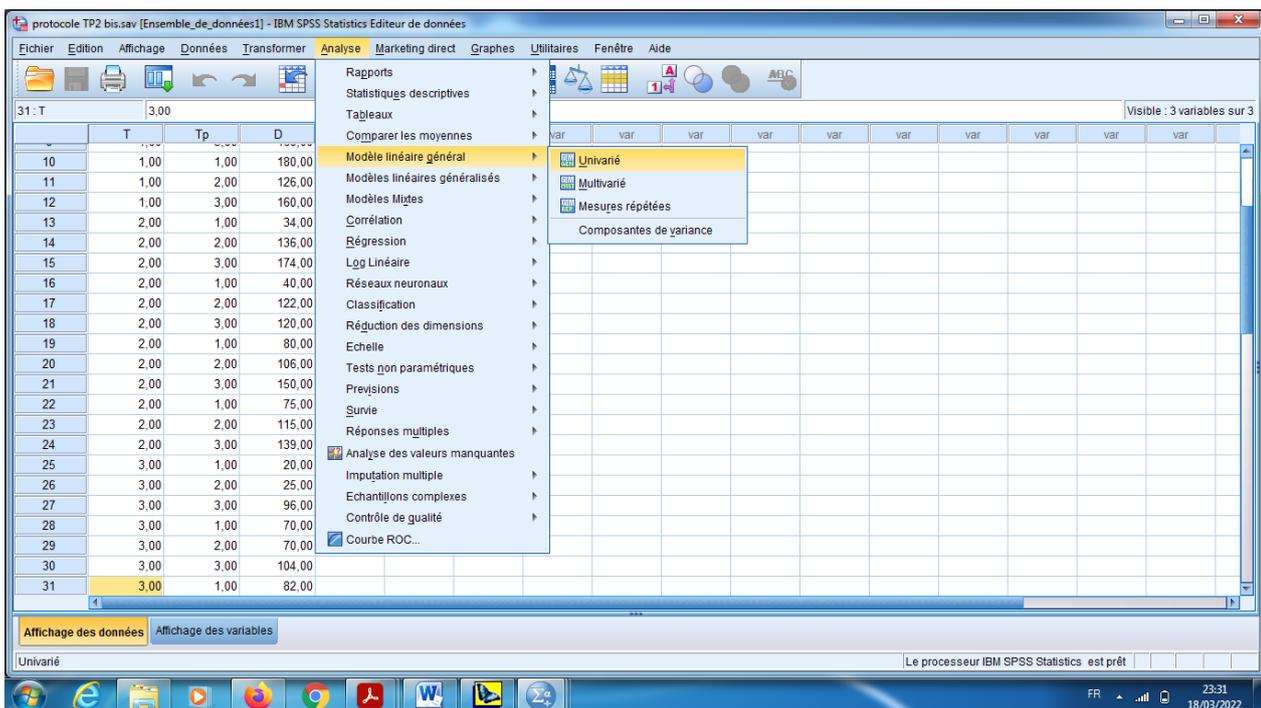
- b) On choisit les modalités pour la variable qualitative qui représente les échantillons de la variable (température) dans « valeurs », on peut prendre comme un exemple (la valeur 1 pour 15°C, et la valeur 2 pour 25°C, et 3 pour 35°C, aussi pour la variable type de solution.
- c) On introduit les données dans la barre « Affichage des données ».



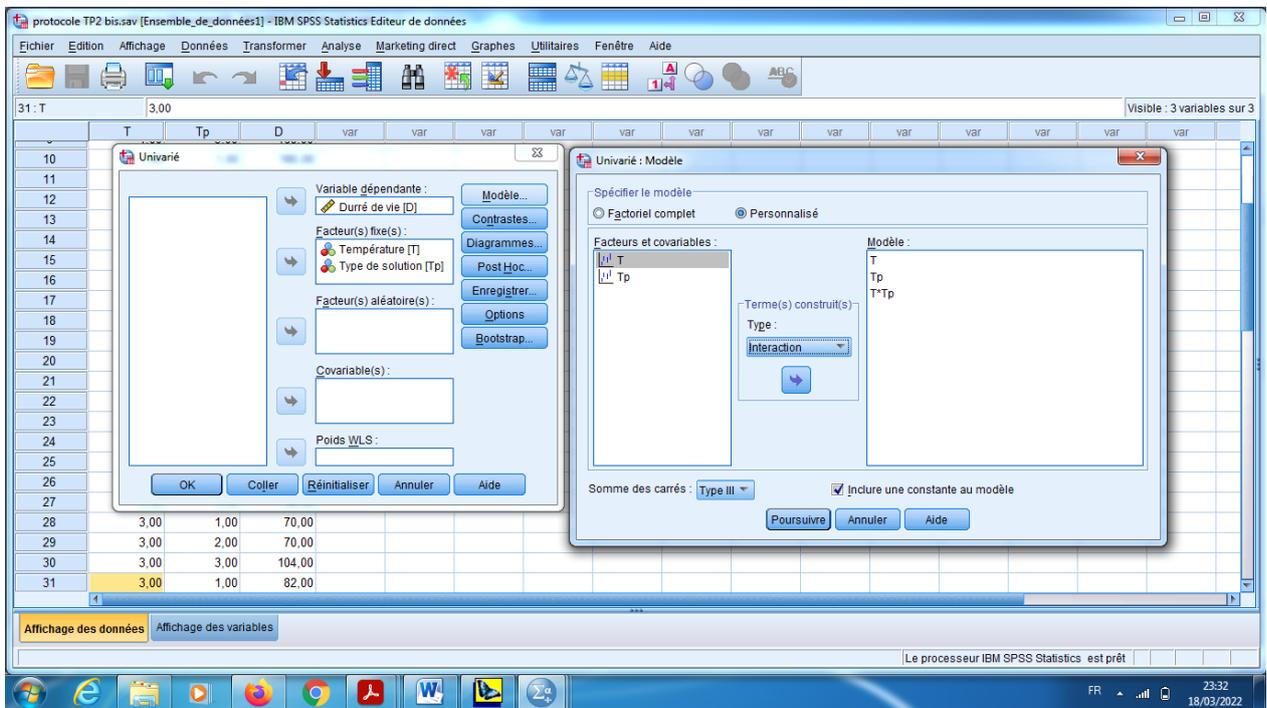
En cliquant sur le bouton « Etiquettes des valeurs » pour visualiser le codage des échantillons.



Pour obtenir le test d'analyse de la variance à deux facteurs, on suivre les étapes : Analyse, puis Modèle linéaire général, puis Uni varié (une seule variable dépendante).

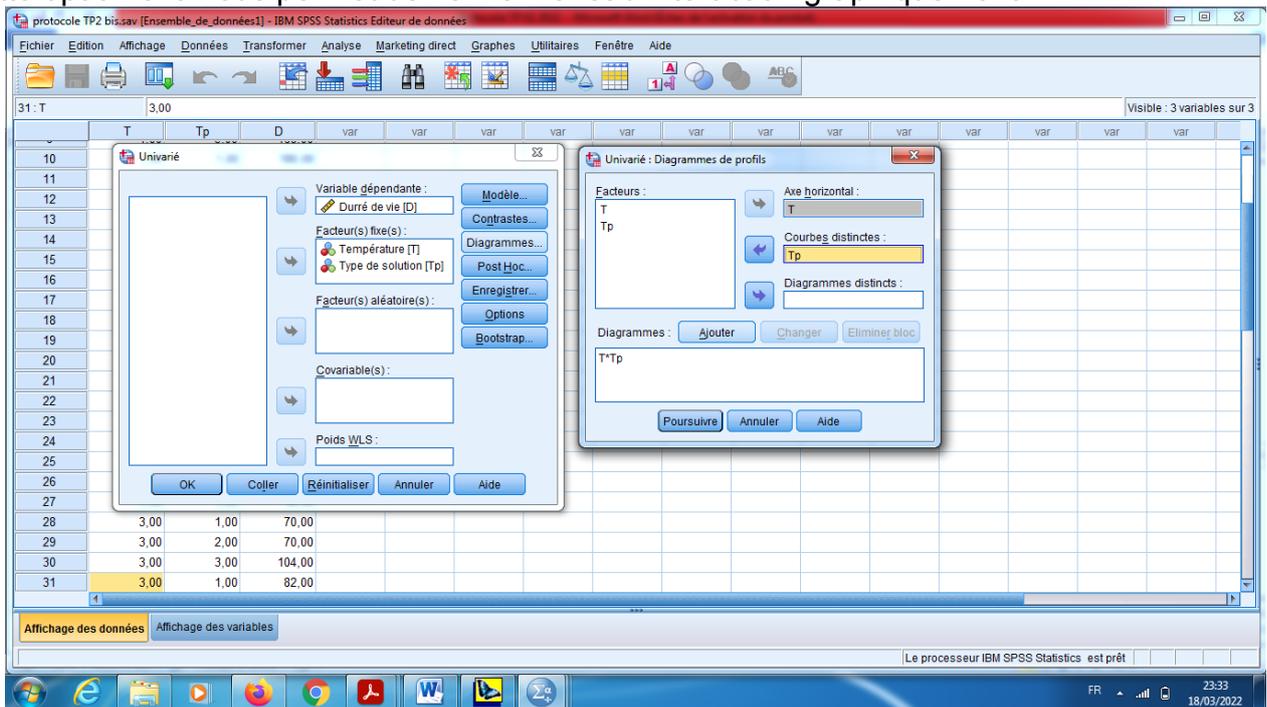


- On pose la variable durée de vie X dans le choix « variables dépendantes », et dans le « facteurs fixes » on pose les variable qui indiquent les deux facteurs A et B.
- Pour « Modèle » faire glisser les deux facteurs dans le modèle d'interaction.

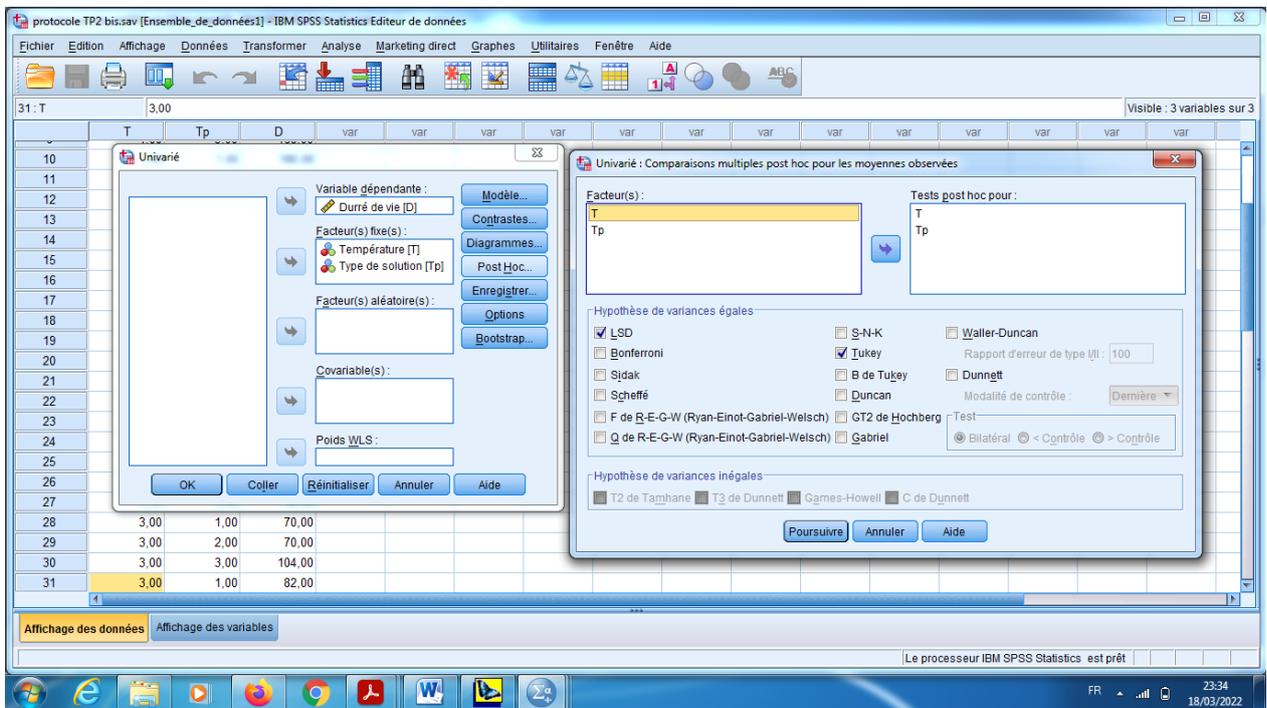


c) Dans le choix « Diagrammes », on pose le caractères (T) sur l'axe horizontal, et caractère (Tp) sur les courbes distinctés, et puis on clique sur « Ajouter ».

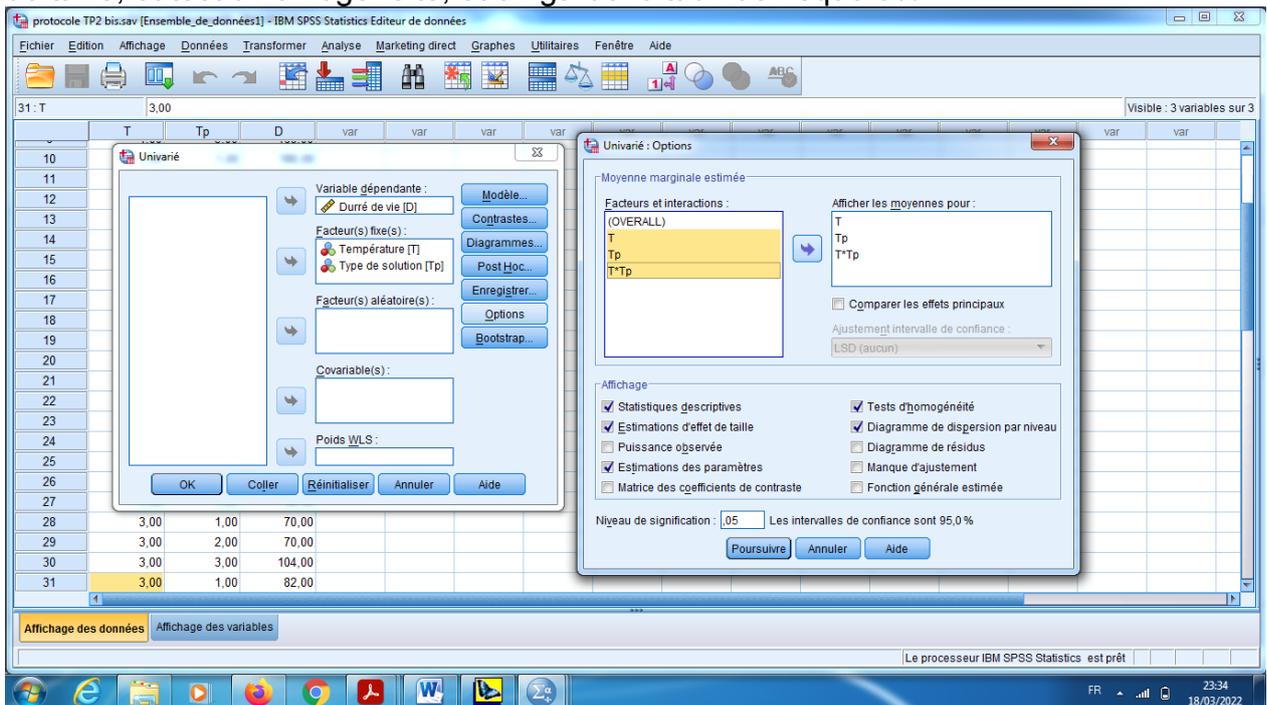
Cette option elle nous permet de vérifier l'effet d'interaction graphiquement.



d) Dans le choix « Post-Hoc », on pose les facteurs sur (test post-hoc pour), et on choisit le test (LSD et Tukey).



e) Dans le choix « Options », on pose les facteurs T et Tp et interaction dans la boîte (Affiché les moyennes pour), et puis on coche sur les choix statistique descriptive, estimation d'effet de taille, et test d'homogénéité, et on garde le taux de risque 5%.



En fin OK.

Interprétation des résultats

Pour test d'homogénéité de la variance, c'est le test de Levene, on remarque que $Sig=0,529 > 0,05$ alors on accepte l'hypothèse de l'homogénéité de la variance (c'est-à-dire que la distribution est homogène de la variance entre les échantillons)

Les deux premiers tableaux nous donnent une vision générale sur la statistique descriptive pour la variable dépendante ainsi que pour les deux variable qualitatives

(N=12=3 modalités×4 répétitions). Pour le deuxième tableau est représenté les moyennes pour chaque case (intersection entre la modalité i et la modalité j).

The screenshot shows the IBM SPSS Statistics Viewer interface. The main window displays two tables: 'Facteurs inter-sujets' and 'Statistiques descriptives'. The 'Statistiques descriptives' table is a two-way ANOVA table showing the mean and standard deviation for the dependent variable 'Durré de vie' across different combinations of 'Température' (15°C, 25°C, 35°C) and 'Type de solution' (Type 1, Type 2, Type 3). The total sample size is N=12.

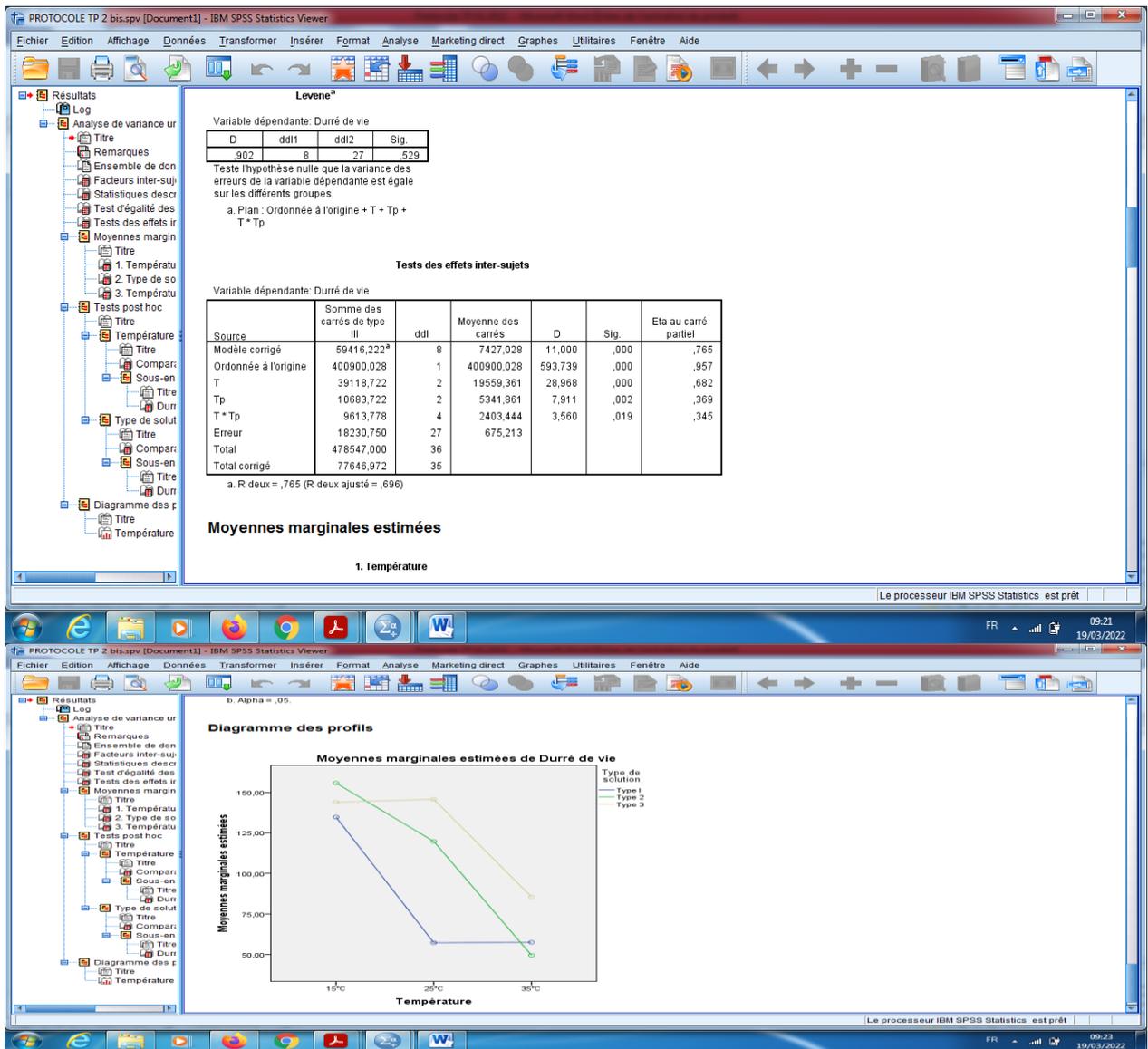
	Etiquette de valeur	N
Température	1,00 15°C	12
	2,00 25°C	12
	3,00 35°C	12
Type de solution	1,00 Type 1	12
	2,00 Type 2	12
	3,00 Type 3	12

Variable dépendante: Durré de vie		Moyenne	Ecart-type	N
15°C	Type 1	134,7500	45,35324	4
	Type 2	155,7500	25,61738	4
	Type 3	144,0000	25,97435	4
	Total	144,8333	31,69409	12
25°C	Type 1	57,2500	23,59908	4
	Type 2	119,7500	12,65899	4
	Type 3	145,7500	22,54440	4
	Total	107,5833	42,88347	12
35°C	Type 1	57,5000	26,85144	4
	Type 2	49,5000	19,26136	4
	Type 3	85,5000	19,27866	4
	Total	64,1667	25,67218	12
Total	Type 1	83,1667	48,56888	12
	Type 2	108,3333	49,47237	12
	Type 3	125,0833	35,76555	12

Tableau d'ANOVA à deux facteurs (test d'effets inter-sujets), qui nous permet d'établir s'il existe un effet du facteur A sur la variable dépendante X, et aussi s'il existe un effet du facteur B sur la variable dépendante X, et s'il existe un effet d'interaction entre deux facteurs A et B. Tout d'abord il est nécessaire de vérifier l'effet d'interaction entre les deux facteurs, alors ($\text{sig}=0,019 < 0,05$), ce qui nous donne l'acceptation d'hypothèse alternative pour l'interaction, alors il y a un effet d'interaction.

Pour le facteur A « température », on remarque que $\text{sig}=0,000 < 0,05$. Alors on accepte l'hypothèse alternative, donc il y a un effet de température sur la moyenne de la variable durée de vie, et on peut dire aussi que le taux d'influence c'est 100% ($1-0,000=1$).

Pour le facteur B « Type de solution alcoolique », on remarque que $\text{sig}=0,002 < 0,05$. Alors on accepte l'hypothèse alternative, donc il y a un effet de type de solution sur la moyenne de la variable durée de vie c'est-à-dire la différence entre les moyennes est significative pour le facteur type de solution, et on peut dire aussi que le taux d'influence c'est 99, 8% ($1-0,002=0,998$).



Cette graphique nous dit qu'il y a un effet d'interaction entre les deux facteurs, il suffit de voir s'il y a une intersection entre les segments.

Remarque importante:

S'il y a un effet d'interaction (acceptation l'hypothèse H_1), dans ce cas on ne peut rien dire. D'autre part, s'il n'y a pas un effet d'interaction entre les deux facteurs (acceptation d'hypothèse H_0), seulement dans ce cas, il est nécessaire d'étudier les effets principaux A et B.

On cherche les groupes d'homogénéité deux à deux par rapport au facteur température, on effectue pour cette raison le test de LSD ou bien le test de Tukey.

On remarque qu'il y a trois groupes différents, et la meilleure modalité c'est la température 15°C.

	(I) Température	(J) Température	Différence des moyennes (I-J)	Erreur standard	Sig.	Intervalle de confiance à 95%	
						Borne inférieure	Limite supérieure
Test de Tukey	15°C	25°C	37,2500	10,60827	,004	10,9477	63,5523
		35°C	80,6667	10,60827	,000	54,3643	106,9690
	25°C	15°C	-37,2500	10,60827	,004	-83,5523	-10,9477
		35°C	43,4167	10,60827	,001	17,1143	69,7190
	35°C	15°C	-80,6667	10,60827	,000	-106,9690	-54,3643
		25°C	-43,4167	10,60827	,001	-69,7190	-17,1143
LSD	15°C	25°C	37,2500	10,60827	,002	15,4836	59,0164
		35°C	80,6667	10,60827	,000	58,9003	102,4330
	25°C	15°C	-37,2500	10,60827	,002	-59,0164	-15,4836
		35°C	43,4167	10,60827	,000	21,6503	65,1830
	35°C	15°C	-80,6667	10,60827	,000	-102,4330	-58,9003
		25°C	-43,4167	10,60827	,000	-65,1830	-21,6503

En fonction des moyennes observées.
Le terme d'erreur est Carré moyen(Erreur) = 675,213.
*. La différence des moyennes est significative au niveau ,05.

Sous-ensembles homogènes

Durée de vie

Température	N	Sous-ensemble		
		1	2	3
Test de Tukey ^{a,b}				
35°C	12	64,1667	107,5833	144,8333
25°C	12			
15°C	12			
Sig.		1,000	1,000	1,000

Les moyennes des groupes dans des sous-ensembles homogènes sont

En effet : On compare par exemple entre (15°C et 25°C), on remarque que $Sig=0,004 < 0,05$. Alors on accepte l'hypothèse H1 : (15°C et 25°C ne sont pas homogènes).

De plus on remarque que les deux bornes d'intervalle de confiance sont positives, c'est-à-dire que $I-J > 0$, alors $I > J$, alors température 15°C est meilleure que 25°C.

Tableaux des moyennes marginales et les moyennes pour chaque case.

1. Température

Variable dépendante: Durée de vie

Température	Moyenne	Erreur standard	Intervalle de confiance à 95%	
			Borne inférieure	Limite supérieure
15°C	144,833	7,501	129,442	160,224
25°C	107,583	7,501	92,192	122,974
35°C	64,167	7,501	48,776	79,558

2. Type de solution

Variable dépendante: Durée de vie

Type de solution	Moyenne	Erreur standard	Intervalle de confiance à 95%	
			Borne inférieure	Limite supérieure
Type 1	83,167	7,501	67,776	98,558
Type 2	108,333	7,501	92,942	123,724
Type 3	125,083	7,501	109,692	140,474

3. Température * Type de solution

Variable dépendante: Durée de vie

Température	Type de solution	Moyenne	Erreur standard	Intervalle de confiance à 95%	
				Borne inférieure	Limite supérieure
15°C	Type 1	134,750	12,992	108,092	161,408
	Type 2	155,750	12,992	129,092	182,408
	Type 3	144,000	12,992	117,342	170,658
25°C	Type 1	57,250	12,992	30,592	83,908
	Type 2	119,750	12,992	93,092	146,408

Pour cela, on cherche les groupes d'homogénéités deux à deux par rapport au facteur Type de solution, on effectue pour cette raison le test de test de Tukey.

On remarque qu'il y a deux groupes différents, et la meilleure modalité c'est la type de solution N° :3.

PROTCOLE TP 2 bis.spv [Document1] - IBM SPSS Statistics Viewer

Fichier Edition Affichage Données Transformer Insérer Format Analyse Marketing direct Graphes Utilitaires Fenêtre Aide

Résultats

- Log
- Analyse de variance ur
 - Titre
 - Remarques
 - Ensemble de don
 - Facteurs inter-suj
 - Statistiques descr
 - Test d'égalité des
 - Tests des effets ir
 - Moyennes margin
 - 1. Températu
 - 2. Type de so
 - 3. Températu
 - Tests post hoc
 - Température
 - Titre
 - Compari
 - Sous-en
 - Durr
 - Type de solut
 - Titre
 - Compari
 - Sous-en
 - Durr
 - Diagramme des p
 - Titre
 - Température

(I) Type de solution	(J) Type de solution	moyennes (I-J)	Erreur standard	Sig.	Borne inférieure	Limite supérieure	
Test de Tukey	Type 1	Type 2	-25,1667	10,60827	,063	-51,4690	1,1357
		Type 3	-41,9167	10,60827	,001	-68,2190	-15,6143
	Type 2	Type 1	25,1667	10,60827	,063	-1,1357	51,4690
		Type 3	-16,7500	10,60827	,272	-43,0523	9,5523
	Type 3	Type 1	41,9167	10,60827	,001	15,6143	68,2190
		Type 2	16,7500	10,60827	,272	-9,5523	43,0523
LSD	Type 1	Type 2	-25,1667	10,60827	,025	-46,9330	-3,4003
		Type 3	-41,9167	10,60827	,001	-63,6830	-20,1503
	Type 2	Type 1	25,1667	10,60827	,025	3,4003	46,9330
		Type 3	-16,7500	10,60827	,126	-38,5164	5,0164
	Type 3	Type 1	41,9167	10,60827	,001	20,1503	63,6830
		Type 2	16,7500	10,60827	,126	-5,0164	38,5164

En fonction des moyennes observées.
Le terme d'erreur est Carré moyen(Erreur) = 675,213.
*. La différence des moyennes est significative au niveau ,05.

Sous-ensembles homogènes

Durré de vie

Type de solution	N	Sous-ensemble	
		1	2
Test de Tukey ^{a,b}			
Type 1	12	83,1667	
Type 2	12	108,3333	108,3333
Type 3	12		125,0833
Sig.		,063	,272

Les moyennes des groupes dans des sous-ensembles homogènes sont affichées.
En fonction des moyennes observées.

Le processeur IBM SPSS Statistics est prêt

FR 10:13 19/03/2022

En effet : On compare entre (Type de solution N° :1 et Type de solution N° :3), on remarque que $Sig=0,001 < 0,05$. Alors on accepte l'hypothèse H_1 : (Type de solution N° :1 et Type de solution N° :3 sont significativement différentes).

De plus on remarque que les deux bornes d'intervalle de confiance sont négatives, c'est-à-dire que $I-J < 0$, alors $I < J$, alors Type de solution N° :3 est meilleure que Type de solution N° :1.