République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

DÉPARTEMENT DE SNV



Polycopié du TP :

Biostatistiques

Régression linéaire sous SPSS (TP 4)

Préparé par ROUBI. A

2020/2021

Table des matières

Ta	able	des matières	2											
Li	ste d	les figures	3											
1	Régression linéaire sous SPSS													
	1.1	Rappel	1											
	1.2	Régression linéaire simple	2											
		1.2.1 Modèle de régression linéaire simple	2											
		1.2.2 Régression linéaire simple sous SPSS	3											
	1.3	Régression linéaire multiple	15											
		1.3.1 Modèle de régression linéaire multiple	15											
		1.3.2 Régression linéaire multiple sous SPSS	16											

Table des figures

1.1	Saisie des données sous SPSS	4
1.2	Présentation graphique des données (Etape 2 partie a)	5
1.3	Présentation graphique des données (Etape2 partie b)	6
1.4	Nuage de points présente la Biomasse en fonction de la Concentration d'azote.	6
1.5	Ajout du droite de régression au nuage de points	7
1.6	Ajout du droite de régression au nuage de points	8
1.7	Nuage de points et droite de régression	8
1.8	Procédures de la réalisation de la régression linéaire simple (Partie 1)	9
1.9	Procédures de la réalisation de la régression linéaire simple (Partie 2)	10
1.10	Résultats de la régression linéaire simple	11
1.11	Tableau récapitule les variables explicatives introduites dans le modèle	12
1.12	Tableau des coefficients de corrélation et de détermination	12
1.13	Table de L'ANOVA. .	14
1.14	Tableau des coefficients de la régression linéaire simple	14
1.15	Saisie des données.	18
1.16	Réalisation de la régression linéaire multiple sous SPSS (Partie 1)	19
1.17	Réalisation de la régression linéaire multiple sous SPSS (Partie 2)	20
1.18	Résultats de la régression linéaire multiple sous SPSS	21
1.19	Procédure de la réalisation de la régression linéaire multiple par étapes (en	
	choisissant la méthode "Pas à Pas")	23

Chapitre 1

Régression linéaire sous SPSS

Dans ce chapitre, on va présenter un rappel sur la notion de la régression linéaire. Puis, à travers d'un exemple d'application, on va donner les principales étapes à suivre pour faire une régression linéaire à l'aide du logiciel SPSS.

1.1 Rappel

Le but de la régression simple (respectivement. multiple) est d'expliquer une variable Y à l'aide d'une variable X (resp. plusieurs variables X_1, X_2, \ldots, X_p). Ou d'une autre façon la régression permet :

- De trouver (modéliser) la relation entre la variable Y et la variable X (ou entre la variable Y et plusieurs variables X_1, X_2, \ldots, X_p).
- De prédire les valeurs de la variable Y à partir des valeurs de X (si la variable X (ou les variables $X_j, j = 1, 2, ..., p$) est connue).

Remarque 1.1 La variable Y est appelée variable dépendante, ou variable à expliquer et les variables X_j (j = 1, 2, ..., p) sont appelées variables indépendantes, ou variables explicatives. C'est- à-dire, dans la régression simple (respectivement. multiple) on cherche d'une fonction f telle que

$$Y = f(X) + \epsilon, (resp.Y = f(X_1, X_2, \dots, X_p) + \epsilon),$$

où ϵ est une variable aléatoire (résidus).

Exemple 1.1 - Étude de la température (Y) en fonction de l'altitude (X).

- Étude de la taille (Y) en fonction du poids (X) ou l'inverse.

- Étude du nombre de morts d'une maladie (Y) en fonction du nombre des infectés (X).

- Étude de la taille (Y) en fonction du poids (X_1) et de l'âge (X_2) .

- Étude du poids de l'enfant à la naissance (Y) en fonction du poids de la mère (X_1) , de son âge (X_2) et de sa taille (X_3) .

Si la fonction f est affinée (la relation est linéaire) on parle sur la régression linéaire. Alors Cas de régression linéaire simple

$$f(x) = a + bX_{z}$$

Cas de régression linéaire multiple

$$f(x_1, x_2, \dots, x_p) = a + b_1 X_1 + b_2 X_2 + \dots + b_p X_p.$$

1.2 Régression linéaire simple

1.2.1 Modèle de régression linéaire simple

Soit un échantillon de n individus. Pour chaque individu, on a les observations x_i et y_i , i = 1, 2, ..., n qu'elles sont les valeurs des réalisations des variables quantitatives X et Y respectivement.

Le modèle de régression linéaire simple est de la forme suivante :

$$Y = a + bX + \varepsilon. \tag{1.1}$$

Pour la $i^{\grave{e}me}$ observation, on peut réécrire le modèle (1.1) sous la forme

$$y_i = a + bx_i + \varepsilon_i$$
, pour $i = \overline{1, n}$. (1.2)

Les hypothèses relatives à ce modèle sont

 $i) \ E(\varepsilon_i) = 0;$ $ii) \ var(\varepsilon_i) = \sigma^2 < \infty \qquad \forall i = \overline{1, n};$ $iii) \ cov(\varepsilon_i, \varepsilon_j) = 0 \qquad \forall i \neq j.$

De plus, une hypothèse complémentaire pour les inférences : les variables aléatoires ε_i sont normalement distribuées, alors $\varepsilon_i \sim N(0, \sigma^2)$.

1.2.2 Régression linéaire simple sous SPSS

Exemple 1.2 (Exercice 13 sérieTP3) (Régression linéaire simple)

Dans le cadre de travaux de recherche sur la Biomasse (mg), d'un certain type de plante, en fonction de la concentration de l'Azote NH_4^+ (μ mol), nous avons réalisée des expériences dont la biomasse moyenne (Y) ainsi que la concertation du l'Azote (X) en question sont données dans le tableau ci-dessus :

Concentration μmol	0	100	200	400	600
Biomasse mg	305	378	458	540	565

Afin de modéliser ces données, nous avons proposé le modèle linéaire suivant :

$$Y = a + bX + \varepsilon.$$

Questions :

1. Pour un seuil de risque $\alpha = 5\%$, le modèle proposé est-il pertinent ?

2. Donner les estimations des paramètres a et b et donner la droite de régression.

- 3. Donner le coefficient de corrélation linéaire. Que peut-on conclure ?
- 4. Quelle Biomasse prévoyez-vous à une concentration 500 μ mol?

Avant de répondre aux questions, on peut **présenter graphiquement le nuage des points** (x_i, y_i) pour faire une idée préliminaire sur la distribution de ces points et si le modèle linéaire (modèle proposé) peut décrire les données.

Dans cet exemple on veut étudier

La biomasse (Y) d'une plante en fonction de la concentration de l'Azote NH_4^+ (X), alors

- La variable dépendante : la biomasse;
- La variable indépendante : la concentration de l'Azote NH_4^+ .

Pour répondre aux questions de cet exercice sous SPSS, il faut suivre les étapes suivantes

Etape 1. Saisie des données

Entrez les données dans SPSS, dont vous avez deux variables quantitatives Y et X à définir séparément dans SPSS (voir figure 1.1).

Fichier	Edition Affichae	e <u>D</u> onnées	Transform	er <u>A</u> nalyse	Marketing of	direct <u>G</u> raph	es <u>U</u> tilitaire	s Fenêtre	Alde			
a 1			~		H T		4	3	14 14	•	-	
	BIOMASSE	CONCENTR ATION	HauteurH	DiamètreD	Absorbance Y	Concentrati onX	HauteurNei ge	Altitude	Rugosité	Pente	Orientation	Latitude
1	305,00	,00	9,2073	,1999	,000	0	95	2768	252	22	324	87602
2	378,00	100,00	9,6794	,3012	,205	20	150	4208	333	29	308	87601
3	458,00	200,00	10,8049	,3791	,331	40	4	4045	62	5	349	87601
4	540,00	400,00	13,4637	,6005	,515	60	0	4572	85	8	14	87601
5	565,00	600,00	14,1540	,6570	,584	80	0	4614	115	10	63	87601
6			-		,671	100	80	4321	176	16	130	87600
7		-	-				95	3886	72	6	199	87600
8			-			-	20	4206	57	5	32	87600
9			-				90	4192	266	23	197	87599
10			-				10	4051	69	6	113	87599
11			-				10	3746	62	5	149	87599
12			-				50	3789	42	3	218	87598
13			-				45	3771	44	4	53	87598
14			-				60	3796	48	4	101	87598
15							55	3885	77	7	332	87598
16							3	4295	113	10	18	87597
17			-				33	4467	147	13	50	87597
18			-		D ·		0	4764	12	1	276	87597
19							35	4313	38	3	350	87597
20			-				45	4387	40	3	46	87596
21												
	1											

FIG. 1.1 – Saisie des données sous SPSS.

Remarque 1.2 Il faut sauvegardez votre fichier.

Etape 2. Présentation graphique des données (Nuage de points)

a- Allez à $Barre de menus \longrightarrow Graphes \longrightarrow Boîtes dedialogue ancienne version$

puis cliquez sur $Dispersion/points \rightarrow Dispersion simple \rightarrow Définir$

()					Série	3 TP3.sav [E	insemble_de	e_données1] - IBM SPS	S Statistics	Editeu	ur de données
<u>Fichier</u>	Edition Affichag	ge <u>D</u> onnées	Transform	ner <u>A</u> nalyse	Marketing of	direct <u>G</u> raph	nes <u>U</u> tilitaire	es Fenêtre	Aide			
		,	∼ 📱				Générateur de Gélecteur de n	diagrammes nodèles de re	 présentations	graphiques		
		E	Boîtes de dia <u>l</u> o	gue ancienne	e version		*	Bâtons				
	BIOMASSE	CONCENTR ATION	HauteurH	DiamètreD	Absorbance Y	Concentrati onX	Hauteurivei ge	Altitude	Rugosite	Pente	Orient	III Bâtons <u>3</u> D
1	305,00	,00	9,2073	,1999	,000	0	95	2768	252	22		Courbes
2	378,00	100,00	9,6794	,3012	,205	20	150	4208	333	29		Aires
3	458,00	200,00	10,8049	,3791	,331	40	4	4045	62	5		Secteur
4	540,00	400,00	13,4637	,6005	,515	60	0	4572	85	8		Plafond-plancher
5	565,00	600,00	14,1540	,6570	,584	80	0	4614	115	10		I Boîte à moustaches
6					,671	100	80	4321	176	16		Barre d'erreur
7							95	3886	72	6		Pyramide de population
8							20	4206	57	5		
9							90	4192	266	23		
10							10	4051	69	6		Histogramme
11							10	3746	(P)		Disp	ersion/Points ×
12							50	3789		Dianamian		
13							45	3771	5	simple	15	Matrice Point simple
14							60	3796		Dispersion de	time L	Dispersion
15							55	3885	898 C	Superpos	100	3D
16							3	4295				
										l	Definir	Annuler Aide

FIG. 1.2 – Présentation graphique des données (Etape 2 partie a).

b- Dans la boîte de dialogue qui va apparaître (figure 1.3) insérez **la variable dépendante** dans la case **Axe des Y** et **la variable indépendante** dans la case **Axe des X** puis cliquez sur **OK**.

FIG. 1.3 – Présentation graphique des données (Etape2 partie b).

une fois que vous exécutez l'étape 2, vous obtiendrez la figure 1.4.



FIG. 1.4 – Nuage de points présente la Biomasse en fonction de la Concentration d'azote.

Au vue du graphique (figure 1.4), il semble que le modèle linéaire est adéquat pour l'explication de biomasse en fonction de la concentration de l'Azote (car le nuage des points est distribué sous une forme linéaire).

Remarque 1.3 Vous pouvez ajouter la droite de régression au nuage de points en suivant ce qui suit

1- cliquez deux fois sur le nuage de points une fenêtre va s'ouvrir appelée "**Editeur de** diagrammes";

2- sur la fenêtre "Editeur de diagrammes" cliquez sur Eléments \longrightarrow ajouter une courbe d'ajustement au total (voir figure 1.5);

3- une autre fenêtre va s'ouvrir appelée **"propriété"** fermez cette fenêtre puis fermez la fenêtre d'Editeur de diagrammes (voir figure 1.6).



FIG. 1.5 – Ajout du droite de régression au nuage de points.

En suivant les étapes citées, vous obtiendrez alors la figure 1.7.



FIG. 1.6 – Ajout du droite de régression au nuage de points.



FIG. 1.7 – Nuage de points et droite de régression.

Etape 3. Réalisation de la régression linéaire sous SPSS

Pour obtenir une régression linéaire simple il faut suivre ces étapes

1- Sélectionnez sur la barre de menu

 $Analyse \longrightarrow R\acute{e}gression \longrightarrow Lin\acute{e}aire$

Eichier Eo	dition Affichag	ge <u>D</u> onnées	Transform	ier	Analyse Ma Rapports Statistics	rketing direct	<u>G</u> rap	hes U		S Fenêtre	Aide			ui de (Jonnees
1 2 3	BIOMASSE 305,00 378,00 458,00	CONCENTR ATION ,00 100,00 200,00	HauteurH 9,2073 9,6794 10,8049	Dia	Tableau Compare Modèle I Modèles Modèles	er les moyenn inéaire génér: linéaires gén Migtes	es al éralisé:	* * * *	Jei 95 50 4	Altitude 2768 4208 4045	Rugosité 252 333 62	Pente 22 29 5	Orient	324 308 349	Latitude 87602 87601 87601
4 5 6 7 8 9 10 11 12 13 14 15 16	540,00 565,00 	400,00 600,00	13,4637 14,1540		Régress Log Liné Réseau Classific Réductio Echelle Tests po Prevision Survie Réponse Malayse	ion aire neuronaux ation n des dimens n paramétriqu ns es myltiples des valeurs m on multiple	iions Jes Janquar	F F F F F F F F		Modélisatio Linéaire Ajustement Moindres c Logistique Orgistique Orginale Modèles de Non linéair Pondératio Doubles m	n linéaire auto de fongtions arrés partiels binaire muttinomiale e.choix binaire. e n estimée oindres carrés		2	14 63 130 199 32 197 113 149 218 53 101 332 18	87601 87600 87600 87600 87599 87599 87599 87598 87598 87598 87598
17 18 19	· ·				Contrôle	Echantilions complexes Contrôle de gualité Courbe ROC		;	0	Codage op 4764 4313	12 38	3) 1 3		50 276 350	87597 87597 87597
20 21	des données	Affichage des	, variables		-				45	4387	40	3		40	87396

FIG. 1.8 – Procédures de la réalisation de la régression linéaire simple (Partie 1).

2- Dans la boîte de dialogue de la figure 1.9 apparaît : sélectionnez, dans la liste des variables, les deux variables que vous souhaitez à analyser, et mettez, en cliquant sur les flèches, la variable dépendante (variable à expliquer) dans la case Dépendant et la variable indépendante (variable explicative) dans la case Variables indépendantes, puis cliquez sur OK.



FIG. 1.9 – Procédures de la réalisation de la régression linéaire simple (Partie 2).

L'application de ces étapes sur les données de l'exercice permet d'obtenir les résultats (4 tableaux) qui sont présentés dans la figure 1.10. Ces résultats contiennent des réponses sur les questions 1, 2, 3 et 4.



FIG. 1.10 – Résultats de la régression linéaire simple.

Modèle	Variables introduites	Variables supprimées	Méthode
1	Concentration d'azote ^b		Entrée

Variables introduites/supprimées^a

a. Variable dépendante : Biomasse d'une plante

b. Toutes variables requises saisies.

FIG. 1.11 – Tableau récapitule les variables explicatives introduites dans le modèle.

	Recapitulatil des modèles											
Modèle		R		R-deux	R-deux ajusté	Erreur standard de l'estimation						
1		,956ª	Π	,914	,885	37,01895						

a. Valeurs prédites : (constantes), Concentration d'azote

FIG. 1.12 – Tableau des coefficients de corrélation et de détermination.

Interprétation des Résultats obtenus

Premier tableau (figure 1.11) : ce tableau récapitule les variables explicatives prises en compte dans le modèle. Ici, il n'y a qu'une seule variable dans la case "variables introduites" (Concentration de d'Azote) tandis qu'il n'ya pas des variables supprimées puisque nous travaillons sur une régression simple.

Deuxième tableau (figure 1.12) : il donne deux valeurs importantes dans le modèle de régression :

• Le coefficient de corrélation : R = 0.956. Donc, le coefficient de corrélation est presque égal à 1, ce qui indique qu'il y a une forte liaison linéaire entre la biomasse (Y) et la concentration de l'azote (X). Le signe positif de R indique que les deux variables varient dans le même sens.

• Le coefficient de détermination : $R^2 = 0.914$, ce qui indique que 91.4% de la variation totale de Y est expliquée par le modèle de régression sur X.

Troisième tableau (figure 1.13) : c'est la table d'analyse de la variance, il indique si le modèle est valide ou non à partir d'un test sur la pente de la droite de régression. C'est à dire, il nous permet de répondre au test d'hypothèses suivant :

$$\begin{cases} H_0: & b=0 \ ((\text{le modèle n'est pas valide}); \\ H_1: & b\neq 0 \ (\text{le modèle est valide}). \end{cases}$$

Alors, à l'aide de la valeur de signification (Sig) et au seuil de risque α , on décide de

$$\begin{cases} \text{ne pas rejeter } H_0 & \text{si } \alpha < Sig; \\ \text{rejeter } H_0 & \text{si } \alpha \ge Sig. \end{cases}$$

Ce tableau donne les valeurs suivantes

- La somme des carrés des résidus (variation qui n'est pas expliquée par le modèle de régression)

$$SCR = 4111.207.$$

- La somme des carrés de régression (la variation expliquée par la régression)

$$SCE = 43483.593.$$

- La réalisation de la statistique de Fisher

$$f_c = 31.731.$$

- La valeur de signification : Sig = 0.011. Il résulte de cette valeur et pour un seuil de risque $\alpha = 5\%$ que le modèle obtenu est pertinent (valide) car $\alpha >$ Sig (0.05 > 0.011), c'est-à-dire, il existe une relation linéaire statistiquement significative entre la Biomasse de la plante et la concentration de l'azote donnée par l'équation : $\hat{Y} = \hat{a} + \hat{b}X$, où \hat{a} et \hat{b} sont donnés dans le quatrième tableau.

Modèle	Somme des carrés	ddl	Moyenne des carrés	D	Sig.
1 Régression	43483,593	1	43483,593	31,731	,011 ^b
Résidu	4111,207	3	1370,402		
Total	47594,800	4			

ANOVA^a

a. Variable dépendante : Biomasse d'une plante

b. Valeurs prédites : (constantes), Concentration d'azote

FIG. 1	1.13 -	Table	de	L'ANOVA
--------	--------	-------	----	---------

Quatrième tableau (figure 1.14) : il donne les estimations des paramètres a et b (coefficients de la droite de régression) dits "A" dans SPSS (Partie encadrée) suivantes :

- La valeur de la constante ou l'ordonnée à l'origine $\hat{a} = 336.638$;
- La valeur de la pente $\hat{b} = 0.433$.

Alors l'équation de la droite de régression sera donnée par

 $\hat{Y} = 336.638 + 0.433X.$

			efficients n	on standardisés	Coefficients standardisés		
Modèle			А	Erreur standard	Bêta	t	Sig.
1 (0	Constante)		336,638	25,950		12,973	,001
С	concentration d'azote		,433	,077	,956	5,633	,011

Coefficients^a

a. Variable dépendante : Biomasse d'une plante

FIG. 1.14 – Tableau des coefficients de la régression linéaire simple.

Pour la dernière question, on veut prédire, sur la base de notre modèle, la biomasse (Y)pour une concentration d'azote égale à 500 μ mol, il suffit alors de substituer la valeur de la concentration d'azote ($X = 500 \ \mu$ mol) dans la dernière équation pour trouver **la valeur** **prédite** (\hat{Y}) de la biomasse de la plante

$$\hat{Y} = 336.638 + 0.433(500) = 553.138.$$

1.3 Régression linéaire multiple

Les idées utilisées en régression multiple sont les mêmes que celles qu'on a vu en régression linéaire simple. La régression linéaire multiple diffère de la régression linéaire simple par le nombre de variables explicatives présentes dans le modèle.

1.3.1 Modèle de régression linéaire multiple

Le modèle de régression linéaire multiple est de la forme suivante :

$$Y = a + b_1 X_1 + b_2 X_2 + \dots + b_p X_p + \varepsilon.$$
(1.3)

Pour la i^{eme} observation, le modèle (1.3) peut être représenté de la manière suivante

$$Y_i = a + b_1 x_{1i} + b_2 x_{2i} + \dots + b_p x_{pi} + \epsilon_i, \qquad i = 1, n;$$
(1.4)

où :

 Y_i est la valeur de la variable dépendante Y (variable quantitative),

 $x_{1i}, x_{2i}, ..., x_{pi}$ sont les valeurs de p autres variables quantitatives (variables explicatives) $X_1, ..., X_p$, pour i = 1, ..., n.

les ε_i sont les termes des erreurs.

Les hypothèses relatives à ce modèle sont

- i) $E(\varepsilon_i) = 0;$
- *ii*) $var(\varepsilon_i) = \sigma^2 < \infty$ $\forall i = \overline{1, n};$
- *iii*) $cov(\varepsilon_i, \varepsilon_j) = 0$ $\forall i \neq j;$

iv) Les termes x_j $(j = \overline{1, p})$ étant déterministes;

de plus, une hypothèse complémentaire pour les inférences : les variables aléatoires ε_i sont normalement distribuées, alors $\varepsilon_i \sim N(0, \sigma^2)$.

On peut aussi écrire le modèle (1.4) sous sa forme matricielle :

$$Y = Xb + \varepsilon;$$

оù

- Y est un vecteur aléatoire de dimension n,
- X est une matrice de taille $n \times (p+1)$ connue,
- β est le vecteur de dimension (p+1) des **paramètres inconnus** du modèle,

• ε est le vecteur de dimension n des erreurs.

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} \end{bmatrix}, b = \begin{bmatrix} a \\ b_1 \\ \vdots \\ b_p \end{bmatrix},$$
$$et \ \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

1.3.2 Régression linéaire multiple sous SPSS

Exemple 1.3 On étudie l'influence des heures de travail (X_1) et du capital utilisé (X_2) sur la production industrielle (Y) ou d'une autre façon, on cherche à établir une relation entre la

Entronrise (i)	Travail (heures)	Capital (machines/heures)	Production (100 tonnes)
	(x_{1i})	(x_{2i})	(Y_i)
1	1100	300	60
2	1200	400	120
3	1430	420	190
4	1500	400	250
5	1520	510	300
6	1620	590	360
7	1800	600	380
8	1820	630	430
9	1800	610	440

production, les heures de travail et le capital utilisé. Pour cela, on dispose des observations de 9 entreprises résumées dans le tableau ci-dessous

On suppose que la production (Y) est expliquée par un modèle de régression linéaire multiple avec deux variables explicatives, le travail (X_1) et le capital (X_2) , c'est à dire par le modèle

$$Y = a + b_1 X_1 + b_2 X_2 + \varepsilon.$$

Questions :

- 1. Donner les estimations des paramètres du modèle proposé.
- **2.** Pour un seuil de risque $\alpha = 5\%$, le modèle proposé est-il pertinent ?
- 3. Tester l'hypothèse nulle $H_0: b_j = 0$ contre l'alternative $H_1: b_j \neq 0$ pour j = 1; 2.
- **4.** Tester l'hypothèse nulle H_0 : a = 0 contre l'alternative H_1 : $a \neq 0$.

La méthode de la régression linéaire multiple se réalise sous le logiciel SPSS en générale en effectuant presque les mêmes étapes de la régression linéaire simple :

Etape 1. Saisie des données

Dans cet exercice on a

- Une variable dépendante ou variable à expliquée qui est la production;
- Deux variables explicatives qu'elles sont : le travail (X_1) et le capital (X_2) .

Entrez les données dans SPSS, dont vous avez 3 variables quantitatives à définir séparément dans SPSS (voir figure 1.15).

¢.				EXEMP	LE POUR	R REGR	ESSION	N MULT	IPLE.sa	av (Ens	emble_	_de_dor	nées0] - IBN	A SPSS St	atistics Edited	ur de donnée
Eichier	Edito	on Affichage	Données Tra	nsformer Ana	ilyse <u>M</u> a	rketing	direct	Graphes	<u>U</u> tili	aires	Fenêtre	Aide				
)		iii 📩	=	h	*,	¥		4		▲ 1여	0	A86		
													-			
		Production	HeursTravail	Capital	var		var	var		var		var	var	var	var	var
1		60	1100	300												
2		120	1200	400												
3		190	1430	420												
- 4		250	1500	400												
5		300	1520	510												
6		360	1620	590												
7		380	1800	600												
8		430	1820	630												
9		440	1800	610												
10																
11																
12																
13																
14																
15																
16																
17																
18																
19																
20																
21																
		4	_	_			_			_	_					
Afficha	Affichana der dennén - Affichana des variables															
Amena	ige des	s connees Au	chage des valiabl													

FIG. 1.15 – Saisie des données.

Etape 2. Réalisation de la régression linéaire multiple sous SPSS

Pour obtenir une régression linéaire multiple il faut suivre ces étapes

1- Sélectionnez sur la barre de menu

$$Analyse \longrightarrow R\acute{e}gression \longrightarrow Lin\acute{e}aire$$

Eches Edition Addichage Données Transformer doalyze Marates galact Graphes Usitaires Fender Add Rapports Rappo	@	EXEMPLE POUR REGRESSION MULTIPLE.sav [Ensemble_de_données0] - IBM SPSS Statistics Editeur de données															
Production HeursTravial Capital 1 60 1100 3 190 1430 4 220 100 5 300 1500 6 360 1600 6 360 1600 7 380 1000 8éaux neuronaux Image automatique. Image automatique. 10 Image automatique. Image automatique. 10 Image automatique. Image automatique. 110 Image automatique. Image automatique. 12 Image automatique. Image automatique. 12 Image automatique. Image automatique. 12 Image automatique. Image automatique. 13 Image automatique. Image automatique. 14 Image automatique. Image automatique. 15 Image automatique. Image automatique. 14 Image automatique. Image automatique. 15 Image automatique. Image automatique. 16 Image automatique. Image automatique. 16 <td< th=""><th>Eichier</th><th>Edition</th><th>Affichage</th><th>Données Tr</th><th>ansformer</th><th>Analyse</th><th>Marketing direct</th><th>Graphes</th><th>Us</th><th>litaires Fer</th><th>lêtre Aide</th><th></th><th></th><th></th><th></th><th></th><th></th></td<>	Eichier	Edition	Affichage	Données Tr	ansformer	Analyse	Marketing direct	Graphes	Us	litaires Fer	lêtre Aide						
Tableaur Production HeursTraval Capita 1 60 1100 0 10 10) 🛄		1	Rag	ports stiqu <u>e</u> s descriptiv	es	;	42	1 1 1	@	ARG				
Production HeursTraval Compare les moyennes Var		Tableaux															
1 60 1100 Modèle linéaire général Image: Section of the section o		Pr	oduction	HeursTravail	Capital	Com	parer les moyen	165		var	var	var	var	var	var	var	var
2 120 1200 Modèles linéaires généraisés Image: Section of the sec	1		60	1100		Mod	èle linéaire génér	al									
3 190 1430 Modèles Migles Image: Section and the se	2		120	1200		Mod	èles linéaires gér	véralisés									
4 250 1500 Qorrélation Image: State in the s	3		190	1430		Mod	èles Migles										
5 300 1520 Régression Itel Modélisgion linéaire automatique 6 360 1620 Log Linéaire Itel Modélisgion linéaire automatique 7 380 1800 Réseaux neuronaux Itel Modélisgion linéaire automatique Itel Modélisgion linéaire automatique 9 440 1800 Régestion Itel Modélisgion linéaire automatique Itel Modélisgion linéaire automatique 10 120 Classification Itel Modélisgion linéaire automatique Itel Modélisgion linéaire automatique 11 120 1800 Régestion automatique Itel Modélisgion linéaire automatique Itel Modélisgion linéaire 112 11 110 110 Itel Modélise de Choix binaire 114 115 Itel Modélises des valeurs manquantes Itel Modélise de Choix binaire Itel Modélise de Choix binaire Itel Modélise de Choix binaire 116 Itel Modélise de choix binaire	4		250	1500		Com	élation										
6 360 1620 Lgg Linéaire Image: Control of the second secon	5		300	1520		Rég	ression			Modél	sation linéair	e automatique					
7 380 1800 Réseaux neuronaux Image: Austement de fongtions 8 430 1820 Classification Image: Austement de fongtions 9 440 1800 Réguidon des dimensions Image: Austement de fongtions 10 Image: Austement de fongtions Image: Austement de fongtions Image: Austement de fongtions 10 Image: Austement de fongtions Image: Austement de fongtions Image: Austement de fongtions 10 Image: Austement de fongtions Image: Austement de fongtions Image: Austement de fongtions 110 Image: Austement de fongtions Image: Austement de fongtions Image: Austement de fongtions 12 Image: Austement de fongtions Image: Austement de fongtions Image: Austement de fongtions 12 Image: Austement de fongtions Image: Austement de fongtions Image: Austement de fongtions 13 Image: Austement de fongtions Image: Austement de fongtions Image: Austement de fongtions 14 Image: Austement de fongtions Image: Austement de fongtions Image: Austement de fongtions 16 Image: Austement de fongtions Image: Austement de fongtions Image: Austement de fongtions 19 </th <th>6</th> <th></th> <th>360</th> <th>1620</th> <th></th> <th>Log</th> <th>Linéaire</th> <th></th> <th></th> <th>Linéai</th> <th>re .</th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th>	6		360	1620		Log	Linéaire			Linéai	re .						
8 430 1820 Classification Image: Classification Image	7		380	1800		Rés	eaux neuronaux			Aiuste	ment de fonct	tions	2				
9 440 1800 Réguction des dimensions Image: Status and the status	8		430	1820		Classification Réguction des dimensions				E Haind	ne comie es	et ale					
10 Echelle Image: Logistique binaire Image: Logistique binaire 11 Tests non paramétriques Image: Logistique multinomiale Image: Logistique multinomiale 12 Previsions Image: Logistique multinomiale Image: Logistique multinomiale 13 Orginale Image: Logistique multinomiale Image: Logistique multinomiale 13 Survie Image: Logistique multinomiale Image: Logistique multinomiale 14 Survie Image: Logistique multinomiale Image: Logistique multinomiale 15 Qurvie Image: Logistique binaire Image: Logistique binaire 16 Image: Logistique binaire Image: Logistique binaire Image: Logistique binaire 16 Image: Logistique binaire Image: Logistique binaire Image: Logistique binaire 17 Echantijion multiple Image: Logistique binaire Image: Logistique binaire Image: Logistique binaire 18 Controlle de gualité Image: Logistique binaire Image: Logistique binaire Image: Logistique binaire Image: Logistique binaire 20 Image: Logistique binaire Image: Logistique binaire Image: Log	9		440	1800						To monores carres parents							
11 Tests non paramétiques Image: Control of the gualité Image: Control of the gualité 12 Previsions Image: Control of the gualité Image: Control of the gualité Image: Control of the gualité 13 Image: Control of the gualité 14 Image: Control of the gualité 16 Image: Control of the gualité 19 Image: Control of the gualité 20 Image: Control of the gualité 21 Image: Control of the gualité Image: Co	10					Ech	olle			Logist	que binaire						
12 Prevásions Image: Control of the guardé	11					Test	s <u>n</u> on paramétriq	ues		Logist	ique <u>m</u> ultinon	miale					
13 Sunvie Image: Sunvie	12					Prev	isions			Crgina Orgina	ile						
14 Réponses multiples Imputation multiple Imputation multiple 16 Imputation multiple Imputation multiple Imputation multiple 17 Imputation multiple Imputation multiple Imputation multiple 18 Imputation multiple Imputation multiple Imputation multiple 19 Imputation multiple Imputation multiple Imputation multiple 20 Imputation multiple Imputation multiple Imputation multiple 21 Imputation	13					Surv	ie			Modèle	es de choix bi	inaire					
15 Implified analyse des valeurs manquantes Implified nultiple Implified nultiple 16 Implified nultiple Implified nultiple Implified nultiple 17 Implified agualité Implified nultiple Implified nultiple 18 Implified agualité Implified nultiple Implified nultiple 19 Implified agualité Implified nultiple Implified nultiple 20 Implified agualité Implified nultiple Implified nultiple 21 Implified agualité Implified nultiple Implified nultiple 4 Implified agualité Implified nultiple Implified nultiple	14					Rép	onses multiples			Non lin	néaire						
16 Imputation multiple Imputation multiple 17 Echantijions complexes Cgdage optimal (CATREG) 18 Contrôle de gualté Cgdage optimal (CATREG) 19 Courbe ROC Courbe ROC 20 Courbe ROC Courbe ROC	15					Anal	yse des valeurs n	nanquantes		R Pondé	ration estimé	ie					
17 Echantijions complexes Cgdage optimal (CATREG) 18 Contrôle de gualté Cgdage optimal (CATREG) 19 Image: Contrôle de gualté Image: Contrôle de gualté 20 Image: Contrôle de gualté Image: Contrôle de gualté 20 Image: Contrôle de gualté Image: Contrôle de gualté 20 Image: Contrôle de gualté Image: Contrôle de gualté 20 Image: Contrôle de gualté Image: Contrôle de gualté 21 Image: Contrôle de gualté Image: Contrôle de gualté 4 Image: Contrôle de gualté Image: Contrôle de gualté	16					Imp	station multiple			Double	es moindres	carrés					
18 Contrôle de gualté Contrôle de gualté 19 Image: Contrôle de gualté Image: Contrôle de gualté 20 Image: Contrôle de gualté Image: Contrôle de gualté 20 Image: Contrôle de gualté Image: Contrôle de gualté 20 Image: Contrôle de gualté Image: Contrôle de gualté 20 Image: Contrôle de gualté Image: Contrôle de gualté 20 Image: Contrôle de gualté Image: Contrôle de gualté 20 Image: Contrôle de gualté Image: Contrôle de gualté 20 Image: Contrôle de gualté Image: Contrôle de gualté 20 Image: Contrôle de gualté Image: Contrôle de gualté 21 Image: Contrôle de gualté Image: Contrôle de gualté 4 Image: Contrôle de gualté Image: Contrôle de gualté 4 Image: Contrôle de gualté Image: Contrôle de gualté 4 Image: Contrôle de gualté Image: Contrôle de gualté 4 Image: Contrôle de gualté Image: Contrôle de gualté 4 Image: Contrôle de gualté Image: Contrôle de gualté 4 Image: Contrôle de gualté Image: Contrôle de gualté 4 Image: Contrôle de gualté Image: Contrôle de gualté	17					Ech	antijions complex	es		Codad	e optimal (C	ATREG					
19 Courbe ROC	18					Con	trôle de gualité			01001	to the second second						
20 21 4 20 21 4 20 20 21 20 20 20 20 20 20 20 20 20 20 20 20 20	19					Cou	rbe ROC										
21 Affichage des variables I i nésien	20					_			-	1							
Affichage des variables	21																
Affichage des données Affichage des variables																	
	Affich	Affichage des données Affichage des variables															
LE DIOLESSEUR IDM OF OS OURSILS	Linéaire	l a processour BM OPCC Createring															

FIG. 1.16 – Réalisation de la régression linéaire multiple sous SPSS (Partie 1).

2- Dans la boîte de dialogue de la figure 1.17 apparaît, il faut réaliser ce qui suit :

a- Transférez dans la liste des variables la variable dépendante (variable à expliquer) dans la case Dépendant ;

b- Transférez les deux variables indépendantes (variables explicatives) dans la case
Variables indépendantes;

c- Dans la case "Méthode" laissez la méthode par défaut c'est à dire "Entrée". Cette méthode (Entrée) est choisie parmi d'autres méthodes de sélection des variables selon notre but dans l'exercice.

d- Cliquez sur OK

Endler Edition Affrehage Données Transformer Analyse Marketing direct Graphes Utiliaires Fenétre Ade Image: Strategy and the strate	ta			EXEMPL	E POUR REGRESSION MULTIPLE.sav [Ensemble_de_données0] - IBM SPSS Statistics Editeur de données
Production HeursTraval Capital Var V	Fichier	Edition Affichage	Données Tra	nsformer Anal	yse Marketing direct Graphes Utilitaires Fenêtre Aide
Production HeursTravail Capital Var					💷 M 🚟 🖾 📟 🖧 🗮 🔐 💊 🧠 🥗
Production HeursTravail Capital var var <th></th> <td></td> <td></td> <td></td> <td></td>					
1 60 1100 300 Régression linéaire X 2 120 1200 400 3 199 1430 420 4 250 1500 400 5 300 1520 510 6 360 1620 599 7 380 1800 600 8 433 1820 630 9 440 1800 610 10 - - - 11 - - - 12 - - - - 13 - - - - 14 - - - - 16 - - - - 16 - - - - - 18 - - - - - - 19 - - - - - - - 20 - - - - - -<		Production	HeursTravail	Capital	var var var var var var var var var
2 120 1200 400 3 190 1430 420 4 250 1500 400 5 300 1520 510 6 360 1620 590 7 380 1800 600 9 440 1800 610 10	1	60	1100	300	🔅 Régression linéaire 💌
3 190 1430 420 Perendant. Biagrammes. Enregister Diagrammes. Enregister Options 5 300 1520 510 Gained and and and and and and and and and an	2	120	1200	400	Disastati
4 250 1500 400 Image: Capital Capit	3	190	1430	420	Production
5 300 1520 510 6 360 1620 590 7 380 1800 600 8 430 1820 630 9 440 1800 610 10	4	250	1500	400	Capital Diagrammes
6 360 1620 590 7 380 1800 600 8 430 1820 630 9 440 1800 610 10	5	300	1520	510	Enregistrer
7 380 1800 600 8 430 1820 630 9 440 1800 610 10	6	360	1620	590	Précédent Suivant Options
8 430 1820 630 9 440 1800 610 10	7	380	1800	600	Variables indépendantes : Bootstrap
9 440 1800 610 10 Image: Contraction of the second and the se	8	430	1820	630	Capital D
10 11 12 13 13 14 15 16 17 18 19 20 21	9	440	1800	610	
11 12 13 14 14 15 16 16 17 18 19 20 21	10				
12 13 14 14 15 16 16 17 18 19 20 21	11				Méthode : C Entrée
13 Image: Second and technic and	12				Variable de filtrage : Pas à pas
14 15 16 16 17 18 19 20 21	13				Eliminer bloc
15 Ascendante 16 Poids WLS : 17 Image: Control of the second seco	14				Etiquettes d'observation : Descendante
16 Poids WLS: 17 d ОК Соцег Réinitialiser Annuler Aide 19	15				Ascendante
17 18 19 20 21	16				Poids WLS :
18 d OK Coller Réinitialiser Annuler Aide 19 20 21	17				
19 20 21	18				OK Coller Réinitialiser Annuler Aide
20 21	19				
	20				
	21				
	-				
Milebone dae dae dae dae dae variablee	Aller	and an densities of the	ichono dec usrich	lac	

FIG. 1.17 – Réalisation de la régression linéaire multiple sous SPSS (Partie 2).

L'application de ces étapes sur les données de l'exemple permet d'obtenir les résultats (4 tableaux) qui sont présentés dans la figure 1.18. Ces résultats contiennent des réponses sur les questions posées.



FIG. 1.18 – Résultats de la régression linéaire multiple sous SPSS.

Interprétation des résultats obtenus

Le premier tableau de la figure 1.18 présente les variables introduites dans le modèle de régression (le capital et l'heur de travail) et les variables qui sont exclues de l'entrée dans le modèle : ici la méthode "entrée" n'exclut pas les variables.

- Du deuxième tableau de la figure 1.18 on constate que les deux variables prises en compte (ou le modèle) expliquent 97.8% ($R^2 = 0.978$) de la variance de la variable production.

- Pour la première question : les estimations des paramètres a, b_1 et b_2 sont données dans le dernier tableau de la figure 1.18, alors

- La valeur de la constante $\hat{a} = -437.714$;
- La valeur du coefficient de la variable Travail (X_1) : $\hat{b}_1 = 0.337$.

• La valeur du coefficient de la variable Capital (X_2) : $\hat{b}_2 = 0.41$.

Alors le modèle de la régression linéaire multiple sera donné par

$$\hat{Y} = -437.714 + 0.337X_1 + 0.41X_2.$$

Remarque 1.4 A la base d'un modèle de régression linéaire multiple, on peut prédire la valeur de la variable Y si on a des valeurs données pour les variables explicatives, donc il suffit de substituer ces valeurs dans l'équation du modèle obtenu.

 Dans la deuxième question, on veut réaliser le test suivant (dit Test de significativité globale du modèle)

$$\begin{cases} H_0: \quad b_1 = b_2 = 0 \\ H_1: \quad \exists j/b_j \neq 0 \ (j = 1, 2) \end{cases};$$

et à l'aide de la valeur de signification (Sig) et au seuil de risque α , on décide de

1

$$\begin{cases} \text{ne pas rejeter } H_0 & \text{si } \alpha < Sig; \\ \text{rejeter } H_0 & \text{si } \alpha \ge Sig. \end{cases}$$

Alors, de la table de l'ANOVA (troisième tableau de la figure 1.18) et au seuil de risque $\alpha = 5\%$, on constate qu'on peut rejeter H_0 car $\alpha > Sig$ (0.05 > 0), donc le modèle obtenu est pertinent (valide).

- Les questions 3 et 4 concernant les tests de nullité de chacun des paramètres du modèle de la régression (Tests de significativité des paramètres du modèle). La réponse à ces questions se trouve dans la colonne signification du quatrième tableau de la figure 1.18, donc au seuil de risque $\alpha = 5\%$, on constate que

• pour le test d'hypothèse nulle $(H_0 : a = 0)$, on rejette H_0 car $\alpha > Sig_0$ (0.05 > 0) (voir la première ligne du tableau 4), alors a est significativement différente de 0.

• pour le test d'hypothèse nulle $(H_0 : b_1 = 0)$, on rejette H_0 car $\alpha > Sig_1$ (0.05 > 0.009) (voir la deuxième ligne du tableau 4), alors b_1 est significativement différente de 0..

• pour le test d'hypothèse nulle $(H_0 : b_2 = 0)$, on ne peut pas rejeter H_0 car $\alpha < Sig_2$ (0.05 < 0.082) (voir la troisième ligne du tableau 4) alors b_2 est significativement égale à 0. Cela veut dire que la variable X_2 (Capital) ne contribue pas significativement à la régression (elle n'a pas une influence significative sur la variable production) donc on peut l'exclue du modèle de la régression linéaire multiple et on fait à nouveau une autre régression en fonction d'une seule variable explicative (X_1) .

Remarque 1.5 Pour choisir automatiquement les variables explicatives ayant une influence sur Y, en choisissant une option "**pas à pas**" dans "Méthode".

t 2				EXEMP	PLE POUR REGRESSION MULTIPLE say (Ensemble de données11 - IBM SPSS Statistics Editeur de donnée
Fichier	Editio	n Affichage	Données Tra	ansformer Ana	alvse Marketing direct Graphes Utilitaires Fenêtre Aide
		0		- E -	🗐 M 🕺 🖳 🐴 🎹 📑 🖓 🔴 🤲
6:					
		Production	HeursTravail	Capital	var var var var var var var var
1		60	1100	300	Régression linéaire
2		120	1200	400	Dépendant:
3		190	1430	420	HeursTravail
4		250	1500	400	Capital Diagrammes
5		300	1520	510	Brésédant Enregistrer
6		360	1620	590	Understein Suivaire Options
7		380	1800	600	Variables independantes : Bootstrap
8		430	1820	630	A Capital
9		440	1800	610	
10					Néthodo : Pasibinas V
11				-	Entrée
12					Variable de filtrage : Pas à pas
13					Eliminer bloc 45
14					Etiquettes d'observation : Descendante
15					Ascendanie
16					Poids WLS :
17					
18					OK Coller Réinitialiser Annuler Aide
19					
20			-		
21					

Afficha	age des	données Affi	chage des variab	les	

FIG. 1.19 – Procédure de la réalisation de la régression linéaire multiple par étapes (en choisissant la méthode "Pas à Pas").

Méthodes de sélection des variables de régression linéaire

La sélection d'une méthode vous permet de spécifier la manière dont les variables indépendantes sont entrées dans l'analyse. En utilisant différentes méthodes, vous pouvez construire divers modèles de régression à partir du même groupe de variables.

Méthode "Entrée" (par défaut) : Méthode qui introduit toutes les variables indépendantes simultanément. A utiliser si on veut déterminer l'équation de la droite de régression avec toutes les variables indépendantes.

Les autres méthodes sont des méthodes hiérarchiques. Seulement à utiliser si on pense qu'une des variables est plus importante que les autres.

Méthode "Pas à pas" : les variables indépendantes sont ajoutées à l'équation une par une et peuvent être enlevées subséquemment si elles ne contribuent plus significativement à la régression. Le processus s'arrête lorsqu'aucune variable ne peut plus être introduite ou éliminée.

Méthode "Eliminer bloc" : toutes les variables dans un bloc sont supprimées en une seule étape.

Méthode "Descendante" : toutes les variables sont entrées initialement dans l'équation et sont ensuite éliminées une à une. La variable ayant la plus petite corrélation avec la variable dépendante est d'abord étudiée pour l'élimination. Si elle est éliminée par le modèle, la prochaine variable avec le plus petit coefficient de corrélation est étudiée, jusqu'à ce qu'aucune variable ne satisfasse plus au critère d'élimination.

Méthode "Ascendante" : les variables sont introduites séquentiellement une par une. Si la première variable est introduite dans l'équation, la variable explicative ne figurant pas dans l'équation et présentant la plus forte corrélation partielle est considérée ensuite. La procédure s'arrête lorsqu'il ne reste plus de variables satisfaisant le critère d'introduction.