Université Mohamed Khider de Biskra. Faculté des Sciences Exactes et de la Vie

Département : Biologie

Module: Biostatistiques.

Année Univ: 2021-2022

Prof Chala Adel

Master 1 LMD Biologie (Tous les spécialités)

Protocole de TP 03 :

Exécution De la Régression Linéaire Simple

Pour valider une modèle de la régression linéaire simple, il faut que les conditions suivantes soient satisfaites :

- a) Deux variables quantitatives mesurées : X=Explicative (indépendante), et la variable à expliquer Y (Dépendante).
- b) La loi des erreurs est Gaussienne.

Il faut comprendre bien que la <u>réponse</u> d'une question posée s'appelle une variable aléatoire.

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

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

Alors, notre objectif dans la curent TP c'est

- 1) Etablir la valeur numérique du coefficient de corrélation entre X et Y (est noté par r).
- 2) Tester la signification du modèle.
- 3) Etablir l'intervalle de confiance pour le coefficient A.
- 4) Etablir le coefficient de détermination de la régression R.
- 5) Etudier la normalité d'erreurs.

I) Brief sur le Modèle de la Régression Linéaire Simple

On peut poser les questions suivantes :

Existe-il une influence de la variable explicative X sur la variation de la variable à expliquer Y ?.

Peut-on assurer que cette influence s'écrit sous forme Linéaire ?.

Voici le tableau des données statistique, les mesures à été fait sur un échantillon de k individus.

Y y ₁ y ₂ y ₃ y ₄ y ₅ y _k	Х	X ₁	X ₂	X ₃	X 4	X 5	 	X _k	
	Y	y 1	У ₂	У 3	y 4	y 5	 	Уĸ	

Pour cela on veut écrire la relation entre les deux variables X et Y sous la forme suivante :

$$Y = f(X).$$

Ou bien pour chaque couple mesurable (xi,yi), avec $i \in \{1,2,3,4,5, ..., k\}$, par la relation

$$y_i = Ax_i + B + \varepsilon_i,.$$

avec ε_i sont les erreurs (les distences qui se trouvent entre le point (xi, yi) et la droite linéare,

Remarque : la loi de la variable aléatoire qui exprime les erreurs suit la loi de Gauss. (Condition « b » pour valider modèle de la régression Linéaire Simple).

Pour cela on doit ordonner les réponses par les étapes suivantes

Proposition d'hypothèses

Hypothèse nulle H₀: (Il n'existe pas une liaison entre X et Y).

Hypothèse alternative H_1 : (Il existe une liaison entre X et Y).

Conclusion (la décision)

Pour la décision, on utilise souvent la règle suivante :

Si Signification inférieure à α %. Alors on rejette H0.

Si Signification supérieure à α%. Alors on accepte H0.

Exemple:

Le tableau suivant, donne les résultats d'une expérience statistique porte sur 10 étudiants, nombre d'heures de la révision dans la semaine avant l'examen, et l'augmentation de notes obtenus sur module Mathématiques.

nombre	10	11	14	15	20	25	46	50	59	70
d'heures										
augmentation	10	10	12	12	13	13	19	15	16	20
de notes										

- 1) Etablir la valeur numérique du coefficient de corrélation entre X et Y (est noté par r).
- 2) Tester la signification du modèle.
- 3) Etablir l'intervalle de confiance pour le coefficient A.
- 4) Etablir le coefficient de détermination de la régression R.
- 5) Etudier la normalité d'erreurs.

<u>Reponse</u>

Pour cela, tout d'abord il faut entrer ces données dans Logiciel SPSS.

Nous suivons les étapes suivantes:

 a) Il faut définir dans la barre en bas « Affichage des variables » : les variables (deux quantitatives de type numérique, et de mesure Echelle) suivantes: X=Volume Horaire, Y=Augmentation des notes.

ta proto	cole TP 03.	.sav [Ensemt	ole_de_donnée	s1] - IBM SPSS	Statistics Ed	liteur de données					-		
Eichier	Edition	Affichage	<u>D</u> onnées	Transformer	Analyse	Marketing direct	<u>G</u> raphes <u>U</u> tilita	aires Fenêtre	Aide				
2) 🗖		¥ 🎇	॑	H 19		<i>≤</i>					
		Nom	Туре	Largeur	Décimales	Etiquette	Valeurs	Manquant	Colonnes	Align	Mesure	Rôle	
1	Н		Numérique	8	2	Heure	Aucun	Aucun	8	Droite	Ichelle 🔗	S Entrée	<u> </u>
2	N		Numérique	8	2	Note	Aucun	Aucun	8	🗃 Droite	Ichelle 🔗	🔪 Entrée	
3													
4													
5													
6													
7													
8													
9													
10													
11													
12													
13													
14													
15													
16													
17													
18													
19													
20													
21													
22													
23													
24	1									1		1	
Afficha	ae des don	nées Affi	chane des vai	riables									
			1										
												Le processe	ur IBM SPSS Statistics est prêt
1	e		0	6) [FR 🔺 .ntl 🗊 17:49 03/05/2022

b) On introduit les données dans la barre « Affichage des données ».

ta proto	🔓 protocole TP 03.sav [Ensemble_de_données1] - IBM SPSS Statistics Editeur de données																
<u>F</u> ichier	Edition Affichage	<u>D</u> onnées	Transformer	Analyse	Marketing dire	ect <u>G</u> raphes	<u>U</u> tilitaires	Fenêtre A	ide								
			1	*	A	¥5 🖬				6 46							
6:															Vi	sible : 2 varia	bles sur 2
	н	N	var	var	var	var	var	var	var	var	var	var	var	var	var	var	v
1	10,00	10,00															-
2	11,00	10,00															
3	14,00	12,00															
4	15,00	12,00															
5	20,00	13,00															
6	25,00	13,00															
	46,00	19,00															
8	50,00	15,00															
10	9 59,00 16,00																
11	10,00	20,00															
12																	
13																	
14																	
15																	
16																	
17																	
18																	
19																	
20																	
21																	
Afficha	Affichage des données Affichage des variables																
	Le processeur IBM SPSS Statistics est prêt																
@	🚱 🌔 🧮 🖸 💕 🔿 📕 📉 🖳 😂																

Pour obtenir la régression Linéaire simple, on suivre les étapes : <u>Analyse</u>, Régression, puis <u>Linéaire</u>.

ta proto	cole TP 0	B.sav [Ensemb	le_de_donnée	s1] - IBM SPSS	Statistics Editeur de	données	_	-										I X	
Eichier	Edition	Affichage	<u>D</u> onnées	Transformer	Analyse Marketi	ing direct <u>G</u> raphes	Uti	litaires	Fenêtre /	Nide									
		₿ Щ		¥ 🎬	Rapports Statistiqu <u>e</u> s	descriptives) }	4	s 🎹 🛛		• 46								
6:					Ta <u>b</u> leaux		•									V	isible : 2 varia	bles sur 2	
		Н	N	var	Co <u>m</u> parer le	s moyennes	•	ar	var	var	var	var	var	var	var	var	var	v	
1		10,00	10,00		Modèle linéa	ire général												4	
2		11,00	10,00		Modèles liné	aires généralisés													
3		14,00	12,00		Modèles Mi <u>x</u> t	es													
4		15,00	12,00		<u>C</u> orrélation														
5		20,00	13,00		<u>R</u> égression		•		Aodélisation li	néaire autom;	atique	1							
6		25,00	13,00		L <u>o</u> g Linéaire				inéaire										
7		46,00	19,00		Réseaux neu	Ironaux			liuctoment de	fonctions									
8		50,00	15,00		Classification	n			yusternent de	Tongaons									
9		59,00	16,00		Réduction de	es dimensions			noindres carre	es partiels									
10	10 70,00 20,00 Echelle Image: Constraint of the con																		
11	Image: Note of the second s																		
12	Tests non paramétriques Image: Cogregation material Im																		
13					Survie			- II.	lodèles de ch	oix binaire									
14					Réponses m	ultiples			on linéaire										
16					Analyse des	valeurs manguantes		B	- Pondération e	stimée									
10	_				Imputation m	ultiple .		Li I	oubles moin	dres carrés									
18					Echantillons	complexes			Codage optim										
19					Contrôle de o	ualité		<u> </u>	Judage opum	ar (CATINEO)	•								
20					Courbe ROC														
21																			
22																			
	4																		
Affichage des données Affichage des variables																			
Linéaire	inéaire Le processeur IBM SPSS Statistics est prêt																		
	E		0	6	>		😨 🤗 🗮 🖸 😆 🔿 📕 Ķ 🖳 🙉												

- c) On pose la variable à expliqué Y =Note dans le choix « dépendante », et dans le « variables indépendantes» on pose la variable explicative X=Heure.
- d) On garde la méthode utilisé (Entrée).
- e) Dans le choix « Statistiques », on choisit sur « Coefficients de régression » (Estimation et intervalle de confiance avec 95% le taux de confiance), et (qualité de l'ajustement).

ta proto	cole TP 03.sav [Ense	mble_de_données1]	- IBM SPSS Statistics Edite	eur de données	-	and the Real		a de production de la constante			
Eichier	Edition Affichag	e <u>D</u> onnées <u>T</u> ra	nsformer <u>A</u> nalyse <u>M</u>	arketing direct <u>G</u> rag	hes <u>U</u> tilitaires	Fenêtre Aide					
			🖺 🕌 📰	H 👪 🛙	i - 4	A 14	0	A86			
6:										Visible : 2	variables sur 2
	Н	N	V ta Régression linéai	re				8	Régression linéaire : Statist	iques	×
1	10,0	0 10,00		-					regression meaner statist	ideo.	
2	11,0	0 10,00		R	épendant :			Statistiques	Coefficients de régression	Qualité de l'ajustement	
3	14,0	0 12,00	Heure [H]		Note [N]			Diagrammes	Estimations	Variation de R-deux	
4	15,0	12,00		Bloc 1 de	1			Epropietror	Intervalles de confiance	e 🔲 Caractéristiques	
5	20,0	0 13,00		Précéd	ent	S	uiva <u>n</u> t	Chregistier	Niveau (%) : 95	Mesure et corrélations par	ielles
6	25,0	0 13,00		,	/ariables indépenda	antes :		Options	Matrice de covariance	Tests de colinéarité	
7	46,0	0 19,00			Heure [H]			Bootstrap	Résidus		
8	50,0	0 15,00		•					Durbin Watson		
9	59,0	0 16,00							Diagnostic dos obson	ations	
10	70,0	20,00			Méthode :	Entrée	~		Points atvniques :	auons a écarts-types	
11									Toutes les observatio		
12					riabl <u>e</u> de filtrage :	Ràc	lo		C Toutes les observau	5115	
13			-			1402	10		Poursuivre	Annuler Aide	
14					iquettes d'observatio	on :					
15											
16					olds WLS :						
1/											
18				OK Colle	r <u>R</u> éinitialiser	Annuler	Aide				
19			(-		_				
20											
21											
22											
	4						_				
Afficha	ge des données	Affichage des variabl	es								
									Le processeur IBM S	PSS Statistics, act prât	
									The processed ribin a	or oo oranonco est pret	
	6) 🔿 🛃							FR 🔺 .all 🗎	18:00 03/05/2022

f) Dans le choix « Diagramme », on pose (ZPRED=X), et (ZRESID=Y), et sutrout n'oblier pas de cocher sur (Diagramme de répartition Gaussien), pour savoir est-ce que deuxième condition de validation est vrai pour assurer que le Modèle de la Régression Linéaire Simple est applicable.

ta proto	cole TP 03.sav [Er	semble_de_do	nnées	ta Régression linéaire	-	-		×							3 22
Eichier	Edition Affich	age <u>D</u> onnée	es :			Désaudant :									
				Heure [H]	ן 🖌	Note [N]		Statistiques							
		- •			Bloc 1	de 1		Diagrammes							
6:		1	_		Pró	cédent	Quivant	Enregistrer		1	1	1	Vi	sible : 2 varial	oles sur 2
	Н	N				Variables indéner	dantes :	<u>O</u> ptions	var	var	var	var	var	var	
	11	00 10	0,00			Heure [H]	duntes .	<u>B</u> ootstrap	-						
3	14	00 12	2 00		•				-						
4	15	.00 12	2,00						-						
5	20	,00 13	3,00			Méthode	: Entrée 🔻								
6	25	,00 13	3,00			Variable de filtrage			-	-		X			
7	46	,00 19	9,00		-	vanabl <u>e</u> de indage	Kegression lineaire : Dia	grammes							
8	50	,00 15	5,00			Etiquettes d'observ	DEPENDNT	Dispers	on 1 de 1						
9	55	,00 16	5,00		*		*ZPRED	Précéd	lent		Suivant				
11		,00 20	,00			Poids WLS :	*DRESID		γ·						
12							*ADJPRED	*	*ZRESID	l.					
13				ОК		Co <u>l</u> ler <u>R</u> éinitialis	*SRESID *SDRESID		<u>x</u> :						
14			U						*ZPRED						
15							Diagrammes des résid	us normalisés 👘	Cánárar	tous los ara	obiques partici				
16							Histogramme	E.	Generer	tous les graj	pinques parie	°			
17			_				Diagramme de répa	intition gaussien							1
10															
20								oursuivre Annuler	Ai	de					
21							C		_						
22															
	1														
Afficha	ide des données	Affichage de	s varia	bles			***								
Aniche	ige des données]													
					(Lep	processeur IBN	I SPSS Statis	ics est prêt		
1	6			🔰 🔘 🔼	K	W 2									18:03 05/2022

- g) En fin OK.
- h) Pour voir Nuage des points, en suivant les étapes :

Graphs, puis Boite de dialogue anciennes version, pui choisir Disperssion/Points, et puis Disperssion Simple.

On pose la variable X et La variable Y sur l'axes X et Y. Puis OK.

ta protocol	e TP 03.sav [Ensembl	e_de_donnée:	s1] - IBM SPS	S Statistics Edi	teur de donné	es										
Fichier E	dition Affichage	Données	<u>T</u> ransformer	<u>A</u> nalyse	<u>M</u> arketing dire	ct <u>G</u> raphes	<u>U</u> tilitaires	Fenêtre /	Nide							
				↓ _	- A-A	🚛 👔 Gér	nérateur de di	agrammes								
						💾 🔟 Sél	ecteur de moo	ièles de repré	sentations gra	aphiques						
						Boî	tes de dialoqu	ie ancienne ve	ersion	4		âtonc		V	sible : 2 variat	oles sur 2
	Н	Ν	var	var	var	var	var	var	var	var		átono 2D	var	var	var	v
1	10,00	10,00										atoris <u>5</u> D				
2	11,00	10,00										ourbes				
3	14,00	12,00									<u>∧</u>	ires				
4	15,00	12,00									🛃 S	ecteur				
5	20,00	13,00									P 👔	lafond-planc <u>h</u> er				
6	25,00	13,00									🗰 B	oîte à moustaches				
7	46,00	19,00									HI B	arre d'erreur				
8	50,00	15,00									- P	vramide de population				
9	9 59,00 16,00 10 70,00 20,00															
10	10 70,00 20,00 11 Image: Comparison Points Image: Comparison Points															
11	11 Histogramme															
12	13															
15	13 14															
14	14 15 16<															
16																
17	-															
18	-															
19																
20																
21																
22																
Affichage Dispersion	Vifichage des variables															
1	FR - ul 🔒 🔿 📕 🔀 💹 🖄															

Interprétation des résultats

1) On constat ici que les résultats sont organiser selon les tableaux :

Dans le tableau 1 « Variables introduits », on remarque qu'il y 'a une seule variable exlicative X=Volume Horraires (c'est pour ca notre modèle prend son nom), et modèle « 1 ».

Dans le tableau 2 « Récapitulatif des modèles », on remarque que la valeur de coefficient de correlation « r : notation de cours » ici dans SPSS est noté par « R =0,913 » ,c'est-à-dire qu'il y a une très bonne corrélation entre Volume Horraire et l'augmentation des notes, de plus la valeur de coefficient de determitation de la régression c'est « R :notation de cours » ici dans SPSS est noté par « R^2 =0,833 » c'est-à-dire que le taux de corrélation c'est 83,3% (parmi 100 points il y a 83 qui se trouvent dans la droite ».

D'autre part, on propose l'hypothèses suivantes :

H0 : « Notre modèle de régression n'est pas corrélé ».

H1 : « Notre modèle de régression est bien corrélé ».

Alors, on peut remarquer dans I tableau 3 de l'Analyse de la Variance que la valeur « Sig=0,000<0,05 », donc on acceptte H1, c'est-à-dire qu'il ya une liaison siginicative entre X=Heure et Y=Notes, ou bien notre modèle est globalement significative, avec une taux de



 Dans le tableau 4, on peut voir les valeurs à éstimer pour la droite de la régression linéaire, avec « A=0,143 et B=9,436 » (dans le cours sont notés respectivements par α, β), c'est-à-dire que sur l'augmentation de 10 heurs de la révision est donne une augmentation de 0,143 sur les notes.

Alors la droite de régression s'écrit sous la forme :

$$y = 0,143x + 9,436.$$

De plus pour l'intervalle de confiance pour le coefficient

 $\alpha \in [0,091; 0,195], et \beta \in [7,444; 11,427].$ « Voir les formules exlicites dans le cours ». Autrement dit que le pente de la droite peut être varié de 0,091 jusqu'à 0,195.

Résultats PROTOCOLE TP 3.spv [Do	cument2] - IBM SPSS Statistic	s Viewer	-									3
Eichier Edition Affichage Insér	er Tableau <u>p</u> ivotant F <u>o</u> rr	mat <u>A</u> nalyse	<u>G</u> raphes <u>U</u> tilita	aires Fenêtre	Aide							
😑 🖩 🖨 🙇 🥑) 🛄 🖛 🛥				🦊 🖥	P 🖻		+ +	+-			
Résultats	1 Régression Résidu Total a. Variable dépenda b. Valeurs prédites :	90,000 18,000 108,000 inte : Note : (constantes), He	1 8 9	90,000 4 2,250 Coefficients ^a	0,001	,000 ^b						
Titre Remarques L'Ensemble de données acti Dispersion de N H		Coefficients no	n standardisés Erreur	Coefficients standardisés Rôto		Qia	95,0% % int confiance Borne	ervalles de e pour B Limite				
E Régression	Modele 1 (Constante) Heure	9,436 ,143	,864 ,023	,913	10,926 6,325	,000 ,000	7,444 ,091	11,427 ,195				
 —	a. Variable dépenda	inte : Note Statistiques	des résidus ^a									
Len ANOVA Coefficients Statistiques des résidus Diagrammes E Diagrammes Coefficients Diagramme gaussien Cm Diagramme de disper:	Min Prévision 10 Résidu -1,1 Erreur Prévision Erreur Résidu -	imum Maximu 0,8621 19,42 85104 3,003 -,992 1,7 -1,234 2,0 wto:Noto	IM Moyenne 00 14,0000 16 ,00000 14 ,000 02 ,000	Ecart-type 3,16228 1,41420 1,000 ,943	N 10 10 10 10							
	Diagrammes Diagram	nme gaussie	n P-P de rég	ression de R	ésidu sta	ndardisé						
			سيتعادا والأسوي					Lep	rocesseur IBM SF	SS Statistics est	prêt H: 147, W: 391	pt.
0 6 🔚 0		B		2.				-			FR •II • 10:11 04/05/2022	2

 Dans le diagramme ci-dessous, on peut établir que condition de validation de la régression linéaire simple est valide, c'est-à-dire que les erreurs sont gaussiennes.



4) Pour Nuage des point et caractéristiques.

Double clic sur le graph, et on clique eur « Element » qui se trouve dans la barre de Menu, et puis choisir « Ajouter une courbe d'ajustement ». R²=0,833.

