الانحدار الخطي البسيط

La régression linéaire simple

مقدمـــــة

المتغيرات الاحصائية

- المتغير المستقل Independent variable: هو المتغير الذي يتم بحث اثره في متغير اخر ويمكن للباحث التحكم فيه للكشف عن تباين هذا الاثر باختلاف قيم ذلك المتغير.

- المتغير التابع dépendent variable: هو ذلك المتغير الذي يرغب الباحث في الكشف عن تأثير المتغير المستقل عليه.

تعريف الانحدار الخطي

- الانحدار أسلوب يمكن بواسطته تقدير قيمة أحد المتغيرين بمعلومية قيمة المتغير الآخر عن طريق معادلة(دالة) الانحدار.
- ويعرف المتغير الأول بالمتغير التابع Dépendent ويرمز له بالحرف Y، في حين يعرف المتغير الآخر بالمتغير المستقل Indépendant ويرمز له بالحرف X.
- تهدف دراسة الانحدار التنبؤ بقيمة متغير Y بمعرفة متغير X. إذن الغاية من استخدام أسلوب تحليل الانحدار الخطي دراسة وتحليل أثر متغير كمي على متغير كمي آخر.

سندرس:

الانحدار الخطي البسيط : فكلمة " بسيط " تعني أن المتغير التابع Y يعتمد على متغير مستقل واحد وهوX وكلمة " خطي" تعني أن العلاقة بين المتغيرين (X , Y) هي علاقة خطية.

$$\hat{y} = a + bx$$

حيث a : ثابت الانحدار أو الجزء المقطوع من محور y

b : ميل الخط المستقيم أو معامل انحدار Y على X (أو Y/X)

- وتحسب القيمتان a و b من العلاقتين التاليتين :

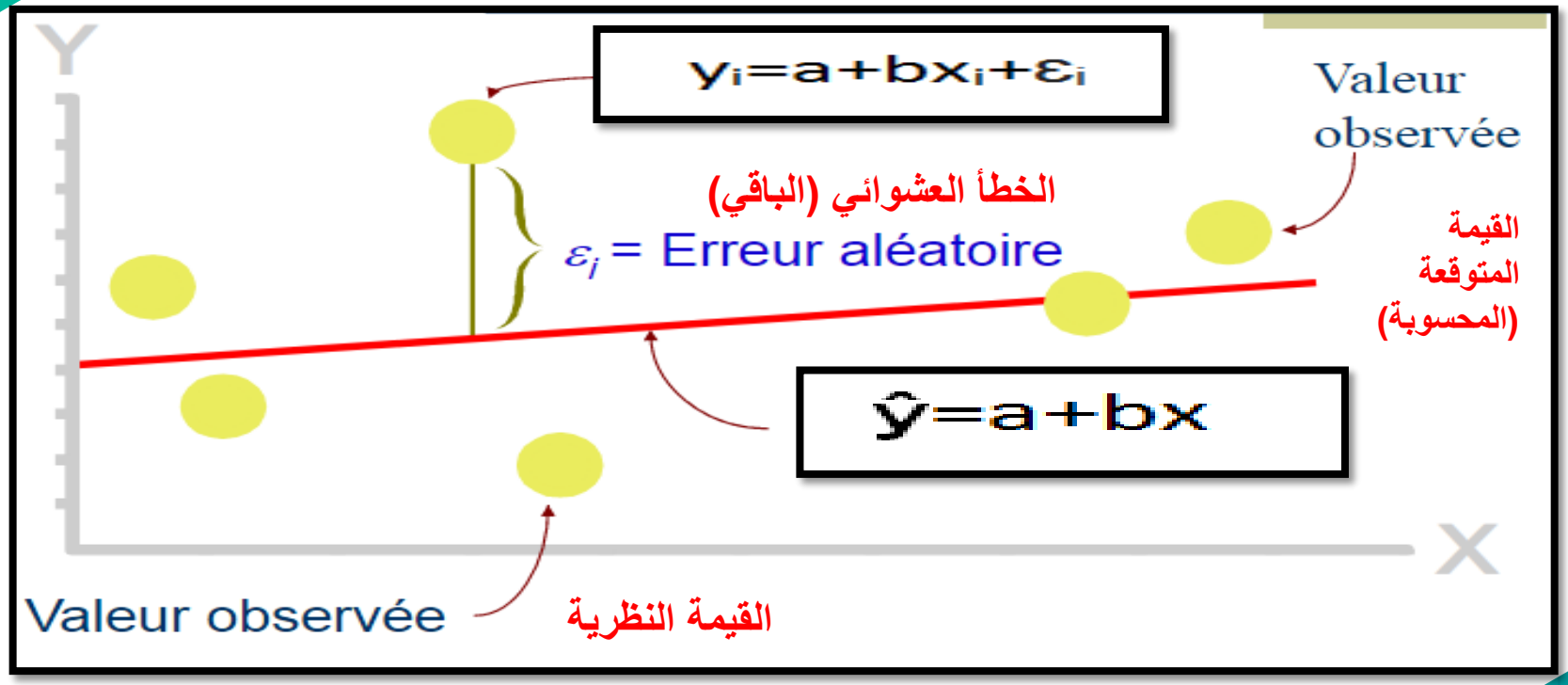
حيث:

$$a = \frac{\sum y - b\sum x}{n} \qquad b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2} =$$

لإيجاد قيمة مقدرة جديدة $\hat{y}_h$ نعوض بقيمة معلومة للمتغير المستقل ولتكن x_h في معادلة تقدير خط الانحدار Y/X

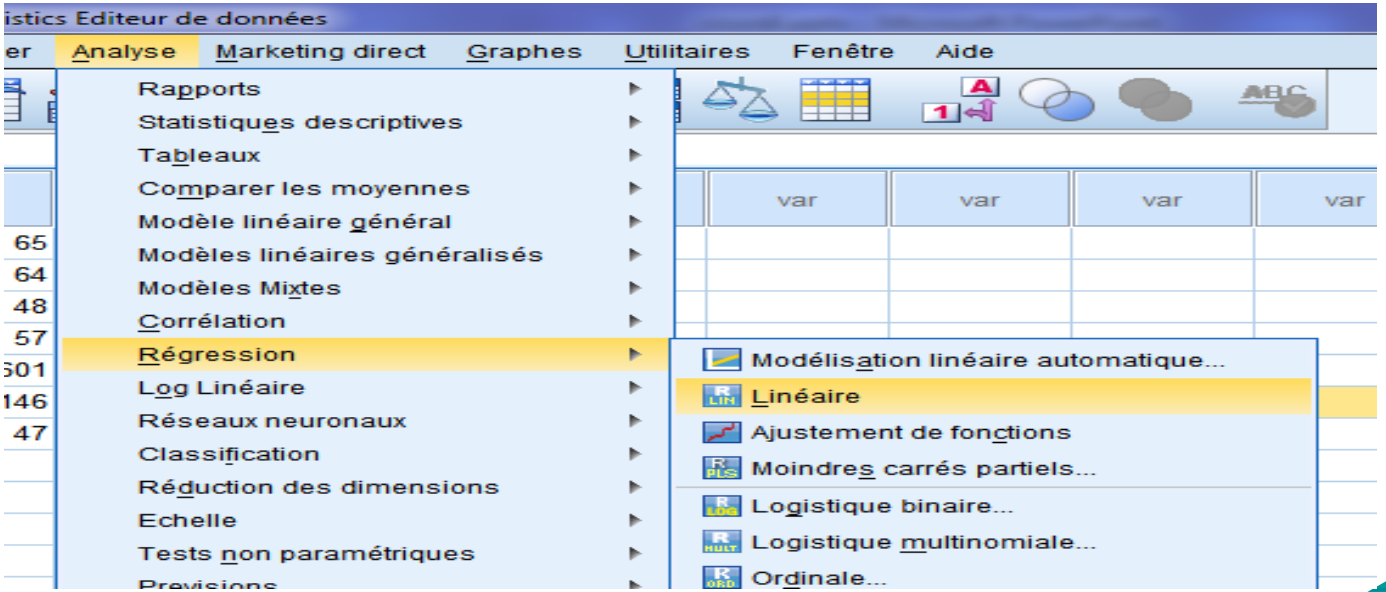
$$\hat{y} = a + bx$$

نعوض

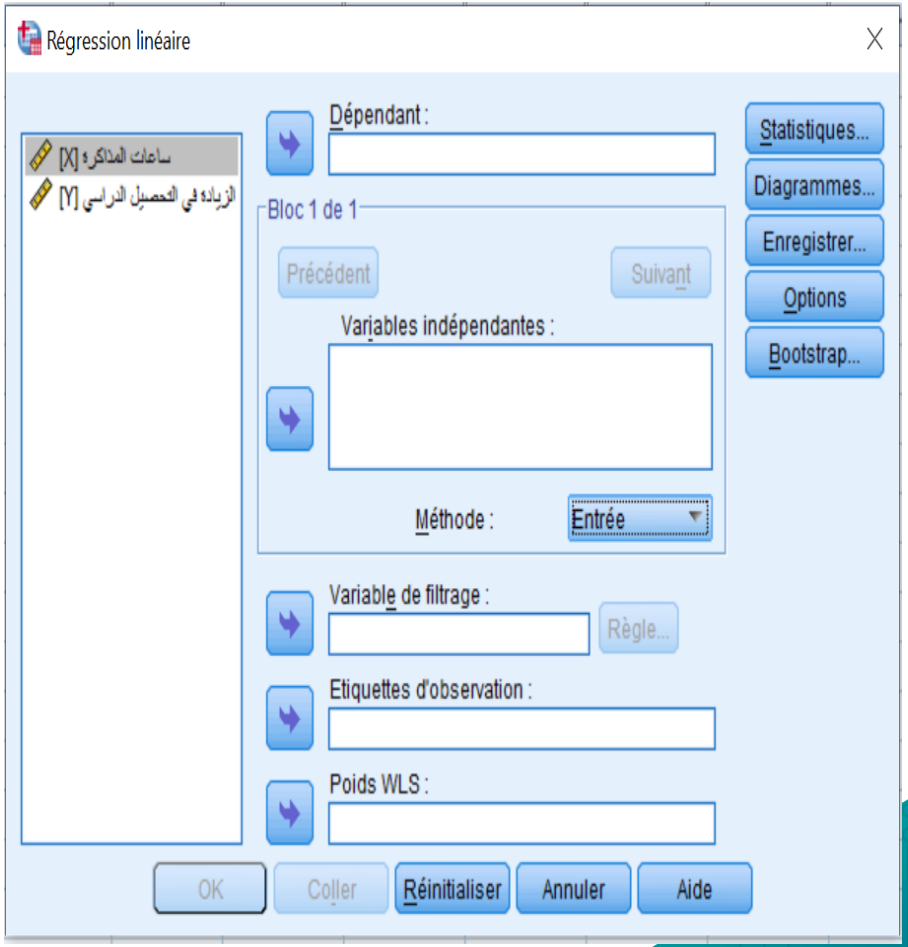

$y_i = a + bx_i + \varepsilon_i$
Valeur observée
الخطأ العشوائي (الباقي)
ε_i = Erreur aléatoire
القيمة المتوقعة (المحسوبة)
$\hat{y} = a + bx$
Y
X
Valeur observée
القيمة النظرية

حساب الانحدار الخطي البسيط باستخدام برنامج الـ SPSS :

Menu Analyse--> Régression--> Linéaire



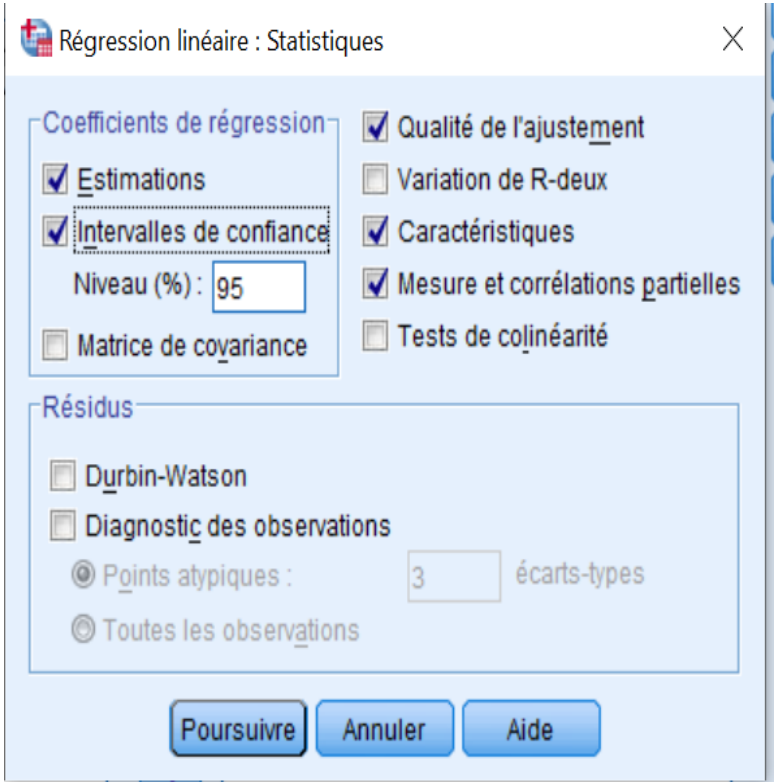
بحيث يتم نقل المتغير التابع الى المجال Dependent والمتغير المستقل الى المجال Independent، ويتوفر الزر Method على عدة خيارات خاصة بطريقة التقدير، والطريقة التلقائية للبرنامج هي طريقة المربعات الصغرى العادية المشار اليها بـ Enter. أما المجال الخاص بالمتغيرات المستقلة فيتوفر فيه امكانية تقدير مجموعة من النماذج باستخدام نفس المتغير التابع ومتغيرات مختلفة مستقلة مثلا استخدام المتغير X1 في النموذج (1) و X2 في النموذج (2) بنفس الإجراء، وذلك بالاستعانة بالزر Next في المجال Block 1 Of 1، بحيث يتم نقل المتغير X1 في Block1 والمتغير X2 في Block2، ونتائج التقدير تكون في شكل نموذجين مختلفين. بينما ادخال المتغيرين معا في المجال Independent فإنه يؤدي الى تقدير نموذج انحدار متعدد (بمتغيرين مستقلين).



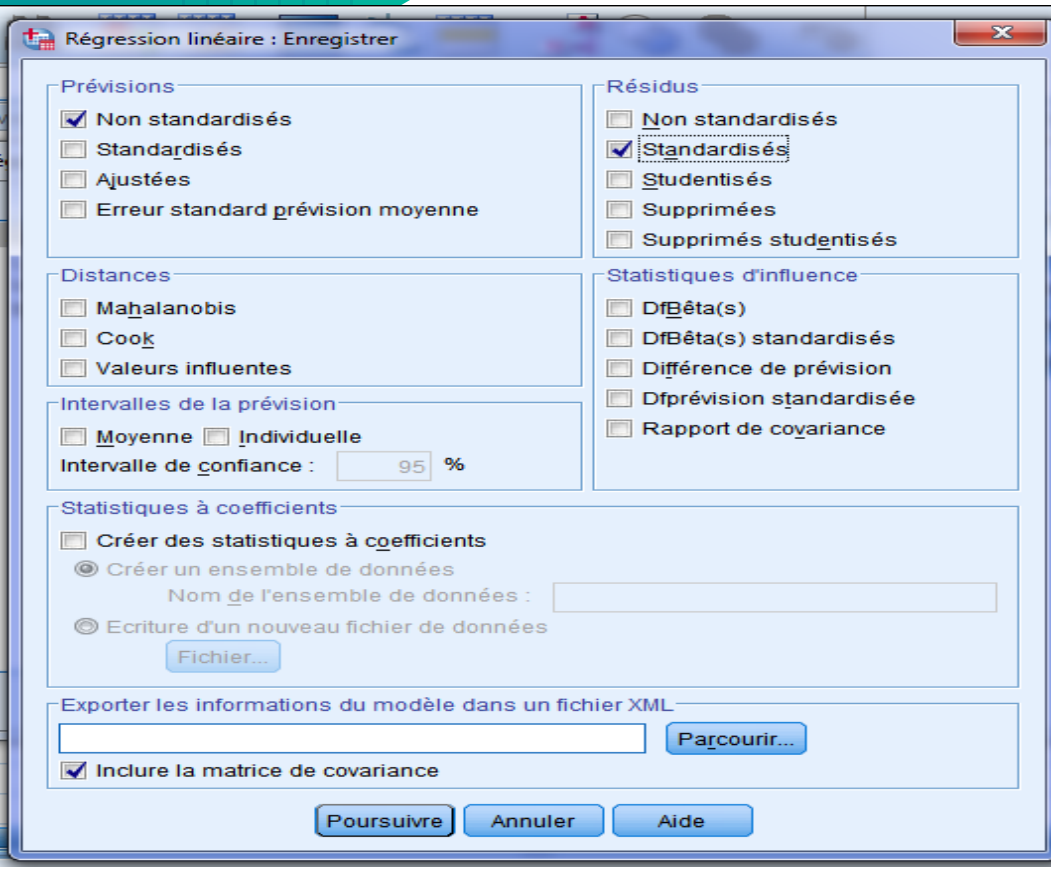
الزر Statistiques

نقوم بتحديد الخيارات التالية:

- ✓**Estimations**: لتقدير معلمتي نموذج الانحدار.
- ✓**Intervalles de confiance** : لتقدير مجال الثقة 95% لمعلمي الانحدار.
- ✓**Qualité de l'ajustement** : لتقدير جودة النموذج (لعرض جدول تحليل التباين ANOVA و معامل التحديدR2).
- ✓**Caractéristiques**: لعرض الإحصاءات الوصفية.
- ✓**Mesure et corrélations partielles**: لحساب معاملات الارتباط البسيط و الجزئي.



Poursuivre



الزر Enregistrer

Non standardisés prévisions

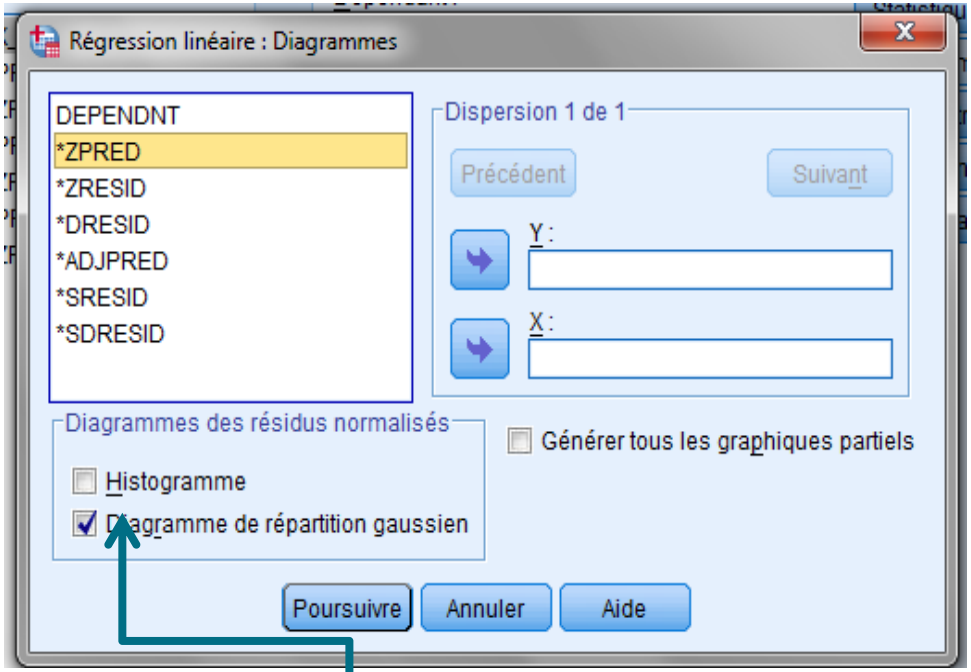
Standardisés résidus

Poursuivre

في هذه النافذة نؤشر على Non standardisésمن الاختيارات الخاصة ب (القيم التنبؤية) Prévisions

كما نؤشر على Standardisés من الاختيارات الخاصة ب (البواقي)، ثم نضغط على poursuivre للعودة إلى النافذة الأساسية.

Diagrammes الزر



مخطط التوزيع الطبيعي (إعتدالية التوزيع الاحتمالي للبواقي)

Poursuivre ← OK

مثال تطبيقي

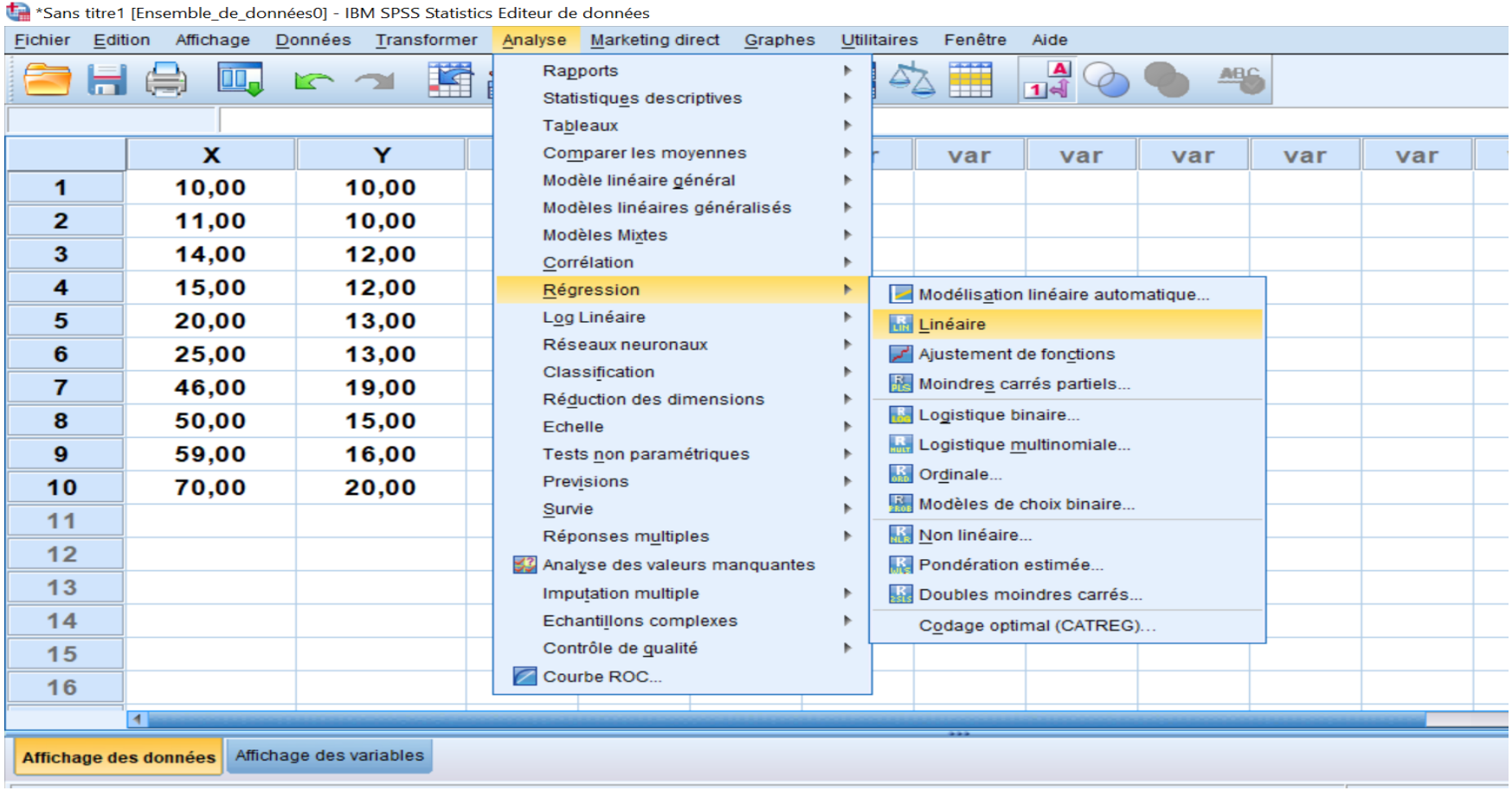
فيما يلي بيانات عن عدد ساعات المذاكرة في الأسبوع لعينه من 10 طلاب و مقدار الزيادة في التحصيل لمادة الرياضيات.

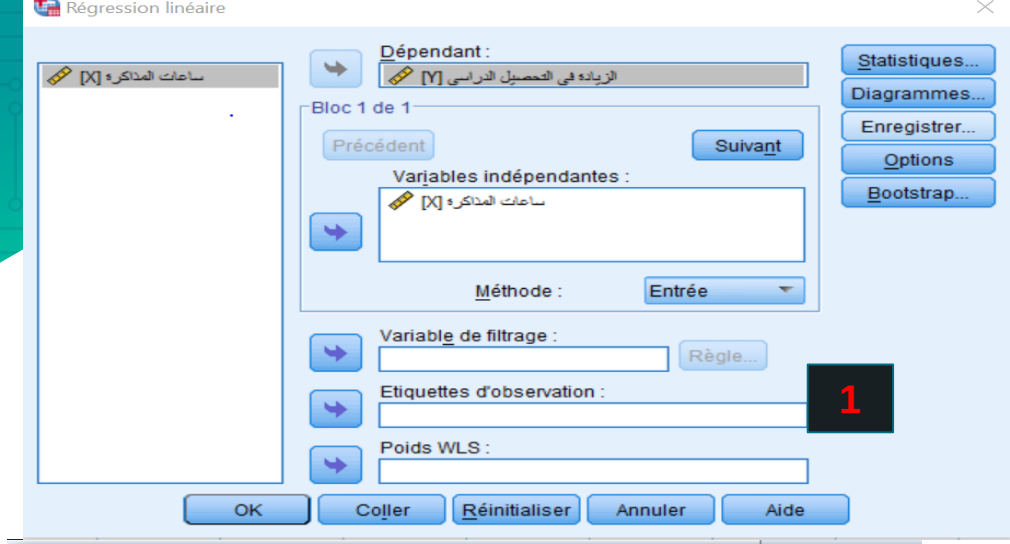
ساعات المذاكرة	10	11	14	15	20	25	46	50	59	70
الزيادة في التحصيل	10	10	12	12	13	13	19	15	16	20

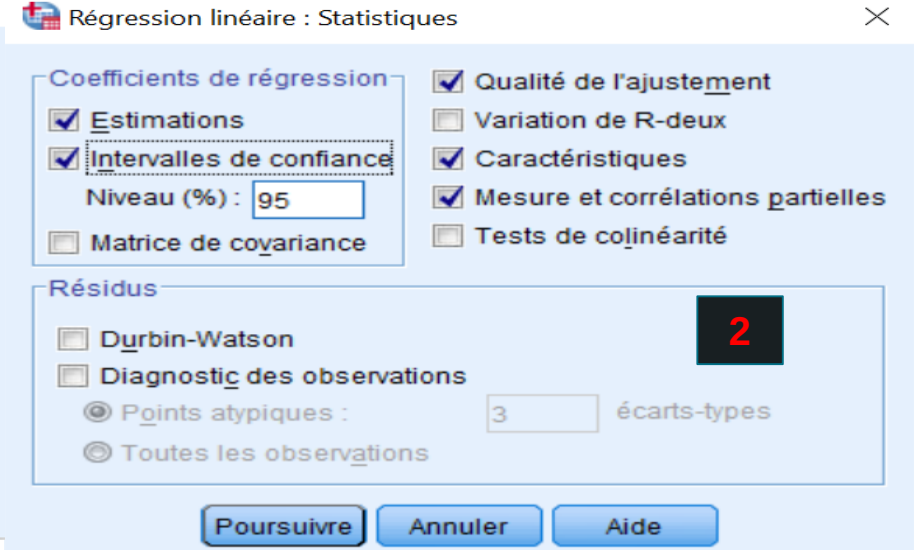
المطلوب:

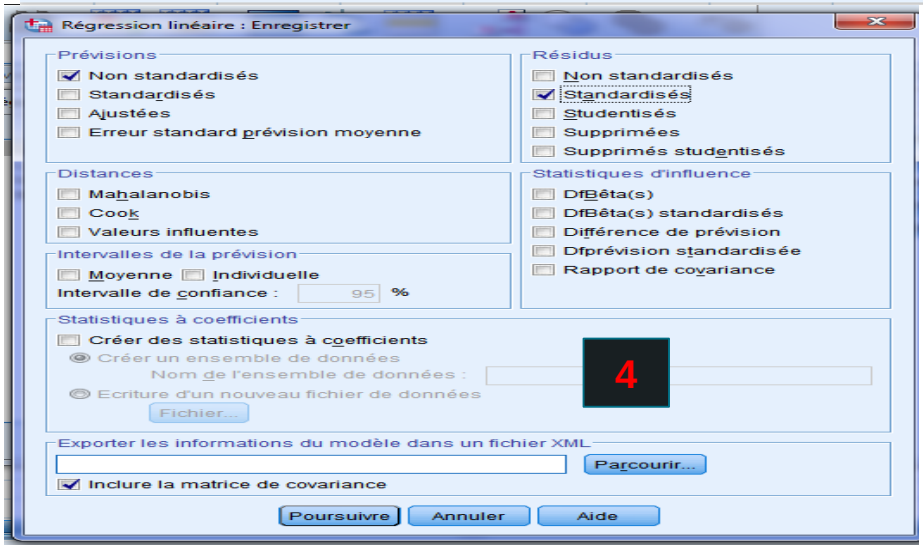
- ارسم مخطط الانتشار.
- قدر معادلة انحدار ساعات المذاكرة على التحصيل.
- فسر معادلة الانحدار.

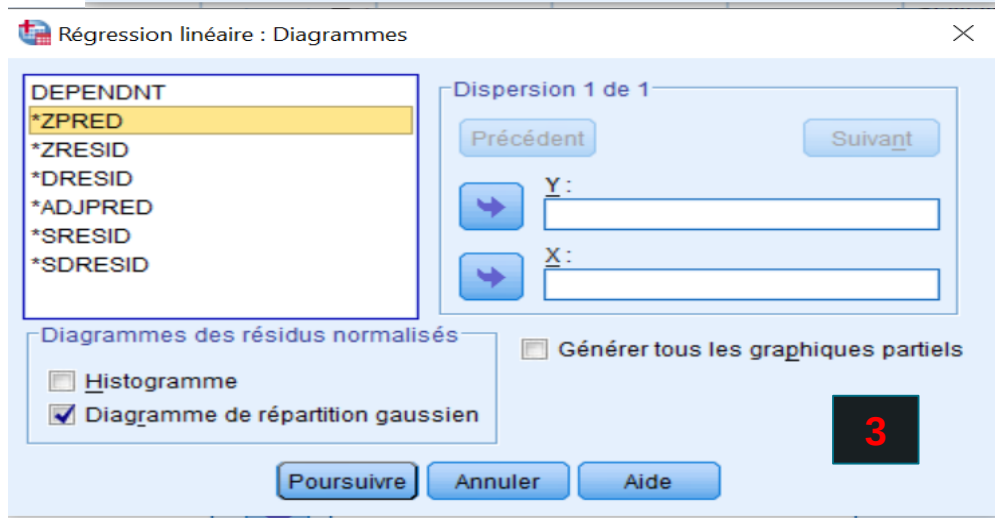
خطـــــــوات الحل




Régression linéaire
ساعات المذاكرة [X]
Dépendant :
الزيادة في التحصيل الدراسي [Y]
Bloc 1 de 1
Précédent
Suivant
Variables indépendantes :
ساعات المذاكرة [X]
Méthode :
Entrée
Variable de filtrage :
Règle...
Etiquettes d'observation :
Poids WLS :
Statistiques...
Diagrammes...
Enregistrer...
Options
Bootstrap...
OK
Coller
Réinitialiser
Annuler
Aide
1


Régression linéaire : Statistiques
Coefficients de régression
Estimations
Intervalles de confiance
Niveau (%) : 95
Matrice de covariance
Qualité de l'ajustement
Variation de R-deux
Caractéristiques
Mesure et corrélations partielles
Tests de colinéarité
Résidus
Durbin-Watson
Diagnostic des observations
Points atypiques : 3 écarts-types
Toutes les observations
2
Poursuivre
Annuler
Aide


Régression linéaire : Enregistrer
Prévisions
Non standardisés
Standardisés
Ajustées
Erreur standard prévision moyenne
Résidus
Non standardisés
Standardisés
Studentisés
Supprimées
Supprimés studentisés
Distances
Mahalanobis
Cook
Valeurs influentes
Statistiques d'influence
DfBêta(s)
DfBêta(s) standardisés
Différence de prévision
Dfprévision standardisée
Rapport de covariance
Intervalles de la prévision
Moyenne
Individuelle
Intervalle de confiance : 95 %
Statistiques à coefficients
Créer des statistiques à coefficients
Créer un ensemble de données
Nom de l'ensemble de données :
Ecriture d'un nouveau fichier de données
Fichier...
4
Exporter les informations du modèle dans un fichier XML
Parcourir...
Inclure la matrice de covariance
Poursuivre
Annuler
Aide


Régression linéaire : Diagrammes
DEPENDNT
*ZPRED
*ZRESID
*DRESID
*ADJPRED
*SRESID
*SDRESID
Dispersion 1 de 1
Précédent
Suivant
Y :
X :
Diagrammes des résidus normalisés
Histogramme
Diagramme de répartition gaussien
Générer tous les graphiques partiels
3
Poursuivre
Annuler
Aide

تحليل النتائـــــــــــج

Statistiques descriptives

	Moyenne	Ecart-type	N
الزياده في التحصيل الدراسي	14,0000	3,46410	10
ساعات المذاكره	32,0000	22,17105	10

الجدول الأول: يوضح الإحصاءات الوصفية للمتغيرين حيث يظهر ان المتوسط الحسابي لمتغير التحصيل قد بلغ 14بانحراف معياري3.46410 في حين أن المتوسط الحسابي لمتغير ساعات الدراسة قد بلغ 32 بانحراف معياري 22.17105 و بما ان كل البيانات كانت كاملة لكل وحدات العينة فان حجم العينة لم يتغير و هو يساوي 10.

Corrélations

		الزيادة في التحصيل الدراسي	ساعات المذاكرة
Corrélation de Pearson	الزيادة في التحصيل الدراسي	1,000	,913
	ساعات المذاكرة	,913	1,000
Sig. (unilatérale)	الزيادة في التحصيل الدراسي	.	,000
	ساعات المذاكرة	,000	.
N	الزيادة في التحصيل الدراسي	10	10
	ساعات المذاكرة	10	10

الجدول الثاني: و يظهر نتائج الارتباط البسيط بين المتغيرين حيث يظهر ارتباطا موجب قوي بين متغير التحصيل و ساعات المذاكرة و قد بلغ 0.913 عند مستوى دلالة 0.05 حيث يظهر أن Sig α > (0.05> 0.000) ، ومنه نرفض فرضية العدم H_0 مما يدل على أن لمعامل الارتباط دلالة إحصائية عند مستوى معنوية 0.05= α.

Variables introduites/supprimées[b]

Modèle	Variables introduites	Variables supprimées	Méthode
1	ساعات المذاكره [a]	.	Entrée

a. Toutes variables requises saisies.
b. Variable dépendante : الزياده في التحصيل الدراسي

الجدول الثالث (جدول نوع الطريقة): يبين أن طريقة المربعات الصغرى هي المتبعة في تحليل الانحدار الخطي و كذلك تم ادخال متغير مفسر واحد (مستقل) و هو متغير ساعات المذاكرة باعتبار انه المتغير الوحيد المدرج في النموذج المفسر للانحدار و بأن المتغير التابع هو متغير التحصيل.

جدول معامل التحديد :

Récapitulatif des modèles[b]

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,913[a]	,833	,813	1,49999

a. Valeurs prédites : (constantes), ساعات المذاكره
b. Variable dépendante : الزياده في التحصيل الدراسي

الجدول الرابع: يتضمن الجدول أهم مؤشرات نموذج الانحدار و المتمثل في معامل التحديد و الذي يرمز له بالرمز R^2(R-deux) و الذي يعتبر مقياسا لدقة النموذج و هو يساوي 0.833 و تفسير ذلك ان 83.3% من التغيرات التي تحدث في المتغير التابع (التحصيل الدراسي سببه المتغير المستقل (ساعات الدراسة) و النسبة الباقية 16.7% ترجع إلى عوامل أخرى، و كلما اقتربت قيمة R^2 من 100% دل ذلك على دقة النموذج و ذلك لان :

$$R=\sqrt{R2}$$

جدول تحليل التباين ANOVA:

ANOVA[b]

Modèle		Somme des carrés	ddl	Moyenne des carrés	D	Sig.
1	Régression	90,000	1	90,000	40,001	,000[a]
	Résidu	18,000	8	2,250		
	Total	108,000	9			

a. Valeurs prédites : (constantes), ساعات المذاكره
b. Variable dépendante : الزياده في التحصيل الدراسي

الجدول الخامس: من خلال جدول تحليل التباين نستخلص العلاقة بين Sig و α فعندما تكون $Sig < \alpha$ نستطيع الحكم على تحقق شرط المعنوية الكلية للنموذج.

و في مثالنا تحقق شرط المعنوية الكلية للنموذج. أي ان هناك واحد على الأقل من معاملات الانحدار تختلف على الصفر.

جدول معاملات الانحدار:

Coefficients[a]

Modèle		Coefficients non standardisés		Coefficients standardisés	t	Sig.	95,0% % intervalles de confiance pour B		Corrélations		
		A	Erreur standard	Bêta			Borne inférieure	Limite supérieure	Corrélation simple	Partielle	Partie
1	(Constante)	9,436	,864		10,926	,000	7,444	11,427			
	ساعات المذاكره	,143	,023	,913	6,325	,000	,091	,195	,913	,913	,913

a. Variable dépendante : الزياده في التحصيل الدراسي

كما يتضح من الجدول أن قيمة Sig= 0.000 الخاصة بالثابت هي أقل من مستوى المعنوية 0.05 مما يدل أن المقدار الثابت في نموذج الانحدار معنوي.

ونجد قيمة Sigالخاصة بالمتغير هي 0.000 وهي أقل من 0.05 مما يدل أن قيمة معامل الانحدار b في نموذج الانحدار معنوي.

و بالتالي فان ساعات المذاكرة يمكنه التنبأ في الزيادة في التحصيل الدراسي

حسب الجدول نستنتج معادلة نموذج الانحدار التالية:

$y=a+bx$ و منه $y=9.436+0.143\ (x)$

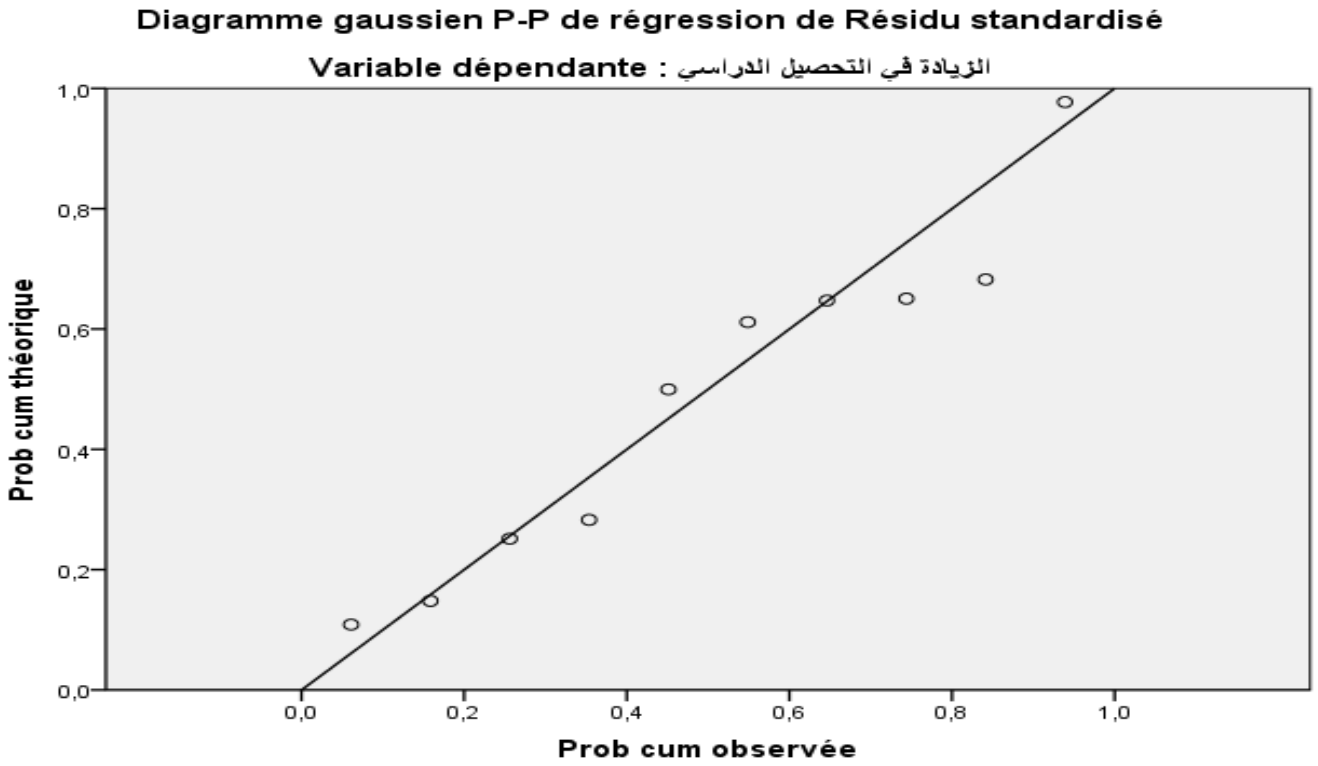
التحصيل الدراسي= 0.143+9.436 (ساعات المذاكرة)

Statistiques des résidus[a]

	Minimum	Maximum	Moyenne	Ecart-type	N
Prévision	10,8621	19,4200	14,0000	3,16228	10
Résidu	-1,85104	3,00316	,00000	1,41420	10
Erreur Prévision	-,992	1,714	,000	1,000	10
Erreur Résidu	-1,234	2,002	,000	,943	10

a. Variable dépendante : الزياده في التحصيل الدراسي

الجدول بعنوان Statistiques des Residus يستخدم لمعرفه بعض المقاييس الخاصة بالبواقي.

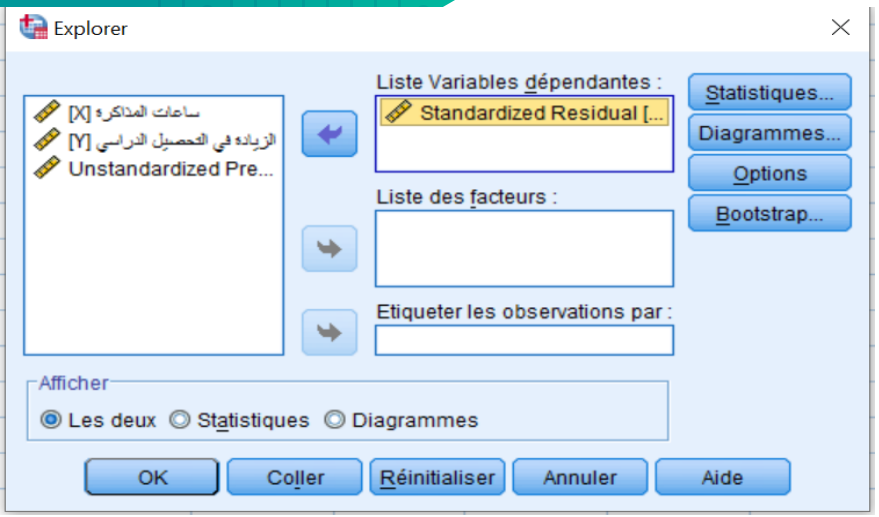


ومن الشكل نجد أن النقاط تتجمع حول الخط وبالتالي فان البيانات (البواقي) تتوزع حسب التوزيع الطبيعي.

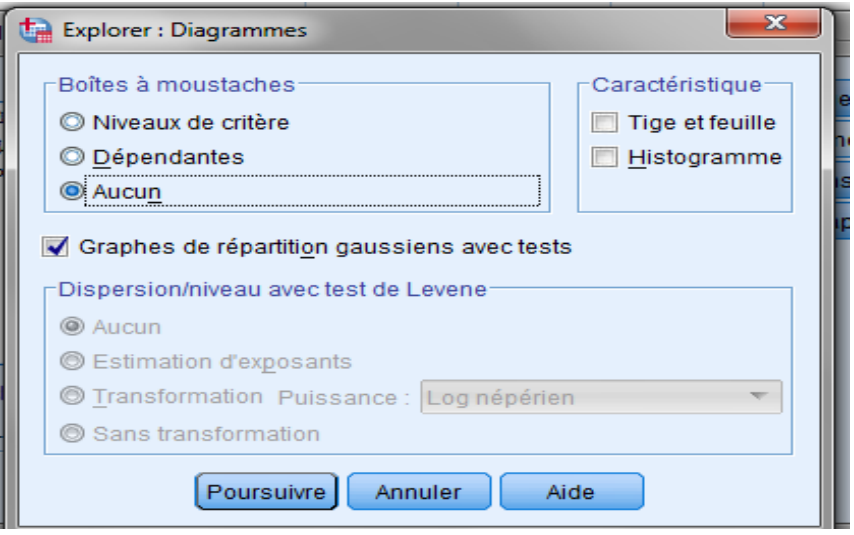
كما يمكن الحكم على مدى اعتدالية البواقي اعتمادا على الطريقة الحسابية وذلك من خلال اختبار (كلومجروف-سمنروف) واختبار (شابيرو-ويليك)، ونحصل عليهما من خلال اتباع الخطوات التالية

Analyse---Statistiques descriptives---Explorer

فتظهر لنا النافذة التالية:



نقوم بإدراج المتغير Standardized Ressiduals
Liste variables dépendantes



ثم ننتقل إلى Diagrammes
فتظهر نافذة جديدة في نفس النافذة كما يلي:

في هذه النافذة نؤشر على Graphes de répartition gaussiens avec tests ثم من الخيار، نؤشر على Boites à moustaches ونلغي التأشير على Tige et feuille بهدف تقليل المخرجات التي Aucun لسنا بحاجة إليها. ثم نضغط على Continue ثم ok نحصل على الجدول التالي:

Tests de normalité

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistique	ddl	Signification	Statistique	ddl	Signification
Standardized Residual	,207	10	,200*	,919	10	,348

a. Correction de signification de Lilliefors
*. Il s'agit d'une borne inférieure de la signification réelle.

يتضح من الجدول السابق أن قيمة Sig في كلا الاختبارين أكبر من 0.05 وهو ما يدل على أن البواقي تتبع التوزيع الطبيعي.

وبالعودة إلى نافذة البيانات Affichage de données نجد أنه قد تم إضافة عمود للبواقي المعيارية (*Standardized Ressiduals*) بعنوان ZRE_1 وعمود آخر للقيم المقدرة للمتغير التابع Yبعنوان PRE_1

ويمكن الاعتماد على البيانات الخاصة بالقيم المعيارية للبواقي عند دراسة اعتدالية التوزيع الاحتمالي للبواقي حسابيا.

X	Y	PRE_1	ZRE_1
10,00	10,00	10,86212	-,57475
11,00	10,00	11,00475	-,66984
14,00	12,00	11,43264	,37824
15,00	12,00	11,57527	,28315
20,00	13,00	12,28843	,47439
25,00	13,00	13,00158	-,00105
46,00	19,00	15,99684	2,00212
50,00	15,00	16,56736	-1,04491
59,00	16,00	17,85104	-1,23403
70,00	20,00	19,41998	,38668

تفسير معادلة الانحدار

- يدل على أنه كلما زادت كمية ساعات المذاكرة ساعة واحدة حدث زيادة في التحصيل بمقدار 0.143 درجة

مثـــــــال2

في الجدول التالي ،المتغير x هو تكلفة إنتاج فيلم سينمائي (بملايين الدولارات) والمتغير y هو مردود هذا الفيلم :

التكلفة (x)	62	90	50	35	200	100	95
المردود (y)	65	64	48	57	601	146	47

أرسم مخطط الانتشار

a أوجد معادلة خط الانحدار.

b قدّر مردود فيلم بلغت تكلفته 55 مليون دولار.

c أوجد مقدار الخطأ لفيلم بلغت تكلفته 90 مليون دولار.