SPSS appliqué en sociologie
مطبق في علم الإجتماع SPSS

السداسي الثاني Semestre 2

Cours :
Régression linéaire multiple : Etapes SPSS
انحدار خطي متعدد:
مراحل
SPSS

Dr Yasmine Musette
Université Alger 2
yasminemusette2@gmail.com
25/12/2019

Régression linéaire multiple

" الإنحدار الخطي المتعدد هي إحدى الحلول لملاحظة العلاقات بين متغير كمي تابع و ن متغير كمي مستقل "

(Dominique Laffly , 2006) (C.Chouquet)

امتغيرات مستقلة

"كمية , يمكن أن تكون كيفية تم إعادة ترميزها"

(Rakotomalala , ND)

"المتغيرات الكيفية كادين , المهارات , الإقامة يجب حفظها على شكل متغيرات ثنائية "

(SPSS)

المتغير المستقل يمكن أن يكون ثنائي

Source : http://eric.univ-lyon2.fr/~ricco/cours/cours/Regression_Lineaire_Multiple.pdf
http://eric.univ-lyon2.fr/~ricco/cours/slides/Reg_Multiple_Exogenes_Qualitatives.pdf
http://www.jybaudot.fr/Correl_regress/regsaison.html

نموذج الانحدار الخطي المتعدد

- $\mathbf{Yi = \beta 0 + \beta 1}\ xi1 + \mathbf{\beta 2}\ xi2 + \ldots \mathbf{Bp}\ xip + \epsilon i$

ϵi : خطأ عشوائي(v.a.r.)

التغير في y
الذي لا يتم تفسيره من طرف الارتباط الخطي

$\beta 0, \beta 1, \ldots, \beta p$: معلمات نموذج الإنحدار

X1i, X2i, X3i, …, Xkj
هي المتغيرات المستقلة

Source : Michaël Genin http://cerim.univ-lille2.fr/fileadmin/user_upload/statistiques/michael_genin/Cours/Modelisation/Regression-lineaire-multiple_DU.pdf

محمود محمد طاهر يونس العبادي2011

https://www.iasj.net/iasj?func=fulltext&aId=28387

Hypothèses فرضيات

- لكل قيمة للمتغير المستقل , التوزيع للمتغير التابع توزيع طبيعي
- يجب أن يكون تباين توزيع المتغير التابع ثابتًا لجميع قيم المتغير المستقل
- العلاقة بين المتغير التابع و كل متغير مستقل هي علاقة خطية "إن تغير المتغير التابع لكل زيادة في وحدة واحدة من المتغير المستقل يتبع خطًا مستقيمًا"
- يجب أن تكون جميع الملاحظات مستقلة.

المصدر : SPSS

فرضيات

- **hypothèse d'homoscédasticité** :

تباين الأخطاء هو نفسه لكل ملاحظة(Polomé ،2016)
(تجانس تباينات الأخطاء)
يجب أن يكون تباين الأخطاء نفسه لجميع مستويات المتغير المستقل.

- **الخطأ مستقل عن المتغيرات المستقلة**

→ $COV(x_{ij}, \epsilon_i) = 0$

Indépendance des erreurs :
(أخطاء مستقلة)
عدم وجود ارتباط بين قيم ألأخطاء
يتم تحقيق هذا الشرط لما قيمة
Durbin-Watson
موجود بين 0 et 4
قيمة أقل من 2 تعني وجود ارتباط طردي
قيمة أكبر من 2 تعني وجود ارتباط عكسي أو سالب, قيمة 2 تعني عدم وجود ارتباط
لا يجب أن تكون قيمة الخطأ أكبر من 3 و أقل من 1

فرضيات

- $\epsilon i \sim N(0, \sigma 2)$ (**normalité des résidus**) → tests dans le modèle (اختبارات في النموذج)

التوزيع الطبيعي للأخطاء : "يجب أن يكون الوسط الحسابي يساوي 0, معظم القيم يجب أن تكون قريبة من الصفر "

"

للتأكد من هذه الفرضية يمكن حفظ قيم الأخطاء في قاعدة البيانات و القيام بالاختبار Shapiro-Wilksأو Kolmogorov-Smirnov

الموجودين في الخيار

Explorer

يجب أن يكون الإختبار غير معنوي للتأكد من صحة الفرضية

"

Source :

http://spss.espaceweb.usherbrooke.ca/pages/stat-inferentielles/regression-multiple.php

Michaël Genin http://cerim.univ-lille2.fr/fileadmin/user_upload/statistiques/michael_genin/Cours/Modelisation/Regression-lineaire-multiple_DU.pdf

http://risques-environnement.universite-lyon.fr/IMG/pdf/ectx_ii_l3_ch3_2015-16.pdf

حجم العينة Taille de l’échantillon

« على الأقل 10 أو 20 مشارك لكل متغير مستقل ”

مثلا : نموذج يحتوي على 5 متغيرات مستقلة يستلزم على الأقل0 5
أو 100مشارك

الكثير من المتغيرات المستقلة

Le problème du surajustement :

”النتيجة تنطبق على العينة و ليس على المجتمع , إذا كررنا الدراسة النتائج لا تتكرر . هذا المشكل يحدث لما نطرح الكثير من الأسئلة التي لا تستطيع البيانات الإجابة عنها ”

(Notre traduction)

(Harvey J Motulsky, 2019)

- Source : https://books.google.dz/books?id=5_e1DwAAQBAJ&pg=PA393&lpg=PA393&dq=surajustement+r%C3%A9gression+lin%C3%A9aire+multiple+taille+de+l%C2%B4%C3%A9chantillon+&source=bl&ots=Qe-5AekAxE&sig=ACfU3U1qSuVFVNe9zNki7Azk0VLS3CyOBA&hl=fr&sa=X&ved=2ahUKEwicp5eS2trmAhVc8uAKHXPFAzYQ6AEwBXoECAoQAQ#v=onepage&q=surajustement%20r%C3%A9gression%20lin%C3%A9aire%20multiple%20taille%20de%20l%C2%B4%C3%A9chantillon&f=false

Source Supplémetaire

- https://www.minitab.com/fr-fr/Published-Articles/Les-dangers-du-Surajustement-des-mod%C3%A8les-de-r%C3%A9gression/
- http://eric.univ-lyon2.fr/~ricco/cours/cours/Regression_Lineaire_Multiple.pdf

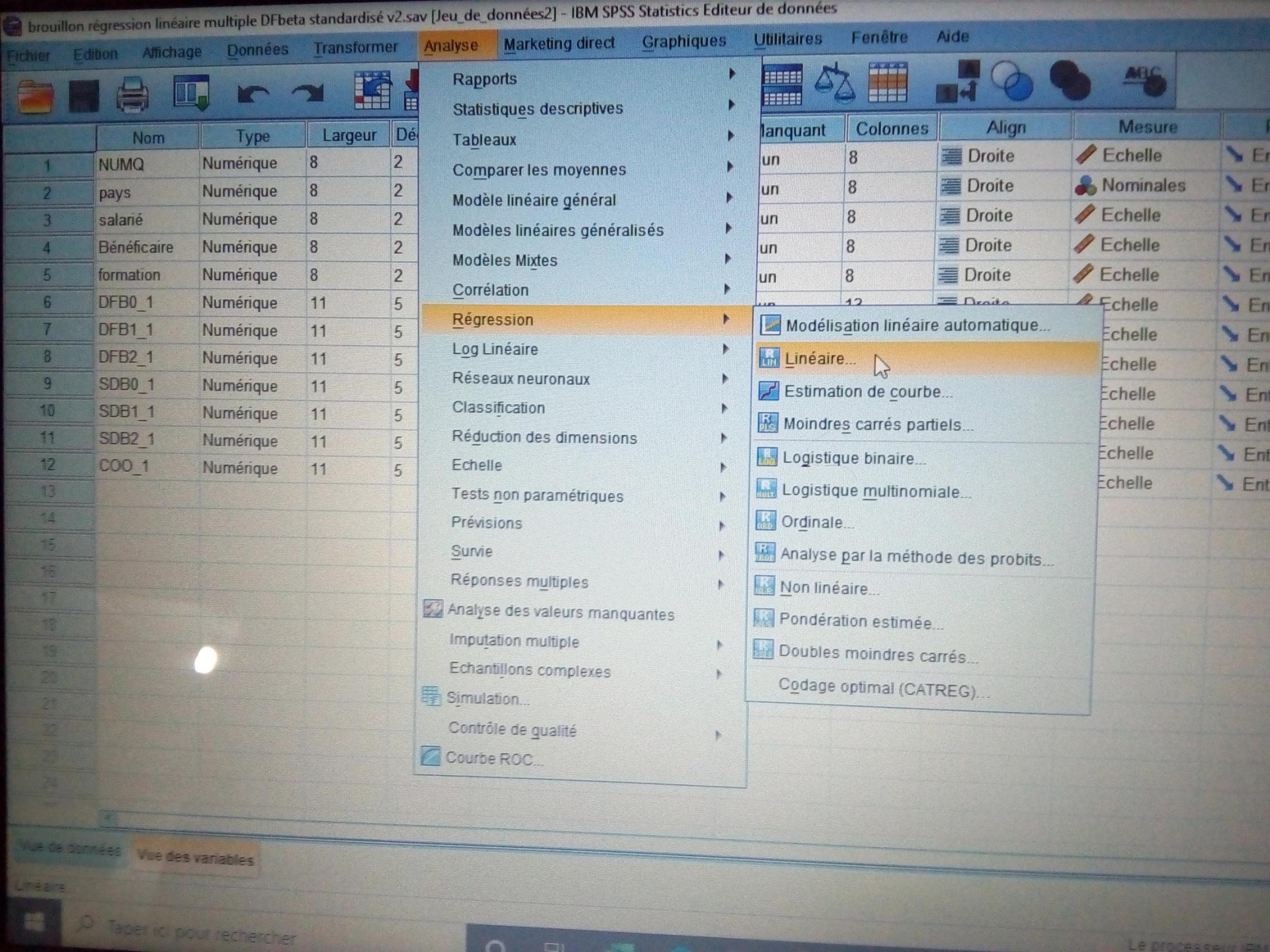
حجم العينة

"مثلا , إذا لديكم عينة تتكون من 20 فيمة و تريد تقييم متوسط حسابي واحد للمجتمع , من المحتمل أن تتحصل على تقييم صحيحة . إذا تريد تقييم متوسطين حسابيين للمجتمع باستعمال نفس حجم العينة , التقييمات لها أقل مصداقية. إذا تريد ثلاث متوسطات حسابية أو أكثر للمجتمع , جودة النتائج يمكن أن تكون سيئة . جودة النتائج تنخفض لما تحاولون الحصول على الكثير من المعلومات من عينة صغيرة . عدد الملاحظات تنخفض لكل عامل..."

(Notre traduction)

(Bruno Scibilia, ND)

مراحل SPSS


brouillon régression linéaire multiple DFbeta standardisé v2.sav [Jeu_de_données2] - IBM SPSS Statistics Editeur de données
Fichier Edition Affichage Données Transformer Analyse Marketing direct Graphiques Utilitaires Fenêtre Aide
Rapports
Statistiques descriptives
Tableaux
Comparer les moyennes
Modèle linéaire général
Modèles linéaires généralisés
Modèles Mixtes
Corrélation
Régression
Log Linéaire
Réseaux neuronaux
Classification
Réduction des dimensions
Echelle
Tests non paramétriques
Prévisions
Survie
Réponses multiples
Analyse des valeurs manquantes
Imputation multiple
Echantillons complexes
Simulation...
Contrôle de qualité
Courbe ROC...
Modélisation linéaire automatique...
Linéaire...
Estimation de courbe...
Moindres carrés partiels...
Logistique binaire...
Logistique multinomiale...
Ordinale...
Analyse par la méthode des probits...
Non linéaire...
Pondération estimée...
Doubles moindres carrés...
Codage optimal (CATREG)...
Nom Type Largeur
1 NUMQ Numérique 8
2 pays Numérique 8
3 salarié Numérique 8
4 Bénéficaire Numérique 8
5 formation Numérique 8
6 DFB0_1 Numérique 11
7 DFB1_1 Numérique 11
8 DFB2_1 Numérique 11
9 SDB0_1 Numérique 11
10 SDB1_1 Numérique 11
11 SDB2_1 Numérique 11
12 COO_1 Numérique 11
Colonnes Align Mesure
Droite Echelle
Droite Nominales
Droite Echelle
Droite Echelle
Droite Echelle
Vue de données
Vue des variables
Linéaire...
Taper ici pour rechercher

brouillon régression linéaire multiple DFbeta standardisé v2.sav [Jeu_de_données2] - IBM SPSS Statistics Editeur de données

Fichier Edition Affichage Données Transformer Analyse Marketing direct Graphiques Utilitaires Fenêtre Aide

	Nom	Type	Largeur	Décimales	Libellé	Valeurs	Manquant	Colonnes	Align	Mesure	Rôle
1	NUMQ	Numérique	8	2	Numéro du que...	Aucun	Aucun	8	Droite	Echelle	Entrée
2	pays	Numérique	8	2							
3	salarié	Numérique	8	2							
4	Bénéficaire	Numérique	8	2							
5	formation	Numérique	8	2							
6	DFB0_1	Numérique	11	5							
7	DFB1_1	Numérique	11	5							
8	DFB2_1	Numérique	11	5							
9	SDB0_1	Numérique	11	5							
10	SDB1_1	Numérique	11	5							
11	SDB2_1	Numérique	11	5							
12	COO_1	Numérique	11	5							
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											

Vue de données Vue des variables

Régression linéaire

NUMQ
pays
Bénéficaire
formation
DFB0_1
DFB1_1
DFB2_1
SDB0_1
SDB1_1
SDB2_1
COO_1

Dépendant :
salarié

Bloc 1 de 1
Précédent Suivant

Variables indépendantes :
Bénéficaire
formation

Méthode : Entrez

Variable de filtrage :
Règle...

Libellés d'observation :

Pondération WLS :

Statistiques...
Tracés...
Enregistrer...
Options...
Style...
Bootstrap...

OK Coller Réinitialiser Annuler Aide

Taper ici pour rechercher

Le processeur IBM SPS

مراحل SPSS

- **Analyse – Régression – linéaire**

يجب وضع المتغير التابع في الخانة« dépendant »
بالضغط على السهم أو سحب المتغير

- وضع المتغيرات المستقلة في الخانة
« variables indépendantes »
بالضغط على السهم أو سحب المتغير

- **اختيار الطريقة :**

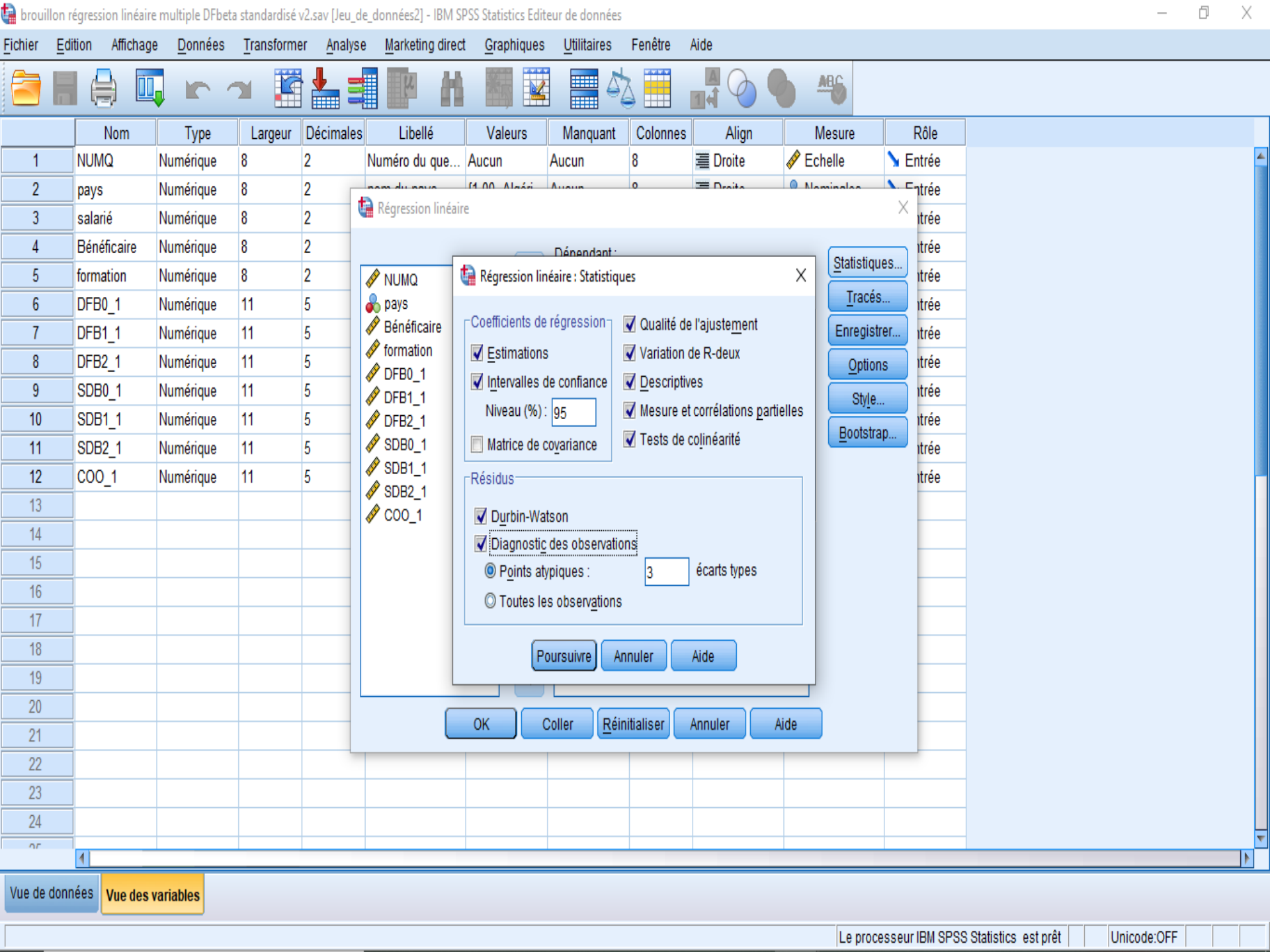
- La méthode entrée :
"تستعمل عادة إلا في حالة معلومات نظرية معينة "

- Pas à pas
- Eliminer bloc
- Descendante
- Ascendante

مراحلSPSS

- **(en plus) Case variable de filtrage :**

” يمكن إختيار متغير لحصر التحليل على جزء من العينة المكونة من المشاركين الذين تحصلوا قيمة أو أكثر معينة لهاذا المتغير“


brouillon régression linéaire multiple DFbeta standardisé v2.sav [Jeu_de_données2] - IBM SPSS Statistics Editeur de données
Fichier Edition Affichage Données Transformer Analyse Marketing direct Graphiques Utilitaires Fenêtre Aide
Nom Type Largeur Décimales Libellé Valeurs Manquant Colonnes Align Mesure Rôle
1 NUMQ Numérique 8 2 Numéro du que... Aucun Aucun 8 Droite Echelle Entrée
2 pays Numérique 8 2 Entrée
3 salarié Numérique 8 2 Entrée
4 Bénéficaire Numérique 8 2 Entrée
5 formation Numérique 8 2 Entrée
6 DFB0_1 Numérique 11 5 Entrée
7 DFB1_1 Numérique 11 5 Entrée
8 DFB2_1 Numérique 11 5 Entrée
9 SDB0_1 Numérique 11 5 Entrée
10 SDB1_1 Numérique 11 5 Entrée
11 SDB2_1 Numérique 11 5 Entrée
12 COO_1 Numérique 11 5 Entrée
Régression linéaire
Dépendant :
NUMQ
pays
Bénéficaire
formation
DFB0_1
DFB1_1
DFB2_1
SDB0_1
SDB1_1
SDB2_1
COO_1
Statistiques...
Tracés...
Enregistrer...
Options
Style...
Bootstrap...
Régression linéaire : Statistiques
Coefficients de régression
Estimations
Intervalles de confiance
Niveau (%) : 95
Matrice de covariance
Qualité de l'ajustement
Variation de R-deux
Descriptives
Mesure et corrélations partielles
Tests de colinéarité
Résidus
Durbin-Watson
Diagnostic des observations
Points atypiques : 3 écarts types
Toutes les observations
Poursuivre Annuler Aide
OK Coller Réinitialiser Annuler Aide
Vue de données Vue des variables
Le processeur IBM SPSS Statistics est prêt
Unicode:OFF

مراحلSPSS

- إضغط على **statistiques**
  - إضغط بجانب

- **Estimation (Estimates)** :

يسمح بالحصول على

المعلمة B (le coefficient de régression *B)*

الخطأ القياسي ل B (l'erreur standard de B)

المقدر القياسي ل B (le coefficient bêta standardisé)

قيمة t ل B

ومستوى معنوية ل t .

le niveau de signification bilatéral de *t*

مراحلSPSS

- **Qualité de l’ajustement (model fit) :**
  - **يوفر الاختبار للتقييم الكلى للنموذج** (F)
  - le R multiple
  - le R^2 correspondant
  - le R^2 ajusté
  - l´erreur standard de l'estimation
  - tableau d'analyse de variance

- **Variance de R-deux (R squared change) :**

”التغير في R-Deux

بعد زيادة مجموعة من المتغيرات المستقلة

مراحلSPSS

- **Descriptives (caractéristiques) :**

le nombre d'observations valides **يوفر عدد الملاحظات**

la moyenne **المتوسط الحسابي**

l'écart type **الإنحراف المعياري**

لكل متغير

le nombre d'observations pour chaque corrélation sont également affichés

عدد الملاحظات لكل إرتباط

Une matrice de corrélations **مصفوفة الإرتباط**

Le niveau de signification unilatéral **مستوى المعنوية**

مراحلSPSS

(**suite statistiques**) :

- **Mesure et corrélations partielles (part and partial correlations) :**

Corrélation partielle.

(الإرتباط الجزئي)

الإرتباط بين المتغير التلبع و المستقل لما تأثير الخطي للمتغيرات الأخرى المستقلة للنموذج تم نزعه من المتغيرين

Mesure

.

الإرتباط بين المتغير التابع و متغير المستقل لما التاثير الخطي للمتغيرات الأخرى المستقلة للنموذج تم نزعها من المتغير المستقل .

هو مرتبط بتغير

R-deux

لما تم إضافة متغير للمعادلة

(Parfois appelée corrélation semi-partielle.)

(الارتباط شبه الجزئي)

مراحلSPSS

- **Tests de colinéarité (colineraity diagnostics) : «**évaluation de la multicolinéarité dans le modèle (VIF). »

La colinéarité (ou multicolinéarité)

(مشكلة التعدد الخطي)

هي حالة وجود إرتباط خطي بين متغير مستقل ومتغيرات اخرى مستقلة

يكون المتغير المستقل دالة خطية للمتغيرات المستقلة الأخرى

يتم توفير:

–Les valeurs propres de la matrice des produits croisés dimensionnés et non centrés

– ***دليل الشرط***Les *indices de condition*

–les proportions de décomposition de variance **نسب تحلل تباين**

–les facteurs d'inflation de la variance ***(VIF)*** ***معامل تضخم التباين***

– et les tolérances ***معامل التباين المسامح***

pour les variables individuelles

مراحل SPSS

(**suite statistiques**) :

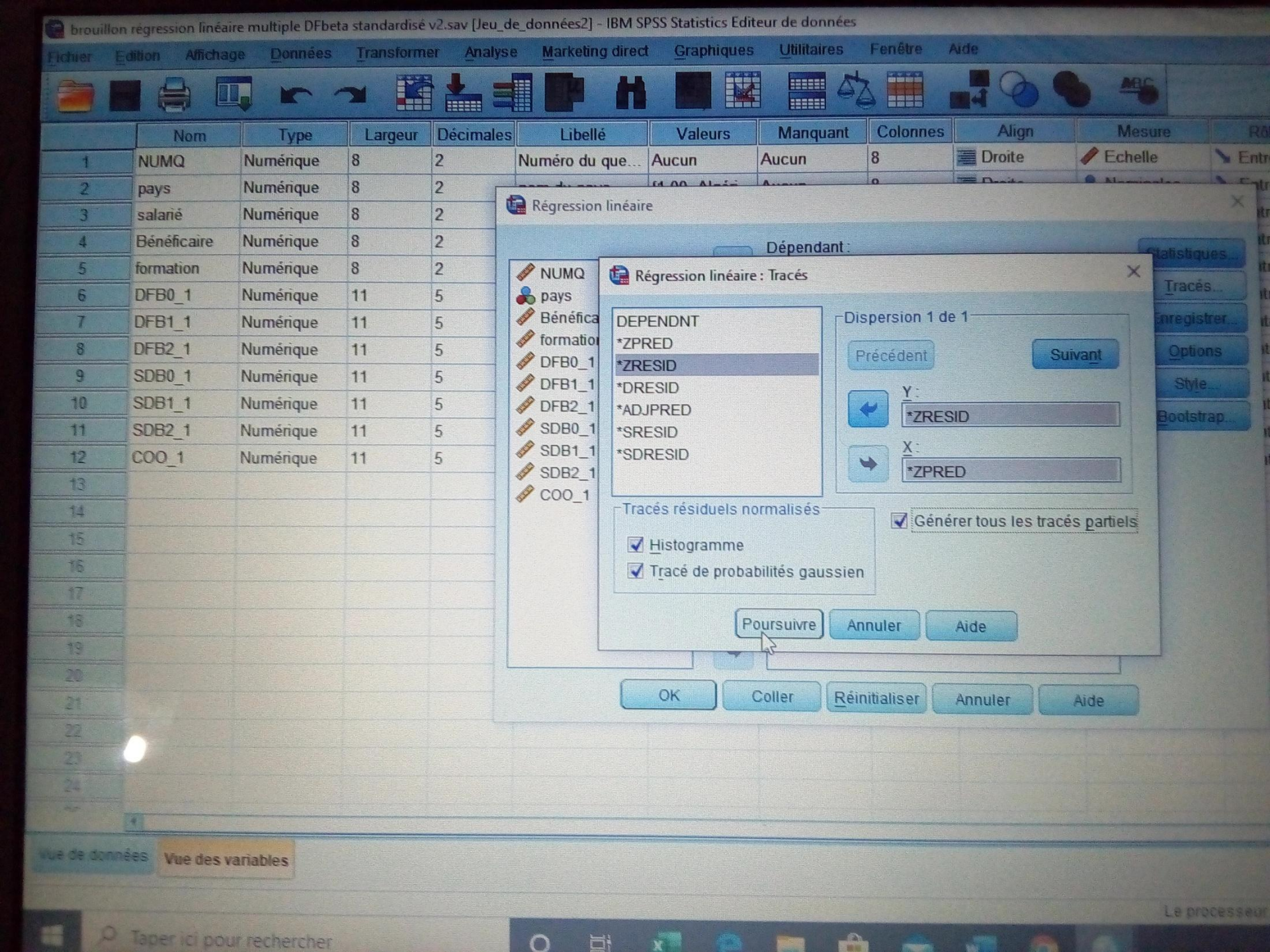
- **Intervalle de confiance :** (فترات الثقة**)**
- **«** *affiche les intervalles de confiance avec le niveau spécifié de confiance pour chaque coefficient de régression ou une matrice de covariance.* »

- **Durbin watson :**"يقيم إستقلال الأخطاء"

- **Diagnostic des observations (casewise diagnostic) :**
(تشخيص المشاهدات)

يوفر "القيمة الملاحظة للمتغير التابع و القيمة المتنبأ بها , الخطأ المعياري لكل ملاحظة
يبين ماهي الملاحظات التي لها خطا معياري أكثر من 2 أو 3 إنحراف معياري "

2 ou 3 é.-t. » (é.-t veut dire écart type)
بمعنى
2 أو 3 إنحراف معياري


brouillon régression linéaire multiple DFbeta standardisé v2.sav [Jeu_de_données2] - IBM SPSS Statistics Editeur de données
Fichier Edition Affichage Données Transformer Analyse Marketing direct Graphiques Utilitaires Fenêtre Aide
Nom Type Largeur Décimales Libellé Valeurs Manquant Colonnes Align Mesure
1 NUMQ Numérique 8 2 Numéro du que... Aucun Aucun 8 Droite Echelle
2 pays Numérique 8 2
3 salarié Numérique 8 2
4 Bénéficaire Numérique 8 2
5 formation Numérique 8 2
6 DFB0_1 Numérique 11 5
7 DFB1_1 Numérique 11 5
8 DFB2_1 Numérique 11 5
9 SDB0_1 Numérique 11 5
10 SDB1_1 Numérique 11 5
11 SDB2_1 Numérique 11 5
12 COO_1 Numérique 11 5
Régression linéaire
Dépendant :
NUMQ
pays
Bénéfica
formatio
DFB0_1
DFB1_1
DFB2_1
SDB0_1
SDB1_1
SDB2_1
COO_1
Statistiques...
Tracés...
Enregistrer...
Options
Style...
Bootstrap...
Régression linéaire : Tracés
DEPENDNT
*ZPRED
*ZRESID
*DRESID
*ADJPRED
*SRESID
*SDRESID
Dispersion 1 de 1
Précédent
Suivant
Y :
*ZRESID
X :
*ZPRED
Tracés résiduels normalisés
Histogramme
Tracé de probabilités gaussien
Générer tous les tracés partiels
Poursuivre
Annuler
Aide
OK
Coller
Réinitialiser
Annuler
Aide
Vue de données
Vue des variables
Taper ici pour rechercher

مراحل SPSS

- **إضغط على : tracés (plots)**

ضع « ZRESID » في الخانة Y
ضع « ZPRED » في الخانة X

إضغط في « Tracé résiduel normalisé » على :

Histogramme
Tracé de probabilités gaussien (Normal probability plot)
Produce all partial plots (Générer tous les tracés partiels)

إضغط على Continue(Poursuivre)

مراحل SPSS

الإطار « Tracé résiduel normalisé »

يسمح بتمثيل توزيع الاخطاء عن طريق

(Histogramme ,Tracé de probabilités gaussien)

ما يسمح بالتحقق من فرضية التوزيع الطبيعي للأخطاء

مراحل SPSS

”التمثيل البياني الذي يربط بين القيم المتنبأ بها (*ZPRED)
و الأخطاء (*ZRESID) المعيارية يبين إحترام أو عدم إحترام شرط الإنسجام (التوزيع العشوائي للنقاط حول الصفر) و الخطي (ميل النقاط إلى الظهور حول خط)“

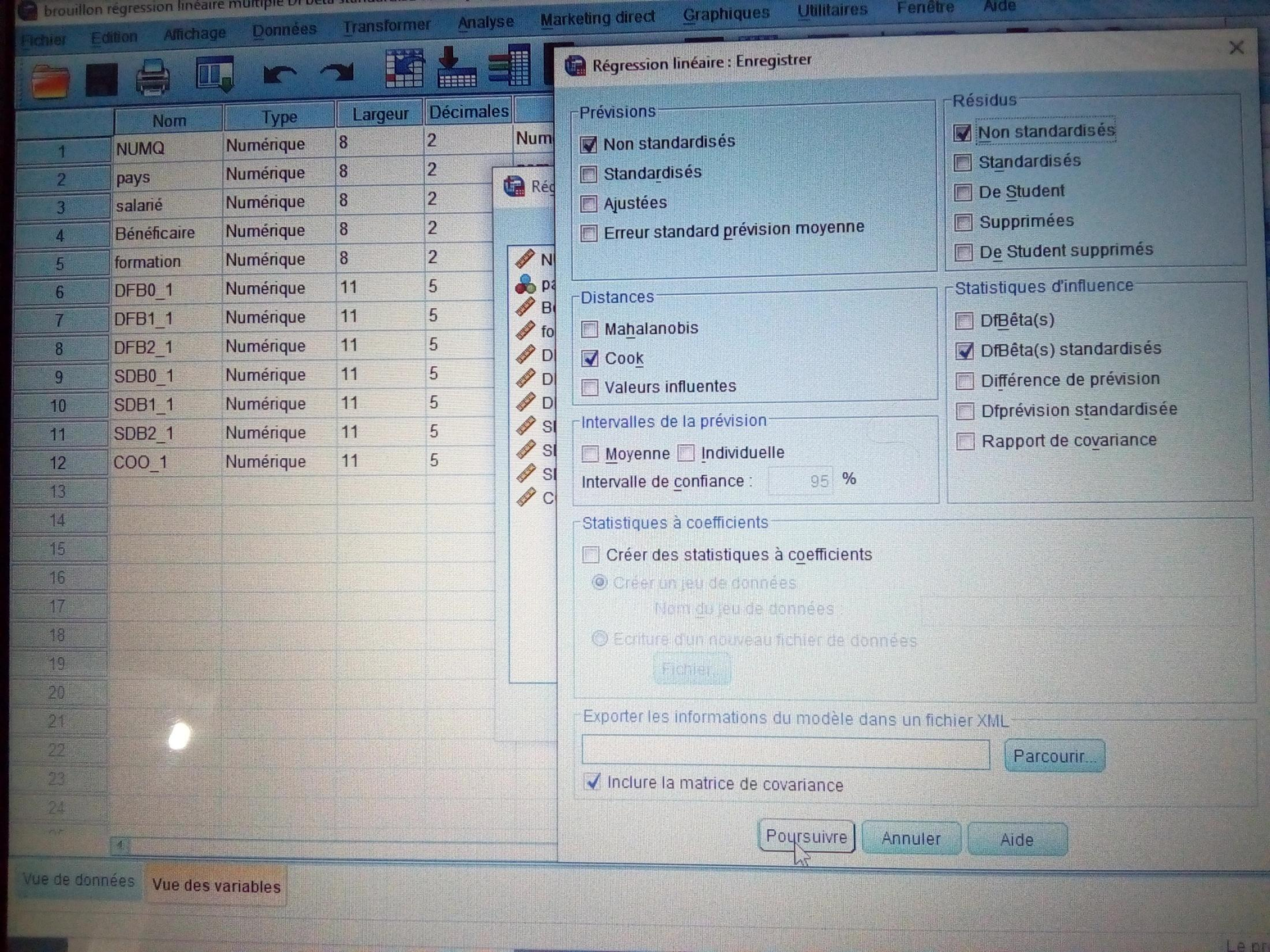
(ترجمتنا)

مراحل SPSS

”يبين أسحاب من النقاط من البواقي لكل متغير مستقل و أخطاء المتغير التابع لما يتم إنحدار خطي بين المتغيرين بدون أخذ بعين إعتبار بقية المتغيرات المستقلة.

يجب وجود على الأقل متغيرين مستقلين في المعادلة لإنتاج التمثيل الجزئي ”

(SPSS)


brouillon régression linéaire multiple DFbeta standardisés
Fichier Edition Affichage Données Transformer Analyse Marketing direct Graphiques Utilitaires Fenêtre Aide
Nom Type Largeur Décimales
1 NUMQ Numérique 8 2
2 pays Numérique 8 2
3 salarié Numérique 8 2
4 Bénéficaire Numérique 8 2
5 formation Numérique 8 2
6 DFB0_1 Numérique 11 5
7 DFB1_1 Numérique 11 5
8 DFB2_1 Numérique 11 5
9 SDB0_1 Numérique 11 5
10 SDB1_1 Numérique 11 5
11 SDB2_1 Numérique 11 5
12 COO_1 Numérique 11 5
Régression linéaire : Enregistrer
Prévisions
Non standardisés
Standardisés
Ajustées
Erreur standard prévision moyenne
Résidus
Non standardisés
Standardisés
De Student
Supprimées
De Student supprimés
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 jeu de données
Nom du jeu de données
Ecriture d'un nouveau fichier de données
Fichier...
Exporter les informations du modèle dans un fichier XML
Parcourir...
Inclure la matrice de covariance
Poursuivre
Annuler
Aide
Vue de données
Vue des variables

مراحلSPSS

إضغط على « enregistrer » (save)
إضغط بجانب :

- **Prévision : Non standardisés**

يسمح بالحصول على "القيمة المتنبأ بها من طرف النموذج للمتغير التابع

- **Distances : Cook**

"تسمح بمعرفة بكم الأخطاء لكل الملاحظات ستتغير في حالة حدف ملاحظة ما من حساب معاملات الإنحدار الخطي المتعدد .

إذا قيمة la distance de Cook مرتفعة ذالك يعني أن حذف ملاحظة سيغير بشكل معتبر قيمة المعاملات

- **Résidus : Standardisé**

Résidus
القيمة الحقيقية للمتغير التابع (الملاحظ) تم الطرح منها القيمة المينبأ بها من طرف معادلة *الإنحدار الخطي المتعدد*

المعيارية :
الخطأ تم تقسيمه على الإنحراف المعياري المقدر .
لديهم وسط حسابس يساوي 0 و إنحراف معياري يساوي 1
الأخطاء المعيارية تسمى *résiduels de Pearson, les résiduels standardisés*

مراحلSPSS

- **Statistique d’influence : Dfbêta(s) standardisés**

DfBêta(s)

الفرق في ما يخص

المتمثل في التغيير في ما يخص معاملات الإنحدار *bêta* لخطي المتعدد

الناتج عن حذف ملاحظة معينة. يتم حساب قيمة لكل معامل في النموذج الخطي المتعدد بما فيها الثابت

DfBêta(s) standardisée.

الفرق المعياري لقيمة *bêta*

التغيير في معامل الإنحدار الخطي المتعدد الناتج عن حذف ملاحظة ما

يتم حساب قيمة لكل معامل في النموذج بما فيها الثابت

- **inclure la matrice de covariance**

مراحلSPSS

في الخانة « Option »نترك كما هي :

la probabilité de F (use of probability of F)

"الخانة الأخيرة تسمح بتعيين معايير إختيار طرق دخول بالتدرج
(Ascendante ou descendante - *stepwise*)
يمكن إستخدام قيمة الإحتمال المتعلق بقيمة F (بمعنى قيمة p)أو قيمة F
لإدخال أو سحب متغيرات , من الأفضل القيم الموجودة إلا إذا أردتم أن تكون معايير إدخال أو
سحب متغيرات من النموذج تسمح بالحصول على عدد أقل أو أكثر من المتغيرات"
يتم إدخال متغير إلى النموذج إذا مستوى المعنوية لقيمة
Fأصغر من القيمة Entrée
يتم حذف المتغير إذا هذا المستوى هوأعلى من القيمة
Suppression

يجب أن تكون القيمة Entréeأقل من القيمة Suppression وكلاهما يجب أن تكون موجبة.
لإدخال عدد أكبر من المتغيرات في النموذج يجب تخفيض القيمة Entrée
لسحب عدد أكبر من المتغيرات من النموذج يجب تخفيض القيمةSuppression (SPSS)

مراحلSPSS

في الخانة **« option »نترك كما هي :** -

- **Inclure terme constant dans l’équation (include constant in equation)**

سيدخل الثابت في المعدلة

- **Exclure toute observation incomplète (Exclude case listwise) :**

يمكن تحديد ما تريدون القيام به بالقيم الضائعة
(les valeurs manquantes)
من الأفضل نترك ما اخيرى من طرف
SPSS
لأن سحب كل ملاحظة غير كاملة (قيمة ضائعة) تسمح بالحفاظ على نفس عدد الملاحظات

- **لا نقوم بتغيير في الخانة « style »**

مراحلSPSS

القيام بإنحدار خطي متعدد بالمتغير

dummy variable

(بعد إعادة ترميز المتغيرات)

https://www.youtube.com/watch?v=ScV2FOlx41A&t=54s