

SOCIO 532.2
STATISTIQUES & INFORMATIQUE APPLIQUÉES AUX SCIENCES SOCIALES
 © El Hadj Touré, Ph.D.

LABO SPSS # 3
Régression linéaire approfondie

Strictement parlant, la régression linéaire simple est appropriée lorsque les deux variables sont quantitatives. Cependant, il est possible d'inclure une VI qualitative dans un modèle de régression. Nous verrons donc comment approfondir l'analyse de régression linéaire simple en incluant une VI qualitative dichotomique, puis une VI qualitative non dichotomique.

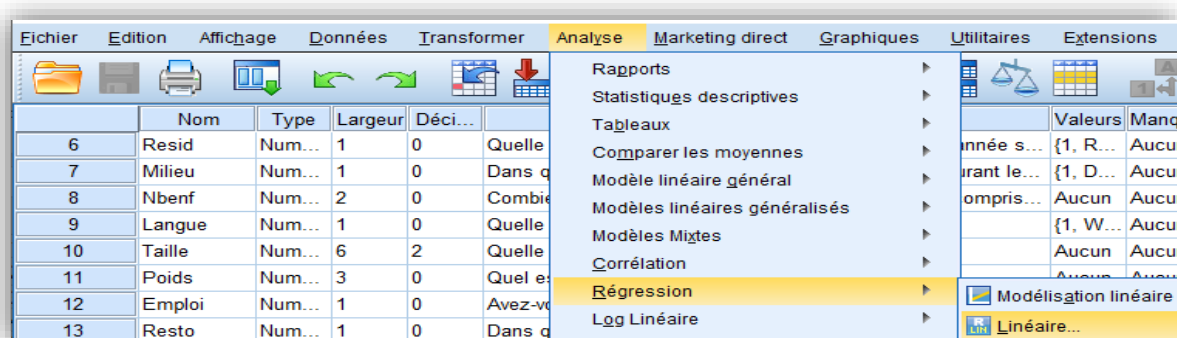
1. Régression linéaire simple : quelques éléments de rappel

1.1. Analyse des données

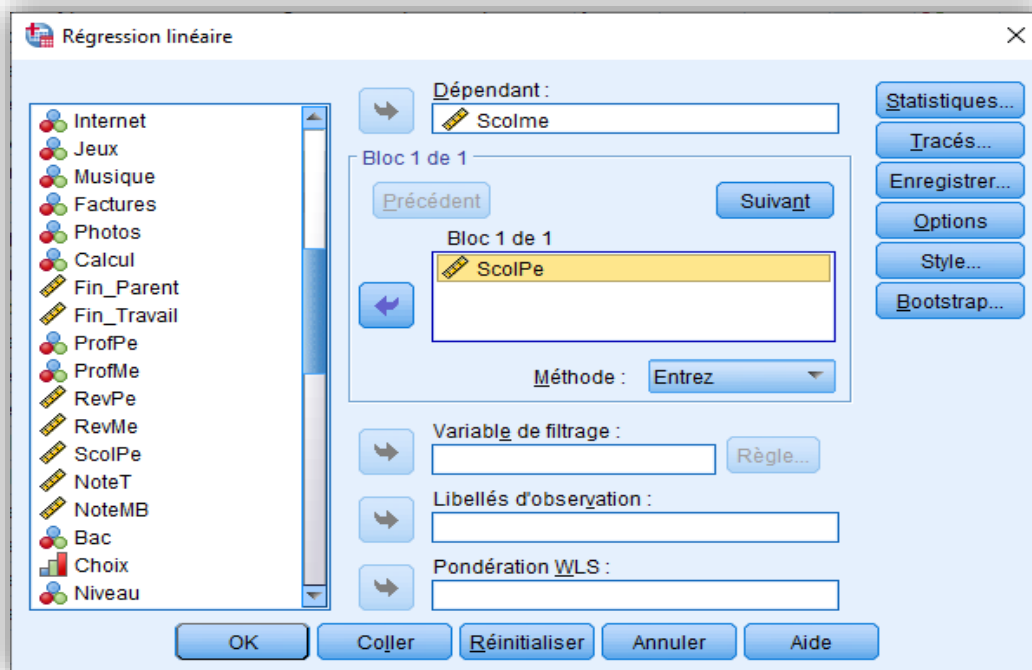
Ouvrons la base de données « [Sondage EtudiantsSocioL2 2021\(labo7\)](#) » et revenons sur la relation entre le nombre d'années de scolarité du père (ScolPe) et le nombre d'années de scolarité de la mère (ScolMe) chez les étudiants inscrits en L2 de sociologie à l'UGB, soit deux variables quantitatives.

Scolarité du père (ScolPe) \longrightarrow Scolarité de la mère (ScolMe)

Analyse
Régression
Linéaire



Une fenêtre de dialogue apparaît ! Cliquez sur la variable dépendante (scolme), puis indépendante (scolpe). Contrairement à la corrélation, la régression accorde une importance à l'identification de la VD et de la VI.



Validez pour obtenir la page des résultats !

Récapitulatif des modèles^b

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	.597 ^a	.357	.350	4.770

a. Prédicteurs : (Constante), ScolPe Combien d'années de scolarité, approximativement, votre père a-t-il complétées ?

b. Variable dépendante : Scolme Combien d'années de scolarité, approximativement, votre mère a-t-elle complétées ?

ANOVA^a

Modèle		Somme des carrés	ddl	Carré moyen	F	Sig.
1	Régression	1198.438	1	1198.438	52.668	.000 ^b
	Résidu	2161.686	95	22.755		
	Total	3360.124	96			

a. Variable dépendante : Scolme Combien d'années de scolarité, approximativement, votre mère a-t-elle complétées ?

b. Prédicteurs : (Constante), ScolPe Combien d'années de scolarité, approximativement, votre père a-t-il complétées ?

Modèle		Coefficients non standardisés		Coefficients standardisés		
		B	Erreur standard	Bêta	t	Sig.
1	(Constante)	3.018	.789		3.826	.000
	ScolPe Scolarité du père	.562	.077	.597	7.257	.000

a. Variable dépendante : Scolme Combien d'années de scolarité, approximativement, votre mère a-t-elle complétées ?

Comme les sorties ont été analysées précédemment, à vous de vous rappeler !

1.2. Le coefficient de régression bêta

Modèle		Coefficients non standardisés		Coefficients standardisés		
		B	Erreur standard	Bêta	t	Sig.
1	(Constante)	3.018	.789		3.826	.000
	ScolPe Scolarité du père	.562	.077	.597	7.257	.000

a. Variable dépendante : Scolme Combien d'années de scolarité, approximativement, votre mère a-t-elle complétées ?

Le coefficient de régression **b** (0,562) est un coefficient non standardisé. Dans le cas de régressions où les prédicteurs analysés ne sont pas mesurés de la même façon, le coefficient bêta permet de déterminer lesquels ont un impact plus important sur la variable dépendante. Le coefficient bêta β est un coefficient de régression standardisé. Il est de **0,597** d'après le tableau ci-dessous. Pour l'obtenir, il suffit de transformer le coefficient de régression b par l'équation :

$$\beta = b (S_x / S_y)$$

Autrement dit, il faut multiplier le coefficient de régression b par le rapport entre l'écart-type de la variable X (VI) et l'écart-type de la variable Y (VD).

	N	Minimum	Maximum	Moyenne	Ecart type
ScolPe	97	0	21	8,04	6,288
Scolme	97	0	18	7,54	5,916

Sachant que le coefficient de régression $b = 0,562$, et que l'écart-type de ScolPe = 6,288 et l'écart-type de ScolMe = 5,916, on peut retrouver ainsi le bêta :

$$\beta = 0,562 * (6,288 / 5,916) = \mathbf{0,597}.$$

Interprétation statistique : Le coefficient bêta est de 0,597. Il suggère que lorsque la scolarité du père augmente d'un écart-type (soit 6,288), la scolarité de la mère

augmente de 0,597 écart-type (soit $0,597 * 5,916 = 3,32$). Ce qui signifie précisément que chaque fois que la scolarité du père augmente de 6,288 années, celle de la mère augmente proportionnellement de 3,32 années.

En régression simple, le coefficient Bêta est l'équivalent du coefficient de corrélation : donc le coefficient de corrélation simple $R = \beta$. Précisément : $\beta = R = 0,597$.

Coefficient standardisé « Bêta » (β) donne la même information que le coefficient de régression (pente **b**), mais sur une base standardisée selon l'écart-type. Le coefficient « Bêta » représente le changement en écart-type de Y pour une augmentation d'un écart-type de X. Il est approprié pour comparer la force de l'effet respectif de plusieurs prédicteurs dont l'échelle de mesure est différente.

2. Régression linéaire simple avec une VI dichotomique

2.1. Analyse des données

La procédure de régression avec une VI qualitative dichotomique est la même que pour la régression linéaire simple. La seule différence est qu'il faut vous assurer que la VI dichotomique est codée 0/1 de façon à faciliter l'interprétation des résultats de l'analyse de régression.

Changeons maintenant de prédicteur (VI), et intéressons-nous à la relation entre la profession du père (**ProfPe2**) et le nombre d'années de scolarité de la mère (**ScolMe**) chez les étudiants inscrits en L2 de sociologie de l'UGB. La question de recherche est la suivante : *La profession du père a-t-elle un effet sur la scolarité de la mère chez les étudiants ? Précisément dit, l'appartenance du père à la catégorie des cadres/entrepreneurs peut-elle prédire la scolarité de la mère chez les étudiants ?*

Profession du père (ProfPe2) —————> Scolarité de la mère (ScolMe)

Étant donné que la VI se présente sous forme dichotomique (Cadre/entrepreneur Vs Autre profession) et que la VD est de nature quantitative (ScolMe), on peut toujours utiliser la régression linéaire. Toutefois, il faut vous assurer que la VI est codée 0/1. Sortons donc la distribution de fréquences et de % de la variable **ProfPe2**.

		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	,00 Autre profession	73	70,9	76,0	76,0
	1,00 Cadre ou entrepreneur	23	22,3	24,0	100,0
	Total	96	93,2	100,0	
Manquant	Système	7	6,8		
Total		103	100,0		

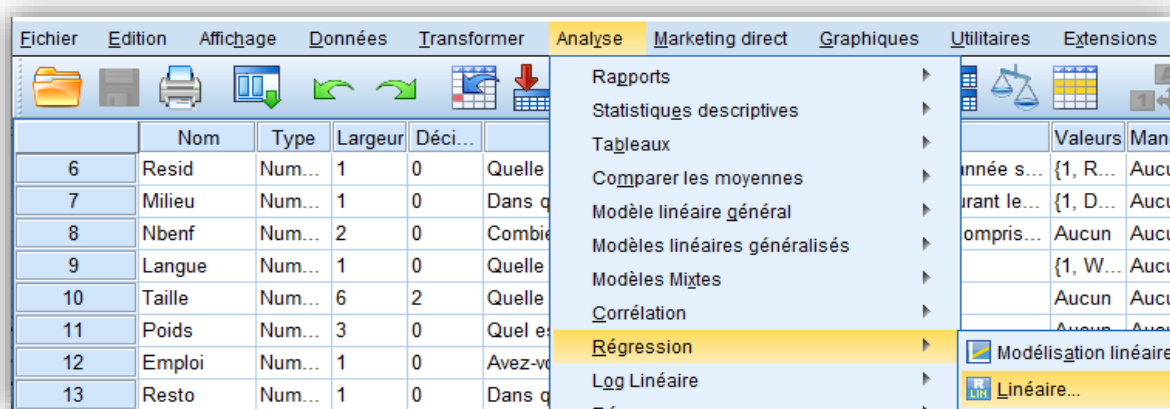
La profession du père comporte deux catégories codées 0/1 : 0 = Autre profession; 1 = Cadre/entrepreneur. CADRE/ENTREPRENEUR réfère à la **catégorie d'intérêt (1)** alors que AUTRE PROFESSION est la **catégorie de référence (0)**. On interprétera ainsi : lorsqu'on passe d'une autre profession à la profession Cadre/entrepreneur du

père (0 à 1), comment régressera-t-il le nombre d'années de scolarité de la mère ?

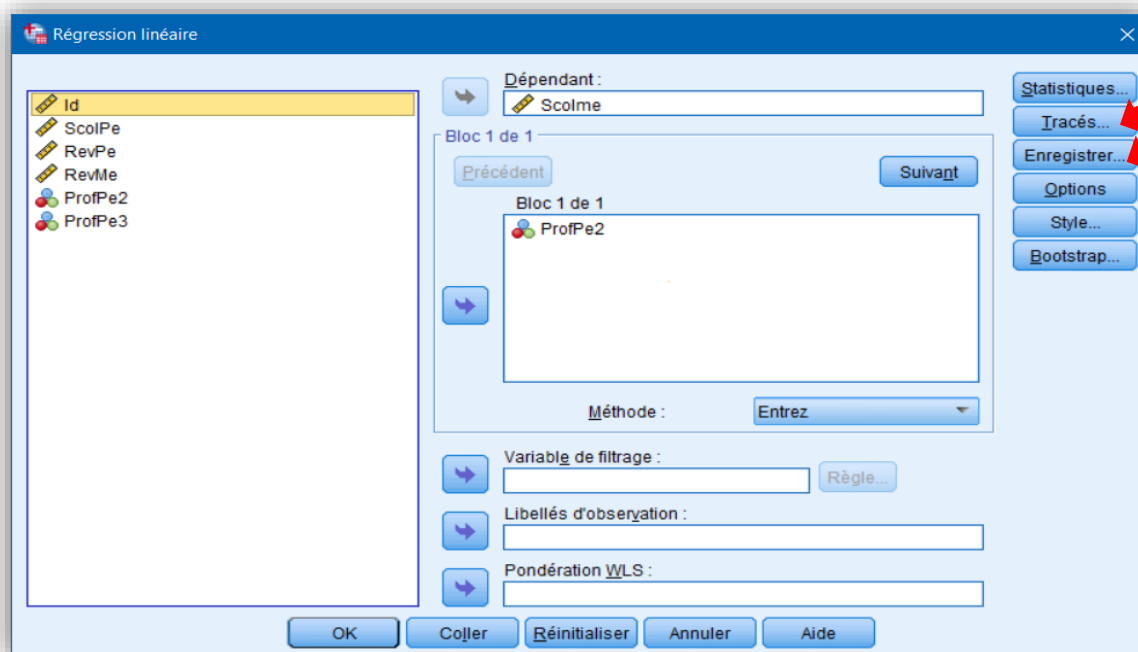
Évidemment, si la VI dichotomique est codée 1 et 2, il faut la recoder 0/1 en utilisant la procédure **CRÉATION DE VARIABLES** de SPSS (cf. labo 6). L'idée est de créer une nouvelle variable recodée en signifiant, par exemple, à SPSS que 1 devient 0 et que 2 devient 1. Ou inversement, selon qu'une catégorie est considérée comme la référence (0) ou la catégorie d'intérêt (1). Cela déterminera le signe du coefficient de régression et donc la façon dont il sera interprété.

Par conséquent, on peut procéder à l'analyse de régression linéaire simple avec la VI dichotomique **ProfPe2**. Suivons à nouveau la procédure :

Analyse Régression Linéaire

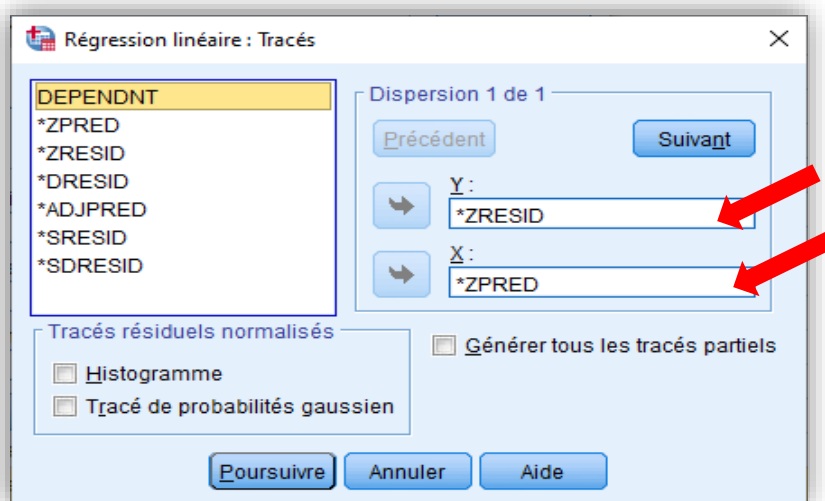


Une fenêtre de dialogue apparaît ! Cliquez sur la variable dépendante (scolme), puis indépendante (CadrEntr).

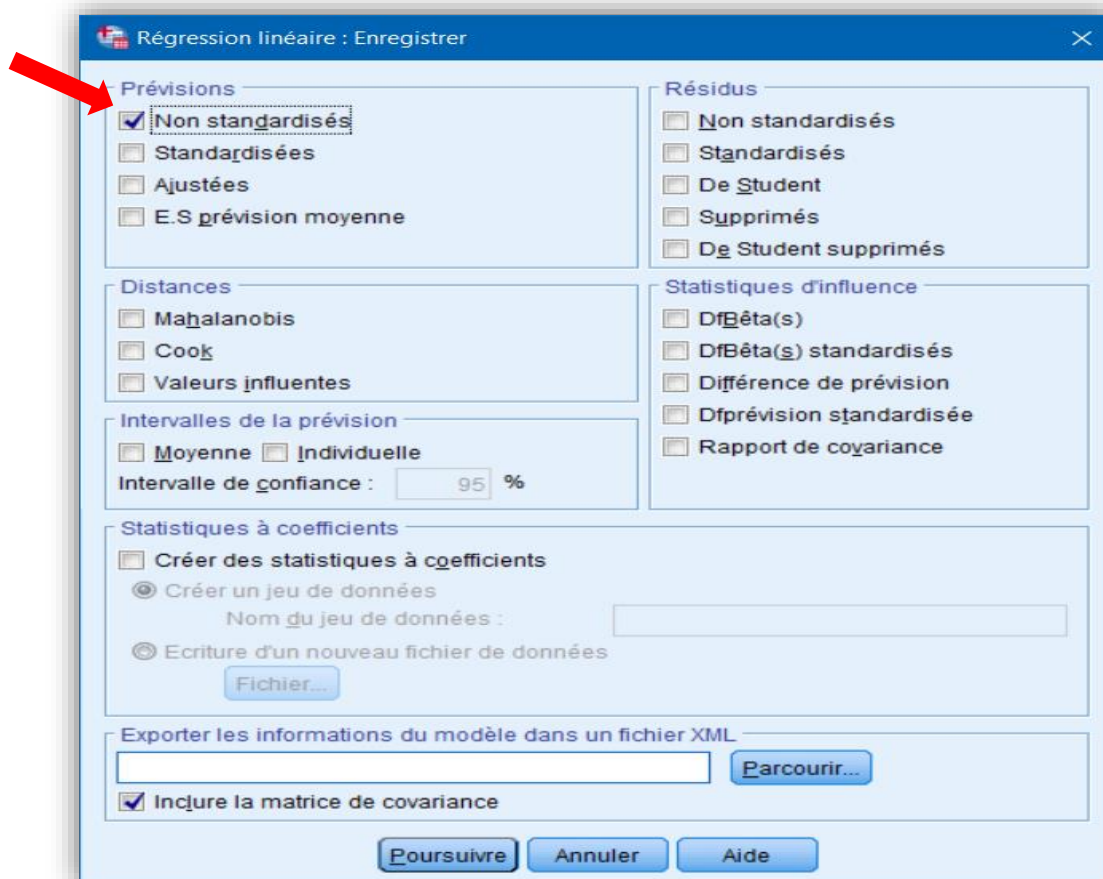


Pour procéder au diagnostic du modèle de régression, cliquez sur Tracés, et mettez les valeurs prédites standardisées (ZPRED) sur l'axe X et les valeurs résiduelles

standardisées (ZRESID) sur l'axe Y.



Poursuivez! Il est possible de sortir les valeurs prédites de la scolarité de la mère (Y), selon que le père est d'une AUTRE PROFESSION (X=0) ou de profession CADRE/ENTREPRENEUR (X=1). Pour ce faire, cliquez sur ENREGISTRER pour voir apparaître la fenêtre ci-dessous ! Puis, cochez la cage **Non standardisés** sous **Prévisions**!



Poursuivez et validez le tout pour voir apparaître les résultats.

Récapitulatif des modèles^b

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,320 ^a	,102	,092	5,671

a. Prédicteurs : (Constante), CadrEntr Profession du père
 b. Variable dépendante : Scolme Combien d'années de scolarité, approximativement, votre mère a-t-elle complétées ?

Le tableau ci-dessus montre que le coefficient de corrélation R est de 0,32, traduisant l'existence d'une relation positive et faible à modérée entre la profession du père et le nombre d'années de scolarité de la mère.

Le R-deux de 0,102 signifie que la profession du père explique une proportion de 10,2% de la variation dans la scolarité de la mère. De même, nous réduisons de 10,2% nos erreurs de prédiction de la scolarité de la mère (VD) quand nous connaissons la profession du père (VI). Finalement, 89,8 % de la variation reste à expliquer.

ANOVA^a

Modèle		Somme des carrés	ddl	Carré moyen	F	Sig.
1	Régression	322,550	1	322,550	10,031	,002 ^b
	Résidu	2829,672	88	32,155		
	Total	3152,222	89			

a. Variable dépendante : Scolme Combien d'années de scolarité, approximativement, votre mère a-t-elle complétées ?
 b. Prédicteurs : (Constante), CadrEntr Profession du père

Le coefficient de corrélation est significatif, puisque la probabilité de commettre une erreur en rejetant l'hypothèse nulle est de 0,002 ($p < 0,01$).

Coefficients^a

Modèle		Coefficients non standardisés		Coefficients standardisés		
		B	Erreur standard	Bêta	t	Sig.
1	(Constante)	6,368	,688		9,260	,000
	CadrEntr Profession du père	4,405	1,391	,320	3,167	,002

a. Variable dépendante : Scolme Combien d'années de scolarité, approximativement, votre mère a-t-elle complétées ?

D'une part, la constante signifie que la valeur estimée de la scolarité de la mère est de 6,368 années lorsque la profession du père est autre ($X=0$).

D'autre part, la pente signifie que lorsque l'on passe d'une profession autre ($X=0$) à une profession de cadre/entrepreneur du père ($X=1$), la scolarité prédite de la mère augmente de 4,405 années, indiquant ainsi une relation positive.

Voici l'équation de la droite de régression :

$$Y = a + b (X)$$

$$\text{Scolarité de la mère} = 3,018 + 0,562 (\text{Profession du père})$$

Pour un étudiant, prédisons la scolarité de la mère si le père est un cadre/entrepreneur ($X=1$).

$$\hat{Y} = 6,368 + 4,405 (1) = 6,368 + 4,405 = 10,773$$

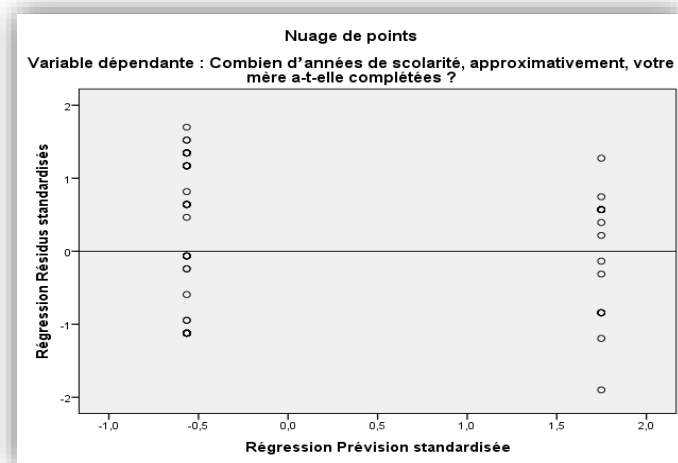
La scolarité prédite des mères est de 10,773 ans chez les pères cadres/entrepreneurs ($X=1$). Or, nous savons déjà que la scolarité prédite des mères est de 6,368 chez les pères d'une autre profession ($X=0$). On retrouve les deux valeurs prédites que nous venons d'enregistrer sous forme de variable (PRE_1) dans la base de données :

1: Id	1	Visible : 7 variables sur 7				
Id	ScolPe	Scolme	CadrEntr	RevPe	RevMe	PRE_1
1	6	10	Cadre ou entrepreneur	250000	175000	10,773
2	0	0	Autre profession	.	.	6,368
3	6	15	Autre profession	.	.	6,368
4	18	6
5	11	6	Autre profession	45000	50000	6,368

Qu'en est-il du diagnostic des résidus du modèle de régression ?

Statistiques des résidus ^a					
	Minimum	Maximum	Moyenne	Ecart type	N
Prévision	6,37	10,77	7,44	1,904	90
Résidu	-10,773	9,632	,000	5,639	90
Prévision standardisée	-,566	1,748	,000	1,000	90
Résidu standardisé	-1,900	1,699	,000	,994	90

a. Variable dépendante : Scolme Combien d'années de scolarité, approximativement, votre mère a-t-elle complétées ?



Analyse ou interprétation statistique des résidus :

1. **Linéarité** de la relation entre la VD et la VI :
Postulat peu pertinent, puisque la VI est dichotomique.
2. **Normalité** de la VD à l'intérieur des niveaux de la VI :
Les points du diagramme des résidus sont répartis de façon assez similaire de part et d'autre de la ligne horizontale autour de 0. Comme le montrent les statistiques des résidus, les résidus standardisés varient de -1.90 à 1.699. Il n'y a donc aucun cas extrêmement déviant, les scores résiduels étant inférieurs à 3 en valeurs absolues. Il semble y avoir une normalité de la distribution de **ScolMe** (VD) à l'intérieur des niveaux de **CadrEntr** (VI).
3. **Homogénéité des variances** de la VD à l'intérieur des niveaux de la VI :
Dans l'ensemble, les points du diagramme des résidus sont répartis de façon assez homogène verticalement. Le nuage de points ne forme pas de groupes qui indiqueraient la présence d'une hétéroscédasticité. Il y a donc homogénéité des variances de **ScolMe** (VD) à l'intérieur des niveaux de **CadrEntr** (VI)

2.2. Présentation et interprétation résultats

Dans le cadre d'un article scientifique, d'un mémoire ou d'une thèse, les résultats de l'analyse de régression et de corrélation sont souvent présentés dans un tableau unique avant de faire l'objet d'une interprétation statistique ou sociologique dans le texte. Voici un exemple d'une façon de faire.

Tableau 2. *Modèle de régression de la scolarité de la Mère selon la profession du père chez les étudiants*

	b (erreur-type)
Profession du père (cadre/entrepreneur)	4.41** (1.39)
Constante	6.37*** (0.69)
n	90
R-deux	0.10

Notes. Les entrées correspondent à des coefficients de régression non standardisés (avec les erreurs-types entre

parenthèses). Pour la profession du père, la catégorie de référence est Autre profession. *** $p < .01$; ** $p < .01$.

Interprétation statistique (analyse succincte)¹: L'analyse de régression montre que la profession du père a un effet significatif sur la scolarité de la mère dans la population étudiante étudiée ($b=4.41$; $\beta=0.32$; $p < 0,01$). Lorsqu'on passe d'un père d'une autre profession à un père cadre/intellectuel, la scolarité complétée par la mère augmente de 4,41 années. Pour un père cadre/intellectuel, on peut prédire le nombre d'années de scolarité de la mère à 10,77 ans. Dans l'ensemble, le modèle linéaire explique 10% de la variation dans la variable dépendante ($R^2=0.10$). L'analyse subséquente des résidus confirme la validité de ce modèle, puisque le diagnostic révèle le respect des postulats de normalité et de d'homogénéité des variances, le postulat de linéarité n'étant pas applicable. Aucun cas extrêmement déviant n'est détecté, les résidus standardisés variant de -1.90 à 1.70.

Interprétation théorique/sociologique (discussion): Les résultats suggèrent que l'hypothèse de l'homogamie est confirmée. Les hommes cadres/entrepreneurs épousent des femmes instruites. Cela peut s'expliquer par le fait que la société valorise l'éducation et la profession comme un marqueur déterminant du statut socioéconomique.

3. Régression linéaire simple avec une VI qualitative non dichotomique ($k \geq 3$)

3.1. Analyse des données

Il est possible d'inclure une VI qualitative non dichotomique dans un modèle de régression. Pour ce faire, il faut créer des variables factices dichotomiques codées 0/1 de façon à faciliter l'interprétation des résultats de l'analyse de régression.

Changeons à nouveau de prédicteur (VI), et intéressons-nous à la relation entre la profession du père (**ProfPe3**) et le nombre d'années de scolarité de la mère (**ScolMe**) chez les étudiants inscrits en L2 de sociologie de l'UGB. La question de recherche est la suivante : *La profession du père a-t-elle un effet sur la scolarité de la mère chez les étudiants ? Précisément dit, l'appartenance du père à la catégorie des cadres/entrepreneurs peut-elle prédire la scolarité de la mère chez les étudiants ?*

Profession du père (ProfPe3) —————> Scolarité de la mère (ScolMe)

Étant donné que la VI se présente sous forme trichotomique (Agriculteur/éleveur Vs Cadre/entrepreneur Vs Autre profession) et que la VD est de nature quantitative (ScolMe), on peut toujours utiliser la régression linéaire. Toutefois, il faut transformer la VI en créant des variables dichotomiques codées 0/1. Sortons d'abord la distribution de fréquences et de % de la variable **ProfPe3**.

¹ On suppose que l'analyse descriptive univariée a été effectuée auparavant en ces termes : Au total, 97 étudiants ont répondu à propos de la scolarité en années de la mère ($M=7.54 \pm 5.92$), alors que parmi les 96 répondants valides, 24% sont des cadres ou entrepreneurs, 76% d'entre eux relevant d'une autre profession.

ProfPe3 Profession du père					
		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	1,00 Cadre ou entrepreneur	23	22,3	24,0	24,0
	2,00 Autre profession	47	45,6	49,0	72,9
	3,00 Agriculteur/pêcheur	26	25,2	27,1	100,0
	Total	96	93,2	100,0	
Manquant	Système	7	6,8		
Total		103	100,0		

Cette variable comporte trois catégories : 1 = Cadre/entrepreneur; 2 = Autre profession; 3=Agriculteur/éleveur. Telle qu'elle se présente, la variable ne peut pas être directement incluse dans le modèle, car cela supposerait l'existence d'un ordre : cadre/entrepreneur supérieur à autre, mais inférieur à Agriculteur/éleveur.

Toutefois, pour prendre en compte toute l'information contenue dans la profession du père (ProfPe3) et considérer l'effet discriminé des catégories sur ScolMe, une procédure souvent utilisée consiste à créer des **variables factices dichotomiques** codées 0/1. Toutefois, retenez que seules k-1 variables factices seront introduites dans l'analyse², alors que l'autre catégorie sert de référence et doit être exclue de l'analyse. Voici comment procéder à la transformation.

- **Étape 1 : Savoir combien de catégories (k) comporte la VI originale et comment ses catégories sont codées.**

La variable ProfPe3 comporte trois catégories non ordonnées (k=3)

1. Cadre/entrepreneur
2. Autre profession
3. Agriculteur/éleveur

- **Étape 2 : Déterminer le nombre de variables factices dichotomiques à inclure dans le modèle de régression (k-1)**

Il s'agit de créer trois (k-1=3-1=3) variables factices à inclure dans l'analyse, étant entendu que la 3^e catégorie sert de référence et réfère donc à la catégorie de référence. Prenons comme catégorie de référence AUTRE PROFESSION. Ce qui signifie que les trois autres catégories constitueront les variables dichotomiques³.

- **Étape 3. Créer les variables factices dichotomiques codées 0/1 en transformant la variable originale**

La catégorie de référence (0) est : AUTRE PROFESSION

² Évidemment, on peut créer k variables factices en autant que le nombre de variables factices, à inclure dans le modèle, soit égal à k-1 ou moins.

³ Bien sûr, on peut créer une variable factice représentant la catégorie de référence. Mais, vous devez savoir que cette variable ne sera pas introduite dans le modèle de régression. On peut aussi créer toutes les variables dichotomiques possibles et choisir une catégorie de référence à tour de rôle.

Les deux variables factices codées 0/1 sont :

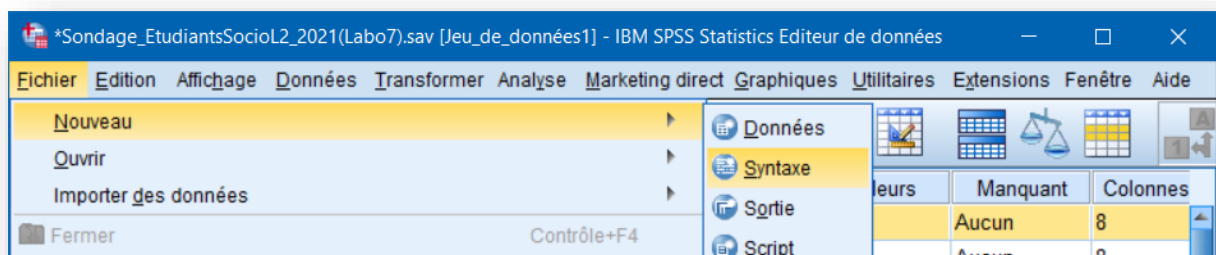
Variables factices	Codes	
Cadre/entrepreneur	1.Cadre/entrepreneur	0.Autrement
Agriculteur/éleveur	1.Agriculteur/éleveur	0.Autrement

Procédons à la transformation en passant par les menus contextuels de SPSS ou en écrivant ou appliquant directement la syntaxe ci-dessous :

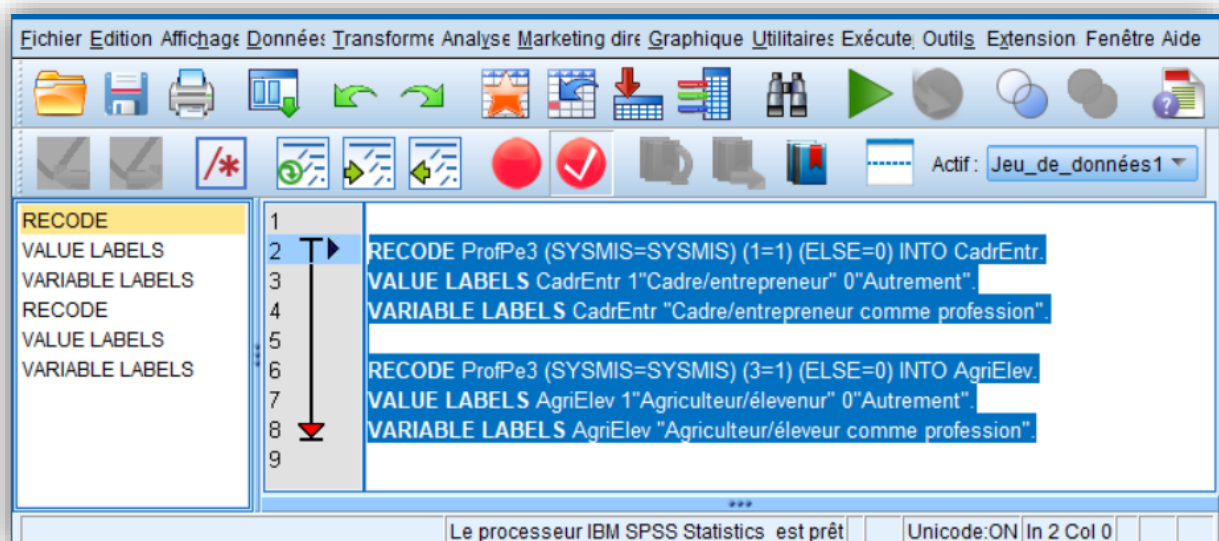
```
RECODE ProfPe3 (SYSMIS=SYSMIS) (1=1) (ELSE=0) INTO CadrEntr.
VALUE LABELS CadrEntr 1"Cadre/entrepreneur" 0"Autrement".
VARIABLE LABELS CadrEntr "Cadre/entrepreneur comme profession".
```

```
RECODE ProfPe3 (SYSMIS=SYSMIS) (3=1) (ELSE=0) INTO AgriElev.
VALUE LABELS AgriElev 1"Agriculteur/éleveur" 0"Autrement".
VARIABLE LABELS AgriElev "Agriculteur/éleveur comme profession".
```

Sélectionner et copier la syntaxe (encadré bleu) en cliquant-droit!
Sur SPSS, cliquez sur le menu FICHIER, puis NOUVEAU, puis SYNTAXE!



Dans la fenêtre qui apparaît, collez la syntaxe ainsi copiée, en cliquant-droit.



Sélectionnez toute la syntaxe de façon à obtenir une surbrillance bleue, et cliquez sur le bouton triangle vert pour exécuter l'instruction et créer les deux variables :



Ci-dessous leur distribution de fréquences et de pourcentages.

CadrEntr Cadre/entrepreneur comme profession					
		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	,00 Autrement	73	70,9	76,0	76,0
	1,00 Cadre/entrepreneur	23	22,3	24,0	100,0
	Total	96	93,2	100,0	
Manquant	Système	7	6,8		
Total		103	100,0		

AgriElev Agriculteur/éleveur comme profession					
		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	,00 Autrement	70	68,0	72,9	72,9
	1,00 Agriculteur/éleveur	26	25,2	27,1	100,0
	Total	96	93,2	100,0	
Manquant	Système	7	6,8		
Total		103	100,0		

Si la VI dichotomique est codée 1 et 2, il faut la recoder 0/1 en utilisant la procédure **CRÉATION DE VARIABLES**. L'idée est de créer une nouvelle variable recodée en signifiant, par exemple, à SPSS que 1 devient 0 et que 2 devient 1. Ou inversement, selon qu'une catégorie est considérée comme la référence (0) ou la catégorie d'intérêt (1). Cela déterminera le signe du coefficient de régression et donc la façon dont il sera interprété. Lorsque la VI qualitative comporte plusieurs catégories, on peut la dichotomiser afin d'utiliser le modèle linéaire, mais on gagnerait à conserver le plus de catégories pour mieux les discriminer et gagner en finesse à propos de l'analyse.

La clé d'interprétation repose sur une double question :

- 1) Lorsqu'on passe d'une autre profession à la profession Cadre/entrepreneur du père (0 à 1), comment régressera-t-il le nombre d'années de scolarité de la mère ?
- 2) Lorsqu'on passe d'une autre profession à la profession Agriculteur/éleveur du père (0 à 1), comment régressera-t-il le nombre d'années de scolarité de la mère ?

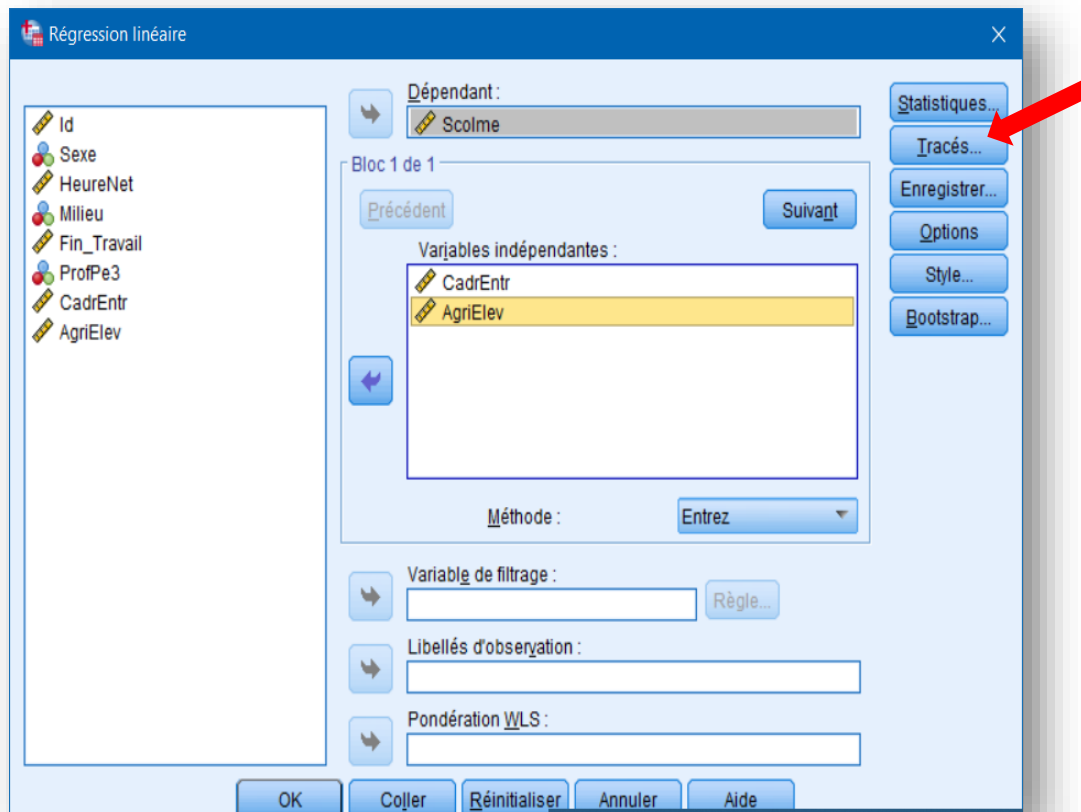
On est maintenant prêt à procéder à l'analyse de régression linéaire simple en incluant les deux variables dichotomiques **CadrEntr** et **AgriElev**. Suivons à nouveau la même procédure.

Analyse

Régression

Linéaire

Une fenêtre de dialogue apparaît ! Cliquez sur la variable dépendante (scolme), puis indépendante (CadrEntr).



Pour procéder au diagnostic du modèle de régression, cliquez sur Tracés, et mettez les valeurs prédites standardisées (ZPRED) sur l'axe X et les valeurs résiduelles standardisées (ZRESID) sur l'axe Y. Poursuivez! Il est possible de sortir les valeurs prédites de la scolarité de la mère (Y), selon que le père est d'une AUTRE PROFESSION (X=0) ou de profession CADRE/ENTREPRENEUR (X=1). Pour ce faire, cliquez sur ENREGISTRER pour voir apparaître la fenêtre ci-dessous ! Puis, cochez la cage **Non standardisés** sous **Prévisions**!

Poursuivez et validez le tout pour voir apparaître les résultats.

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,341 ^a	,116	,096	5,659

a. Prédicteurs : (Constante), AgriElev Agriculteur/éleveur comme profession, CadrEntr Cadre/entrepreneur comme profession

b. Variable dépendante : Scolme Combien d'années de scolarité, approximativement, votre mère a-t-elle complétées ?

Le tableau ci-dessus montre que le coefficient de corrélation R est de 0,34, traduisant l'existence d'une relation faible à modérée entre la profession du père et le nombre d'années de scolarité de la mère.

Le R-deux de 0,12 signifie que la profession du père explique une proportion de 12% de la variation dans la scolarité de la mère. De même, nous réduisons de 12% nos erreurs de prédiction de la scolarité de la mère (VD) quand nous connaissons la profession du père (VI). Finalement, 88% de la variation reste à expliquer.

Modèle		Somme des carrés	ddl	Carré moyen	F	Sig.
1	Régression	365,942	2	182,971	5,713	,005 ^b
	Résidu	2786,280	87	32,026		
	Total	3152,222	89			

a. Variable dépendante : Scolme Combien d'années de scolarité, approximativement, votre mère a-t-elle complétées ?

b. Prédicteurs : (Constante), AgriElev Agriculteur/éleveur comme profession, CadrEntr Cadre/entrepreneur comme profession

Nous rejetons l'hypothèse nulle voulant qu'il n'y ait pas de relation entre la profession du père et la scolarité de la mère, puisque la probabilité de commettre une erreur en rejetant l'hypothèse nulle est inférieure à 0,05 : elle est précisément de 0,005. Il existe une corrélation statistiquement significative entre la profession du père et la scolarité de la mère dans la population étudiante étudiée.

Modèle		Coefficients non standardisés		Coefficients standardisés		
		B	Erreur standard	Bêta	t	Sig.
1	(Constante)	6,977	,863		8,084	,000
	CadrEntr Cadre/entrepreneur comme profession	3,796	1,483	,276	2,559	,012
	AgriElev Agriculteur/éleveur comme profession	-1,657	1,423	-,125	-1,164	,248

a. Variable dépendante : Scolme Combien d'années de scolarité, approximativement, votre mère a-t-elle complétées ?

La constante (6,977) signifie que lorsque la profession du père est autre ($X=0$), la scolarité prédite de la mère est de 6,977 années. La constante s'avère statistiquement significative à au moins 99,9% ($t=8,08$; $p<0,001$).

CadrEntr : Le coefficient de régression (3,796) signifie que lorsqu'on passe de la profession autre du père ($X=0$) à la profession cadre/entrepreneur ($X=1$), la scolarité

prédite de la mère augmente de 3,796 années. Le coefficient de régression s'avère statistiquement significatif à au moins 95% ($t=3,30$; $p<0,05$).

AgriElev : Le coefficient de régression (-1,657) signifie que lorsqu'on passe de la profession autre du père ($X=0$) à la profession agriculteur/éleveur ($X=1$), la scolarité prédite de la mère diminue de 1,657 années. Toutefois, le coefficient de régression ne s'avère pas statistiquement significatif ($p=0,248$).

Voici les deux équations de régression linéaire de la relation entre Profession du père et Scolarité de la mère.

1) CadrEntr ($X=1$)

$$\text{Scolme} = 6,977 + 3,796 (\text{CadrEntr})$$

$$\text{Scolme} = 6,977 + 3,796 (1) = 10,773$$

Pour un père cadre/entrepreneur, la scolarité prédite de la mère est de 10,773 années.

2) AgriElev ($X=1$)

$$\text{Scolme} = 6,977 - 1,657 (\text{AgriElev})$$

$$\text{Scolme} = 6,977 - 1,657 (1) = 5,32$$

Pour un père agriculteur/éleveur, la scolarité prédite de la mère est de 5,32 années.

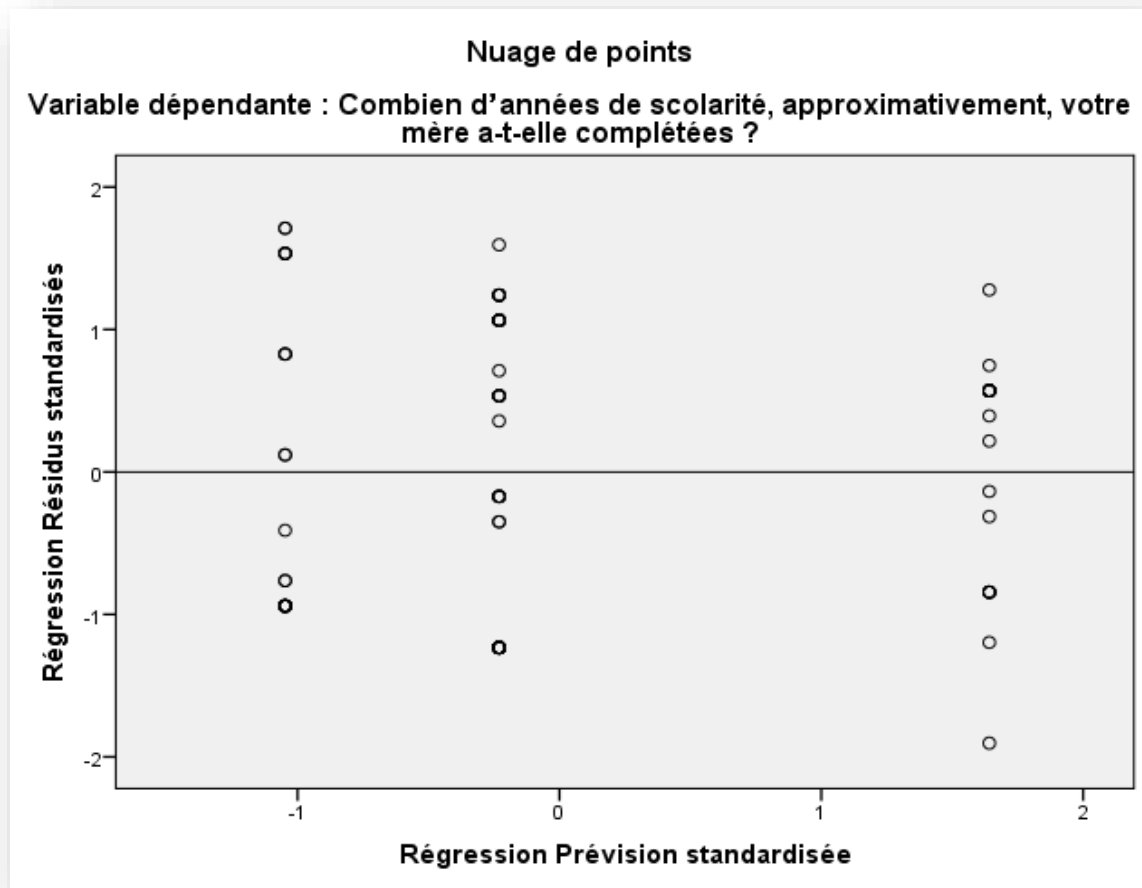
La scolarité prédite des mères est de 10,773 ans chez les pères cadres/entrepreneurs ($X=1$), de 5,32 années chez les pères agriculteurs/éleveurs ($X=1$) et de 6,977 chez les pères d'une autre profession ($X=0$). On retrouve les trois valeurs prédites que nous venons d'enregistrer sous forme de variable (PRE_1) dans la base de données :

	Scolme	CadrEntr	AgriElev	PRE_1	var	var
1	10	Cadre/entrepr...	Autrement	10,773		
2	0	Autrement	Autrement	6,977		
3	15	Autrement	Agriculteur/éle...	5,320		

Qu'en est-il du diagnostic des résidus du modèle de régression ?

Statistiques des résidus ^a					
	Minimum	Maximum	Moyenne	Ecart type	N
Prévision	5,32	10,77	7,44	2,028	90
Résidu	-10,773	9,680	,000	5,595	90
Prévision standardisée	-1,048	1,641	,000	1,000	90
Résidu standardisé	-1,904	1,710	,000	,989	90

a. Variable dépendante : Scolme Combien d'années de scolarité, approximativement, votre mère a-t-elle complétées ?



À vous d'analyser!

3.2. Présentation et interprétation résultats

Tableau 2. *Modèle de régression de la scolarité de la Mère selon la profession du père chez les étudiants*

	b (erreur-type)
Profession du père	
Cadre/entrepreneur	3,80* (1.48)
Agriculteur/éleveur	-1,66 (1.42)
Constante	6,98*** (0.86)
n	90
R-deux	0.12

Notes. Les entrées correspondent à des coefficients de régression non standardisés (avec les erreurs-types entre parenthèses). Pour la profession du père, la catégorie de référence est Autre profession. * $p < .05$; *** $p < .001$.

Interprétation statistique (analyse succincte): L'analyse de régression montre que la profession cadre/entrepreneur chez le père a un effet significatif sur la scolarité de la mère dans la population étudiante étudiée ($b=5.45$; $\beta=0.40$; $p < 0,001$). Lorsqu'on

passé d'un père d'une autre profession à un père cadre/intellectuel, la scolarité complétée par la mère augmente de 5,45 années. Pour un père cadre/intellectuel, on peut prédire le nombre d'années de scolarité de la mère à 10,77 ans. Toutefois, la profession agriculteur/éleveur n'est pas associée à la scolarité de la mère. Dans l'ensemble, le modèle linéaire explique 12% de la variation dans la variable dépendante ($R^2=0.12$). L'analyse subséquente des résidus confirme la validité de ce modèle, puisque le diagnostic révèle le respect des postulats de normalité et de l'homogénéité des variances, le postulat de linéarité n'étant pas applicable. Aucun cas extrêmement déviant n'est détecté, les résidus standardisés variant de -1.90 à 1.71.

Interprétation théorique/sociologique (discussion): Même interprétation!

4. Exercice pratique

Vous voulez savoir si le milieu d'origine des étudiants (**milieu**) a un effet sur la scolarité de la mère (**Scolme**).

- a) Vous souhaitez utiliser la régression. Pour ce faire, vous décidez de transformer la variable Milieu en créant des variables factices dichotomiques, notamment avec **Dakar urbain** comme catégorie de référence.
- b) Sortez les résultats de l'analyse de régression linéaire simple incluant les deux variables dichotomiques;
- c) Interprétez statistiquement le coefficient de corrélation R, le R-deux, le test de la corrélation.
- d) Interprétez chacun des paramètres des deux équations de régression.
- e) Rédiger l'équation de régression de la relation entre Milieu rural et Scolme. Quelle est la scolarité prédite de la mère lorsque l'étudiant réside en milieu rural plutôt que dans la zone urbaine de Dakar?
- f) Rédiger l'équation de régression de la relation entre Autre milieu urbain et Scolme. Quelle est la scolarité prédite de la mère lorsque l'étudiant réside dans un autre milieu urbain plutôt que dans la zone urbaine de Dakar?