**REPUBLIQUE DU BENIN**

🙡🙡🙡

**UNIVERSITE DE PARAKOU (UP)**

🙡🙡🙡🙡🙡

**ECOLE NATIONIONALE DE LA STATISTIQUE DE LA PLANIFICATION ET DE LA DEMOGRAPHIE (ENSPD)**

**Master 1**

**COURS :** Technique et méthode d’échantillonnage et de rééchantillonnage



**Membres du groupe :**

1. MAMA Abdou Moukadas
2. KINSA Ulrich F
3. SANTA Mathias
4. DONKO S Mounirath
5. AMADJI Setonde
6. SABI KOTE B Kpenté

**Chargé du cours :**

Dr Ir. Epiphane SODJINOU

Agroéconomiste, Biostatisticien

(Maitre de conférences)

ANNEE ACADEMIQUE : 2024 -2025

**THEME : Identifier dans le logiciel SPSS les procédures et commandes disponible pour la vérification des conditions d’applications des régressions linéaires**

Table des matières

[**Introduction** 3](#_Toc200025022)

[**1.** **Condition de validation de la régression linéaire** 3](#_Toc200025023)

[**1.1.** **Caractère aléatoire et simple** 3](#_Toc200025024)

[**1.2.** **Indépendance des résidus** 3](#_Toc200025025)

[**1.3.** **Normalité des résidus** 6](#_Toc200025026)

[**1.4.** **Homogénéité des variances résiduelles (homoscédasticité)** 9](#_Toc200025027)

[**Conclusion** 10](#_Toc200025028)

# **Introduction**

La régression linéaire est une méthode statistique essentielle qui permet d’analyser la relation entre une variable dépendante et une ou plusieurs variables indépendantes. Elle est largement utilisée dans les domaines de la science, de l’économie, de la psychologie et bien d’autres disciplines pour modéliser et prédire des phénomènes. L’objectif principal d’une régression linéaire est de trouver une équation qui représente au mieux la relation entre les variables, permettant ainsi de faire des prédictions ou d’examiner l’impact des facteurs explicatifs sur une variable cible. Pour que les résultats du modèle soient valides, certaines conditions doivent être respectées, comme la normalité des résidus, l’indépendance des observations et l’homoscédasticité des erreurs. Grâce à des logiciels comme SPSS, il est possible de réaliser une régression linéaire facilement, en appliquant des tests statistiques et en analysant graphiquement la pertinence du modèle.

1. **Condition de validation de la régression linéaire**

La régression linéaire vise à expliquer une variable dépendante par une ou un ensemble de variables indépendantes quantitatives. Lorsque le problème implique une seule variable indépendante, la technique statistique est appelée régression simple. Lorsque le problème implique plusieurs variables indépendantes, il s’agit d’une régression multiple. La régression est utilisée pour l’explication et la prédiction. Elles se reposent sur plusieurs conditions essentielles pour garantir la validité des estimations et des tests statistiques. Ainsi, il est dénombré quatre conditions à savoir la vérification du caractère aléatoire, l’indépendance des résidus, la normalité des résidus et l’homoscédasticité. Voici les principales conditions :

## **Caractère aléatoire et simple**

Les observations doivent être sélectionnées de manière aléatoire pour éviter les biais et assurer la représentativité de l’échantillon. Chaque observation doit être indépendante des autres. Un échantillon **aléatoire et représentatif** garantit que les résultats peuvent être généralisés à une population plus large. Si les données sont **corrélées ou biaisées**, les estimations du modèle peuvent être faussées.

Dans une régression linéaire, cette condition est souvent associée à l'**indépendance des observations**, assurant que chaque donnée est mesurée sans influence des autres.

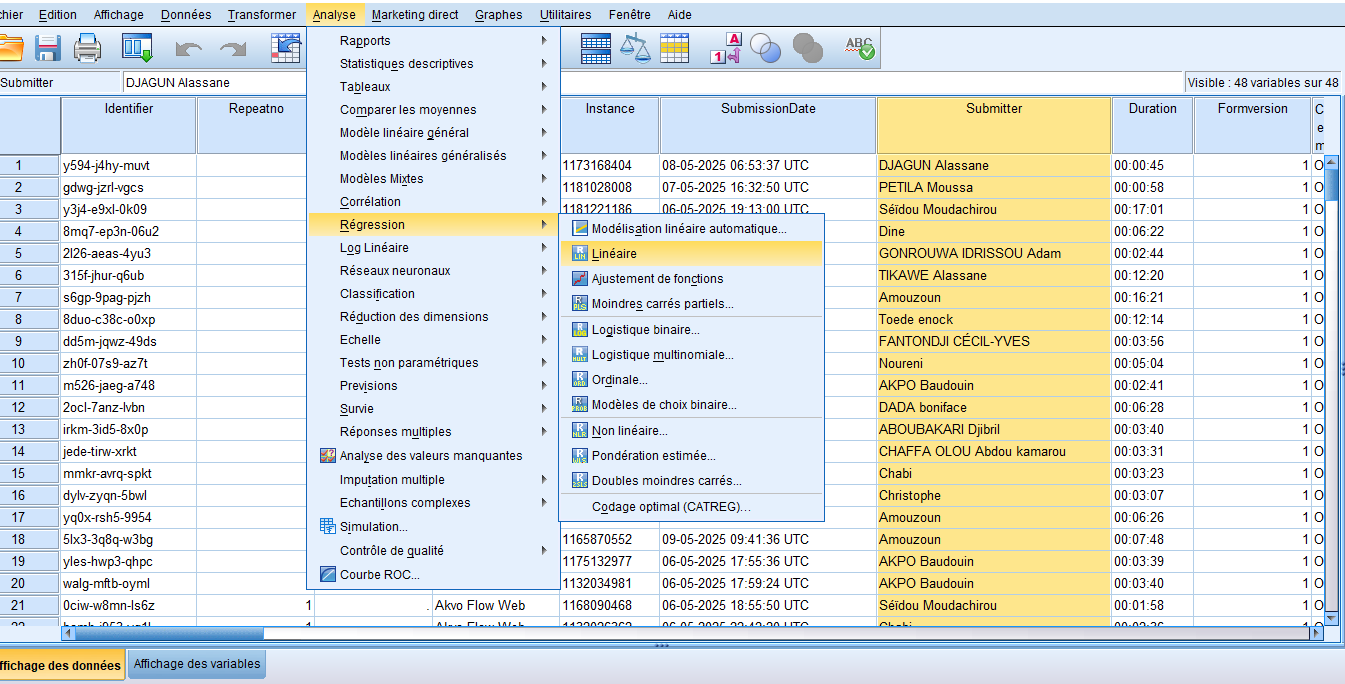
* 1. **Indépendance des résidus**

L’indépendance des résidus est une condition essentielle pour assurer la validité d’une régression linéaire. Elle signifie que les erreurs de prédiction (résidus) ne doivent pas être corrélées entre elles. En cas de corrélation entre les erreurs, il peut y avoir une violation de l'hypothèse d'indépendance, souvent détectée via le test de Durbin-Watson. Ce qui peut entraîner des estimations biaisées et des conclusions incorrectes. L’autocorrélation des erreurs est particulièrement problématique dans les données temporelles et les séries chronologiques.

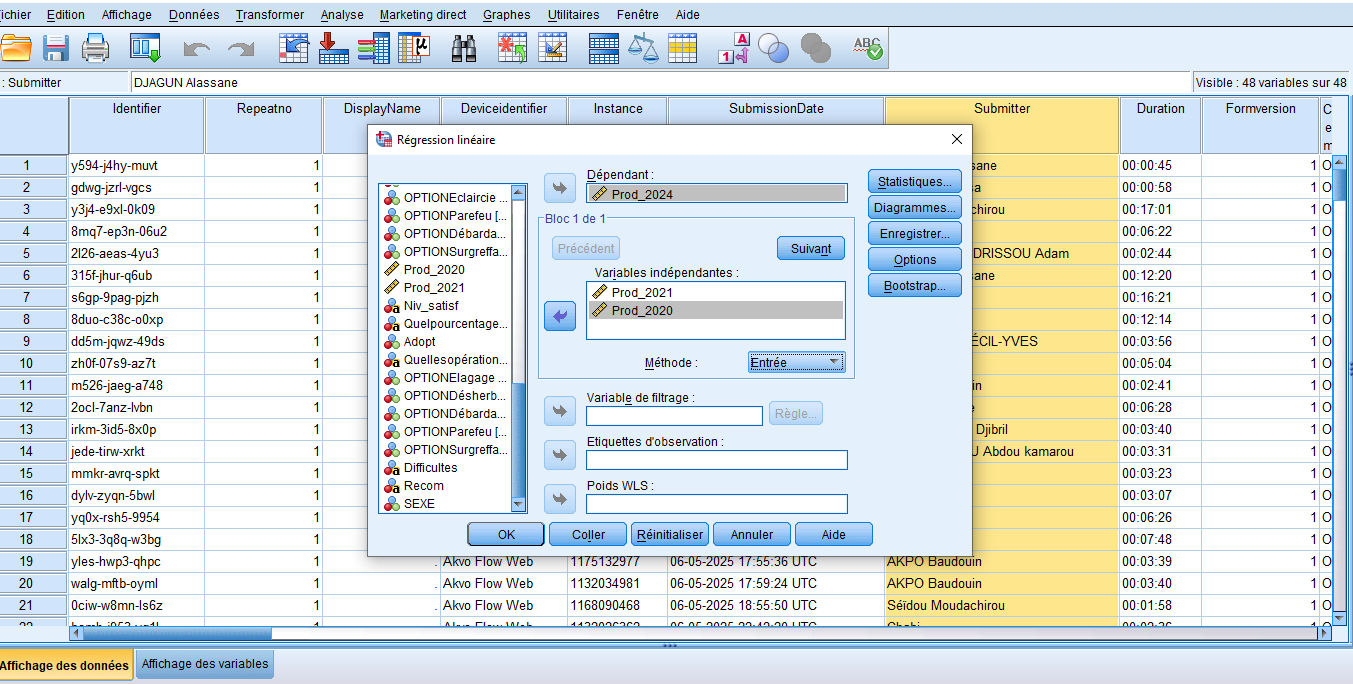
Dans SPSS, vous pouvez tester l’indépendance des observations dans une régression linéaire en utilisant le test de Durbin-Watson. Ce test permet de détecter une éventuelle autocorrélation des résidus, ce qui violerait l’hypothèse d’indépendance.

Étapes pour effectuer le test de Durbin-Watson dans SPSS

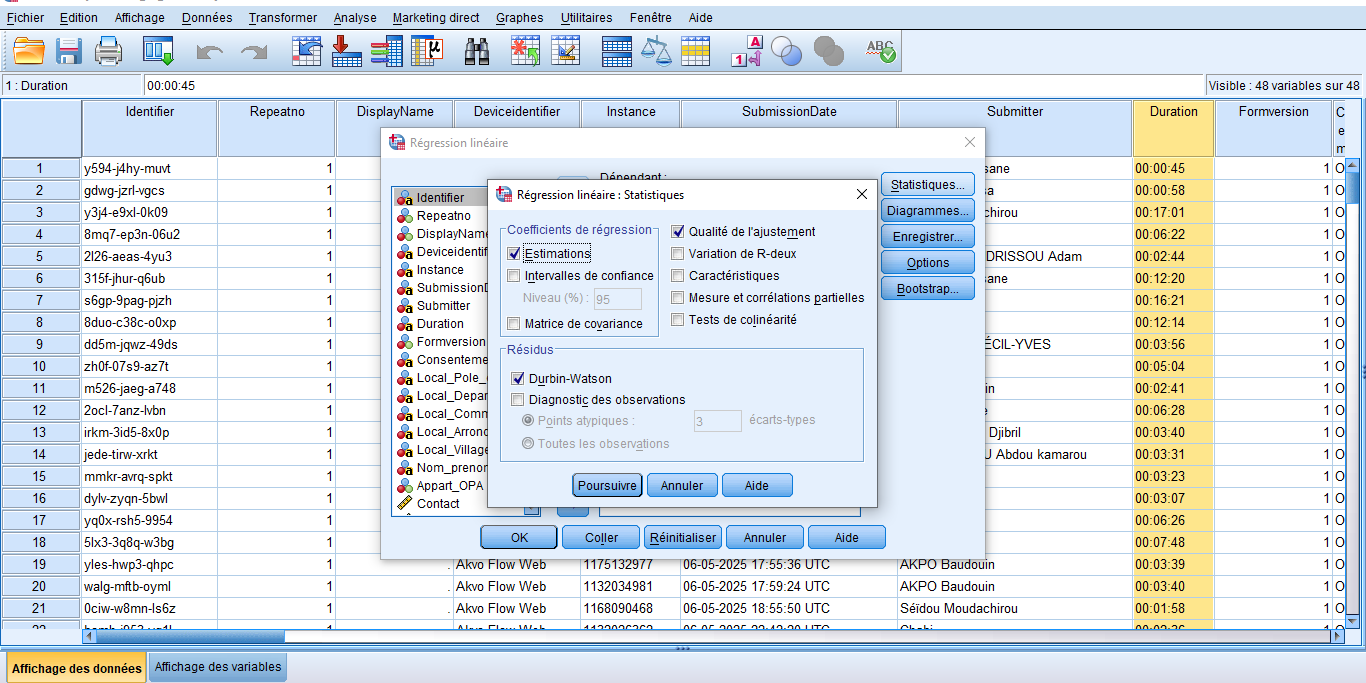
1. Ouvrir SPSS et charger votre jeu de données.
2. Aller dans Analyse → Régression → Linéaire.



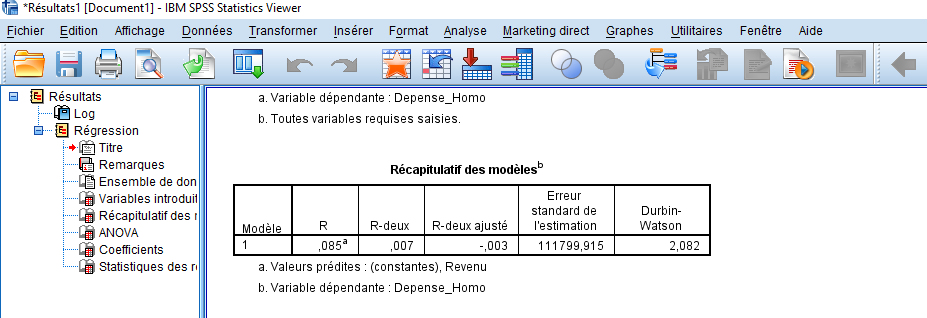
1. Sélectionner votre variable dépendante et vos variables indépendantes.



1. Cliquer sur Statistiques et cocher Durbin-Watson.



1. Exécuter l’analyse et vérifier la valeur du test dans les résultats.



Interprétation du test de Durbin-Watson

* Valeur proche de 2 → Pas d’autocorrélation des résidus (condition respectée).
* Valeur < 1,5 → Présence d’une autocorrélation positive (les erreurs sont liées entre elles).
* Valeur > 2,5 → Autocorrélation négative (les erreurs alternent de manière systématique).

Si l’indépendance est violée, des solutions existent :

* Ajouter des variables explicatives pour mieux modéliser la structure des données.
* Utiliser des modèles spécifiques comme l’utilisation de modèles de régression adaptés aux données temporelles les modèles ARIMA ou modèles de régression robuste.
  1. **Normalité des résidus**

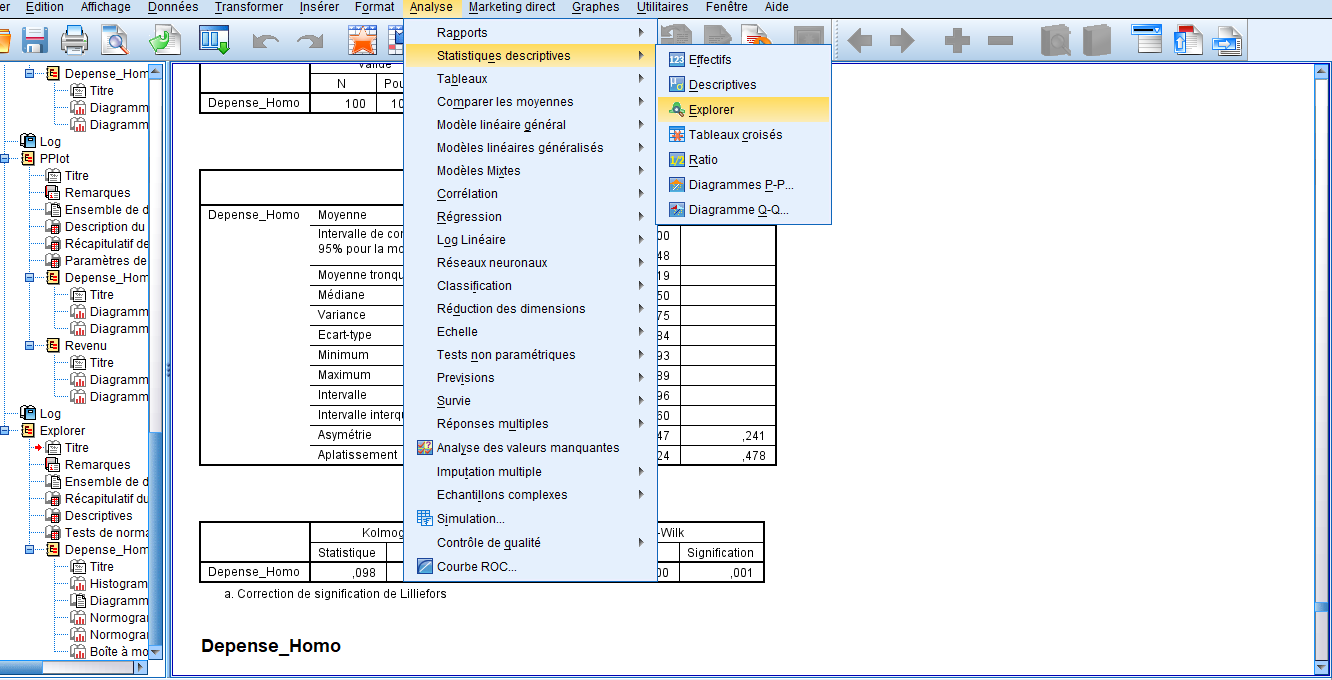
La normalité des résidus est une condition clé en régression linéaire, bien qu'elle soit souvent moins stricte que d'autres hypothèses. Elle permet d'assurer la validité des tests statistiques et des intervalles de confiance des coefficients. Cette condition stipule que les erreurs de régression doivent suivre une distribution normale et peut etre vérifiée par des tests comme Shapiro-Wilk ou par des graphiques Q-Q plot. En resumé :

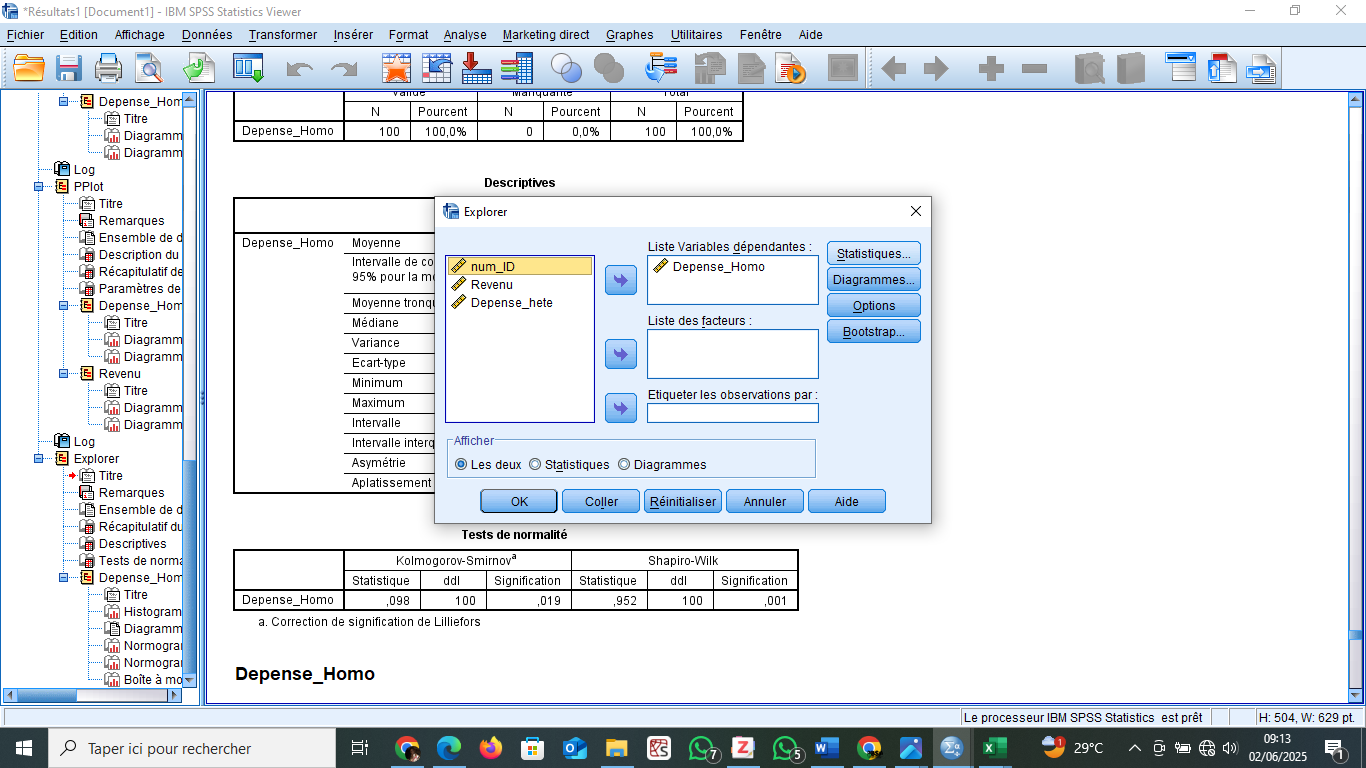
* Si les résidus suivent une distribution normale, les tests de significativité (comme le test de Student) sont plus fiables.
* La normalité est particulièrement critique lorsque l'échantillon est petit, car les estimations deviennent instables en cas de non-normalité.
* Si les résidus ne sont pas normaux, les p-valeurs des tests peuvent être biaisées, ce qui peut conduire à des conclusions incorrectes.

Dans SPSS, voici quelques méthodes courantes pour tester la normalité des résidus :

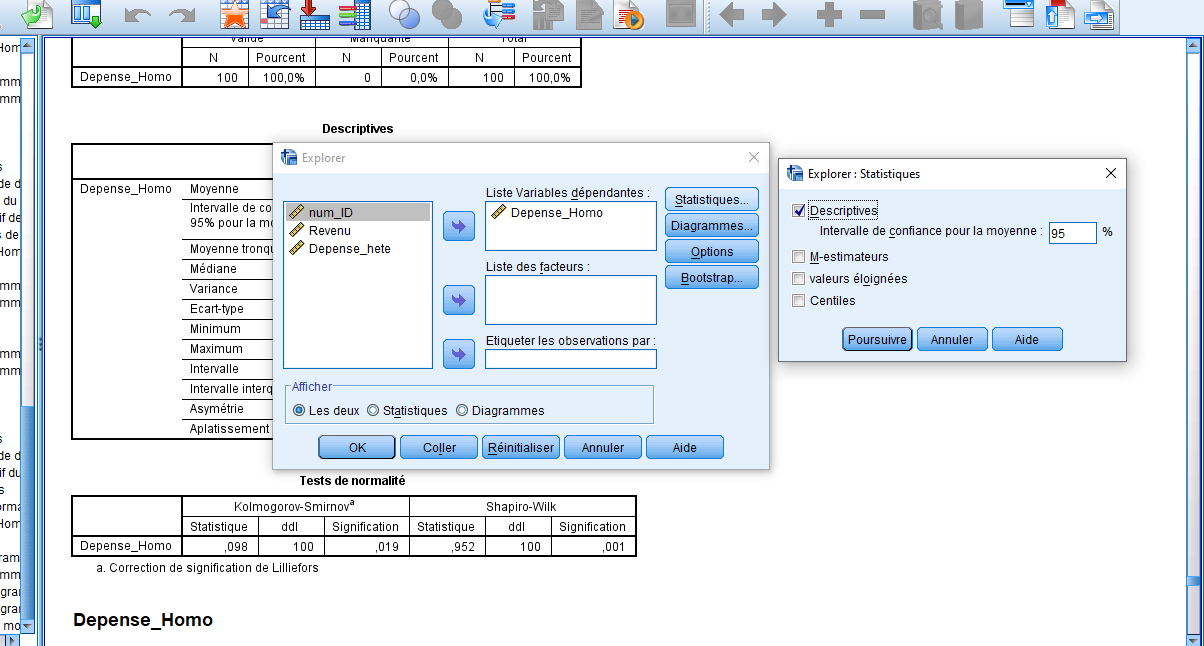
1. Histogramme des résidus → Permet une première visualisation de la distribution.
2. Graphique Q-Q (Quantile-Quantile) → Compare les résidus à une distribution normale.
3. Test de Shapiro-Wilk ou Kolmogorov-Smirnov → Vérifie statistiquement la normalité.

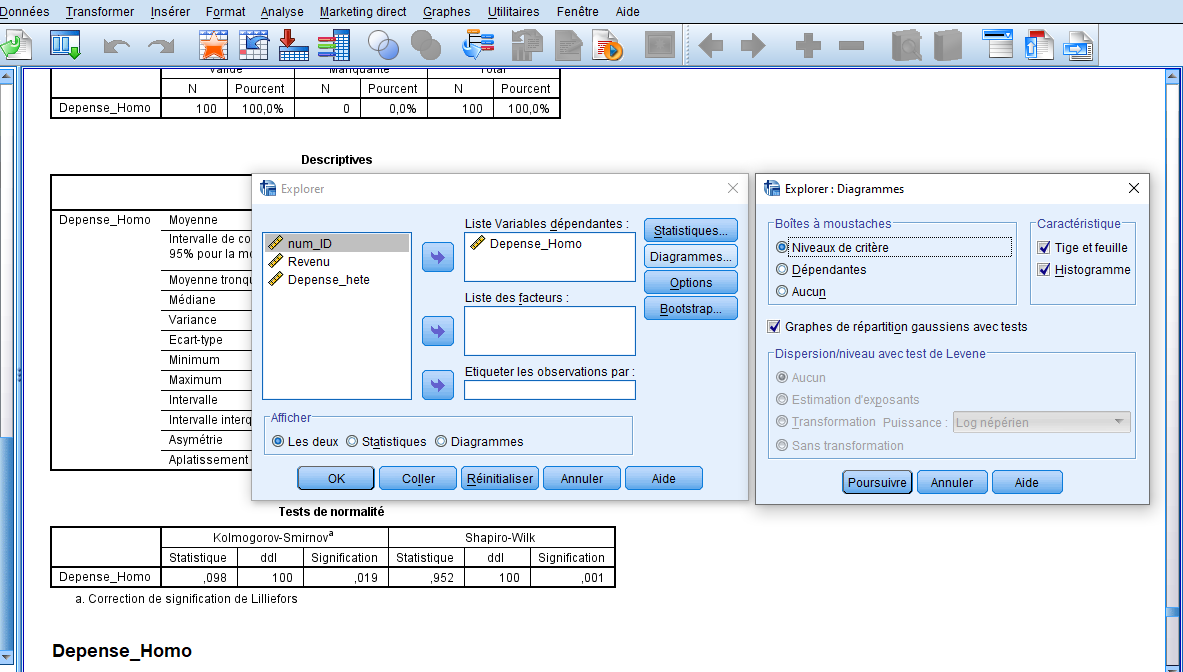
ETAPE

1. Ouvrir SPSS et chargé la base
2. Pour accéder au test de normalité, il faut :
   * Cliquer sur Analyse > Statistique descriptive> Explorer (Par la suite configurer la boite de dialogue qui s’affiche)
3. 
   * + Déplacer la variable dépendante dans la zone variable dépendante
     + S’il existe une variable de regroupement alors déplacer celle-ci dans Facteur regroupement



* + Cliquer sur statistiques et assurer que Descriptives est croché



* + Cliquer sur Graphique
    - * Cocher Histogramme
      * Cocher normalité avec tests
      * Cliquer sur ok
* 

Interprétation des résultats

Avec le test de Shapiro Wilk, pour p> 0.05, la variable suit une distribution normale (on ne rejette pas l’hypothèse nulle)

P<0.05 alors la variable ne suit pas une distribution de la loi normale (on rejette l’hypothèse nulle)

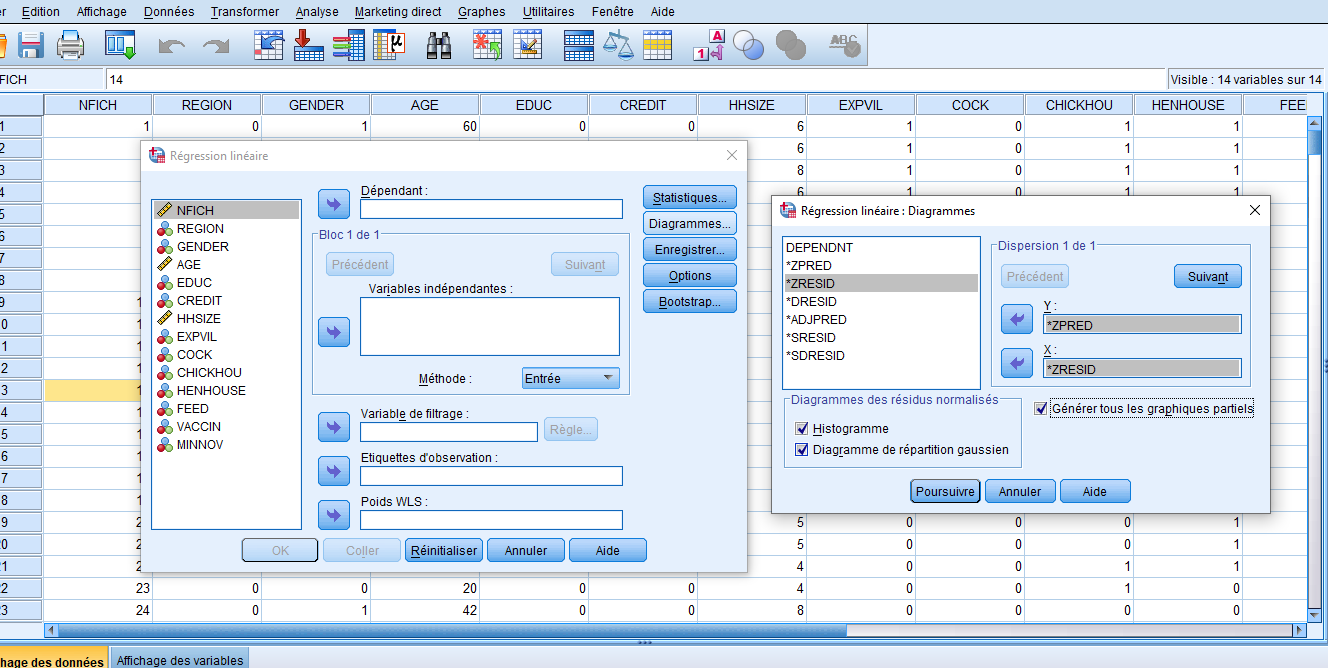
Dans le cas où, cette condition est viollé c’est-à-dire que les résidus ne suivent pas une loi normale, voici quelques solutions :

* Transformation des données (logarithme, racine carrée) pour réduire l’asymétrie.
* Utilisation d’une régression robuste ou d’une approche non paramétrique.
* Vérification que la violation de la normalité n’affecte pas les estimations des coefficients, car la régression linéaire peut être robuste à certaines violations.
  1. **Homogénéité des variances résiduelles (homoscédasticité)**

L'homoscédasticité est une condition essentielle en régression linéaire, qui signifie que la variance des résidus doit être constante pour toutes les valeurs des variables indépendantes. Si cette condition est violée, on parle d'hétéroscédasticité, ce qui peut fausser les estimations des coefficients et les tests de significativité.

Avec le logiciel SPSS, on peut tester l'homoscédasticité par la méthode des graphiques ou par le test de Pagan. Voici une description de la démarche :

1. Graphique des résidus
   * Aller dans Analyse → Régression → Linéaire.
   * Sélectionner la variable dépendante et les variables indépendantes.
   * Dans Statistiques, cocher Résidus standardisés et Graphiques.
   * Choisir Résidus standardisés vs valeurs prédites.
   * Si les points sont uniformément dispersés, l’homoscédasticité est respectée. Si une forme en cône apparaît, il y a hétéroscédasticité.



En cas de hétéroscédasticité, des méthodes correctives comme la transformation des données ou l’utilisation de modèles robustes peuvent être nécessaires. Il s’agit de :

* + Transformation des données (logarithme, racine carrée).
  + Utilisation d’une régression robuste ou d’un modèle de régression pondéré.

# **Conclusion**

La régression linéaire repose sur plusieurs hypothèses dont la vérification est essentielle pour assurer la validité et fiabilité des estimateurs. Ces conditions concernent tant les résidus du model que la structure des donnes leurs respect permet une interprétation correcte des coefficients, une estimation efficace des paramètre Le logiciel SPSS est l’un des meilleures applications pour la vérification et tant sur l’utilisation.