**REPUBLIQUE DU BENIN**

**\*\*\*\*\*\*\*\***

**MINISTERE DE L’ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE**

**\*\*\*\*\*\*\*\*\***

**UNIVERSITE DE PARAKOU**

**\*\*\*\*\*\*\*\*\*\*\*\***

**ECOLE NATIONAL DE PLANIFICATION, DE STATISTIQUE ET DE DEMOGRAPHIE**

**\*\*\*\*\*\*\*\*\*\*\*\*\*\***

**Master 1 PSE / SA**

**Cours : Techniques et méthodes d’échantillonnage et de rééchantillonnage**

**Travaux de Groupe N°8**

**THEME**

**Identifier dans les logiciels Stata les procédures et commandes disponibles pour la vérification des conditions d’application des méthodes statistiques paramétriques.**

**Chargé du cours :**

**Dr. SODJINOU Epiphane**

**Groupe 8**

**Membres du groupe :**

1. ATCHA Issere
2. BOUKARI Yasmina
3. KOTCHEDJO Akpo Lucien
4. OUENDO Austella Carline Abissola
5. PADONOU Frejus

**Mai 2025**

# Plan

# 1. Introduction

# 2. Conditions d’application typiques à vérifier

# 3. Commandes Stata pour chaque vérification

# 4. Exemple d’analyse complète avec régression linéaire

# 5. Cas pratique

# 6. Conclusion

# 1. Introduction

Les méthodes statistiques paramétriques (comme la régression linéaire, l’ANOVA, le test t de Student, etc.) reposent sur un ensemble de conditions à vérifier avant application. Ces conditions assurent la validité des résultats. Stata offre une large gamme de commandes pour tester ces hypothèses.

# 2. Conditions d’application typiques à vérifier

Voici les principales hypothèses à vérifier avant d’appliquer une méthode paramétrique :

|  |  |
| --- | --- |
| **Hypothèse à vérifier** | **Méthodes concernées** |
| Normalité des résidus / variables | t-test, ANOVA, régression |
| Homogénéité des variances | ANOVA, test t à variances égales |
| Indépendance des observations | Toutes les méthodes paramétriques |
| Linéarité | Régression |
| Absence de multicolinéarité | Régression multiple |
| Homoscédasticité des résidus | Régression |
| Absence d’autocorrélation | Régression avec données temporelles |

**3. Commandes Stata pour chaque vérification**

* **Normalité (test de normalité)**

summarize varname, detail

sktest varname

swilk varname // Test de Shapiro-Wilk (n < 2000)

histogram varname, normal

qnorm varname

pnorm varname

* **Homogénéité des variances**

oneway y groupe, tabulate

robvar y, by(groupe) // Test de Bartlett

oneway y groupe, bf // Test de Brown-Forsythe

* **Indépendance des observations**

xtset id temps

tsset temps

* **Linéarité (relation linéaire entre X et Y)**

twoway (scatter y x) (lfit y x)

avplot

* **Multicolinéarité (corrélation entre variables explicatives)**

reg y x1 x2 x3

vif // variance inflation factor

* **Homoscédasticité (variance constante des résidus)**

rvfplot

estat hettest // test de Breusch-Pagan

estat imtest, white

* **Autocorrélation des résidus**

estat dwatson // Test de Durbin-Watson

estat bgodfrey // Test de Breusch-Godfrey

**4. Exemple d’analyse complète avec régression linéaire**

reg y x1 x2

estat hettest // homoscédasticité

vif // colinéarité

predict resid, resid

swilk resid // normalité des résidus

rvfplot // graphique des résidus

estat dwatson // autocorrélation

# 5. Cas pratique

# 5.1. Objectif

Une chercheuse souhaite comparer les niveaux de stress (échelle de 0 à 100) entre deux groupes de salariés :  
- Groupe 1 : salariés en télétravail  
- Groupe 2 : salariés en présentiel  
  
Avant d'appliquer un test t de Student, elle doit vérifier les conditions d'application du test paramétrique

# Étapes de l'analyse dans Stata

## 5.2.1. Création des données simulées

Commandes Stata à exécuter :  
clear  
set obs 100  
  
// Création du groupe : 50 télétravail (1), 50 présentiel (2)  
gen mode = cond(\_n <= 50, 1, 2)  
  
// Génération du stress  
set seed 123  
gen stress = .  
replace stress = rnormal(55, 10) if mode == 1 // Télétravail  
replace stress = rnormal(60, 12) if mode == 2 // Présentiel  
  
label define mode\_lbl 1 "Télétravail" 2 "Présentiel"  
label values mode mode\_lbl  
  
save stress\_data.dta, replace

## 5.2.2. Vérification des conditions paramétriques

### A. Normalité (test de Shapiro-Wilk)

Commandes :

swilk stress if mode == 1  
swilk stress if mode == 2

Résultat attendu : p = 0.53 (télétravail), p = 0.41 (présentiel) → Normalité non rejetée

### B. Égalité des variances

Commande :

sdtest stress, by(mode)

Résultat attendu : p = 0.12 → variances considérées égales

## 3. Test t de Student

Commande :

ttest stress, by(mode)

**Interprétation** : Le stress moyen est significativement plus élevé chez les salariés en présentiel que ceux en télétravail.

# 5. Conclusion

Avant d’appliquer une méthode paramétrique dans Stata, utilisez les tests statistiques (Shapiro, Bartlett, Breusch-Pagan…), analysez les graphiques de résidus et de dispersion, et vérifiez que les assomptions clés sont remplies. Cela garantit la validité des tests et des estimations obtenus.