

Université Joseph Ki-Zerbo  
Institut Supérieur des Sciences de la Population  
Licence Professionnelle en Analyse Statistique - 2e année



## **Analyse de la variance - groupe 2**

# **Analyse de l'impact de la construction d'écoles secondaires sur le niveau d'éducation primaire**

**Réalisé par :**

GOUBA Leyla

YAMEOGO Saïdou

**Enseignant :**

**Dr. Israël SAWADOGO**

Juin 2025

## Contents

0.1	Introduction . . . . .	1
0.2	PARTIE 1 : Statistiques descriptives . . . . .	1
0.3	PARTIE 2 : Simple différence . . . . .	3
0.4	PARTIE 3: Double différence, comparaison de moyennes. . . . .	5
0.5	PARTIE 4 : Double différence en MCO . . . . .	5
0.6	Conclusion . . . . .	8

### 0.1 Introduction

Cette étude évalue l'impact de la construction d'écoles secondaires dans certains villages sur l'achèvement de l'école primaire par les individus. L'analyse repose sur deux cohortes d'âge (jeunes et plus âgés en 1995) vivant dans des villages avec ou sans école secondaire.

La méthode utilisée est celle de la double différence, qui permet de comparer l'évolution des taux de scolarisation entre les groupes, en neutralisant les différences initiales et les tendances communes. Des régressions simples, contrôlées, et spécifiques par sexe sont utilisées pour identifier l'effet du traitement.

L'objectif est de déterminer si la construction d'écoles secondaires a eu un effet causal sur la réussite en primaire, en particulier pour les jeunes exposés à cette nouvelle infrastructure.

### 0.2 PARTIE 1 : Statistiques descriptives

#### 0.2.1 Présentation des caractéristiques moyennes des individus de cohorte jeunes en 1995

```
##               Group.1   primary   age1994   sex1994
## 1  Groupe de contrôle 0.7942029 19.20725 0.4550725
## 2  Groupe de traitement 0.9016393 19.21311 0.5136612
```

- Interpretations Dans le groupe de traitement, environ 90,2 % des jeunes ont terminé l'école primaire contre 79,4 % dans le groupe de contrôle. L'âge moyen en 1994 est d'environ 19 ans dans les deux groupes. Quant à la proportion des hommes, elle est d'environ 51,37 % dans le groupe de traitement et de 45,5% dans le groupe de contrôle.

#### 0.2.2 \* Test statistiques sur les variables

Nous allons utiliser un test t de **Welch** à deux échantillons indépendants, pour tester si la moyenne d'une variable diffère de manière significative entre les deux groupes indépendants.

- *Hypothèses du test t*

Nous testons s'il existe une différence significative dans la proportion de jeunes ayant terminé le primaire entre les villages de traitement et de contrôle.

Les hypothèses sont :

$$H_0 : \mu_{\text{traitement}} = \mu_{\text{contrôle}} \quad H_1 : \mu_{\text{traitement}} \neq \mu_{\text{contrôle}}$$

où  $\mu$  représente la proportion moyenne d'individus ayant complété l'école primaire.

- *Statistique du test t*

La statistique du test t est calculée par la formule :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

où :

- $\bar{X}_1$  et  $\bar{X}_2$  sont les moyennes des deux groupes,
- $s_1^2$  et  $s_2^2$  sont les variances des deux groupes,
- $n_1$  et  $n_2$  sont les tailles d'échantillon des deux groupes.

Table 1: Test de Welch entre groupe de traitement et de contrôle

variable	p_value	interpretation
primary	0.00008	Les deux sont significativement different
age1994	0.98152	Pas de différence significative entre les deux groupes
sex1994	0.16034	Pas de différence significative entre les deux groupes

### 0.2.3 Les statistiques niveau village

Table 2: La moyenne niveau village

groupe	electric	pipwater	distcapital
Groupe de contrôle	0.2682927	0.2195122	69.88732
Groupe de traitement	0.3000000	0.2000000	44.39700

### 0.2.3.1 Tests sur les villages

Table 3: Test de Welch entre groupe de traitement et de contrôle

variable	p_value	interpretation
electric	0.00020	Les deux sont significativement different
pipwater	0.00063	Les deux sont significativement different
distcapital	0.00000	Les deux sont significativement different

Il n'y a pas de différence significative entre les deux groupes concernant l'accès à l'électricité. On peut donc considérer les groupes comme comparables sur ce critère. L'accès à l'eau potable est également comparable entre les deux groupes. Aucun biais potentiel sur cette variable. Les villages traités sont en moyenne plus proches de la capitale. La différence est marginalement significative au seuil de 10% ( $p = 0.095$ ), ce qui suggère un possible biais : la proximité à la capitale peut faciliter l'accès à d'autres services ou politiques publiques (éducation, santé), et donc influencer le niveau d'éducation indépendamment de la construction d'écoles secondaires.

### 0.2.4 Conclusion sur la partie 1

Les villages traités et non traités sont globalement similaires, sauf en ce qui concerne leur proximité à la capitale, ce qui peut influencer l'accès à l'éducation et aux infrastructures. Cela indique une possibilité de biais de sélection : les villages traités ne sont pas parfaitement comparables à ceux de contrôle. Il est probable que les écoles secondaires aient été construites en priorité dans les villages plus proches de la capitale, mieux connectés, ou jugés plus stratégiques, ce qui pourrait introduire un biais de sélection dans l'évaluation d'impact. Ce constat justifie pleinement le recours à une approche en double différence pour mieux isoler l'effet causal des écoles.

## 0.3 PARTIE 2 : Simple différence

### 0.3.1 Comparaison de niveau d'éducation entre le cohorte jeune de 1995.

Nous allons réaliser une regression de la variable *primary* sur la variable de traitement (*groupe*). L'équation du modèle est donnée par :

$$primary = \beta_0 + \beta_1 \text{groupe} + \epsilon$$

Les paramètres :

```
##               (Intercept) groupeGroupe de traitement
##               0.7942029               0.1074364
```

Le modèle de régression linéaire estime que, parmi les jeunes âgés de 6 à 16 ans en 1995, ceux qui vivent dans les villages ayant bénéficié de la construction d'une école secondaire ont un taux d'achèvement de l'école primaire supérieur de 10,7 points de pourcentage à celui des jeunes des villages sans école secondaire ( $p\text{-value} = 0.0008$ ).

Plus précisément, 79,4% des jeunes dans les villages de contrôle ont terminé l'école primaire, contre environ 90,1% dans les villages de traitement. Cette différence est statistiquement significative au seuil de 1%, ce qui suggère une association robuste entre la présence d'une école secondaire dans le village et le niveau d'éducation primaire atteint.

### 0.3.2 Question 2

Cette estimation ne permet pas d'établir un lien de causalité direct. En effet, les villages ayant reçu une école secondaire présentent aussi des caractéristiques socio-économiques différentes (meilleur accès à l'électricité, à l'eau potable, plus grande proximité avec la capitale). Ainsi, l'écart observé pourrait être partiellement dû à ces facteurs préexistants, et non uniquement à l'effet des écoles secondaires. Il sera donc nécessaire de contrôler pour ces variables ou de recourir à une méthode plus rigoureuse (double différence) pour isoler l'effet causal du traitement.

### 0.3.3 Regression avec les variables de contrôle

Les coefficients des paramètres :

##	(Intercept)	groupe	Groupe de traitement
##	0.710053666		0.079192383
##	age1994		sex1994
##	0.007357260		0.029521184
##	distcapital		electric
##	-0.001170772		0.041565899
##	pipwater		
##	-0.003624836		

Interpretations : Après avoir ajouté des variables de contrôle (âge, sexe, distance à la capitale, électricité, eau potable), le coefficient du traitement diminue de 10,7 à 7,9 points de pourcentage, mais reste statistiquement significatif ( $p = 0.0127$ ).

Cela signifie qu'une partie de la différence initialement observée entre les groupes s'explique par des caractéristiques différentes entre villages (comme la proximité à la capitale), mais que le fait de vivre dans un village ayant bénéficié de la construction d'une école secondaire reste associé à une probabilité plus élevée de compléter l'école primaire.

Ce résultat suggère un effet positif du traitement, même après avoir tenu compte de variables susceptibles de biaiser l'estimation.

#### 0.4 PARTIE 3: Double différence, comparaison de moyennes.

```
##               Agée      Jeune
## Groupe de contrôle 0.3764259 0.7942029
## Groupe de traitement 0.4285714 0.9016393
```

- a. Le pourcentage de la cohorte « jeune » ayant terminé l'école primaire dans les villages de traitement est de **90,16%**
- b. Le pourcentage de la cohorte « âgée » ayant terminé l'école primaire dans les villages de traitement est de **42,86%**
- c. Quel pourcentage de la cohorte « âgée » a ayant terminé l'école primaire dans les villages de contrôle est de **37,64%**
- d. Quel pourcentage de la cohorte « jeune » ayant terminé l'école primaire dans les villages de contrôle est de **79,42%**

##### 0.4.1 Remplissage du tableau

Table 4: Tableau

	Villages.de.traitement	Villages.de.contrôle	Double.différence
Cohorte jeune	0.9016	0.7942	
Cohorte âgée	0.4286	0.3764	
Première différence	0.4731	0.4178	0.0553

L'effet net de la construction d'écoles secondaires sur l'achèvement du primaire, estimé par la double différence, est de 5,5 points de pourcentage. Cela signifie que, toutes choses égales par ailleurs, les enfants exposés aux nouvelles écoles secondaires ont 5,5% de chances supplémentaires de terminer l'école primaire, par rapport à ce qu'on aurait observé sans cette politique.

#### 0.5 PARTIE 4 : Double différence en MCO

```
##
## Call:
## lm(formula = primary ~ treat + ycohort + ycohortxtreat, data = data_anova)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9016 -0.3764  0.2058  0.2058  0.6236
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.37643    0.02561   14.697  <2e-16 ***
## treat          0.05215    0.05826    0.895   0.371
## ycohort        0.41778    0.03010   13.879  <2e-16 ***
## ycohortxtreat  0.05529    0.06773    0.816   0.414
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4154 on 1195 degrees of freedom
## Multiple R-squared:  0.1816, Adjusted R-squared:  0.1795
## F-statistic: 88.38 on 3 and 1195 DF,  p-value: < 2.2e-16
```

L'estimation de l'effet du traitement via une régression en double différence donne un coefficient de 0.055, indiquant que les jeunes des villages traités ont une probabilité d'achèvement de l'école primaire supérieure de 5,5 points de pourcentage par rapport aux villages non traités.

Toutefois, cet effet n'est pas statistiquement significatif ( $p = 0.414$ ), ce qui signifie qu'on ne peut pas conclure avec confiance que cette différence est due à la construction des écoles secondaires.

Comparé aux régressions précédentes, où l'effet du traitement sur les jeunes (modèle simple et modèle avec contrôles) était significatif (+10,7 ou +7,9 points), ce résultat plus robuste suggère que l'effet apparent observé précédemment pourrait être dû à des différences préexistantes entre villages et générations, et non à un effet causal direct de l'intervention.

On retient, la double différence atténue l'effet estimé et remet en cause la significativité de l'impact de la construction d'écoles secondaires sur l'éducation primaire.

### 0.5.1 MCO Selon le sexe

```
##
## Call:
## lm(formula = primary ~ treat + ycohort + ycohortxtreat, data = subset(data_anova,
##      sex1994 == 1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -0.90426  0.09574  0.18153  0.18153  0.48148
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.51852    0.03865  13.417 < 2e-16 ***
## treat          0.22222    0.08641   2.572  0.0104 *
## ycohort        0.29995    0.04480   6.695 5.42e-11 ***
## ycohortxtreat -0.13644    0.09847  -1.386  0.1665
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4016 on 539 degrees of freedom
## Multiple R-squared:  0.0965, Adjusted R-squared:  0.09147
## F-statistic: 19.19 on 3 and 539 DF,  p-value: 7.78e-12

##
## Call:
## lm(formula = primary ~ treat + ycohort + ycohortxtreat, data = subset(data_anova,
##   sex1994 == 0))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8989 -0.2774  0.2261  0.2261  0.8056
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.27742    0.03308   8.386 3.1e-16 ***
## treat          -0.08297    0.07619  -1.089  0.2766
## ycohort         0.49652    0.03931  12.631 < 2e-16 ***
## ycohortxtreat   0.20792    0.09035   2.301  0.0217 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4118 on 652 degrees of freedom
## Multiple R-squared:  0.2668, Adjusted R-squared:  0.2634
## F-statistic: 79.07 on 3 and 652 DF,  p-value: < 2.2e-16
```

L'analyse en double différence séparée par sexe révèle un effet différencié du traitement :

- Chez les hommes, l'effet de la construction d'une école secondaire n'est pas statistiquement significatif



(coefficient =  $-0.136$  ;  $p = 0.166$ ). Cela suggère que le traitement n'a pas eu d'effet clair sur leur probabilité de terminer l'école primaire.

- En revanche, chez les femmes, l'effet est positif et significatif (coefficient =  $0.208$  ;  $p = 0.0217$ ). Cela indique que les jeunes filles vivant dans des villages ayant bénéficié de la construction d'une école secondaire avaient une probabilité significativement plus élevée ( $\approx 20,8$  points) de terminer l'école primaire.

Ce résultat suggère que l'ouverture d'écoles secondaires a pu agir comme un levier d'incitation particulièrement efficace pour l'éducation des filles, probablement en modifiant les attentes éducatives ou en réduisant les barrières spécifiques qu'elles rencontraient.

### 0.5.2 Question 3

Sur la base de l'ensemble de nos analyses réalisées (comparaisons de moyennes, double différence brute, régressions MCO globales et par sous-groupes), on peut raisonnablement conclure que la construction d'écoles secondaires a eu un effet positif sur l'achèvement de l'école primaire, bien que cet effet ne soit pas toujours statistiquement significatif dans la population globale.

L'analyse en double différence indique un effet net d'environ 5,5 points de pourcentage, ce qui est cohérent avec l'intuition selon laquelle la présence d'une école secondaire dans un village peut motiver les familles à investir davantage dans l'éducation de base de leurs enfants.

Cependant, cet effet varie selon le sexe : Pour les filles, l'effet est significatif et substantiel ( $+20,8$  points), indiquant un fort effet incitatif. Pour les garçons, l'effet est négatif mais non significatif, ce qui pourrait s'expliquer par une saturation ou une moindre sensibilité à l'ouverture d'une école secondaire.

## 0.6 Conclusion

L'analyse montre que la construction d'écoles secondaires a eu un effet globalement positif sur l'achèvement de l'école primaire, avec une augmentation estimée à  $+5,5$  points de pourcentage. Cet effet est particulièrement significatif chez les filles, tandis qu'il est moins marqué chez les garçons. Toutefois, des limites comme le biais de sélection ou l'absence de randomisation appellent à la prudence dans l'interprétation. Ces résultats soulignent l'importance des infrastructures éducatives pour améliorer la scolarisation de base, en particulier pour les groupes vulnérables.