



Rapport de projet

TP ECONOMETRIE DES MODELES LINEAIRES



ENSEIGNANT

DR Fabrice YAMEOGO

29 juillet 2024

MEMBRES DU GROUPe 6

BATIONO Martinien

CISSE Oumarou

NANA Hervé

Table des matières

[Introduction 2](#_Toc173180797)

[Chargement des packages importants pour l’analyse 2](#_Toc173180798)

[Importation de la base 3](#_Toc173180799)

[Description de la base 3](#_Toc173180800)

[Choix de la méthode 4](#_Toc173180801)

[Choix des variables 5](#_Toc173180802)

[Variable dépendante: COUT 5](#_Toc173180803)

[Variables indépendantes: Superficie, Cout\_m2, Site, Usage, Taxe\_jouissance, Type\_option 5](#_Toc173180804)

[Création des bases par année et modélisation 6](#_Toc173180805)

[Création des bases 6](#_Toc173180806)

[Importation 6](#_Toc173180807)

[Modélisation 8](#_Toc173180808)

[Vérification et correction des hypothèses 15](#_Toc173180809)

[Vérification 15](#_Toc173180810)

[Pour le modèle de 2018 15](#_Toc173180811)

[Pour le modèle de 2019 16](#_Toc173180812)

[Pour le modèle de 2020 17](#_Toc173180813)

[Pour le modèle de 2021 18](#_Toc173180814)

[Pour le modèle de 2022 19](#_Toc173180815)

[Pour le modèle de 2023 20](#_Toc173180816)

[Pour le modèle de 2024 21](#_Toc173180817)

[Correction des hypothèses 22](#_Toc173180818)

[Pour le modèle de 2018 22](#_Toc173180819)

[Pour le modèle de 2019 23](#_Toc173180820)

[Pour le modèle de 2020 23](#_Toc173180821)

[Pour le modèle de 2021 24](#_Toc173180822)

[Pour le modèle de 2022 24](#_Toc173180823)

[Pour le modèle de 2023 25](#_Toc173180824)

[Pour le modèle de 2024 25](#_Toc173180825)

[Calcul de l’indice 26](#_Toc173180826)

[Conclusion 28](#_Toc173180827)

# Introduction

Ouagadougou, la capitale du Burkina Faso, est une ville en pleine expansion. En tant que centre administratif, économique et culturel du pays, elle attire de plus en plus de résidents et d’investisseurs. Cette dynamique de croissance influence directement le marché foncier, faisant des parcelles de terrain une ressource de plus en plus prisée. Le prix des parcelles à Ouagadougou varie considérablement en fonction de plusieurs facteurs. Les quartiers proches du centre-ville ou des zones d’activités économiques majeures tendent à avoir des prix plus élevés en raison de leur attractivité et de leur accessibilité. En revanche, les zones périphériques ou en cours de développement offrent des terrains à des prix plus abordables, mais avec des perspectives de valorisation future.

# Chargement des packages importants pour l’analyse

library(readr)  
library(tidyverse)

library(lubridate)  
library(car)

library(performance)  
library(lmtest)

library(questionr)  
library(dplyr)  
library(performance)

# Importation de la base

setwd("C:/Users/cisse/Desktop/Projet\_quanti/Projet\_quanti")  
data <- read\_delim("Parcelles.csv", delim = ";", escape\_double = FALSE, trim\_ws = TRUE)

## Rows: 1811 Columns: 15

names(data)

## [1] "Numero" "Ville" "Site"   
## [4] "Usage" "Superficie" "Cout\_m2"   
## [7] "COUT" "Taxe\_Jouissance" "Type\_option"   
## [10] "Date\_vente" "Date\_fin\_contrat" "attestation\_etablie"  
## [13] "plan\_etablie" "Presence\_ONEA" "Presence\_SONABEL"

Dans notre jeu de données, on retrouve les différentes variables ci-dessus mentionnées notamment le COUT de la parcelle, sa Superficie, le Cout\_m2, le Site, l’Usage,…

## Description de la base

summary(data)

## Numero Ville Site Usage   
## Min. : 1.0 Length:1811 Length:1811 Length:1811   
## 1st Qu.: 453.5 Class :character Class :character Class :character   
## Median : 906.0 Mode :character Mode :character Mode :character   
## Mean : 906.0   
## 3rd Qu.:1358.5   
## Max. :1811.0   
## Superficie Cout\_m2 COUT Taxe\_Jouissance   
## Min. : 82 Min. : 0 Min. : 0 Min. : 0.0   
## 1st Qu.: 300 1st Qu.: 26000 1st Qu.: 7644000 1st Qu.: 250.0   
## Median : 320 Median : 26000 Median : 8320000 Median : 500.0   
## Mean : 495 Mean : 27350 Mean : 16068350 Mean : 514.8   
## 3rd Qu.: 488 3rd Qu.: 36000 3rd Qu.: 17107500 3rd Qu.: 500.0   
## Max. :13611 Max. :190000 Max. :722376000 Max. :3000.0   
## Type\_option Date\_vente Date\_fin\_contrat attestation\_etablie  
## Length:1811 Length:1811 Length:1811 Length:1811   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## plan\_etablie Presence\_ONEA Presence\_SONABEL   
## Length:1811 Length:1811 Length:1811   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character

Parmi les variables de notre jeu de données, quatre sont quantitatives à savoir le COUT de la parcelle, sa Superficie, le Cout\_m2 et la Taxe\_jouissance. Les autres variables notamment Usage, Site, Site, … sont catégorielles. On remarque que parmi les variables quantitatives, le minimum du COUT est 0 ; A Ouagadougou, on ne peut avoir une parcelle à 0f. Pour ce faire nous allons supprimer les observations qui ont 0 comme COUT de parcelle.

data <- subset(data,COUT != 0)

## Choix de la méthode

La méthode hédonique adoptée par notre groupe est la méthode des caractéristiques. Cette méthode est appropriée car elle permet de modéliser le cout d’un bien en fonction des différentes caractéristiques. Elle permet ainsi de faire une comparaison de la variation de ce cout par année à travers le calcul d’indices notamment l’indice des prix de Laspeyres et de Paasche.

## Choix des variables

### Variable dépendante: COUT

L’analyse consiste à comparer l’indice d’évolutions des prix des parcelles à Ouagadougou entre 2018 et 2024; la variable dépendante est naturellement le COUT de la parcelle.

### Variables indépendantes: Superficie, Cout\_m2, Site, Usage, Taxe\_jouissance, Type\_option

Notre choix de variables indépendantes s’est tourné vers les variables : Superficie, Cout\_m2, Site, Usage, Taxe\_jouissance, Type\_option; car selon nous le prix des parcelles à Ouagadougou serait influencé par la superficie, le quartier s’il est proche du centre-ville ou des périphéries, l’usage pour habitation ou commerce, la taxe liée au terrain et la modalité de paiement : au comptant ou par tranche. Ainsi, nous allons donc exclure certaines qui sont suceptibles de ne pas nous servir dans l’analyse notamment les variables à une modalité : Ville, présence ONEA(qui est présent partout) et SONABEL(aussi présent partout) et certaines variables: attestation établie, plan établie que nous jugeons pas trop influents le prix des parcelles.

data = data[,-c(1,2,12,13,14,15)]  
names(data)

## [1] "Site" "Usage" "Superficie" "Cout\_m2"   
## [5] "COUT" "Taxe\_Jouissance" "Type\_option" "Date\_vente"   
## [9] "Date\_fin\_contrat"

table(data$Usage)

##   
## COMMERCE COMMERCE A L'ANGLE COMMERCE ANGLE   
## 104 193 2   
## COMMERCE ANGLE 1 BITUME COMMERCE ANGLE 2 VOIES COMMERCE ORDINAIRE ANGLE   
## 55 15 27   
## COMMUNAUTAIRE HABITATION HABITATION ANGLE   
## 20 865 224   
## STATION SERVICE   
## 2

table(data$Site)

##   
## BASSINKO SITE - BA CISSIN 2020 - SITE G OUAGA 2000 - SITE A   
## 97 9 158   
## OUAGA 2000 - SITE AA SECTEUR 16 OUAGA SILMIOUGOU   
## 376 1 866

Après tabulation des deux variables catégorielles : Usage et Site, il se trouve qu’on peut regrouper certaines modalités de ces variables notamment harmoniser commerce dans usage et Ouaga 2000 dans site et profiter raffiner les différentes modalités.

## Recoding data$Usage into data$Usage\_rec  
data$Usage <- data$Usage %>%  
 fct\_recode(  
 "COMMERCE" = "COMMERCE A L'ANGLE",  
 "COMMERCE" = "COMMERCE ANGLE",  
 "COMMERCE" = "COMMERCE ANGLE 1 BITUME",  
 "COMMERCE" = "COMMERCE ANGLE 2 VOIES",  
 "COMMERCE" = "COMMERCE ORDINAIRE ANGLE",  
 "HABITATION" = "HABITATION ANGLE")  
  
## Recoding data$Site into data$Site\_rec  
data$Site <- data$Site %>%  
 fct\_recode(  
 "BASSINKO" = "BASSINKO SITE - BA",  
 "CISSIN" = "CISSIN 2020 - SITE G",  
 "OUAGA 2000" = "OUAGA 2000 - SITE A",  
 "OUAGA 2000" = "OUAGA 2000 - SITE AA")

Nous allons transformer les variables Date\_vente et Date\_contrat en format date et créer une nouvelle variable annee qui va prendre les années de la variable Date\_vente en modalité. Cette variable va servir à scinder la base selon les années.

#Conversion de certianes variables date en format date  
data <- data %>%  
 mutate(Date\_vente = dmy(Date\_vente),Date\_fin\_contrat = dmy(Date\_fin\_contrat))  
  
# Création de la variable annee  
data$annee = year(data$Date\_vente)

## Création des bases par année et modélisation

### Création des bases

Le code suivant scinde notre base en fonction de l’année de vente de la parcelle.

# Création d'un fichier CSV pour chaque année----  
data %>%  
 split(.$annee) %>%  
 walk(~ write\_csv(.x, paste0("parcelles\_", unique(.x$annee), ".csv")))

### Importation

Nous allons à présent importer toutes les bases par année crées.

## Importation des bases  
parcelles\_2018 <- read\_csv("parcelles\_2018.csv")

## Rows: 168 Columns: 10

parcelles\_2019 <- read\_csv("parcelles\_2019.csv")

## Rows: 104 Columns: 10

parcelles\_2020 <- read\_csv("parcelles\_2020.csv")

## Rows: 219 Columns: 10

parcelles\_2021 <- read\_csv("parcelles\_2021.csv")

## Rows: 78 Columns: 10

parcelles\_2022 <- read\_csv("parcelles\_2022.csv")

## Rows: 45 Columns: 10

parcelles\_2023 <- read\_csv("parcelles\_2023.csv")

## Rows: 823 Columns: 10

parcelles\_2024 <- read\_csv("parcelles\_2024.csv")

## Rows: 70 Columns: 10

### Modélisation

Nous allons modéliser le COUT des parcelles par année.

#### Pour 2018

# Modélisation  
model\_2018 <- lm(COUT ~ Superficie + Cout\_m2 + Taxe\_Jouissance + as.factor(Type\_option) +  
 as.factor(Site) + as.factor(Usage),data = parcelles\_2018)  
  
# Résumé du modèle  
summary(model\_2018)

##   
## Call:  
## lm(formula = COUT ~ Superficie + Cout\_m2 + Taxe\_Jouissance +   
## as.factor(Type\_option) + as.factor(Site) + as.factor(Usage),   
## data = parcelles\_2018)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -59514839 -1851367 -872481 1787868 55807544   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -4.559e+07 1.219e+07 -3.740 0.000257 \*\*\*  
## Superficie 4.058e+04 1.500e+03 27.050 < 2e-16 \*\*\*  
## Cout\_m2 9.442e+02 9.762e+01 9.673 < 2e-16 \*\*\*  
## Taxe\_Jouissance 4.006e+03 2.161e+03 1.854 0.065636 .   
## as.factor(Type\_option)ACOMPTE 50% -1.217e+07 4.695e+06 -2.592 0.010434 \*   
## as.factor(Type\_option)COMPTANT -1.269e+07 3.855e+06 -3.291 0.001229 \*\*   
## as.factor(Site)OUAGA 2000 1.680e+07 8.914e+06 1.885 0.061297 .   
## as.factor(Usage)COMMUNAUTAIRE -5.700e+07 8.652e+06 -6.588 6.28e-10 \*\*\*  
## as.factor(Usage)HABITATION 5.609e+06 2.374e+06 2.362 0.019378 \*   
## as.factor(Usage)STATION SERVICE -1.907e+05 1.104e+07 -0.017 0.986236   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10800000 on 158 degrees of freedom  
## Multiple R-squared: 0.895, Adjusted R-squared: 0.889   
## F-statistic: 149.7 on 9 and 158 DF, p-value: < 2.2e-16

Les variables superficie, cout\_m2, taxe\_jouissance,type\_optionACOMPTE 50%,site sont significatives. Ces variables expliquent vraiment le cout des parcelles en 2018(ces variables ont un impact sur le cout des parcelles) La superficie des parcelles augmente leurs couts de 0.000006229 unités. Le prix des parcelles situées à OUAGA 2000 augmente de 1567unités R2=0.8999 alors les variables explicatives ajustent très bien le cout des parcelles (89.99%)

#### Pour 2019

model\_2019 <- lm(COUT ~ Superficie + Cout\_m2 + Taxe\_Jouissance + as.factor(Type\_option) +  
 as.factor(Site) + as.factor(Usage),data = parcelles\_2019)  
  
# Résumé du modèle  
summary(model\_2019)

##   
## Call:  
## lm(formula = COUT ~ Superficie + Cout\_m2 + Taxe\_Jouissance +   
## as.factor(Type\_option) + as.factor(Site) + as.factor(Usage),   
## data = parcelles\_2019)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -101907776 -4192112 2092558 3053075 36441562   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -4.677e+07 1.679e+07 -2.785 0.00646 \*\*   
## Superficie 6.917e+04 1.697e+03 40.760 < 2e-16 \*\*\*  
## Cout\_m2 3.643e+02 1.232e+02 2.958 0.00391 \*\*   
## Taxe\_Jouissance -1.490e+03 3.328e+03 -0.448 0.65542   
## as.factor(Type\_option)ACOMPTE 50% 1.694e+07 1.583e+07 1.070 0.28722   
## as.factor(Type\_option)COMPTANT 9.847e+06 9.272e+06 1.062 0.29093   
## as.factor(Site)SECTEUR 16 OUAGA 7.686e+06 1.328e+07 0.579 0.56423   
## as.factor(Usage)COMMUNAUTAIRE -1.491e+08 1.451e+07 -10.277 < 2e-16 \*\*\*  
## as.factor(Usage)HABITATION 1.262e+07 1.572e+07 0.803 0.42389   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 12710000 on 95 degrees of freedom  
## Multiple R-squared: 0.9705, Adjusted R-squared: 0.9681   
## F-statistic: 391.3 on 8 and 95 DF, p-value: < 2.2e-16

Les variables superficie, cout\_m2, UsageCOMMUNAUTAIRE sont significatives. Ces variables expliquent vraiment le cout des parcelles en 2019. La superficie des parcelles augmente leurs couts de 0.0002893 unités. Les parcelles à usage communautaire ont des prix augmentés de 1319 unités. Le R2=0.9049, ce qui veut dire que les variables indépendantes expliquent très bien le cout des parcelles (90.49%).

#### Pour 2020

model\_2020 <- lm(COUT ~ Superficie + Cout\_m2 + Taxe\_Jouissance + as.factor(Type\_option) +  
 as.factor(Site) + as.factor(Usage) ,data = parcelles\_2020)  
  
# Résumé du modèle  
summary(model\_2020)

##   
## Call:  
## lm(formula = COUT ~ Superficie + Cout\_m2 + Taxe\_Jouissance +   
## as.factor(Type\_option) + as.factor(Site) + as.factor(Usage),   
## data = parcelles\_2020)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -85810303 -2962870 504561 1017229 85810303   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -3.485e+07 9.585e+06 -3.636 0.000349 \*\*\*  
## Superficie 3.381e+04 1.015e+03 33.319 < 2e-16 \*\*\*  
## Cout\_m2 9.903e+02 1.215e+02 8.154 3.23e-14 \*\*\*  
## Taxe\_Jouissance -1.914e+03 1.516e+03 -1.263 0.208097   
## as.factor(Type\_option)ACOMPTE 50% -4.643e+06 7.692e+06 -0.604 0.546752   
## as.factor(Type\_option)COMPTANT -7.521e+06 6.653e+06 -1.130 0.259561   
## as.factor(Site)OUAGA 2000 1.368e+07 2.850e+06 4.801 3.00e-06 \*\*\*  
## as.factor(Usage)COMMUNAUTAIRE -1.210e+08 1.021e+07 -11.853 < 2e-16 \*\*\*  
## as.factor(Usage)HABITATION -7.464e+06 7.993e+06 -0.934 0.351484   
## as.factor(Usage)STATION SERVICE 1.395e+07 1.119e+07 1.247 0.213822   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10750000 on 209 degrees of freedom  
## Multiple R-squared: 0.9158, Adjusted R-squared: 0.9122   
## F-statistic: 252.7 on 9 and 209 DF, p-value: < 2.2e-16

#### Pour 2021

model\_2021 <- lm(COUT ~ Superficie + Cout\_m2 + Taxe\_Jouissance + as.factor(Type\_option) +  
 as.factor(Usage) + as.factor(Site),data = parcelles\_2021)  
  
# Résumé du modèle  
summary(model\_2021)

##   
## Call:  
## lm(formula = COUT ~ Superficie + Cout\_m2 + Taxe\_Jouissance +   
## as.factor(Type\_option) + as.factor(Usage) + as.factor(Site),   
## data = parcelles\_2021)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -49150562 -3691447 683218 3216960 42214006   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -5.329e+07 1.293e+07 -4.121 0.000102 \*\*\*  
## Superficie 2.156e+04 1.130e+03 19.080 < 2e-16 \*\*\*  
## Cout\_m2 1.064e+03 2.841e+02 3.744 0.000368 \*\*\*  
## Taxe\_Jouissance 1.340e+04 3.616e+03 3.705 0.000419 \*\*\*  
## as.factor(Type\_option)ACOMPTE 50% -6.656e+06 6.930e+06 -0.960 0.340188   
## as.factor(Usage)COMMUNAUTAIRE -2.412e+07 1.191e+07 -2.026 0.046603 \*   
## as.factor(Usage)HABITATION NA NA NA NA   
## as.factor(Site)CISSIN 2.454e+08 4.035e+07 6.082 5.6e-08 \*\*\*  
## as.factor(Site)OUAGA 2000 2.149e+07 6.400e+06 3.358 0.001271 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 14020000 on 70 degrees of freedom  
## Multiple R-squared: 0.9853, Adjusted R-squared: 0.9839   
## F-statistic: 671.8 on 7 and 70 DF, p-value: < 2.2e-16

Le R2 ajusté est de 0.9913; ce qui veut dire que le modèle s’ajuste aux données à 99.13%. En 2021, les caractéristiques qui influençaient le prix des parcelles à Ouagadougou étaient la superficie, le Cout\_m2 et la Taxe\_jouissance. Une augmentation de la superficie d’une unité entrainait une hausse du prix de la parcelle de 29550. Aussi une hausse du prix du m2 d’une unité augmentait le cout de la parcelle de 574 tandis qu’une augmentation de la Taxe\_jouissance diminuait le prix de 858.

#### Pour 2022

model\_2022 <- lm(COUT ~ Superficie + Cout\_m2 + Taxe\_Jouissance + as.factor(Type\_option) +  
 as.factor(Site) + as.factor(Usage),data = parcelles\_2022)  
  
# Résumé du modèle  
summary(model\_2022)  
## Call:  
## lm(formula = COUT ~ Superficie + Cout\_m2 + Taxe\_Jouissance +   
## as.factor(Type\_option) + as.factor(Site) + as.factor(Usage),   
## data = parcelles\_2022)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8741199 -6013027 1448894 1974750 39541417   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6.335e+07 1.382e+07 -4.584 4.82e-05 \*\*\*  
## Superficie 1.343e+04 1.386e+03 9.687 8.20e-12 \*\*\*  
## Cout\_m2 1.988e+03 3.606e+02 5.513 2.65e-06 \*\*\*  
## Taxe\_Jouissance -1.360e+03 3.391e+03 -0.401 0.69050   
## as.factor(Type\_option)ACOMPTE 50% -1.771e+07 7.631e+06 -2.320 0.02580 \*   
## as.factor(Site)OUAGA 2000 1.407e+07 4.378e+06 3.214 0.00267 \*\*   
## as.factor(Usage)COMMUNAUTAIRE 3.836e+07 1.218e+07 3.150 0.00318 \*\*   
## as.factor(Usage)HABITATION NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8174000 on 38 degrees of freedom  
## Multiple R-squared: 0.8406, Adjusted R-squared: 0.8154   
## F-statistic: 33.4 on 6 and 38 DF, p-value: 1.064e-13

Pour l’année 2022 nous appliquons un modèle linéaire multiple pour estimer le coût des parcelles (COUT) à partir des variables telles que la superficie, le coût par mètre carré, l’attestation établie, le site, et l’usage. Ainsi Superficie et Cout\_m2 ont des coefficients élevés et significatifs (p < 2e-16). Cela suggère que ces variables ont un impact assez fort et positif sur le coût des parcelles. Chaque unité de la superficie et le coût par m² augmente le coût de la parcelle de manière équivalente. SiteSILMIOUGOU a un coefficient significatif (p < 0.001), ce qui suggère que ce site a un impact positif sur le coût.

#### Pour 2023

model\_2023 <- lm(COUT ~ Superficie + Cout\_m2 + as.factor(Type\_option) + Taxe\_Jouissance +  
 as.factor(Site) + as.factor(Usage), data = parcelles\_2023)  
  
# Résumé du modèle  
summary(model\_2023)

##   
## Call:  
## lm(formula = COUT ~ Superficie + Cout\_m2 + as.factor(Type\_option) +   
## Taxe\_Jouissance + as.factor(Site) + as.factor(Usage), data = parcelles\_2023)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3649165 -272652 199516 234372 6316587   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.766e+07 2.887e+05 -61.176 < 2e-16 \*\*\*  
## Superficie 2.949e+04 3.018e+02 97.714 < 2e-16 \*\*\*  
## Cout\_m2 5.706e+02 1.034e+01 55.174 < 2e-16 \*\*\*  
## as.factor(Type\_option)ACOMPTE 50% 6.227e+05 1.575e+05 3.953 8.39e-05 \*\*\*  
## Taxe\_Jouissance -9.877e+02 4.733e+02 -2.087 0.03722 \*   
## as.factor(Site)OUAGA 2000 -1.636e+06 5.197e+05 -3.148 0.00171 \*\*   
## as.factor(Site)SILMIOUGOU 2.043e+06 2.948e+05 6.928 8.67e-12 \*\*\*  
## as.factor(Usage)HABITATION NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 594200 on 816 degrees of freedom  
## Multiple R-squared: 0.9916, Adjusted R-squared: 0.9915   
## F-statistic: 1.596e+04 on 6 and 816 DF, p-value: < 2.2e-16

Comme dans le modèle précédent, les variables superficie et cout\_m2 ont des coefficients élevés et significatifs (p < 2e-16). Ce qui suggère que ces variables ont un impact très fort et positif sur le coût des parcelles. Chaque unité de superficie ou de coût par m² augmente le coût de la parcelle de manière équivalente. Concernant les sites de localisation des parcelles , SiteSILMIOUGOU à un coefficient significatif (p < 0.01), ce qui suggère que ce site a un impact notable sur le coût des parcelles. Le modèle est globalement significatif (R2 ajusté = 0.9915)

#### Pour 2024

model\_2024 <- lm(COUT ~ Superficie + Cout\_m2 + Taxe\_Jouissance + as.factor(Type\_option)  
 + as.factor(Site) + as.factor(Usage),data = parcelles\_2024)  
  
# Résumé du modèle  
summary(model\_2024)

##   
## Call:  
## lm(formula = COUT ~ Superficie + Cout\_m2 + Taxe\_Jouissance +   
## as.factor(Type\_option) + as.factor(Site) + as.factor(Usage),   
## data = parcelles\_2024)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1579460 -166247 -35218 218630 2623498   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -9.423e+06 1.408e+06 -6.693 6.41e-09 \*\*\*  
## Superficie 2.854e+04 8.811e+02 32.389 < 2e-16 \*\*\*  
## Cout\_m2 4.903e+02 4.509e+01 10.874 3.50e-16 \*\*\*  
## Taxe\_Jouissance -8.610e+03 1.132e+03 -7.608 1.58e-10 \*\*\*  
## as.factor(Type\_option)ACOMPTE 50% 8.554e+06 1.343e+06 6.369 2.35e-08 \*\*\*  
## as.factor(Site)SILMIOUGOU NA NA NA NA   
## as.factor(Usage)HABITATION NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 480200 on 64 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.9952, Adjusted R-squared: 0.9949   
## F-statistic: 3322 on 4 and 64 DF, p-value: < 2.2e-16

Le modèle est globalement significatif : Le R2 ajusté est de 0.99, ce qui veut dire le modèle explique le cout des parcelles à 99.52% en 2024. Pour ce modèle, c’est la superficie, le cout\_m2, la Taxe\_jouissance et l’option acompte 50% qui sont significatifs au seuil de 5% ; Ce qui veut dire qu’en 2024, toute chose étant égale par ailleurs, les prix des parcelles dépendaient plus de la superficie, du Cout\_m2 et de la Taxe\_jouissance et l’option de paiement.

# Vérification et correction des hypothèses

## Vérification

### Pour le modèle de 2018

# Test d'homoscédasticité  
check\_heteroscedasticity(model\_2018)

## Warning: Heteroscedasticity (non-constant error variance) detected (p < .001).

# Test d'autocorrélation  
check\_autocorrelation(model\_2018)

## Warning: Autocorrelated residuals detected (p = 0.044).

# Test de normalité  
check\_normality(model\_2018)

## Warning: Non-normality of residuals detected (p < .001).

# Test de multicollinéarité  
check\_collinearity(model\_2018)

## # Check for Multicollinearity  
##   
## Low Correlation  
##   
## Term VIF VIF 95% CI Increased SE Tolerance  
## Superficie 1.54 [1.32, 1.93] 1.24 0.65  
## Cout\_m2 3.30 [2.64, 4.24] 1.82 0.30  
## Taxe\_Jouissance 2.46 [2.00, 3.13] 1.57 0.41  
## as.factor(Type\_option) 1.75 [1.47, 2.19] 1.32 0.57  
## as.factor(Site) 3.30 [2.64, 4.24] 1.82 0.30  
## as.factor(Usage) 3.23 [2.58, 4.14] 1.80 0.31

Pour ce modèle, les hypothèses d’autocorrélation et de normalité des résidus ne sont pas vérifiées. Les résidus ne suivent donc pas une loi normale et ils sont autocorrélés.

### Pour le modèle de 2019

# Test d'homoscédasticité  
check\_heteroscedasticity(model\_2019)

## Warning: Heteroscedasticity (non-constant error variance) detected (p < .001).

# Test d'autocorrélation  
check\_autocorrelation(model\_2019)

## OK: Residuals appear to be independent and not autocorrelated (p = 0.762).

# Test de normalité  
check\_normality(model\_2019)

## Warning: Non-normality of residuals detected (p < .001).

# Test de multicollinéarité  
check\_collinearity(model\_2019)

Pour ce modèle, les hypothèses d’homoscédasticité et de normalité des résidus ne sont pas vérifiées. Aussi, certaines variables notamment Usage et Type\_option ont une forte corrélation.

### Pour le modèle de 2020

# Test d'homoscédasticité  
check\_heteroscedasticity(model\_2020)

## Warning: Heteroscedasticity (non-constant error variance) detected (p < .001).

# Test d'autocorrélation  
check\_autocorrelation(model\_2020)

## Warning: Autocorrelated residuals detected (p = 0.006).

# Test de normalité  
check\_normality(model\_2020)

## Warning: Non-normality of residuals detected (p < .001).

# Test de multicollinéarité  
check\_collinearity(model\_2020)

Pour ce modèle, les hypothèses d’autocorrélations et de normalité des résidus ne sont pas vérifiées ainsi que l’homoscédasticité. Aussi, certaines variables notamment Usage et Type\_option ont une forte corrélation.

### Pour le modèle de 2021

# Test d'homoscédasticité  
check\_heteroscedasticity(model\_2021)

## Warning: Heteroscedasticity (non-constant error variance) detected (p < .001).

# Test d'autocorrélation  
check\_autocorrelation(model\_2021)

## OK: Residuals appear to be independent and not autocorrelated (p = 0.664).

# Test de normalité  
check\_normality(model\_2021)

## Warning: Non-normality of residuals detected (p < .001).

# Test de multicollinéarité  
check\_collinearity(model\_2021)

Pour ce modèle, certaines des hypothèses sont vérifiées : les résidus sont non auto corrélés. Par contre, les variables Cout\_m2 et Site ont une forte corrélation.

### Pour le modèle de 2022

# Test d'homoscédasticité  
check\_heteroscedasticity(model\_2022)

## Warning: Heteroscedasticity (non-constant error variance) detected (p < .001).

# Test d'autocorrélation  
check\_autocorrelation(model\_2022)

## OK: Residuals appear to be independent and not autocorrelated (p = 0.350).

# Test de normalité  
check\_normality(model\_2022)

## Warning: Non-normality of residuals detected (p < .001).

# Test de multicollinéarité  
check\_collinearity(model\_2022)

Pour ce modèle, les résidus ne suivent pas une loi normale mais sont non autocorrélés et la variable Cout\_m2 a une forte corrélation.

### Pour le modèle de 2023

# Test d'homoscédasticité  
check\_heteroscedasticity(model\_2023)

## Warning: Heteroscedasticity (non-constant error variance) detected (p < .001).

# Test d'autocorrélation  
check\_autocorrelation(model\_2023)

## Warning: Autocorrelated residuals detected (p < .001).

# Test de normalité  
check\_normality(model\_2023)

## Warning: Non-normality of residuals detected (p < .001).

# Test de multicollinéarité  
check\_collinearity(model\_2023)

Pour ce modèle, aucune hypothèse n’est vérifiée et la variable Taxe\_jouissance a une forte corrélation.

### Pour le modèle de 2024

# Test d'homoscédasticité  
check\_heteroscedasticity(model\_2024)

## Warning: Heteroscedasticity (non-constant error variance) detected (p < .001).

# Test d'autocorrélation  
check\_autocorrelation(model\_2024)

## Warning: Autocorrelated residuals detected (p = 0.034).

# Test de normalité  
check\_normality(model\_2024)

## Warning: Non-normality of residuals detected (p < .001).

# Test de multicollinéarité  
check\_collinearity(model\_2024)

Pour ce modèle, seulement l’hypothèse de non autocorrélation des résidus est vérifiée et il y a une forte corrélation entre les variables Cout\_m2, Taxe\_jouissance et Type\_option.

## Correction des hypothèses

Pour l’homoscédasticité, nous avions appliqué le logarithme au cout, prendre sa racine carrée aussi la transformation de boxcox ; mais dans certains modèles cette hypothèse n’est pas vérifiée et cela joue beaucoup sur nos indices, du coup nous avons décidé de laisser juste le modèle brut sans aucune transformation. Quand on applique par exemple le logarithme au cout, l’homoscédasticité est vérifiée pour le modèle de 2018 (non autocorrélation aussi), 2020 ,2021 et 2022 où toutes les hypothèses sont vérifiées : seulement certaines variables ont une forte corrélation (Usage, Cout\_m2). Au vu de l’impact de la transformation sur les indices et une pertinence assez moindre, nous avons décidé de prendre le modèle brut. Nous allons donc procéder à correction de la multicollinéarité dans les différents modèles : le calcul des indices est très sensible à la multicollinéarité; pour preuve il ne fait pas de prévision pour 2024 car la multicollinéarité est forte.

### Pour le modèle de 2018

L’hypothèse de multicollinéarité est vérifiée pour 2018.

check\_collinearity(model\_2018)

## # Check for Multicollinearity  
##   
## Low Correlation  
##   
## Term VIF VIF 95% CI Increased SE Tolerance  
## Superficie 1.54 [1.32, 1.93] 1.24 0.65  
## Cout\_m2 3.30 [2.64, 4.24] 1.82 0.30  
## Taxe\_Jouissance 2.46 [2.00, 3.13] 1.57 0.41  
## as.factor(Type\_option) 1.75 [1.47, 2.19] 1.32 0.57  
## as.factor(Site) 3.30 [2.64, 4.24] 1.82 0.30  
## as.factor(Usage) 3.23 [2.58, 4.14] 1.80 0.31  
## Tolerance 95% CI  
## [0.52, 0.76]  
## [0.24, 0.38]  
## [0.32, 0.50]  
## [0.46, 0.68]  
## [0.24, 0.38]  
## [0.24, 0.39]

### Pour le modèle de 2019

model\_2019 <- lm(COUT ~ Superficie + Cout\_m2 + Taxe\_Jouissance + as.factor(Type\_option) +  
 as.factor(Site) ,data = parcelles\_2019)  
  
check\_collinearity(model\_2019)

## # Check for Multicollinearity  
##   
## Low Correlation  
##   
## Term VIF VIF 95% CI Increased SE Tolerance  
## Superficie 1.60 [1.32, 2.14] 1.26 0.63  
## Cout\_m2 1.47 [1.23, 1.96] 1.21 0.68  
## Taxe\_Jouissance 1.64 [1.35, 2.20] 1.28 0.61  
## as.factor(Type\_option) 1.28 [1.10, 1.74] 1.13 0.78  
## as.factor(Site) 1.08 [1.01, 2.08] 1.04 0.93  
## Tolerance 95% CI  
## [0.47, 0.76]  
## [0.51, 0.82]  
## [0.45, 0.74]  
## [0.58, 0.91]  
## [0.48, 0.99]

Après avoir supprimé la variable Usage, il n’y a plus de multilcollinéarité dans le modèle.

### Pour le modèle de 2020

model\_2020 <- lm(COUT ~ Superficie + Cout\_m2 + Taxe\_Jouissance + as.factor(Type\_option) +  
 as.factor(Site) ,data = parcelles\_2020)  
  
check\_collinearity(model\_2020)

## # Check for Multicollinearity  
##   
## Low Correlation  
##   
## Term VIF VIF 95% CI Increased SE Tolerance  
## Superficie 1.41 [1.23, 1.72] 1.19 0.71  
## Cout\_m2 2.99 [2.45, 3.73] 1.73 0.33  
## Taxe\_Jouissance 2.25 [1.88, 2.79] 1.50 0.44  
## as.factor(Type\_option) 3.93 [3.18, 4.94] 1.98 0.25  
## as.factor(Site) 1.97 [1.66, 2.42] 1.40 0.51  
## Tolerance 95% CI  
## [0.58, 0.81]  
## [0.27, 0.41]  
## [0.36, 0.53]  
## [0.20, 0.31]  
## [0.41, 0.60]

Après avoir supprimé la variable Usage qui avait une forte corrélation, il n’y a plus de forte corrélation dans le modèle.

### Pour le modèle de 2021

model\_2021 <- lm(COUT ~ Superficie + Cout\_m2 + Taxe\_Jouissance + as.factor(Type\_option) +  
 as.factor(Usage),data = parcelles\_2021)  
  
check\_collinearity(model\_2021)

## Model matrix is rank deficient. VIFs may not be sensible.

## # Check for Multicollinearity  
##   
## Low Correlation  
##   
## Term VIF VIF 95% CI Increased SE Tolerance  
## Superficie 1.88 [1.48, 2.63] 1.37 0.53  
## Cout\_m2 2.04 [1.58, 2.86] 1.43 0.49  
## Taxe\_Jouissance 4.41 [3.18, 6.33] 2.10 0.23  
## as.factor(Type\_option) 2.65 [1.99, 3.75] 1.63 0.38  
## as.factor(Usage) 3.11 [2.30, 4.42] 1.76 0.32  
## Tolerance 95% CI  
## [0.38, 0.68]  
## [0.35, 0.63]  
## [0.16, 0.31]  
## [0.27, 0.50]  
## [0.23, 0.44]

Après avoir supprimé la variable Site, il n’ya plus de forte corrélation dans le modèle.

### Pour le modèle de 2022

model\_2022 <- lm(COUT ~ Superficie + Taxe\_Jouissance + as.factor(Type\_option) +  
 as.factor(Site) + as.factor(Usage), data = parcelles\_2022)  
  
check\_collinearity(model\_2022)

## Model matrix is rank deficient. VIFs may not be sensible.

## # Check for Multicollinearity  
##   
## Low Correlation  
##   
## Term VIF VIF 95% CI Increased SE Tolerance  
## Superficie 2.73 [1.95, 4.15] 1.65 0.37  
## Taxe\_Jouissance 2.78 [1.98, 4.23] 1.67 0.36  
## as.factor(Type\_option) 1.11 [1.01, 2.46] 1.05 0.90  
## as.factor(Site) 2.49 [1.80, 3.77] 1.58 0.40  
## as.factor(Usage) 2.81 [2.00, 4.28] 1.68 0.36  
## Tolerance 95% CI  
## [0.24, 0.51]  
## [0.24, 0.51]  
## [0.41, 0.99]  
## [0.27, 0.56]  
## [0.23, 0.50]

Après avoir supprimé la variable Cout\_m2, il n’ya plus de forte corrélation dans le modèle.

### Pour le modèle de 2023

model\_2023 <- lm(COUT ~ Superficie + Cout\_m2 + as.factor(Type\_option) +  
 as.factor(Site),data = parcelles\_2023)  
  
check\_collinearity(model\_2023)

## # Check for Multicollinearity  
##   
## Low Correlation  
##   
## Term VIF VIF 95% CI Increased SE Tolerance  
## Superficie 2.07 [1.87, 2.30] 1.44 0.48  
## Cout\_m2 2.87 [2.57, 3.22] 1.69 0.35  
## as.factor(Type\_option) 2.98 [2.67, 3.34] 1.72 0.34  
## as.factor(Site) 2.77 [2.49, 3.11] 1.67 0.36  
## Tolerance 95% CI  
## [0.43, 0.53]  
## [0.31, 0.39]  
## [0.30, 0.37]  
## [0.32, 0.40]

Après avoir supprimé les variables Taxe\_Jouissance et Usage, il n’y a plus de forte corrélation dans le modèle.

### Pour le modèle de 2024

model\_2024 <- lm(COUT ~ Superficie + Cout\_m2 + as.factor(Site),data = parcelles\_2024)  
  
check\_collinearity(model\_2024)

## # Check for Multicollinearity  
##   
## Low Correlation  
##   
## Term VIF VIF 95% CI Increased SE Tolerance Tolerance 95% CI  
## Superficie 2.57 [1.89, 3.75] 1.60 0.39 [0.27, 0.53]  
## Cout\_m2 3.05 [2.21, 4.49] 1.75 0.33 [0.22, 0.45]  
## as.factor(Site) 1.32 [1.11, 1.96] 1.15 0.76 [0.51, 0.90]

Après avoir supprimé les variables Type\_Option,Taxe\_Jouissance, il n’ya plus de forte corrélation dans le modèle.

# Calcul de l’indice

Pour comparer les COUT par année, nous allons calculer les indices élémentaires d’une année à l’autre. Ainsi nous allons comparer les indices d’une année par rapport à la précédente.

# Fonction pour prédire les prix et calculer le coût moyen  
predict\_mean\_cost <- function(model, data) {  
 mean(predict(model, newdata = data))  
}  
  
# Calcul du coût moyen prédit pour chaque année  
mean\_cost\_2018 <- predict\_mean\_cost(model\_2018, parcelles\_2018)  
mean\_cost\_2019 <- predict\_mean\_cost(model\_2019, parcelles\_2019)  
mean\_cost\_2020 <- predict\_mean\_cost(model\_2020, parcelles\_2020)  
mean\_cost\_2021 <- predict\_mean\_cost(model\_2021, parcelles\_2021)

## Warning in predict.lm(model, newdata = data): prediction from a rank-deficient  
## fit may be misleading

mean\_cost\_2022 <- predict\_mean\_cost(model\_2022, parcelles\_2022)

## Warning in predict.lm(model, newdata = data): prediction from a rank-deficient  
## fit may be misleading

mean\_cost\_2023 <- predict\_mean\_cost(model\_2023, parcelles\_2023)  
mean\_cost\_2024 <- predict\_mean\_cost(model\_2024, parcelles\_2024)  
  
# Les coûts moyens dans un dataframe  
mean\_costs <- data.frame(  
 annee = 2018:2024,  
 mean\_cost = c(mean\_cost\_2018, mean\_cost\_2019, mean\_cost\_2020, mean\_cost\_2021, mean\_cost\_2022, mean\_cost\_2023, mean\_cost\_2024))  
  
# Calcul des indices élémentaires d'une année à l'autre  
mean\_costs <- mean\_costs %>%  
 mutate(  
 indice\_elementaire\_annee\_a\_annee = (mean\_cost / lag(mean\_cost)) \* 100,  
 indice\_2018\_vs\_2024 = ifelse(annee == 2024, (mean\_cost / mean\_cost\_2018) \* 100, NA))  
  
# Afficher les résultats  
print(mean\_costs)

## annee mean\_cost indice\_elementaire\_annee\_a\_annee indice\_2018\_vs\_2024  
## 1 2018 26726071 NA NA  
## 2 2019 34892793 130.55714 NA  
## 3 2020 26341607 75.49297 NA  
## 4 2021 50073660 190.09341 NA  
## 5 2022 30303111 60.51707 NA  
## 6 2023 11175059 36.87760 NA  
## 7 2024 10652586 95.32465 39.8584

Par rapport à l’année 2018, les couts des parcelles ont connu une hausse de 30% en 2019. Cela pourrait être logique dans la mesure où le cout maximal en 2019 est largement supérieur à celui de 2018. En ce qui concerne le prix des parcelles en 2020, on remarque une baisse de 24,51% par rapport à 2019. Par contre, on note une hausse de 90% des prix en 2021 par rapport à 2020 témoigné par le fait que le prix maximal en 2021 était évalué à plus de 500 millions tandis que celui de 2020 était de plus de 250 millions. En 2022, le prix maximal est passé de plus 510 millions à plus de 114 millions. Ceci se fait ressentir au niveau de l’indice qui est évalué à 60,51% En 2023, on note une baisse drastique de 73,13 % des prix des parcelles par rapport à 2022 ; se faisant ressentir au niveau des couts maximaux qui passent de plus de 114 millions en 2022 à plus de 46 millions en 2023. En 2024, on note une légère baisse de de 4,68% par rapport à 2023. Globalement, les prix des parcelles ont chuté de 60,15% entre 2018 et 2024. Le cout maximal est passé de plus de 256 millions à 41 millions en 2024.

# Conclusion

Le marché foncier de Ouagadougou est un reflet de la vitalité et de la croissance de la ville. Avec l’afflux constant de nouveaux résidents et l’intérêt croissant des investisseurs, le prix des parcelles à Ouagadougou est en constante évolution. Les facteurs tels que l’urbanisation rapide, les infrastructures améliorées, et les politiques gouvernementales jouent tous un rôle crucial dans la détermination des prix. Les quartiers centraux continuent de voir une hausse des prix en raison de leur attractivité et de leur accessibilité. Les zones périphériques, bien que plus abordables, offrent également des opportunités intéressantes pour ceux qui cherchent à investir dans des secteurs en développement.