APLICATIA 2 ECONOMETRIE

CERINTELE 1-2

1. Se va configura setul de date pentru un esantion de unitati si pentru un interval de timp, creandu-se o structura cu date de tip panel adecvata fenomenului analizat.

# 1. Instalarea și încărcarea pachetelor necesare

install.packages("readxl") # Pentru citirea fișierelor Excel

install.packages("plm") # Pentru analiza datelor de tip panel

library(readxl)

library(plm)

# 2. Importul fișierului Excel

file\_path <- "AdidasSalesdata.xlsx"

sales\_data <- read\_excel(file\_path)

# 3. Verificarea numelor coloanelor înainte de curățare

cat("Numele originale ale coloanelor:\n")

print(colnames(sales\_data))

# 4. Curățarea și standardizarea numelor coloanelor

colnames(sales\_data) <- make.names(colnames(sales\_data))

# 5. Verificarea numelor coloanelor după curățare

cat("\nNumele curățate ale coloanelor:\n")

print(colnames(sales\_data))

# 6. Conversia coloanei de dată într-un format de tip Date

sales\_data$Invoice.Date <- as.Date(sales\_data$Invoice.Date)

# 7. Configurarea setului de date ca panel

panel\_data <- pdata.frame(sales\_data, index = c("Retailer.ID", "Invoice.Date"))

# 8. Verificarea structurii datelor panel

cat("\nRezumatul datelor panel:\n")

print(summary(panel\_data))

# 9. Model de regresie pe date panel

# Regresie cu "Total Sales" ca dependentă și "Units Sold" ca explicativă

model <- plm(Total.Sales ~ Units.Sold, data = panel\_data, model = "within")

# 10. Rezultatele modelului

cat("\nRezultatele modelului de regresie:\n")

print(summary(model))

Rezultatele datelor panel:

|  |  |
| --- | --- |
| Variabilă | Rezumat |
| Retailer | Lungime: 9648, Clasa: caracter, Mod: caracter |
| Retailer ID | 1128299: 2370, 1185732: 5265, 1189833: 360, 1197831:1653 |
| Invoice Date | 2021-01-17: 77, 2021-03-16: 72, 2021-04-17: 72, 2021-08-17: 72, 2021-11-17: 71, 2021-07-16: 70, (Other) :9214 |
| Region | Lungime: 9648, Clasa: caracter, Mod: caracter |
| State | Lungime: 9648, Clasa: caracter, Mod: caracter |
| City | Lungime: 9648, Clasa: caracter, Mod: caracter |
| Gender type | Lungime: 9648, Clasa: caracter, Mod: caracter |
| Product Category | Lungime: 9648, Clasa: caracter, Mod: caracter |
| Price per Unit | Min.: 7.00, 1st Qu.: 35.00, Median: 45.00, Mean: 45.22, 3rd Qu.: 55.00, Max.: 110.00 |
| Units Sold | Min.: 0.0, 1st Qu.: 106.0, Median: 176.0, Mean: 256.9, 3rd Qu.: 350.0, Max.: 1275.0 |
| Total Sales | Min.: 0, 1st Qu.: 4254, Median: 9576, Mean: 93273, 3rd Qu.: 150000, Max.: 825000 |
| Operating Profit | Min.: 0, 1st Qu.: 1922, Median: 4371, Mean: 34425, 3rd Qu.: 52063, Max.: 390000 |
| Operating Margin | Min.: 0.100, 1st Qu.: 0.350, Median: 0.410, Mean: 0.423, 3rd Qu.: 0.490, Max.: 0.800 |
| Sales Method | Lungime: 9648, Clasa: caracter, Mod: caracter |

Rezultatele modelului de regresie:

**Unbalanced Panel:**

* n = 4
* T = 360-5265
* N = 9648

**Reziduuri:**

* Min.: -261666.9
* 1st Qu.: -30793.3
* Median: 2655.4
* 3rd Qu.: 30574.0
* Max.: 414044.6

**COEFICIENTII:**

**Units Sold**

* + Estimate: 611.1147
  + Std. Error: 2.6479
  + t-value: 230.8
  + p-value: < 2.2e-16

**Suma pătratelor totale:** 1.9257e+14

**Suma pătratelor reziduale:** 2.9518e+13

**R-Squared:** 0.84672

**R-Squared ajustat:** 0.84665

**F-statistic:** 53266.7 pe 1 și 9643 DF, p-value: < 2.22e-16

1. Sa se defineasca modelul si sa se specifice variabilele cu precizarea transformarilor aduse.

# 1. Instalarea si incarcarea pachetelor necesare

install.packages("readxl") # Pentru citirea fisierelor Excel

install.packages("plm") # Pentru analiza datelor de tip panel

library(readxl)

library(plm)

# 2. Importul fisierului Excel

file\_path <- "AdidasSalesdata.xlsx"

sales\_data <- read\_excel(file\_path)

# 3. Curatarea si standardizarea numelor coloanelor

colnames(sales\_data) <- make.names(colnames(sales\_data))

# 4. Conversia coloanei de data intr-un format de tip Date

sales\_data$Invoice.Date <- as.Date(sales\_data$Invoice.Date)

# 5. Verificarea si eliminarea duplicatelor

sales\_data <- sales\_data[!duplicated(sales\_data[, c("Retailer.ID", "Invoice.Date")]), ]

# 6. Eliminarea valorilor Total.Sales <= 0 pentru a evita probleme cu log-transformarea

sales\_data <- sales\_data[sales\_data$Total.Sales > 0, ]

# 7. Configurarea setului de date ca panel

panel\_data <- pdata.frame(sales\_data, index = c("Retailer.ID", "Invoice.Date"))

# 8. Specificarea modelului si transformarea variabilelor

# Variabila dependenta: Total.Sales

# Variabile independente: Units.Sold si Price.per.Unit

# Vom aplica log-transformarea pentru variabilele numerice pentru a surprinde elasticitati

panel\_data$log\_Total\_Sales <- log(panel\_data$Total.Sales)

panel\_data$log\_Units\_Sold <- log(panel\_data$Units.Sold + 1) # +1 pentru a evita log(0)

panel\_data$log\_Price\_per\_Unit <- log(panel\_data$Price.per.Unit)

# 9. Definirea modelului de regresie cu efecte fixe

model\_fixed <- plm(log\_Total\_Sales ~ log\_Units\_Sold + log\_Price\_per\_Unit,

data = panel\_data, model = "within")

# 10. Rezultatele modelului de regresie

cat("\nRezultatele modelului de regresie cu efecte fixe:\n")

print(summary(model\_fixed))

# 11. Testarea efectelor fixe vs. aleatorii

model\_random <- plm(log\_Total\_Sales ~ log\_Units\_Sold + log\_Price\_per\_Unit,

data = panel\_data, model = "random")

hausman\_test <- phtest(model\_fixed, model\_random)

cat("\nRezultatele testului Hausman:\n")

print(hausman\_test)

# Interpretarea:

# Daca p-valoarea testului Hausman este semnificativa (< 0.05), folosim modelul cu efecte fixe.

Rezultatele modelului de regresie cu efecte fixe:

Oneway (individual) effect Within Model

**Unbalanced Panel:**

* n=4
* T=141−469
* N=1207

**Residuals:**

* Min.: -2.16572
* 1st Qu.: -0.57732
* Median: 0.12967
* 3rd Qu.: 0.54346
* Max.: 2.50862

**COEFICIENTII:**

**log\_Units\_Sold:**

* + Estimate: 1.881643
  + Std. Error: 0.032218
  + t-value: 58.4039
  + p-value: < 2.2e-16

**log\_Price\_per\_Unit:**

* + Estimate: 0.864041
  + Std. Error: 0.088388
  + t-value: 9.7755
  + p-value: < 2.2e-16

**Suma pătratelor totale:** 3130.1

**Suma pătratelor reziduale:** 714.41

**R-Squared:** 0.77176

**R-Squared ajustat:** 0.77081

**F-statistic:** 2030.53 on 2 and 1201 DF, p-value: < 2.22e-16

**Rezultatele Testului Hausman**

**Hausman Test**

* **Data:** log\_Total\_Sales ~ log\_Units\_Sold + log\_Price\_per\_Unit
* **Chi-squared statistic (chisq):** 0.049918
* **Degrees of Freedom (df):** 2
* **p-value:** 0.9753
* **Alternative Hypothesis:** One model is inconsistent

Cerinta 3. Sa se testeze alegerea tipului de model RE sau FE cu ajutorul testelor specifice.

Deoarece p-value = 0.9753 in testul Hausman, fiind mult mai mare decat 0.05 inseamna ca nu respingem ipoteza nula H0, ce reprezinta ca modelul cu efecte aleatorii (RE) este consistent si adecvat.

Nu există o corelație semnificativă între efectele individuale neobservate și variabilele explicative. Modelul cu efecte aleatorii este preferat deoarece oferă estimări ne-biasate și este mai eficient decât modelul cu efecte fixe (în special pentru eșantioane mari).

Alegerea modelului cu efecte aleatoare (RE) înseamnă că presupunem că efectele specifice entităților nu sunt corelate cu variabilele explicative (log\_Units\_Sold și log\_Price\_per\_Unit).

Această abordare este adecvată atunci când există variabilitate între entități, dar efectele specifice nu au o relație cauzală directă cu factorii din model.

Modelul cu efecte fixe de timp este:

Cod R:

fixed\_time <- plm(log\_Total\_Sales ~ log\_Units\_Sold + log\_Price\_per\_Unit + factor(Invoice.Date),

data = panel\_data,

index = c("Invoice Date"),

model = "within")

summary(fixed\_time)

# Model de bază fără efecte fixe

pool <- plm(log\_Total\_Sales ~ log\_Units\_Sold + log\_Price\_per\_Unit,

data = panel\_data,

index = c("year"),

model = "pooling")

summary(pool)

Alte teste specifice pentru alegerea modelului RE sau FE sunt:

1. Testarea necesitatii efectelor fixe de timp:

Testul F pentru efecte fixe de timp:

pFtest(fixed\_time, pool)

Acest test a aratat ca valoarea lui p-value este 0.8351, considerabil mai mare decat pragul de 0.05, ceea ce inseamna ca dateke de timp nu sunt relevante.

1. Testul Lagrange Multiplier (Breusch-Pagan) pentru efecte de timp.

Testul verifica daca exista variatie semnificativa de-a lungul timpului

Cod R: plmtest(pool, effect = "time", type = "bp")

Aici avem nevoie tot de valoarea lui p-value pentru a cloncluziona ca datele de timp nu sunt relevante, valoarea calculata fiind 0.1114 > 0.05

1. Testarea dependentei seriale (serial correlation) – acest test indica corelatia dintre erori

Am folosit urmatorul cod R:

fixed <- plm(log\_Total\_Sales ~ log\_Units\_Sold + log\_Price\_per\_Unit,

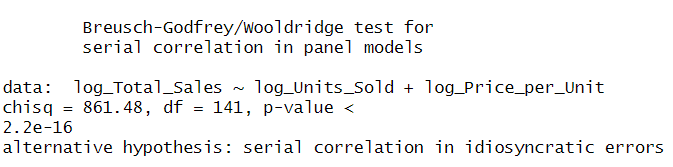
data = panel\_data,

index = c("year"),

model = "within")

pbgtest(fixed)

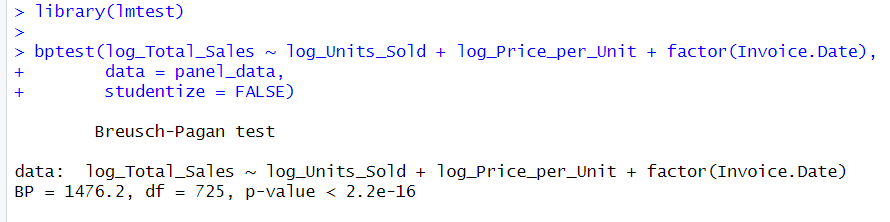
Rezultatul:



Valoarea lui p-value este considerata foarte mica, spre null ceea ce inseamna ca este mai mica de 0,05, existand autocorelatie si trebuiesc corectate standardele erorii.

1. Testarea heteroscedaticitatii:

Codul R si valorile rezultate:



Cum p-value este iar o valoare <0.05 rezulta ca avem heteroscedasticitate și trebuie să folosim erori robuste.

Corecatarea pentru heteroscedasticitate si autocorelatie:

# Corectarea erorilor standard pentru heteroscedasticitate și autocorelație

library(sandwich)

library(lmtest)

# Model final cu erori robuste

coeftest(fixed\_time, vcovHC(fixed\_time, method = "arellano"))

In final concluzionam ca alegerea finală a modelului este cel cu efecte aleatorii (RE), bazată pe rezultatele testului Hausman (p=0.9753).

Testele diagnostice indică necesitatea ajustării erorilor standard pentru autocorelare și heteroscedasticitate, lucru pe care l-am realizat prin utilizarea erorilor robuste.

Cerinta 4. Să se estimeze modelul adecvat obtinut din analiza etapei 3.

Pe baza analizei din etapa 3 am selectat modelul cu efecte aleatorii (RE) deoarece testul Hausman a aratat ca acesta este consistent si adecvat si nu exista o corelatie semnificativa intre efectele neobservate si variabilele explicative.

Modelul RE este estimat utilizand următoarea specificatie:

ln(Total Sales)=β0​+β1​ln(Units Sold)+β2​ln(Price per Unit)+ui​+ϵit​

Unde:

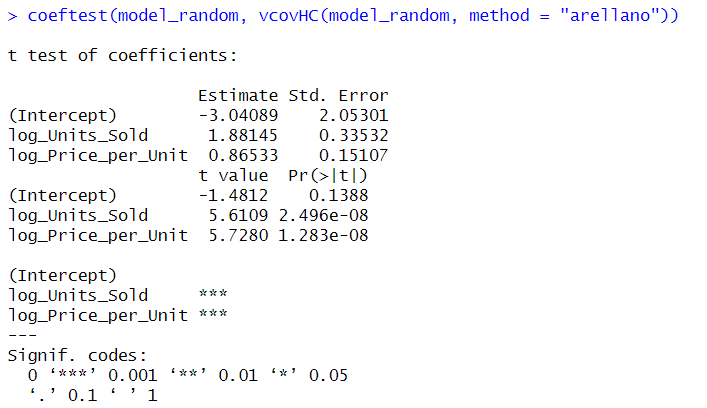
* Ln(Total Sales): logaritmul vanzarilor totale (variabila dependenta).
* Ln(Units Sold): logaritmul unitatilor vandute (variabila explicativa).
* Ln(Price per Unit): logaritmul pretului pe unitate (variabila explicativa)
* ui​: efectele aleatorii specifice entitatilor
* ϵit​: termenul de eroare

Modelul a fost estimat în R folosind biblioteca plm si specificatia pentru efecte aleatorii. In plus, erorile standard au fost corectate pentru heteroscedasticitate si autocorelare.

# Corectarea erorilor standard (erori robuste)

coeftest(model\_random, vcovHC(model\_random, method = "arellano"))

Dupa estimarea modelului RE si corectarea erorilor standard, rezultatele sunt urmatoarele:



Interceptul nu este semnificativ statistic (p-value = 1388), insemnand ca la valori egale cu 0 ale variabilelor explicative, valoarea precisa a vanzarilor nu poate fi interpretata ca relevanta.

Ln(Units Sold) = 188145 inseamna ca este extrem de semnificativ statistic, ceea ce indica faptul ca o crestere cu 1% in numarul unitatilor vandute determina, in medie, o crestere de aproximativ 1,88% in vanzarile totale, presupunand ca toate celelalte variabile raman constante.

Coeficientul pentru logaritmul pretului pe unitate este 0.86533 si este, de asemenea, extrem de semnificativ statistic (p-value =1.283×10 ^-8, marcat cu \*\*\*). Aceasta arată că o creștere de 1% a prețului pe unitate duce, în medie, la o creștere de aproximativ 0.87% în vânzările totale, ceea ce sugerează o elasticitate pozitivă față de preț.

Ambele variabile explicative sunt extrem de semnificative din punct de vedere statistic, avand valori de p-value mult sub pragul de 0.001. Acest lucru indică o relatie solida intre variabilele explicative si vanzările totale.

Elasticitatea ridicata fata de unitatile vandute (1.88) subliniaza importanta volumului vanzarilor asupra performantei totale.

Elasticitatea față de preț (0.87) sugereaza ca o creștere a pretului nu doar ca nu reduce vanzarile totale, ci chiar are un impact pozitiv. Acest lucru poate fi interpretat ca o reflectie a perceptiei calitatii asociate pretlui produsului sau a unui comportament inelastic al cererii.

Corectarea erorilor standard asigura ca testele de semnificatie statistica sunt robuste in prezenta heteroscedasticitatii si autocorelarii.