

Proiect Econometrie

Scheme logice

Echipa

Dumitriu Ana Maria
Dumitrescu Teodora
Florea Gabriela Mihaela

Profesori coordonatori

Alexandru Adriana Ana Maria
Manta Eduard Mihai

Teme analizate

1. Modele care analizeaza influenta specificatiilor masinilor electrice in pretul acestora.
2. Model Panel care analizeaza variabilitatea GDP in functie de consumul din diverse categorii de surse energetice.

Sursa datelor

www.kaggle.com

Informatii despre date

PriceEuro - prețul mașinii în euro
Accel (s) - timpul în care ajunge la 100km/h
TopSpeed (km/h) - viteza maximă
Range (km) - autonomie
Caroserie - forma mașinii (Hatchback=1, 0 restul)
Dummy: Tipul de tracțiune (1=FWD, 0 restul)
Numar de locuri (1=5 locuri, 0 restul)

Regresii

Panel

GDP (\$) - PIB real, ajustat în funcție de inflație
nuclear_consumption (TW/h) - Consumul de energie primară din energia nucleară
hydro_consumption (TW/h) - Consumul de energie primară din hidroenergie
renewables_consumption - Consumul de energie primară din surse regenerabile

Regresii

Colectarea si prelucrarea si
definirea datelor

Studiul cunoasterii
Analiza articolelor de specialitate

Regresia simpla
Testarea ipotezelor

```
rs_topspeed <- lm(PriceEuro ~ TopSpeed, price1)
```

Aplicarea corectiilor

Proгноze

```
> out_of_sample <- data.frame(Topspeed = c(200,250,270))  
> # Proгноza  
> y_pred_outsample <- predict(model_2, newdata = out_of_sample)  
> y_pred_outsample  
1 2 3  
67076.92 98795.46 111482.88
```

Regresia multipla
Testarea ipotezelor

```
model_0 <- lm(PriceEuro ~ DummySeats5+  
TopSpeed+ Range, price1)
```

Aplicarea corectiilor

Proгноze

	fit	lwr	upr
1	65060.27	60029.99	70091.24
2	97244.11	89627.94	104860.28
3	110580.26	103723.64	117456.87
4	122706.78	116745.81	128667.71
5	130904.85	125008.53	136801.17
6	147922.15	138545.37	157298.93

Model de interactiune cu variabile dummy

```
model_interactiune_dummy <- lm(PriceEuro ~ DummyBodyStyle +  
DummyFWD + DummyBodyStyleXDummyFWD, hprice1)
```

Testul Chow

Modele de regularizare
LASSO, Ridge, Elastic Net Regression

Boruta

Panel

Colectarea si prelucrarea si
definirea datelor

Studiul cunoasterii
Analiza articolelor de specialitate

Hausmann
Compararea modelelor OLS, FE, RE

Testarea ipotezelor

Concluzii

Proiect Econometrie

Scheme logice

Regresia simplă

Prețul în funcție de viteză maximă

PriceEuro = -60 570.5 + 649.47 * TopSpeed

rs_topSpeed <- lm(PriceEuro ~ TopSpeed, price1)

Bonitate: 68.73%
Model semnificativ la 99%
Coeficient semnificativ la 99%

p-val < 0.1
Homoscedasticitate
White_test bptest
p-val > 0.1
Hetero. Homo.
p-value = 0.34 0.16

p-val < 0.1
Autocorelare
dwtest bgtest
p-val > 0.1
Rez. autocorelate
Ord 1: dw = 0.59
BG - ord1 ord2 ord3
0.7 0.93 0.97

p-val < 0.1
Normalitate
jarque.bera. test
p-val > 0.1
2.2e-16
Nu sunt normal distrib.
Normal distrib.

Corecție:
elim. outlier cu
distanța Cook
0.01183 < 0.1
Nu sunt normal distrib,
limitare a datelor

Proгноза TopSpeed = 250 km/h
Set de antrenare -> 98 795 euro
Interval de încredere (90%): 98795 84798 102792

Regresia multiplă

Pretul în funcție de viteză maximă, numărul de locuri

(dummy), autonomie
PriceEuro = -49734.44 + (-9186.03) * DummySeats5 + 538.22

* TopSpeed + 45.55 * Range

model_0 <- lm(PriceEuro ~ DummySeats5 +
TopSpeed + Range, price1)

Bonitate: 70.9%
Model semnificativ la 99%
Coef. semnificativi la 90% și 99%

val > 10
Multicoliniaritate
vif
NU: val < 10
1.07 2.41 2.38

p-val < 0.05
Homoscedasticitate
White_test bptest
p-val > 0.05
Hetero. Homo. la 95%
p-value = 0.059 0.079

p-val < 0.1
Autocorelare
dwtest bgtest
p-val > 0.1
Rez. autocorelate
Ord 1: dw = 0.83
BG - ord1 ord2 ord3
0.38 0.56 0.7

p-val < 0.1
Normalitate
jarque.bera. test
p-val > 0.1
2.2e-16
Nu sunt normal distrib.
Normal distrib.

Corecție:
elim. outlier cu
distanța Cook
0.044 < 0.1
Nu sunt normal distrib,
limitare a datelor

Proгноза 5 locuri, TopSpeed = 200, Range 250
Set de antrenare -> 65 060 euro
Interval de încredere (90%): 65 060 60029 70091

Model de interacțiune cu variabile dummy

Variabile x1 = HB = hatchback

Dummy x2 = FWD = tracțiune față

hprice1 %<>% mutate(HB_FWD = DummyBodyStyle * DummyFWD,
notHB_FWD = notBodyStyle * DummyFWD,
HB_notFWD = DummyBodyStyle * notFWD)

Creare model cu cele 3 variabile

PriceEuro = 70 232 + (-31940)*x1 + (-33478)*x2 + 26681*x1*x2

model_interactiune_dummy <- lm(PriceEuro ~ DummyBodyStyle +
DummyFWD + DummyBodyStyleXDummyFWD, hprice1)

Bonitate: 38.8%
Model semnificativ la 99%
Coef. semnificativi la 99%

Rezultate

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	70232	3486	20.143	< 2e-16 ***
DummyBodyStyle	-31940	6638	-4.812	5.99e-06 ***
DummyFWD	-33478	6338	-5.284	4.29e-07 ***
DummyBodyStyleXDummyFWD	26681	10074	2.648	0.00955 **

Model de interacțiune cu variabile dummy

model_10 <- lm(PriceEuro ~ TopSpeed + Range
+ Accel + TopSpeedXHB +
RangeXHB + AccelXHB, hprice1)

SSR mod. nerestricționat

SSR mod. cu term. de interacțiune, q = 4

F-stat 0.95
(F_stat <- ((ssr_r - ssr_ur)/q)
/ (ssr_ur/df_resid))
F-critic 2.47
qf(0.95, q, df_resid)

F-stat < F-critic
Coef. nu sunt semnificativi simultan
Coef. sunt semnificativi simultan

Testul Chow - calculăm F-stat pe subseturile HB = 1 și HB = 0
model_12 <- update(model_11, subset = (DummyBodyStyle == 1))
calculăm SSR pe cele două modele

F-Chow-stat < F-critic
Coef. nu sunt semnificativi simultan
Chow F-stat 1.54 F-stat = 3.54
F-Chow-stat > F-critic
Coef. sunt semnificativi simultan

Proiect Econometrie

Scheme logice

Modele de regularizare

Pretul in functie de viteza maxima, numarul de locuri (dummy), autonomie

Cream modelele Ridge, LASSO si Elastic Net Regression si comparam rezultatele obtinute

Ridge Regression

Alpha

Lambda

R-squared

Pret

0

2640.67

87.3%

70 519 euro

```
(Intercept) ~46615.8972
DummySeats5 ~3859.0923
TopSpeed ~506.8466
Range ~35.6543
```

LASSO Regression

1

131

88.45%

72 343 euro

```
(Intercept) ~9008.9467
DummySeats5 ~1669.0442
TopSpeed ~635.5829
Range ~
```

Elastic Net Regression

0.5

150.41

88.45%

72 773 euro

```
(Intercept) ~5.82305e-04
DummySeats5 ~2.17003e+05
TopSpeed ~6.38419e-07
Range ~2.38066e-01
```

Cream setul pentru care vom realiza prognoza

```
prognoza <- data.frame(DummySeats3 = c(1,1),
TopSpeed = c(200), Range = c(300))
```

Definim predictorii

Stabilim alpha si estimam model <- glmnet(x, y, alpha = 0)
alpha=0, Ridge, 1 LASSO, 0.5 EL. Net

Realizam prognoza

```
predict(best.model, s = best.lambda,
newx = new)
```

Rescriem modelul

```
y_predicted <- predict(model,
s = best.lambda, newx = x)
```

Identificam lambda (pentru MSE minimizat)

Determinam bonitatea modelului

Comparam cele trei modele

Boruta

Identificam variabilele independente importante

```
boruta.bank_train <- Boruta(PricelEuro--,
data = hprice1, doTrace = 2)
```

Afiam variabilele importante si cream modele

```
getSelectedAttributes(boruta.bank_train, withtentative = T)
#(1) "Brand" "Accel" "TopSpeed" "Range" "Efficiency" "PowerTrain" "Segment"
#(8) "Seats" "PriceRat" "Year" "Year1" "Year2" "Year3" "Year4" "Year5"
```

Modelul realizat cu toate variabilele importante:

```
model_boruta1 <- lm(PriceEuro ~ Brand + Accel + TopSpeed + Range + Efficiency
+ PowerTrain + Segment + Seats, hprice1) # R-squared = 98.82%
```

Model realizat cu variabilele numerice importante:

```
model_boruta2 <- lm(PriceEuro ~ Accel + TopSpeed + Efficiency + Range, hprice1) #R-squared=90.76%
```

Model realizat cu var. numerice si cu var. dummy create folosind var. calitative importante

```
model_boruta3 <- lm(PriceEuro ~ Accel + TopSpeed + Efficiency + Range +
DummySegmentR + DummySeatsS, hprice1) # Bonitatea modelului = 91.17%
```

Analizam coeficientii

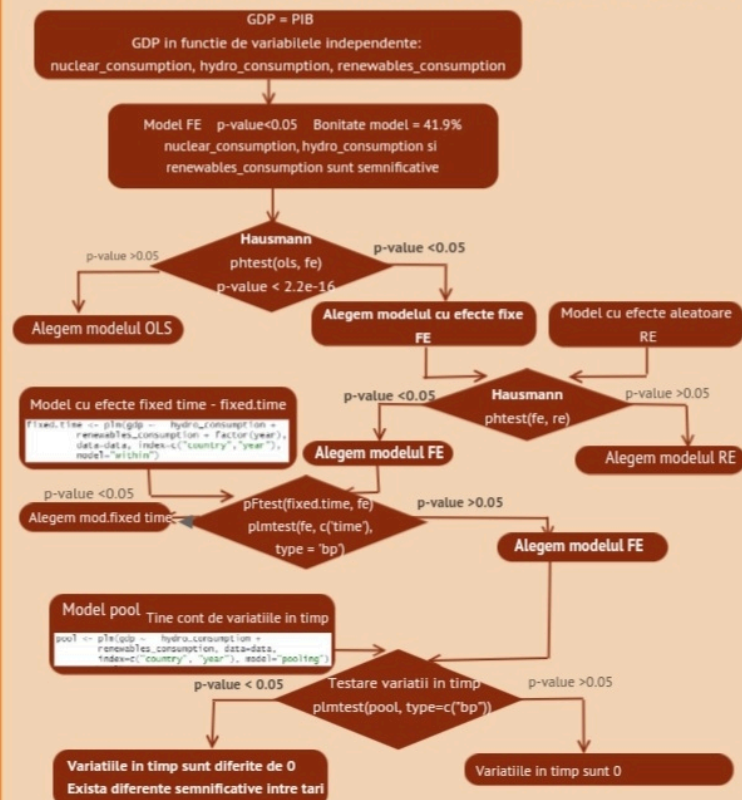
```
Coefficients:
(Intercept) Estimate Std. Error t value Pr(>|t|)
Accel 1.057e+05 1.261e+04 8.384 8.17e-13 ***
TopSpeed 6.307e+02 4.715e+01 13.381 < 2e-16 ***
Efficiency 2.072e+02 4.073e+01 5.083 2.10e-06 ***
Range 6.171e+00 1.453e+01 0.427 0.6704
DummySegmentR 3.252e+03 2.397e+03 1.399 0.1655
DummySeatsS 2.922e+05 2.114e+03 1.365 0.1763
```

Proiect econometrie

Scheme logice

Panel

GDP in functie de consumul de energie nucleara, hidroenergie si energie din resurse regenerabile



Testare

