HD Dataanalyse, efterår 2018 Opgaveark 9



Baggrund og databeskrivelse (opgave 1-3)

Denne opgave beskæftiger sig med udbudsprisen på brugte biler. Opgaven tager udgangspunkt i datamaterialet i filen Biltorvet.jmp, hvis indhold er beskrevet på opgaveark 1.

Opgave 1 (brugtvognsprisen)

a). Angiv de estimerede parametre i en lineær regressionsmodel med brugtvognsprisen (Pris (kr) ) som responsvariabel og bilens kørte kilometer (Km kørt) som forklarende

variabel.

JMP-vink: “Analyze” -> “Fit Model”. Placér Pris (kr) under “Y”, Km kørt under “Construct Model Effects”.



Modellen bliver:

Angiv modellens forklaringsgrad *R*2.



36% af variationen i Y-variablen Pris kan forklares ved variationen i X-variablen Km kørt

b). Angiv forklaringsgraderne for nedenstående tre lineære regressionsmodeller, der alle har brugtvognsprisen som forklarede variabel. Giv desuden mulig intuition bag forskellene i forklaringsgraden i de tre modeller ved at se på scatterplot af data:

1. Bilens kørte kilometer (Km kørt) som forklarende variabel, men hvor modellen kun estimeres for biler af mærket Citroén (Producent).

JMP-vink: Begræns datamaterialet v.hj.a. “Rows” -> “Data Filter”.





Godt fit ikke megen variation

1. Bilens kørte kilometer (Km kørt) som forklarende variabel, men hvor modellen kun estimeres for biler af mærket VW (Producent).





VW Få observationer med en del variation, forkert angivne priser

1. Bilens indregistreringsmåned (Indregistrering (måned) ) som forklarende variabel (for alle producenter).





Vi tester:

Vi kan ikke forkaste nulhypotesen, vi kan ikke afvise at hældningskoefficienten er 0.



HD Dataanalyse, efterår 2018 Opgaveark 9



Opgave 2 (brugtvognsprisen)

Opgave 2 omhandler udelukkende biler af mærket Kia.

JMP-vink: “Rows” -> “Data Filter”

a). Angiv de estimerede parametre i en lineær regressionsmodel med brugtvognsprisen (Pris (kr) ) som responsvariabel og bilens kørte kilometer (Km kørt) som forklarende

variabel.







JMP-vink: “Analyze” -> “Fit Model”. Placér Pris (kr) under “Y”, Km kørt under “Construct Model Eﬀects”.

Giv en fortolkning af de estimerede parametre.

Ligningen bliver



En fabriksny Kia koster 185312 ifølge modellen

For hver km den kører falder prisen med 93 øre

b). Beregn et 95%-konfidensinterval for og giv en fortolkning af intervallet.

JMP-vink: “Regression Reports” -> “Show All Confidence Intervals” via den røde trekant øverst i JMPs

regressions output-vindue.



Vi kan med 95% sikkerhed sige at en fabriksny Kia koster mellem 171178 kr. og 199445 kr.

c). Beregn et 95%-konfidensinterval for betydningen for den forventede brugtvognspris af en stigning på 1 kørt kilometer og giv en fortolkning af intervallet.

JMP-vink: “Regression Reports” -> “Show All Confidence Intervals” via den røde trekant øverst i JMPs

regressions output-vindue.

Vi kan med 95% sikkerhed sige at en prisen på en Kia falder mellem 0.7 kr. og 1.16 kr. pr kørt km

Beregn et 95%-konfidensinterval for betydningen for den forventede brugtvognspris af en stigning på 1.000 kørte kilometer og giv en fortolkning af intervallet.

Vi kan med 95% sikkerhed sige at en prisen på en Kia falder mellem 700 kr. og 1160 kr. pr kørte 1000 km

d). Beregn på baggrund af den estimerede regressionsmodel den forventede brugtvognspris for en bil af mærket Kia, der har kørt 50.000 km.

JMP-vink: “Factor Profiling” -> “Profiler” via den røde trekant øverst i JMPs regressions output-vindue.



Den forventede brugtvognspris for en bil af mærket Kia, der har kørt 50.000 km er 138763 kr.

Beregn på baggrund af den estimerede regressionsmodel den forventede brugtvognspris for en bil af mærket Kia, der har kørt 120.000 km.



Den forventede brugtvognspris for en bil af mærket Kia, der har kørt 120.000 km er 73595 kr.

Beregn på baggrund af den estimerede regressionsmodel den forventede brugtvognspris for en bil af mærket Kia, der har kørt 300.000 km.



Den forventede brugtvognspris for en bil af mærket Kia, der har kørt 300.000 km er -93979 kr.



2/4

HD Dataanalyse, efterår 2018 Opgaveark 9



Opgave 3 (kørte kilometer)

a). Angiv for henholdsvis diesel- og benzinbiler (Drivmiddel) de estimerede parametre i en lineær regressionsmodel med bilens kørte kilometer (Km kørt) som responsvariabel og bilens årgang (Årgang) som forklarende variabel.

JMP-vink: “Analyze” -> “Fit Model”. Placér Km kørt under “Y”, Årgang under “Construct Model Eﬀects”,

Drivmiddel under “By”.

Vi fjerner outliers dvs. biler hvor årgang

Benzin:



Diesel:



Sammenlign parametrene i de to estimerede modeller.

Intercept svarer til biler år 0 så dette estimat giver det ikke mening at tale om

Vi ser at benzinbiler kører 13575 km om året ifølge modellen.

Vi ser at dieselbiler kører 19469 km om året ifølge modellen.

b). Gør rede for om der er statistisk belæg for en nulhypotese om, at bilens årgang ikke har nogen betydning for det forventede antal kørte kilometer for brugte benzinbiler (brug = 5% som signifikansniveau). Angiv i den forbindelse hypoteser samt teknisk og let forståelig konklusion.

p-værdien er mindre end 0.0001 hvilket klart er mindre end signifikansniveauet 0.05, vi forkaster derfor nulhypotesen. Bilens årgang har altså betydning for det forventede antal kørte kilometer for brugte benzinbiler

c). Gør rede for om der er statistisk belæg for en nulhypotese om, at for hvert år bilen er nyere (dvs. for hver ekstra 1 af variablen Årgang) reduceres det forventede antal kørte kilometer for brugte benzinbiler med 15.000 km (brug = 5% som signifikansniveau). Angiv i den forbindelse hypoteser samt teknisk og let forståelig konklusion.

JMP-vink: “Regression Reports” -> “Show All Confidence Intervals” via den røde trekant øverst i JMPs

regressions output-vindue.



Vi ser -15000 IKKE er indholdt i 95% KI, det betyder vi forkaster nulhypotesen.

Der er IKKE statistisk belæg for en nulhypotese om, at for hvert år bilen er nyere (dvs. for hver ekstra 1 af variablen Årgang) reduceres det forventede antal kørte kilometer for brugte benzinbiler med 15.000 km

d). Gør rede for om der er statistisk belæg for en nulhypotese om, at for hvert år bilen er nyere (dvs. for hver ekstra 1 af variablen Årgang) reduceres det forventede antal kørte kilometer for brugte benzinbiler med 13.000 km (brug = 5% som signifikansniveau). Angiv i den forbindelse hypoteser samt teknisk og let forståelig konklusion.



Vi ser -13000 er indholdt i 95% KI, det betyder vi IKKE forkaster nulhypotesen.

Der er derfor statistisk belæg for en nulhypotese om, at for hvert år bilen er nyere (dvs. for hver ekstra 1 af variablen Årgang) reduceres det forventede antal kørte kilometer for brugte benzinbiler med 13.000 km



3/4

HD Dataanalyse, efterår 2018 Opgaveark 9



Kortfattede talløsninger

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Opgave 1 | | |  |  |  |  |  |  |  |  |
| a): | ˆ |  |  | ˆ |  | ˆ = 74*:*794, | | *R* | 2 | = 0*;*3632 |
| 0 = 183*:*364, | | | 1 = 0*;*6787, | |  |
| b): | i): *R*2 = 0*;*6232, | | | ii): *R*2 = 0*;*1656, iii): *R*2 = 0*;*0003 | | | | | | |
| Opgave 2 | | |  |  |  |  |  |  |  |  |
| a): | ˆ |  |  | ˆ |  | ˆ = 67*:*155 | |  |  |  |
| 0 = 185*:*312, | | | 1 = 0*;*9310, | |  |  |  |
| b): [171*:*178; 199*:*445] | | | | |  |  |  |  |  |  |
| c): [ 1*;*1607; 0*;*7013] (1 km), | | | | | [ 1*:*160*;*7; 701*;*3] (1.000 km) | | | | | |
| d): 138*:*763 kr. (50.000 km), | | | | | 73*:*595 kr. (120.000 km), 93*:*979 kr. (300.000 km) | | | | | |
| Opgave 3 | | |  |  |  |  |  |  |  |  |
|  |  | ˆ | = 27*:*373*:*356, | | ˆ | = 13*:*575, | | ˆ = 43*:*238 | | |
| a): Benzin : 0 | | | 1 |
|  |  | ˆ | = 39*:*261*:*808, | | ˆ | = 19*:*469, | | ˆ = 44*:*818 | | |
|  | Diesel : 0 | | 1 |
| b): *H*0 : 1 = 0, *Ha* : 1 , 0. Nulhypotesen forkastes | | | | | | | | | | |
| c): *H*0 : 1 = 15*:*000, *Ha* : 1 , 15*:*000. | | | | | | | Nulhypotesen forkastes | | | |
| d): *H*0 : 1 = 13*:*000, *Ha* : 1 , 13*:*000. | | | | | | | Nulhypotesen forkastes ikke | | | |



4/4