HD Dataanalyse, efterår 2018 Opgaveark 10



Baggrund og databeskrivelse (opgave 1-5)

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

Opgave 1 (brugtvognsprisen)

Opgave 1 omhandler udelukkende biler af mærket Peugeot.

JMP-vink: “Rows” -> “Data Filter”

a). Tegn en figur der viser, hvordan brugtvognsprisen (Pris (kr) ) afhænger af bilens år-gang (Årgang).

JMP-vink: “Graph” -> “Graph Builder”

Gør rede for på baggrund af figuren, om der ser ud til at være en lineær eller en kvadratisk sammenhæng mellem brugtvognsprisen og bilens årgang.



b). Opskriv en lineær regressionsmodel med brugtvognsprisen (Pris (kr) ) som responsva-riabel og en kvadratisk effekt af bilens årgang (Årgang), dvs. en regressionsmodel med bilens årgang og bilens årgang opløftet i 2. potens (dvs. Årgang2) som forklarende variable.

Angiv de estimerede parametre i modellen.

JMP-vink: “Analyze” -> “Fit Model”. Placér Pris (kr) under “Y”. Fjern “Center Polynomials” under “Model

Specification” (klik på den røde trekant). Markér Årgang under “Model Specification” og vælg “Macros” -> “Polynomial to Degree” under “Construct Model Eﬀects”.





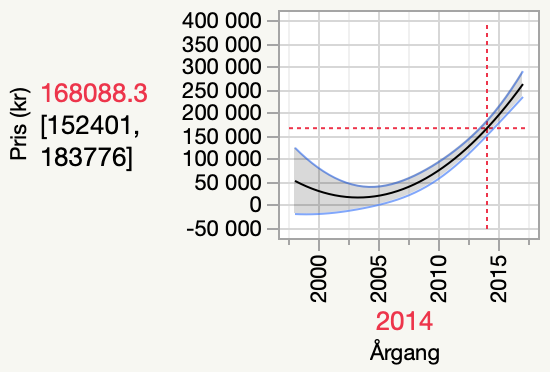
c). Beregn på baggrund af den i delspg. b) estimerede model hvor meget den forventede brugtvognspris ændres, afhængig af om bilen er årgang 2014 eller årgang 2016.

JMP-vink: Beregn for hvert af de to tilfælde den forventede brugtsvognspris v.hj.a. “Factor Profiling” ->

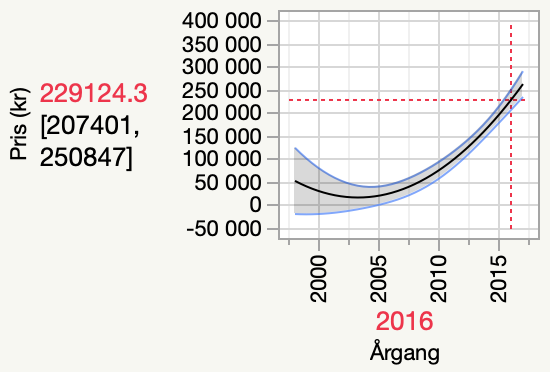
“Profiler” via den røde trekant øverst i JMPs regressions output-vindue.



**Prediction Profiler**



**Prediction Profiler**



Årgang 2014: 168.088 kr., årgang 2016: 229.124 kr. forskel: 61.036 kr.

1/7

HD Dataanalyse, efterår 2018 Opgaveark 10



Opgave 2 (brugtvognsprisen)

Opgave 2 omhandler udelukkende biler af mærket Skoda.

JMP-vink: “Rows” -> “Data Filter”

a). Reducér en lineær regressionsmodel med brugtvognsprisen (Pris (kr) ) som responsvariabel og variablene Årgang, Indregistreringsår, Km kørt, Cylindre, Airbags, Heste-kræfter, Tankkapacitet (l) , Totalvægt (kg) , Leveringsomkostninger (kr) som forklarende variable til kun at indeholde signifikante variable (brug 5% som signifikans-niveau).

JMP-vink: “Analyze” -> “Fit Model”. Placér Pris (kr) under “Y” og alle øvrige variable under “Construct Model Eﬀects”. Fjern insignifikante variable én ad gangen ved hele tiden at fjerne den mest insignifi-kante variabel. Brug f.eks. “’Eﬀect Summary” i JMPs regressions output-vindue.

Opskriv den reducerede model.

Angiv de estimerede parametre i den reducerede model.



Reduceretmodel: Pris(kr)=+·Totalvægt(kg)+·Hestekræfter+·Årgang+ · ·Indregistreringsår + · Km kørt +

= −8.349.210, = 88.271, = −84.326, = −0,3566, = −1.125, = 374, 58

b). Fortolk parametrene i den reducerede regressionsmodel.

Interceptet svarer til at alle øvrige værdier er lig med 0. Det betyder en Skoda med 0 hestekræfter der vejer 0 kg og har slet ikke har kørt, er indregistreret i år 0 og stammer fra år 0 ifølge modellen koster -8.349.219 kr. Man bliver altså mangemillionær af at hente bilen som jo ikke findes Extrapolation.

Når totalvægten på en Skoda stiger med 1 kilo stiger prisen med 375 kr

Når en Skoda har en hestekraft mere falder prisen med 1125 kr

Når en Skoda er indregistreret 1 år senere stiger prisen med 84.326 kr

For hver kørt km falder prisen på en Skoda med 36 øre

Når en Skoda er 1 år nyere stiger prisen med 88.271 kr

c). Beregn ændringen i forklaringsgraden mellem den oprindelige og den reducerede regressionsmodel.

Vink: Træk de to tal fra hinanden.

Reducerede model fra a)





Oprindelig model : = 0,9206, Reduceret model : = 0,8860, Forskel = 0,0346,

Oprindelig model : = 0,8785 Reduceret model : = 0,8589 Forskel = 0,0196

Beregn ændringen i den justerede forklaringsgrad mellem den oprindelige og den reducerede regressionsmodel.

Gør rede for eventuelle ligheder og forskelle mellem ændringerne i henholdsvis for-klaringsgrad og justeret forklaringsgrad mellem den oprindelige og den reducerede regressionsmodel.

Jo flere variable desto større værdi af , derfor kan vi benytte i multipel lineær regression, værdien giver en vis penalty for at tilføje yderligere variable, der ikke tilfører yderligere information til modellen. Ofte ville vi se at når vi reducerer en model stiger , dette sker ikke her. Det kan skyldes at vi kun har 27 observationer (for få beviser mod nulhypotesen) hvilket påvirker p-værdierne, havde vi et større datasæt ville måske se mindre p-værdier, således at nogle af de fjernede variable var signifikante.



HD Dataanalyse, efterår 2018 Opgaveark 10



Opgave 3 (brugtvognsprisen)

Opgave 3 omhandler udelukkende biler af mærket Peugeot.

JMP-vink: “Rows” -> “Data Filter”

a). Opskriv en lineær regressionsmodel med brugtvognsprisen (Pris (kr) ) som responsva-riabel og bilens årgang (Årgang) og drivmiddel (Drivmiddel) som forklarende variable.

Angiv de estimerede parametre i modellen.

JMP-vink: “Analyze” -> “Fit Model”. Vælg “Estimates” -> “Indicator Parameterization Estimates“ i JMPs

regressions output-vindue.





b). Fortolk parametrene i den estimerede regressionsmodel.

En meeeeget gammel Peugeot diesel (basis) fra år 0 (da Jesus blev født) får man 31 842 445 kr for at købe (extrapolation vi kan ikke regne med dette resultat, der fandtes ikke biler dengang)

For hver år nyere dieselbilen, bliver stiger prisen med 15885 kr

Er benzindrevet Peugeot er 29349 kr billigere ifølge modellen.

c). Tegn en figur der viser, om der ser ud til at være en forskel på årgangens betydning for brugtvognsprisen for henholdsvis benzin- og dieselbiler.

JMP-vink: “Graph” -> “Graph Builder”. Placér Pris (kr) under “Y”, Årgang under “X”, Drivmiddel under

“Group Y”.



Vi ser ældre dieselbiler er mindre værd, dette kan skyldes de kører flere km om året.

d). Opskriv en udvidelse af regressionsmodellen fra delspg. a), der også indeholder en interaktionseﬀekt mellem bilens årgang og bilens drivmiddel.

Angiv de estimerede parametre i modellen.

JMP-vink: “Analyze” -> “Fit Model”. Placér Pris (kr) under “Y”. Placér Årgang, Drivmiddel under “Construct Model Eﬀects”. Fjern “Center Polynomials” under “Model Specification” (klik på den røde trekant). Markér Årgang og Drivmiddel under “Model Specification” og vælg “Cross” under “Construct Model Effects”. Vælg “Estimates” -> “Indicator Parameterization Estimates“ i JMPs regressions output-vindue.



Model: Pris (kr) =

e). Fortolk parametrene i den estimerede udvidede regressionsmodel.

Vi har en dummy variabel, der kan være tændt eller slukket, når den er slukket (med et fag-ord er dummy variablen i basis) er der tale om en dieselbil. Det betyder i alle fortolkninger af koefficienter (undtagen når vi taler om drivmiddel[Benzin] dummy ) vil vi tale om en dieselbil.

En Dieselbil fra år 0 koster ifølge modellen -44.238.092 kr (man får mange millioner for at afhente den)

For hvert år nyere en dieselbil er stiger prisen med 22.064 kr

Er der tale om en benzinbil til salg i år 0 stiger prisen med 25.080.599 kr, dvs en benzinbil i år 0 koster så kun -44.238.092 +25.080.599 = -19.157.493 kr (man får mange millioner for at afhente den, men ikke så mange som hvis det var en dieselbil)

For hvert år nyere en benzinbil er falder prisen med 12.490 kr i forhold til en et år nyere dieselbil.

Samlet stiger prisen på en benzinbil med 22.064-12490=9574 kr for hvert år nyere bilen er.

Vore problemer med tolkningerne af interceptet og de billigere ældre dieselbiler, skyldes at vi ikke har ændret årstal til en variabel alder og har taget højde for eventuelt flere kørte km i dieselbiler

f). Opskriv nulhypotesen om at årgangens betydning for brugtvognsprisen er den samme for henholdsvis benzin- og dieselbiler.

Gør rede for om der er statistisk belæg for en påstand om at årgangens betydning for brugtvognsprisen er den samme for henholdsvis benzin- og dieselbiler (brug 5% som signifikansniveau).

Vink: Angiv P-værdi og konkludér på baggrund af den.



Vi ser på p-værdien <0.0001 for F-testet Årgang\*Drivmiddel.

Vi forkaster klart nuhypotesen da p-værdien <0.0001 er mindre end 5% signifikansniveauet.

Årgangens betydning for brugtvognsprisen er altså ikke den samme for henholdsvis benzin- og dieselbiler



HD Dataanalyse, efterår 2018 Opgaveark 10



Opgave 4 (brugtvognsprisen)

Opgave 4 omhandler udelukkende biler af mærket Peugeot.

JMP-vink: “Rows” -> “Data Filter”

a). Opskriv en lineær regressionsmodel med brugtvognsprisen (Pris (kr) ) som responsvariabel og bilens årgang (Årgang) og bodytype (Bodytype) som forklarende variable. Angiv de estimerede parametre i modellen.

JMP-vink: “Analyze” -> “Fit Model”.



Model: Pris (kr) =

b). Fortolk parametrene i den estimerede regressionsmodel.

Basis for Bodytypedummies er Van, så udgangspunktet er en Van.

En Van i år 0 koster -31,88 mio. kr.

Prisen for en Van stiger med 15897 kr. når den er et år nyere.

Er der tale om en Cabriolet stiger prisen med 105.563

Er der tale om en Hatchback falder prisen med 14.505

Er der tale om en MPV stiger prisen med 53.140

Er der tale om en Sedan stiger prisen med 32.616

Er der tale om en Stationcar stiger prisen med 53.280

c). Opskriv nulhypotesen om at bodytype ikke har nogen betydning for brugtvognsprisen. Gør rede for at der ikke er statistisk belæg for en påstand om at bodytype ikke har nogen betydning for brugtvognsprisen (brug 5% som signifikansniveau).

Vink: Angiv P-værdi og konkludér på baggrund af den.



Teknisk konklusion:

Da p-værdien <0.0001 er klart mindre end 0.05 signifikansniveauet forkaster vi nulhypotesen.

Let konklusion:

Der er således statistisk belæg for en påstand om at bodytype har betydning for brugtvognsprisen

d). Gør rede for om der er belæg for en påstand om, at der ikke er forskel i den forventede

brugtvognspris for henholdsvis...

1. en Van og en Hatchback
2. en Van og en MPV
3. en Van og en Stationcar
4. en MPV og en Stationcar

Brug i alle fire tilfælde 5% som signifikansniveau og såfremt der ikke er belæg for påstanden angiv da den forventede forskel i brugtvognsprisen.

JMP-vink: Vælg “Estimates” -> “Indicator Parameterization Estimates“ i JMPs regressions output-vindue.

Angiv for hvert tilfælde en P-værdi og konkludér på baggrund af den. JMP-vink vedr. tilfælde iv): Markér Bodytype i datafilen. Højreklik på variablen og vælg “Column Properties” -> “Value Ordering”. Flyt værdien “MPV” ned som den nederste ved hjælp af “Move Down”. Klik “OK” og estimér regressionsmodellen igen.



i)



P-Værdi=0,5359 dvs. større end 0.05 signifikansniveauet.

Nulhypotesen om ingen forskel mellem Van og Hatchback forkastes ikke. Der er ikke forskel på Van og Hatchback

ii)

P-værdi = 0,0422 dvs. mindre end 0.05 signifikansniveauet.

Nulhypotesen om ingen forskel mellem Van og MPV forkastes. En MPV forventes at være 53.140 kr. dyrere end en Van.

iii)

P-værdi = 0,0009 dvs. klart mindre end 0.05 signifikansniveauet.

Nulhypotesen om ingen forskel mellem Van og Stationcar forkastes. En Stationcar forventes at være 83.487 kr. dyrere end en Van.



iv)

P-værdi = 0,0902 dvs. større end 0.05 signifikansniveauet.

Nulhypotesen om ingen forskel mellem MPV og Stationcar forkastes ikke.

Der er ikke forskel mellem MPV og Stationcar.

HD Dataanalyse, efterår 2018 Opgaveark 10



Opgave 5 (brugtvognsprisen)

Opgave 5 omhandler udelukkende biler med et ikke-negativt antal kørte kilometer, dvs. Km kørt ≥ 0.

JMP-vink: “Rows” -> “Data Filter”

a). Reducér en lineær regressionsmodel med brugtvognsprisen (Pris (kr) ) som responsva-riabel og variablene Producent, Årgang, Bodytype, Drivmiddel, Km kørt, Cylindre, Døre, Sæder, Gear, Airbags, Hestekræfter, Tophastighed (km/t) , Tankkapacitet

(l) , Køreklar vægt (kg) , Anhængervægt (kg) , Totalvægt (kg) , Årlig grøn ejerafgift, Energiklasse, Leveringsomkostninger (kr) som forklarende variable til kun at inde-holde signifikante variable (brug 5% som signifikansniveau).

JMP-vink: “Analyze” -> “Fit Model”. Placér Pris (kr) under “Y” og alle øvrige variable under “Construct Model Eﬀects”. Fjern insignifikante variable én ad gangen ved hele tiden at fjerne den mest insignifi-kante variabel. Brug f.eks. “’Eﬀect Summary” i JMPs regressions output-vindue.



a): Variable i reduceret model:

Km kørt, Bodytype, Årlig grøn ejerafgift, Totalvægt (kg), Hestekræfter, Gear, Drivmiddel, Airbags, Årgang, Producent, Energiklasse, Leveringsomkostninger (kr)

b). Reducér samme regressionsmodel som i delspg. a) men brug i stedet 1% som signifikansniveau).

JMP-vink: Vælg signifikansniveau til 1% i “Analyze” -> “Fit Model” under “Model Selection” -> “Set Alp-

ha level” inden modellen estimeres.



b): Variable i reduceret model 1% signifikansniveau:

Km kørt, Bodytype, Årlig grøn ejerafgift, Totalvægt (kg), Hestekræfter, Gear, Drivmiddel, Airbags, Årgang, Producent, Energiklasse

Sammenlign den reducerede model med den reducerede model i delspg. a).

Leveringsomkostninger (kr) fjernes i b) da vi tester på 1% signifikansniveau. ændres ikke nævneværdigt til 81.26%

c). Beregn på baggrund af den i delspg. a) fundne regressionsmodel et 95%-konfidensinterval for den forventede betydning for brugtvognsprisen af en stigning i antal kørte kilometer på 10.000 km. 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 prisen for en km kørt ekstra ligger mellem 34 og 47 øre derfor kan vi med 95% sikkerhed sige at prisen for 10000 km kørt ekstra ligger mellem 3383 og 4737 kr.

Vi ganger bare med 10000 for at enhederne passer.

d). Beregn på baggrund af den i delspg. a) fundne regressionsmodel den forventede brugtvognspris for en 5-dørs Kia Hatchback årgang 2015, der har en 4-cylinders 110 hk benzinmotor, en tophastighed på 180 km/t, 5 gear, 5 døre, 5 sæder, 2 airbags, en 45 liters tank, en køreklar vægt på 1.300 kg, en anhængervægt på 500 kg, en totalvægt på 1.700 kg, er energiklasse B, med en årlig grøn ejerafgift på 1.500 kr., leveringsom-kostninger på 3.500 kr. og som 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 5-dørs Kia Hatchback årgang 2015, der har en 4-cylinders 110 hk benzinmotor, en tophastighed på 180 km/t, 5 gear, 5 døre, 5 sæder, 2 airbags, en 45 liters tank, en køreklar vægt på 1.300 kg, en anhængervægt på 500 kg, en totalvægt på 1.700 kg, er energiklasse B, med en årlig grøn ejerafgift på 1.500 kr., leveringsomkostninger på 3.500 kr. og som har kørt 50.000 km. bliver 149.957 kr.