

Copenhagen Business School  
Dataanalyse, HD 1. del

# EKSAMENSOPGAVE

SYGE-/OMPRØVE, EFTERÅR 2017

4 timers skriftlig eksamen

Opgaven består af i alt **32** delspørgsmål fordelt på **7** opgaver. Ved bedømmelsen vægtes alle opgavens delspørgsmål ligeligt med undtagelse af delspørgsmål 7 a), der vægtes dobbelt. Vægtene skal alene opfattes som vejledende med henblik på disponering af tiden ved besvarelsen af eksamensopgaven.

NB:

- Hvis værktøjslinjen i JMP ikke fremgår af programmet, kan den genskabes v.hj.a. <Ctrl>-<K> -> "Windows Specific" -> "Auto-hide menu and toolbars" -> "never"
- Husk at medtage relevant JMP-output i opgavebesvarelsen (vedlæg ikke alt output men kun de dele, der refereres til i opgavebesvarelsen)
- Husk at JMP-output kun tæller med i bedømmelsen, såfremt det er eksplicit kommenteret/refereret.



## SALGET AF EL-PÆRER I DANMARK

### BAGGRUND OG DATABESKRIVELSE

Salget af el-pærer i den danske detailhandel er fordelt på to produkttyper, sparepærer og almindelige pærer. I denne opgave vil vi se på, hvorledes det er muligt at bruge historisk salgsdata til at få et mere præcist billede af de vigtigste markedstendenser indenfor detailsalget af el-pærer i Danmark.

Datamaterialet i JMP-filen *Elpærer.jmp* indeholder 3 års 4-ugentlige observationer af salget af el-pærer, enten sparepærer eller almindelige pærer, i henholdsvis discountbutikker og supermarkeder. For nemheds skyld omtales observationerne igennem hele opgaven som “månedlige observationer”, selv om der reelt “kun” er tale om observationer over 4 uger (28 dage).<sup>1</sup>

Datamaterialet indeholder information om den solgte mængde el-pærer og om den gennemsnitlige salgspris per pære. Derudover indeholder datamaterialet information om to variable, som kunne tænkes at have en sammenhæng med salget af el-pærer, nemlig dagens gennemsnitlige længde i den givne periode (målt som tiden fra solopgang til solnedgang) samt den gennemsnitlige pris per kiloWatt-time elforbrug.

Eksamensopgaven skal besvares på baggrund af datamaterialet i *Elpærer.jmp*, hvis indhold er nærmere beskrevet i tabellen på næste side.

---

<sup>1</sup>Der er én enkelt periode, som kun dækker 21 dage. Alle øvrige perioder omfatter 28 dage.



VARIABELNAVN	FORKLARING
<i>Generelt</i>	
<i>Dato</i>	Sidste dag i den betragtede 4 ugers periode
<i>År</i>	Årstal for den sidste dag i den betragtede 4 ugers periode
<i>Årstid</i>	Årstid for den betragtede 4 ugers periode (med de mulige værdier "Forår", "Sommer", "Efterår", "Vinter")
<i>Butikstype</i>	Hvilken butikstype omsætningstallene stammer fra (med de mulige værdier "Discount", "Supermarkeder")
<i>Pæretype</i>	Hvilken type el-pærer omsætningstallene stammer fra (med de mulige værdier "Spare" og "Almindelig")
<i>Omsætningsinformation</i>	
<i>Stk</i>	Antal solgte el-pærer (i stk) over den betragtede periode
<i>Pris</i>	Pris (i kr.) på én solgt el-pære (gennemsnit over den betragtede periode)
<i>Øvrig information</i>	
<i>Dagslængde</i>	Dagslængde (i antal minutter) i København målt fra solopgang til solnedgang (gennemsnit over den betragtede periode)
<i>Elpris</i>	Samlet elpris inkl. moms (i øre) per kiloWatt-time (kWh) for forbrugere og små virksomheder (gennemsnit over den betragtede periode)



### OPGAVE 1 (VOLUMEN)

For at få et overblik over salget af el-pærer i den danske dagligvarehandel ser vi på den omsatte mængde (variablen *Stk*) af hhv. almindelige og sparepærer (variablen *Pæretype*) i hhv. supermarkeder og discountbutikker (variablen *Butikstype*).

- a). **Beregn** middelværdien for det månedlige salg (i stk) af hhv. almindelige og sparepærer i hhv. supermarkeder og discountbutikker.

*Vink: Sæt begge de to variable Butikstype og Pæretype under "By" i JMPs "Distribution"-funktion.*

- b). **Gør rede for** om der i gennemsnit sælges flest sparepærer i supermarkeder eller i discountbutikker.

**Gør rede for** om der i supermarkeder i gennemsnit sælges flest almindelige pærer eller flest sparepærer.

I de resterende delspørgsmål i denne opgave ser vi udelukkende på salget (i stk) af *almindelige* pærer i hhv. supermarkeder og discountbutikker.

- c). **Tegn** et histogram og et boxplot over salget (i stk) af almindelige pærer i discountbutikker.

**Giv** en kort beskrivelse af fordelingen på baggrund af histogram, boxplot og relevante nøgletal for centrum og variation.

- d). **Angiv og fortolk** medianen i fordelingen af salget (i stk) af almindelige pærer i discountbutikker.

**Gør rede for** om median eller middelværdi (eller begge) er retvisende mål for centrum i fordelingen.

- e). **Tegn** den kumulerede fordelingsfunktion (CDF) for salget (i stk) af almindelige pærer i discountbutikker.

**Bestem** ud fra CDF-plottet omtrent hvor stor en andel af årets måneder, der sælges mere end 100.000 stk. almindelige pærer i discountbutikker.

**Bestem** ud fra CDF-plottet omtrent hvor stor en andel af årets måneder, der sælges mellem 100.000 og 150.000 stk. almindelige pærer i discountbutikker.

- f). **Gør rede for** i hvilken af de to butikstyper, der oftest sælges mere end 150.000 stk. almindelige pærer i en given måned.

*Vink: Sammenlign CDF-plottene for almindelige pærer i de to butikstyper.*

### OPGAVE 2 (PRIS)

Salget af sparepærer i supermarkeder udgør en væsentlig andel af det samlede salg af el-pærer i den danske detailhandel. Prisen (variablen *Pris*) på sparepærer solgt i supermarkeder giver derfor en god indikation på det generelle prisniveau for el-pærer i den danske detailhandel. Vi ser i denne opgave nærmere på prisen på sparepærer solgt i supermarkeder.



- a). **Tegn** et histogram over fordelingen af prisen på sparepærer solgt i supermarkeder.  
**Estimer** middelværdi og standardafvigelse i den normalfordeling, der bedst beskriver prisen på sparepærer solgt i supermarkeder.  
*Vink: Brug f.eks. "Rows" -> "Data Filter" i JMP til at begrænse datamaterialet til kun at omhandle sparepærer solgt i supermarkeder.*
- b). **Beregn** på baggrund af den i delspg. a) beregnede normalfordeling sandsynligheden for, at prisen på en sparepære solgt i et supermarked højst er 35 kr.  
**Beregn** på baggrund af den i delspg. a) beregnede normalfordeling sandsynligheden for, at prisen på en sparepære solgt i et supermarked ligger mellem 33 kr. og 35 kr.
- c). **Angiv** på baggrund af den i delspg. a) beregnede normalfordeling et interval som med 95% sandsynlighed indeholder prisen på en sparepære solgt i et supermarked.

### OPGAVE 3 (PRIS OG BUTIKSTYPER)

Sædvanligvis vil man forvente, at de fleste forbrugsvarer kan købes til en lavere pris i discountbutikker end i supermarkeder. Vi ser i denne opgave på, om det også gør sig gældende for et så standardiseret produkt som sparepærer.

- a). **Tegn** for hver af de to butikstyper et boxplot over fordelingen af prisen på sparepærer.  
**Gør rede for** hvad de to boxplots indikerer om eventuelle forskelle i prisen på sparepærer i de to butikstyper.
- b). **Beregn** middelværdi og standardafvigelse for prisen på sparepærer i hver af de to butikstyper.  
**Gør rede for** hvorledes de beregnede nøgletal indikerer de samme tendenser som de i delspg. a) optegnede boxplots.
- c). **Gør rede for** om der er statistisk belæg for en nulhypotese om, at den forventede pris på sparepærer er den samme i supermarkeder og i discountbutikker. **Angiv** i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.
- d). **Beregn** et 95%-konfidensinterval for den forventede prisforskel på sparepærer i supermarkeder og discountbutikker og **giv** en fortolkning af intervallet.  
**Gør rede for** hvordan og hvorfor konfidensintervallet ændres, såfremt der i stedet beregnes et 90%-konfidensinterval.



#### OPGAVE 4 (PRISUDVIKLING)

Prisen på sparepærer har været forholdsvis stabil gennem de seneste år (variablen *År*), mens der er sket væsentlige ændringer i prisen på almindelige pærer. Vi ser i denne opgave nærmere på prisudviklingen på almindelige pærer.

- Tegn** for hvert år et boxplot af fordelingen af prisen på almindelige pærer solgt i supermarkeder.  
**Tegn** for hvert år et boxplot af fordelingen af prisen på almindelige pærer solgt i discountbutikker.  
*Vink: Sæt f.eks. År under "X, Factor" og begge de to variable Butikstype og Pæretype under "By" i JMPs "Fit Y by X"-funktion.*
- Gør rede for** hvad de i delspg. a) optegnede boxplots indikerer om prisudviklingen på almindelige pærer i de to butikstyper.
- Gør rede for** om der er statistisk belæg for en nulhypotese om, at den forventede pris på almindelige pærer i supermarkeder har været uændret over perioden 2013 til 2016. **Angiv** i den forbindelse antagelser, hypoteser og signifikanssandsynlighed.
- Beregn** hvor meget prisen på almindelige pærer i supermarkeder i gennemsnit har ændret sig over perioden 2013 til 2016.  
**Gør rede for** i hvilken af de to butikstyper prisen på almindelige pærer i gennemsnit har ændret sig mest over perioden 2013 til 2016.
- Beregn** et 95%-konfidensinterval for den forventede pris på almindelige pærer solgt i supermarkeder i 2016.  
**Gør rede for** hvordan og hvorfor konfidensintervallet ændres, såfremt det i stedet baseres på data for en længere periode (end blot årene 2013 til 2016).

#### OPGAVE 5 (VOLUMEN, ELPRIS OG DAGSLÆNGDE)

Salget af el-pærer afhænger udover pærernes pris forventeligt også af en række andre faktorer, som eksempelvis hvilket tidspunkt på året, vi ser på (variablen *Årstid*), samt hvor meget man skal betale for at bruge pærerne (variablen *Elpris*). Vi ser i denne opgave på både almindelige og sparepærer og på salget i både discountbutikker og i supermarkeder.

- Lav** passende valgte beregninger og/eller figurer der viser, hvorvidt der er forskel i elprisen henover de fire årstider.
- Gør rede for** hvorfor det ikke er muligt at undersøge en eventuel sammenhæng mellem elpris og årstider via beregning af en korrelation.
- Tegn** scatterplots af variablene *Stk*, *Dagslængde* og *Elpris* mod hinanden.  
**Beregn** variablenes indbyrdes korrelation og **gør rede for** mulig intuition bag fortegnene på de



beregnedede korrelationer.

- d). **Gør rede for** i hvilket omfang korrelationerne i delspg. c) kan siges at udtrykke en kausal sammenhæng.

#### OPGAVE 6 (VOLUMEN, SAMLET ANALYSE)

Mange faktorer påvirker salget af el-pærer, og der er således behov for en samlet vurdering af de enkelte faktorerers påvirkning. Vi ser derfor på en lineær regressionsmodel med variabelen *Stk* som responsvariabel og variablene *Pris*, *År*, *Årstid*, *Dagslængde*, *Elpris* som forklarende variable. For nemheds skyld ser vi i denne opgave udelukkende på salget af sparepærer i supermarkeder.

- a). **Reducér** den angivne regressionsmodel til kun at indeholde signifikante variable.  
*Vink: Fjern insignifikante variable fra modellen én ad gangen ved hele tiden at fjerne den mest insignifikante variabel. Som signifikansniveau anvendes som sædvanlig 5%. Husk at bruge resultaterne under "Indicator Function Parameterization".*
- b). **Gør rede for** hvordan og hvorfor determinationskoefficienten ændres, hvis der fjernes forklarende variable fra modellen.
- c). **Angiv** estimater for regressionskoefficienter hørende til variablene i den reducerede model fra delspg. a) og **giv** en fortolkning af estimaterne.
- d). **Beregn** et 95%-konfidensinterval for betydningen af en prisstigning på 1 kr. per sparepære og **giv** en fortolkning af intervallet.
- e). **Beregn** et 95%-konfidensinterval for betydningen af et prisfald på 1 kr. per sparepære og **giv** en fortolkning af intervallet.
- f). **Beregn** på baggrund af den estimerede reducerede regressionsmodel det forventede månedlige salg (i stk) af sparepærer i supermarkeder i vinteren 2016, forudsat at sparepærerne sælges til 35 kr. pr. stk., at den gennemsnitlige dagslængde i perioden er 8 timer, samt at elprisen per kWh er 225 øre.
- g). **Gør rede for** hvordan det forventede månedlige salg i delspg. f) ændres, såfremt der i stedet er tale om salget i efteråret 2016 (og alle øvrige oplysninger er uændrede).  
**Gør rede for** hvordan det forventede månedlige salg i delspg. f) ændres, såfremt elprisen per kWh i stedet er 230 øre (og alle øvrige oplysninger er uændrede).
- h). **Opskriv** nulhypotesen om at regressionskoefficienten hørende til året 2013 ikke har nogen betydning i den opstillede regressionsmodel.  
**Gør rede for** i let forståeligt sprog, hvilket udsagn der er udtrykt i nulhypotesen.
- i). **Beregn** om det er fra 2013 til 2014, fra 2014 til 2015 eller fra 2015 til 2016, at der ifølge modellen fra delspg. a) er sket det største fald i salget af sparepærer i supermarkeder.  
*Vink: Se på de estimerede regressionsparametre hørende til variabelen År*



## OPGAVE 7 (SAMLET VOLUMEN)

- a). **Gennemfør** en passende valgt analyse som undersøger udviklingen i det samlede salg af el-pærer i den danske detailhandel over perioden 2013 til 2016 (dvs. salget af både sparepærer og almindelige pærer i både discountbutikker og supermarkeder).

## DATAKILDE

Datamaterialet i ovenstående opgaver er venligst stillet til rådighed af markedsanalysevirksomheden Nielsen og er en del af Nielsen ScanTrack Dagligvareindeks<sup>TM</sup> (nielsen.com). Virksomheden giver følgende beskrivelse af sig selv:

Markedsanalysevirksomheden Nielsen er grundlagt i USA i 1923. Virksomheden samarbejder indenfor detail- og produktionssektoren med over 9.000 kunder fordelt på 100 lande.

Virksomhedens dagligvaredatabase ScanTrack indeholder salgsinformationer for udvalgte produkter fra dagligvarehandlen. Informationerne i den danske del af ScanTrack-databasen indsamles via virksomhedens samarbejde med alle de største kæder indenfor dagligvarehandlen i Danmark foruden benzinstationer samt kæderne 7-Eleven og Matas.

Nielsen indsamler på ugentlig basis detaljeret salgsinformation om udvalgte dagligvarer på butiksniveau, hvor der for hver enkelt vare – identificeret ved sin unikke EAN-kode ("stregkode") – registreres information om den samlede omsætning opgjort både i stk. og i kr. Informationerne grupperes herefter på baggrund af kriterier som eksempelvis producent, mærke, vægt, type, farve etc. til aggregerede størrelser for den samlede detailhandel i Danmark.

## DISCLAIMER:

Datamaterialet i nærværende opgave må udelukkende anvendes i forbindelse med faget Dataanalyse på CBS. Datamaterialet er fortroligt og må ikke videredistribueres eller offentliggøres andet end overfor studerende, der følger faget Dataanalyse på CBS, hverken skriftligt, mundtligt eller på anden vis, og ej heller de resultater og indsigter, der måtte komme ud af arbejdet med datamaterialet.





# VEJLEDENDE LØSNING

Bemærk at nedenstående løsning...

- udelukkende er **vejledende**, og at der i visse tilfælde kan være mere end én måde at besvare en given opgave på.
- visse steder har anført teoretiske formler sammen med det aflæste JMP-output for at fremme læsbarheden. Det er ikke et krav, at man gør noget tilsvarende i sin eksamensbesvarelse. Det er tilstrækkeligt at anføre de relevante tal direkte fra JMPs output samt inkludere det relevante output i besvarelsen.
- visse steder kommenterer på anførte antagelser. Det er ikke et krav, at man kommenterer på antagelser, medmindre det eksplicit efterspørges i opgaveteksten.



## OPGAVE 1

**1A):** Middelværdierne for det månedlige salg af hhv. sparepærer og almindelige pærer i hhv. supermarkeder og discountbutikker er

$$\hat{\mu}_{\text{spare}, \text{disc}} = 173.823 \text{ stk} \quad \hat{\mu}_{\text{alm}, \text{disc}} = 104.305 \text{ stk} \quad \hat{\mu}_{\text{spare}, \text{super}} = 213.197 \text{ stk} \quad \hat{\mu}_{\text{alm}, \text{super}} = 107.417 \text{ stk}$$

Distributions Butikstype=Discount, Pærtype=Spare	
Stk	
Summary Statistics	
Mean	173822,82
Std Dev	71788,121
Std Err Mean	11495,299
Upper 95% Mean	197093,84
Lower 95% Mean	150551,8
N	39

Distributions Butikstype=Discount, Pærtype=Almindelig	
Stk	
Summary Statistics	
Mean	104305,08
Std Dev	46767,088
Std Err Mean	7488,7275
Upper 95% Mean	119465,21
Lower 95% Mean	89144,941
N	39

Distributions Butikstype=Supermarkeder, Pærtype=Spare	
Stk	
Summary Statistics	
Mean	213197,26
Std Dev	102025,12
Std Err Mean	16337,094
Upper 95% Mean	246269,98
Lower 95% Mean	180124,55
N	39

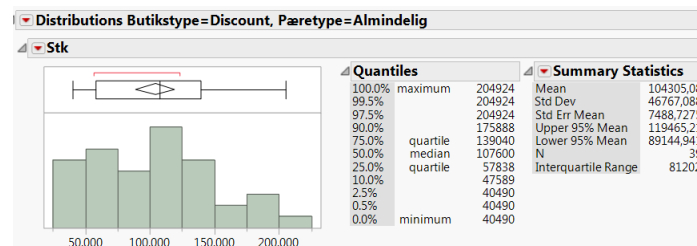
  

Distributions Butikstype=Supermarkeder, Pærtype=Almindelig	
Stk	
Summary Statistics	
Mean	107416,94
Std Dev	43796,112
Std Err Mean	7012,9905
Upper 95% Mean	121614
Lower 95% Mean	93219,886
N	39

**1B):** I gennemsnit sælges der (jf. delspg. 1a)) hver måned ca. 174.000 stk sparepærer i discountbutikker og ca. 213.000 stk sparepærer i supermarkeder. Der sælges således flere sparepærer i supermarkeder end i discountbutikker.

I gennemsnit sælges der i supermarkeder (jf. delspg. 1a)) hver måned ca. 107.000 stk almindelige pærer. I supermarkeder sælges der således flere sparepærer (ca. 213.000 stk pr. måned) end almindelige pærer.

**1C):** Histogrammet over fordelingen af salget af almindelige pærer i discountbutikker viser, at der er tale om en nogenlunde symmetrisk fordeling med en let tendens til højreskævhed, idet der enkelte uger sælges væsentligt mere (omkring 200.000 stk) end i alle øvrige perioder.



Fordelingens symmetri fremgår også delvist af boxplottet, som blandt andet viser, at middelværdien (104.305 stk) omtrent er lig medianen (107.600 stk), hvilket alt andet lige er en indikation på symmetri.

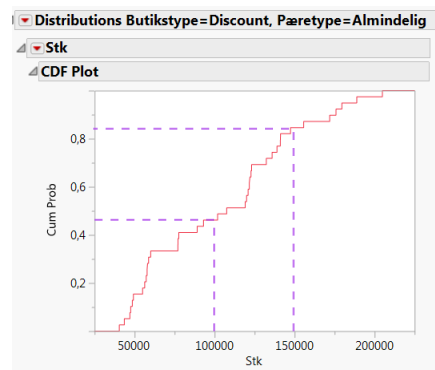


Ligeledes ses det, at fordelingen er uden væsentlige outliers idet eftersom standardafvigelsen, som er følsom overfor outliers, er mindre (46.767 stk) end interkvartilbredden (på  $139.040 - 57.838 = 81.202$  stk), der ikke er følsom overfor outliers.

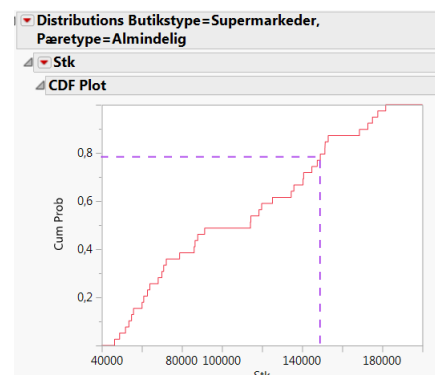
**1d):** Medianen i fordelingen af det månedlige salg af almindelige pærer i discountbutikker er 107.600 stk., svarende til at 50% af årets 12 måneder sælges 107.600 stk eller færre almindelige pærer i discountbutikker. På grund af fordelingens nogenlunde symmetriske form er middelværdi og median af omtrent samme størrelsesorden (jf. delspg. 1c)) og kan dermed begge bruges som mål for centrum af fordelingen.

**1e):** Den kumulerede fordelingsfunktion for det månedlige salg af almindelige pærer i discountbutikker viser, at ca. 45% af årets måneder sælges mindre end 100.000 stk. pærer. Dermed er det ca. 55% af årets måneder, der sælges 100.000 stk. pærer eller mere.

Den kumulerede fordelingsfunktion viser ligeledes, at ca. 85% af årets måneder sælges mindre end 150.000 stk. almindelige pærer. Dermed er det ca.  $85\% - 45\% = 40\%$  af årets måneder, der sælges mellem 100.000 og 150.000 stk. almindelige pærer i discountbutikker. M.a.o.: ca. 40% af tiden ligger det månedlige salg af almindelige pærer i discountbutikker på mellem 100.000 og 150.000 stk.



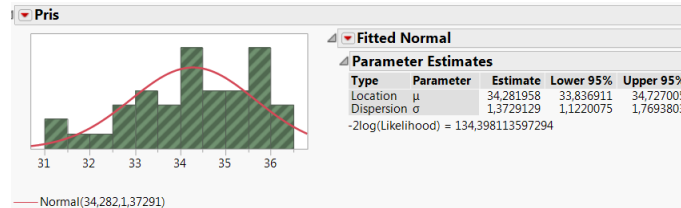
**1f):** Den kumulerede fordelingsfunktion for det månedlige salg af almindelige pærer i supermarkeder viser, at ca. 80% af årets måneder sælges mindre end 150.000 stk. pærer. Det tilsvarende tal for discountbutikker er ca. 85%, jf. delspg. 1e). Dermed er det ca. 20% (ca. 15%) af årets måneder, der i supermarkeder (discountbutikker) sælges 150.000 stk almindelige pærer eller mere. Det er således i supermarkeder, at der oftest sælges 150.000 stk eller flere almindelige pærer i løbet af en måned.





## OPGAVE 2

**2A):** Histogram samt middelværdi (34,28 kr.) og standardafvigelse (1,37 kr.) for fordelingen af prisen på sparepærer i supermarkeder:



**2B):** Sandsynligheden for at prisen på en sparepære solgt i et supermarked er højst 35 kr. er

$$P(\text{Pris} \leq 35) = 70,0\%$$

hvor sandsynligheden er beregnet i en normalfordeling med middelværdi og std.afv. som angivet i delspg. a).

Sandsynligheden for at prisen på en sparepære solgt i et supermarked er mellem 33 kr. og 35 kr. er

$$P(33 \leq \text{Pris} \leq 35) = P(\text{Pris} \leq 35) - P(\text{Pris} \leq 33) = 70,0\% - 17,5\% = 52,5\%$$

hvor sandsynligheden er beregnet i en normalfordeling med middelværdi og std.afv. som angivet i delspg. a).

**2c):** I en normalfordeling ligger 95% af sandsynligheden i intervallet fra "middelværdi-1,96·standardafvigelse" til "middelværdi+1,96·standardafvigelse", dvs. i intervallet

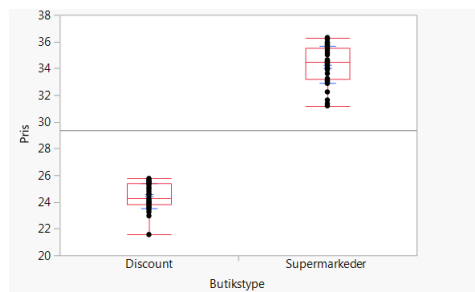
$$[34,28 - 1,96 \cdot 1,37; 34,28 + 1,96 \cdot 1,37] = [31,59; 36,97]$$

Med 95% sandsynlighed ligger prisen på en sparepære, såfremt den købes i et supermarked, således mellem 31,59 kr. og 36,97 kr.



### OPGAVE 3

**3A):** Figuren viser, at prisniveauet for sparepærer er markant lavere i discountbutikker end i supermarkeder med en pris på 24-26 kr. pr. stk for sparepærer købt i discountbutikker og 33-36 kr. for en sparepære købt i et supermarked<sup>2</sup>. Der er således en klar indikation på, at det er væsentligt billigere at købe sparepærer i discountbutikker end i supermarkeder. Variationen fra måned til måned i prisen på sparepærer ser ud til at være en anelse mindre i discountbutikker end i supermarkeder (idet boxplottet for prisen i discountbutikker er lidt smallere).



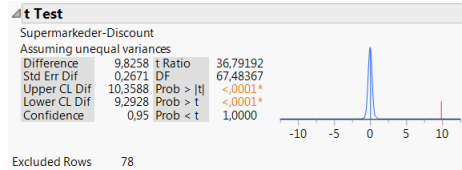
**3B):** Middelværdi og standardafvigelse for fordelingen af prisen på sparepærer i hhv. discountbutikker og supermarkeder understøtter tendenserne i boxplottene beskrevet i delspg. a). En lavere middelværdi i fordelingen for discountbutikker (24,46 kr. mod 34,28 kr. for supermarkeder) indikerer, at sparepærer sælges væsentligt billigere i discountbutikker, mens en lidt lavere standardafvigelse (0,95 kr. mod 1,37 kr. for supermarkeder) ligeledes indikerer en lidt mindre variation fra måned til måned i prisen på sparepærer i discountbutikker sammenlignet med variationen i prisen i supermarkeder.

Means and Std Deviations						
Level	Number	Mean	Std Dev	Std Err		
Discount	39	24,4561	0,94696	0,15164	Lower 95%	Upper 95%
Supermarkeder	39	34,2820	1,37291	0,21984	33,837	34,727
Excluded Rows		78				

<sup>2</sup>Bemærk: Datamaterialet giver ikke mulighed for at undersøge, om det er forskellige mærker pærer, og dermed også forskellige kvalitet af pærer, der sælges i de to butikstyper.



**3c):** Test af nulhypotese om at den forventede pris på sparepærer er den samme i discountbutikker og supermarkeder:



**Antagelser:** Observationerne af de enkelte måneders priser er indbyrdes uafhængige og approksimativt normalfordelte; normalfordelte indenfor hver butikstype.

**Hypoteser:**  $H_0 : \mu_{discount} = \mu_{supermarked}$

$H_a : \mu_{discount} \neq \mu_{supermarked}$

**Signifikansssh:**  $p < 0,0001$

**Tekn. konklusion:** Signifikanssandsynligheden er mindre end 0,05 og dermed forkastes  $H_0$ .

**Let konklusion:** Der er ikke belæg for en påstand om, at den forventede pris på sparepærer er den samme i discountbutikker og supermarkeder. På baggrund af resultaterne i delspg. a) og b) er den konklusion ikke overraskende, eftersom den forventede pris på en sparepære estimeres til 24,46 kr. i discountbutikker og 34,28 kr. i supermarkeder. Det er således væsentligt billigere at købe sparepærer i discountbutikker end i supermarkeder.

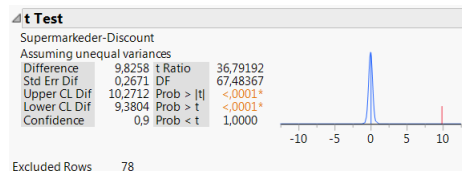
**3d):** Et 95%-konfidensinterval for den forventede prisforskel på sparepærer i discountbutikker og supermarkeder er (jf. JMP-outputtet i delspg. 3c))

[9,29 kr; 10,36 kr]

Med 95% sandsynlighed vil den forventede prisforskel på sparepærer købt i henholdsvis en discountbutik og et supermarked således ligge mellem 9,29 kr og 10,36 kr.

Hvis der i stedet beregnes et 90%-konfidensinterval, betyder det, at intervallet med mindre sandsynlighed (90% i stedet for 95%) vil indeholde den sande prisforskel på sparepærer i de to butikstyper. En mindre grad af sikkerhed for at konfidensintervallet fanger den sande prisforskel betyder, at konfidensintervallet bliver smallere. Såfremt intervallet beregnes (ikke et krav til besvarelsen) fås et 90%-konfidensinterval på

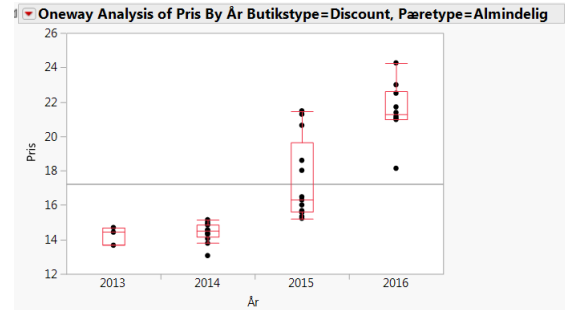
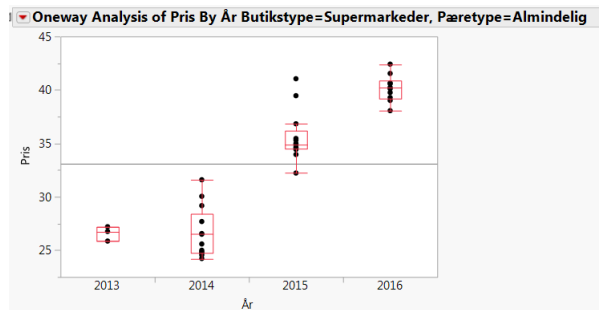
[9,38 kr; 10,27 kr]





## OPGAVE 4

**4A):** Boxplots af fordelingen af prisen på almindelige pærer solgt i supermarkeder (venstre) og discountbutikker (højre) i perioden 2013-2016:



**4B):** Boxplottene i delspg. 4b) viser, at prisen på almindelige pærer er steget væsentligt over perioden 2013-2016 i både supermarkeder og i discountbutikker. I supermarkeder er prisen på almindelig el-pærer steget fra ca. 26 kr. i 2013 til ca. 40 kr i 2016, mens prisen i discountbutikker er steget fra ca. 15 kr. i 2013 til ca. 22 kr. i 2016. Der er således i begge butikstyper tale om prisstigninger på ca. 50% i løbet af perioden.

**4c):** Test af nulhypotese om at den forventede pris på almindelige pærer solgt i supermarkeder har været uændret over perioden 2013-2016:

Oneway Anova					
Summary of Fit					
Rsquare			0.895082		
Adj Rsquare			0.886089		
Root Mean Square Error			2.035785		
Mean of Response			33.10023		
Observations (or Sum Wgts)			39		
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
År	3	1237.5021	412.501	99.5316	<.0001*
Error	35	145.0547	4.144		
C. Total	38	1382.5568			
Means for Oneway Anova					
Level	Number	Mean	Std Error	Lower 95%	Upper 95%
2013	3	26.6076	1.1754	24.221	28.994
2014	13	26.6982	0.5646	25.552	27.844
2015	13	35.5656	0.5646	34.419	36.712
2016	10	40.1656	0.6438	38.859	41.473

Std Error uses a pooled estimate of error variance

Antagelser:

Observationerne af de enkelte måneders priser er indbyrdes uafhængige og approksimativt normalfordelte; normalfordelte indenfor hvert år og med samme std.afvigelse for alle årene. Hvorvidt antagelsen om normalfordeling indenfor hvert år er opfyldt kan ikke vurderes ud fra de i delspg. a) optegnede box plots, men antagelsen om samme std.afvigelse ser iflg. boxplottene ikke helt ud til at være opfyldt (boxplottene er ikke lige brede).

Hypoteser:

$$H_0: \mu_{2013} = \mu_{2014} = \mu_{2015} = \mu_{2016}$$

$H_a$ : Mindst to af de forventede værdier  $\mu_{2013}, \mu_{2014}, \mu_{2015}, \mu_{2016}$  er ikke ens

Signifikansssh.:

$$p < 0,0001$$

Tekn. konklusion:

Signifikanssandsynligheden er mindre end 0,05 og dermed forkastes  $H_0$ .

Let konklusion:

Der er ikke belæg for en påstand om, at den forventede pris på almindelige pærer solgt i supermarkeder er den samme i perioden 2013-2016. Konklusionen er ikke overraskende på baggrund af resultaterne i delspg. b), som viser at den gennemsnitlige pris på almindelige pærer er steget fra ca. 26 kr. i 2013 til ca. 40 kr. i 2016.



**4d):** Middelværdi af prisen på almindelige pærer solgt i supermarkeder (venstre) og discountbutikker (højre) i perioden 2013-2016:

Oneway Analysis of Pris By År Butikstype=Supermarkeder, Pærtype=Almindelig						
Means and Std Deviations						
Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
2013	3	26,6076	0,68514	0,39556	24,906	28,310
2014	13	26,6982	2,31757	0,64278	25,298	28,099
2015	13	35,5656	2,33854	0,64859	34,152	36,979
2016	10	40,1656	1,24887	0,39493	39,272	41,059

Oneway Analysis of Pris By År Butikstype=Discount, Pærtype=Almindelig						
Means and Std Deviations						
Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
2013	3	14,2609	0,53433	0,30849	12,934	15,588
2014	13	14,4433	0,56883	0,15776	14,100	14,787
2015	13	17,4022	2,35223	0,65239	15,981	18,824
2016	10	21,5127	1,60196	0,50658	20,367	22,659

I gennemsnit er prisen på almindelige pærer solgt i supermarkeder steget fra 26,61 kr. i 2013 til 40,17 kr. i 2016, dvs. prisen er steget med 13,56 kr.

I gennemsnit er prisen på almindelige pærer solgt i discountbutikker steget fra 14,26 kr. i 2013 til 21,51 kr. i 2016, dvs. prisen er steget med 7,25 kr. Prisen på almindelige pærer er således steget mest i supermarkeder.

**4e):** Et 95%-konfidensinterval for den forventede pris på almindelige pærer solgt i supermarkeder i 2016 er (jf. JMP-outputtet i delspg. 4d))

[39,27 kr; 41,06 kr]

Med 95% sandsynlighed vil den forventede pris på en almindelig el-pære købt i et supermarked i 2016 således være mellem 39,27 kr og 41,06 kr. M.a.o.: Har man købt en almindelig el-pære i et supermarked engang i løbet af 2016, så har man med 95% sandsynlighed betalt mellem 39,27 kr. og 41,06 kr. for den.

Hvis konfidensintervallet baseres på data for en længere periode (end blot 2016) kan følgende to effekter påvirke konfidensintervallet:

- Såfremt der eksempelvis har været en trend i prisudviklingen over den valgte periode, vil inkludering af data for en længere periode betyde, at variationen i data (målt ved standardafvigelsen) vil blive større og dermed vil konfidensintervallet blive bredere.
- Såfremt der ikke har været en trend i prisudviklingen over den valgte periode, vil inkludering af data for en længere periode betyde, at det er muligt at estimere den forventede pris med større sikkerhed, og dermed vil konfidensintervallet blive smallere.

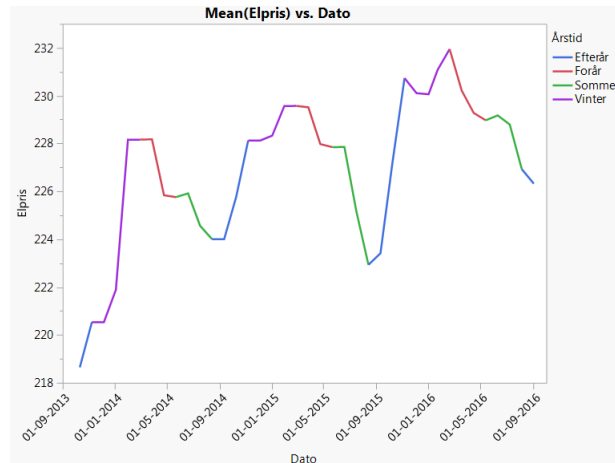
Hvilken af de to effekter der vil dominere er det ikke muligt at sige noget generelt om.



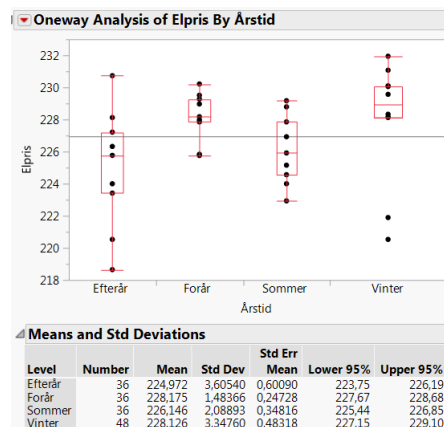


## OPGAVE 5

**5A):** Nedenstående figur af udviklingen i el-prisen over tid med farvemarkering af de fire årstider viser, at el-prisen alt andet lige er højest i vinter- og forårsperioderne.



Samme konklusion fremgår af nedenstående boxplots af fordelingen af el-prisen for hver af de fire årstider, ligesom de beregnede middelværdier viser, at den gennemsnitlige el-pris i vinter- og forårsperioderne er ca. 228 kr. per kWh, mens den om sommeren er ca. 226 kr og om efteråret er ca. 224 kr. per kWh. Tallene er beregnet over hele perioden 2013-2016 og skal derfor tages med et væsentligt forbehold, idet figuren ovenfor viser, at der har været en generel stigning i el-prisen over perioden 2013-2016.



**5B):** Beregning af en korrelation mellem to variable er kun mulig, såfremt begge variable er kvantitative variable. Eftersom variabelen Årstid er kategorisk, er det derfor ikke muligt at beregne en korrelation mellem Elpris og Årstid.

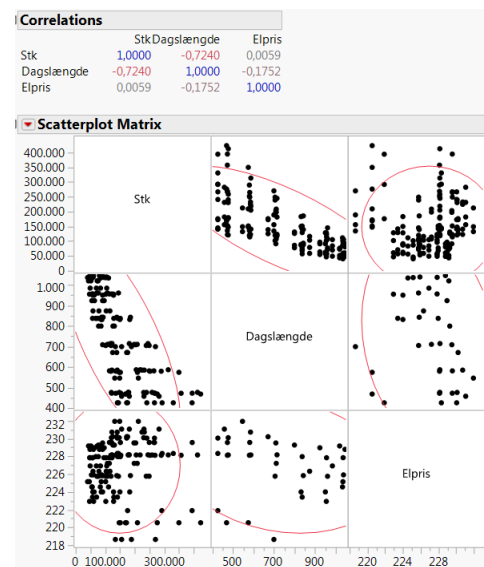


**5c):** Scatterplottene af det solgte antal el-pærer (*Stk*) mod henholdsvis den gennemsnitlige dagslængde (*Dagslængde*) og den gennemsnitlige elpris per kWh (*Elpris*) afspejler ganske klart de beregnede korrelationer.

Korrelationen mellem det solgte antal el-pærer og dagslængden er  $-0,72$  og indikerer en stærk negativ sammenhæng, hvilket også fremgår af det tilhørende scatterplot. Den stærke negative korrelation afspejler, at der sælges langt flere el-pærer i de perioder af året med færrest lyse timer (= kortest dagslængde). M.a.o.: Der sælges flest el-pærer om vinteren og færrest el-pærer om sommeren.

Korrelationen mellem det solgte antal el-pærer og elprisen er  $0,01$ , hvilket indikerer at der ikke er nogen nævneværdig sammenhæng mellem prisen på el og salget af el-pærer. Det afspejles også i det tilhørende scatterplot, idet der ikke er nogen tydelig sammenhæng mellem de optegnede punkter. Idet de fleste forbrugere næppe i væsentligt omfang vil være påvirket af udviklingen i elprisen i forhold til hvor, hvornår og hvor mange el-pærer de køber, så virker det ikke overraskende, at korrelationen er tæt på  $0$ .

Korrelationen mellem dagslængde og elpris er  $-0,18$ , hvilket indikerer en svag negativ sammenhæng, hvilket også fremgår af det tilhørende scatterplot. Den negative sammenhæng afspejler, at prisen på el alt andet lige er højere i de perioder af året med færrest lyse timer (fordi der er en større efterspørgsel efter el i de mørkeste - og koldeste - perioder af året).



**5d):** Korrelation mellem to variable udtrykker ikke nødvendigvis en årsagssammenhæng (kausal sammenhæng) mellem de to variable. I tilfældet med de tre variable *Stk*, *Dagslængde* og *Elpris* kan man argumentere for, at korrelationen...

- mellem *Stk* og *Dagslængde* udtrykker en kausal sammenhæng, idet efterspørgslen efter (og dermed også salget af) el-pærer er størst i de perioder af året med færrest lyse timer, og *Dagslængde* dermed har en indflydelse på *Stk*
- mellem *Stk* og *Elpris* ikke udtrykker nogen kausal sammenhæng, fordi korrelationen er tæt på  $0$
- mellem *Dagslængde* og *Elpris* i et vist omfang udtrykker en kausal sammenhæng, idet efterspørgslen efter (og dermed også prisen på) el er størst i de perioder af året med færrest lyse timer, og *Dagslængde* dermed har en indflydelse på *Elpris*



## OPGAVE 6

**6A):** Variablene i den reducerede regressionsmodel (ved et signifikansniveau på 5%) er *Pris*, *År*, *Årstid*, *Dagslængde*, idet regressionsparameteren hørende til variabelen *Elpris* bliver insignifikant og variabelen derfor udelades af modellen.

Effect Tests					
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Pris	1	1	8823474713	16,3408	0,0003*
År	3	3	2,0692e+10	12,7734	<,0001*
Årstid	3	3	5969324891	3,6850	0,0227*
Dagslængde	1	1	2,5791e+10	47,7646	<,0001*

**6B):** Determinationskoefficienten  $R^2$  udtrykker, hvor stor en del af variationen i responsvariabelen *Stk* der kan beskrives v.h.j.a. regressionsmodellens forklarende variable. Jo færre variable der inkluderes i en regressionsmodel, desto færre muligheder er der for at forklare variationen i responsvariabelen, og desto mindre vil determinationskoefficienten derfor være.

**6C):**

Parameterestimerne hørende til variabelen *År* er...

- $\hat{\beta}_{\text{År},2013} = 92.346$
- $\hat{\beta}_{\text{År},2014} = 56.567$
- $\hat{\beta}_{\text{År},2015} = 41.562$
- $\hat{\beta}_{\text{År},2016} = 0$

Resultatet af den estimerede model:

Indicator Function Parameterization						
Term	Estimate	Std Error	t Ratio	Prob>  t	Lower 95%	Upper 95%
Intercept	965672.92	128820.4	7.50	<,0001*	702586.49	1228759.4
Pris	-15332.2	3792.868	-4.04	0,0003*	-23078.27	-7586.125
År(2013)	92346.29	18940.09	4.88	<,0001*	53665.463	131027.12
År(2014)	56566.816	10667.13	5.30	<,0001*	34781.636	78351.996
År(2015)	41561.568	10015.75	4.15	0,0003*	21106.683	62016.454
Årstid(Efterår)	-38007.54	15359.55	-2.47	0,0192*	-69375.94	-6639.148
Årstid(Forår)	-34563.33	18945.39	-1.82	0,0781	-73254.98	4128.3307
Årstid(Sommer)	-22952.47	26822.9	-0.86	0,3989	-77732.15	31827.208
Dagslængde	-332,1021	48,05276	-6,91	<,0001*	-430,2389	-233,9653

Det betyder, at det forventede månedlige salg af sparepærer i supermarkeder alt andet lige er...

- 92.346 stk større i år 2013 end i år 2016
- 56.567 stk større i år 2014 end i år 2016
- 41.562 stk større i år 2015 end i år 2016

Parameterestimerne hørende til variabelen *Årstid* er...

- $\hat{\beta}_{\text{Årstid,Efterår}} = -38.008$
- $\hat{\beta}_{\text{Årstid,Forår}} = -34.563$
- $\hat{\beta}_{\text{Årstid,Sommer}} = -22.952$
- $\hat{\beta}_{\text{Årstid,Vinter}} = 0$

Det betyder, at det forventede månedlige salg af sparepærer i supermarkeder alt andet lige er...

- 38.008 stk mindre om efteråret end om vinteren
- 34.563 stk mindre om foråret end om vinteren
- 22.952 stk mindre om sommeren end om vinteren



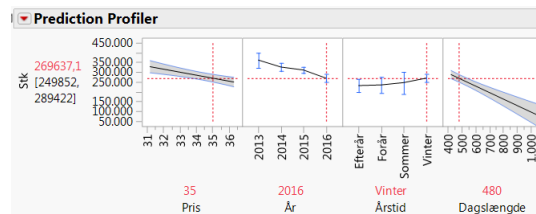
Parameterestimatet hørende til variablen...

- *Pris* er  $\hat{\beta}_{\text{Pris}} = -15.332$ , således at det forventede månedlige salg af sparepærer i supermarkeder falder med 15.332 stk, hvis den gennemsnitlige pris per pære stiger med 1 kr.
- *Dagslængde* er  $\hat{\beta}_{\text{PrisRoyal}} = -332$ , således at det forventede månedlige salg af sparepærer i supermarkeder falder med 332 stk, for hvert minut dagen tiltager i længde.

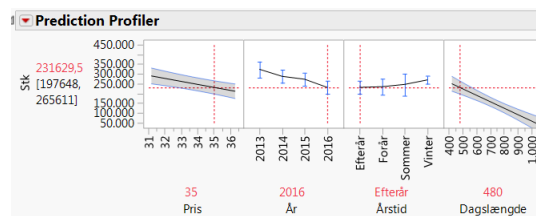
**6d):** Et 95%-konfidensinterval for betydningen af en prisstigning på 1 kr. per sparepære er (jf. JMP-outputtet på forrige side)  $[-23.078; -7.586]$ . Med 95% sandsynlighed vil det forventede månedlige salg af sparepærer i supermarkeder dermed falde med mellem 7.586 og 23.078 stk, såfremt prisen per pære stiger med 1 kr.

**6e):** På baggrund af tallene fra besvarelsen af delspg. d) ses det, at det forventede månedlige salg af sparepærer i supermarkeder vil stige med mellem  $-7.586 \cdot (-1) = 7.586$  og  $-23.078 \cdot (-1) = 23.078$  stk, såfremt prisen per sparepærer falder med 1 kr. Det efterspurgte 95%-konfidensinterval bliver dermed  $[7.586; 23.078]$ .

**6f):** Det forventede månedlige salg af sparepærer i supermarkeder i vinteren 2016, forudsat at prisen er 35 kr, den gennemsnitlige dagslængde er 8 timer (= 480 minutter) og elprisen er 225 øre per kWh er 269.637 stk.<sup>3</sup>



**6g):** Det forventede månedlige salg af sparepærer i supermarkeder i efteråret 2016, forudsat at prisen er 35 kr, den gennemsnitlige dagslængde er 8 timer (= 480 minutter) og elprisen er 225 øre per kWh er 231.629 stk. Det forventede salg falder således med  $269.637 - 231.629 = 38.008$  stk (idet  $\hat{\beta}_{\text{Årstid,Efterår}} = -38.008$ ).



Det forventede månedlige salg af sparepærer i supermarkeder er uændret, såfremt elprisen stiger fra 225 øre til 230 øre per kWh, fordi elprisen ikke indgår i den reducerede regressionsmodel.

<sup>3</sup>Fordi variablen *Elpris* ikke indgår i modellen er oplysningen om en elpris på 225 øre per kWh irrelevant for løsningen af dette delspg.



**6h):** Nulhypotesen om at regressionskoefficienten hørende til året 2013 ikke har nogen betydning i den opstillede regressionsmodel er

$$H_0 : \beta_{\hat{A}_r, 2016} = 0$$

Nulhypotesen udtrykker, at der ikke er forskel på salget af sparepærer i supermarkeder i år 2013 og år 2016 (fordi år 2016 er "baseline" værdien for variabelen  $\hat{A}_r$ , dvs. den værdi af variabelen, der ikke optræder en dummy for i den opstillede model).

**6i):** Ifølge den opstillede regressionsmodel er det estimerede fald i salget af sparepærer i supermarkeder (jf. delspg. 6c))...

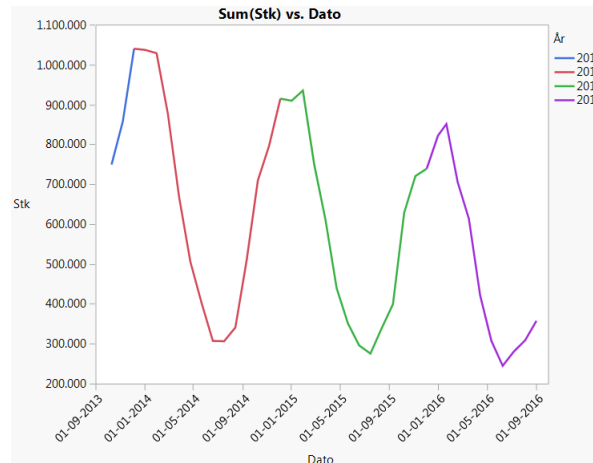
- $92.346 - 56.567 = 35.779$  stk fra år 2013 til år 2014
- $56.567 - 41.562 = 15.005$  stk fra år 2014 til år 2015
- 41.562 stk fra år 2015 til år 2016

Det største fald i salget af sparepærer i supermarkeder (på 41.562 stk) sker således fra år 2015 til år 2016.



### OPGAVE 7

Det samlede månedlige salg af el-pærer i den danske detailhandel har været faldende over perioden 2013-2016, hvilket bl.a. fremgår af nedenstående figur over det samlede månedlige salg.



Alternativt kan man f.eks. se på det gennemsnitlige månedlige salg (gennemsnit over pæretype og butikstype) for hvert af de fire år. Heraf fremgår det ligeledes, at det gennemsnitlige månedlige salg er faldet fra 220.533 stk i 2013 til 122.760 stk i 2016.

Distributions År=2013	
Stk	
Summary Statistics	
Mean	220532.79
Std Dev	90970.98
Std Err Mean	26261.06
Upper 95% Mean	278332.99
Lower 95% Mean	162732.59
N	12
Distributions År=2014	
Stk	
Summary Statistics	
Mean	161560.32
Std Dev	89525.587
Std Err Mean	12414.965
Upper 95% Mean	186484.4
Lower 95% Mean	136636.24
N	52
Distributions År=2015	
Stk	
Summary Statistics	
Mean	142173.6
Std Dev	79063.8
Std Err Mean	10964.176
Upper 95% Mean	164185.11
Lower 95% Mean	120162.1
N	52
Distributions År=2016	
Stk	
Summary Statistics	
Mean	122759.61
Std Dev	63914.93
Std Err Mean	10105.838
Upper 95% Mean	143200.6
Lower 95% Mean	102318.63
N	40



Hvis man ser nærmere på omsætningstallene (ikke et krav til besvarelsen), fremgår det, at det er salget af sparepærer, der er faldet væsentligt og primært salget af sparepærer i supermarkeder.

