

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.