

Copenhagen Business School
Dataanalyse, HD 1. del

EKSAMENSOPGAVE
ORDINÆR PRØVE, EFTERÅR 2017

4 timers skriftlig eksamen

Opgaven består af i alt **31** delspørgsmål fordelt på **9** opgaver. Ved bedømmelsen vægtes alle opgavens delspørgsmål ligeligt med undtagelse af delspørgsmål 9 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.



ØLSALG I DANMARK

BAGGRUND OG DATABESKRIVELSE

Salget af øl på det danske marked er præget af stor konkurrence, både mellem forskellige produkter og mellem forskellige butikskæder. 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 øl i Danmark.

Datamaterialet i JMP-filen *ØlsalgDK.jmp* indeholder 3 års ugentlige observationer af salget af 33 cl dåseøl af fire forskellige mærker (Grøn Tuborg, Carlsberg Pilsner, Heineken Premium og Royal Export) i to forskellige pakkestørrelser: 6 stk ("six pack") og 24 stk ("en ramme") i syv forskellige supermarkeds-kæder (Bilka, Føtex, Discount¹, Købmandshuset, Kvickly, Dagli/Lokal/Irma² og SuperBrugsen).

Datamaterialet indeholder information om den omsatte mængde (dvs. én six pack = 2 liter og én ramme = 8 liter), om den gennemsnitlige ugentlige salgspris (angivet i kr. pr. liter) samt information om i hvilket omfang, salget er sket til en midlertidigt nedsat pris ("rabatpris"). Hvis mindst halvdelen af ugens omsatte mængde (målt i liter) er solgt til rabatpris, betegnes ugen som en "kampagneuge", dvs. en uge hvor der har været ekstraordinært omfattende prisreduktion i forhold til normalprisen.³

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

¹'Discount' indeholder det samlede salg i Netto, Fakta, Rema og Kiwi butikkerne.

²"Daglig/Lokal/Irma" indeholder det samlede salg i Dagli'Brugsen, LokalBrugsen og Irma butikkerne.

³Hvorvidt en given uge er en kampagneuge eller ej for et givet ølmærke, afhænger både af hvilken pakkestørrelse og hvilken supermarkeds-kæde omsætningstallene stammer fra. Det kan således sagtens være, at en given uge er kampagneuge for Royal Export 6 stk. pakker i SuperBrugsen, mens den samme uge ikke er kampagneuge for Royal Export 24 stk. pakker i SuperBrugsen (fordi der i så fald kun er tilbud i SuperBrugsen på Royal Export i 6 stk. pakker og ikke i 24 stk. pakker i den pågældende uge).



VARIABELNAVN	FORKLARING
<i>Generelt</i>	
<i>Dato</i>	Sidste dag i den betragtede uges omsætning
<i>Kæde</i>	Hvilken supermarkeds-kæde omsætningstallene stammer fra (med de mulige værdier "Bilka", "Føtex", "Discount", "Købmands-huset", "Kvickly", "Dagli/Lokal/Irma", "SuperBrugsen")
<i>Pakkestørrelse</i>	Antal 33 cl dåser i den solgte vare (med de mulige værdier "6 stk", "24 stk")
<i>Prisinformation (angivet i kr. pr. liter)</i>	
<i>PrisCarlsberg</i>	Pris på Carlsberg Pilsner
<i>PrisTuborg</i>	Pris på Grøn Tuborg
<i>PrisHeineken</i>	Pris på Heineken Premium
<i>PrisRoyal</i>	Pris på Royal Export
<i>Omsætningsinformation (angivet i 100 liter)</i>	
<i>VolumeCarlsberg</i>	Ugentligt omsætning af Carlsberg Pilsner
<i>VolumeTuborg</i>	Ugentligt omsætning af Grøn Tuborg
<i>VolumeHeineken</i>	Ugentligt omsætning af Heineken Premium
<i>VolumeRoyal</i>	Ugentligt omsætning af Royal Export
<i>Reklamekampagneinformation I (angivet i procent)</i>	
<i>ShPromoCarlsberg</i>	Andel af den ugentlige omsætning (i liter) af Carlsberg Pilsner, der er solgt til rabatpris
<i>ShPromoTuborg</i>	Andel af den ugentlige omsætning (i liter) af Grøn Tuborg, der er solgt til rabatpris
<i>ShPromoHeineken</i>	Andel af den ugentlige omsætning (i liter) af Heineken Premium, der er solgt til rabatpris
<i>ShPromoRoyal</i>	Andel af den ugentlige omsætning (i liter) af Royal Export, der er solgt til rabatpris
<i>Reklamekampagneinformation II</i>	
<i>UgeCarlsberg</i>	Hvorvidt den betragtede uge er en kampagneuge for Carlsberg Pilsner (med de mulige værdier "Almindelig", "Kampagne")
<i>UgeTuborg</i>	Hvorvidt den betragtede uge er en kampagneuge for Grøn Tuborg (med de mulige værdier "Almindelig", "Kampagne")
<i>UgeHeineken</i>	Hvorvidt den betragtede uge er en kampagneuge for Heineken Premium (med de mulige værdier "Almindelig", "Kampagne")
<i>UgeRoyal</i>	Hvorvidt den betragtede uge er en kampagneuge for Royal Export (med de mulige værdier "Almindelig", "Kampagne")



OPGAVE 1 (VOLUMEN)

For at få et overblik over hvor meget øl af mærket Royal Export der sælges på det danske marked, ser vi på den ugentligt omsatte mængde af Royal Export, som angivet i variabelen *VolumeRoyal*, for hver af de to pakkestørrelser, som angivet i variabelen *Pakkestørrelse*.

- a). **Tegn** for hver af de to pakkestørrelser et histogram og et boxplot over fordelingen af variabelen *VolumeRoyal*.
Giv en kort beskrivelse af de to fordelinger af variabelen *VolumeRoyal* på baggrund af histogram, boxplot og relevante nøgletal for centrum og variation.
- b). **Angiv og fortolk** medianen i fordelingen af variabelen *VolumeRoyal* for 6 stk. pakker.
Gør rede for om median eller middelværdi (eller begge) er retvisende mål for centrum i fordelingen af variabelen *VolumeRoyal* for 6 stk. pakker.

I de resterende delspørgsmål i denne opgave ser vi udelukkende på omsætningen af Royal Export i 6 stk. pakker i SuperBrugsen.

*Vink: Sæt f.eks. begge de to variable *Kæde* og *Pakkestørrelse* under "By" i JMPs "Distribution"-funktion.*

- c). **Tegn** den kumulerede fordelingsfunktion (CDF) for variabelen *VolumeRoyal*.
Gør rede for hvorvidt den kumulerede fordelingsfunktion indikerer, at der (i gennemsnit) mindst hver fjerde uge sælges mere end 7.000 liter Royal Export i 6 stk. pakker i SuperBrugsen.
- d). **Beregn** et 95%-konfidensinterval for den forventede omsatte mængde af Royal Export i 6 stk. pakker i SuperBrugsen.
Gør rede for hvorledes konfidensintervallet ændres, såfremt standardafvigelsen af variabelen *VolumeRoyal* stiger.
- e). **Gør rede for** om der er statistisk belæg for en nulhypotese om, at den forventede ugentlige omsatte mængde af Royal Export i 6 stk. pakker i SuperBrugsen er 5.250 liter. **Angiv** i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.

OPGAVE 2 (VOLUMEN OG PRIS HHV. REKLAME)

Hvor meget Royal Export øl der sælges i en given uge hænger blandt andet sammen med hvilken pris, produktet sælges til. Derudover er det naturligt at forvente, at salget også påvirkes af, i hvilket omfang der gennemføres tilbudskampagner for produktet samt af salget af konkurrerende produkter.

- a). **Tegn** scatterplots af variablene *VolumeRoyal*, *PrisRoyal* og *ShPromoRoyal* mod hinanden.
Beregn variablenes indbyrdes korrelation og **gør rede for** mulig intuition bag fortegnene på de beregnede korrelationer.



- b). **Tegn** scatterplots af variablene *VolumeRoyal*, *VolumeHeineken* og *ShPromoHeineken* mod hinanden.
Beregn variablenes indbyrdes korrelation og **gør rede for** mulig intuition bag fortegnene på de beregnede korrelationer.
- c). **Gør rede for** i hvilket omfang korrelationerne i delspg. b) kan siges at udtrykke en kausal sammenhæng.

OPGAVE 3 (PRIS OG PAKKESTØRRELSE)

Royal Export sælges i forskellige pakkestørrelser og et oplagt spørgsmål er derfor, hvorvidt der er forskel på prisen pr. liter afhængig af hvilken pakkestørrelse, man køber.

- a). **Tegn** for hver af de to pakkestørrelser et histogram over fordelingen af *PrisRoyal* og en tilhørende normalfordelingskurve.
Gør for hver af de to pakkestørrelser **rede for** (på baggrund af figurene) om normalfordelingen ser ud til at give en rimelig beskrivelse af fordelingen af prisen pr. liter Royal Export.
- b). **Gør rede for** om der er statistisk belæg for en nulhypotese om, at den forventede pris pr. liter Royal Export er den samme for pakker med henholdsvis 6 og 24 stk. **Angiv** i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.
- c). **Beregn** et 95%-konfidensinterval for forskellen i den forventede pris pr. liter Royal Export for pakker med henholdsvis 6 og 24 stk. og **giv** en fortolkning af intervallet.

OPGAVE 4 (PRIS OG KÆDE)

Prisen på Royal Export varierer ikke kun som følge af reklamekampagner og påvirkning fra konkurrerende produkter. Royal Export sælges i en række forskellige supermarkeds kæder, og ikke altid til samme pris i de forskellige kæder. Det er derfor relevant at undersøge, om der er systematiske forskelle i prisen på tværs af de forskellige kæder.

- a). **Tegn** et boxplot over fordelingen af variablen *PrisRoyal* for hver værdi af variablen *Kæde*.
Beregn middelværdi og standardafvigelse af variablen *PrisRoyal* for hver værdi af variablen *Kæde*.
- b). **Gør rede for** at der ikke er statistisk belæg for en nulhypotese om, at den forventede pris pr. liter Royal Export er den samme i alle kæder. **Angiv** i den forbindelse antagelser, hypoteser og signifikanssandsynlighed.
Vink: Gennemfør et passende valgt statistisk test og konkluder på baggrund heraf, at den relevante nulhypotese forkastes.



- c). **Gør rede for** at der er statistisk belæg for en nulhypotese om, at den forventede pris pr. liter Royal Export er den samme i mindst tre af de betragtede kæder. **Angiv** i den forbindelse antagelser, hypoteser og signifikanssandsynlighed.

Vink: Vælg tre kæder hvis middelværdier jf. delspg. a) ligger tæt på hinanden og udfør et statistisk test kun med data for disse tre kæder.

OPGAVE 5 (VOLUMEN, SAMLET ANALYSE)

Mange faktorer påvirker salget af Royal Export, og der er således behov for en samlet vurdering af de enkelte faktorerers påvirkning. Vi ser derfor på en lineær regressionsmodel med variabelen *VolumeRoyal* som responsvariabel og variablene *PrisRoyal*, *Kæde*, *Pakkestørrelse*, *VolumeCarlsberg*, *VolumeTuborg*, *VolumeHeineken*, *ShPromoCarlsberg*, *ShPromoTuborg*, *ShPromoHeineken*, *ShPromoRoyal* som forklarende variable.

- 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%.

- b). **Gør rede for** om den reducerede regressionsmodel ville indeholde flere eller færre forklarende variable, hvis vi ændrede signifikansniveauet fra det sædvanlige 5% til 1% og **forklar** hvorfor.
- 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. pr. liter Royal Export og **giv** en fortolkning af intervallet.
- e). **Beregn** et 95%-konfidensinterval for betydningen af en prisstigning på 0,50 kr. pr. liter Royal Export.
- f). **Beregn** på baggrund af den estimerede reducerede regressionsmodel det forventede salg (i liter) af Royal Export 6 stk-pakker i Føtex i en given uge, forudsat at regressionsmodellens øvrige forklarende variable sættes lig deres respektive middelværdier (beregnet på baggrund af hele datasættet).
- Vink: Ved brug af "Profiler" i JMPs "Fit Model" regressionsoutput sættes alle numeriske forklarende variable pr. default lig deres respektive middelværdier beregnet på baggrund af hele datasættet.*
- g). **Gør rede for** hvordan det forventede salg i delspg. f) ændres, såfremt der i stedet er tale om salget af 24 stk.-pakker.
- Gør rede for** hvordan den forventede salg i delspg. f) ændres, såfremt der i stedet er tale om salget i SuperBrugsen.



OPGAVE 6 (VOLUMEN OG REKLAME)

Reklamekampagner er i mange sammenhænge et yderst effektivt middel til at øge salget af et givet produkt. For at få en mere detaljeret forståelse af, præcis hvor meget reklamekampagner⁴ påvirker salget af Royal Export, ser vi på fordelingen af salget af Royal Export (variablen *VolumeRoyal*) i 24 stk-pakker i discountbutikker (*Pakkestørrelse*="24 stk", *Kæde*="Discount") i henholdsvis almindelige uger og kampagneuger, som angivet i variablen *UgeRoyal*.

- a). **Tegn** et histogram over fordelingen af salget af Royal Export i 24 stk-pakker i discountbutikker i en kampagneuge. *Vink: Sæt f.eks. alle tre variable Kæde, Pakkestørrelse, UgeRoyal under "By" i JMPs "Distribution"-funktion.*
Estimér middelværdi og standardafvigelse i den normalfordeling, der bedst beskriver salget af Royal Export i 24 stk.-pakker i discountbutikker i en kampagneuge.
- b). **Beregn** på baggrund af den i delspg. a) beregnede normalfordeling sandsynligheden for, at der sælges mindst 20.000 liter Royal Export i 24 stk-pakker i discountbutikker i løbet af en kampagneuge.
Beregn på baggrund af den i delspg. a) beregnede normalfordeling sandsynligheden for, at der sælges mellem 20.000 og 70.000 liter Royal Export i 24 stk-pakker i discountbutikker i løbet af en kampagneuge.
- c). **Beregn** på baggrund af en passende valgt normalfordeling sandsynligheden for, at der sælges mellem 20.000 og 70.000 liter Royal Export i 24 stk-pakker i discountbutikker i løbet af en almindelig uge.

OPGAVE 7 (REKLAME OG KÆDE)

De forskellige supermarkeds kæder henvender sig til delvist forskellige kundesegmenter, og som følge heraf er det naturligt at forvente, at der bl.a. er forskel i deres valg af markedsføringsstrategi. Det er derfor interessant at undersøge, om der reelt er forskel på, hvor ofte Royal Export er på tilbud (dvs. sælges til rabatpris) i de forskellige kæder.

- a). **Opstil** en krydstabel for de to variable *Kæde* og *UgeRoyal*.
Gør rede for hvilken af de betragtede supermarkeds kæder, der i gennemsnit har færrest kampagneuger på et år.
- b). **Gør rede for** om der er statistisk belæg for en nulhypotese om, at der er uafhængighed mellem hvilken kæde, omsætningen stammer fra, og om omsætningen stammer fra en kampagneuge eller ej. **Angiv** i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.

⁴I hele denne eksamensopgave opfattes "reklamekampagne" som synonymt med en midlertidig reduktion i den normale salgspris.



- c). **Gør rede for** om konklusionen i delspg. b) ændres, såfremt vi ser bort fra omsætningen i kæderne "Discount" og "Daglig/Lokal/Irma". **Angiv** i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.

OPGAVE 8 (REKLAME OG PAKKESTØRRELSE)

Ligesom de forskellige supermarkeds kæder har forskellige kundesegmenter, kan man overveje, om også de to forskellige pakkestørrelser henvender sig til forskellige segmenter. Hvis det er tilfældet, kan det være relevant med forskellige markedsføringsstrategier for de to pakkestørrelser (uafhængigt af i hvilken kæde de sælges).

- a). **Opstil** en krydstabel for de to variable *Pakkestørrelse* og *UgeRoyal*.
Estimer andelen af kampagneuger for Royal Export for hver af de to pakkestørrelser.
- b). **Beregn** et 95%-konfidensinterval for andelen af kampagneuger på et år for Royal Export i 24-stk pakker. **Angiv** antagelser bag konfidensintervallet og **giv** en fortolkning af intervallet.
- c). **Gør rede for** om der er statistisk belæg for en nulhypotese om, at der ikke er forskel på andelen af kampagneuger for de to pakkestørrelser. **Angiv** i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.
Gør rede for hvad der ud fra krydstabellen ser ud til at være forklaringen på, at nulhypotesen accepteres/forkastes.

OPGAVE 9 (PRIS OG REKLAME)

- a). **Gennemfør** en passende valgt analyse som undersøger, hvorvidt datamaterialet understøtter en påstand om, at der er forskel på markedsføringsstrategierne for Royal Export og Heineken Premium.

*Vink: Man kan f.eks. basere sin analyse på de to variable *ShPromoRoyal* og *ShPromoHeineken* ELLER man kan i stedet basere sin analyse på de to variable *UgeRoyal* og *UgeHeineken*.*



DATAKILDE

Datamaterialet i ovenstående opgaver er venligst stillet til rådighed af bryggeriet **Royal UnibrewTM** (royalunibrew.com). Virksomheden giver følgende beskrivelse af sig selv:

We produce, market, sell and distribute quality beverages with focus on branded products within beer, malt beverages and soft drinks as well as cider and long drinks.

Our main markets are Denmark, Finland, Italy and Germany as well as Latvia, Lithuania and Estonia. To these should be added the international markets comprising a number of established markets in the Americas region and major cities in Europe and North America as well as emerging markets in for example Africa.

In our multi-beverage markets, we offer our customers strong and locally based brands. Based on continuous development and innovation, it is our objective to meet customer demand for quality beverages.

In addition to our own brands, we offer licence-based international brands of the PepsiCo and Heineken Groups.

Datamaterialet er en del af Nielsen ScanTrack DagligvareindeksTM (nielsen.com).

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.

