



BAGGRUND OG DATABESKRIVELSE

Denne opgave undersøger, hvorledes det ved hjælp af simple grafiske og numeriske værktøjer er muligt at skabe overblik over et kompliceret datamateriale. Opgaven omhandler møbelvirksomheden Montana, der producerer og sælger sine møbler i en række lande verden over.

Opgaven tager udgangspunkt i datamaterialet i filen `MontanaVaresalg.jmp`. Datamaterialet indeholder information om Montanas månedsvise varesalg for regnskabsåret 2015-2016 (perioden 1/7 2015 til 30/6 2016) opgjort i antal solgte varer og opdelt på marked, produktgruppe, varegruppe og varenummer.

Bemærk:

I opgave 1 og 2 nedenfor opgør vi for nemheds skyld ikke virksomhedens salg i antal solgte varer men i stedet som antallet af varenumre, hvorefter der er solgt mindst ét styk i en given måned (hvert varenummer hvorefter der er solgt mindst ét styk i en given måned svarer til én række i datafilen `MontanaVaresalg.jmp`). Det faktiske salg i antal styk er indeholdt i variablen `AntalVarer`, som behandles nærmere i opgave 3 og 4.



EKSEMPEL 1 (SALG FORDELT PÅ PRODUKTGRUPPER)

- a). **Tegn** et søjlediagram over fordelingen af virksomhedens salg opdelt på produktgrupper (variablen *Produktgruppe*).
Bestem hvor stor en andel produktgruppen “12 mm reoler” udgør af virksomhedens samlede salg.
- b). **Tegn** et cirkeldiagram over fordelingen af virksomhedens salg opdelt på produktgrupper.
Angiv på det optegnede cirkeldiagram de procentvise andele hørende til hver produktgruppe.

EKSEMPEL 2 (SALG FORDELT PÅ MARKEDER)

- a). **Tegn** et søjlediagram over fordelingen af virksomhedens salg på de forskellige markeder (variablen *Marked*).
Angiv virksomhedens 5 største markeder.
Angiv hvor stor en andel hvert af disse 5 markeder udgør af det samlede salg, samt hvor meget de tilsammen udgør af det samlede salg.
- b). **Tegn** et cirkeldiagram der viser fordelingen af salget på de 5 største markeder (dvs. betragt kun salget af varer på de 5 største markeder).
Angiv hvor stor en andel salget på det danske marked udgør af salget på de 5 største markeder.

EKSEMPEL 3 (SALG FORDELT PÅ MÅNEDER)

- a). **Tegn** for hver måned et histogram og et boxplot af salget per varenummer (dvs. tegn figurer af variablen *AntalVarer* opdelt efter værdien af variablen *Måned*).
 NB: Det er tilstrækkeligt at angive figurerne hørende til april måned i besvarelsen.
- b). **Lav** en summarisk fordelingsbeskrivelse af salget per varenummer i april måned (inddrag relevante nøgletal samt figurerne fra delspg. a)).
- c). **Angiv** og fortolk 1. kvartil, median og 3. kvartil i fordelingen af salget per varenummer i april måned.
Angiv hvor mange varer per varenummer der *mindst* blev solgt i april måned blandt de 10% *bedst* sælgende varenumre.



Eksempel 4 (Salg for udvalgte varenumre)

I denne opgave ser vi udelukkende på virksomhedens salg af de 2 populære varenumre 001111 og 001112 (to af modulerne i 12 mm reolsystemet) på de 3 markeder Norge, Sverige og Tyskland.

- a). **Tegn** et histogram og et boxplot af det samlede månedlige salg af de to varenumre.
- b). **Lav** en summarisk fordelingsbeskrivelse af det månedlige salg af de to varenumre (inddrag relevante nøgletal samt figurerne fra delspg. a)).
- c). **Gør rede for** hvorvidt det er muligt at bruge den empiriske regel til beskrivelse af det månedlige salg af de to varenumre.

Gør rede for hvad den empiriske regel fortæller om forventningerne til det månedlige salg for hvert af de to varenumre (idet vi her forudsætter, at den empiriske regel finder anvendelse).



DATAKILDE

Datamaterialet i ovenstående eksempler er venligst stillet til rådighed af møbelvirksomheden **Montana** (montana.dk). Virksomheden giver følgende beskrivelse af sig selv:

Montana Møbler A/S producerer og markedsfører Montana reolen samt borde, stole og baderumsmøbler til hjem og moderne kontormiljøer. Vi har hovedkontor og egen produktion i højteknologiske produktionslokaler i Haarby på Fyn, mens vores showroom og afdeling for Design & Communication er at finde i Pakhus 48 i Københavns Frihavn. Derudover har vi egen salgsorganisation i Danmark og på de fleste europæiske markeder. Vi beskæftiger i dag ca. 160 medarbejdere.

DISCLAIMER:

Datamaterialet i nærværende eksempler må udelukkende anvendes i forbindelse med undervisningen i 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.

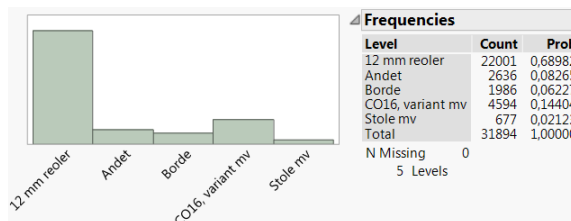


VEJLEDENDE LØSNINGER

EKSEMPEL 1 (► Løsning i JMP)

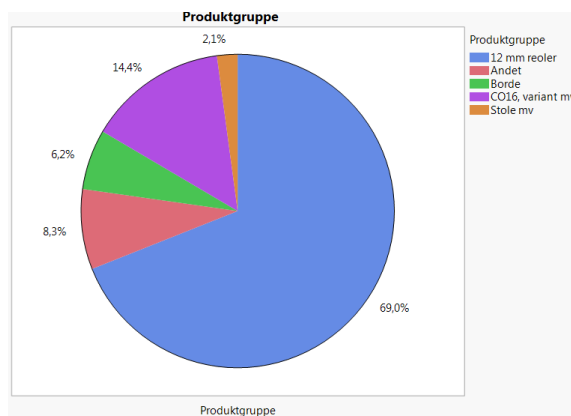
1A):

Søjlediagrammet over salget fordelt på produktgrupper viser, at størstedelen af virksomhedens salg (målt på antal varenumre) ligger indenfor produktgruppen "12 mm reoler". Produktgruppen udgør 69,0% af det samlede salg.



1B):

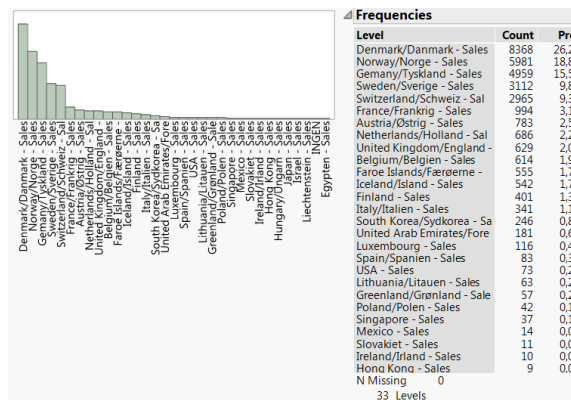
Repræsenteres salgets fordeling på produktgrupper i stedet i et cirkeldiagram ser det således ud.



EKSEMPEL 2 (► Løsning i JMP)

2A):

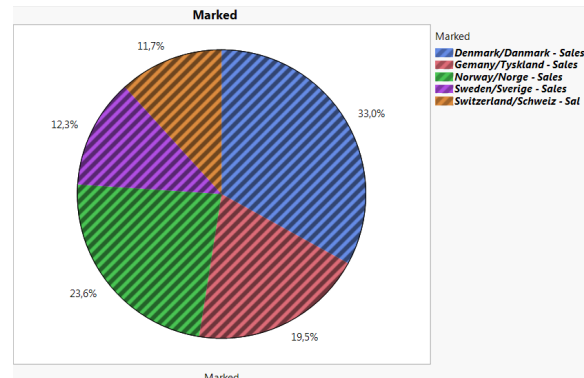
Søjlediagrammet over salget fordelt på markeder viser, at størstedelen af virksomhedens salg (målt på antal varenumre) ligger i Danmark (andel: 26,2%), Norge (andel: 18,8%), Tyskland (andel: 15,5%), Sverige (andel: 9,8%) og Schweiz (andel: 9,3%). Tilsammen udgør salget på de 5 markeder $26,2\% + 18,8\% + 15,5\% + 9,8\% + 9,3\% = 79,6\%$ af det samlede salg.





2B):

Fordelingen af salget på de fem største markeder viser, at salget i Danmark udgør 33,0% ($= \frac{26,2\%}{79,6\%}$) af det samlede salg på de fem største markeder.

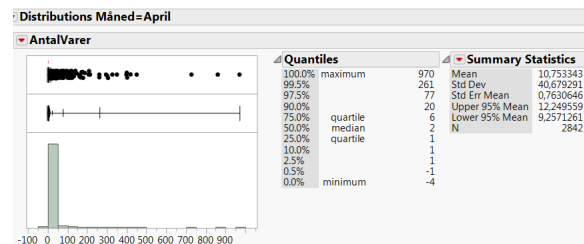


Eksempel 3 (► Løsning i JMP)

3A):

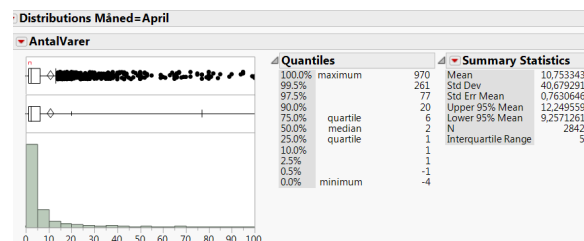
Histogram og boxplots af antallet af solgte varer per varenummer i løbet af april måned viser, at der kun sælges ganske få styk af langt størstedelen af virksomhedens varenumre, mens der sælges mange enheder af nogle enkelte varenumre.

NB: Det er tilstrækkeligt blot at tegne ét af de to boxplots.



3B):

Eftersom salget indenfor langt de fleste varenumre ikke overstiger 50-100 enheder, kan det være nyttigt - som supplement til figurerne i delspg. a) - at "zoome" ind på fordelingen i dette interval.



Histogrammet bekræfter indtrykket fra delspg. a) af en meget højreskæv fordeling, hvilket også understreges af, at det gennemsnitlige antal solgte enheder per varenummer (10,8 styk) langt overstiger median (2 styk) antal solgte enheder per varenummer. Det er umiddelbart svært at se ret meget ud af boxplottene, fordi de ligger koncentreret tæt inde omkring 0.



Fordelingen af solgte antal varer per varenummer har en lang, men ikke særligt tung højre hale, der primært skyldes salget indenfor tre varenumre: 727 styk af varenummer 001112, 860 styk af varenummer 0602 og 970 styk af varenummer 00006X. Disse tre observationer adskiller sig så markant fra alle øvrige, at de naturligt bør betragtes som outliers.

På grund af de tre outliers har fordelingen en meget stor standardafvigelse (40,7), der langt fra er retvisende for variationen i datamaterialet. I stedet indikerer interkvartilbredden (5), at langt størstedelen af varenumrene ikke varierer ret meget i antallet af solgte enheder. Det understreges yderligere af at 3. kvartil kun er på 6 styk, dvs. at mindst 75% af varenumrene maksimalt har solgt 6 styk i løbet af april måned.

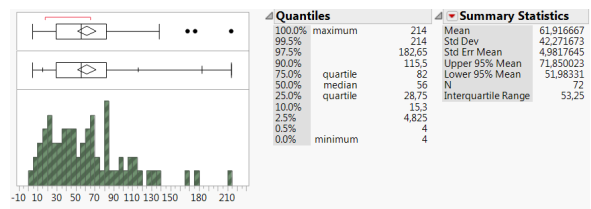
3c):

Solgte antal varer per varenummer har en 1. kvartil på $q_{25\%} = 1$, en median på $q_{50\%} = 2$ og en 3. kvartil på $q_{75\%} = 6$. Af mindst 25% af virksomhedens varenumre blev der således kun solgt 1 styk i løbet af april måned. Af mindst halvdelen af varenumrene blev der højst solgt 2 styk og af mindst 75% blev der højst solgt 6 styk (dvs. af højst 25% af varenumrene blev der solgt mere end 6 styk). Endelig bemærkes det, at de 10% bedst sælgende varenumre solgte mindst 20 styk hver.

EKSEMPEL 4 (► Løsning i JMP)

4A):

Histogram og boxplots af det samlede antal solgte enheder af de to varenumre 001111 og 001112 i Tyskland, Norge og Sverige i løbet af april måned viser, at der kun sælges ganske få styk af langt størstedelen af virksomhedens varenumre, mens der sælges mange enheder af nogle enkelte varenumre.



**4b):**

Histogrammet ovenfor viser en fordeling med en vis højreskævhed, omend ikke værre end at det gennemsnitlige antal solgte enheder (61,9 styk) er omtrent lig median (56 styk) antal solgte enheder. Det er umiddelbart svært at se ret meget ud af boxplottene, fordi de ligger koncentreret tæt inde omkring 0.

Fordelingen af solgte antal varer per varenummer har en lang, men ikke særligt tung højre hale, der primært skyldes tre observationer, der alle vedrører varenummer 001112: 168 styk solgt i Norge i november måned, 176 styk solgt i Tyskland i september måned og 214 styk solgt i Tyskland i oktober måned. Disse tre observationer adskiller sig fra alle øvrige, omend forskellen ikke er større, end at de ikke naturligvis kan betragtes som outliers.

Hvis vi ser bort fra de tre observationer i den højre hale er fordelingen nogenlunde symmetrisk, hvilket også understreges af de to boxplots, der er nogenlunde symmetriske omkring medianen. Ligeledes viser histogrammet svage (men kun svage!) tegn på klokkeform. Både standardafvigelsen (42,3) og interkvartilbredden (53,3) er dermed fornuftige mål for variationen i datamaterialet.

4c):

Tendensen til klokkeform i histogrammet samt indikationerne på en vis grad af symmetri (når vi ser bort fra de tre observationer længst ude i højre hale) gør, at man den empiriske regel ikke kan forventes at give en retvisende beskrivelse af fordelingen.

Hvis vi alligevel forsøger at anvende den empiriske regel vil vi forvente, at det månedlige antal solgte enheder af hvert af de to varenumre 001111 og 001112...

- med ca. 68% sandsynlighed vil ligge i intervallet $[61,9 - 42,3; 61,9 + 42,3] = [19,6\text{stk.}, 104,2\text{stk.}]$
- med ca. 96% sandsynlighed vil ligge i intervallet $[61,9 - 2 \cdot 42,3; 61,9 + 2 \cdot 42,3] = [-22,7\text{stk.}, 146,5\text{stk.}]$
- med næsten 100% sandsynlighed vil ligge i intervallet $[61,9 - 3 \cdot 42,3; 61,9 + 3 \cdot 42,3] = [-65,0\text{stk.}, 188,8\text{stk.}]$



Det giver selvsagt ikke megen mening at tale om systematisk negative salg (der kan være tale om enkelte returvarer, men derudover giver negativt salg ikke mening). Det faktum at de ovenfor beregnede intervaller indikerer negativt salg illustrerer blot, at fordelingen ikke er tilstrækkeligt pæn symmetrisk og klokkeformet til, at den empiriske regel for alvor finder anvendelse.