



### BAGGRUND OG DATABESKRIVELSE (EKSEMPEL 1-2)

Denne opgave beskæftiger sig med omsætningen af skrabespil på det danske marked, nærmere bestemt salget af de fire skrabespil “Million Quick”, “Mini Quick”, “Dobbelt Bogstavjagt” og “Quick Ekstra Månedsløn” udbudt af Danske Spil.

Opgaven tager udgangspunkt i datamaterialet i filen `DanskeSpil.jmp`. Datamaterialet indeholder information om den ugentlige omsætning (i 1.000 kr) for hver af de fire skrabespil i perioden uge 1, 2014 til uge 36, 2015. Salget er opgjort både som det samlede salg (“i alt”) såvel som opdelt på salget i henholdsvis dagligvarebutikker (“dvh”) og “convenience”-butikker (“conv”), dvs. tankstationer, kiosker etc.

### Eksempel 1 (Million Quick)

- Tegn** et histogram over fordelingen af den samlede ugentlige omsætning af “Million Quick” (*Million Quick (i alt)*).  
**Estimér** middelværdi og standardafvigelse i den normalfordeling, der bedst beskriver den samlede ugentlige omsætning af “Million Quick”.
- Beregn** et 95%-konfidensinterval for middelværdien i fordelingen af den samlede ugentlige omsætning af “Million Quick” og **giv** en fortolkning af intervallet.
- Gør rede for** hvordan konfidensintervallet i delspg. b) ændres, såfremt middelværdien kun estimeres på baggrund af salget over 12 uger (og alle andre størrelser antages uændrede).
- Gør rede for** hvordan konfidensintervallet i delspg. b) ændres, såfremt den ugentlige omsætning hver uge var 100.000 kr. højere (og alle andre størrelser antages uændrede).

### Eksempel 2 (Dobbelt Bogstavjagt)

- Estimér** middelværdi og standardafvigelse i den normalfordeling, der bedst beskriver den samlede ugentlige omsætning af “Dobbelt Bogstavjagt” (*Dobbelt Bogstavjagt (i alt)*).  
**Beregn** på baggrund af den fundne normalfordeling gennemsnitligt hvor mange uger om året, den samlede ugentlige omsætning overstiger 600.000 kr.



- b). **Beregn** et 95%-konfidensinterval for middelværdien i fordelingen af den samlede ugentlige omsætning af "Million Quick".
- c). **Gør rede for** hvordan konfidensintervallet i delspg. b) ændres, såfremt konfidensniveauet ændres til 99%.
- d). **Gør rede for** hvor mange ugers omsætningsdata der er nødvendig, hvis den forventede samlede ugentlige omsætning med 95% sandsynlighed skal kunne bestemmes med en usikkerhed på plus/minus 25.000 kr.

#### DATAKILDE (OPGAVE 1-2)

Datamaterialet i ovenstående opgaver er venligst stillet til rådighed af markedsanalysevirksomheden Nielsen og er en del af Nielsen ScanTrack Dagligvareindeks™ (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 opgaver 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.*



### BAGGRUND OG DATABESKRIVELSE (EKSEMPEL 3)

Denne opgave beskæftiger sig med befolkningens holdning til en eventuel samfundsøkonomisk gevinst ved at flere unge gennemfører en videregående uddannelse. Opgaven tager udgangspunkt i filen `VideregåendeUddannelse.jmp`, hvis indhold er beskrevet på eksempelark 4.

### EKSEMPEL 3 (TID UNDER UDDANNELSE)

- a). **Estimér** sandsynligheden for at være enig i spørgsmål 4 (*Spg 4 (grupperet)*).  
**Beregn** et 95%-konfidensinterval for sandsynligheden for at være enig i spørgsmål 4 og **giv** en fortolkning af intervallet.
- b). **Gør rede for** hvordan konfidensintervallet i delspg. a) ændres, såfremt konfidensniveauet i stedet ændres til 80%.  
**Gør rede for** hvordan konfidensintervallet i delspg. a) ændres, såfremt antallet af adspurgte i stedet halveres (og alle andre størrelser antages uændrede).
- c). **Beregn** hvor mange personer, der skal svare på spørgsmål 4, for at sandsynligheden for at være enig i spørgsmålet med 95% sandsynlighed kan bestemmes med en usikkerhed på plus/minus 2%.  
**Beregn** hvor mange personer, der skal svare på spørgsmål 4, for at sandsynligheden for at være enig i spørgsmålet med 99% sandsynlighed kan bestemmes med en usikkerhed på plus/minus 2%.

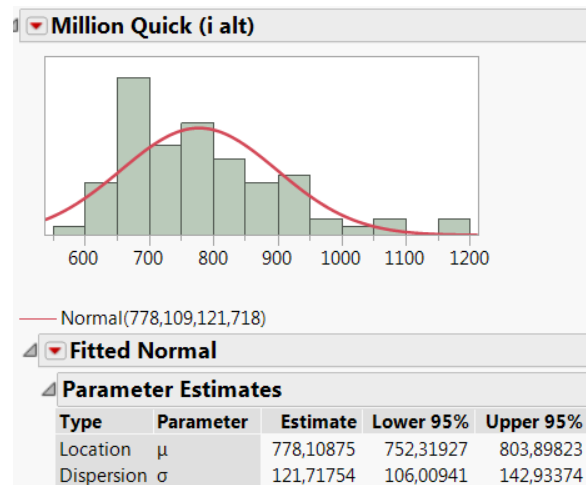


## VEJLEDENDE LØSNINGER

### EKSEMPEL 1 (► Løsning i JMP)

1A):

Den normalfordeling der bedst beskriver den samlede ugentlige omsætning af "Million Quick" er givet ved  $\hat{\mu} = 778,109$  og  $\hat{\sigma} = 121,718$ .



1B):

Et 95%-konfidensinterval for den forventede ugentlige omsætning af "Million Quick" er givet ved [752.319 kr.; 803.898 kr.]. Med 95% sandsynlighed vil den forventede ugentlige omsætning af "Million Quick" således ligge mellem 752.319 kr. og 803.898 kr.

1c):

Hvis den forventede ugentlige omsætning kun estimeres ud fra 12 ugentlige observationer, så bliver konfidensintervallet brede. Det skyldes, at estimationen nu baseres på færre observationer (datasættet har 88 observationer), dvs. på mindre information, og dermed alt andet lige bliver mere usikker.

Summary Information	
Sample Average	778,109
Sample Standard Deviation	121,718
Sample Size	12
Confidence Interval	
Confidence Level	0,95
Result	Value
t multiplier	2,20099
Standard Error of the Mean	35,137
Lower Limit	700,773
Upper Limit	855,445

Helt specifikt, så ændres 95%-konfidensintervallet til [700.773 kr.; 855.445 kr.].



**1d):**

Hvis den ugentlige omsætning af "Million Quick" hver uge var 100.000 kr. højere ville det betyde, at estimatet af den forventede omsætning ( $\mu$ ) ville være 100 højere, mens det ikke ville påvirke variationen ( $\sigma$ ) i omsætningen fra uge til uge (fordi omsætningen stiger med 100.000 kr. hver eneste uge).

Normalfordelingens estimerede parametre ville dermed ændres til

$$\hat{\mu} = 878,109 \quad \hat{\sigma} = 121,718$$

og 95%-konfidensintervallet for  $\mu$  ville derfor ændres til

$$[852.319 \text{ kr.}; 903.898 \text{ kr.}]$$

fordi man ud fra formelen for konfidensintervallet kan se, at der blot skal lægges 100 til hvert af intervallets to endepunkter (idet  $\hat{\sigma}$ ,  $\alpha$  og  $n$  er uændrede).

$$\left[ \hat{\mu} + t_{\alpha/2}(n-1) \cdot \frac{\hat{\sigma}}{\sqrt{n}}; \quad \hat{\mu} - t_{\alpha/2}(n-1) \cdot \frac{\hat{\sigma}}{\sqrt{n}} \right]$$



## Eksempel 2 (► Løsning i JMP)

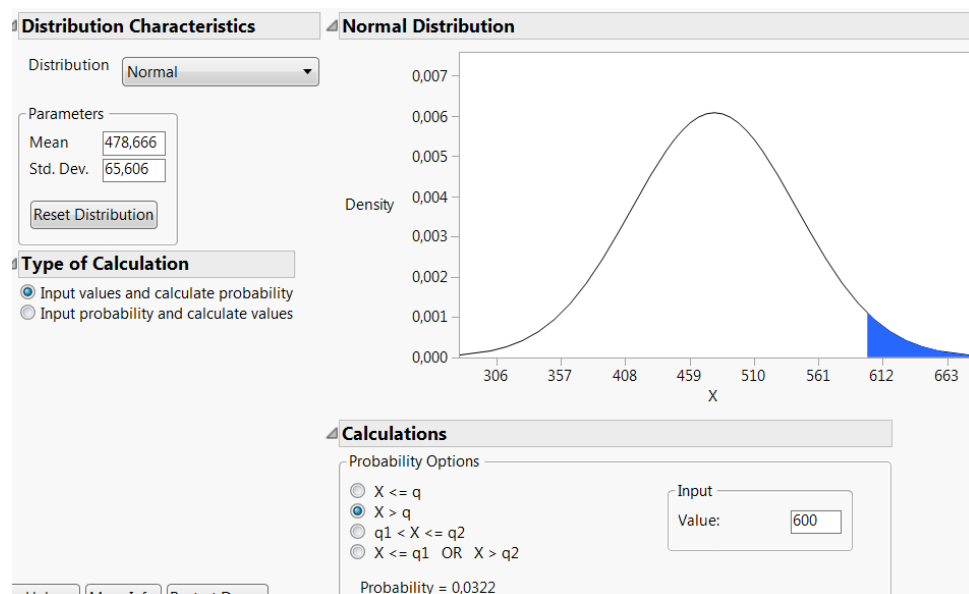
2A):

Den normalfordeling der bedst beskriver den samlede ugentlige omsætning af "Million Quick" er givet ved  $\hat{\mu} = 478,666$  og  $\hat{\sigma} = 65,606$ .

Fitted Normal				
Parameter Estimates				
Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	$\mu$	478,66636	464,76586	492,56686
Dispersion	$\sigma$	65,605612	57,138951	77,04112

Sandsynligheden for at den samlede ugentlige omsætning overstiger 600.000 kr. er 3,22%.

Det betyder at 3,22% af tiden, svarende til ca. 2 uger om året ( $3,22\% \cdot 52 = 1,67 \approx 2$ ), vil omsætningen overstige 600.000 kr.



2B):

Et 95%-konfidensinterval for den forventede samlede ugentlige omsætning af "Dobbelt Bogstavjagt" er givet ved [464.766 kr.; 492.567 kr.].



2c):

Hvis konfidensintervallet for den forventede ugentlige omsætning i stedet beregnes med et konfidensniveau på 99%, så bliver konfidensintervallet bredere. Det skyldes, at intervallet nu med større sikkerhed (99% frem for 95%) skal indeholde den sande (men ukendte) middelværdi  $\mu$ .

Helt specifikt, så bliver 99%-konfidensintervallet [460.249 kr.; 497.084 kr.].

Confidence Intervals				
Parameter	Estimate	Lower CI	Upper CI	1-Alpha
Mean	478,6664	460,2486	497,0842	0,990
Std Dev	65,60561	54,79445	81,21096	0,990

2d):

Hvis den forventede ugentlige omsætning med 95% sandsynlighed skal kunne bestemmes med en usikkerhed på  $\pm 25.000$  kr., så kræver det 27 ugers omsætningsdata.

Sample Size for Estimation of the Mean	
Inputs	
Confidence Level	0,95
Population Std. Dev. (Planning Value)	65,606
Desired Margin of Error (C.I. 1/2-Width)	25
Results	
Calculated Value	26,4547
Calculated Value (rounded up)	27

**Eksempel 3** (► Løsning i JMP)**3A):**

Sandsynligheden for at være enig i spørgsmål 4 estimeres til  $\hat{p} = 62,6\%$ .

Et 95%-konfidensinterval for denne sandsynlighed er givet ved  $[57,6\%; 67,4\%]$ .

Med 95% sandsynlighed vil sandsynligheden for at en tilfældigt valgt person er enig i spørgsmål 4 således ligge mellem 57,6% og 67,4%. Alternativt formuleret: Med 95% sandsynlighed vil andelen af den danske befolkning, der er enig i spørgsmål 4, ligge mellem 57,6% og 67,4%.

Spg. 4 (grupperet)

Frequencies

Level	Count	Prob
Enig	233	0,62634
Uenig	139	0,37366
Total	372	1,00000

N Missing 142

2 Levels

Confidence Intervals

Level	Count	Prob	Lower CI	Upper CI	1-Alpha
Enig	233	0,62634	0,576127	0,673979	0,950
Uenig	139	0,37366	0,326021	0,423873	0,950
Total	372				





3b):

Hvis konfidensintervallet for sandsynligheden for at være enig i spørgsmål 4 i stedet beregnes med et konfidensniveau på 80%, så bliver konfidensintervallet smallere. Det skyldes, at intervallet nu med mindre sikkerhed (80% frem for 95%) skal indeholde den sande (men ukendte) sandsynlighed  $p$ .

Confidence Intervals					
Level	Count	Prob	Lower CI	Upper CI	1-Alpha
Enig	233	0,62634	0,59371	0,657867	0,800
Uenig	139	0,37366	0,342133	0,40629	0,800
Total	372				

Helt specifikt, så bliver 80%-konfidensintervallet [59,4%; 65,8%].

Hvis sandsynligheden for at være enig i spørgsmål 4 kun estimeres ud fra halvt så mange spørgeskemasvar, så bliver konfidensintervallet bredere. Det skyldes, at estimationen nu baseres på færre observationer, dvs. på mindre information, og dermed alt andet lige bliver mere usikker.

Number of Successes	116
Sample Size	186
Confidence Interval	
Confidence Level	0,95
Result	Value
Estimated Proportion	0,62366
Standard Error of Proportion	0,06962
Lower Limit	0,55403
Upper Limit	0,69328

Helt specifikt, så ændres 95%-konfidensintervallet til [55,4%; 69,3%], idet stikprøvestørrelsen nu er  $n = \frac{372}{2} = 186$ , mens antallet af tilfredse er  $\hat{p} \cdot 186 = 62,6\% \cdot 186 \approx 116$ .



3c):

Hvis sandsynligheden for at en tilfældigt valgt person er enig i spørgsmål 4 med 95% sandsynlighed skal kunne bestemmes med en usikkerhed på  $\pm 2\%$ , så kræver det spørgeskemasvar fra 2.249 personer.<sup>1</sup>

Sample Size for Estimation of the Proportion	
<b>Inputs</b>	
Confidence Level	0,95
Expected Proportion (Planning Value)	0,626
Desired Margin of Error (C.I. 1/2-Width)	0,02
<b>Results</b>	
Calculated Value	2248,44
Calculated Value (rounded up)	2249

Hvis sandsynligheden for at en tilfældigt valgt person er enig i spørgsmål 4 med 99% sandsynlighed skal kunne bestemmes med en usikkerhed på  $\pm 2\%$ , så kræver det spørgeskemasvar fra 3.884 personer.

Sample Size for Estimation of the Proportion	
<b>Inputs</b>	
Confidence Level	0,99
Expected Proportion (Planning Value)	0,626
Desired Margin of Error (C.I. 1/2-Width)	0,02
<b>Results</b>	
Calculated Value	3883,47
Calculated Value (rounded up)	3884

<sup>1</sup> Sætningens ordlyd kan være lidt forvirrende. Alternativt kan sætningen formuleres som: "Hvis andelen af befolkningen, der er enig i spørgsmål 4, med 95% sandsynlighed skal kunne bestemmes med en usikkerhed på  $\pm 0,02$ , så kræver det...".