



## BAGGRUND OG DATABESKRIVELSE (EKSEMPEL 1-2)

Denne opgave beskæftiger sig med den danske turismeindustri og ser på udviklingen i antallet af hotelovernatninger i Hovedstadsområdet. Opgaven tager udgangspunkt i datamaterialet i filen `HotelKbh.jmp`. Datamaterialet indeholder information om det månedlige antal hotelovernatninger i Hovedstadsområdet opdelt på hotelgæstens nationalitet for perioden januar 2010 til maj 2017.

### Eksempel 1 (Overnatninger)

Opgave 1 omhandler udelukkende hotelovernatninger blandt svenske og norske gæster.

- a). Tegn for hver af de to nationaliteter det månedlige antal hotelovernatninger (*Overnatninger*) over tid (*Periode*).

Gør på baggrund af figurerne rede for udviklingen i antallet af hotelovernatninger over den betragtede tidsperiode.

- b). Tegn for hver af de to nationaliteter et histogram over fordelingen af antal hotelovernatninger samt en tilhørende normalfordelingskurve.

Gør på baggrund af figurerne rede for om normalfordelingen kan forventes at give en rimelig beskrivelse af fordelingen af antal hotelovernatninger over den betragtede tidsperiode.

- c). Tegn for hver af de to nationaliteter et histogram over fordelingen af ændringen i antal overnatninger i forhold til samme måned året før (*Overnatninger (ændring ift sidste år)*) samt en tilhørende normalfordelingskurve.

Gør på baggrund af figurerne rede for om normalfordelingen ser ud til at give en rimelig beskrivelse af fordelingen af ændringen i antal hotelovernatninger.

### Eksempel 2 (Ændring i Overnatninger)

Eksempel 2 omhandler udelukkende hotelovernatninger blandt svenske gæster.

- a). **Estimer** middelværdi og standardafvigelse i den normalfordeling, der bedst beskriver ændringen i antal hotelovernatninger blandt svenske gæster i forhold til samme måned året før (*Overnatning (ændring ift sidste år)*).

- b). **Beregn** på baggrund af en passende valgt normalfordeling sandsynligheden for, at antal hotelovernatninger blandt svenske gæster stiger med mellem 2.000 og 4.000 i



forhold til samme måned året før.

**Beregn** på baggrund af en passende valgt normalfordeling sandsynligheden for, at antal hotelovernatninger blandt svenske gæster falder med mindst 5.000 i forhold til samme måned året før.

- c). **Beregn** hvilket niveau ændringen i antal hotelovernatninger blandt svenske gæster overstiger 1 ud af 4 måneder om året (dvs. beregn 75%-fraktilen).

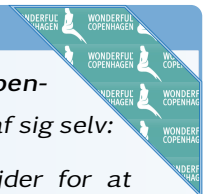
#### DATAKILDE

Datamaterialet i ovenstående eksempler er venligst stillet til rådighed af **Wonderful Copenhagen** ([copenhagen.tourismbarometer.com](http://copenhagen.tourismbarometer.com)). Virksomheden giver følgende beskrivelse af sig selv:

*“Wonderful Copenhagen er hovedstadsregionens turismeorganisation, som arbejder for at fremme og udvikle erhvervs- og ferieturismen i Region Hovedstaden til gavn for almennyttige interesser.*

*Derudover er Wonderful Copenhagen en netværksorganisation med over 300 kommercielle samarbejdspartnere og arbejder på at skabe forretning for aktørerne inden for turismebranchen og oplevelsesindustrien i København og hele Region Hovedstaden.*

*Wonderful Copenhagen er en erhvervsdrivende fond og finansieres gennem bidrag fra erhvervslivet, organisationer og det offentlige, herunder flerårige finansieringsaftaler med Region Hovedstaden.”*





### 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*, der indeholder svar på fem konkrete spørgsmål vedrørende sammenhængen mellem uddannelse og samfundsøkonomi. Data-materialet er indsamlet i september 2014 af analysevirksomheden Webpol og indeholder svar fra i alt 514 respondenter på følgende spørgsmål:

- 1) *Er du enig eller uenig i, at det vil øge den økonomiske vækst i samfundet, såfremt flere unge gennemfører en videregående uddannelse?*
- 2) *Er du enig eller uenig i, at man gennem en videregående uddannelse tilegner sig brugbar viden, som man ikke lige såvel kan opnå gennem praktisk erfaring?*
- 3) *Er du enig eller uenig i, at Statens Uddannelsesstøtte (5.839 kr. pr. måned før skat for udeboende) er for lav til, at man som studerende kan klare sig økonomisk gennem en videregående uddannelse uden at have et arbejde ved siden af?*
- 4) *Er du enig eller uenig i, at studerende på en videregående uddannelse vil komme hurtigere gennem uddannelsen, såfremt der er en økonomisk gevinst forbundet med at gennemføre på normeret tid?*
- 5) *Er du enig eller uenig i, at man ved at tage en videregående uddannelse styrker sine egne forudsætninger for social og økonomisk velstand på lang sigt?*

For hvert af de fem spørgsmål er svarene angivet i to variable: en variabel der angiver respondenternes oprindelige svar (med de seks svarmuligheder “meget enig”, “enig”, “hverken enig eller uenig”, “uenig”, “meget uenig”, “ved ikke”) og en grupperet variable (hvor svarmulighederne “meget enig” og “enig” samles i værdien “enig”, svarmulighederne “uenig” og “meget uenig” samles i værdien “uenig”, mens der ses bort fra de sidste to svarmuligheder “hverken enig eller uenig” og “ved ikke”).

### Eksempel 3 (Statens Uddannelsesstøtte (SU))

- a). **Angiv** passende valgte parametre i en binomialfordeling, der beskriver fordelingen af de adspurgtes holdning til spørgsmål 3 (*Spg 3 (grupperet)*).



- b). **Diskutér** i hvilket omfang antagelser bag anvendelsen af binomialfordelingen i delspg. a) kan antages at være opfyldte.
- c). **Beregn** sandsynligheden for at mere end 400 ud af 500 tilfældigt udvalgte danskere mener, at SU'en er for lav.
- Beregn** sandsynligheden for at netop 400 ud af 500 tilfældigt udvalgte danskere mener, at SU'en er for lav.
- Beregn** sandsynligheden for at mellem 400 og 450 (begge værdier inklusiv) ud af 500 tilfældigt udvalgte danskere mener, at SU'en er for lav.

#### DATAKILDE

Datamaterialet i ovenstående eksempler er venligst stillet til rådighed af **Webpol** (webpol.dk).

Virksomheden giver følgende beskrivelse af sig selv:

*“Webpol er en analysevirksomhed, der er ekspert i online interviews og markedsanalyse. Webpol er ejet af JP/Politikens Hus og fungerer som uafhængig analysepartner for koncernen med hovedvægten på Politiken, Jyllands-Posten og Ekstra Bladet.*

*Webpol gennemfører også undersøgelser for store danske virksomheder og organisationer som for eksempel Danske Spil, RUC, Det Kongelige Teater og Nykredit.”*

#### 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.*



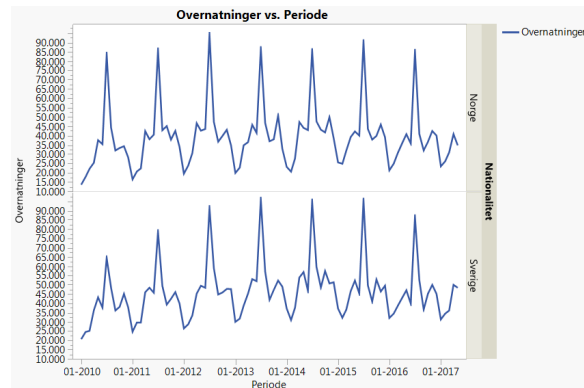


## VEJLEDENDE LØSNINGER

### EKSEMPEL 1 (► Løsning i JMP)

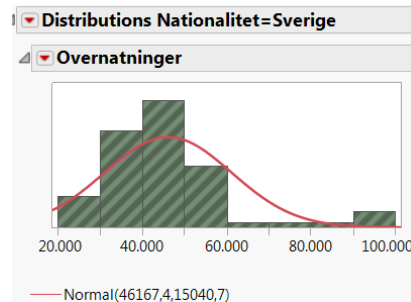
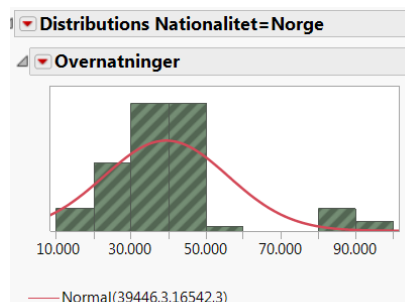
#### 1A):

Figurer over udviklingen i antal hotelovernatninger blandt svenske og norske gæster viser, at der er en meget stærk sæsonvariation med langt flere overnattende gæster om sommeren, og i særdeleshed i juli måned. Samtidig er der muligvis også en svagt stigende tendens over tid, omend den er svær at se af figurerne på grund af den stærke sæsonvariation.



#### 1B):

Histogrammer over fordelingen af antal hotelovernatninger blandt svenske og norske gæster:

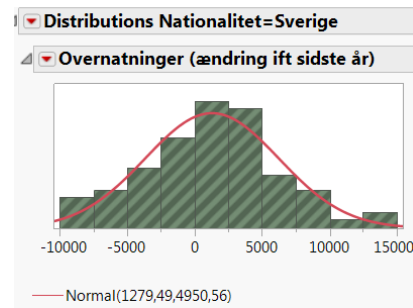
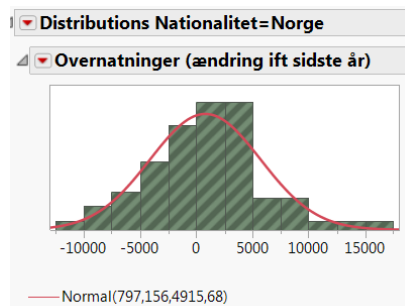


Det fremgår tydeligt, at variablenes fordeling ikke følger normalfordelingens klokkeform, hvilket blandt andet skyldes problemerne med sæson (og evt. trend) i delspg. a). Eksistens af sæson og trend gør det som udgangspunkt svært at bruge en normalfordeling med en konstant (dvs. ikke tids-varierende) middelværdi  $\mu$  til at beskrive en variabel med.



1c):

Histogrammer over fordelingen af ændringen i antal hotelovernatninger blandt svenske og norske gæster i forhold til samme måned året før:

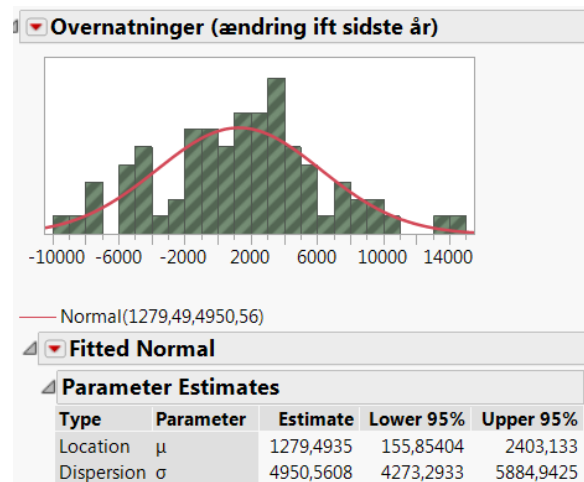


Det fremgår tydeligt, at variablenes fordeling nogenlunde følger normalfordelingens klokkeform, hvilket blandt andet skyldes, at problemer med sæson (og evt. trend) i en variabel delvist forsvinder, når vi i stedet ser på ændringer. Det betyder, at normalfordelingen får lettere ved at give en god beskrivelse.

## Eksempel 2 (► Løsning i JMP)

2a):

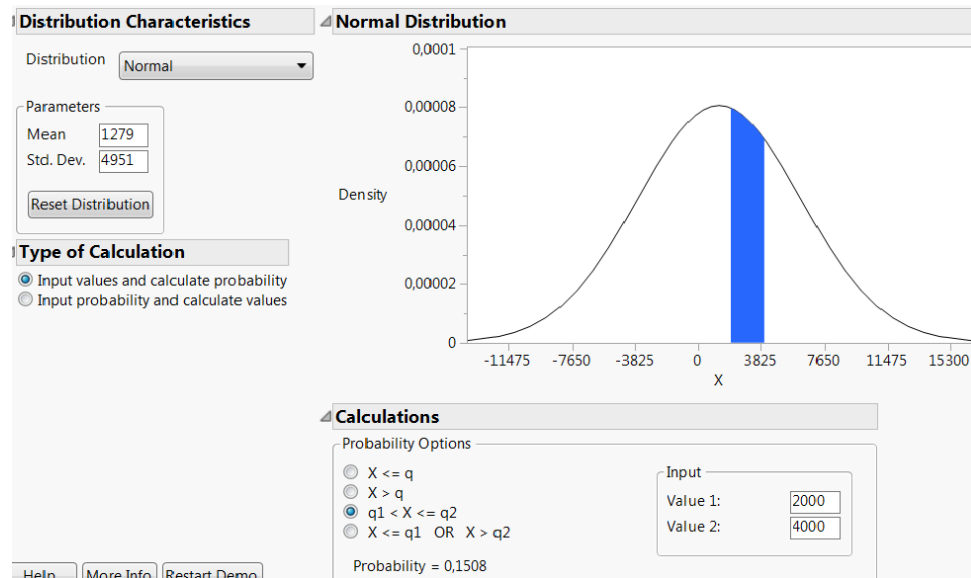
Den normalfordeling der bedst beskriver ændringen i antal overnatninger blandt svenske gæster er givet ved  $\hat{\mu} = 1.279$  og  $\hat{\sigma} = 4.951$ .





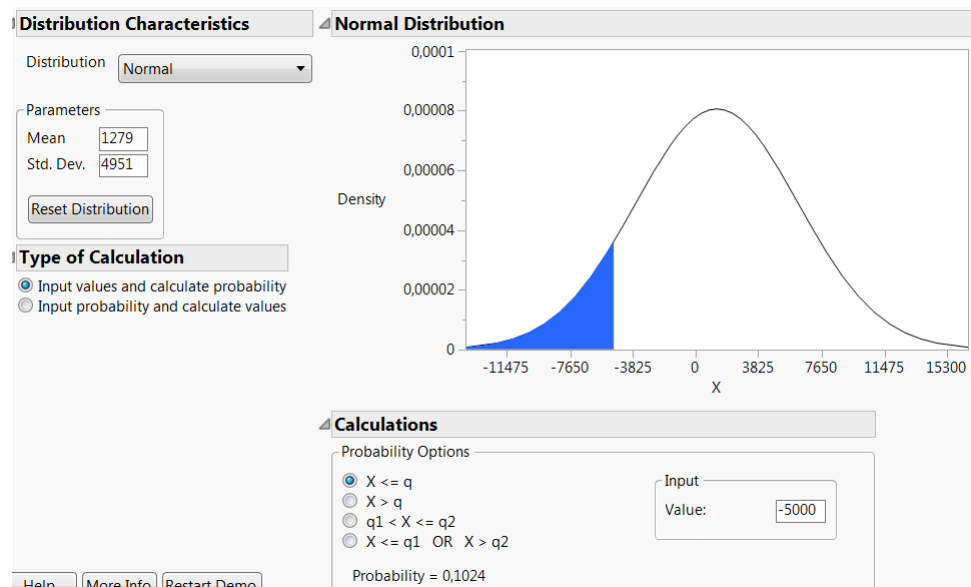
2B):

Sandsynligheden for at antallet af hotelovernatninger blandt svenske gæster stiger med mellem 2.000 og 4.000 i forhold til samme måned året før er:



$$P(2.000 \leq \text{Ændring i antal overnatninger} \leq 4.000) = 15,1\%$$

Sandsynligheden for at antallet af hotelovernatninger blandt svenske gæster falder med mindst 5.000 i forhold til samme måned året før er:

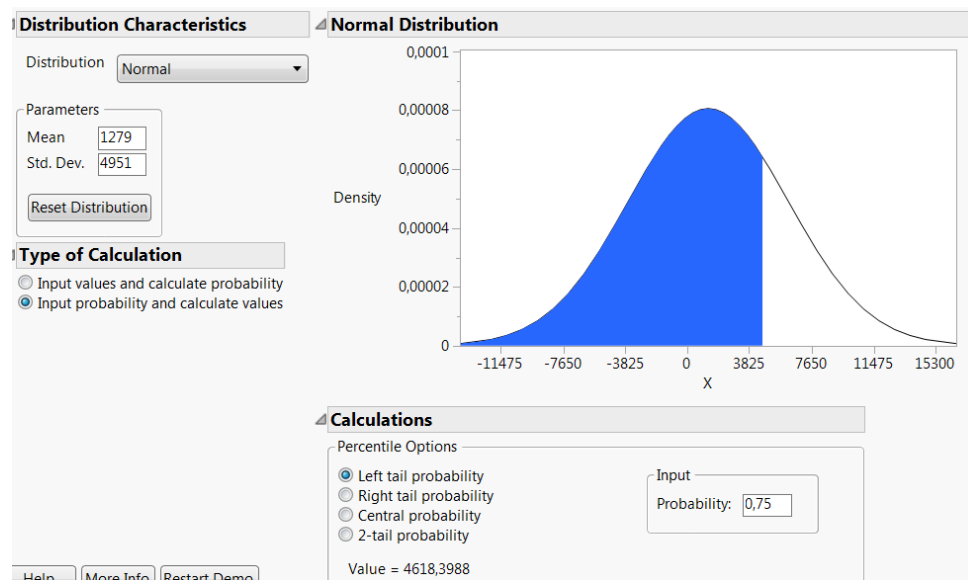


$$P(\text{Ændring i antal overnatninger} \leq -5.000) = 10,2\%$$



2c):

Det niveau som overstiges 1 ud af 4 måneder, dvs.  $\frac{1}{4} = 25\%$  af tiden, er den værdi i normalfordelingen, som har 25% sandsynlighed til højre for sig, og dermed har 75% sand-



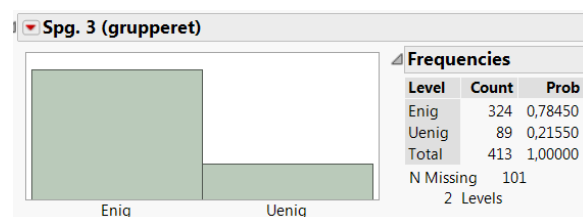
synlighed til venstre for sig. Det er dermed 75%-fraktilen (som det også fremgår af opgaveteksten). Værdien beregnes til 4.618.

M.a.o. så vil antallet af hotelovernatninger blandt svenske gæster 25% af tiden være mindst 4.618 overnatninger flere end samme måned året før. Hvis der eksempelvis i indeværende måned sidste år var 30.000 overnatninger, så vil vi med 25% sandsynlighed forvente, at antal overnatninger i indeværende måned i år vil være på  $30.000 + 4.618 = 34.618$  eller mere.

### Eksempel 3 (► Løsning i JMP)

3A):

Hvis vi sætter 1="Enig" og 0="Uenig", så kan svarene på spørgsmål 3 beskrives ved hjælp af en binomialfordeling med  $n = 413$  og  $\hat{p} = 78,5\%$ .





**3b):**

Antagelserne bag binomialfordelingen betyder, at vi antager...

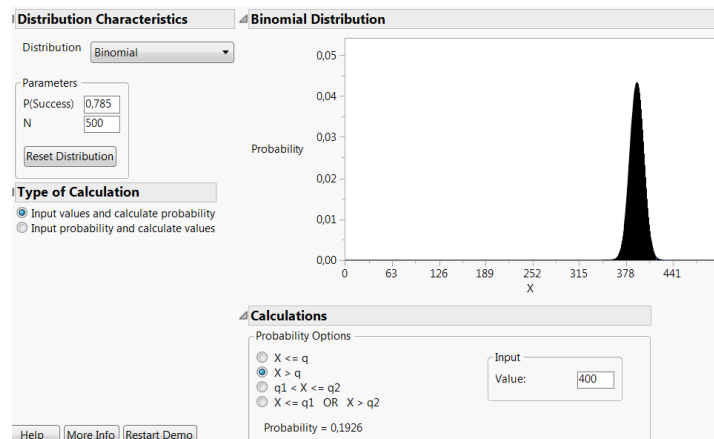
- at der kun er to mulige svar, hvilket vi har sikret os er opfyldt per konstruktion, idet vi ser på den grupperede variabel, der kun indeholder svarmulighederne “enig” og “uenig”.
- at hver respondent svarer “enig” med sandsynlighed  $p$ , dvs. at alle respondenter har samme sandsynlighed for at være enige i, at SU'en er for lav. Vi tager således ikke hensyn til, at der kan være ting som alder, køn m.m., som kan have en indflydelse på, hvor tilbøjelig en person er til at være enig hhv. uenig i, at SU'en er for lav.
- at antallet af respondenter  $n = 413$  er kendt, dvs. at det er fastlagt hvor mange respondenter man vil stille spørgsmålet til, inden undersøgelsen gennemføres (til forskel fra at man i stedet stiller sig ned på gaden og spørger dem, der nu tilfældigvis kommer forbi). Vi har ikke umiddelbart information, der gør os i stand til at vurdere, hvorvidt denne antagelse er opfyldt. Normalt når vi anvender binomialfordelingen, vil vi helt per automatik antage, at antal respondenter ( $n$ ) er kendt på forhånd.
- at de enkelte respondents svar er uafhængige af hinanden, dvs. at én respondents svar ikke er påvirket af hvad en eller flere af de øvrige respondenter svarer. Som udgangspunkt virker det rimeligt at tro, at respondenterne ikke ved, hvad de øvrige har svaret, og derfor virker uafhængighedsantagelsen umiddelbart naturlig.



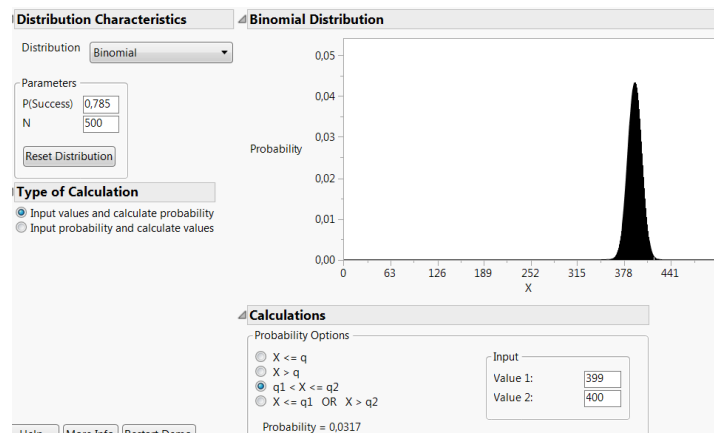
3c):

Vi ser nu på, hvad vi på baggrund af datamaterialet skal forvente, hvis vi spørger 500 tilfældigt udvalgte danskere om de mener, at SU'en er for lav (dvs. vi sætter her  $n = 500$ ).

Sandsynligheden for at mere end 400 danskere mener, at SU'en er for lav er  $P(\# \text{ SU er for lav} > 400) = 19,3\%$ .



Sandsynligheden for at netop 400 danskere mener, at SU'en er for lav er  $P(\# \text{ SU er for lav} = 400) = 3,2\%$ .



Sandsynligheden for at mellem 400 og 450 danskere (begge værdier inklusiv) mener, at SU'en er for lav er  $P(400 \leq \# \text{ SU er for lav} \leq 450) = 22,4\%$ .

