HD Dataanalyse

Note 4: Teoretiske fordelinger

Copenhagen Business School

Emne i dette notesæt

For en variabel i et datamateriale er vi først og fremmest interesserede i at få et overblik over variablens fordeling, dvs. finde ud af...

- hvilke værdier variablen kan antage
- hvor ofte variablen antager en given værdi (dvs. med hvilken sandsynlighed)

Derfor tegner vi som regel et histogram af variablen og får dermed vist variablens såkaldte ***empiriske fordeling***.

Næste skridt er at lave analyser af variablen og drage konklusioner om variablen ud fra det. Her er det nødvendigt at have en teoretisk model, der beskriver, hvordan vi teoretisk set forventer, at variablen vil opføre sig.

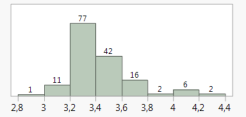
Det er netop, hvad en **teoretisk fordeling** gør. Det er en matematisk model, der giver en stiliseret, teoretisk beskrivelse af en variabel ud fra nogle få centrale karakteristika, som vi vil kalde for **parametre**.

Der findes mange forskellige teoretiske fordelinger. Notesæt 4 beskæftiger sig med de to hyppigst anvendte, **normalfordelingen** og **binomialfordelingen**, som tilsammen viser sig at kunne beskrive mange af de variable, man støder på i praksis.

Emne i dette notesæt

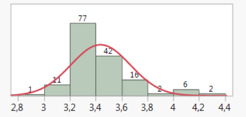
Eksempel: Ølsalg

Ser vi på prisen for 1 stk. Grøn Tuborg (33 cl glasflaske) i supermarkedskæden Føtex, får vi følgende histogram over variablens (empiriske) fordeling



Datamaterialet består af 157 ugers priser, og det fremgår af figuren, at en Grøn Tuborg typisk koster lidt over 3 kr. i Føtex (der ligger flest observationer mellem 3,20 kr. og 3,60 kr.). Eksempelvis ligger prisen i intervallet fra 3,20 kr. til 3,40 kr. i 77 ud af de 157 uger, datamaterialet omhandler.

For variablen "prisen på 1 stk. Grøn Tuborg i Føtex" viser det sig, at man kan bruge en normalfordeling til beskrivelse af variablen. Normalfordelingen giver en stiliseret beskrivelse af variablens fordeling indikeret ved den røde kurve i figuren nedenfor



Normalfordelingen beskriver, hvordan vi alt andet lige vil regne med, at prisen på 1 stk. Grøn Tuborg i Føtex vil fordele sig (dvs. hvordan den vil variere fra uge til uge).

JMP-video [Analyze -> Distribution]

INDHOLDSFORTEGNELSE

1. Normalfordelingen
2. Binomialfordelingen
3. OPSUMMERING

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

1 Normalfordelingen
Karakteristika • *Beregning af sandsynligheder* • *Transformation* • *Fraktiler i* $N(0,1)$ • *Fraktiler i* $N(\mu,\sigma)$ • *Symmetri*

2 Binomialfordelingen

3 OPSUMMERING

Karakteristika

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

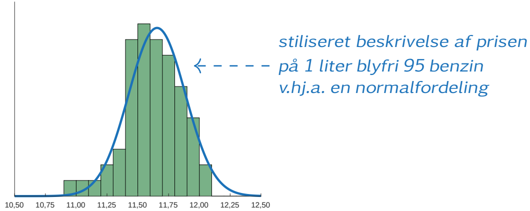
Normalfordelingen er den ubetinget mest anvendte teoretiske fordeling. Det skyldes blandt andet, at den i praksis viser sig at kunne beskrive mange forskellige typer af data.

Mange variable viser sig nemlig – overraskende nok – at have en fordeling, der omtrent har form som en (kirke)klokke, når man tegner et histogram af variablen.

I sådanne tilfælde giver normalfordelingen ofte en god beskrivelse af variablen og kan dermed bruges som udgangspunkt for statistiske analyser af variablen.

Et eksempel på en variabel, hvis fordeling omtrent har en klokkeform, er prisen på en liter blyfri 95 benzin.

(datamaterialet i figuren er ugentlige vejledende priser fra Q8 for perioden 13/7 2017 til 13/7 2018)

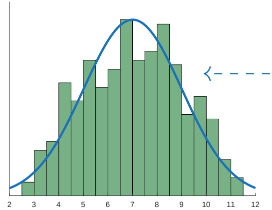


KARAKTERISTIKA

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Et andet eksempel er karaktergennemsnittet for HD 1. delsstuderende.

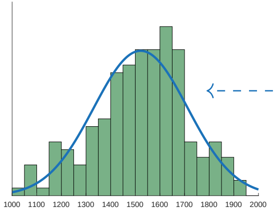
(datamaterialet i figuren er karaktergennemsnit for studerende, der har færdiggjort HD 1. del på CBS i studieåret 2017-2018)



stiliseret beskrivelse af HD 1. dels karaktergennemsnittet v.hj.a. en normalfordeling

Et tredje eksempel er antallet af passagerer med den københavnske Metro.

(datamaterialet i figuren er antal påstigninger på Nørreport station i løbet af et 20 minutters interval (00-20, 20-40, 40-00) i tidsrummet kl. 15-18 på hverdage i januar 2016)



stiliseret beskrivelse af antal passagerer med Metroen v.hj.a. en normalfordeling

KARAKTERISTIKA

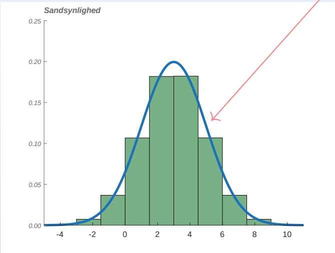
Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Normalfordelingen...

- beskriver fordelingen for en kvantitativ, ***kontinuert*** variabel
- er kendetegnet ved en glat, ***symmetrisk***, ***klokkeformet*** kurve

Nedenstående eksempel viser histogrammer med varierende søjlebredder for en variabel, der er konstrueret sådan, at den er normalfordelt. Eksemplet viser, at man intuitivt kan tænke på den glatte normalfordelingskurve (den blå kurve) som et histogram med meget smalle søjler.

ANIMATION: Normalfordelingen



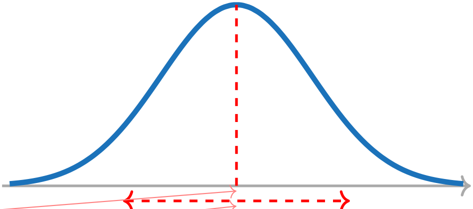
Start animation

KARAKTERISTIKA

Normalfordelingen er fastlagt ud fra to parametre:

- ***middelværdi*** μ (udtales "my")
- ***standardafvigelse*** σ (udtales "sigma")

der tilsammen beskriver normalfordelingens symmetriske, klokkeformede kurve.



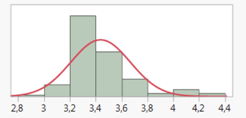
Parameteren μ beskriver, hvor normalfordelingen har sit centrum, mens parameteren σ beskriver, hvor meget normalfordelingen varierer omkring sit centrum (dvs. hvor bred/smal den klokkeformede kurve er).

Parametrene μ og σ tænker vi på som ukendte tal. Når vi har et konkret datasæt, kan vi imidlertid estimere (dvs. gætte på) deres værdier på baggrund af data. Disse estimerede værdier betegner vi med henholdsvis $\hat{\mu}$ og $\hat{\sigma}$.

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Eksempel: Ølsalg

Ser vi igen på prisen på 1 stk. Grøn Tuborg (33 cl glasflaske) i Føtex og den normalfordeling, der bedst beskriver prisen (den røde kurve i figuren nedenfor)



så viser denne normalfordeling sig at være bestemt ved $\hat{\mu} = 3,44$ og $\hat{\sigma} = 0,23$.

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	3,4381017	3,4018514	3,474352
Dispersion	σ	0,2299489	0,207018	0,2586376

Vi vil således regne med, at prisen vil ligge omkring 3,44 kr. og vil variere omkring denne værdi (dvs. nogle gange være højere end 3,44 kr. og nogle gange lavere) i et omfang svarende til værdien af standardafvigelsen $\hat{\sigma} = 0,23$. Mere specifikt, vil vi på baggrund af den empiriske regel regne med, at prisen på Grøn Tuborg i Føtex med ca. 95% sandsynlighed vil ligge i intervallet

$$[\hat{\mu} - 2\hat{\sigma};\ \hat{\mu} + 2\hat{\sigma}] = [3,44 - 2 \cdot 0,23;\ 3,44 + 2 \cdot 0,23] = [2,98\,\text{kr.};\ 3,90\,\text{kr.}]$$

BEMÆRK: Som figuren ovenfor viser, så giver normalfordelingen ikke en perfekt beskrivelse af prisen på Grøn Tuborg i Føtex, idet histogrammet ikke stemmer fuldstændig overens med den klokkeformede røde kurve. Derfor skal den betragtede normalfordeling blot ses som en grov model for, hvad vi regner med prisen på Grøn Tuborg i Føtex vil være i fremtiden.

JMP-video [Analyze -> Distribution]

Karakteristika

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

For hver af normalfordelingens to parametre μ og σ, gælder det, at deres betydning nemt kan visualiseres, fordi begge parametre – hver på sin måde – har en meget direkte indflydelse på, hvorledes normalfordelingens klokkeformede kurve ser ud.

I praksis betyder det, at når man sammenligner to variable (f.eks. i forbindelse med en konkret statistisk analyse), så kan forskelle mellem de to variable i enten middelværdi eller standardafvigelse ofte ses direkte ved blot at sammenligne figurerne af variablenes normalfordelinger.

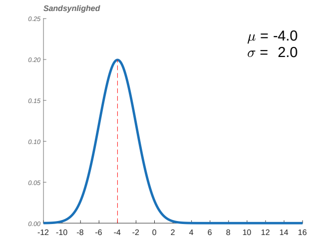
Det viser sig at være nyttigt, når vi i de kommende notesæt skal se på, hvordan vi udfører og fortolker forskellige statistiske analyser. At vi kan visualere betydningen af μ og σ hjælper med til bedre at forstå, hvad der er årsagen til de konklusioner, der kommer ud af vores statistiske analyser.

Karakteristika

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Middelværdien μ angiver normalfordelingens centrum. Nedenstående eksempel viser, hvordan normalfordelingen ændrer sig, når værdien af middelværdien μ ændres (den stiplede røde hjælpelinje angiver værdien af μ).

ANIMATION: Normalfordelingen – ændring af middelværdien μ



Start animation

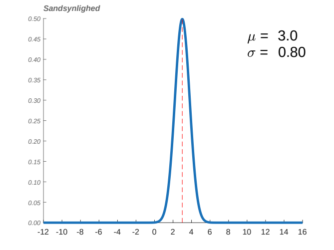
At øge (mindske) middelværdien μ svarer således til at rykke normalfordelingens centrum til højre (venstre), men uden at ændre på variationen i fordelingen (dvs. ændre på hvor bred den klokkeformede kurve er).

Karakteristika

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Standardafvigelsen σ er tilsvarende et mål for normalfordelingens variation. Nedenstående eksempel viser, hvordan normalfordelingen ændrer sig, når værdien af standardafvigelsen σ ændres (den stiplede røde hjælpelinje angiver værdien af μ).

ANIMATION: Normalfordelingen — ændring af standardafvigelsen σ



Start animation

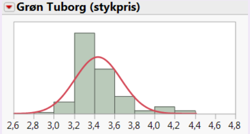
At øge (mindske) standardafvigelsen σ svarer således til at gøre variationen i fordelingen større (mindre) (dvs. ændre på hvor bred den klokkeformede kurve er), men uden at ændre på fordelingens centrum.

KARAKTERISTIKA

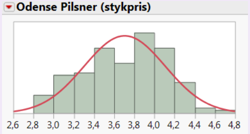
Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Eksempel: Ølsalg

Vi ser fortsat på prisen på 1 stk. Grøn Tuborg (33 cl glasflaske) i Føtex



hvis normalfordeling er givet ved $\hat{\mu}_{Tuborg} = 3,44$ og $\hat{\sigma}_{Tuborg} = 0,23$, og sammenligner den nu med prisen på 1 stk. Odense Pilsner (33 cl glasflaske) i Kvickly



hvis normalfordeling er givet ved $\hat{\mu}_{Odense} = 3,70$ og $\hat{\sigma}_{Odense} = 0,40$.

Det fremgår tydeligt af de to normalfordelinger (de røde kurver), at prisen på Odense Pilsner i Kvickly generelt er højere end prisen på Grøn Tuborg i Føtex, idet centrum i fordelingen for Odense Pilsner (3,70 kr.) ligger længere mod højre end centrum for Grøn Tuborg (3,44 kr.).

Tilsvarende er det også tydeligt, at prisen på Odense Pilsner i Kvickly varierer mere fra uge til uge end prisen på Grøn Tuborg i Føtex, idet normalfordelingen er bredere for Odense Pilsner (std.afv. på 0,40 kr.) end normalfordelingen for Grøn Tuborg (std.afv. på 0,23 kr.).

JMP-video [Analyze -> Distribution]

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Kort om notation:
Vi tænker på en variabels teoretiske fordeling som en beskrivelse af, hvordan vi regner med at variablen vil opføre sig, dvs. vi tænker på det, som om vi endnu ikke ved præcis hvilke værdier af variablen, vi observerer i vores datasæt.

For at understrege, at vi lader som om, vi ikke kender variablens værdier i datasættet, når vi ser på dens teoretiske fordeling, betegner vi variablen med et stort bogstav, som regel X, Y, Z eller lignende.

Hvis vi antager, at variablens teoretiske fordeling er en normalfordeling med parametre μ og σ, vil vi ofte skrive det som $X \sim N(\mu, \sigma)$.

Her er μ variablens middelværdi og σ er variablens standardafvigelse.

I det specielle tilfælde hvor middelværdien (μ) er 0 og standardafvigelsen (σ) er 1, får vi normalfordelingen $N(0,1)$, der også kaldes for **standardnormalfordelingen** *("standard normal distribution")*.

Beregning af sandsynligheder

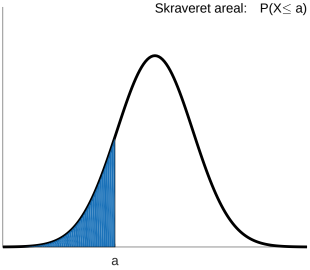
Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Beregning af en ***sandsynlighed*** for en normalfordelt variabel $X \sim N(\mu, \sigma)$ foregår ved at beregne størrelsen af et ***areal under den klokkeformede kurve***.

Det samlede areal under den klokkeformede kurve er lig 1 (=100%).

Arealet til venstre for middelværdien μ er lig 50%.

For en hvilken som helst talværdi a er arealet til venstre for a lig sandsynligheden $P(X \leq a)$, dvs. sandsynligheden for at variablen X er mindre end eller lig talværdien a.



Beregning af sandsynligheder

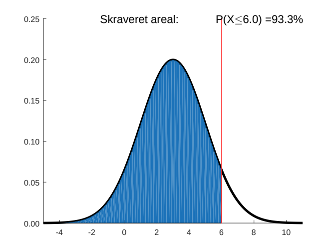
Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Nedenstående eksempel viser den klokkeformede kurve for en normalfordelt variabel $X \sim N(3,2)$ (dvs. en normalfordeling med $\mu = 3$ og $\sigma = 2$).

Heraf fremgår det eksempelvis, at sandsynligheden for, at variablen X er...

- mindre end 0,5 er $P(X \leq 0,5) = 10,6\%$
- mindre end 3 er $P(X \leq 3) = 50,0\%$
- mindre end 6 er $P(X \leq 6) = 93,3\%$

ANIMATION: Normalfordeling - Beregning af sandsynlighed



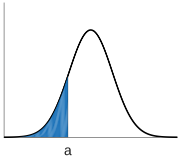
Start animation

Beregning af sandsynligheder

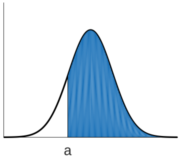
Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

I praksis er det en af følgende tre typer af sandsynligheder for en normalfordelt variabel $X \sim N(\mu, \sigma)$, vi har behov for at kunne beregne:

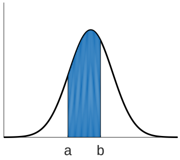
$P(X \leq a)$:



$P(X > a)$:



$P(a \leq X \leq b)$:



hvor a og b er vilkårlige talværdier.

Bemærk: Ud fra den første type af sandsynlighed, $P(X \leq \ldots)$, kan man beregne de to øvrige, idet

- $P(X > a) = 1 - P(X \leq a)$
- $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$

Beregning af sandsynligheder

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

I en normalfordeling er sandsynligheden for at ramme ét bestemt punkt a lig med nul, dvs. $P(X = a) = 0$.

Derfor er det ***ligegyldigt***, når man skal regne en sandsynlighed ud, om man erstatter $\leq$ med $<$, og tilsvarende om man erstatter $\geq$ med $>$.

De tre typer af sandsynligheder på forrige side, kan derfor også skrives som...

- $P(X \leq a) = P(X < a)$
- $P(X \geq a) = P(X > a)$
- $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$

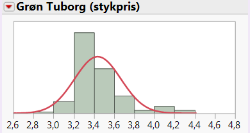
Bemærk: Det er ***kun*** ved beregning af sandsynligheder ***for en normalfordeling***, at man må udskifte $\leq$ med $<$ (og $\geq$ med $>$), som man har lyst til.

BEREGNING AF SANDSYNLIGHEDER

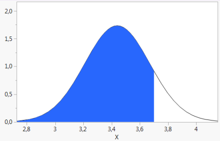
Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

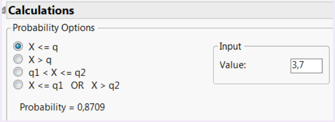
Eksempel: Ølsalg

Ser vi igen på den normalfordeling $X \sim N(\mu,\sigma)$, der beskriver prisen på 1 stk. Grøn Tuborg i Føtex



og som er givet ved de estimerede parametre $\hat{\mu} = 3,44$ og $\hat{\sigma} = 0,23$, så kan vi eksempelvis beregne sandsynligheden for, at prisen i en given uge vil være lavere end 3,70 kr. til $P(X \leq 3,70) = 87,1\%$.





Med andre ord, så vil prisen på 1 stk. Grøn Tuborg i Føtex være under 3,70 kr. 87,1% af ugerne om året (dvs. ca. 7 ud af 8 uger om året, idet $\frac{7}{8} = 87,5\% \approx 87,1\%$).

Ligeledes kan vi eksempelvis beregne sandsynligheden for, at prisen vil ligge mellem 3,50 kr. og 3,70 kr. til $P(3,50 \leq X \leq 3,70) = 26,8\%$. Med andre ord, så vil prisen på 1 stk. Grøn Tuborg i Føtex ligge mellem 3,50 kr. og 3,70 kr. 26,8% af ugerne om året (dvs. ca. 1 ud af 4 uger om året, idet $\frac{1}{4} = 25\% \approx 26,8\%$).

JMP-video [Help -> Sample Data -> Teaching Scripts -> Interactive Teaching Modules -> Distribution Calculator]

Transformation

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Normalfordelingen har en særlig transformationsegenskab, som betyder, at man kan flytte fordelingen til højre eller venstre, og gøre den brede eller smallere, og stadigvæk bevare normalfordelingen.

Kort fortalt så består egenskaben i, at man med simple transformationer (dvs. omregninger) kan...

- gå ***fra*** en normalfordeling $N(\mu,\sigma)$ med vilkårlig middelværdi μ og standardafvigelse σ ***til*** en standardnormalfordeling $N(0,1)$
- gå ***fra*** en standardnormalfordeling $N(0,1)$ ***til*** en normalfordeling $N(\mu,\sigma)$

Med andre ord, så kan man skifte fra en hvilken som helst normalfordeling $N(\mu,\sigma)$ til en standardnormalfordeling $N(0,1)$ og ligeså den anden vej, stort set som man har lyst til.

Transformationsegenskaben er omdrejningspunktet både når vi skal lave konkrete beregninger (af fraktiler eller sandsynligheder) i en normalfordeling, og ikke mindst når vi skal fortolke normalfordelingens parametre μ og σ.

TRANSFORMATION

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Transformationsegenskaben kan opskrives formelt i form af nedenstående to resultater.

Resultat [transformation fra $N(\mu,\sigma)$ til $N(0,1)$]

Hvis $X \sim N(\mu,\sigma)$ er en normalfordelt variabel, så er variablen

$$Z = \frac{X-\mu}{\sigma}$$

standardnormalfordelt $Z \sim N(0,1)$.

Resultat [transformation fra $N(0,1)$ til $N(\mu,\sigma)$]

Hvis $Z \sim N(0,1)$ er en standardnormalfordelt variabel, så er variablen

$$X = \mu + \sigma \cdot Z$$

normalfordelt $X \sim N(\mu,\sigma)$.

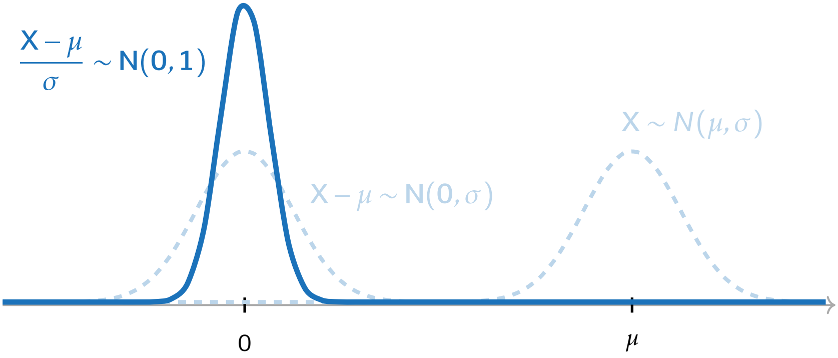
Selv om det ikke er nødvendigt at forstå matematikken bag disse resultater, er det vigtigt at forstå intuitionen bag.

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Intuitionen bag at gå ***fra*** $N(\mu,\sigma)$- ***til*** $N(0,1)$-fordelingen:

Hvis vi har en normalfordeling med middelværdi μ og standardafvigelse σ (variablen X), kan vi...

- lave den om til en normalfordeling med centrum omkring 0 ved at trække middelværdien μ fra (variablen $X-\mu$)
- ændre hvor meget normalfordelingen varierer omkring sit centrum ved at dividere med standardafvigelsen (variablen $Z = \frac{X-\mu}{\sigma}$)

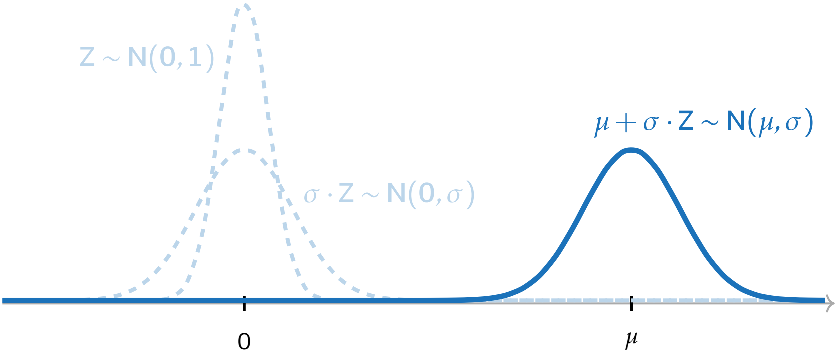


Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Intuitionen bag at gå ***fra*** $N(0,1)$- ***til*** $N(\mu,\sigma)$-fordelingen:

Hvis vi har en standardnormalfordeling (variablen Z), kan vi...

- ændre hvor meget normalfordelingen varierer omkring sit centrum ved at gange med standardafvigelsen (variablen $\sigma \cdot Z$)
- lave den om til en normalfordeling med middelværdi μ ved at lægge μ til (variablen $X = \mu + \sigma \cdot Z$)



Fraktiler i $N(0,1)$

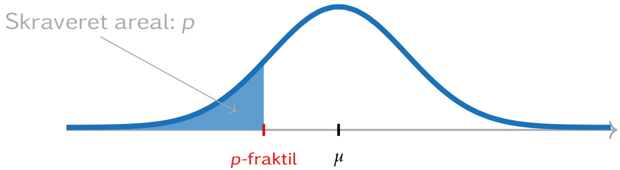
Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Transformationsresultatet viser blandt andet, at man kan få en normalfordeling $N(\mu,\sigma)$ frem blot ved at ændre lidt på en standardnormalfordeling $N(0,1)$.

Det viser sig nyttigt, når man skal beregne fraktiler i en normalfordeling $N(\mu,\sigma)$. Resultatet betyder nemlig, at man kan nøjes med at beregne fraktiler i standardnormalfordelingen $N(0,1)$, og så blot transformere (omregne) disse til fraktiler i $N(\mu,\sigma)$-fordelingen.

Derfor starter vi med at se på, hvordan man beregner fraktiler i $N(0,1)$-fordelingen, og ser derefter på, hvordan man tilsvarende gør det i $N(\mu,\sigma)$-fordelingen. Men inden da lad så lad os lige minde om, hvad en fraktil er for noget:

> *p-fraktilen i en normalfordeling $N(\mu,\sigma)$ er den talværdi, der har sandsynlighed p til venstre for sig (= det skraverede areal).*

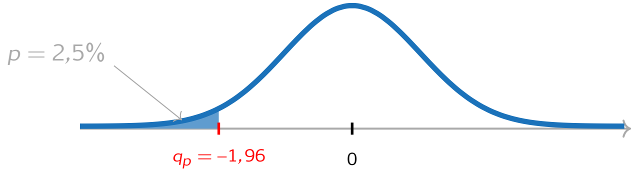


FRAKTILER I $N(0,1)$

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Praktisk anvendelse af en p-fraktil q_p i $\mathbf{N(0,1)}$-fordelingen foregår på én af to måder.

1. Vi ***vælger*** en værdi for ***sandsynligheden*** p, og ønsker så at ***bestemme*** hvilken ***fraktilværdi*** q_p, det svarer til.



Hvis vi eksempelvis vælger $p = 2{,}5\%$, så finder vi ved beregning at $q_p = -1{,}96$.

JMP-video [Help -> Sample Data -> Teaching Scripts -> Interactive Teaching Modules -> Distribution Calculator]

De hyppigst forekommende tilfælde, hvor vi har behov for at vælge sandsynligheden p og derudfra finde den tilhørende fraktil q_p, er angivet i tabellen nedenfor:

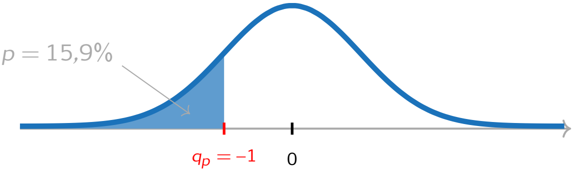
p	0,5%	1%	2,5%	5%
p-fraktil	-2,58	-2,33	-1,96	-1,64

Fraktiler i $N(0,1)$

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

I stedet for at vælge p og derudfra bestemme q_p kan vi også gøre det modsatte, nemlig vælge q_p og derudfra bestemme p.

2. Vi ***vælger*** en værdi for ***fraktilen*** q_p, og ønsker så at ***bestemme*** hvilken ***sandsynlighed*** p, det svarer til.



Hvis vi eksempelvis vælger $q_p = -1$, så finder vi ved beregning at $p = 15{,}9\%$.

▶ JMP-video [Help -> Sample Data -> Teaching Scripts -> Interactive Teaching Modules -> Distribution Calculator]

De hyppigst forekommende tilfælde, hvor vi har behov for at vælge sandsynligheden p og derudfra finde den tilhørende fraktil q_p, er angivet i tabellen nedenfor:

p-fraktil	-3	-2	-1	-0,5
p	0,1%	2,3%	15,9%	30,9%

Fraktiler i $N(\mu,\sigma)$

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Indtil nu har det ikke været helt klart, hvordan størrelsen af standardafvigelsen σ i en normalfordeling $N(\mu,\sigma)$ præcis skal fortolkes. En større værdi af σ betyder en større variation i normalfordelingen $N(\mu,\sigma)$ (dvs. en bredere normalfordelings-klokke), men hvor meget større?

Resultatet på de foregående sider om transformation fra $N(0,1)$ til $N(\mu,\sigma)$ har imidlertid som konsekvens, at vi nu bliver i stand til at give en konkret fortolkning af standardafvigelsen σ. Fortolkningen hænger sammen med den måde vi beregner fraktiler på i normalfordelinger.

Transformationsresultatet betyder nemlig, at en fraktil i standardnormalfordelingen $N(0,1)$ kan bruges til at beregne den tilsvarende fraktil i enhver anden normalfordeling $N(\mu,\sigma)$. Som vi skal se, gør det os i stand til at give en fortolkning af σ.

FRAKTILER I $N(\mu,\sigma)$

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Transformationsresultatet viser, at hvis Z er standardnormalfordelt $N(0,1)$, så er $X = \mu + \sigma \cdot Z$ normalfordelt $N(\mu,\sigma)$. En konsekvens af det er, at vi kan beregne fraktiler i en hvilken som helst normalfordeling $N(\mu,\sigma)$ som angivet nedenfor.

Resultat [beregning af p-fraktil i $N(\mu,\sigma)$-fordelingen]

p-fraktilen i $N(\mu,\sigma)$-fordelingen kan beregnes som

$$\mu + \sigma \cdot q_p$$

hvor q_p er p-fraktilen i $N(0,1)$-fordelingen.

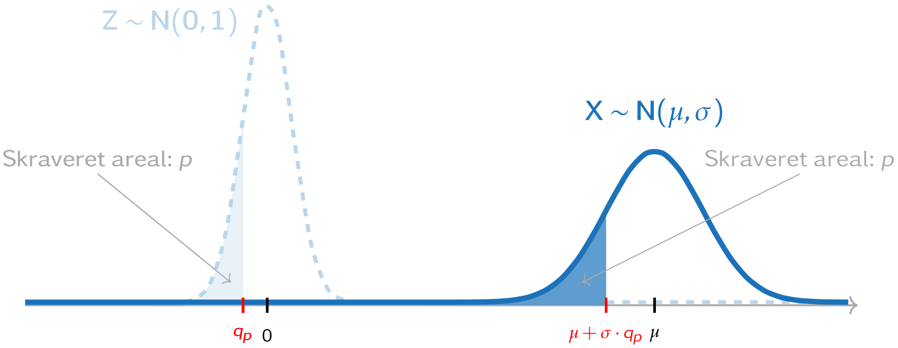
Resultatet er nyttigt, fordi det viser, hvordan en p-fraktil $\mu + \sigma \cdot q_p$ i normalfordelingen $N(\mu,\sigma)$ afhænger af parametrene μ og σ. Dermed viser det, hvordan p-fraktilen i $N(\mu,\sigma)$-fordelingen ændrer sig, såfremt μ eller σ ændrer sig.

FRAKTILER I $N(\mu,\sigma)$

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Intuitionen bag resultatet:

- Ud fra standardnormalfordelingen $Z \sim N(0,1)$ kan vi gå til en normalfordeling $X \sim N(\mu,\sigma)$ ved at sætte $X = \mu + \sigma \cdot Z$
- Hvis vi beregner en p-fraktil q_p i standardnormalfordelingen $N(0,1)$...
- kan vi derfor beregne den tilsvarende p-fraktil i normalfordelingen $N(\mu,\sigma)$ ved at sætte $\mu + \sigma \cdot q_p$.

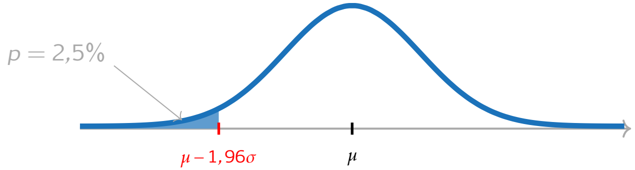


FRAKTILER I $N(\mu,\sigma)$

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Praktisk anvendelse af en p-fraktil q_p i $\mathbf{N}(\mu,\sigma)$-fordelingen foregår på én af to måder.

1. Vi ***vælger*** en værdi for ***sandsynligheden*** p, og ønsker så at ***bestemme*** hvilken ***fraktilværdi*** $\mu+\sigma\cdot q_p$, det svarer til.



Hvis vi eksempelvis vælger $p = 2{,}5\%$, så finder vi ved beregning fraktilen til $\mu - 1{,}96\sigma$.

JMP-video [Help -> Sample Data -> Teaching Scripts -> Interactive Teaching Modules -> Distribution Calculator]

De hyppigst forekommende tilfælde, hvor vi har behov for at vælge sandsynligheden p og derudfra finde den tilhørende fraktil q_p, er angivet i tabellen nedenfor:

p	0,5%	1%	2,5%	5%
p-fraktil	$\mu-2{,}58\sigma$	$\mu-2{,}33\sigma$	$\mu-1{,}96\sigma$	$\mu-1{,}64\sigma$

Normalfordelingen

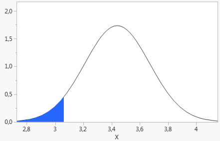
Fraktiler i $N(\mu,\sigma)$

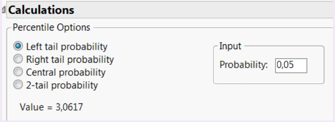
Eksempel: Ølsalg

Vi ser igen på den normalfordeling $X \sim N(\mu,\sigma)$, der beskriver prisen på 1 stk. Grøn Tuborg i Føtex, og som er givet ved de estimerede parametre $\hat{\mu} = 3,44$ og $\hat{\sigma} = 0,23$.

Her kan vi eksempelvis beregne 5%-fraktilen i fordelingen til

$$\hat{\mu} + q_{5\%} \cdot \hat{\sigma} = 3,44 - 1,64 \cdot 0,23 = 3,06 \text{ kr.}$$





Med andre ord, så vil prisen på 1 stk. Grøn Tuborg i Føtex 5% af ugerne om året (dvs. 1 ud af 20 uger om året, idet $\frac{1}{20} = 5\%$) være 3,06 kr. eller lavere.

Vi kan også beregne f.eks. 75%-fraktilen i fordelingen til

$$\hat{\mu} + q_{75\%} \cdot \hat{\sigma} = 3,44 + 0,67 \cdot 0,23 = 3,59 \text{ kr.}$$

Med andre ord, så vil prisen på 1 stk. Grøn Tuborg i Føtex 75% af ugerne om året (dvs. 3 ud af 4 uger om året, idet $\frac{3}{4} = 75\%$) være 3,59 kr. eller lavere.

Hvis variationen i prisen fra uge til uge pludselig stiger, f.eks. pga. større udsving i efterspørgselen, og standardafvigelsen derfor eksempelvis fordobles, dvs. $\hat{\sigma} = 0{,}46$, så ændres 75%-fraktilen til $3,44 + q_{75\%} \cdot 0,46 = 3,44 + 0,67 \cdot 0,46 = 3,75$ kr.

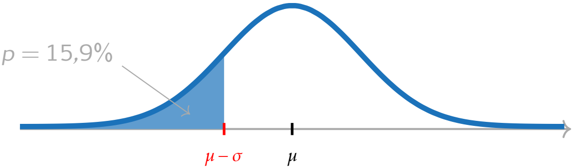
▶ JMP-video [Help -> Sample Data -> Teaching Scripts -> Interactive Teaching Modules -> Distribution Calculator]

Fraktiler i $N(\mu,\sigma)$

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

I stedet for at vælge p og derudfra bestemme q_p kan vi også gøre det modsatte, nemlig vælge q_p og derudfra bestemme p.

2. Vi ***vælger*** en værdi for ***fraktilen*** $\mu+\sigma\cdot q_p$, og ønsker så at ***bestemme*** hvilken ***sandsynlighed*** p, det svarer til.



Hvis vi eksempelvis vælger fraktilen $\mu-\sigma\cdot q_p$ (dvs. $q_p=-1$), så finder vi ved beregning at $p=15{,}9\%$.

▶ JMP-video [Help -> Sample Data -> Teaching Scripts -> Interactive Teaching Modules -> Distribution Calculator]

De hyppigst forekommende tilfælde, hvor vi har behov for at vælge sandsynligheden p og derudfra finde den tilhørende fraktil q_p, er angivet i tabellen nedenfor:

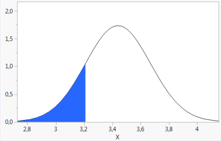
p-fraktil	$\mu-3\sigma$	$\mu-2\sigma$	$\mu-\sigma$	$\mu-0{,}5\sigma$
p	0,1%	2,3%	15,9%	30,9%

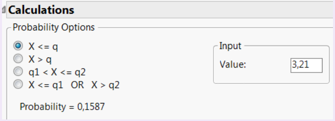
FRAKTILER I $N(\mu, \sigma)$

Eksempel: Ølsalg

Vi ser fortsat på den normalfordeling $X \sim N(\mu, \sigma)$, der beskriver prisen på 1 stk. Grøn Tuborg i Føtex, og som er givet ved de estimerede parametre $\hat{\mu} = 3,44$ og $\hat{\sigma} = 0,23$.

Her kan vi eksempelvis beregne sandsynligheden for, at variablen vil være mere end 1 standardafvigelse lavere end sin middelværdi, dvs. lavere end $3,44 - 0,23 = 3,21$ (svarende til at sætte $q_p = -1$) til 15,9%.





Med andre ord, så vil prisen på 1 stk. Grøn Tuborg i Føtex 15,9% af ugerne om året være 3,21 kr. eller lavere.

På samme måde kan vi f.eks. også beregne sandsynligheden for, at variablen vil være mindre end 1 standardafvigelse større end sin middelværdi, dvs. mindre end $3,44 + 0,23 = 3,67$ (svarende til at sætte $q_p = 1$) til 84,1%. Med andre ord, så vil prisen på 1 stk. Grøn Tuborg i Føtex 84,1% af ugerne om året være 3,67 kr. eller lavere.

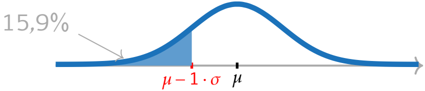
JMP-video [Help -> Sample Data -> Teaching Scripts -> Interactive Teaching Modules -> Distribution Calculator]

SYMMETRI

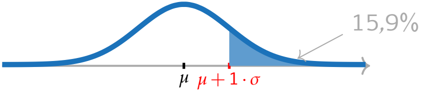
Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Normalfordelingen $N(\mu,\sigma)$ er symmetrisk omkring sin middelværdi μ. Sætter vi eksempelvis $q_p = -1$, hvilket svarer til $p = 15{,}9\%$, så ligger der...

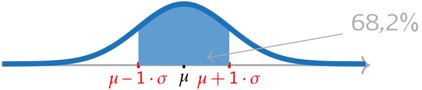
- $15{,}9\%$ sandsynlighed *til venstre* for værdien $\mu - 1 \cdot \sigma$



- ... og pga. symmetrien også $15{,}9\%$ sandsynlighed *til højre* for værdien $\mu + 1 \cdot \sigma$



- ... og dermed ligger der $68{,}2\%$ $(= 1 - 2 \cdot 15{,}9\%)$ sandsynlighed i intervallet $[\mu - 1 \cdot \sigma;\ \mu + 1 \cdot \sigma]$



Konklusion: I normalfordelingen $N(\mu,\sigma)$ ligger ca. 68% af sandsynligheden mindre end 1 standardafvigelse σ væk fra middelværdien μ.

Symmetri

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

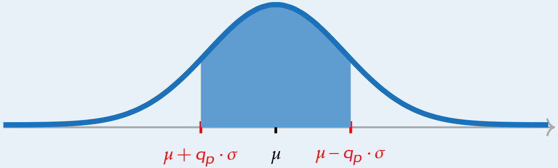
Mere generelt, så giver normalfordelingens symmetri anledning til følgende resultat:

Resultat [fortolkning af σ]

I $N(\mu,\sigma)$-fordelingen indeholder intervallet

$$\left[\mu + q_p \cdot \sigma;\ \mu - q_p \cdot \sigma\right]$$

$1-2p$ sandsynlighed, hvor $0 < p < \frac{1}{2}$ og q_p er p-fraktilen i $N(0,1)$-fordelingen.



Bemærk: For $0 < p < \frac{1}{2}$ er $q_p < 0$ og derfor er $\mu + \sigma \cdot q_p < \mu - \sigma \cdot q_p$, således at intervallet ovenfor altid er veldefineret.

Symmetri

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

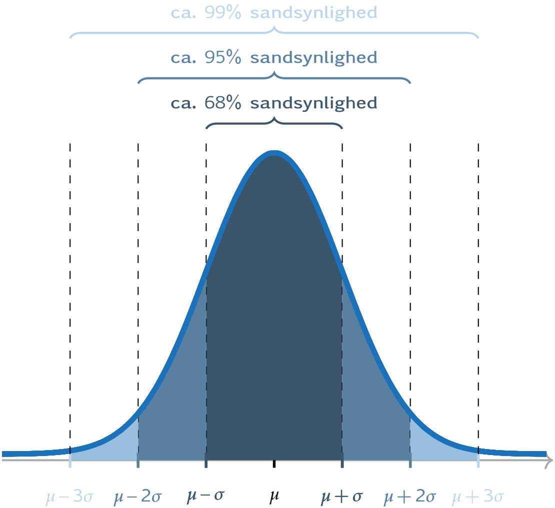
Vi ser for eksemplets skyld på resultatet i tre specielle tilfælde.

- $q_p = -1$ (svarende til $p = 15{,}9\%$):
  I $N(\mu,\sigma)$-fordelingen er der $1-2p = 1-2\cdot 15,9\% = 68,1\%$ sandsynlighed i intervallet
  $$\left[\mu - 1\cdot\sigma;\ \ \mu-(-1)\cdot\sigma\right] = \left[\mu-\sigma;\ \ \mu+\sigma\right]$$
- $q_p = -2$ (svarende til $p = 2{,}3\%$):
  I $N(\mu,\sigma)$-fordelingen er der $1-2p = 1-2\cdot 2,3\% = 95,4\%$ sandsynlighed i intervallet
  $$\left[\mu - 2\cdot\sigma;\ \ \mu-(-2)\cdot\sigma\right] = \left[\mu-2\sigma;\ \ \mu+2\sigma\right]$$
- $q_p = -3$ (svarende til $p = 0{,}1\%$):
  I $N(\mu,\sigma)$-fordelingen er der $1-2p = 1-2\cdot 0,1\% = 99,8\%$ sandsynlighed i intervallet
  $$\left[\mu - 3\cdot\sigma;\ \ \mu-(-3)\cdot\sigma\right] = \left[\mu-3\sigma;\ \ \mu+3\sigma\right]$$

Symmetri

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Grafisk kan vi illustrere de tre eksempler på følgende måde:



Figuren viser, hvor meget størrelsen af standardafvigelsen σ betyder for en normalfordeling $N(\mu,\sigma)$. Det er ofte denne figur, der bruges, når man skal give en fortolkning af σ.

Symmetri

Normal
Karakteristika
Beregning af ssh.
Transformation
Fraktiler i $N(0,1)$
Fraktiler i $N(\mu,\sigma)$
Symmetri
Binomial
OPSUMMERING

Eksempel: Ølsalg

Ser vi igen på den normalfordeling $X \sim N(\mu,\sigma)$, der beskriver prisen på 1 stk. Grøn Tuborg i Føtex, og som er givet ved de estimerede parametre $\hat{\mu} = 3,44$ og $\hat{\sigma} = 0,23$, så vil prisen på 1 stk. Grøn Tuborg i Føtex i en tilfældigt valgt uge...

- med ca. 68% sandsynlighed ligge i intervallet

$$\left[\hat{\mu} - \hat{\sigma};\ \hat{\mu} + \hat{\sigma}\right] = \left[3,44 - 0,23;\ 3,44 + 0,23\right] = \left[3,21 \text{ kr.};\ 3,67 \text{ kr.}\right]$$

- med ca. 95% sandsynlighed ligge i intervallet

$$\left[\hat{\mu} - 2\hat{\sigma};\ \hat{\mu} + 2\hat{\sigma}\right] = \left[3,44 - 2 \cdot 0,23;\ 3,44 + 2 \cdot 0,23\right] = \left[2,98 \text{ kr.};\ 3,90 \text{ kr.}\right]$$

- med ca. 99% sandsynlighed ligge i intervallet

$$\left[\hat{\mu} - 3 \cdot \hat{\sigma};\ \hat{\mu} + 3 \cdot \hat{\sigma}\right] = \left[3,44 - 3 \cdot 0,23;\ 3,44 + 3 \cdot 0,23\right] = \left[2,75 \text{ kr.};\ 4,13 \text{ kr.}\right]$$

Vi kan således på baggrund af vores analyse konkludere...

- at prisen på Grøn Tuborg i Føtex mere end halvdelen af tiden (faktisk ca. 7 ud af 10 uger om året, idet $\frac{7}{10} = 70\% \approx 68\%$) vil ligge mellem 3,21 kr. og 3,67 kr.
- at prisen det meste af tiden (faktisk ca. 19 ud af 20 uger om året, idet $\frac{19}{20} = 95\%$) vil ligge mellem 2,98 kr. og 3,90 kr.
- at prisen stort set altid (ca. 99% af tiden) vil ligge mellem 2,75 kr. og 4,13 kr.

1 Normalfordelingen

2 Binomialfordelingen

Karakteristika • *Beregning af sandsynligheder*

3 OPSUMMERING

Karakteristika

Normal
Binomial
Karakteristika
Beregning af ssh.
OPSUMMERING

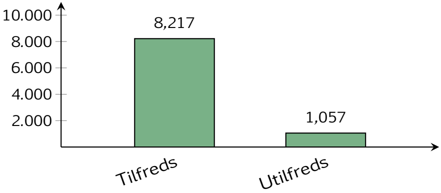
Ofte har vi behov for at analysere situationer, hvor der kun er ***2 mulige udfald***.

I sådanne tilfælde giver binomialfordelingen som regel en god beskrivelse og kan derfor bruges som udgangspunkt for forskellige statistiske analyser.

Et eksempel på en situation med kun 2 mulige udfald er svarene fra en kundetilfredshedsundersøgelse, hvor hver kunde kan svare enten "Tilfreds" eller "Utilfreds".

Region Hovedstaden

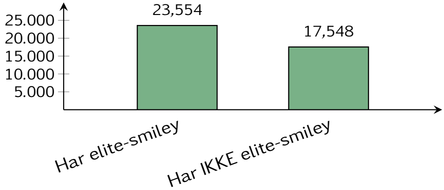
(datamaterialet i figuren er svarene på spørgsmålet *"Er du alt i alt tilfreds med forløbet, fra du blev indlagt, til du blev udskrevet?"* fra 9.274 patienter, der har været akut indlagt på et hospital i Region Hovedstaden i løbet af 2015)



KARAKTERISTIKA

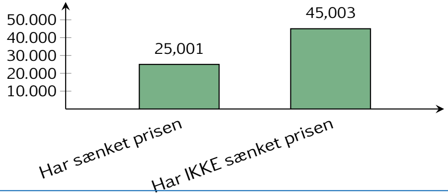
Et andet eksempel er Fødevarestyrelsens elite-smiley for danske fødevarevirksomheder.

(datamaterialet i figuren angiver for hver af 41.102 danske detail fødevarevirksomheder, hvorvidt de har en elite-smiley eller ej pr. 14/8 2018)



Et tredje eksempel er en opgørelse af hvilke boliger udbudt til salg, der har fået sænket udbudsprisen i forhold til den oprindelige udbudspris.

(datamaterialet i figuren angiver om den nuværende udbudspris er lavere end den oprindelige udbudspris for samtlige 70.004 boliger udbudt til salg via boligsiden boliga.dk pr. 17/11 2015)



Karakteristika

Normal
Binomial
Karakteristika
Beregning af ssh.
OPSUMMERING

Alle de tre ovennævnte eksempler omhandler situationer med kun 2 mulige udfald, nemlig...

- at en kunde enten er tilfreds eller er utilfreds
- at en fødevarevirksomhed enten har eller ikke har en elite-smiley
- at en bolig enten har eller ikke har fået sænket sin nuværende pris i forhold til den oprindelige udbudspris

I alle tre eksempler er vi interesserede i at beskrive, ***hvor mange gange*** hver af de 2 mulige udfald indtræffer, dvs. ...

- hvor mange kunder, der er tilfredse hhv. utilfredse
- hvor mange fødevarevirksomheder, der har hhv. ikke har en elite-smiley
- hvor mange boliger, der har hhv. ikke har fået sænket deres udbudspris

Binomialfordelingen er en teoretisk fordeling, der netop modellerer sådanne situationer. Den giver en teoretisk beskrivelse af, hvor mange gange vi skal regne med at observere hver af de 2 mulige udfald, når vi ser på et antal gentagelser (dvs. et antal kunder, et antal virksomheder, et antal boliger).

Karakteristika

Normal
Binomial
Karakteristika
Beregning af ssh.
OPSUMMERING

De to mulige udfald i binomialfordelingen betegnes ofte for nemheds skyld med 1 og 0.

Det er ligegyldigt, hvilket af de to udfald man betegner med 1, og hvilket man betegner med 0.
(dvs. at det f.eks. er ligegyldigt, om vi sætter 1=*"kunden er tilfreds"* og 0=*"kunden er utilfreds"* eller om vi bytter om)

Antallet af gentagelser opfattes som kendt på forhånd.
(dvs. at vi på forhånd har besluttet os for, hvor mange kunder vi vil spørge i vores kundetilfredshedsundersøgelse)

Helt konkret, så beskriver binomialfordelingen det antal gange, vi skal regne med at få udfaldet 1.
(dvs. beskriver antal tilfredse kunder, såfremt vi har sat 1=*"kunden er tilfreds"*. Har vi i stedet sat 1=*"kunden er utilfreds"*, så beskriver binomialfordelingen antal utilfredse kunder)

KARAKTERISTIKA

Definition (binomialfordelingen)

*En variabel X er **binomialfordelt** med parametre n og p, hvis følgende betingelser er opfyldt:*

- *Vi betragter en situation med kun 2 mulige udfald (1 og 0)*
- *Udfaldet 1 indtræffer med samme sandsynlighed p i hver gentagelse*
- *Vi ser på n gentagelser*
- *Alle n gentagelser er indbyrdes uafhængige*

Binomialfordelingen er således fastlagt ud fra de to parametre:

- antalsparameter n
- sandsynlighedsparameter p

Hvis en variabel X er binomialfordelt med parametre n og p, så beskriver X ***antal gange vi får udfaldet 1 i løbet af*** n ***gentagelser***.

Med binomialfordelingen kan vi for hvert $k = 0, \dots, n$ beregne ***sandsynligheden for at få udfaldet 1 netop*** k ***gange*** i løbet af de n gentagelser.

Karakteristika

Normal
Binomial
Karakteristika
Beregning af ssh.
OPSUMMERING

Den klassiske anvendelse af en binomialfordeling er til beskrivelse af n kast med en mønt, hvor...

- 1="krone" og 0="plat"
- der i hvert møntkast er sandsynlighed p for at få udfaldet "krone"
- variablen X angiver antal gange vi får "krone" (ud af de n møntkast)

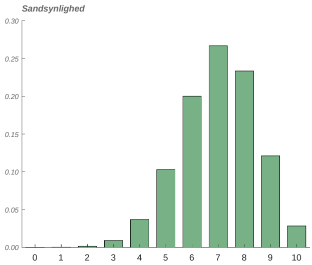
Det er klart, at vi ikke får det samme antal "krone", hver gang vi udfører n møntkast. Det er således noget tilfældigt over variablen X.

Ved hjælp af binomialfordelingen kan vi for hvert $k = 0, \ldots, n$ beregne sandsynligheden for at få "krone" netop k gange i løbet af de i alt n møntkast.

Karakteristika

Normal
Binomial
Karakteristika
Beregning af ssh.
OPSUMMERING

Sandsynlighederne er illustreret i figuren nedenfor i tilfældet med $n = 10$ møntkast og sandsynlighed $p = 0{,}7$ for at få "krone" i hvert enkelt kast:



Eksempelvis er der...

- en sandsynlighed på 10,3% for at få $k = 5$ "krone" i de $n = 10$ møntkast
- en sandsynlighed på 20,0% for at få $k = 6$ "krone" i de $n = 10$ møntkast
- en sandsynlighed på 26,7% for at få $k = 7$ "krone" i de $n = 10$ møntkast

Vi skal senere hen se, hvordan disse sandsynligheder beregnes.

Karakteristika

Normal
Binomial
Karakteristika
Beregning af ssh.
OPSUMMERING

Eksempel: Skat

Vi ser nu på svarene på spørgsmålet *"Er topskatten for høj?"* fra en spørgeskemaundersøgelse vedr. velfærd og skat baseret på svar fra 975 respondenter. Her kan vi bruge binomialfordelingen som teoretisk model til at beskrive de afgivne svar, idet der kun er tilladt to svarmuligheder på det stillede spørgsmål: "Ja" eller "Nej".

Antalsparameteren n er i dette tilfælde lig de i alt 975 indkomne svar. Sætter vi 1 til at repræsentere svaret "Ja", og 0 til at repræsentere "Nej", så angiver sandsynlighedsparameteren p sandsynligheden for, at en given respondent svarer "Ja" på spørgsmålet.

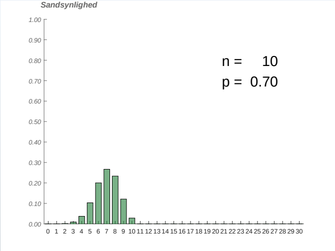
Antagelserne bag binomialfordelingen betyder, at vi antager...

- at der kun er to mulige svar, hvilket i dette tilfælde er opfyldt per definition af de svarmuligheder, respondenterne er blevet givet
- at hver respondent svarer "Ja" med sandsynlighed p, dvs. at alle respondenter har samme sandsynlighed for at svare "Ja". Det kan retfærdiggøres med, at vi som udgangspunkt ikke ved noget om de enkelte respondenter, og derfor rimeligvis må formode, at de alle har samme tilbøjelighed til at svare "Ja" hhv. "Nej"
- at antallet af respondenter $n = 975$ er kendt, dvs. at det er fastlagt hvor mange respondenter man vil stille spørgsmålet til, inden undersøgelsen gennemføres
- at de enkelte respondenters svar er uafhængige, dvs. at én respondents svar ikke er påvirket af hvad en eller flere af de øvrige respondenter svarer

Karakteristika

Antalsparameteren n i binomialfordelingen angiver antal gentagelser af et eksperiment med de to mulige udfald 1 og 0. Nedenstående eksempel viser, hvordan binomialfordelingen ændrer sig, når værdien af n ændres.

ANIMATION: Binomialfordelingen – ændring af antalsparameteren n



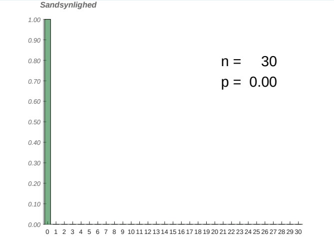
Start animation

At øge (mindske) antalsparameteren n medfører således, at binomialfordelingen rykker til højre (venstre), samtidig med at det også øger (mindsker) variationen i fordelingen (dvs. ændrer på hvor bred fordelingen er).

KARAKTERISTIKA

Sandsynlighedsparameteren p angiver tilsvarende sandsynligheden for at få udfaldet 1 i hver enkelt gentagelse. Nedenstående eksempel viser, hvordan binomialfordelingen ændrer sig, når værdien af p ændres.

ANIMATION: Binomialfordelingen – ændring af sandsynlighedsparameteren p



Start animation

At øge (mindske) sandsynlighedsparameteren p medfører således, at binomialfordelingen rykker til højre (venstre).

KARAKTERISTIKA

Normal
Binomial
Karakteristika
Beregning af ssh.
OPSUMMERING

Diverse facts om binomialfordelingen:

- Binomialfordelingen beskriver per konstruktion fordelingen af en kvantitativ, diskret variabel
- Hvis en variabel X kan beskrives v.hj.a. en binomialfordeling med parametre n og p, vil vi ofte skrive det som $X \sim Bin(n,p)$
- Hvis en variabel er binomialfordelt $X \sim Bin(n,p)$, så er variablen $n-X$ tilsvarende binomialfordelt $X \sim Bin(n, 1-p)$

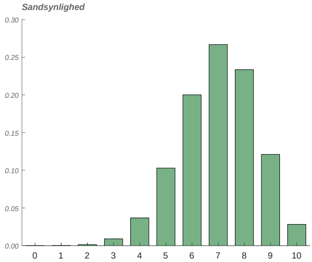
  Intuitionen bag at det forholder sig sådan, fremgår eksempelvis af møntkasteksemplet:

  - Hvis X beskriver antal gange vi får "krone"(=1) ved n møntkast
  - ...og der ved hvert af de n kast er sandsynlighed p for at få "krone"
  - ...så beskriver $n-X$ antal gange vi får "plat" (=0)
  - ...og ved hvert af de n kast er der sandsynlighed $1-p$ for at få "plat"
  - Altså er $n-X \sim Bin(n, 1-p)$
- Hvis en variabel er binomialfordelt $X \sim Bin(n,p)$, så har variablen middelværdi np og standardafvigelse $\sqrt{np(1-p)}$

Beregning af sandsynligheder

Beregning af en ***sandsynlighed*** for en binomialfordelt variabel $X \sim Bin(n,p)$ foregår ved at se på sandsynligheder af formen $P(X=k)$, hvor $k=0,\ldots,n$.

Figuren nedenfor viser sandsynlighederne $P(X=k)$ for $k=0,\ldots,10$ i tilfældet med $n=10$ g $p=0,7$.



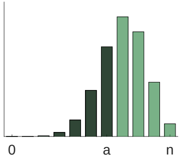
Det samlede grøntskravede areal i figuren, dvs. summen af alle sandsynlighederne $P(X=0), P(X=1),\ldots,P(X=n)$ er lig 1 (=100%).

Højden af den k'te søjle angiver den tilhørende sandsynlighed $P(X=k)$.

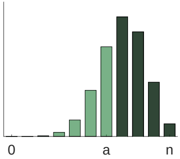
Beregning af sandsynligheder

I praksis er det en af følgende tre typer af sandsynligheder for en binomialfordelt variabel $X \sim Bin(n,p)$, vi har behov for at kunne beregne:

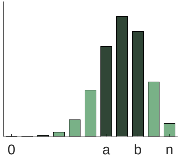
$P(X \leq a)$:



$P(X > a)$:



$P(a \leq X \leq b)$:



hvor a og b er vilkårlige talværdier mellem 0 og n.

Bemærk: Ud fra den første type af sandsynlighed, $P(X \leq \ldots)$, kan man beregne de to øvrige, idet

- $P(X > a) = 1 - P(X \leq a)$
- $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a - 1)$

BEREGNING AF SANDSYNLIGHEDER

Normal
Binomial
Karakteristika
Beregning af ssh.
OPSUMMERING

I en binomialfordeling er sandsynligheden for at ramme ét bestemt punkt a ALDRIG lig med nul, dvs. $P(X = a) > 0$.

Derfor er det ***IKKE ligegyldigt***, når man skal regne en sandsynlighed ud, om man erstatter $\leq$ med $<$, og tilsvarende ikke ligegyldigt om man erstatter $\geq$ med $>$.

Forskellen på om man vælger $\leq$ eller $<$ (og tilsvarende $\geq$ eller $>$) fremgår af nedenstående:

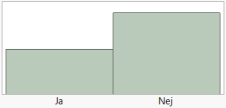
- $P(X \leq a) = P(X < a) + P(X = a)$
- $P(X \geq a) = P(X > a) + P(X = a)$

Normal
Binomial
Karakteristika
Beregning af ssh.
OPSUMMERING

Eksempel: Skat

Ser vi igen på svarene fra spørgmålet vedr. topskat og sætter 1="Ja" og 0="Nej", så kan svarene beskrives af en binomialfordeling med antalsparameter $n = 975$ og estimeret sandsynligheds-parameter $\hat{p} = 35,5\%$.

Er topskatten for høj?

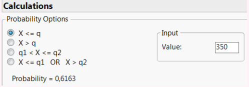


Frequencies

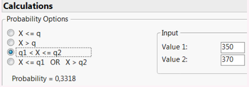
Level	Count	Prob
Ja	346	0,35487
Nej	629	0,64513
Total	975	1,00000

N Missing 0
2 Levels

Lader vi variablen X betegne antallet af "Ja"-svar fra de i alt 975 adspurgte personer, dvs. $X \sim Bin(975, 35,5\%)$, kan vi eksempelvis beregne sandsynligheden, for at højst 350 ud af de 975 adspurgte personer svarer "Ja", til $P(X \leq 350) = 61,63\%$.



Ligeledes kan vi eksempelvis beregne sandsynligheden, for at mere end 350 men højst 370 personer svarer "Ja" (begge værdier inklusiv), til $P(350 < X \leq 370) = P(X \leq 370) - P(X \leq 350) = 94,81\% - 61,63\% = 33,18\%$.



JMP-video [Analyze -> Distribution ; Help -> Sample Data -> Teaching Scripts -> Interactive Teaching Modules -> Distribution Calculator]

1 Normalfordelingen

2 Binomialfordelingen

3 OPSUMMERING

OPSUMMERING

Kort opsummering af dette notesæt:

Normalfordelingen $N(\mu,\sigma)$ er...

- en fordeling for en kontinuert variabel
- en symmetrisk, klokkeformet fordeling
- karakteriseret ved middelværdi μ og standardafvigelse σ

p-fraktilen i en normalfordeling $N(\mu,\sigma)$ kan beregnes som $\mu+\sigma\cdot q_p$, hvor q_p er p-fraktilen i en standard-normalfordeling $N(0,1)$.

I normalfordelingen $N(\mu,\sigma)$ er der $1-2p$ sandsynlighed indeholdt i intervallet $[\mu+\sigma\cdot q_p;\ \mu-\sigma\cdot q_p]$ for $0<p<\frac{1}{2}$.

Ved beregning af sandsynligheder i normalfordelingen, er det ligegyldigt, om man erstatter $\leq$ med $<$ og $\geq$ med $>$.

Binomialfordelingen $Bin(n,p)$...

- er en fordeling for en diskret variabel
- beskriver n uafhængige gentagelser af et eksperiment med 2 mulige udfald (1 og 0)
- har samme sandsynlighed p for at få udfaldet 1 i hver gentagelse
- er karakteriseret ved antalsparameter n og sandsynlighedsparameter p

Normal

Binomial

OPSUMMERING

INDEKS

Binomialfordeling *s. 43*

Binomialfordeling, antalsparameter (n) *s. 47*

Binomialfordeling, sandsynlighedsparameter (p) *s. 48*

Normalfordeling *s. 7*

Normalfordeling, middelværdi (μ) *s. 10*

Normalfordeling, standardafvigelse (σ) *s. 11*

Normalfordeling, transformationsegenskab *s. 20*

Normalfordeling, symmetri *s. 34*

Normal
Binomial
OPSUMMERING

Nye funktionaliteter i dette notesæt:

- *Analyze -> Distribution:*
  - Normalfordeling – estimation
- *Help -> Sample Data -> Teaching Scripts -> Interactive Teaching Modules -> Distribution Calculator:*
  - Binomialfordeling – beregning af sandsynligheder
  - Normalfordeling – beregning af sandsynligheder
  - Normalfordeling – beregning af fraktiler

JMP-videoer:

s. 2: ▶ [Analyze -> Distribution] (normalfordeling)

s. 12: ▶ [Analyze -> Distribution] (normalfordeling)

s. 18: ▶ [Help -> Sample Data -> Teaching Scripts -> Interactive Teaching Modules -> Distribution Calculator] (normalfordeling, sandsynligheder)

s. 24: ▶ [Help -> Sample Data -> Teaching Scripts -> Interactive Teaching Modules -> Distribution Calculator] (standardnormalfordeling, fraktiler)

s. 25: ▶ [Help -> Sample Data -> Teaching Scripts -> Interactive Teaching Modules -> Distribution Calculator] (standardnormalfordeling, sandsynligheder)

s. 30: ▶ [Help -> Sample Data -> Teaching Scripts -> Interactive Teaching Modules -> Distribution Calculator] (normalfordeling, fraktil)

s. 32: ▶ [Help -> Sample Data -> Teaching Scripts -> Interactive Teaching Modules -> Distribution Calculator] (normalfordeling, sandsynlighed)

s. 53: ▶ [Help -> Sample Data -> Teaching Scripts -> Interactive Teaching Modules -> Distribution Calculator] (binomialfordeling)