Note: Normalfordeling med Nspire

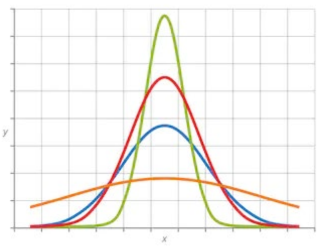
KBJ, december 2019

Normalfordelingens grundlag

En *kontinuert* stokastisk variabel X beskriver udfaldet af et sandsynlighedseksperiment som kan resultere i et hvilket som helst reelt tal. Udfaldsrummet for X er altså i princippet hele $\mathbb{R}$.

En særligt vigtig kontinuert sandsynlighedsfordeling er *normalfordelingen*, som er entydigt beskrevet ved en middelværdi μ og en spredning σ, skrevet $X \sim N(\mu, \sigma)$. For ethvert tal t gælder da at $P(X = t) = 0$, mens der om ethvert interval $[a; b]$ gælder at $P(a \leq X \leq b) > 0$. Vi kan således ikke beskrive sandsynligheden for et enkelt tal, men kun sandsynlighedstætheden. Tætheden af sandsynlighed ved et sted x på den reelle tal akse er da beskrevet ved tæthedsfunktionen

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \mathbf{e}^{-\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2},$$



hvis graf er en ”klokkeformet” Gauss-kurve som vist til højre.

For at bestemme sandsynligheden for at X lander i et interval $[a; b]$ benyttes integration:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Den tilhørende fordelingsfunktion $F(x)$ angiver således $P(X \leq x)$ og er defineret ved

$$F(x) = \int_{-\infty}^x f(t)\, dt$$

Heraf følger at $P(a \leq X \leq b) = F(b) - F(a)$. $F(x)$ kan ikke skrives ved en forskrift.

En særlig vigtig normalfordeling er *standardnormalfordelingen* $X \sim N(0,1)$ med tæthedsfunktion

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \mathbf{e}^{-\frac{1}{2}x^2},$$

Med fordelingsfunktion $\Phi(x) = \int_{\infty}^x \phi(t)dt$. Der gælder for $F(x)$ der beskriver $X \sim N(\mu, \sigma)$ at:

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Kommandoer i Nspire: Standardnormalfordeling

Tæthedsfunktionen for standardnormalfordelingen $X \sim N(0,1)$ kan defineres som

$$\phi(x) = \texttt{normpdf(x)}$$

Fordelingsfunktionen for tilhørende fordelingsfunktion kan defineres som

$$\Phi(x) = \texttt{normcdf(-infinity,x)}$$

Sandsynligheden $P(a \leq X \leq b)$ kan bestemmes:

$$P(a \leq X \leq b) = \texttt{normcdf(a,b)}$$

<u>Eksempler på Nspire-kommander</u>

Definition af $\phi(x)$:	Definition af $\Phi(x)$:	Bestemmelse af $P(-1 \leq X \leq 1)$:
$f(x):=\text{normpdf}(x)$	$ff(x):=\text{normcdf}(-\text{infinity},x)$	$\text{normcdf}(-1,1)$
$f(x):=\text{normPdf}(x)$	$ff(x):=\text{normCdf}(-\infty,x)$	$\text{normCdf}(-1,1) \quad 0.682689$

Kommandoer i Nspire: Normalfordeling generelt

Tæthedsfunktionen for normalfordelingen $X \sim N(\mu, \sigma)$ kan defineres som

$$f(x) = \texttt{normpdf(x,}\mu\texttt{,}\sigma\texttt{)}$$

Fordelingsfunktionen for tilhørende fordelingsfunktion kan defineres som

$$F(x) = \texttt{normcdf(-infinity,x,}\mu\texttt{,}\sigma\texttt{)}$$

Sandsynligheden $P(a \leq X \leq b)$ kan bestemmes:

$$P(a \leq X \leq b) = \texttt{normcdf(a,b,}\mu\texttt{,}\sigma\texttt{)}$$

<u>Eksempler på Nspire-kommandoer for $X \sim N(10,2)$:</u>

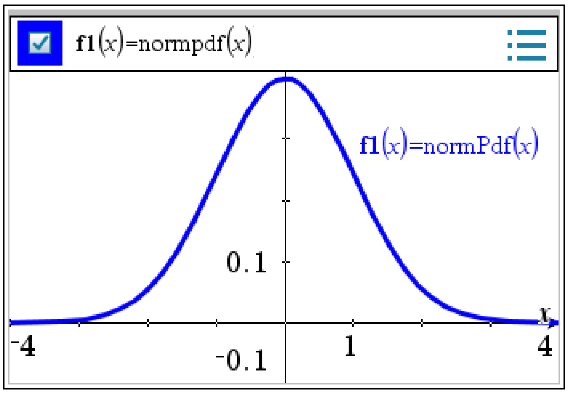
Definition af $f(x)$:	Definition af $F(x)$:	Bestemmelse af $P(9 \leq X \leq 13)$:
$f(x):=\text{normpdf}(x,10,2)$	$ff(x):=\text{normcdf}(-\text{infinity},x,10,2)$	$\text{normcdf}(9,13,10,2)$
$f(x):=\text{normPdf}(x,10,2)$	$ff(x):=\text{normCdf}(-\infty,x,10,2)$	$\text{normCdf}(9,13,10,2) \quad 0.624655$

Grafer for tætheds- og fordelingsfunktioner

Grafen for $\phi(x)$ kan tegnes i et grafer-vindue med kommandoen:

```
f1(x)=normpdf(x)
```

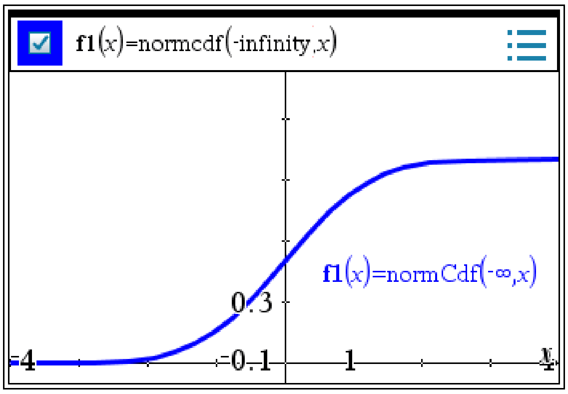
Det giver som vist til højre, den klokkeformede kurve med y-aksen som symmetriakse. Specielt gælder at $\phi(0) = \frac{1}{\sqrt{2\pi}}$.



Grafen for $\Phi(x)$ kan tegnes i et grafer-vindue med kommandoen:

```
f1(x)=normcdf(-infinity,x)
```

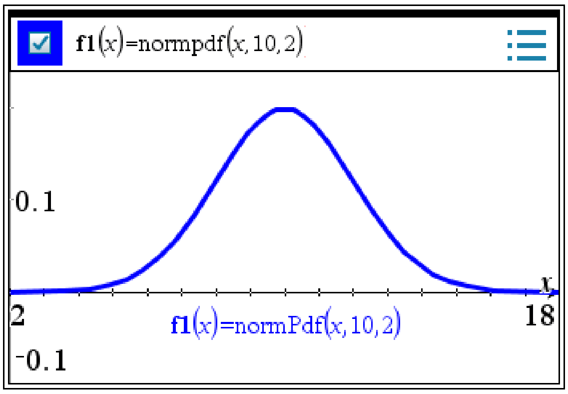
Det giver som vist til højre en kurve der konvergerer mod 0 for $x \to -\infty$ og mod 1 for $x \to \infty$. Specielt gælder at $\phi(0) = 0{,}5$.



For $X \sim N(\mu, \sigma)$ gælder at tæthedsfunktionen $f(x)$ kan tegnes i et grafer-vindue med kommandoen:

```
f1(x)=normpdf(x,μ,σ)
```

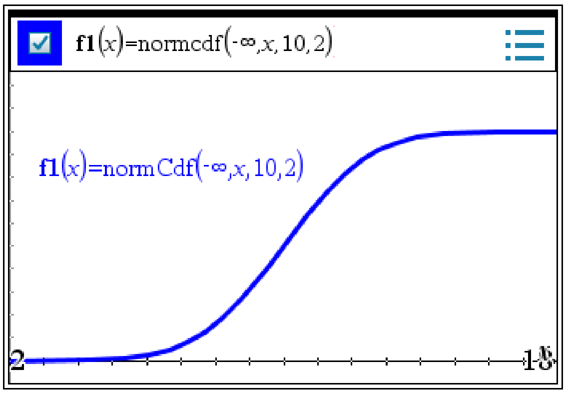
For $X \sim N(10{,}2)$ fås den klokkeformede Gauss-kurve vist til højre her, symmetrisk omkring linjen $x = 10$.



For $X \sim N(\mu, \sigma)$ gælder at fordelingsfunktionen $F(x)$ kan tegnes i et grafer-vindue med kommandoen:

```
f1(x)=normcdf(-infinity,x,μ,σ)
```

For $X \sim N(\mu, \sigma)$ fås en kurve der konvergerer mod 0 for $x \to -\infty$ og mod 1 for $x \to 1$. Specielt gælder at $F(10) = 0{,}5$.



Fraktil og fraktilfunktion

For standardnormalfordelingen defineres q-fraktilen som x-værdien x_q om hvilken der gælder:

$$\mathrm{P}(X < x_q) = \Phi(x_q) = q$$

Til fordelingsfunktionen Φ hører inversfunktionen Φ^{-1}, kaldet *fraktilfunktionen*. Det gælder da:

$$x_q = \Phi^{-1}(q)$$

I Nspire kan q-fraktilen bestemmes med kommandoen:

$$x_q = \texttt{invnorm(q)}$$

For fordelingsfunktionen F for $X \sim N(\mu, \sigma)$ er q-fraktilen defineret som x-værdien x_q således at

$$P(X < x_q) = F(x_q) = q$$

Til fordelingsfunktionen F hører inversfunktionen F^{-1} kaldet *fraktilfunktionen*. Det gælder da

$$x_q = F^{-1}(q)$$

I Nspire kan q-fraktilen bestemmes med kommandoen:

$$x_q = \texttt{invnorm(q,}\mu\texttt{,}\sigma\texttt{)}$$

Til bestemmelse af σ eller μ ud fra q, x_q og enten μ eller σ benyttes:

$$\begin{aligned} F(x_q) &= q \\ \Phi\left(\frac{x_q - \mu}{\sigma}\right) &= q \\ \Phi^{-1}\left(\Phi\left(\frac{x_q - \mu}{\sigma}\right)\right) &= \Phi^{-1}(q) \\ \frac{x_q - \mu}{\sigma} &= \Phi^{-1}(q) \end{aligned}$$

Denne ligning kan da løses for μ eller σ. Kendes to fraktiler p og q kan såvel μ som σ bestemmes ved at løses ligningssystemet

$$\frac{x_p - \mu}{\sigma} = \Phi^{-1}(p) \text{ og } \frac{x_q - \mu}{\sigma} = \Phi^{-1}(q)$$

for μ og σ.

Med Nspire bestemmes 75%-fraktil for standardnormalfordelingen:

invnorm(0.75)

invNorm(0.75)	0.67449

Det ses at $x_{0,75} = 0{,}6745$

Med Nspire bestemmes 75%-fraktilen for normalfordelingen $X \sim N(10{,}2)$:

invnorm(0.75,10,2)

invNorm(0.75,10,2)	11.349

Det ses at $x_{0,75} = 11{,}35$.

Med Nspire bestemmes μ for $X \sim N(\mu, 4)$ når $x_{0,2} = 40$ ved at løse $\frac{40-\mu}{4} = \Phi^{-1}(0{,}2)$:

solve((40−m)/4=invnorm(0.2),m)

$\text{solve}\left(\frac{40-m}{4}=\text{invNorm}(0.2),m\right)$	$m=43.3665$

Det er altså bestemt at $\mu = 43{,}37$

Med Nspire bestemmes σ for $X \sim N(55, \sigma)$ når $x_{0,85} = 68$ ved at løse $\frac{68-55}{\sigma} = \Phi^{-1}(0{,}85)$:

$\text{solve}\left(\frac{68-55}{s}=\text{invnorm}(0.85),s\right)$

$\text{solve}\left(\frac{68-55}{s}=\text{invNorm}(0.85),s\right)$	$s=12.543$

Det er altså bestemt at $\sigma = 12{,}54$

For $x_{0.2} = 25$ og $x_{0,7} = 46$ bestemmes σ og μ ved at løse $\frac{25-\mu}{\sigma} = \Phi^{-1}(0{,}2)$ og $\frac{46-\mu}{\sigma} = \Phi^{-1}(0{,}7)$:

solve((25−m)/s=invnorm(0.2) and (46−m)/s=invnorm(0.7),m,s)

$\text{solve}\left(\frac{25-m}{s}=\text{invNorm}(0.2) \text{ and } \frac{46-m}{s}=\text{invNorm}(0.7),m,s\right)$	$m=37.9383$ and $s=15.3731$

Det er altså bestemt at $\mu = 37{,}94$ og $\sigma = 15{,}37$.

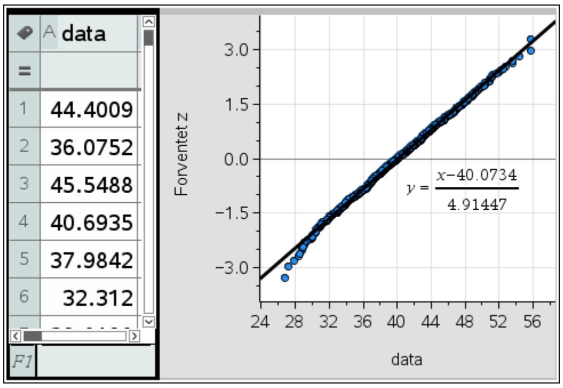
Undersøgelse af om data er normalfordelte

Hvis et større datasæt med n observationer sorteres voksende kan man tildele hver observation en kumuleret frekvens $F(x)$ ved formlen $F(x_i) = \frac{2i-1}{2n}$. Hvis data er normalfordelte, vil vi forvente en lineær sammenhæng mellem x og forventet $F(x)$-fraktil, z, i standardnormalfordelingen:

$$z = \Phi^{-1}(F(x)) = \Phi^{-1}\left(\Phi\left(\frac{x-\mu}{\sigma}\right)\right) = \frac{x-\mu}{\sigma} = \frac{1}{\sigma} \cdot x - \frac{\mu}{\sigma}$$

Et plot af alle (x_i, z_i) vil give et *fraktilplot*. Hvis data er normalfordelte kan vi endvidere estimere at $\hat{\mu} = \bar{x}$ (middelværdi af data) og $\hat{\sigma} = \sqrt{\frac{(x_1-\hat{x})^2+\cdots+(x_n-\hat{x})^2}{n-1}}$ (stikprøvespredningen af data).

Hvis vi plotter linjen $y = \frac{x-\hat{\mu}}{\hat{\sigma}}$ sammen med fraktilplottet, vil fraktilplottets punkter tilnærmelsesvist ligge på linjen, såfremt de er tilnærmelsesvist normalfordelte.



I Nspire kan dette gøres simpelt, f.eks. ved at oprette en liste kaldet *data*, i et "Lister og regneark"-vindue og derpå placere data på førsteaksen i et "Diagrammer og statistik" vindue og højreklikke på diagrammet og vælge "normalfordelingsplot". På figuren ses at fraktilplottet tilnærmelsesvist ligger på linjen:

$$y = \frac{x-\mu}{\sigma}$$

På eksemplet ovenfor kan altså aflæses at $\mu = 40{,}07$ og $\sigma = 4{,}91$.

Hvis der gives obervationsintervaller $I_1, I_2, \dots, I_n$ med intervalendepunkter $a_1, a_2, \dots, a_n$ og kumulerede frekvenser $F(I_1)$, $F(I_2)$,…,$F(I_n)$ kan beregnes tilhørende standardnormalfordelingsfraktiler og således tegnes et punktplot over alle punkter $\left(a_i, \Phi^{-1}(F(I_i))\right)$ på hvilke der kan laves regression

Eksempel: Et grupperet datasæt er givet med nedenstående frekvenser:

Interval (I)	Intervalfrekvens $f(I)$	Kumuleret frekvens $F(I)$	$\Phi^{-1}(F(I))$
$]35;40]$	2,5%	2,5%	$-1{,}960$
$]40;45]$	17,5%	20%	$-0{,}842$
$]45;50]$	27,0%	47%	$-0{,}075$
$]50;55]$	36%	83%	$0{,}954$
$]55;60]$	15%	98%	$2{,}054$
$]60;65]$	2%	100%	∞

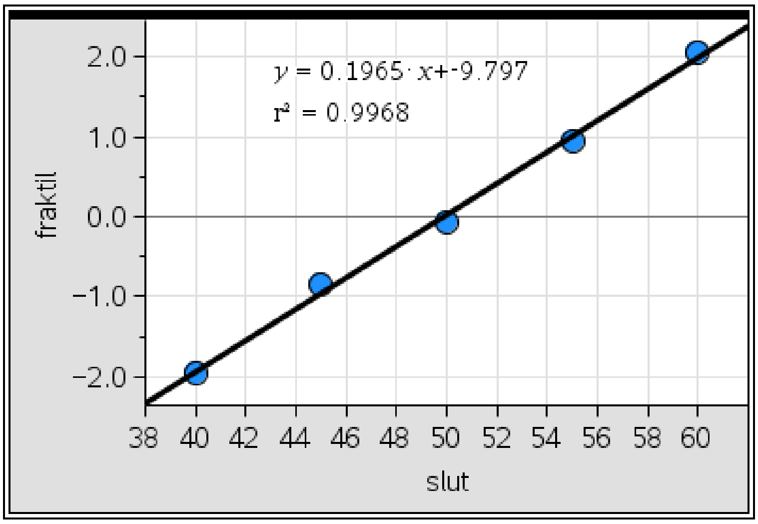
I et “Lister og regneark”-vindue kan data indtastet med eksempelvis ”start” og ”slut” værdier fra intervallerne i separate søjler samt en søjle med frekvens. I en søjle beregnes kumulerede frekvenser med:

	A start	B slut	C frk	D kumfrk	E fraktil
=				=cumulativesum(frk)	=invnorm(kumfrk)
1	35.	40.	0.025	0.025	-1.95996
2	40.	45.	0.175	0.2	-0.841621
3	45.	50.	0.27	0.47	-0.07527
4	50.	55.	0.36	0.83	0.954165
5	55.	60.	0.15	0.98	2.05375

```
cumulativesum(frk)
```

Ud fra kumulerede frekvenser bestemmes normalfordelingen ved

```
invnorm(kumfrk)
```



På figuren oven for ses indtastningen. Bemærk at sidste interval må undlades, da 100%-fraktilen er ∞. De øvrige punkter indtegnes i et punktplot med intervalendepunkter på førsteaksen og fraktiler på andenaksen.

Det ses i eksemplet at bedste rettelinje bliver: $y = 0{,}1965 \cdot x - 9{,}797$.

Det ses at punkterne ligger tilnærmelsesvist på en ret linje, så data er tilnærmelsesvist normalfordelte med $\sigma = \frac{1}{a} = \frac{1}{0{,}1965} = 5{,}09$ og $\mu = -\frac{b}{a} = -\frac{-9{,}797}{0{,}1965} = 49{,}86$.

Generering af normalfordelte data

Hvis der er behov for en liste med tilfældigt genererede tilnærmelsesvist normalfordelte data med middelværdi μ og spredning σ kan listen genereres med kommandoen:

```
randnorm(μ,σ,#Antal data)
```

Således kan en liste på 1000 værdier normalfordelte med $\mu = 50$ og $\sigma = 7$ genereres med kommandoen:

```
randnorm(50,7,1000)
```

	A data
=	=randnorm(50,7,1000)
1	53.175
2	36.5747
3	45.8197
4	53.2483
5	41.5111

data:=randNorm(50,7,1000)
{31.0675,37.6452,53.655,43.7314▸

OVERSIGT OVER NSPIRE-KOMMANDOER

I det følgende er X en normalfordelt stokastisk variabel $X \sim N(\mu, \sigma)$ hvor andet ikke fremgår

Se præsentation i screencast: https://youtu.be/IA3fUSJ0ZOE

Beskrivelse	Kommando
$\phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \mathbf{e}^{-\frac{1}{2} \cdot x^2}$	`normpdf(x)`
$\Phi(x) = \int_{-\infty}^{x} f(t)dt$	`normcdf(-infinity,x)`
$\Phi(b) - \Phi(a) = \int_a^b f(x)dx$	`normcdf(a,b)`
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \mathbf{e}^{-\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2}$	normpdf(x,μ,σ)
$F(x) = \int_{-\infty}^{x} f(t)dt = P(X \leq x)$	normcdf(-infinity,x,μ,σ)
$F(b) - F(a) = \int_a^b f(x)dx = P(a \leq X \leq b)$	normcdf(a,b,μ,σ)
Fraktilfunktion standardnormalfordeling: $\Phi^{-1}(a) = x_a$	`invnorm(a)`
Fraktilfunktion for $X \sim N(\mu, \sigma)$: $F^{-1}(a) = x_a$	invnorm(a,μ,σ)
Liste med N tilfældigt genererede talværdier tilnærmelsesvist normalfordelt med middelværdi μ og spredning σ.	randnorm(μ,σ,N)