

KAPITEL 4

KONTINUERTE FORDELINGER

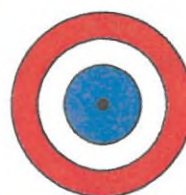
4.1 Tæthedsfunktioner og fordelingsfunktioner

Kontinuerte sandsynlighedsfordelinger optræder, når udfaldsrummet er et interval. De enkelte tal i intervallet har sandsynlighed 0, så i stedet for at lægge punkt-sandsynligheder sammen vil vi typisk bestemme sandsynligheder som længder eller arealer.

Eksempel 4.1.1

En primitiv dartskive med radius 35 cm er inddelt i fire felter, som er dannet ud fra cirkler med radius 2 cm, 15 cm, 25 cm og 35 cm (figur 4.1.1).

En dartpil, som rammer skiven, kan ramme i et vilkårligt punkt på skiven, og dens afstand X til centrum (i cm) kan vi angive som et tal i intervallet $[0; 35]$. Sandsynligheden for at ramme et udvalgt punkt er 0, men sandsynligheden for at ramme inden for et felt, antager vi, er bestemt ved feltets areal i forhold til hele skivens areal.



Figur 4.1.1

Dartskivens areal er $\pi \cdot (35 \text{ cm})^2 = 1225\pi \text{ cm}^2$. Sandsynlighederne for at ramme det sorte, det blå, det hvide og det røde felt er henholdsvis

$$P(\text{sort}) = P(0 \leq X \leq 2) = \frac{\pi \cdot 2^2}{1225\pi} = 0,00327 = 0,327 \%$$

$$P(\text{blå}) = P(2 < X \leq 15) = \frac{\pi(15^2 - 2^2)}{1225\pi} = 0,1804 = 18,04 \%$$

$$P(\text{hvid}) = P(15 < X \leq 25) = \frac{\pi(25^2 - 15^2)}{1225\pi} = 0,3265 = 32,65 \%$$

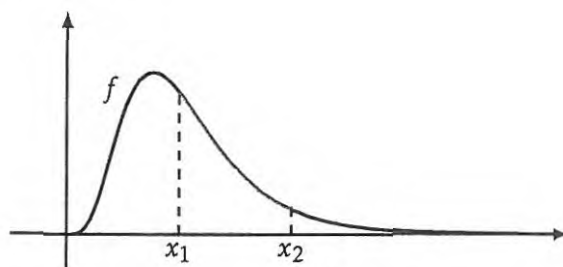
$$P(\text{rød}) = P(25 < X \leq 35) = \frac{\pi(35^2 - 25^2)}{1225\pi} = 0,4898 = 48,98 \%$$

En stokastisk variabel X , der kan antage alle mulige værdier i et interval I (begrænset eller ubegrænset) i $\mathbb{R}$, kalder vi en *kontinuert stokastisk variabel*, og den stokastiske variable X i Eksempel 4.1.1 er en kontinuert stokastisk variabel.

Vi vil koncentrere os om nogle vigtige kontinuerte stokastiske variable, hvis sandsynlighedsfordelinger er givet ved en kontinuert tæthedsfunktion defineret på et interval.

Tæthedsfunktioner

En *tæthedsfunktion* f hørende til en stokastisk variabel X er en ikke-negativ, kontinuert funktion defineret på et interval $I = Vm(X)$. Tæthedsfunktionen giver os sandsynligheden for at få et udfald mellem x_1 og x_2 svarende til arealet af det område, der er afgrænset af grafen for f , x -aksen og de to lodrette linjer med $x = x_1$ og $x = x_2$, se Figur 4.1.2.



Figur 4.1.2

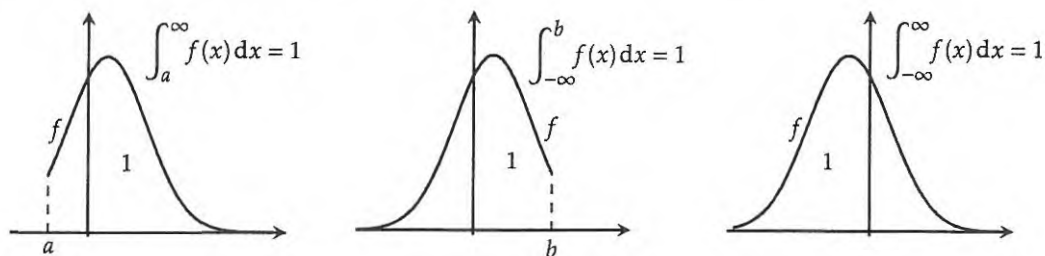
Sandsynligheden bliver dermed

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(t) dt$$

Da den samlede sandsynlighed skal være 1, må arealet af hele området mellem grafen og x -aksen være 1. Hvis f f.eks. er defineret på $[a; b]$, skal

$$\int_a^b f(x) dx = 1$$

Er f i stedet defineret på et ubegrænset interval, bliver de tilhørende integraler uegentlige. Disse uegentlige integraler skal eksistere og have værdien 1.



Figur 4.1.3

Når en stokastiske variabel X har tæthedsfunktion f , kan vi udtrykke sandsynligheder som $P(\alpha \leq X \leq \beta)$, $P(\alpha \leq X)$ og $P(X \leq \beta)$ som integraler.

Hvis f.eks. tæthedsfunktionens definitionsmængde er alle reelle tal $]-\infty; \infty[$, er

$$P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} f(x) dx$$

$$P(X = \alpha) = P(\alpha \leq X \leq \alpha) = 0$$

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x) dx$$

$$P(X \leq \beta) = P(-\infty < X \leq \beta) = \int_{-\infty}^{\beta} f(x) dx$$

$$P(X \geq \alpha) = P(\alpha \leq X < \infty) = \int_{\alpha}^{\infty} f(x) dx$$

Eksempel 4.1.2

For en speciel type elektriske komponenter kan vi beskrive restlevetiden målt i år ved en stokastisk variabel X med tæthedsfunktionen f givet ved

$$f(x) = 0,2 \cdot e^{-0,2x}, \quad x > 0$$

Med denne tæthedsfunktion er

$$P(X \leq \alpha) = \int_0^{\alpha} 0,2 \cdot e^{-0,2t} dt = [-e^{-0,2t}]_0^{\alpha} = 1 - e^{-0,2\alpha}$$

$$P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} 0,2 \cdot e^{-0,2t} dt = [-e^{-0,2t}]_{\alpha}^{\beta} = e^{-0,2\alpha} - e^{-0,2\beta}$$

$$P(X \geq \beta) = \int_{\beta}^{\infty} 0,2 \cdot e^{-0,2t} dt = \lim_{b \rightarrow \infty} ([-e^{-0,2t}]_{\beta}^b) = \lim_{b \rightarrow \infty} (e^{-0,2\beta} - e^{-0,2b}) = e^{-0,2\beta}$$

Sandsynligheden for, at en komponent har en restlevetid på mindre end 5 år, er

$$P(X < 5) = P(X \leq 5) = 1 - e^{-0,2 \cdot 5} = 1 - e^{-1} = 0,6321 = 63,21 \%$$

Sandsynligheden for, at en komponent har en restlevetid på over 7 år, er

$$P(X > 7) = P(X \geq 7) = e^{-0,2 \cdot 7} = e^{-1,4} = 24,66 \%$$

▣ Øvelse 157, Øvelse 158, Øvelse 159, Øvelse 160, Øvelse 161, Øvelse 162

Når en stokastiske variabel X har tæthedsfunktion f , kan vi udtrykke sandsynligheder som $P(\alpha \leq X \leq \beta)$, $P(\alpha \leq X)$ og $P(X \leq \beta)$ som integraler.

Hvis f.eks. tæthedsfunktionens definitionsområde er alle reelle tal $]-\infty; \infty[$, er

$$P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} f(x) dx$$

$$P(X = \alpha) = P(\alpha \leq X \leq \alpha) = 0$$

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x) dx$$

$$P(X \leq \beta) = P(-\infty < X \leq \beta) = \int_{-\infty}^{\beta} f(x) dx$$

$$P(X \geq \alpha) = P(\alpha \leq X < \infty) = \int_{\alpha}^{\infty} f(x) dx$$

Eksempel 4.1.2

For en speciel type elektriske komponenter kan vi beskrive restlevetiden målt i år ved en stokastisk variabel X med tæthedsfunktionen f givet ved

$$f(x) = 0,2 \cdot e^{-0,2x}, \quad x > 0$$

Med denne tæthedsfunktion er

$$P(X \leq \alpha) = \int_0^{\alpha} 0,2 \cdot e^{-0,2t} dt = [-e^{-0,2t}]_0^{\alpha} = 1 - e^{-0,2\alpha}$$

$$P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} 0,2 \cdot e^{-0,2t} dt = [-e^{-0,2t}]_{\alpha}^{\beta} = e^{-0,2\alpha} - e^{-0,2\beta}$$

$$P(X \geq \beta) = \int_{\beta}^{\infty} 0,2 \cdot e^{-0,2t} dt = \lim_{b \rightarrow \infty} ([-e^{-0,2t}]_{\beta}^b) = \lim_{b \rightarrow \infty} (e^{-0,2\beta} - e^{-0,2b}) = e^{-0,2\beta}$$

Sandsynligheden for, at en komponent har en restlevetid på mindre end 5 år, er

$$P(X < 5) = P(X \leq 5) = 1 - e^{-0,2 \cdot 5} = 1 - e^{-1} = 0,6321 = 63,21 \%$$

Sandsynligheden for, at en komponent har en restlevetid på over 7 år, er

$$P(X > 7) = P(X \geq 7) = e^{-0,2 \cdot 7} = e^{-1,4} = 24,66 \%$$

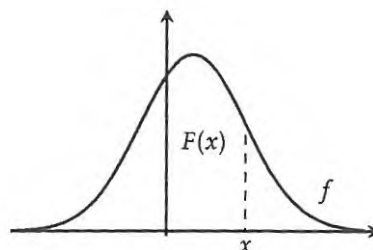
▣ Øvelse 157, Øvelse 158, Øvelse 159, Øvelse 160, Øvelse 161, Øvelse 162

Fordelingsfunktioner

For en tæthedsfunktion f defineret på $\mathbb{R}$ er den tilhørende *fordelingsfunktion* F defineret ved

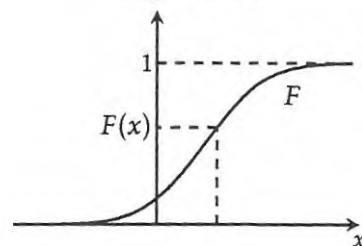
$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt, \quad x \in \mathbb{R}$$

Funktionsværdien $F(x)$ kan vi beskrive som arealet af den ubegrænsede punktmængde vist på Figur 4.1.4. Dermed kan vi opfatte fordelingsfunktionen $F(x)$ som „arealfunktionen med udgangspunkt i $-\infty$ “.



Figur 4.1.4

Fordelingsfunktionen er voksende, og dens graf vil have x -aksen og den vandrette linje med ligningen $y = 1$ som asymptoter, som illustreret på Figur 4.1.5.



Figur 4.1.5

Hvis tæthedsfunktionen er defineret på et begrænset interval, f.eks. $[a; b]$, sætter vi $F(x)$ til 0, når x er mindre end a , og $F(x)$ til 1, når x er større end b , dvs.

$$F(x) = \begin{cases} 0 & \text{for } x < a \\ \int_a^x f(t) dt & \text{for } x \in [a; b] \\ 1 & \text{for } x > b \end{cases}$$

Eksempel 4.1.3

For den stokastiske variabel X med tæthedsfunktion

$$f(x) = 0,75 \cdot e^{-0,75x}, \quad x > 0$$

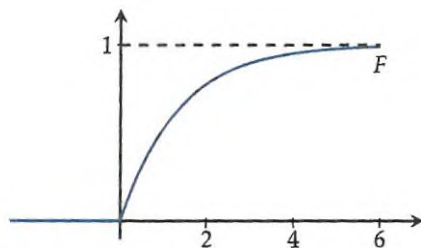
er fordelingsfunktionens værdi for $x > 0$

$$F(x) = \int_0^x f(t) dt = [-e^{-0,75t}]_0^x = 1 - e^{-0,75x}$$

For $x \leq 0$, sætter vi $F(x) = 0$, så

$$F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ 1 - e^{-0,75x} & \text{for } x > 0 \end{cases}$$

Grafen er vist i Figur 4.1.6.



Figur 4.1.6

➤ Øvelse 163

For en stokastisk variabel X med en tæthedsfunktion f defineret på $[a; b]$ er *midelværdien* $E(X)$, som vi også betegner med μ givet ved

$$\mu = E(X) = \int_a^b x \cdot f(x) dx$$

og *variansen* $\text{Var}(X)$ er

$$\text{Var}(X) = \int_a^b (x - \mu)^2 \cdot f(x) dx$$

De samme udtryk gælder også, hvis f er defineret på et interval på formen $]a; b]$, $[a; b[$ eller $]a; b[$.

Når f er defineret på $] -\infty; \infty[$ er *middelværdien* $E(X)$ eller μ givet ved

$$\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

og *variansen* $\text{Var}(X)$ er

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx$$

Spredningen σ er i begge tilfælde defineret som

$$\sigma = \sqrt{\text{Var}(X)}$$

Integranderne i udtrykkene for $E(X)$ og $\text{Var}(X)$ indeholder produkter, og til integration af disse har vi hidtil kun haft substitutionsmetoden til rådighed.

En anden grundlæggende integrationsmetode til integration af produkter er *partiel integration* også kaldet *delvis integration*.

Sætning 4.1.1. Antag, at f er kontinuert og g er differentiabel med kontinuert afledet g' . Så er

$$\int f(x)g(x)dx = F(x)g(x) - \int F(x)g'(x)dx \quad (4.1.1)$$

hvor F er en stamfunktion til f .

Bevis

Når F er en stamfunktion til $f(x)$, får vi ved differentiation

$$\begin{aligned} \left(F(x)g(x) - \int F(x)g'(x)dx \right)' &= (F(x)g(x))' - \left(\int F(x)g'(x)dx \right)' \\ &= F'(x)g(x) + F(x)g'(x) - F(x)g'(x) \\ &= F'(x)g(x) = f(x)g(x) \end{aligned}$$

Altså er $F(x)g(x) - \int F(x)g'(x)dx$ en stamfunktion til $f(x)g(x)$, og sætningen er bevist. $\square$

Udtrykket i (4.1.1) kalder vi *formlen for partiel integration* eller *formlen for delvis integration*.

Ved brug af formelen får vi et udtryk, som indeholder en ny integrand $F(x)g'(x)$, som også er et produkt. I visse tilfælde er dette produkt muligt at integrere.

Eksempel 4.1.4

Vi vil bestemme

$$\int xe^x dx$$

ved partiel integration.

Vi sætter $f(x) = e^x$ og $g(x) = x$ og vælger $F(x) = e^x$. Så giver (4.1.1)

$$\int xe^x dx = e^x \cdot x - \int e^x \cdot 1 dx = xe^x - \int e^x dx = xe^x - e^x$$

Eksempel 4.1.5

Vi vil bestemme en stamfunktion til $h(x) = 4x \cos(x)$ ved partiel integration.

Med $f(x) = \cos(x)$, $g(x) = 4x$ og $F(x) = \sin(x)$ får vi

$$\int 4x \cos(x) dx = 4x \sin(x) - \int 4 \sin(x) dx = 4x \sin(x) + 4 \cos(x)$$

Altså er $H(x) = 4x \sin(x) + 4 \cos(x)$ en stamfunktion til h .

Eksempel 4.1.6

Med partiel integration finder vi

$$\begin{aligned}\int \ln(x) dx &= \int 1 \cdot \ln(x) dx = x \ln(x) - \int x \cdot \frac{1}{x} dx \\ &= x \ln(x) - \int 1 dx \\ &= x \ln(x) - x\end{aligned}$$

Altså er $x \ln(x) - x$ en stamfunktion til $\ln(x)$.

Vi kan også benytte formelen for partiel integration flere gange for at bestemme et integral.

Eksempel 4.1.7

$$\begin{aligned}\int x^2 e^x dx &= e^x \cdot x^2 - \int e^x \cdot 2x dx \\ &= x^2 e^x - 2 \int x e^x dx \\ &= x^2 e^x - 2(x e^x - \int e^x dx) \\ &= x^2 e^x - 2x e^x + 2e^x = (x^2 - 2x + 2)e^x\end{aligned}$$

Formlen for partiel integration kan vi også anvende i forbindelse med bestemte integraler.

Så gælder

$$\int_a^b f(x)g(x)dx = [F(x)g(x)]_a^b - \int_a^b F(x)g'(x) dx \quad (4.1.2)$$

Eksempel 4.1.8

$$\begin{aligned}\int_0^3 x e^{-x} dx &= [-x e^{-x}]_0^3 - \int_0^3 -e^{-x} dx \\ &= -3e^{-3} + \int_0^3 e^{-x} dx \\ &= -3e^{-3} + [-e^{-x}]_0^3 \\ &= -3e^{-3} + e^0 - e^{-3} = 1 - 4e^{-3}\end{aligned}$$

➤ Øvelse 164, Øvelse 165, Øvelse 166

4.2 Normalfordelingen

Normalfordelingen er den vigtigste fordeling for kontinuerte stokastiske variable, da den ofte bliver brugt til at beskrive fordelingen af målte størrelser.

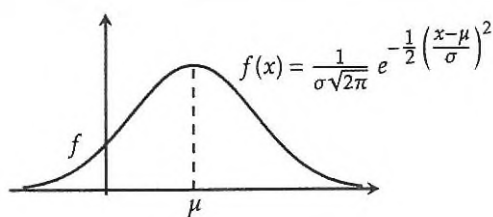
En normalfordelt stokastisk variabel X har en tæthedsfunktion på formen

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R} \quad (4.2.1)$$

hvor parameteren μ kan være et vilkårligt tal, og parameteren σ er et positivt tal.

Parametrene μ og σ kalder vi henholdsvis normalfordelingens *middelværdi* og *spredning*, og vi kalder X normalfordelt med middelværdi μ og spredning σ og skriver $X \sim N(\mu, \sigma)$.

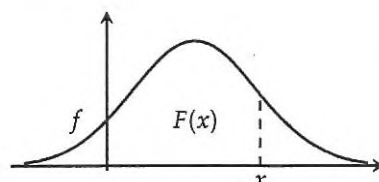
Tæthedsfunktionens graf er den „klokkeformede“ *Gauss-kurve*, som er symmetrisk omkring den lodrette linje $x = \mu$ med maksimum i punktet $\left(\mu, \frac{1}{\sigma\sqrt{2\pi}}\right)$.



Figur 4.2.1

Fordelingsfunktionen F er givet ved

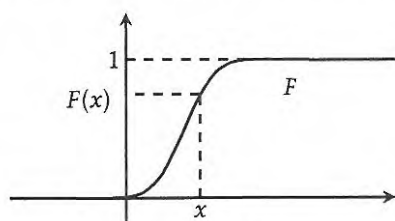
$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt, \quad x \in \mathbb{R}$$



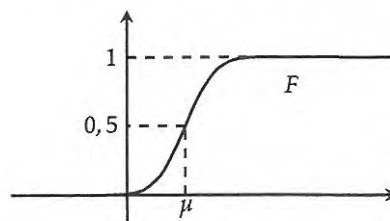
Figur 4.2.2

Sandsynligheden $F(x) = P(X \leq x)$ er illustreret i Figur 4.2.2.

Figur 4.2.3 viser en typisk graf for fordelingsfunktionen for en normalfordeling, og Figur 4.2.4 illustrerer grafens symmetri omkring middelværdien μ .



Figur 4.2.3



Figur 4.2.4

Vi vil ikke bevise, at (4.2.1) rent faktisk fastlægger en tæthedsfunktion, dvs. at

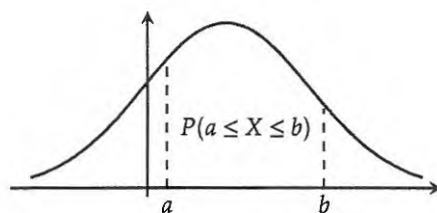
$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt = 1$$

da det er umuligt at angive stamfunktionen til tæthedsfunktionen med et analytisk udtryk.

Af samme grund kan vi heller ikke angive et analytisk udtryk for $F(x) = P(X \leq x)$, som giver os de eksakte sandsynligheder.

For at beregne normalfordelingssandsynligheder som f.eks.

$$\begin{aligned} P(X \leq b) &= F(b) \\ P(a \leq X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= F(b) - F(a) \end{aligned}$$



Figur 4.2.5

må vi i stedet anvende CAS eller tabeller.

Eksempel 4.2.1

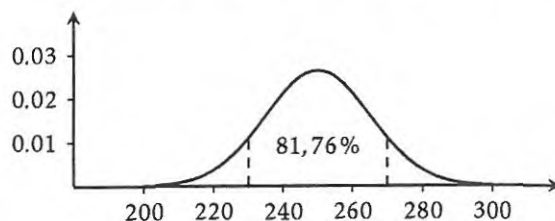
En kaffemaskine er indstillet til at levere et krus kaffe på 250 mL, men det faktiske volumen af kaffe varierer noget fra gang til gang.

Vi antager, at volumenet X af leveret kaffe er normalfordelt med middelværdi $\mu = 250$ mL og spredning $\sigma = 15$ mL.

Sandsynligheden for, at et tilfældigt valgt krus kaffe fra maskinen indeholder mellem 230 mL og 270 mL, er

$$P(230 \leq X \leq 270) = 0,8176 = 81,76 \%$$

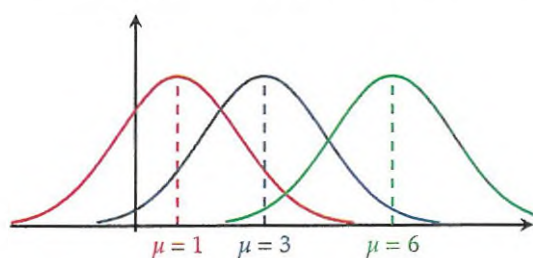
Sandsynligheden er her fundet ved brug af CAS og er illustreret i Figur 4.2.6.



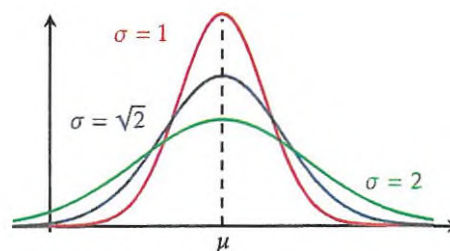
Figur 4.2.6

➤ Øvelse 167, Øvelse 168, Øvelse 169

Gauss-kurver med samme spredning σ , men forskellige middelværdier μ , fremkommer af hinanden ved vandret parallelforskydning, se Figur 4.2.7. Spredningen bestemmer Gauss-kurvens form, og jo større spredning jo bredere Gauss-kurve, som illustreret på Figur 4.2.8.

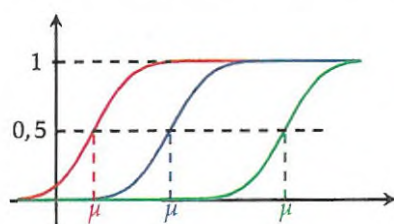


Figur 4.2.7

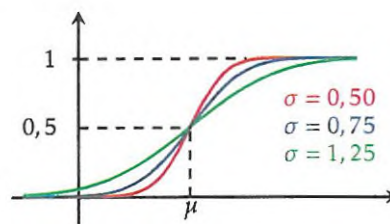


Figur 4.2.8

Fordelingsfunktioner for normalfordelinger med samme spredning fremkommer således ligeledes af hinanden ved vandret parallelforskydning, mens spredningen bestemmer fordelingsfunktionernes form.



Figur 4.2.9



Figur 4.2.10

Normeringsfaktoren $\frac{1}{\sigma\sqrt{2\pi}}$ sikrer, at den samlede sandsynlighed er 1, og er baseret på værdien af det uegentlige integral

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

Ved brug substitutionen $t = \frac{1}{\sqrt{2}}x$ med $dt = \frac{1}{\sqrt{2}}dx$ får vi

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = \int_{-\infty}^{\infty} e^{-(\frac{1}{\sqrt{2}}x)^2} dx = \sqrt{2} \int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{2}\sqrt{\pi}$$

dvs.

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi} \quad (4.2.2)$$

Parametrene μ og σ har vi omtalt som middelværdi og spredning. At denne sprogbrug er korrekt, er indholdet i den følgende sætning.

Sætning 4.2.1. Antag, at X er $N(\mu, \sigma)$ -fordelt.

Så er

$$E(X) = \mu \quad \text{og} \quad \sigma = \sqrt{\text{Var}(X)}$$

Bevis

Når $X \sim N(\mu, \sigma)$ er sandsynlighedsfordelingen for X bestemt ved tæthedsfunktionen

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R}$$

Så er

$$E(X) = \int_{-\infty}^{\infty} xf(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Vi benytter nu substitutionen

$$t = \frac{x-\mu}{\sigma} = \frac{1}{\sigma}x - \frac{\mu}{\sigma} \quad \text{med} \quad dt = \frac{1}{\sigma}dx$$

Da $t \rightarrow \pm\infty$ for $x \rightarrow \pm\infty$, kan vi omforme integralet

$$\begin{aligned} E(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \cdot \sigma \int_{-\infty}^{\infty} (\sigma t + \mu) e^{-\frac{1}{2}t^2} dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma t + \mu) e^{-\frac{1}{2}t^2} dt \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} te^{-\frac{1}{2}t^2} dt + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt \end{aligned}$$

Integranden $te^{-\frac{1}{2}t^2}$ i det første integral er en ulige funktion, og ved at benytte (4.2.2) får vi så

$$E(X) = 0 + \frac{\mu}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = \mu$$

Altså er middelværdien $E(X)$ af X parameteren μ i tæthedsfunktionen.

Vi kan nu bestemme $\text{Var}(X)$ som

$$\text{Var}(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Med samme substitution $t = \frac{x-\mu}{\sigma}$ kan vi omforme integralet

$$\begin{aligned}\text{Var}(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \cdot \sigma \int_{-\infty}^{\infty} (\sigma t)^2 e^{-\frac{1}{2}t^2} dt \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{1}{2}t^2} dt\end{aligned}$$

For at bestemme dette uegentlige integral undersøger vi

$$\int_a^b t^2 e^{-\frac{1}{2}t^2} dt$$

for $b \rightarrow \infty$ og $a \rightarrow -\infty$.

Ved partiel integration får vi

$$\begin{aligned}\int_a^b t^2 e^{-\frac{1}{2}t^2} dt &= \int_a^b t \cdot t e^{-\frac{1}{2}t^2} dt \\ &= \left[t(-e^{-\frac{1}{2}t^2}) \right]_a^b + \int_a^b e^{-\frac{1}{2}t^2} dt \\ &= -be^{-\frac{1}{2}b^2} + ae^{-\frac{1}{2}a^2} + \int_a^b e^{-\frac{1}{2}t^2} dt \\ &= -\frac{b}{e^{\frac{1}{2}b^2}} + \frac{a}{e^{\frac{1}{2}a^2}} + \int_a^b e^{-\frac{1}{2}t^2} dt\end{aligned}$$

For $b \rightarrow \infty$ vil både tæller og nævner i det første led gå mod ∞ , men $e^{\frac{1}{2}b^2}$ bliver uendelig stor i forhold til b , og brøken og det første led går mod 0.

For $a \rightarrow -\infty$ vil tæller gå mod $-\infty$ og nævner mod ∞ i det andet led. Men igen vil $e^{\frac{1}{2}a^2}$ blive uendelig stor i forhold til den numeriske værdi af a , og brøken og det andet led går mod ligeledes 0.

Dermed er

$$\begin{aligned}\text{Var}(X) &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{1}{2}t^2} dt \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \lim_{a \rightarrow -\infty} \left(\lim_{b \rightarrow \infty} \left(\int_a^b t^2 e^{-\frac{1}{2}t^2} dt \right) \right) \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \lim_{a \rightarrow -\infty} \left(\lim_{b \rightarrow \infty} \left(\int_a^b e^{-\frac{1}{2}t^2} dt \right) \right) \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = \sigma^2\end{aligned}$$

Da spredningen er kvadratroden af variansen er

$$\sigma = \sqrt{\text{Var}(X)}$$

og sætningen er bevist. □

➤ Øvelse 170, Øvelse 171

4.3 Standardnormalfordelingen

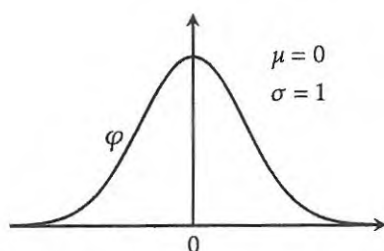
Normalfordelingen $N(0,1)$ med middelværdi 0 og spredning 1 kalder vi *standardnormalfordelingen*.

Den tilhørende tæthedsfunktion betegner vi med φ , dvs.

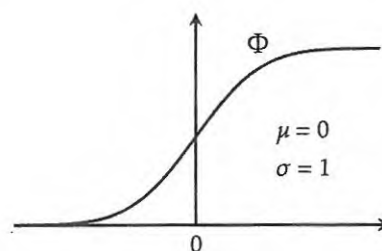
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad x \in \mathbb{R}$$

Fordelingsfunktionen, som vi betegner med Φ , er givet ved

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt, \quad x \in \mathbb{R}$$



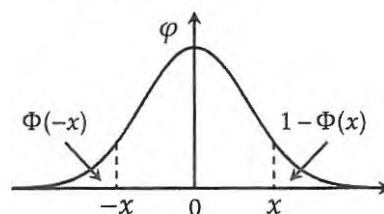
Figur 4.3.1



Figur 4.3.2

Idet $\varphi(-x) = \varphi(x)$, er φ en lige funktion, og Gauss-kurven er symmetrisk omkring y-aksen. Så er

$$\Phi(-x) = 1 - \Phi(x) \quad \text{og} \quad \Phi(0) = \frac{1}{2}$$



Figur 4.3.3

For en $N(0,1)$ -fordelt stokastisk variabel X er

$$P(X \leq a) = \Phi(a)$$

$$P(X \geq a) = 1 - P(X \leq a) = 1 - \Phi(a)$$

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = \Phi(b) - \Phi(a)$$

Det er uhyre sjældent at møde standardnormalfordelingen anvendt til at modellere virkelige data. I stedet anvender vi standardnormalfordelingen som en „referencefordeling“, som vi benytter til at bestemme egenskaber ved de øvrige normalfordelinger.

Ud fra fordelingsfunktionen Φ kan vi bestemme sandsynligheder i alle andre normalfordelinger ved brug af den følgende sætning.

Sætning 4.3.1. Antag, at $X \sim N(\mu, \sigma)$, og at Φ er fordelingsfunktionen for standardnormalfordelingen. Så er

- i) $P(X \leq a) = \Phi\left(\frac{a - \mu}{\sigma}\right)$
- ii) $P(X \geq a) = 1 - \Phi\left(\frac{a - \mu}{\sigma}\right)$
- iii) $P(a \leq X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$

Bevis

Når X er $N(\mu, \sigma)$ -fordelt, er

$$P(X \leq a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} \frac{1}{\sigma} dt$$

Vi vælger nu substitutionen

$$s = \frac{t - \mu}{\sigma} \quad \text{med} \quad ds = \frac{1}{\sigma} dt$$

og bemærker, at

$$s = \frac{t - \mu}{\sigma} \rightarrow -\infty \quad \text{for} \quad t \rightarrow -\infty$$

Substitutionen giver

$$P(X \leq a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{a-\mu}{\sigma}} e^{-\frac{1}{2}s^2} ds = \Phi\left(\frac{a - \mu}{\sigma}\right)$$

som beviser i).

Ved brug af i) får vi nu

$$\begin{aligned} P(X \geq a) &= 1 - P(X \leq a) = 1 - \Phi\left(\frac{a - \mu}{\sigma}\right) \\ P(a \leq X \leq b) &= P(X \leq b) - P(X \leq a) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right) \end{aligned}$$

som viser ii) og iii). □

Eksempel 4.3.1

Vægten af hjernen hos danske kvinder kan vi beskrive ved en normalfordelt stokastisk variabel X med middelværdi $\mu = 1300$ g og spredning $\sigma = 150$ g.

Sandsynligheden for, at hjernen hos en tilfældigt valgt dansk kvinde vejer mindre end 1500 g, er

$$P(X \leq 1500) = \Phi\left(\frac{1500 - 1300}{150}\right) = \Phi\left(\frac{4}{3}\right) = 0,9088 = 90,88 \%$$

Sandsynligheden for at hjernen vejer mellem 1 200 g og 1 400 g, er

$$\begin{aligned} P(1200 \leq X \leq 1400) &= P(X \leq 1400) - P(X \leq 1200) \\ &= \Phi\left(\frac{1400 - 1300}{150}\right) - \Phi\left(\frac{1200 - 1300}{150}\right) \\ &= \Phi\left(\frac{2}{3}\right) - \Phi\left(-\frac{2}{3}\right) \\ &= 0,7475 - 0,2525 = 0,4950 = 49,5 \% \end{aligned}$$

Sætning 4.3.2. Antag, at $X \sim N(\mu, \sigma)$, og lad k være et positivt tal. Så er

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) = 2\Phi(k) - 1$$

hvor Φ er fordelingsfunktionen for standardnormalfordelingen.

Bevis

Ved at benytte iii) i Sætning 4.3.1 får vi

$$\begin{aligned} P(\mu - k\sigma \leq X \leq \mu + k\sigma) &= \Phi\left(\frac{\mu + k\sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - k\sigma - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{k\sigma}{\sigma}\right) - \Phi\left(\frac{-k\sigma}{\sigma}\right) \\ &= \Phi(k) - \Phi(-k) \\ &= \Phi(k) - (1 - \Phi(k)) \\ &= 2\Phi(k) - 1 \end{aligned}$$

□

Sætning 4.3.2 viser os, at sandsynligheden for at en normalfordelt stokastisk X antager værdier i intervallet fra $\mu - k\sigma$ til $\mu + k\sigma$ er den samme for alle normalfordelinger.

For $k = 1$, $k = 2$ og $k = 3$ giver Sætning 4.3.2 os

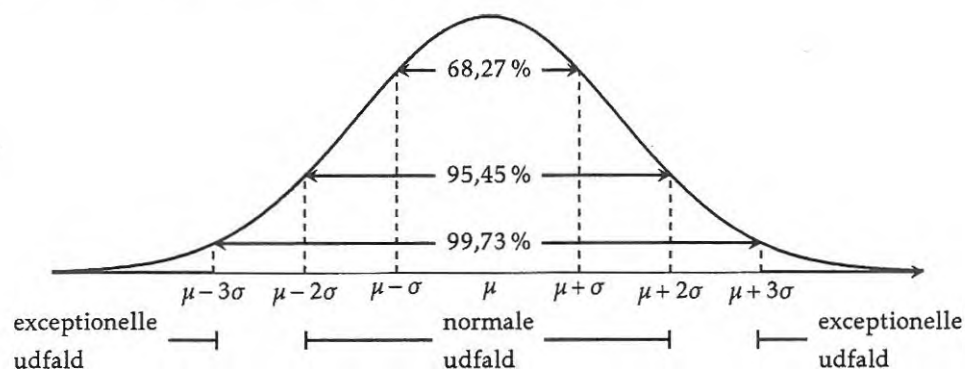
$$P(\mu - \sigma \leq X \leq \mu + \sigma) = 2\Phi(1) - 1 = 68,27 \%$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 2\Phi(2) - 1 = 95,45 \%$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 2\Phi(3) - 1 = 99,73 \%$$

Området inden for $\mu \pm 2\sigma$ anser vi i mange anvendelsessammenhænge som normalområdet, mens vi opfatter området uden for $\mu \pm 3\sigma$ som et ekstremområde, da dette (ubegrænsede) områdes samlede sandsynlighed kun udgør 0,3 %. Et udfald i dette område kalder vi et *exceptionelt* udfald.

De tilsvarende arealer er illustreret på Figur 4.3.4.



Figur 4.3.4

Eksempel 4.3.2

Ved produktion af chokoladepastiller regner producenten med, at massen af chokoladepastillerne er normalfordelt med middelværdi 0,93 g og spredning 0,03 g.

Normalområdet $\mu \pm 2\sigma$ svarer her til en masse fra $0,93 \text{ g} - 2 \cdot 0,03 = 0,87 \text{ g}$ til $0,93 \text{ g} + 2 \cdot 0,03 = 0,99 \text{ g}$, dvs. til en masse i området 0,87 g – 0,99 g.

En masse på under 0,84 g vil være et exceptionelt udfald.

☞ Øvelse 172, Øvelse 173, Øvelse 174

4.4 Fraktilfunktion og fraktilplot

Fraktilfunktion og fraktiler

Fordelingsfunktionen F for en normalfordelt stokastisk variabel er kontinuert og voksende på $\mathbb{R}$ med værdimængde $]0;1[$. Funktionen har dermed en omvendt funktion F^{-1} , defineret på $]0;1[$, som vi kalder *fraktilfunktionen* hørende til F .

Fraktilfunktionen hørende til Φ betegner vi med Φ^{-1} .

Eksempel 4.4.1

For den normalfordelte stokastiske variabel $X \sim N(25, 4)$ vil vi bestemme tallet q , så

$$P(X \leq q) = 0,80$$

Når F er fordelingsfunktionen hørende til X , skal vi løse ligningen $F(q) = 0,80$. Vi får med CAS

$$F(q) = 0,80 \quad \Leftrightarrow \quad q = F^{-1}(0,80) \approx 28,37$$

Eksempel 4.4.2

En normalfordelt stokastisk variabel X opfylder

$$P(X \leq 20) = 0,30 \quad \text{og} \quad P(X \geq 70) = 0,15$$

De to sandsynligheder bestemmer middelværdien μ og spredningen σ for X .

Af $P(X \leq 20) = 0,30$, får vi

$$\begin{aligned} \Phi\left(\frac{20-\mu}{\sigma}\right) &= 0,30 \\ \Leftrightarrow \frac{20-\mu}{\sigma} &= \Phi^{-1}(0,30) \\ \Leftrightarrow 20 &= \Phi^{-1}(0,30) \cdot \sigma + \mu \\ \Leftrightarrow 20 &= -0,5244 \cdot \sigma + \mu \end{aligned}$$

Når $P(X \geq 70) = 0,15$, er $P(X \leq 70) = 0,85$. Så finder vi

$$\begin{aligned} \Phi\left(\frac{70-\mu}{\sigma}\right) &= 0,85 \\ \Leftrightarrow 70 &= \Phi^{-1}(0,85) \cdot \sigma + \mu \\ \Leftrightarrow 70 &= 1,0364 \cdot \sigma + \mu \end{aligned}$$

Vi har således de to ligninger

$$20 = -0,5244 \cdot \sigma + \mu \quad \text{og} \quad 70 = 1,0364 \cdot \sigma + \mu$$

i μ og σ , som ved løsning giver os

$$\mu \approx 36,8 \quad \text{og} \quad \sigma \approx 32,03$$

➤ Øvelse 175, Øvelse 176, Øvelse 177, Øvelse 178

Definition 4.4.1. Givet en normalfordelt stokastisk variabel $X \sim N(\mu, \sigma)$ med fordelingsfunktion F og et tal $\alpha \in]0; 1[$.

Så er α -fraktilen det entydigt bestemte tal x_α , som opfylder

$$P(X \leq x_\alpha) = \alpha$$

α -fraktilen x_α beskriver vi også som $100 \cdot \alpha$ %-fraktilen.

0,25-fraktilen eller 25 %-fraktilen kalder vi den *nedre kvartil*.

0,50-fraktilen eller 50 %-fraktilen kalder vi *medianen*.

0,75-fraktilen eller 75 %-fraktilen kalder vi den *øvre kvartil*.

Bemærk, at medianen er middelværdien μ .

Sætning 4.4.1. Givet en normalfordelt stokastisk variabel $X \sim N(\mu, \sigma)$ med fordelingsfunktion F . Så er α -fraktilen for X

$$x_\alpha = F^{-1}(\alpha)$$

Bevis

Vi har

$$P(X \leq x_\alpha) = \alpha \Leftrightarrow F(x_\alpha) = \alpha \Leftrightarrow x_\alpha = F^{-1}(\alpha)$$

□

Ifølge Sætning 4.4.1 kan vi bestemme fraktiler som funktionsværdier for fraktilfunktionen; deraf dens navn.

Eksempel 4.4.3

For den normalfordelte stokastiske variabel $X \sim N(25, 4)$ i Eksempel 4.4.1 fandt vi

$$F^{-1}(0,80) = 28,37$$

Altså er 0,80-fraktilen eller 80 %-fraktilen for X lig med 28,37.

Endvidere er

$$F^{-1}(0,10) = 19,87$$

$$F^{-1}(0,25) = 22,30$$

$$F^{-1}(0,75) = 27,70$$

$$F^{-1}(0,90) = 30,13$$

så den nedre kvartil for X er 22,30, og den øvre kvartil er 27,70.

10 %-fraktilen på 19,87 og 90 %-fraktilen på 30,13 er eksempler på *deciler*.

Eksempel 4.4.4

For raske 10-uger gamle huskattekillinger er vægten normalfordelt med middelværdi 690 g og spredning 140 g. Vi kan således beskrive vægten af en tilfældigt valgt kattekillling ved en normalfordelt stokastisk variabel $X \sim N(690, 140)$.

Medianen er 690 g, og kvartilerne er

$$F^{-1}(0,25) = 595,6$$

$$F^{-1}(0,75) = 784,4$$

Der er således 25 % sandsynlighed for, at en tilfældigt valgt kattekillling vejer under 595,6 g, og 75 % sandsynlighed for at en tilfældigt valgt kattekillling vejer under 784,4 g.

Sandsynligheden for, at en tilfældigt valgt kattekillling vejer over 784,4 g, er så 25 %.

➤ Øvelse 179, Øvelse 180, Øvelse 181

Når F er fordelingsfunktionen for normalfordelingen $N(\mu, \sigma)$, er

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

som viser, at punkterne $(x, \Phi^{-1}(F(x)))$ ligger på en ret linje med hældningskoefficient $\frac{1}{\sigma}$ og konstantled $-\frac{\mu}{\sigma}$.

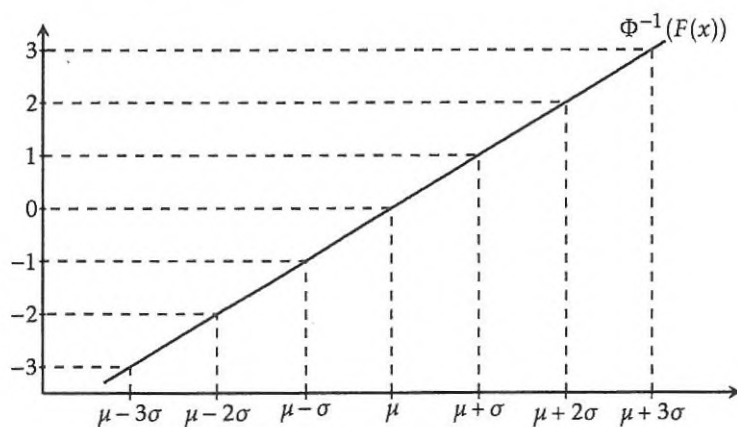
Nogle udvalgte funktionsværdier fremgår af Tabel 4.4.1 nedenfor.

x	$\mu - 3\sigma$	$\mu - 2\sigma$	$\mu - \sigma$	μ	$\mu + \sigma$	$\mu + 2\sigma$	$\mu + 3\sigma$
$\Phi^{-1}(F(x))$	-3	-2	-1	0	1	2	3

Tabel 4.4.1

Den rette linje, som er graf for $\Phi^{-1}(F(x))$, tegner vi typisk i et *fraktilplot*.

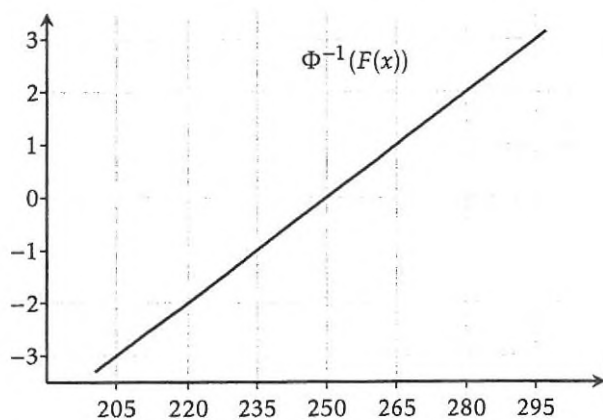
I et fraktilplot er begge koordinatakser ækvidistante, og Figur 4.4.1 viser den generelle opbygning af et fraktilplot.



Figur 4.4.1

Eksempel 4.4.5

Vi antager, at X er $N(250, 15)$ -fordelt. Fraktilplottet for X kan vi tegne som den rette linje, der går gennem $(250, 0)$ og $(280, 2)$, se Figur 4.4.2.



Figur 4.4.2

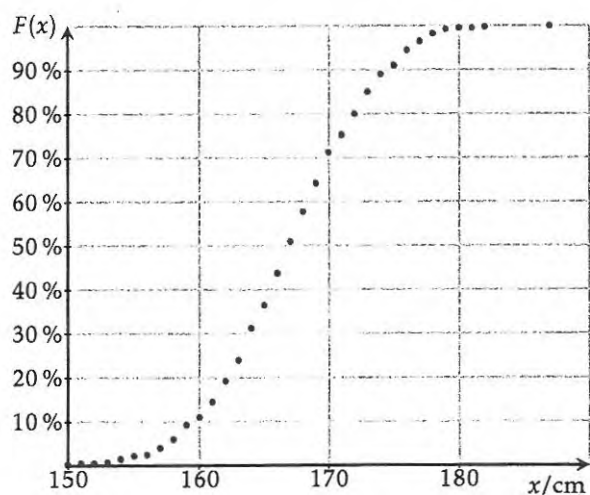
Eksempel 4.4.6

En opmåling af højderne på 400 gymnasiepiger har resulteret i de kumulerede frekvenser i Tabel 4.4.2.

x	$h(x)$	$F(x)$	x	$h(x)$	$F(x)$
150	1	0,25 %	167	29	51,00 %
151	1	0,50 %	168	27	57,75 %
152	0	0,50 %	169	26	64,25 %
153	1	0,75 %	170	28	71,25 %
154	3	1,50 %	171	16	75,25 %
155	3	2,25 %	172	19	80,00 %
156	1	2,50 %	173	20	85,00 %
157	6	4,00 %	174	16	89,00 %
158	8	6,00 %	175	8	91,00 %
159	13	9,25 %	176	14	94,50 %
160	7	11,00 %	177	8	96,50 %
161	14	14,50 %	178	7	98,25 %
162	19	19,25 %	179	4	99,25 %
163	19	24,00 %	180	1	99,50 %
164	29	31,25 %	181	0	99,50 %
165	21	36,50 %	182	1	99,75 %
166	29	43,75 %	187	1	100,00 %

Tabel 4.4.2

De kumulerede frekvenser giver os punkterne på sumkurven vist i Figur 4.4.3.



Figur 4.4.3

Punkterne ligger ret symmetrisk, og formen på en tegnet kurve gennem punkterne minder meget om grafen for fordelingsfunktionen for en normalfordeling. Vi vil derfor modellere data ved hjælp af en normalfordeling.

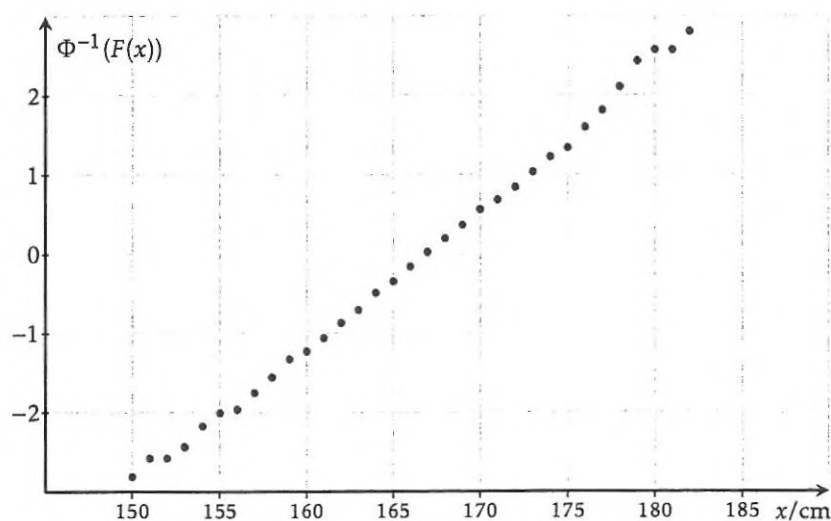
Vi tegner et fraktilplot hørende til de kumulerede data.

Med tallene fra Tabel 4.4.2, får vi Tabel 4.4.3.

x	$\Phi^{-1}(F(x))$	x	$\Phi^{-1}(F(x))$	x	$\Phi^{-1}(F(x))$
150	-2,80703	161	-1,05812	172	0,841621
151	-2,57583	162	-0,86872	173	1,036433
152	-2,57583	163	-0,7063	174	1,226528
153	-2,43238	164	-0,48878	175	1,340755
154	-2,17009	165	-0,34513	176	1,598193
155	-2,00465	166	-0,15731	177	1,811911
156	-1,95996	167	0,025069	178	2,108358
157	-1,75069	168	0,195502	179	2,432379
158	-1,55477	169	0,365149	180	2,575829
159	-1,32552	170	0,560703	181	2,575829
160	-1,22653	171	0,682378	182	2,807034

Tabel 4.4.3

Vi indsætter tabellens tal i et fraktilplot som vist i Figur 4.4.4.



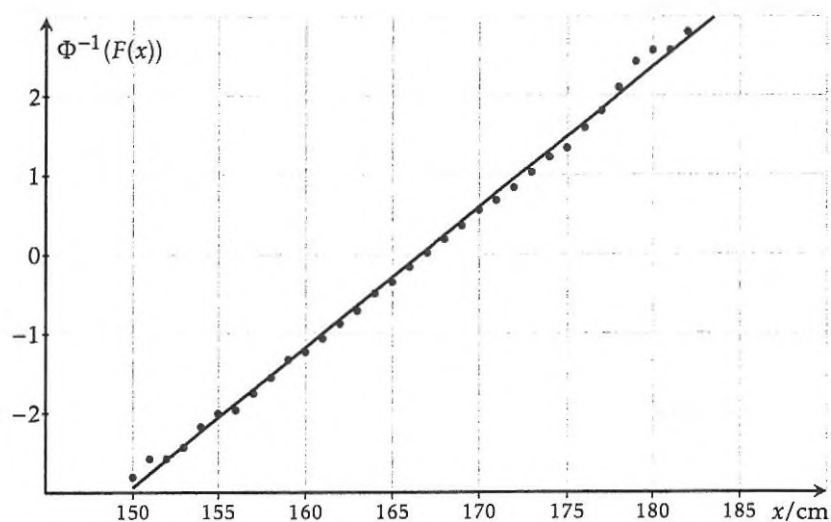
Figur 4.4.4

Punkterne beskriver med god tilnærmelse en ret linje, og lineær regression giver os sammenhængen

$$y = 0,1771x - 29,50$$

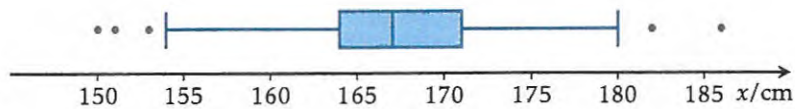
Heraf får vi

$$\sigma = \frac{1}{0,1771} = 5,65 \quad \text{og} \quad \mu = \frac{29,50}{0,1771} = 166,6$$



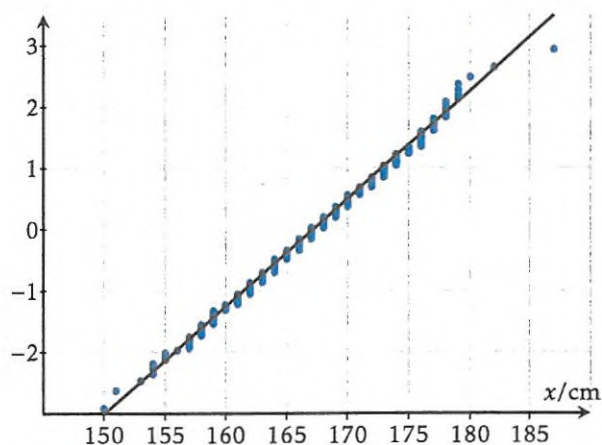
Figur 4.4.5

Ser vi på hyppighederne i Tabel 4.4.2, kan vi få en mistanke om, at de målte pigehøjder indeholder outliers. Vi benytter derfor et program til at tegne et boksplot ud fra de 400 rådata. Boksplottet, som er vist i Figur 4.4.6, afslører fem outliers.



Figur 4.4.6

Ud fra rådata kan vi ligeledes benytte et program til at tegne et fraktilplot.

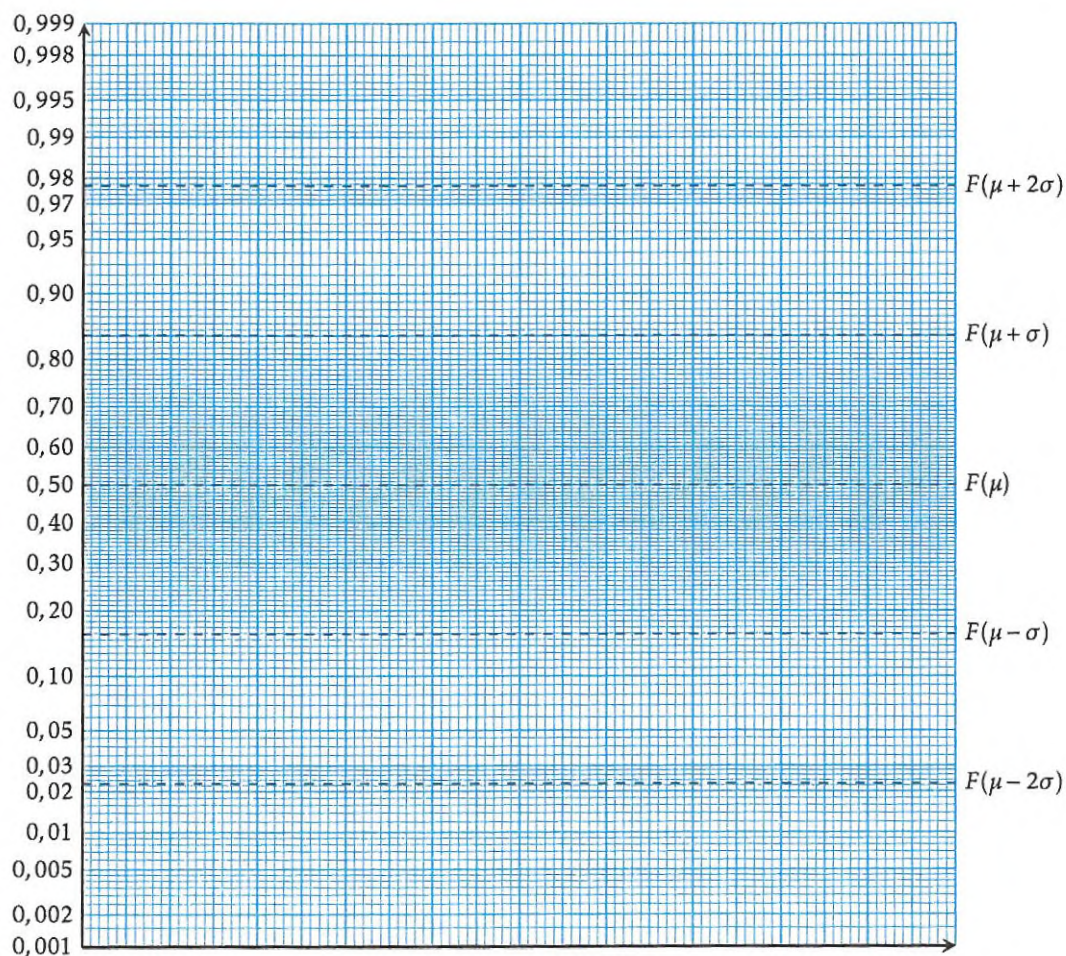


Figur 4.4.7

Normalfordelingspapir

Vi kan tegne grafen for fordelingsfunktionen F for en normalfordeling i et koordinatsystem, hvor andenaksen indeholder sandsynligheder, og hvis inddeling er baseret på normalfordelingssandsynligheder. Førsteaksen er en sædvanlig akse med ækvidistant inddeling.

Koordinatsystemet finder vi på specialpapiret kaldet normalfordelingspapir.



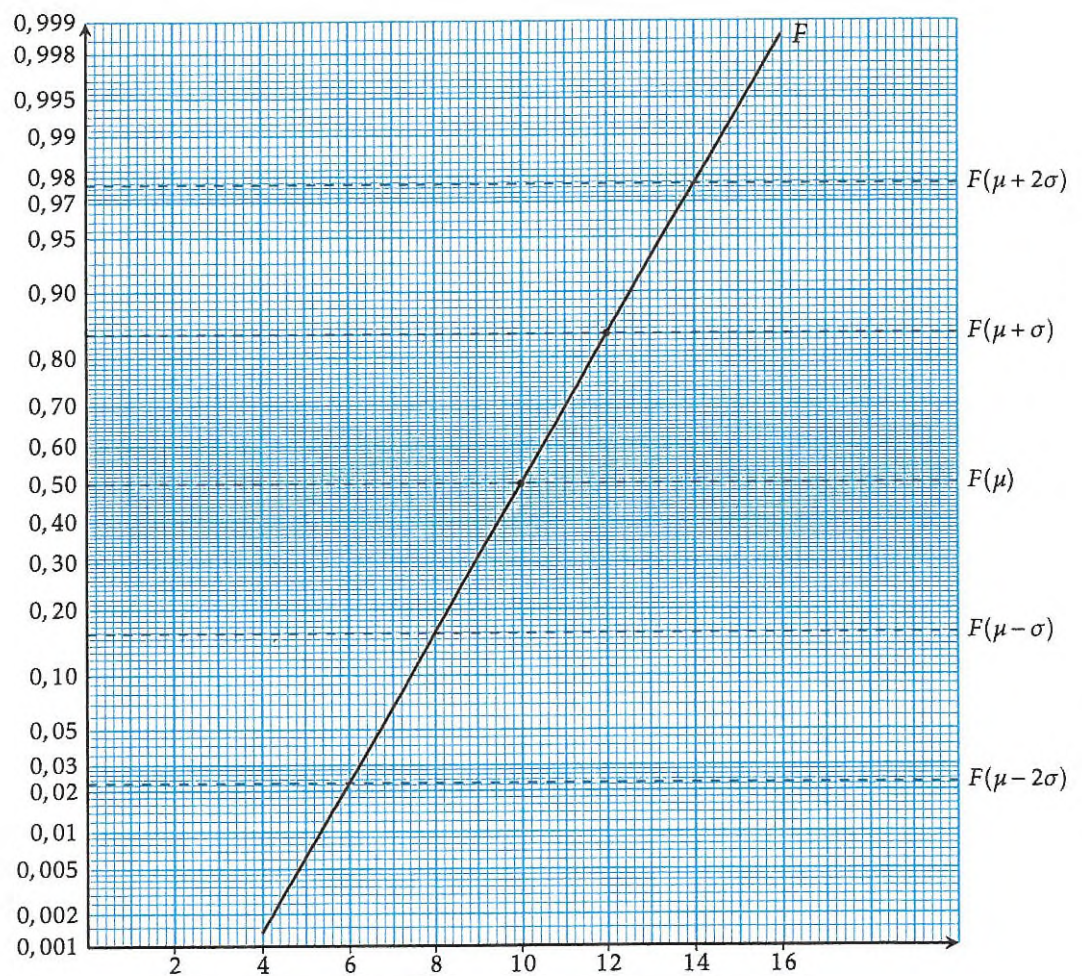
Figur 4.4.8

På grund af andenaksens skalering kan vi ikke plotte hele fordelingsfunktionen, idet halerne ikke kommer med, men vi kan til fulde tegne grafen på normalområdet $\mu \pm 2\sigma$.

Grafen for fordelingsfunktionen F for en normalfordeling er en ret linje i dette koordinatsystem, som vi f.eks. kan tegne som den rette linje gennem punkterne $(\mu, F(\mu))$ og $(\mu + \sigma, F(\mu + \sigma))$.

Eksempel 4.4.7

Grafen for fordelingsfunktionen F for den stokastiske variable $X \sim N(10, 2)$ er den rette linje indtegnet på Figur 4.4.9.



Figur 4.4.9

Generelt vil en voksende ret linje tegnet på et normalfordelingspapir være graf for netop én fordelingsfunktion. Ud fra grafen kan vi nemlig ved aflæsning fastlægge normalfordelingens middelværdi μ og bestemme dens spredning σ .

Middelværdien aflæser vi som den værdi x_1 på x -aksen, der svarer til 50% på y -aksen, da $F(\mu) = 50\%$.

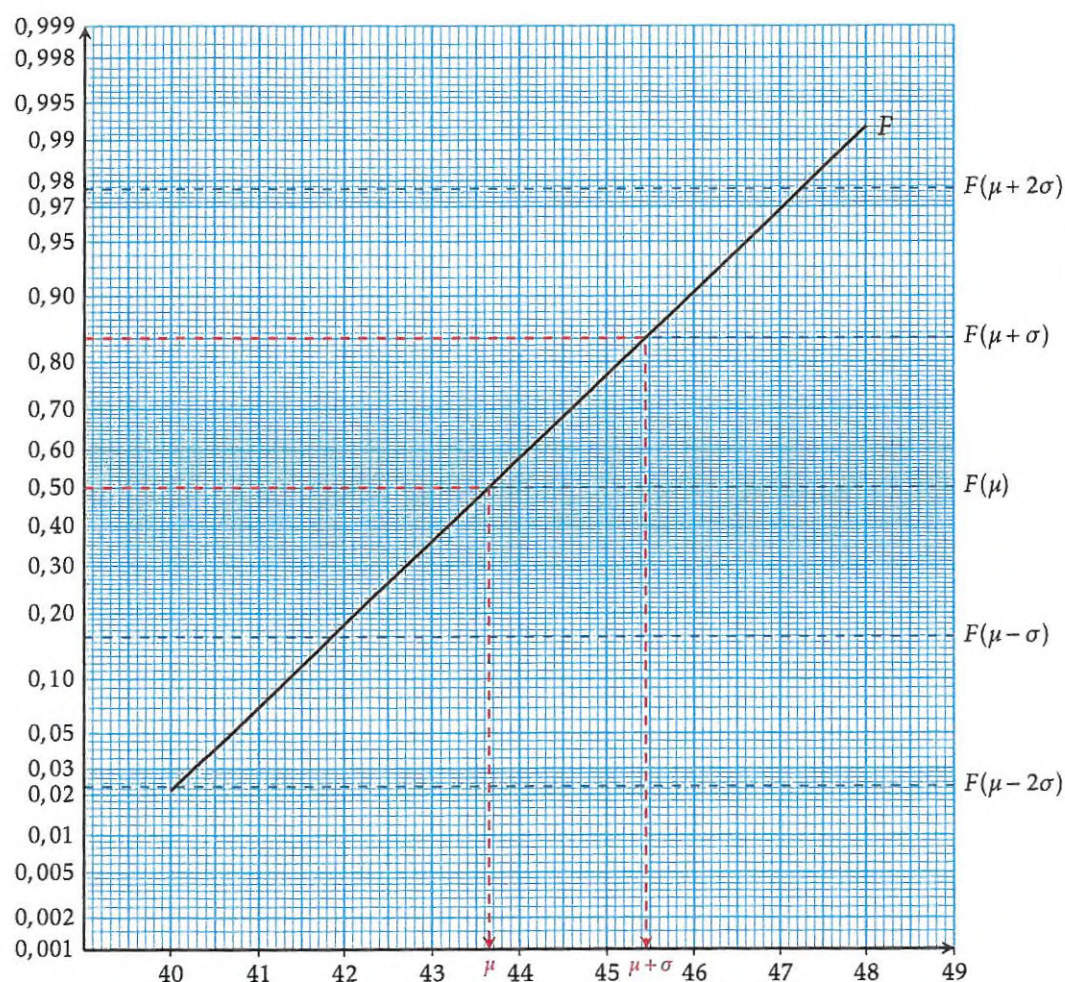
Herefter finder vi spredningen ved f.eks. at aflæse ud fra $F(\mu + \sigma)$ på y -aksen. Den tilhørende x -værdi x_2 er så $\mu + \sigma$.

Dermed er

$$x_1 = \mu \quad \text{og} \quad \sigma = x_2 - x_1$$

Eksempel 4.4.8

Den rette linje i Figur 4.4.10 er graf for fordelingsfunktionen F for en normalfordeling.



Figur 4.4.10

Af grafen aflæser vi

$$\mu = 43,65 \quad \text{og} \quad \mu + \sigma = 45,45$$

Den rette linje svarer således til en normalfordeling med middelværdi $\mu = 43,65$ og spredning $\sigma = 45,45 - 43,65 = 1,8$.

Eksempel 4.4.9

Ved en klinisk undersøgelse har vi undersøgt kropstemperaturen x (i $^{\circ}\text{C}$) hos raske mennesker.

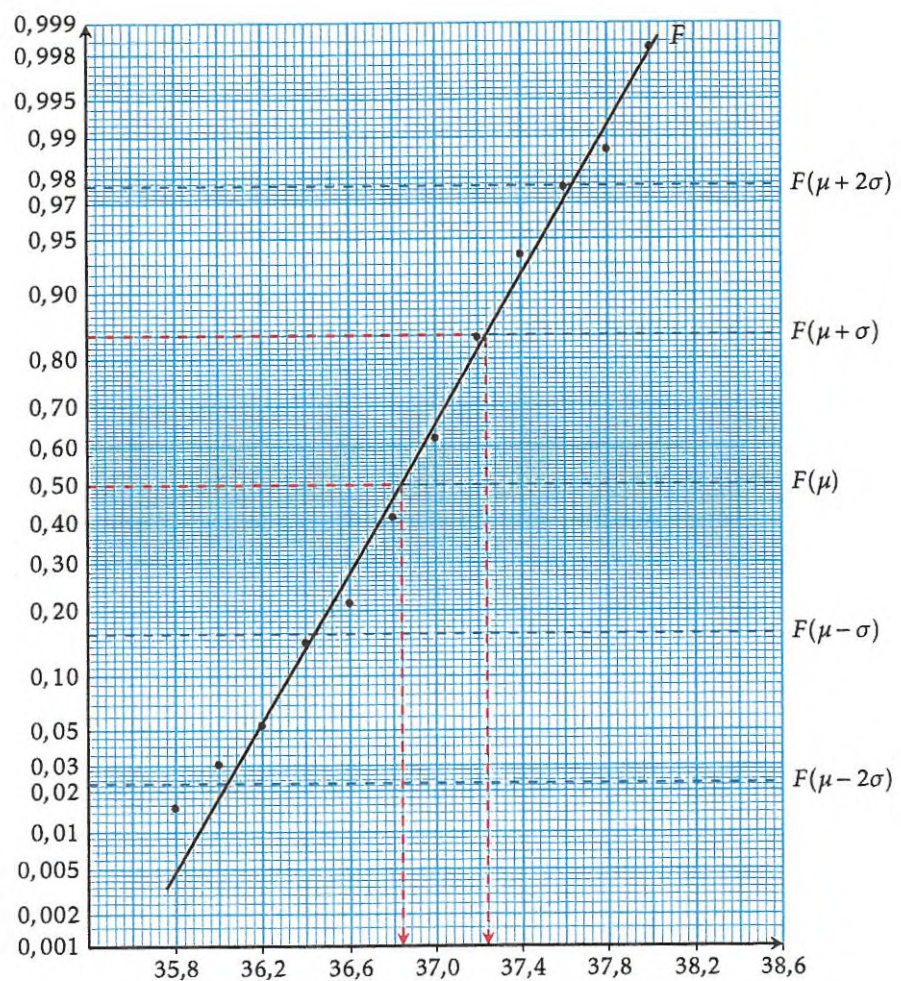
De kumulerede frekvenser af observationerne er vist i Tabel 4.4.4, og disse data indtegner vi på et normalfordelingspapir.

Hvis det er muligt at indtegne en ret linje, som punkterne tilnærmelsesvis ligger på, beskriver vi punkterne ved hjælp af den normalfordeling, der er fastlagt ved den indtegnede linje.

Dette har vi gjort i Figur 4.4.11.

x_i °C	$F(x_i)$ %
35,8	1,54
36,0	3,08
36,2	5,38
36,4	14,62
36,6	21,54
36,8	41,54
37,0	62,31
37,2	83,85
37,4	93,85
37,6	97,69
37,8	98,76
38,0	99,83

Tabel 4.4.4



Figur 4.4.11

Ved aflæsning på figuren finder vi

$$\mu = 36,84 \quad \text{og} \quad \sigma = 37,24 - 36,84 = 0,40$$

Normalområdet for denne fordeling er så bestemt ved grænserne

$$\mu + 2 \cdot \sigma = 36,84 + 2 \cdot 0,4 = 37,64$$

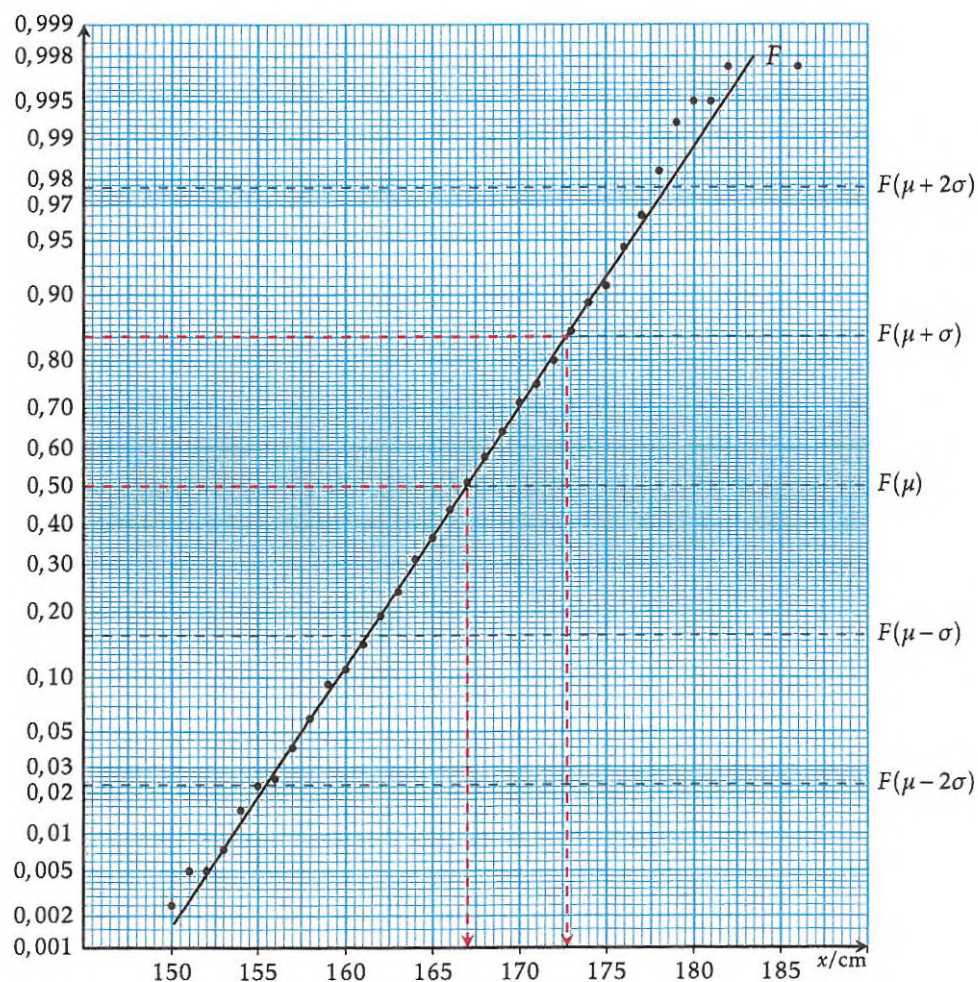
$$\mu - 2 \cdot \sigma = 36,84 - 2 \cdot 0,4 = 36,04$$

Dvs. at normalområdet for kropstemperaturer er intervallet 36,0 – 37,6 °C.

Temperaturer under 35,6°C og over 38,0°C er exceptionelle.

Eksempel 4.4.10

Vi plotter de kumulerede frekvenser for pigehøjder fra Eksempel 4.4.6 og indtegner „den bedste rette linje“ på øjemål. De plottede punkter og et bud på linjen fremgår af figur 4.4.12.



Figur 4.4.12

Ved brug af figuren aflæser vi

$$\mu = 167 \quad \text{og} \quad \sigma = 172,75 - 167 = 5,75$$

Vi kan derfor i vor model beskrive pigehøjderne som normalfordelte med midelværdi 167 cm og spredning 5,75 cm.