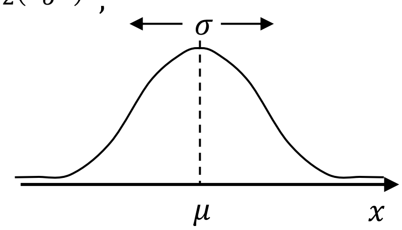
Normaalijakauma

TODENNÄKÖISYYS JA TILASTOT, MAA10

Normaalijakauma on todennäköisyysjakaumista ”suurin ja kaunein” eli tärkein. Sen keksi/ löysi Carl Friedrich Gauss (1777-1855, Saksa) tutkiessaan planeettojen ratojen mittausvirheitä.

Määritelmä, *normaalijakauma*:
Jatkuva satunnaismuuttuja $\underline{x}$ noudattaa normaalijakaumaa parametrein μ ja σ , jos sen tiheysfunktio f on muotoa

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



jolloin merkitään $\underline{x} \sim N(\mu, \sigma)$. Funktion f kuvaaja on symmetrinen suoran $x = \mu$ suhteen. Käyrää $y = f(x)$ kutsutaan *Gaussin käyräksi* tai *kellokäyräksi*.

Huomautus: Huippu on odotusarvon μ kohdalla ja mitä isompi on σ sitä leveämpi jakauma, eli σ antaa normaalijakauman muodon.

Carl Friedrich Gauss

Normaalijakauman tiheysfunktiossa oleva kirjain e on symboli *Neperin luvulle* (kurssit 8 →)

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^{n} \approx 2{,}71828.$$

Normaalijakauman tärkeys ilmenee esimerkiksi yleisyytenä luonnon ja käytännön ilmiöille: *"Keskimääräisyydet" ovat tavallisimpia ja äärimmäisyydet harvinaisempia.*

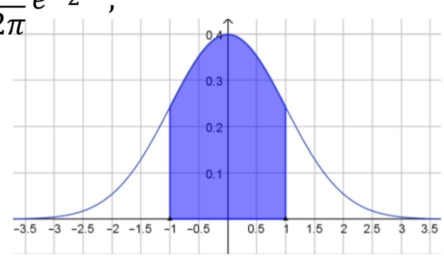
Erityisesti tärkeys ilmenee todennäköisyyslaskennan (yhtenä) tärkeimpänä tuloksena: *Keskeinen raja-arvolause,* jossa on lyhyesti kyse siitä, että samoin jakautuneiden riippumattomien (jatkuvien tai diskreettien) satunnaismuuttujien $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ *summan jakauma on likimain normaalisti jakautunut*, vaikka satunnaismuuttujien $\underline{x}_i$ itsessään ei tarvitse olla normaalijakautuneita. Huom: "samoin jakautuneisuus" tarkoittaa, että kaikilla $\underline{x}_i$ on sama odotusarvo μ ja keskihajonta σ (vaihtoehtoisesti varianssi σ^2).

Määritelmä, *normitettu eli standardoitu normaalijakauma*:
Jos $\underline{x} \sim N(0{,}1)$, niin sen tiheysfunktiolle φ (eikä f) pätee

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

missä φ on pieni fii.

Eli $N(0{,}1)$ jakauman odotusarvona on 0 ja keskihajonta 1.



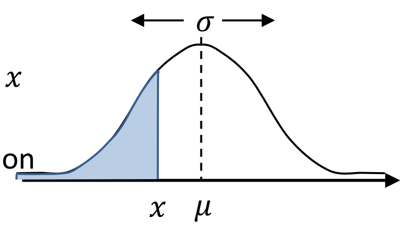
Huomautus:
- Yleisen normaalijakauman kertymäfunktiota merkitään F, mutta standardinormaalijakauman kertymäfunktiota Φ (iso fii). Syy: vakiintunut merkintätapa
- Tärkeää! Usein (kirjallisuus) satunnaismuuttujan $\underline{x}$ normaalijakaantuneisuus ilmoitetaan parametrien $\mu =$ odotusarvo ja $\sigma^2 =$ varianssi. **Tarkista** mitä parametreja kyseinen kirja käyttää ja muista keskihajonnan ja varianssin välinen yhteys, $\sigma = \sqrt{\sigma^2}$.

Huomautus (jatkuu):

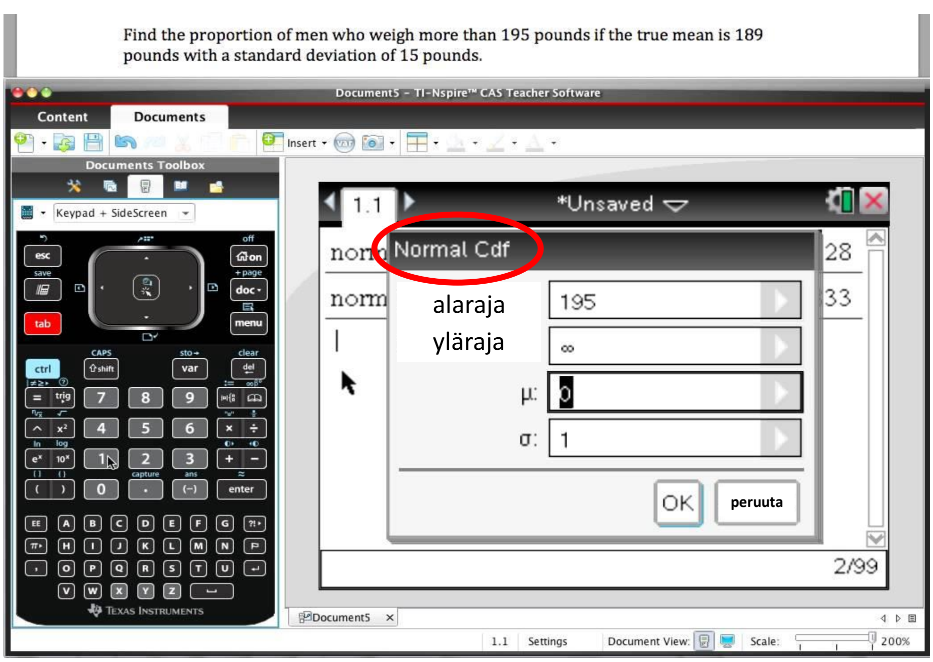
- Kaikille normaalijakaumille $N(\mu, \sigma)$ pätevät seuraavat ominaisuudet:
  - välillä $\mu \pm \sigma$ on noin 68 % muuttujan arvoista,
  - välillä $\mu \pm 2\sigma$ on noin 95 % muuttujan arvoista,
  - välillä $\mu \pm 3\sigma$ on noin 99,7 % muuttujan arvoista.

  Standardinormaalijakaumalle ko. välit ovat $[-1,1]$, $[-2,2]$ ja $[-3,3]$.

Todennäköisyydet:
Kertymäfunktion F tai Φ arvo kohdassa x tarkoittaa todennäköisyyttä $P(\underline{x} \leq x)$. Geometrinen tulkinta on pinta-ala, jota on ”kertynyt” kohtaan x asti.



Kertymäfunktion Φ arvot saadaan MAOLista tai laskimesta.
TI-nspire: Laskentatila → menu→ 5:Todennäköisyyslaskenta→ 5:Jakaumat…→1:Normaali Pdf (ei tarvita) TAI 2:Normaali Cdf (tarvitaan!).

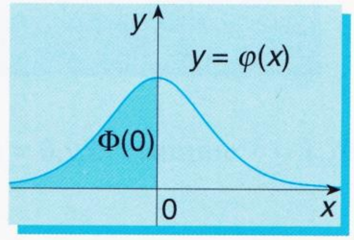


Miesten paino (massa) on normaalijakautunut parametrein 189 paunaa (odotusarvo) ja 15 paunaa (keskihajonta). Etsi yli 199 paunaa painavien miesten osuus. $1\text{lb} \triangleq 453{,}59\text{g}$.

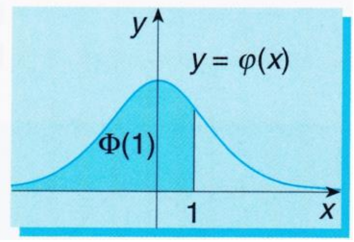
Esimerkki 1: Olkoon $\underline{x} \sim N(0,1)$. Laske **a)** $P(\underline{x} \leq 0)$, **b)** $P(\underline{x} \leq 1)$ ja **c)** $P(\underline{x} \geq 1)$.

a) Tiheysfunktion φ kuvaaja on symmetrinen y-akselin suhteen, joten "todennäköisyysmassasta" on puolet y-akselin vasemmalla puolella. Siis

$$P(\underline{x} \leq 0) = \Phi(0) = \frac{1}{2}.$$

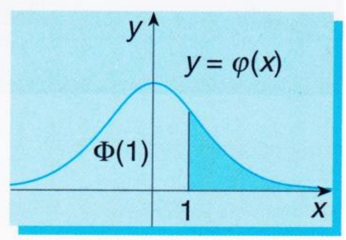


b) Taulukosta näemme tai laskimella saamme, että $P(\underline{x} \leq 1) = \Phi(1) = 0{,}8413$. (Tällaisessa yhteydessä on mukava käyttää yhtäsuuruusmerkkiä, vaikka kyseessä on likiarvo.)



c) Näemme taulukosta suoraan vain muotoa $P(\underline{x} \leq a) = \Phi(a)$ olevat todennäköisyydet, kun $a \geq 0$. Siksi käytämme komplementtisääntöä

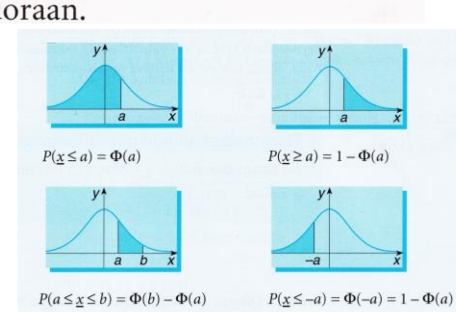
$$\begin{aligned} P(\underline{x} \geq 1) &= 1 - P(\underline{x} < 1) \\ &= 1 - P(\underline{x} \leq 1) = 1 - \Phi(1) \\ &= 1 - 0{,}8413 = 0{,}1587. \end{aligned}$$



Laskimella saamme tuloksen suoraan.

Yleisesti normitetun normaalijakauman todennäköisyydet saadaan yhtälöistä (katso myös kuva oikealla)

$$P(\underline{x} \leq a) = \Phi(a)$$
$$P(\underline{x} \geq a) = 1 - \Phi(a)$$
$$P(a \leq \underline{x} \leq b) = \Phi(b) - \Phi(a)$$
$$P(\underline{x} \leq -a) = \Phi(-a) = 1 - \Phi(a)$$

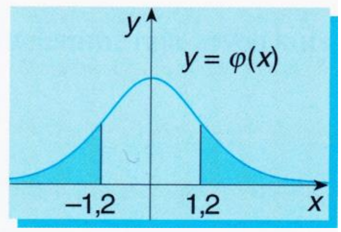


Esimerkki 2: Olkoon $\underline{x} \sim N(0{,}1)$. Laske **a)** $P(\underline{x} \leq 1{,}2)$, **b)** $P(0{,}64 \leq \underline{x} \leq 2{,}37)$ ja **c)** $P(-0{,}64 \leq \underline{x} \leq 2{,}37)$.

Laskimella saadaan kaikki tulokset suoraan. Taulukkoa käyttäen menetellään seuraavasti (MAOL).

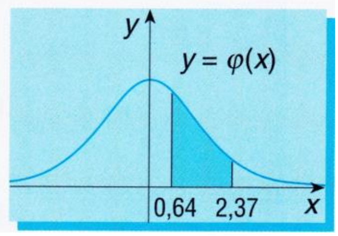
a) Symmetrisyyden perusteella

$$\begin{aligned} P(\underline{x} \leq -1{,}2) &= P(\underline{x} \geq 1{,}2) \\ &= 1 - P(\underline{x} < 1{,}2) \\ &= 1 - \Phi(1{,}2) \\ &= 1 - 0{,}8849 = 0{,}1151. \end{aligned}$$



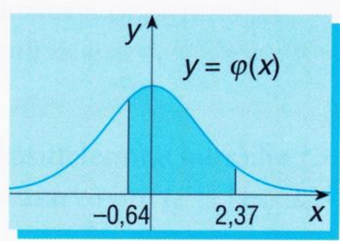
b) $P(0{,}64 \leq \underline{x} \leq 2{,}37)$

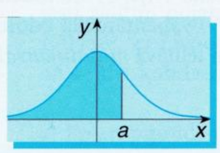
$$\begin{aligned} &= \Phi(2{,}37) - \Phi(0{,}64) \\ &= 0{,}9911 - 0{,}7389 = 0{,}2522. \end{aligned}$$



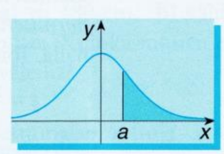
c) $P(-0{,}64 < \underline{x} < 2{,}37)$

$$\begin{aligned} &= P(-0{,}64 \leq \underline{x} \leq 2{,}37) \\ &= \Phi(2{,}37) - \Phi(-0{,}64) \\ &= \Phi(2{,}37) - (1 - \Phi(0{,}64)) \\ &= \Phi(2{,}37) + \Phi(0{,}64) - 1 \\ &= 0{,}9911 + 0{,}7389 - 1 = 0{,}7300. \end{aligned}$$

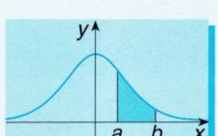




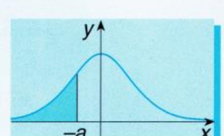
$P(\underline{x} \leq a) = \Phi(a)$



$P(\underline{x} \geq a) = 1 - \Phi(a)$



$P(a \leq \underline{x} \leq b) = \Phi(b) - \Phi(a)$



$P(\underline{x} \leq -a) = \Phi(-a) = 1 - \Phi(a)$