Statistiek MBW/KW deel 1 week 4

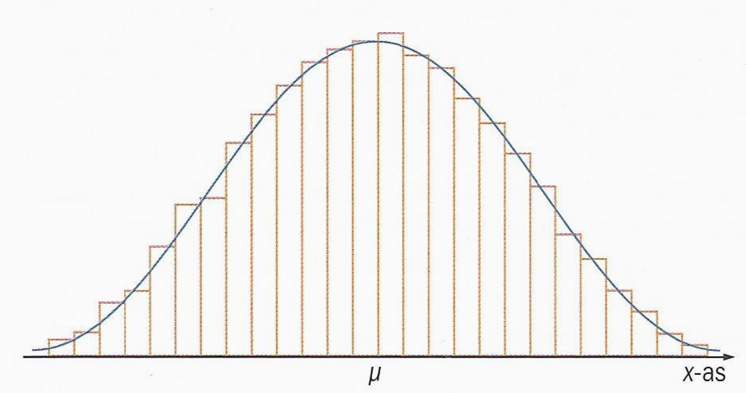
- Normale verdeling
- Standaard normale verdeling, z-waarden
- Toetsje

Normale verdeling H5

De meest gebruikte continue kansverdeling is de **normale verdeling**. Deze verdeling heeft twee parameters: de verwachtingswaarde $\boldsymbol{\mu}$ en de standaarddeviatie $\boldsymbol{\sigma}$. Het beschrijft meetwaarden die zijn samengesteld uit zeer veel kleine random effecten, zoals de leeftijdsverdeling van vrouwen van 22 jaar oud, de snelheid van een passerende auto, het vulgewicht van een kilopak suiker, het jaarinkomen van een Nederlands huishouden.

De kansdichtheidsfunctie van de normale verdeling heeft de volgende vorm:

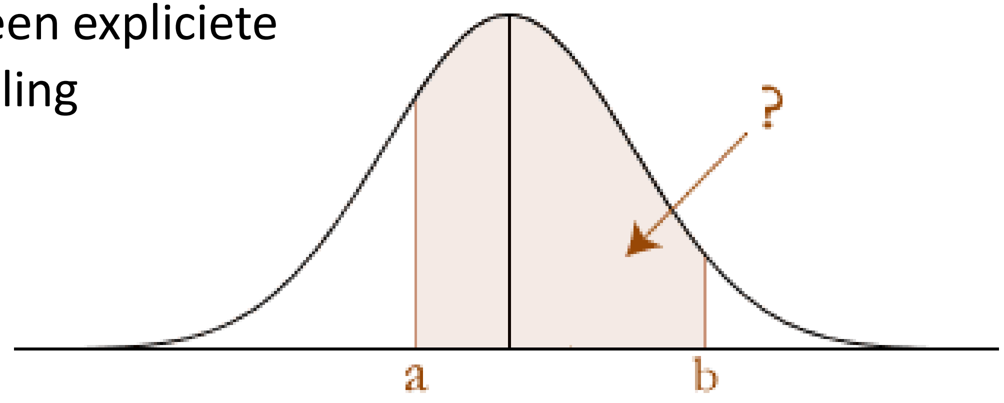
FIGUUR 5.1 Kenmerkende vorm van een normale verdeling



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

Voor de cumulatieve kansverdelingsfunctie $F(x) = P(\underline{x} \leq x)$ is geen expliciete uitdrukking. In het algemeen kun je kansen voor de normale verdeling op de GR uitrekenen als

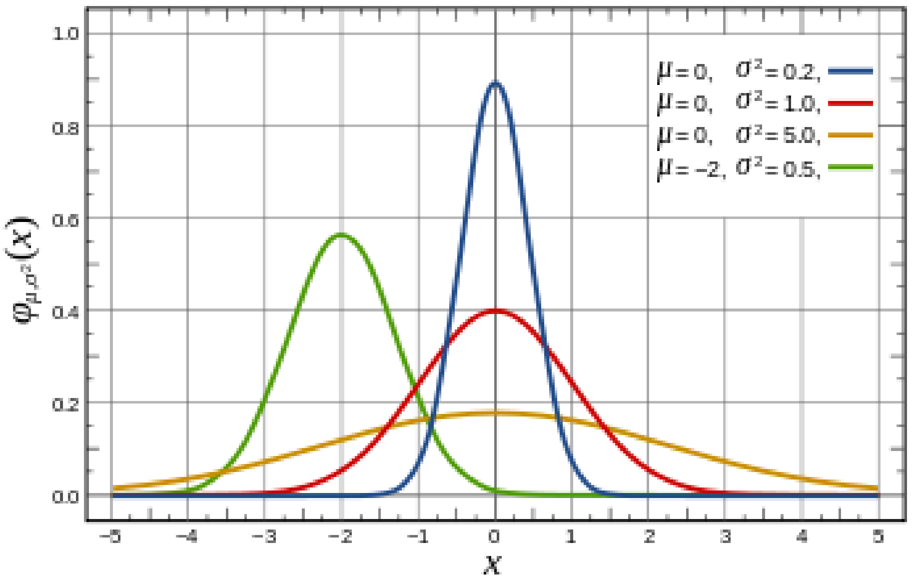
$$P(a \leq \underline{x} \leq b) = \text{normalcdf}(x = a, x = b, \mu, \sigma)$$



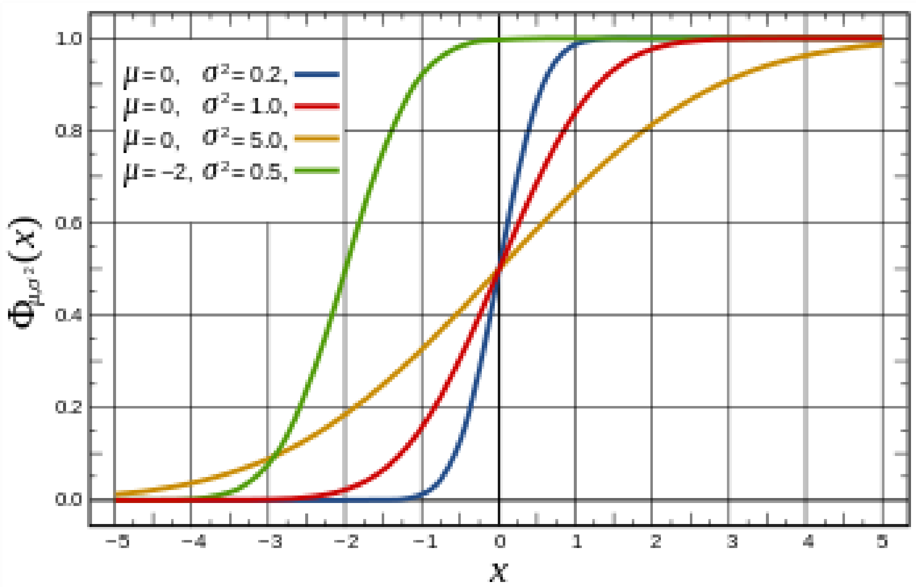
De GR functie normalpdf zul je nooit gebruiken.

Notatie om aan te geven dat een variabele $\underline{x}$ zich gedraagt als een normale verdeling is: $\underline{x} \sim N(\mu; \sigma)$.

Normale verdeling H5



PDF Probability Density Function $f(x)$



CDF Cumulative Density Function $F(x)$

De grafiek van de PDF is symmetrisch en heeft een maximum bij μ. De breedte van de grafiek wordt bepaald door de standaarddeviatie σ.

Normale verdeling H5

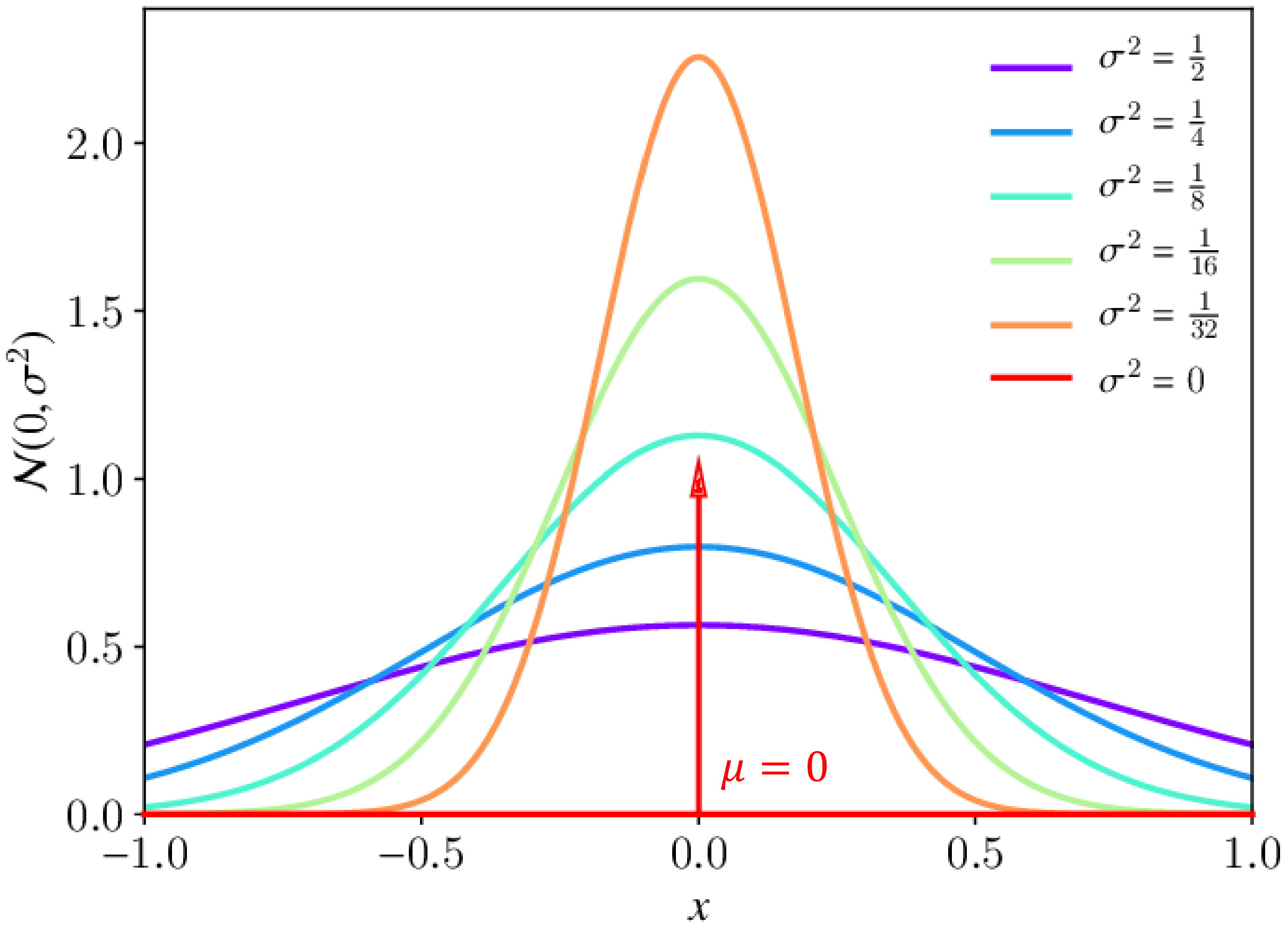
Eigenschappen van de normale (ofwel Gauss-) verdeling zijn:

De maximale waarde van de kansdichtheidsfunctie ligt bij $x = \mu$. De kans is groot dat een meetwaarde hier dicht in de buurt ligt.

De verdeling is symmetrisch rond $x = \mu$. Dus: $P\left(\underline{x} \leq \mu\right) = P\left(\underline{x} \geq \mu\right) = 0{,}5$.

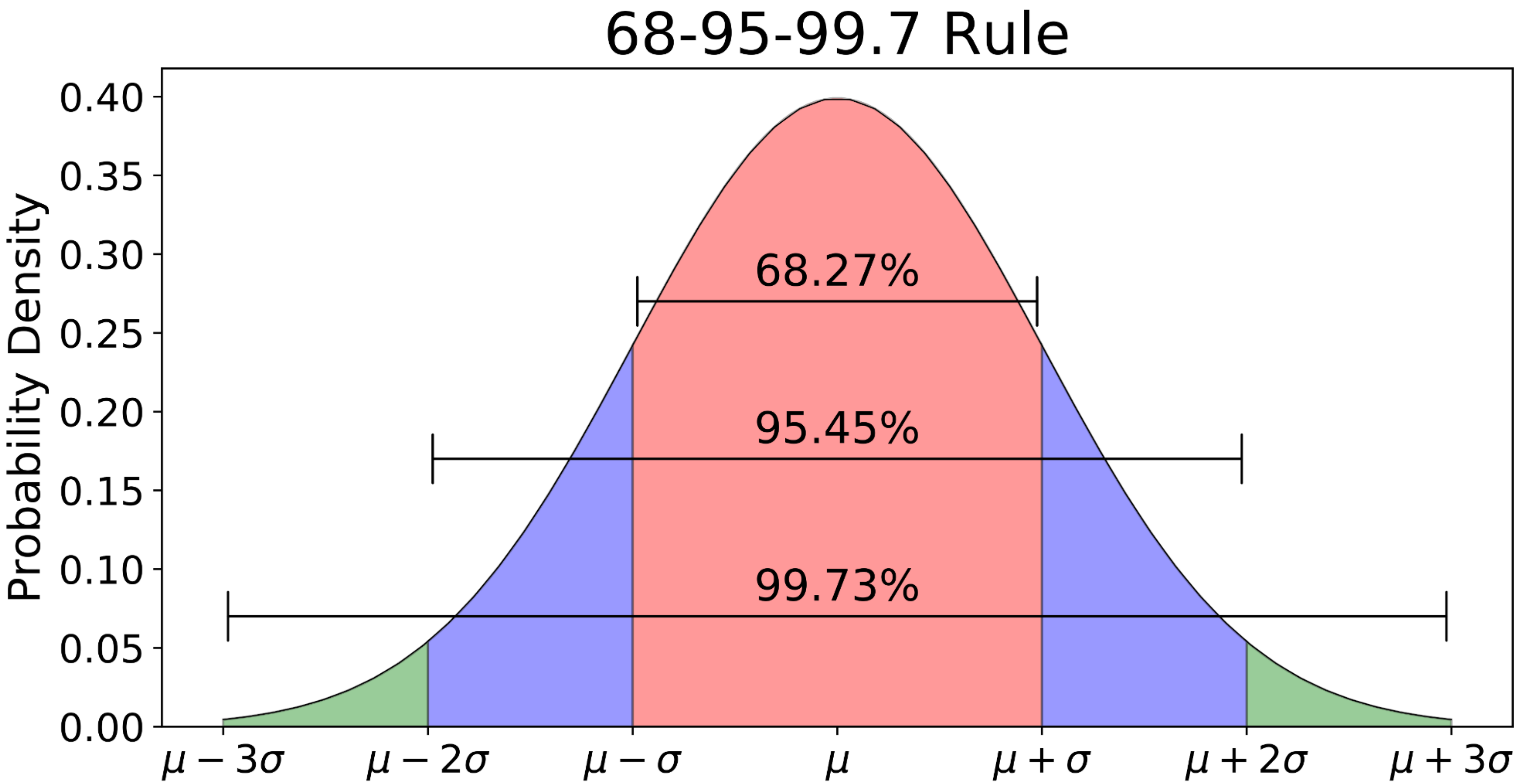
De breedte van de verdeling wordt bepaald door σ.

Er zijn twee buigpunten bij $x = \mu \pm \sigma$.



De **standaard normale verdeling** hoort bij $\mu = 0$ en $\sigma = 1$. (GR: $\text{normalcdf}(x = a, x = b)$ zonder μ, σ).

Normale verdeling H5



$$P\left(-1 \leq \underline{x} \leq 1\right) = \text{normalcdf}(x = -1, x = 1) = 0{,}682689$$

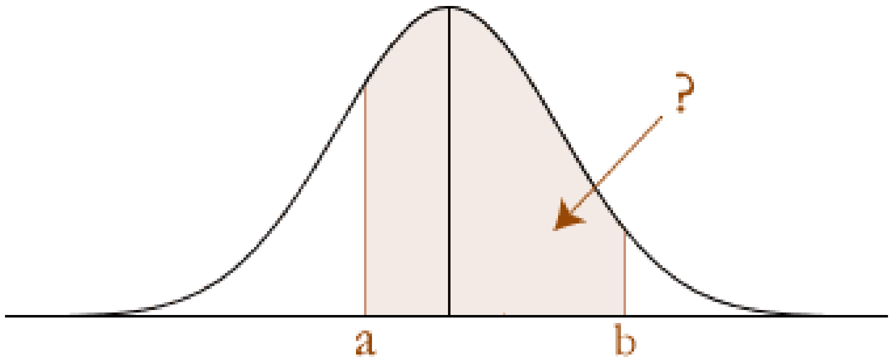
68% van alle waarden zit minder dan σ van het gemiddelde.

Vuistregel: De effectieve breedte van de normale verdeling is ongeveer 6σ (je mist dan maar 0,3%).

Vuistregel: Voor meetwaarden uit de normale verdeling geldt $\boldsymbol{\sigma \approx (\textbf{grootste} - \textbf{kleinste})/6}$.

Normale verdeling H5

$$P(a \leq \underline{x} \leq b) = \text{normalcdf}(x = a, x = b, \mu, \sigma)$$



De waarden van een normaal verdeelde variabele liggen altijd tussen $-\infty$ en ∞.
Bij berekeningen met de GR kun je ∞ vervangen door 10^{10} (of 10^{99} als je het niet vertrouwt).
Zo is $\text{normalcdf}(-10^{10}, 10^{10}) = 1$.

Let op: Gebruik bij -10^{10} het minteken (-) op de rekenmachine, niet de normale knop – voor aftrekken!

TI-84 emulator: https://mn.testnav.com/client/index.html#
Voorbeelden:

$\mu = 3, \sigma = 1{,}5,$ $P(-1 \leq \underline{x} \leq 4) = \text{normalcdf}(x_l = -1, x_u = 4, \mu = 3, \sigma = 1.5) = 0{,}7437$

$\mu = 3, \sigma = 1{,}5,$ $P(\underline{x} \leq 4) = \text{normalcdf}(x_l = -10^{10}, x_u = 4, \mu = 3, \sigma = 1.5) = 0{,}7475$

$\mu = 3, \sigma = 1{,}5,$ $P(\underline{x} > 3{,}5) = \text{normalcdf}(x_l = 3.5, x_r = 10^{10}, \mu = 3, \sigma = 1.5) = 0{,}3694$

$\mu = 12{,}7\ , \sigma = -1{,}3,$ $P(12 \leq \underline{x} \leq 14) = \text{normalcdf}(x_l = 12, x_u = 14, \mu = 12.7, \sigma = -1.3) = ????$

Normale verdeling H5

Wat is het kleinste getal x waarvoor $\text{normalcdf}(x_l = -x, x_u = x\ ) = 1$?

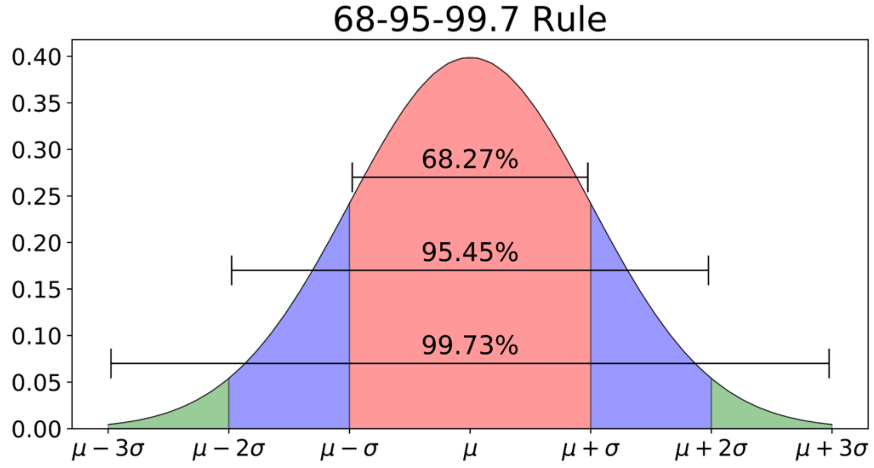
Solve: $\text{normalcdf}(x_l = -x, x_u = x\ ) = 1$

Oplossing met de GR: $x = 13{,}4217728$

Waarom niet $x = \infty$??

Omdat $\text{normalcdf}(x_l = -13.4217728, x_u = 13.4217728) = 1$

Maar ook: $\text{normalcdf}(x_l = -6.58, x_u = 6.58) = 1$



Normale verdeling H5 Voorbeeld

Voorbeeld:

De vraag naar een product is normaal verdeeld.
De gemiddelde vraag bedraagt 100 stuks per week.
De standaarddeviatie is gelijk aan 10 stuks per week. Er liggen 112 stuks op voorraad. Hoe groot is de kans dat deze voorraad voor één week genoeg is (dit is een voorbeeld van de **servicegraad**)?

Uitwerking:

$\underline{x}$ is de vraag in een specifieke week

$$P(\text{genoeg voorraad}) = P(\underline{x} \leq 112) = \text{normalcdf}(-10^{10}, 112, 100, 10) = 0{,}8849$$

Die kans is dus 0,8849, best groot, maar het is niet onmogelijk dat er niet geleverd kan worden. Hoe erg is dat?
Hoe groot moet de weekvoorraad zijn voor een servicegraad van 99%

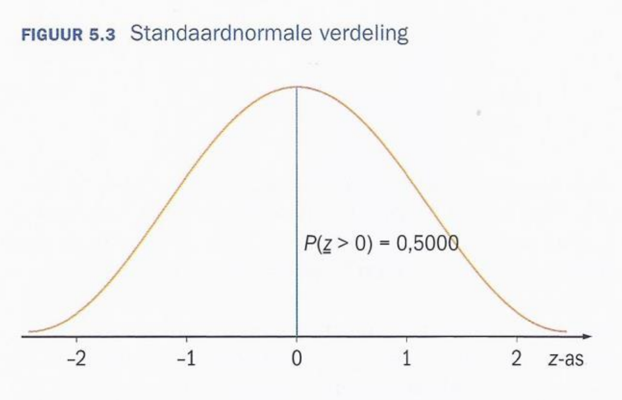
$$\text{normalcdf}(-10^{10}, v?, 100, 10) = 0{,}99$$

Los op met MATH -> Solver. Vul in $0 = \text{normalcdf}(-10^{10}, v, 100, 10) - 0{,}99$
Startwaarde: $v = 100$. Start Solve (knopje alfanumeriek rechtsonder op TI84).
Oplossing: $v = 123{,}26$. Voor alle zekerheid naar boven afronden: $v = 124$, het gaat over gehele stuks.

Standaard normale verdeling en de z-waarde

Een speciaal geval van de normale verdeling is de **standaard normale verdeling**. Hierbij is $\mu = 0$ en $\sigma = 1$.

Notatie: $\underline{x} \sim N(0{,}1)$



Kansen voor de normale verdeling kun je uitrekenen als

$$P\left(a \leq \underline{x} \leq b\right) = \text{normalcdf}(x = a, x = b)$$

Je vult dan de waarden van μ en σ niet in.

Je kunt een berekening voor een algemene $\underline{x} \sim N(\mu, \sigma)$ altijd transformeren naar een standaard normale verdeling met de transformatie

$$\underline{z} = \frac{\underline{x} - \mu}{\sigma}$$

Deze ***z*-waarde** is dan standaard normaal verdeeld $\underline{z} \sim N(0; 1)$

Normale verdeling Voorbeeld 5.2

Voorbeeld 5.2 (Boek Buijs): De wekelijkse afzet van Euro-loodvrij benzine van een tankstation wordt gegeven door een normaal verdeelde kansvariabele $\underline{x}$ met gemiddelde $\mu = 7000$ liter en standaarddeviatie $\sigma = 800$ l.

Bereken de kans dat de afzet in een bepaalde week minder is dan 6400 liter.

Uitwerking: De gevraagde kans is

$$P(\underline{x} < 6400) = \text{normalcdf}(-10^{10}\,, 6400\,, 7000\,, 800) = 0{,}2266$$

Maar klopt dit wel, want de afzet kan niet negatief zijn, dus waarom is het niet

$$P(\mathbf{0} \leq \underline{x} < 6400) = \text{normalcdf}(0\,, 6400\,, 7000\,, 800)?$$

De kans op een negatieve afzet is volgens de normale verdeling

$$P(\underline{x} < 0) = \text{normalcdf}(-10^{10}\,, 0\,, 7000\,, 800) = 1{,}0667 \cdot 10^{-18}$$

Dit is praktisch gelijk aan 0. Het maakt dus niet uit als je dit verwaarloost.

Normale verdeling Voorbeeld 5.3 en InvNorm

Voorbeeld 5.3 (Boek Buijs): Bij een garagebedrijf blijkt de tijdsduur van een reparatie te worden beschreven door een normaal verdeelde kansvariabele $\underline{x}$ met een verwachtingswaarde van 15 minuten en een standaarddeviatie van 2 minuten.

De chef wil bepalen hoeveel tijd hij moet plannen (g minuten), zodat de kans dat dit te krap is ≤10% is.

Uitwerking: We moeten g zo bepalen dat

$$P(\underline{x} > g) = 0{,}10$$

ofwel

$$P(\underline{x} \leq g) = 0{,}90$$

Dit kun je oplossen met het commando InvNorm. Hierbij is de **linkeroverschrijdingskans** $P(\underline{x} \leq g) = 0{,}90$ gegeven en wordt de bovengrens uitgerekend.

$$g = \text{InvNorm}(p = 0.90\,, \mu = 15\,, \sigma = 2) = 17{,}5631 \text{ minuten}$$

Normale verdeling Mogelijke berekeningen

Met de commando's normalcdf en invNorm kun je verschillende berekeningen uitvoeren aan normale verdelingen. In het algemeen is

$$P\left(a \leq \underline{x} \leq b\right) = \text{normalcdf}(x = a, x = b, \mu, \sigma) = p$$

Hierin staan vijf parameters: a, b, μ, σ, p. Elk van deze vijf kan in principe onbekend zijn, bij bekende anderen.
Een paar situaties kun je met normalcdf en invNorm oplossen:

1. p is onbekend en a, b, μ, σ bekend: Standaardberekening $p = \text{ normalcdf}(x = a, x = b, \mu, \sigma)$
2. a is onbekend en p, b, μ, σ bekend: $P\left(a \leq \underline{x} \leq b\right) = p$, dus $P\left(a \leq \underline{x}\right) = p + P\left(b \leq \underline{x}\right)$, dus

   $$P\left(\underline{x} \leq a\right) = 1 - \left(p + \text{ normalcdf}(x = b, x = \infty, \mu, \sigma)\right)\text{, dus}$$

   $$a = \text{invNorm}(1 - p - \text{normalcdf}(x = b, x = \infty, \mu, \sigma), \mu, \sigma).$$

3. b is onbekend en p, a, μ, σ bekend: $P\left(a \leq \underline{x} \leq b\right) = p$, dus $P\left(\underline{x} \leq b\right) = p + P\left(\underline{x} \leq a\right)$, dus

   $$P\left(\underline{x} \leq b\right) = p + \text{ normalcdf}(x = -\infty, x = a, \mu, \sigma)\text{, dus}$$

   $$b = \text{invNorm}(p + \text{ normalcdf}(x = -\infty, x = a, \mu, \sigma), \mu, \sigma).$$

Als één van a, b, μ, σ onbekend is, is het vaak het eenvoudigst om de vergelijking

$$\text{normalcdf}(x = a, x = b, \mu, \sigma) - p = 0$$

op te lossen met de Solver (MATH) van de grafische rekenmachine.

Statistiek MBW/KW deel 1 week 4 Toets

1. $\underline{k}$ is een kansvariabele met $f(1) = 0{,}2$, $f(2) = 0{,}4$, $f(3) = 0{,}3$, $f(4) = 0{,}1$.

Dan is $P(\underline{k} > 3) =$

a. 0,2
b. 0,4
c. 0,3
d. 0,1
e. Ik weet het niet zeker

Statistiek MBW/KW deel 1 week 4 Toets

2. $\underline{k}$ is een kansvariabele die de waarden 1, 2, 3 en 4 kan aannemen. Er geldt $\mathrm{F}(1) = 0{,}2$, $\mathrm{F}(3) = 0{,}3$.

Dan geldt voor $\mathrm{F}(2)$:

a. $\mathrm{F}(2) = 0{,}2$
b. $\mathrm{F}(2) = 0{,}25$
c. $\mathrm{F}(2) = 0{,}3$
d. $\mathrm{F}(2)$ ligt tussen 0,2 en 0,3
e. Ik weet het niet zeker

3. $\underline{x}$ is normaal verdeeld met $\sigma = 1$. Dan geldt voor μ:

a. $-1 \leq \mu \leq 1$

b. $\mu = 0$

c. $-1 \leq \mu \leq 1$ met kans 0,68

d. μ kan alle waarden aannemen

e. Ik weet het niet zeker

Statistiek MBW/KW deel 1 week 4 Toets

4. $\underline{x} \sim N(4\,;1)$ Dan geldt $P(\underline{x} > 4) =$

a. $0{,}32$
b. $0{,}16$
c. $0{,}50$
d. $1{,}00$
e. Ik weet het niet zeker

5. $\underline{x}$ is een continue kansvariabele die uniform verdeeld is op [0,4]. Dan geldt

a. $F(0) = 0{,}5$

b. $f(0) = 0{,}5$

c. $F(2) = 0{,}5$

d. $f(2) = 0{,}5$

e. Ik weet het niet zeker

Statistiek MBW/KW deel 1 week 4 Toets

6. $\underline{x}$ is een continue kansvariabele die uniform verdeeld op [0,4]. Dan geldt

a. $E(\underline{x}) = 0{,}5$

b. $E(\underline{x}) = 2$

c. $E(\underline{x} = 2) = 0{,}5$

d. $E(2) = 0{,}5$

e. Ik weet het niet zeker

Statistiek MBW/KW deel 1 week 4 Toets

Goede antwoorden:

1d
2d
3d
4c
5c
6b