Normale verdeling. (pag. 103-110):

- De belangrijkste en meest voorkomende verdeling in de statistiek.
- De kansveranderlijke heeft een normale verdeling als de kansdichtheid de volgende vorm heeft
  $$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(x-\mu)^2}$$
  Een normale veranderlijke kan elk reëel getal als waarde aannemen.

- De normale verdeling is symmetrisch t.o.v $x = \mu$.
- Parameters:
  - $\mu \in \mathbb{R}$: de symmetriewaarde = de verwachtingswaarde
  - $\sigma > 0$: de standaardafwijking of σ^2 : de variantie

- Notatie: $X \sim N(\mu, \sigma^2)$
  Merk op dat we de variantie noteren en niet de standaardafwijking!

- De grafiek van de normale kansdichtheid is een Gauss of klok curve die dus een symmetrieas heeft bij $x = \mu$. (zie figuur 7.3 p 104)

- Cumulatieve verdelingsfunctie
  - Kan niet expliciet berekend worden!

- Standaardnormale verdeling = Normale verdeling met $\mu = 0$ en $\sigma = 1$

  - Deze wordt meestal met Z genoteerd, dus $Z \sim N(0,1)$
  - Is symmetrisch t.o.v. 0.
  - De kansdichtheid wordt genoteerd als ϕ en is: $\phi(z) = f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

  - De verdelingsfunctie is genoteerd als Φ, is gelijk aan $\Phi(z) = P(Z \le z)$ maar kan niet expliciet berekend worden.

  - De kwantiel van orde *p* is de waarde *z* waarvoor $\Phi(z) = P(Z \le z) = p$ en wordt genoteerd als $\Phi^{-1}(p)$

  - Eigenschappen:

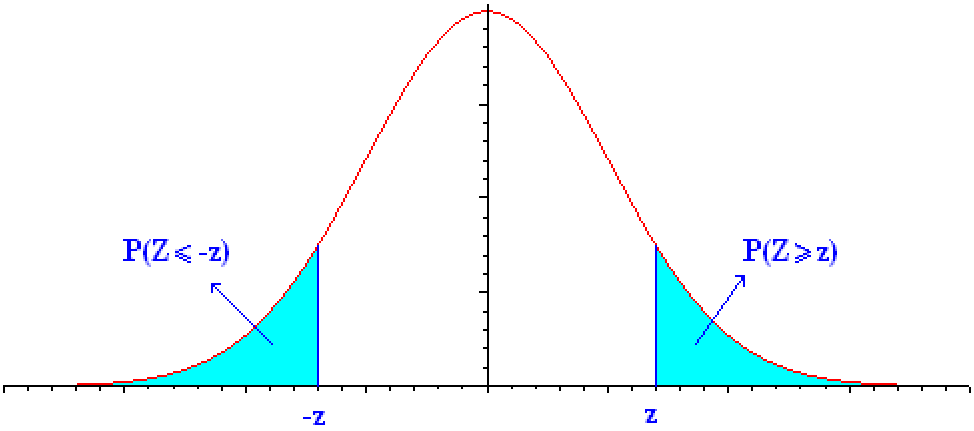
    - Cumulatieve kans voor negatieve *z*-waarden:

      Uit de symmetrie van de standaard normale verdeling, voorgesteld in onderstaande figuur volgt:

      $$P(Z \le -z) = P(Z \ge z) = 1 - P(Z < z) = 1 - P(Z \le z)$$

      of uitgedrukt aan de hand van de verdelingsfunctie:

      $$\Phi(-z) = 1 - \Phi(z)$$



- Bovendien is:

  $$P(-z \le Z \le z) = P(Z \le z) - P(Z \le -z) = P(Z \le z) - (1 - P(Z \le z)) = 2P(Z \le z) - 1$$

  of:

  $$P(-z \le Z \le z) = 2\Phi(z) - 1$$

- Om kwantielen $\Phi^{-1}(p)$ te berekenen gebruik je het commando invNorm(p) dat je vind onder 2nd DISTR

- Als $p < 0.5$ kunnen we ook gebruik van de symmetrie om het kwantiel te herleiden tot het kwantiel van 1-*p* > 0.5 als volgt:
  Als $p < 0.5$ dan is het kwantiel zeker negatief, want $\Phi(0) = P(Z \le 0) = 0.5$,
  noem het kwantiel dus –z.
  Uit: $\Phi(-z) = P(Z \le -z) = p$
  volgt met bovenstaande eigenschap:
  $1 - P(Z \le z) = p \Leftrightarrow P(Z \le z) = 1 - p$
  En dus is z het kwantiel van orde $1 - p$: $z = \Phi^{-1}(1-p)$
  Het kwantiel van orde *p* is : $\Phi^{-1}(p) = -z = -\Phi^{-1}(1-p)$

- Willekeurige normale verdeling: eigenschappen

  - Een lineaire transformatie van een normale verdeling is terug normaal verdeeld:

    Als $X \sim N(\mu, \sigma^2)$ dan is $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$

    De verwachtingswaarde is dus dezelfde lineaire transformatie van de verwachtingswaarde van X, de variantie wordt met het kwadraat van de coëfficiënt van X vermenigvuldigd.

    Uitleg bij het bewijs voor $a > 0$:
    De verdeling van Y kunnen we vinden door de cumulatieve verdeling te bepalen of de distributiefunctie. Voor de cumulatieve verdeling vinden we:

    $$F_Y(y) = P(Y \le y) = P(aX + b \le y) = P(X \le \frac{y-b}{a}) = F_X(\frac{y-b}{a})$$

Omdat we geen expliciete uitdrukking hebben voor de verdelingsfunctie moeten we overgaan op de distributiefunctie. Dit is de afgeleide van de verdelingsfunctie. (Fris nog een op)

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(\frac{y-b}{a}) = F_X^{'}(\frac{y-b}{a}) \frac{d}{dy}(\frac{y-b}{a})$$

$$= f_X(\frac{y-b}{a}) \cdot \frac{1}{a}$$

$$= \frac{1}{a} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\frac{y-b}{a}-\mu}{\sigma}\right)^2}$$

$$= \frac{1}{\sqrt{2\pi} a\sigma} e^{-\frac{1}{2}\left(\frac{y-a\mu-b}{a\sigma}\right)^2}$$

En dit heeft de vorm van de distributiefunctie van een normaal verdeelde veranderlijke met verwachtingswaarde $\mu_Y = a\mu + b$, standaardafwijking $\sigma_Y = a\sigma$ en dus variantie $\sigma_Y^2 = a^2\sigma^2$.

- Standaardiseren van een normale veranderlijke.

  Als $X \sim N(\mu, \sigma^2)$ dan is $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

  Bewijs:

  De vorige eigenschap toepassen op $Z = \frac{X-\mu}{\sigma} = \frac{1}{\sigma} X - \frac{\mu}{\sigma}$ geeft:

  $$Z \sim N(\frac{\mu}{\sigma} - \frac{\mu}{\sigma}, \frac{1}{\sigma^2}\sigma^2) = N(0,1)$$

  Elke normale verdeling kan gestandaardiseerd worden:
  Als $X \sim N(\mu, \sigma^2)$ dan is:

  - $P(X \le x) = P(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}) = P(Z \le \frac{x-\mu}{\sigma})$

    of in termen van verdelingsfuncties:

    $F_X(x) = \Phi(\frac{x-\mu}{\sigma})$
  - Het kwantiel van orde p van X is de waarde x waarvoor $P(X \le x) = p$.
    Met het voorgaande wordt dit:

    $P(Z \le \frac{x-\mu}{\sigma}) = p$

    Dus is $\frac{x-\mu}{\sigma}$ het kwantiel van orde p van Z: $\frac{x-\mu}{\sigma} = \Phi^{-1}(p)$

    En dus: $x = \mu + \sigma \cdot \Phi^{-1}(p)$

- Percentage van waarden gelegen binnen een veelvoud van de standaardafwijking.

  Via de ongelijkheid van Chebyshev kunnen we een ondergrens van deze percentages bepalen. Immers volgt uit deze ongelijkheid door het compliment te nemen dat:
  $P(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2}$
  Voor respectievelijk: $k = 1, 2, 3$ geeft dit als schatting:
  $P(\mu - \sigma \leq X \leq \mu + \sigma) \geq 1 - 1 = 0$
  $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \geq 1 - \frac{1}{4} = \frac{3}{4} = 0.75$
  $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \leq 1 - \frac{1}{9} = \frac{8}{9} = 0.8889$

  Voor de normale verdeling kunnen we deze kansen exact bepalen, omdat ze niet afhangen van $\mu \; en \; \sigma$ zoals blijkt als we standaardiseren.
  $P(\mu - k\sigma \leq X \leq \mu + k\sigma) = P(\frac{\mu - k\sigma - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{\mu + k\sigma - \mu}{\sigma}) = P(-k \leq Z \leq k)$
  Dus:

  $$P(\mu - \sigma \leq \mathrm{X} \leq \mu + \sigma) = \mathrm{P}(-1 \leq \mathrm{Z} \leq 1) = \mathrm{normalcdf}(-1,1) = 68.27\%$$

  $$P(\mu - 2\sigma \leq \mathrm{X} \leq \mu + 2\sigma) = \mathrm{P}(-2 \leq \mathrm{Z} \leq 2) = \mathrm{normalcdf}(-2,2) = 95.45\%$$

  $$P(\mu - 3\sigma \leq \mathrm{X} \leq \mu + 3\sigma) = \mathrm{P}(-3 \leq \mathrm{Z} \leq 3) = \mathrm{normalcdf}(-3,3) = 99.73\%$$

  Minder dan 0.3 % van de waarden van een normale verdeling ligt dus verder dan 3 keer de standaardafwijking verwijderd van de verwachtingswaarde.