Data Advanced

Populatie – steekproef
Dichtheidsfunctie
Normale verdeling

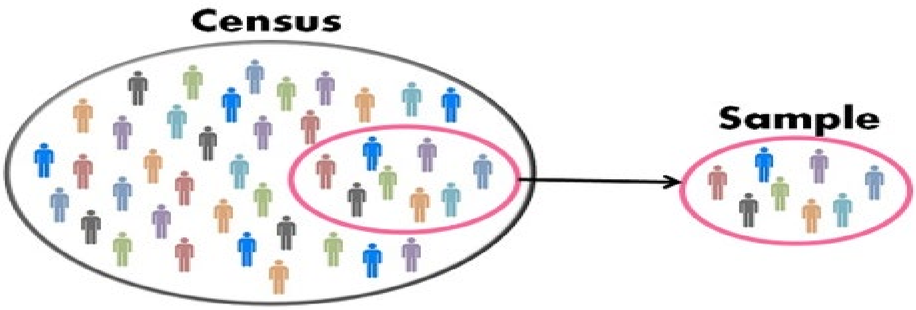
-IT

Lector:
Heidi Tans

1. Populatie – steekproef

Veel onderzoek is gebaseerd op steekproeven: op basis van gegevens van een kleine groep observaties worden conclusies getrokken die gelden voor de hele bevolking (= populatie). De populatie is het grotere geheel waarvan we bepaalde karakteristieken te weten willen komen, maar die we om allerlei redenen (van praktische, financiële, … aard) niet volledig kunnen observeren.

Die kleine groep moet dan wel goed gekozen worden, zodat de groep representatief is voor de populatie waarover uitspraken worden gedaan.



Van deze steekproeven bekijken we enkele belangrijke grootheden: het steekproefgemiddelde (schatter voor het populatiegemiddelde), de steekproefvariantie (schatter voor de populatievariantie) en de steekproefproportie (schatter voor de populatieproportie).

Voorbeeld:

Een sigarettenproducent wil de gemiddelde hoeveelheid teer meten in een nieuw merk sigaretten. Hij doet hiervoor een steekproef van 100 sigaretten, laat de hoeveelheid teer meten mg; en berekent het gemiddelde. Dit leidt tot een steekproefgemiddelde van 14.8mg; wat een schatting geeft van het gemiddelde van de hele populatie (= alle sigaretten van dit merk). Het is duidelijk dat deze schatting toevallig is, een andere steekproef zal waarschijnlijk een ander gemiddelde opleveren en dus een andere schatting.

Kunnen we zomaar de steekproefwaarden gebruiken als 'schatter' voor de populatiewaarden?

Gemiddelde

Stel:

We willen het gemiddelde maandelijks inkomen van een Belg schatten. Alle Belgen ondervragen neemt teveel tijd in beslag en zou ons onderzoek te duur maken. Indien we een steekproef nemen die representatief is voor de populatie, dan mogen we de resultaten van de steekproef veralgemenen naar de populatie.

Voorbeeld:

We nemen één steekproef van 1000 Belgen en berekenen het gemiddelde inkomen $\bar{x}$ en vinden €1350. Dit gemiddelde is één schatting voor het werkelijke gemiddelde μ (over alle Belgen heen, de populatie). Telkens we een andere steekproef van 1000 Belgen nemen, zullen we een ander gemiddeld inkomen vinden (bvb $\bar{x}_1 =$ €1350 , $\bar{x}_2 =$ €1500 , $\bar{x}_3 =$ €1220 , …). Zo verkrijgen we een rij van waarden voor $\bar{X}$ (het gemiddeld inkomen van 1000 Belgen, het steekproefgemiddelde).

Kunnen we deze steekproefwaarden zomaar gebruiken als schatter voor het ongekende populatiegemiddelde? Zijn de steekproefwaarden representatief voor de populatiewaarde?

Ja, men kan aantonen (maar dat ligt buiten het bestek van deze cursus) dat

het gemiddelde van het steekproefgemiddelde gelijk is aan het populatiegemiddelde: Gemiddelde ($\bar{X}$) = μ

De variantie van het steekproefgemiddelde gelijk is aan de populatievariantie / n:

Var ($\bar{X}$) = $\frac{\sigma^2}{n}$

Nemen we nu een aantal steekproeven van grootte 100 (ipv 1000) en berekenen we de gemiddelde inkomens van deze steekproeven, dan zit er veel meer variatie op deze gemiddeldes. Eén enkele schatting van het werkelijk gemiddelde inkomen (populatie gemiddelde μ) aan de hand van een steekproef van 100 personen, kan veel verder verwijderd zijn van het werkelijk gemiddeld inkomen dan wanneer we zouden schatten op basis van een steekproef van 1000 personen. Naarmate de steekproef groter wordt, zal $\overline{x}$ een betere schatter zijn voor μ.

Proportie

Stel:
We willen de proportie rokers in België schatten. We stellen hierbij een ja – neen vraag en delen de ondervraagden in twee groepen in. Alle Belgen ondervragen neemt teveel tijd in beslag en zou ons onderzoek te duur maken. Indien we een steekproef nemen die representatief is voor de populatie, mogen we de resultaten van de steekproef mogen veralgemenen naar de populatie.

Voorbeeld:
We nemen één steekproef van 1000 Belgen en bepalen de proportie rokers $\overline{p}$ en vinden 0.18. Deze proportie is één schatting voor de werkelijke proportie π (over alle Belgen heen, de populatie). Telkens we een andere steekproef van 1000 Belgen nemen, zullen we een andere proportie rokers vinden (bvb $\overline{p}_1 = 0.21$, $\overline{p}_2 = 0.17$, $\overline{p}_3 = 0.2$, …). Zo verkrijgen we een rij van waarden voor $\overline{P}$ (de proportie Belgische rokers, de steekproefproportie).

Kunnen we deze steekproefwaarden zomaar gebruiken als schatter voor het ongekende populatiegemiddelde? Zijn de steekproefwaarden representatief voor de populatiewaarde?

Ja, men kan aantonen (maar dat ligt buiten het bestek van deze cursus) dat

Het gemiddelde van de steekproefproportie gelijk is aan de populatieproportie: Gemiddelde $(\overline{P})$ = π

De variantie van de steekproefproportie gelijk is: Var $(\overline{P})$ = $\frac{\pi(1-\pi)}{n}$

Merk op:

Voor elke proportie (zowel steekproef als populatie) geldt: $0 \leq \overline{p} \leq 1 \quad 0 \leq \pi \leq 1$

	Steekproef	Populatie
Grootte	n	N
Gemiddelde	$\overline{X}\ (\overline{x})$	μ
Variantie	$S^2\ (s^2)$	σ^2
Standaard afwijking	$S\ (s)$	σ
Proportie	$\overline{P}\ (\overline{p})$	π

2. De dichtheidsfunctie

In datarepresentatie hebben we datasets beschreven aan de hand van

kengetallen voor locatie - kengetallen voor spreiding - grafische methodes: frequentietabel, staafdiagram, boxplot, …

Bij grotere datasets (de populatie in het achterhoofd houdend) krijgen we niet altijd een goed beeld mbv bovenstaande methoden:

frequentietabellen worden (te) lang waardoor details onderdrukt worden of afhankelijk worden van de klasseindeling. Staafdiagrammen / histogrammen worden te uitgebreid…

De vraag die we ons kunnen stellen is of er een manier bestaat die toelaat aan de hand van één uitdrukking de volledige dataset (ook wel verdeling genoemd) te beschrijven. Deze "ideale" beschrijving (uitschieters en kleine onregelmatigheden buiten beschouwing gelaten) noemen we de **dichtheidsfunctie**.
De dichtheidsfunctie volgt het algemeen patroon van de gegevens maar verwaarloost onregelmatigheden en uitschieters.

Voorbeeld

Volgend histogram geeft het aantal mensen (in 10-tallen) op een observatie van 500 mensen dat bereid is gedurende een bepaalde tijd te wachten wanneer je naar een call-center telefoneert.

We zien dat deze gegevens een regelmatig patroon vertonen.

Het histogram is min of meer symmetrisch: de top ligt ongeveer in het midden en beide staarten gaan geleidelijk naar 0.

Op basis van dit histogram kunnen we een antwoord geven op volgende vragen:

- Hoeveel % van de mensen wil ten hoogste 6 minuten wachten?

- Hoeveel % van de mensen wacht tussen 4 en 6 minuten?

De vloeiende curve (**dichtheidsfunctie**) die doorheen dit histogram getekend is, is een goede beschrijving van het algemeen patroon van deze gegevens.

Nu blijkt dat wanneer we de oppervlakte onder dichtheidsfunctie gaan bereken, we ongeveer dezelfde resultaten uitkomen.

3. Soorten dichtheidsfuncties en verdelingen

Verdelingen en dus ook dichtheidsfuncties kunnen in verschillende gedaanten voorkomen. De totale oppervlakte onder de dichtheidsfunctie is steeds gelijk aan 100% of 1.

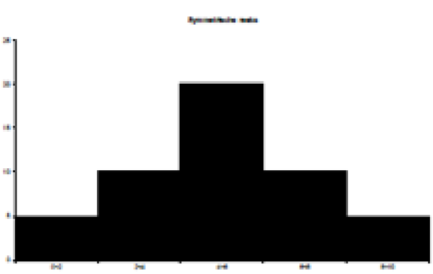
Voorbeeld

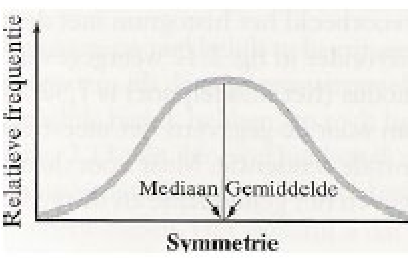
We houden de score bij op 3 verschillende OLOD's voor een steekproef van 50 studenten. Dit geeft volgende frequentietabellen met klassenindeling met bijhorende histogrammen.

Symmetrische verdelingen en dichtheidsfuncties

- Deze gedragen zich op dezelfde wijze aan de linkse en rechtste zijde van de figuur
- Mediaan en gemiddelde zijn (ongeveer) gelijk
- De linkertak en rechtertak vormen elkaars spiegelbeeld

Klassen	m_i	f_i
0<2	1	5
2<4	3	10
4<6	5	20
6<8	7	10
8<10	9	5
		50

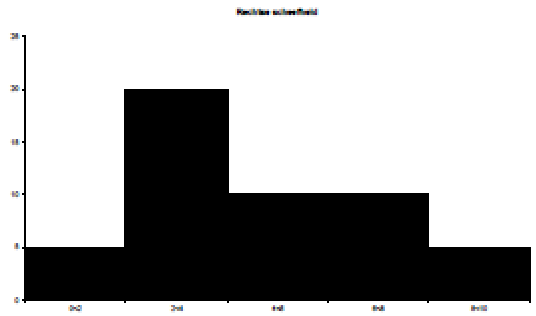


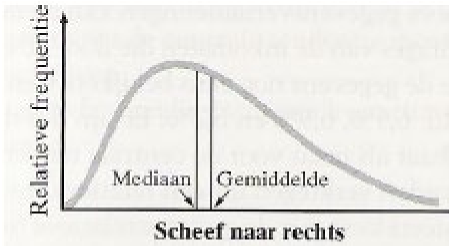


Rechts-scheve verdelingen en dichtheidsfuncties

- Het gemiddelde is groter dan de mediaan
- De dichtheidsfunctie en het histogram hebben een ‘rechterstaart’

Klassen	m_i	f_i
0<2	1	5
2<4	3	20
4<6	5	10
6<8	7	10
8<10	9	5
		50

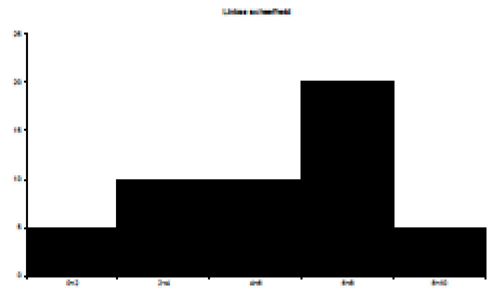


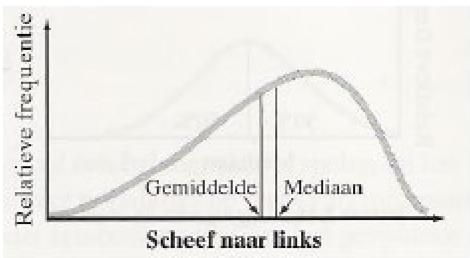


Links-scheve verdelingen en dichtheidsfuncties

- Het gemiddelde is kleiner dan de mediaan
- De dichtheidsfunctie en het histogram hebben een ‘linkerstaart’

Klassen	m_i	f_i
0<2	1	5
2<4	3	10
4<6	5	10
6<8	7	20
8<10	9	5
		50





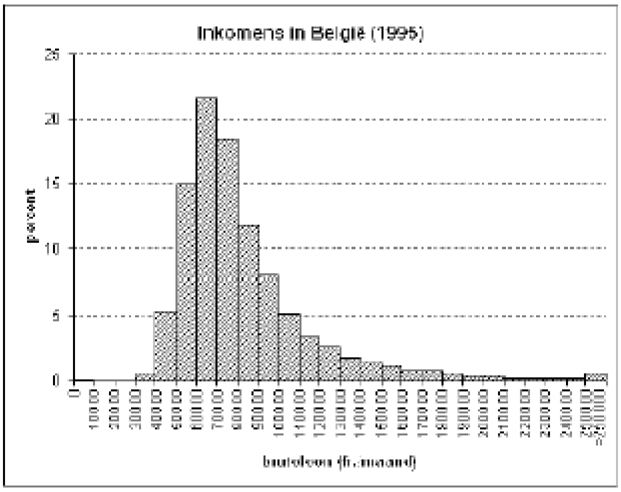
Besluit

- We hebben te maken met een symmetrische verdeling met bijhorende symmetrische dichtheidsfunctie als $\overline{x} \approx$ mediaan $\approx$ modus. Het bijhorende histogram en de dichtheidsfunctie zijn zo goed als symmetrisch.
- We hebben te maken met een links-scheve verdeling met bijhorende symmetrische dichtheidsfunctie als $\overline{x} \leq$ mediaan $\leq$ modus. Het bijhorende histogram en de dichtheidsfunctie vertonen een linkerstaart.
- We hebben te maken met een rechts-scheve verdeling met bijhorende symmetrische dichtheidsfunctie als $\overline{x} \geq$ mediaan $\geq$ modus. Het bijhorende histogram en de dichtheidsfunctie vertonen een rechterstaart.

Om te beoordelen of een observatie symmetrisch, linksscheef of rechtsscheef is, kan je best steeds het bijhorende histogram maken.

Voorbeeld

Histogram van de Belgische gezinsinkomens in het jaar 1995 (n = 474)



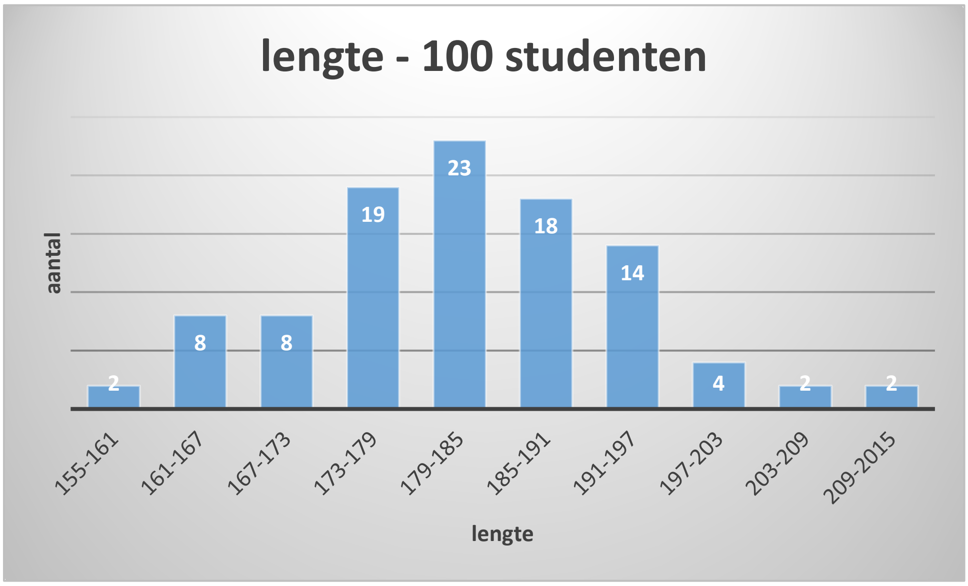
- Welke centrummaat beklemtoont dat een groot aantal gezinnen een laag inkomen hebben? Modus
- Welke maat drukt het best uit hoeveel de doorsnee Belg verdient? Mediaan
- Welke maat houdt rekening met elke inkomen (zowel van de armen als van de rijken)? Gemiddelde
- Is bovenstaande een symmetrische verdeling? nee, rechstscheef

4. De normale verdeling

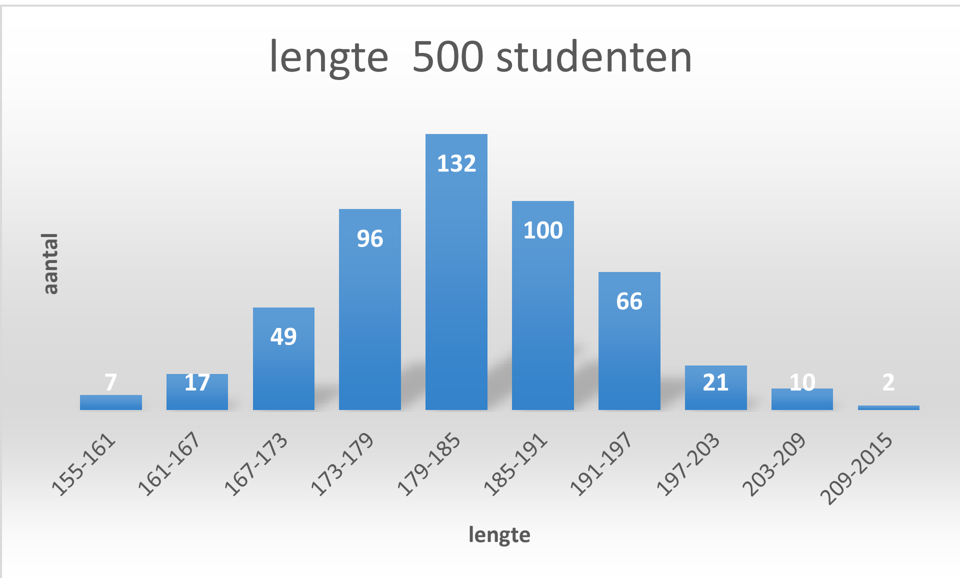
Een veel voorkomende dichtheidsfunctie is deze van de normale verdeling omdat veel gegevens normaal verdeeld zijn. In de praktijk wordt de normale verdeling veel toegepast als het gaat om variabelen als lengte, gewicht, IQ – scores, …

Voorbeeld

- We hebben van 100 studenten de lichaamslengte uitgezet in een staafdiagram

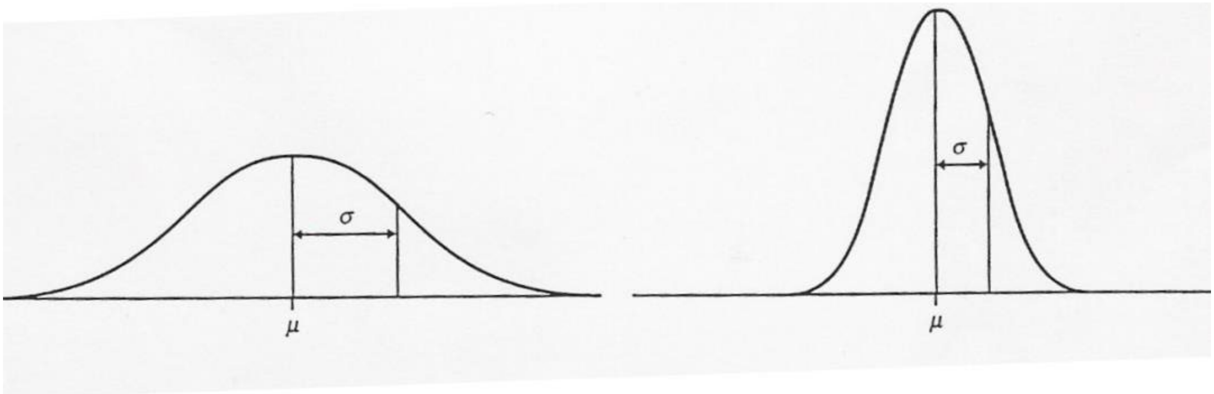


- We hebben van 500 studenten de lichaamslengte uitgezet in een staafdiagram

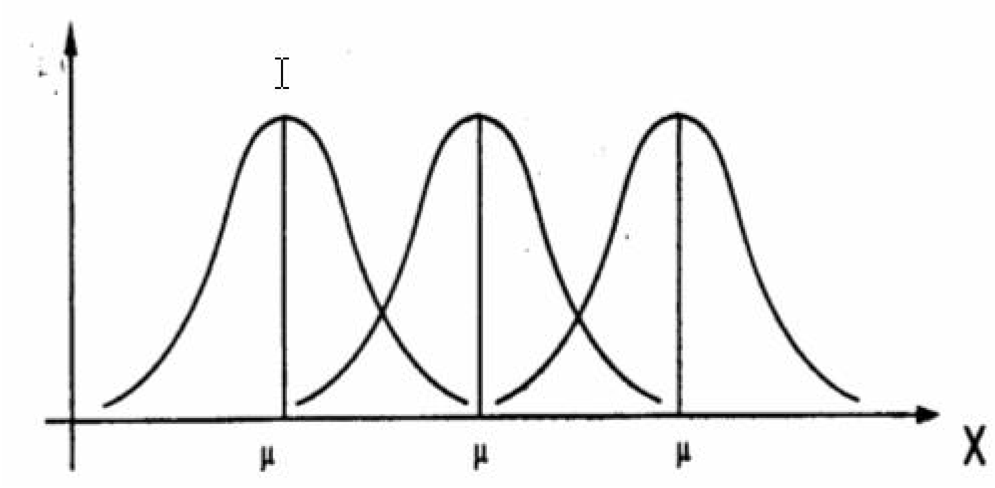


Grafiek

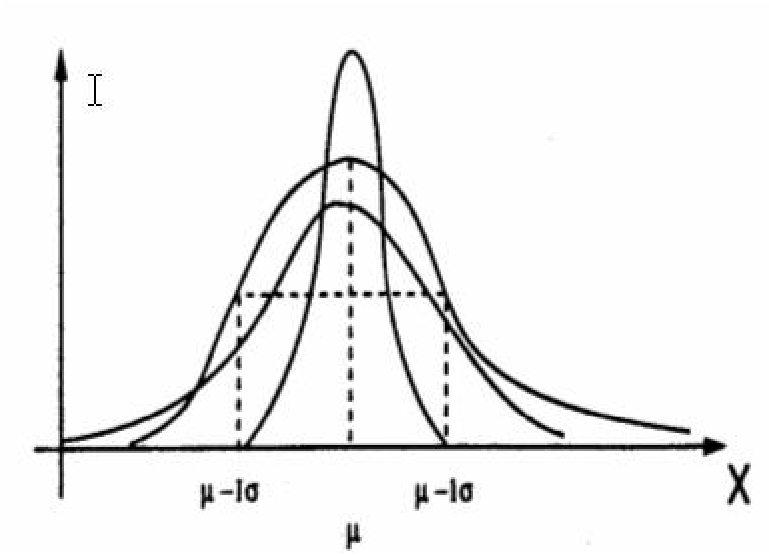
De dichtheidsfunctie van een normaal verdeelde stochastische variabele X ~ N (μ, σ^2) heeft een typische klokvorm (en wordt Gausscurve genoemd). Ze is symmetrisch rond het gemiddelde μ. Bovendien varieert de breedte en de hoogte van de klok afhankelijk van de waarde van de standaard afwijking σ.



Indien we μ zouden wijzigen zonder σ te veranderen, dan verkrijgen we een horizontale verschuiving:



De standaardafwijking σ bepaalt de breedte van de curve. Zoveel te groter σ, zoveel te breder en vlakker de curve, wat een grotere spreiding van de gegevens aangeeft.



- De totale oppervlakte onder de curve is gelijk aan 1 of 100%

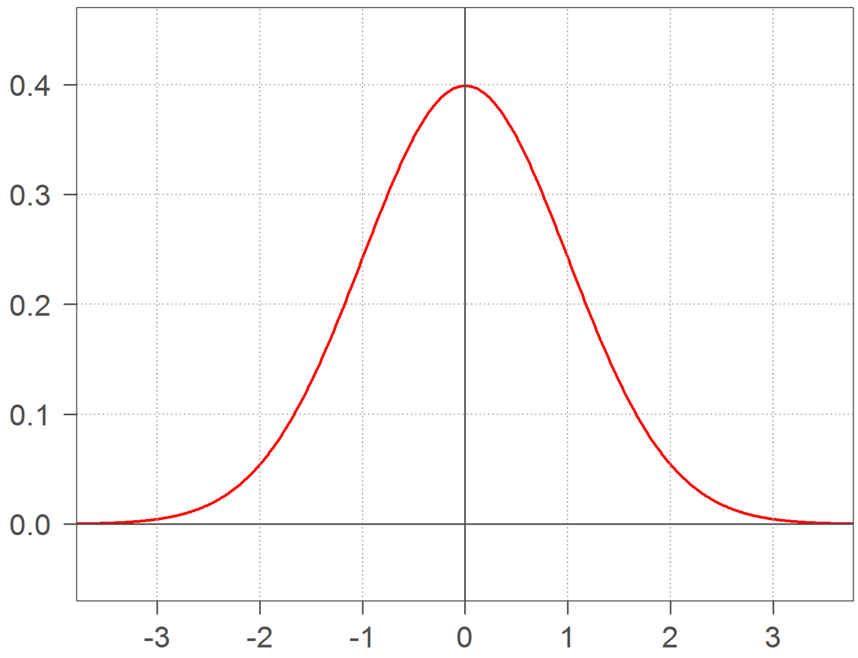
- Bij een normaal verdeelde stochastische variabele geldt het volgende: 68 % van de gegevens bevindt zich in het interval $[\mu - \sigma, \mu + \sigma]$

  95 % van de gegevens bevindt zich in het interval $[\mu - 2\sigma, \mu + 2\sigma]$

  99.7% van de gegevens bevindt zich in het interval $[\mu - 3\sigma, \mu + 3\sigma]$

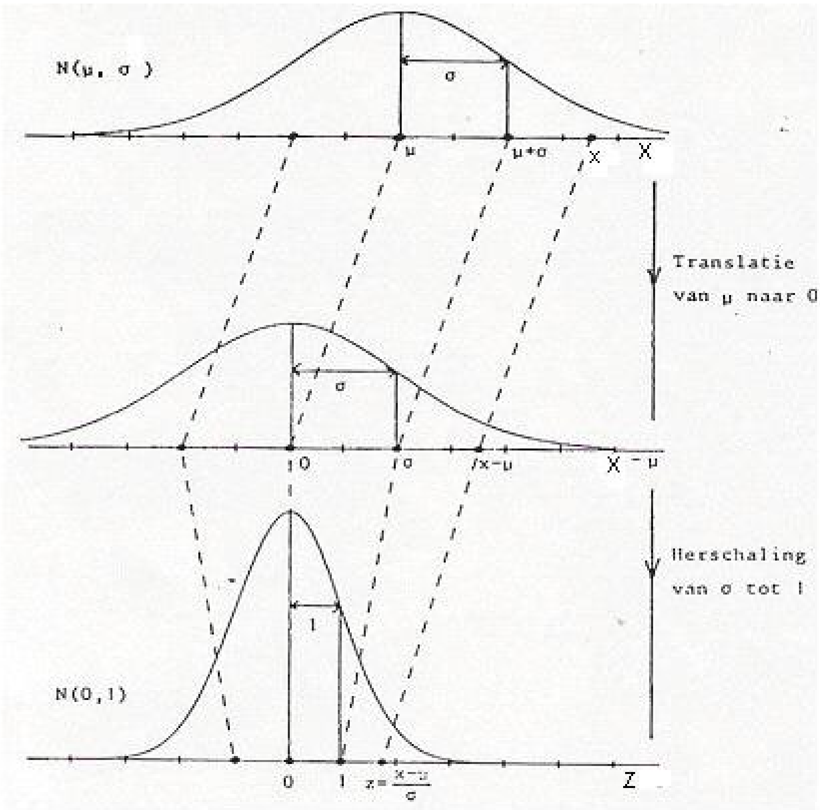
Definitie: De standaard normale verdeling

Een normale verdeling met $\mu = 0$ en $\sigma^2 = 1$ noemt men de standaard normale verdeling. Voor Z ~ N (0, 1) is de kansdichtheidsfunctie gegeven door :

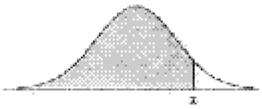


Eigenschap (omzetten van een normaal verdeelde stochastische variabele naar een standaard normaal verdeelde stochastische variabele)

Als X ~ N (μ , σ^2) dan $Z = \dfrac{X - \mu}{\sigma}$ ~ *N (0, 1)*



Tables of the Normal Distribution



Probability Content from -oo to Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Het berekenen van een bepaalde oppervlakte onder de grafiek van een normale verdeling, gebeurt via integralen. Voor de standaard normale verdeling zijn deze oppervlaktes (kansen) getabelleerd voor welbepaalde waarden van x. De tabel die wij hanteren, geeft $P(Z \leq x)$.

Voorbeeld

$P(Z \leq 1{,}15) = 0{,}8749$

Voorbeelden

1. Z ~ N (0, 1), bereken:

$P(Z \leq 2.42) =$ 0.99922

$P(Z > 2.09) =$ 1 - 0.9817 = 0.0183

$P(Z < -1.23) =$ 1 - 0.8907 = 0.1043

$P(1.21 < Z < 2.85) =$ 0.9978 - 0.8869 = 0,1109

2. Z ~ N (0, 1) Bereken a als:

$P(Z < a) = 0.9936$ 2.49

$P(Z > a) = 0.9887$ -2.28

$P(Z \leq a) = 0.0281$ 1-0.0281 = 0.9719 --> -1,91

$P(Z > -a) = 0.8708$ 1.13

3. X ~ N (20, 4)

$P(X < 24) =$ p((x-2)/2 < (24-20)/2) = p(7<2) = 0.97772

$P(X > 18) =$ 1 - P(x<18)=
1-normcof(18,20,2)
=0.841...

4. Zij X de lichaamslengte van een PXL-student. De verdeling kan benaderd worden door een normale verdeling met gemiddelde 170.6cm en standaardafwijking 6.75cm.

a. Hoeveel % van de PXL-studenten hebben een lichaamslengte kleiner dan 180 cm? P(X<180)= nomr cof(180,170.6,6.75) = 0.918

b. Hoeveel % van de PXL-studenten hebben een lichaamslengte tussen 160 en 175 cm?

P(160<x<175) = normcof(175,170.6,6.75) - norm. cof(160,170.6,6.75) = 0.684

c. Hoeveel % van de PXL-studenten heeft een lichaamslengte gelijk aan 180cm?

P(x=180) = 0
P(179.5<180.5) = 0.026

d. 60% van de PXL-studenten heeft een lichaamslengte kleiner dan of gelijk aan ……………cm.

P(X<a) = 0.6
a= norm ppf(0.6,170.6,6.75)
=172.32....

5. De punten van een leestest zijn $N(430; 100)$ verdeeld. Hoe hoog moet een leerling scoren om bij de 10% beste leerlingen te behoren?

P(X>a) = 0.1
P(x<a)=0.9
norm.ppf(0.9,430,10)= 442.8

4.1 Verdeling steekproef – gemiddelde / proportie

Er kan aangetoond worden (maar ook dit ligt buiten het bestek van deze cursus) dat het steekproefgemiddelde $\bar{X}$ een normale verdeling heeft van zodra de onderliggende populatie X normaal verdeeld is voor kleine steekproeven (n < 30). Voor grote steekproeven is het steekproefgemiddelde $\bar{X}$ sowieso normaal verdeeld.

	n $\geq$ 30	n < 30 X normaal verdeeld
σ^2**gekend**	$\bar{X} \sim \mathrm{N}(\mu, \frac{\sigma^2}{n}) \rightarrow$ $\frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$	$\bar{X} \sim \mathrm{N}(\mu, \frac{\sigma^2}{n}) \rightarrow$ $\frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$
σ^2 **niet gekend**	$\bar{X} \sim \mathrm{N}(\mu, \frac{S^2}{n}) \rightarrow$ $\frac{\bar{X}-\mu}{\sqrt{\frac{S^2}{n}}} \sim N(0,1)$	~~Behandelen we niet in deze cursus~~

Ook de steekproefproportie $\overline{P}$ is normaal verdeeld, op voorwaarde dat voldaan is aan

n$\pi \geq$ 5 en n(1-π) $\geq$ 5
$\overline{P} \sim \mathrm{N}\ (\pi\ ;\ \pi\ (1\text{-}\pi)/n\ ) \rightarrow \frac{\overline{P}-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \sim \mathrm{N}\ (0\ ;\ 1\ )$

:

Voorbeeld 1

Stel dat we een steekproef nemen met n=35 uit een populatie met een gemiddelde gelijk aan 80 en een standaardafwijking gelijk aan 5. De verdeling van de populatie is niet gekend.

a) Wat kan je zeggen over de verdeling van het steekproefgemiddelde?

N(80,5^2/35)

b) Bepaal de kans dat het steekproefgemiddelde tussen de 79 en 82 ligt.

P(79<x<82)
= norm.coef(82,80,sqrt(5^2/35) = 0.8726

Voorbeeld 2

Onderzoek heeft uitgewezen dat een bibliotheek gemiddeld 320 boeken per dag uitleent, met een standaardafwijking van 75 boeken. Neem een steekproef van 100 dagen. Hoe groot is de kans dat het steekproefgemiddelde tussen de 300 en 340 ligt?

U=320
v=75
n=100

P(300<x<340) ========= norm.coef(340,320,sqrt(75^2/100) - norm.coef(300,320,sqrt(75^2/100)

nN(320,75^2/100)

Voorbeeld 3

Onderzoek heeft uitgewezen dat een bibliotheek gemiddeld 120 boeken per dag uitleent (de populatie stochastische variabele is normaal verdeeld). Voor een steekproef van 10 dagen krijgen we volgende gegevens omtrent het aantal uitgeleende boeken:

100 130 110 90 100 120 120 130 110 100

Hoe groot is de kans dat het steekproefgemiddelde tussen de 100 en 140 ligt?

P(100<X<140)
--> niet oplosbaar

Voorbeeld 4

Een klas in ‘data advanced’ heeft normaal verdeelde punten met een gemiddelde van 52 op 100 en een standaardafwijking van 9.

a) Wat is de kans dat een individuele toevallig genomen student minder dan 50 heeft? P(x<50) = norm.coef(50,52,9) = 0.41207

b) Wat is de kans dat een toevallig getrokken groepje van 10 studenten een gemiddelde van minder dan 50 heeft?

P($\overline{X}$<50)

norm.coef(50,52,sqrt(91/10)) = 0.2411

Voorbeeld 5:

Uit vroegere onderzoeken weet men dat 70% van alle tickets die verkocht worden door een vliegtuigmaatschappij retourtickets zijn. Men doet een steekproef van 100 passagiers. Wat is de kans dat minstens 75% van deze passagiers een retourticket heeft?

$P(\overline{P}>0.75)$
= 1 - norm.cdf(0.75,0.7,sqrt((0.7*0.3)/100) = 0.1376

5. PYTHON

jupyter norm waarden (autosaved)

File Edit View Insert Cell Kernel Widgets Help

```
Normale waarden berekenen in Python
```

```
Standaard normale verdeling (gemiddelde = 0; variantie = 1)
```

```
In [2]: import scipy.stats as stats
        stats.norm.ppf(0.975)

Out[2]: 1.959963984540054

In [3]: stats.norm.cdf(1.959963984540054)

Out[3]: 0.975

In [4]: stats.norm.ppf(0.975, 0, 1 )

Out[4]: 1.959963984540054

In [5]: stats.norm.cdf(1.959963984540054,0,1)

Out[5]: 0.975
```

Willekeurig normale verdeling

```
In [6]: stats.norm.cdf(24,20,2)

Out[6]: 0.9772498680518208

In [7]: stats.norm.ppf(0.9772498680518208,20,2)

Out[7]: 24.0
```

6. Oefeningen

Oefening 1

Veronderstel dat de IQ - scores van studenten normaal verdeeld zijn met gemiddelde 110 en standaardafwijking 25. Van hoeveel van de 200 studenten verwacht je dat ze een IQ - score van meer dan 140 hebben?

23.014

Oefening 2

Van een bepaald type oplaadbare batterijen is de gemiddelde levensduur 800 uur, met een standaardafwijking van 38 uur. Er wordt een steekproef van 80 batterijen genomen. Wat is de kans dat de batterijen in deze steekproef gemiddeld meer dan 815 uur meegaan?

Oefening 3

Een populatie heeft een gemiddelde van 100 en een standaardafwijking van 16. Hoe groot is de kans dat een steekproefgemiddelde minder dan 2 afwijkt van het populatiegemiddelde als n = 50

0.6232

Oefening 4

30% van de orders van een distributiebedrijf is afkomstig van nieuwe klanten. Er wordt een willekeurige steekproef van 100 orders gedaan.

(a) Hoe groot is de kans dat de steekproefproportie tussen 0,2 en 0,4 ligt?

(b) Hoe groot is de kans dat de steekproefproportie minder dan 0,05 afwijkt van de populatieproportie?

Oefening 5

Een cardioloog beweert dat 20% van de mensen die een hartaanval krijgen, binnen het jaar een tweede hartaanval krijgt. Om dit te checken doet men een steekproef bij 275 mensen die het afgelopen jaar een hartaanval gehad hebben. Neem aan dat de bewering van de cardioloog correct is. Hoe groot is de kans dat minder dan 17% van hen binnen het jaar een tweede hartaanval krijgt?

0.10679

Oefening 6

De tijd die nodig is om een examen van een bepaalde cursus af te leggen is normaal verdeeld met een gemiddelde van 80 minuten en een variantie van 100.

a. Hoe groot is de kans dat een student het examen aflegt in meer dan 60 maar minder dan 75 minuten?

b. Neem aan dat in een klas 60 studenten zitten en dat ze maar 90 minuten de tijd krijgen om het examen te maken. Van hoeveel studenten verwacht je dat ze het examen niet afkrijgen?

c. Hoeveel tijd moet voorzien worden als 90% van de studenten voldoende tijd moet krijgen om het examen af te leggen?

Oefening 7

De levensduur van een machineonderdeel is normaal verdeeld met een gemiddelde van 5 jaar en een standaardafwijking van 1 jaar. De fabrikant vervangt het onderdeel gratis zolang het in garantie is. Hoeveel jaar garantie kan hij geven als hij niet meer dan 4% gratis wil vervangen?

3.249

Oefening 8

Een anesthesist beschikt over twee soorten medicaties om een patiënt te verdoven. De slaapduur bij de eerste medicatie is normaal verdeeld met een gemiddelde van 6 uur en een standaardafwijking van 1 uur. De slaapduur bij de tweede medicatie is normaal verdeeld met een gemiddelde van 5 uur en een standaardafwijking van 1.5 uur. Gezien de zwakte van de patiënt zou de anesthesist graag de tweede medicatie gebruiken, maar bovendien wil hij nog 99% zekerheid hebben dat de patiënt minstens 2.5 uur slaapt. Kan de anesthesist elk van deze medicaties gebruiken?

Oefening 9

De NV Maesen bestelt bouten bij de NV Publico. Als optimale lengte stelt Maesen een lengte van 60mm voor. Nochtans staat zij ook toe dat er bouten worden geleverd die 7mm afwijken van deze optimale lengte. Publico levert een grote vracht bouten die een gemiddelde lengte blijken te hebben van 58mm en een standaardafwijking van 4mm. We mogen ervan uitgaan dat de lengte van de bouten normaal verdeeld is. Hoeveel procent 'slechte' bouten werd er geleverd?

11.71

Oefening 10

Dirk behaalde 680 voor het wiskunde onderdeel in het toegangsexamen geneeskunde aan de universiteit. De puntenverdeling van dit examen is normaal verdeeld met gemiddelde 500 en standaardafwijking 100. Wouter doet mee aan het ingangsexamen burgerlijk ingenieur (deel wiskunde) en behaalt 27. Ook dit examen is normaal verdeeld en heeft als gemiddelde 18 en standaardafwijking 6.

In de veronderstelling dat beide testen in staat zijn in dezelfde mate de kennis te meten, wie behaalde de hoogste score?

Oefening 11

Een bandenfabrikant heeft een nieuwe band ontwikkeld. De managers denken dat een garantie op duurzaamheid belangrijk kan zijn voor de acceptatie van de band door de consumenten. Daarom willen ze inzicht krijgen in de kansverdeling van het aantal kilometers dat de band meegaat. Op grond van testen schat het technisch bureau van de firma dat de band gemiddeld 36.500 kilometer meegaat met een standaardafwijking van 5000 kilometer. Verder valt uit de gegevens op te maken dat het redelijk is ervan uit te gaan dat het aantal kilometers dat de band meegaat normaal verdeeld is.

a. Hoe groot is de kans dat een band meer dan 40.000 kilometer meegaat?
b. De bandenfabrikant overweegt een aantal kilometers te garanderen, en een korting te geven op de vervanging van een band als deze niet langer meegaat. Hoeveel kilometers kan de firma garanderen als ze niet wil dat meer dan 10% van alle banden in aanmerking komt voor een korting?