



Standaardisatie en z-scores

Prof. dr. Herman Callaert

Inhoudstafel

| | | |
|-------|---|---|
| 1 | Standaardisatie bij concreet cijfermateriaal..... | 1 |
| 1.1 | Een eerste voorbeeld: de punten van Pol..... | 1 |
| 1.1.1 | De ruwe score | 1 |
| 1.1.2 | Vergelijken met het klasgemiddelde | 1 |
| 1.1.3 | Variabiliteit rond het gemiddelde..... | 2 |
| 1.1.4 | z-scores | 2 |
| 1.2 | Een tweede voorbeeld: de punten van Emma..... | 3 |
| 1.2.1 | De ruwe score | 3 |
| 1.2.2 | Vergelijken met het klasgemiddelde | 3 |
| 1.2.3 | Variabiliteit rond het gemiddelde en z-scores | 4 |
| 1.2.4 | Hoort Emma bij dezelfde topgroep? | 4 |
| 2 | Standaardisatie bij kansmodellen | 5 |
| 2.1 | Normaal verdeelde populaties en z-scores | 5 |
| 2.1.1 | Vergelijken op basis van normale verdelingen | 5 |
| 2.1.2 | Vergelijken op basis van z-scores | 7 |
| 2.2 | Populaties die niet normaal verdeeld zijn..... | 8 |
| 2.2.1 | Een t-verdeling en een normale verdeling..... | 8 |
| 2.2.2 | De familie van de t-verdelingen. | 9 |

Als je resultaten (zoals examenpunten op verschillende vakken) met elkaar wil vergelijken, dan heb je dikwijls nood aan standaardisatie. Een 8/10 op geschiedenis en een 8/10 op fysica, dat is telkens een 8/10, maar is dat wel gelijkwaardig? Om dat te beoordelen moet je weten hoe de punten op die verschillende vakken gegeven worden. Je zal “punten op geschiedenis” en “punten op fysica” op een of andere manier standaardiseren om ze zo goed mogelijk met elkaar “vergelijkbaar” te maken.

Bij populaties die normaal verdeeld zijn, standaardiseer je met z-scores om resultaten “vergelijkbaar” te maken. Als populaties niet normaal verdeeld zijn, dan kan je de techniek van de z-scores niet zomaar toepassen. En je ziet z-scores ook opduiken bij de interpretatie van concreet cijfermateriaal. Soms is dat zinvol, soms ook niet.

1 Standaardisatie bij concreet cijfermateriaal

1.1 Een eerste voorbeeld: de punten van Pol

In een klas met 10 leerlingen zijn de resultaten op Nederlands, Frans en Duits als volgt:

| | | | | | | | | | | |
|------------|---|---|---|---|---|---|----|---|---|---|
| Nederlands | 7 | 5 | 8 | 7 | 6 | 3 | 5 | 2 | 4 | 3 |
| Frans | 9 | 7 | 5 | 7 | 6 | 9 | 10 | 4 | 5 | 8 |
| Duits | 7 | 5 | 7 | 8 | 7 | 7 | 8 | 7 | 7 | 7 |

Pol heeft op die drie vakken telkens 5 op 10 gehaald.

Er zijn nu verschillende mogelijkheden om de resultaten van Pol te interpreteren.

1.1.1 De ruwe score

Hierbij kijk je gewoon naar de behaalde punten, zonder enige verdere context. Je houdt hierbij geen rekening met wat de andere leerlingen gedaan hebben en ook niet met de manier waarop verschillende leerkrachten punten geven.

Pol haalde drie keer 5 op 10 en dus besluit je dat hij drie keer even goed presteerde.

1.1.2 Vergelijken met het klasgemiddelde

Wat Pol presteerde is één ding, maar wat zijn medeleerlingen op diezelfde toetsen presteerden zegt toch ook iets. Je kan bijvoorbeeld kijken naar de globale prestatie van de hele klas en het klasgemiddelde als een referentiepunt nemen.

Op Nederlands haalde de klas een gemiddelde van 5, op Frans was dat 7 en op Duits ook 7.

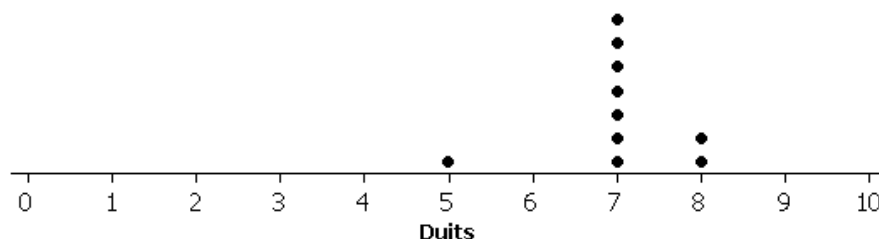
Zowel op Frans als op Duits scoorde Pol 2 punten lager dan het klasgemiddelde en dus heeft hij (in vergelijking met het gemiddelde) voor deze twee vakken op dezelfde manier gepresteerd. Zijn prestatie op Nederlands was beter want daar scoorde hij even hoog als het klasgemiddelde.

1.1.3 Variabiliteit rond het gemiddelde

Als je alleen het klasgemiddelde als referentie neemt, dan zie je bij Pol geen verschil tussen Frans en Duits (de score van Pol is telkens 5 en het klasgemiddelde is telkens 7). Zijn resultaten zal je nochtans heel anders interpreteren als je niet alleen naar het klasgemiddelde kijkt, maar ook naar de spreiding van de scores rond dat gemiddelde. Dat zie je op een eenvoudig puntendiagram.



Bij de toets Frans liggen de scores nogal gespreid. Twee leerlingen haalden een 5, er was ook een leerling met een 4 maar er waren er ook met 9 en 10. Voor de punten van die 10 leerlingen is het gemiddelde 7 en de standaardafwijking is 2.



De toets Duits ziet er helemaal anders uit. Iedereen haalde daar een 7 of een 8, behalve... Pol, die had een 5. Bij deze toets is het gemiddelde 7 en de standaardafwijking is (afgerond) 0.8.

Een getal uit een dataset zomaar vergelijken met het gemiddelde vertelt niet het hele verhaal. Soms geeft dit zelfs een verkeerd beeld. De variabiliteit rond dat gemiddelde speelt ook een rol. Bij Frans behaalde Pol een score die 1 standaardafwijking onder het gemiddelde ligt, want $5 = 7 - (1) \times (2)$. Bij Duits scoorde Pol 2.5 standaardafwijkingen onder het gemiddelde want $5 = 7 - (2.5) \times (0.8)$.

De standaardafwijking van een dataset is dikwijls een goede meetlat om punten uit die dataset te vergelijken met hun gemiddelde. Zo houd je ook rekening met de variabiliteit van de gegevens.

Als je de standaardafwijking als meetlat neemt dan heeft Pol “-1” op Frans en “-2.5” op Duits. In vergelijking met zijn medeleerlingen is zijn prestatie op Duits veel lager dan op Frans.

Frans: score van Pol = gemiddelde - 1 standaardafwijking

Duits: score van Pol = gemiddelde - 2.5 standaardafwijkingen.

1.1.4 z-scores

Naar analogie met de benaming en de notatie bij populaties die normaal verdeeld zijn, spreekt men ook hier over z-scores wanneer je bepaalt hoeveel standaardafwijkingen een oorspronkelijke ruwe score verwijderd is (in positieve of negatieve zin) van het gemiddelde.

Als je de scores op Frans noteert als $x_1, x_2, x_3, \dots, x_{10}$ met gemiddelde $\bar{x} = 7$ en standaardafwijking $s = 2$ dan heb je voor de score van Pol (genoteerd als x) dat:

score van Pol = gemiddelde - 1 standaardafwijking $\rightarrow 5 = 7 + (-1)(2)$ of $x = \bar{x} + z \cdot s$

zodat de z-score gelijk is aan: $z = \frac{x - \bar{x}}{s}$.

Notatie.

Om de punten op Frans en op Duits uit elkaar te houden gebruik je de volgende notatie:

Frans:

de punten: $x_1, x_2, x_3, \dots, x_{10}$ en algemeen $x_1, x_2, \dots, x_i, \dots, x_n$

het punt op Frans behaald door Pol: x

het gemiddelde: \bar{x}

de standaardafwijking: s_x

de z-score van Pol die hoort bij zijn punt x op Frans: z_x

Duits:

de punten: $y_1, y_2, y_3, \dots, y_{10}$ en algemeen $y_1, y_2, \dots, y_i, \dots, y_n$

het punt op Duits behaald door Pol: y

het gemiddelde: \bar{y}

de standaardafwijking: s_y

de z-score van Pol die hoort bij zijn punt y op Duits: z_y

Voor Pol geldt dan:

Frans: $x = 5$, $\bar{x} = 7$, $s_x = 2$ zodat $z_x = \frac{x - \bar{x}}{s_x} = \frac{5 - 7}{2} = -1$. Pol heeft op Frans een z-score = **-1**.

Duits: $y = 5$, $\bar{y} = 7$, $s_y = 0.8$ zodat $z_y = \frac{y - \bar{y}}{s_y} = \frac{5 - 7}{0.8} = -2.5$. Pol heeft op Duits een z-score = **-2.5**.

1.2 Een tweede voorbeeld: de punten van Emma

In een klas met 20 leerlingen zijn de resultaten op fysica en geschiedenis als volgt:

| | | | | | | | | | | | | | | | | | | | | |
|--------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fysica | 7 | 7 | 8 | 5 | 7 | 9 | 7 | 8 | 8 | 6 | 7 | 7 | 7 | 7 | 6 | 7 | 5 | 6 | 9 | 7 |
| Geschiedenis | 8 | 8 | 7 | 8 | 6 | 5 | 7 | 5 | 7 | 8 | 8 | 8 | 6 | 7 | 6 | 8 | 6 | 8 | 8 | 6 |

Emma heeft zowel op fysica als op geschiedenis een 8 gehaald. Hoe vergelijk je die 2 resultaten?

Nota. Gebruik een notatie met x voor fysica en y voor geschiedenis.

1.2.1 De ruwe score

Volgens dit criterium zijn de 2 prestaties identiek want:

punt op fysica: $x = 8$

punt op geschiedenis: $y = 8$

1.2.2 Vergelijken met het klasgemiddelde

Volgens dit criterium zijn de 2 prestaties identiek want:

klasgemiddelde op fysica: $\bar{x} = 7$ zodat $x - \bar{x} = 8 - 7 = 1$.

Emma scoort 1 punt boven het klasgemiddelde bij fysica.

klasgemiddelde op geschiedenis: $\bar{y} = 7$ zodat $y - \bar{y} = 8 - 7 = 1$.

Emma scoort 1 punt boven het klasgemiddelde bij geschiedenis.

1.2.3 Variabiliteit rond het gemiddelde en z-scores

Volgens dit criterium zijn de 2 prestaties identiek want:

standaardafwijking bij fysica: $s_x = 1.08$ zodat $\frac{x - \bar{x}}{s_x} = \frac{8 - 7}{1.08} = 0.93$.

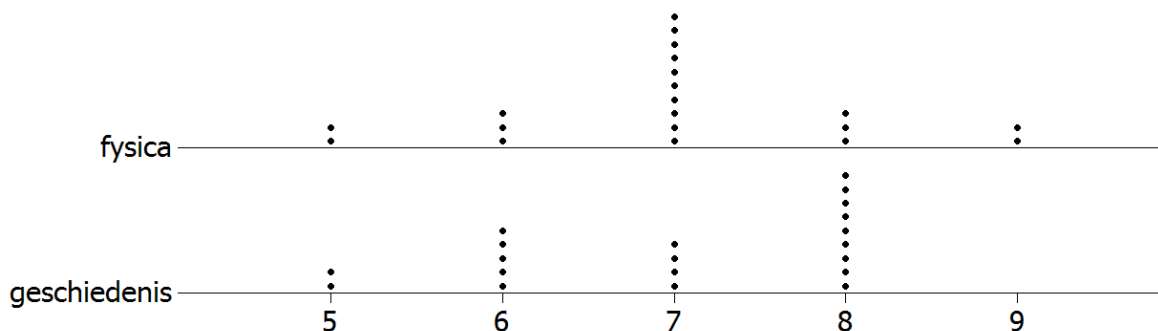
Op fysica haalt Emma een z-score van $z_x = 0.93$.

standaardafwijking bij geschiedenis: $s_y = 1.08$ zodat $\frac{y - \bar{y}}{s_y} = \frac{8 - 7}{1.08} = 0.93$.

Op geschiedenis haalt Emma een z-score van $z_y = 0.93$.

1.2.4 Hoort Emma bij dezelfde topgroep?

Zelfs als je rekening houdt met zowel het klasgemiddelde als met de variabiliteit rond dat gemiddelde, dan nog kan het gebeuren dat je het hele verhaal niet te pakken hebt. Ook de “vorm” van de puntenverdeling speelt een rol. Dat zie je goed op een figuur waar je de puntendiagrammen van beide vakken met elkaar vergelijkt.



Op fysica scoort Emma 8/10. In haar klas zijn er van de 20 leerlingen slechts 5 die het minstens even goed doen. Emma behoort tot de top 5 van haar klas of, als je dat in percent wil uitdrukken, dan behoort zij tot de 25 % beste van haar klas: 25 % van de leerlingen scoort minstens even goed als Emma en 75 % van de leerlingen scoort lager.

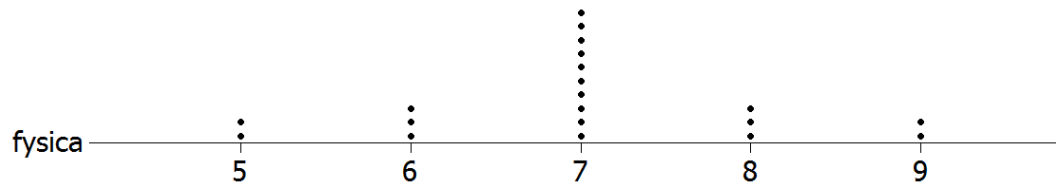
Voor geschiedenis kan je niet zeggen dat Emma tot de 25 % beste behoort. Van de 20 leerlingen zijn er hier 9 die (minstens) een 8 halen. Voor geschiedenis scoort 45 % van de leerlingen even goed als Emma en 55 % heeft een lagere score.

Naast de z-score speelt ook de vorm van de puntenverdeling een rol. Bij fysica zie je een hoge piek bij 7 (het gemiddelde) en de andere punten liggen daarrond symmetrisch gespreid. Bij geschiedenis is het gemiddelde ook 7 maar de punten liggen niet symmetrisch gespreid rond dit gemiddelde. De hoogste piek zie je bij 8 en daar breekt de verdeling af (voorbij 8 ligt er niets meer).

Als je op 2 verschillende vakken dezelfde z-score haalt, dan behoort je wel tot dezelfde topgroep als je te maken hebt met normale verdelingen. Dat leer je hieronder.

2 Standaardisatie bij kansmodellen

Kansmodellen zijn “theoretische” modellen die je gebruikt om een “concrete” werkelijkheid te benaderen. Die benadering heb je nodig omdat de werkelijkheid te complex is of omdat je bepaalde concrete gegevens in een meer globale context wil plaatsen.



In de klas van Emma zitten 20 leerlingen en hun punten op fysica zie je op de figuur. Emma heeft 8/10 gehaald en dus hoort zij bij de top 5 van haar klas. Om dit te zien heb je geen kansmodel (zoals een normale verdeling) nodig. Waarom zou je een normale verdeling gebruiken en zeggen dat Emma “benaderend” tot de beste 18 % van haar klas behoort als je hier kan aflezen dat zij “exact” bij de 25 % beste van haar klas zit?

Vraagstukken die beroep doen op “onderliggende theoretische kansmodellen” gaan ervan uit dat je werkt in een kader dat concrete opmetingen overstijgt.

Als je weet dat punten op fysica “over de jaren heen” en “opgemeten in zeer veel klassen van zeer veel scholen” een “globaal beeld” geven dat goed lijkt op een klokvormige curve, dan kan het verstandig zijn dat je “een normale verdeling” gebruikt om “de theoretische populatie van punten op fysica” te bestuderen.

Het resultaat van Emma bekijk je dan in het grotere kader van “resultaten op fysica”, eerder dan ten opzichte van de concrete 20 leerlingen in haar klas.

2.1 Normaal verdeelde populaties en z-scores

2.1.1 Vergelijken op basis van normale verdelingen

Bij de studie van normale verdelingen ontmoet je nogal eens een vraagstuk dat er als volgt uitziet.

Victor heeft op een toets aardrijkskunde 8/10 gehaald. Het klasgemiddelde was 7 en de standaardafwijking 1. Op biologie haalde Victor 7 en voor dat vak was het klasgemiddelde 5 en de standaardafwijking 1.5. Op welk vak heeft Victor, in vergelijking met zijn klasgenoten, het best gepresteerd? Je mag hierbij onderstellen dat de punten op aardrijkskunde en op biologie normaal verdeeld zijn.

Bovenstaand vraagstuk geeft de indruk dat het over een concrete leerling (Victor) gaat en over een concrete klas waarbij je de resultaten van Victor moet interpreteren in het kader van de scores van zijn medeleerlingen. Niets is minder waar.

Welke scores de medeleerlingen gehaald hebben weet je niet en je weet zelfs niet hoeveel leerlingen er in die klas zitten. Hoe kan je dan de score van Victor interpreteren in het kader van de scores van zijn medeleerlingen? Het enige wat je weet is het gemiddelde en de standaardafwijking van de klas. Daarmee kan je nog alle kanten uit zoals je zag in het voorbeeld over Emma. Daar hebben zowel fysica als geschiedenis hetzelfde gemiddelde en dezelfde standaardafwijking, terwijl Emma tot de top 25 % behoort bij fysica en tot de top 45 % bij geschiedenis.

Het vraagstuk, zoals het gesteld is, gaat niet over een concrete dataset van punten op aardrijkskunde of biologie. Door te zeggen dat de punten normaal verdeeld zijn stap je over op een onderliggend **populatiemodel** voor “punten op aardrijkskunde” [noteer dit model als X] en voor “punten op biologie” [noteer dit model als Y].

Voor de populaties is er dan gegeven dat de vorm van de verdeling normaal is:

$$\text{aardrijkskunde: } X \sim N(\mu_x; \sigma_x)$$

$$\text{biologie: } Y \sim N(\mu_y; \sigma_y)$$

De echte gemiddelden en standaardafwijkingen van de populaties ken je niet en daarom schat je die uit de opmetingen waar je vond: $\bar{x} = 7$, $s_x = 1$ en $\bar{y} = 5$, $s_y = 1.5$.

Dit brengt er je toe om voor het populatiegemiddelde μ_x de waarde 7 te nemen, voor σ_x neem je 1, voor μ_y 5 en voor σ_y 1.5.

Op die manier heb je een voorstel voor **het gedrag van de populatie van punten**:

$$\text{aardrijkskunde: } X \sim N(7; 1)$$

$$\text{biologie: } Y \sim N(5; 1.5)$$

De vraag over het punt van Victor wordt nu vertaald naar een vraag over een populatiewaarde: “tot welke topgroep behoort de waarde 8 bij een normaal verdeelde populatie met gemiddelde 7 en standaardafwijking 1?” of in formulevorm: $P(X \geq 8) = ???$ voor $X \sim N(7; 1)$.

En analoog: $P(Y \geq 7) = ???$ voor $Y \sim N(5; 1.5)$.

Nota.

Meer info over de normale verdeling kan je vinden in de tekst “Normaal verdeelde kansmodellen” op <http://www.uhasselt.be/lesmateriaal-statistiek>.

Oplossing:

| | | | |
|--|---|--|---|
| <pre>normalcdf lower:8 upper:10^99 μ:7 σ:1 Paste</pre> | <pre>normalcdf(8,10^99) .1586552596</pre> | <pre>normalcdf lower:7 upper:10^99 μ:5 σ:1.5 Paste</pre> | <pre>normalcdf(7,10^99) .0912112819</pre> |
|--|---|--|---|

$P(X \geq 8) = 0.16$ voor $X \sim N(7; 1)$: een 8 op aardrijkskunde hoort bij de top 16 %.

$P(Y \geq 7) = 0.09$ voor $Y \sim N(5; 1.5)$: een 7 op biologie hoort bij de top 9 %.

Besluit.

Als je de gevonden resultaten (bij kansmodellen voor populaties) mag gebruiken om iets te zeggen over de score van Victor, dan is zijn prestatie op biologie beter dan die op aardrijkskunde.

2.1.2 Vergelijken op basis van z-scores

Je weet dat, bij **normale verdelingen** zoals $X \sim N(7; 1)$ of $Y \sim N(5; 1.5)$, overstappen op z-scores leidt tot één unieke verdeling: de standaard normale verdeling $Z \sim N(0; 1)$. Je moet dan enkel deze verdeling gebruiken om de prestatie van Victor te beoordelen, in het kader van de populatieverdelingen van “punten op aardrijkskunde” en “punten op biologie”.

Voor punten op aardrijkskunde start je met $X \sim N(7; 1)$ zodat een populatiewaarde x gestandaardiseerd wordt tot: $z_x = \frac{x - \mu_x}{\sigma_x} = \frac{x - 7}{1}$. Bij $x = 8$ hoort $z_x = \frac{8 - 7}{1} = 1$.

Voor biologie geldt: $z_y = \frac{y - \mu_y}{\sigma_y} = \frac{y - 5}{1.5}$. Bij $y = 7$ hoort $z_y = \frac{7 - 5}{1.5} = \frac{2}{1.5} \cong 1.33$.

Bij een vaste verdeling (dat is hier de standaard normale $Z \sim N(0; 1)$) betekent een hogere z-waarde dat je tot een “hogere topklasse” behoort.

Als je alleen maar moet vergelijken dan moet je verder niets uitrekenen. De z-score op biologie ($z_y = 1.33$) is groter dan de z-score op aardrijkskunde ($z_x = 1$) en dus is de prestatie op biologie beter dan die op aardrijkskunde.

Als je precies wil weten tot welke topgroepen die z-scores behoren, dan haal je dat uit de standaard normale verdeling.

| | | | |
|--|---|--|---|
| <pre>normalcdf lower:1 upper:10^99 μ:0 σ:1 Paste</pre> | <pre>normalcdf(1,10^99) .1586552596</pre> | <pre>normalcdf lower:2/1.5 upper:10^99 μ:0 σ:1 Paste</pre> | <pre>normalcdf(2/1.5) .0912112819</pre> |
|--|---|--|---|

$P(Z \geq 1) = 0.16$: voor aardrijkskunde hoort Victor bij de top 16 %.

$P(Z \geq \frac{2}{1.5}) = 0.09$: voor biologie hoort Victor bij de top 9 %.

2.2 Populaties die niet normaal verdeeld zijn

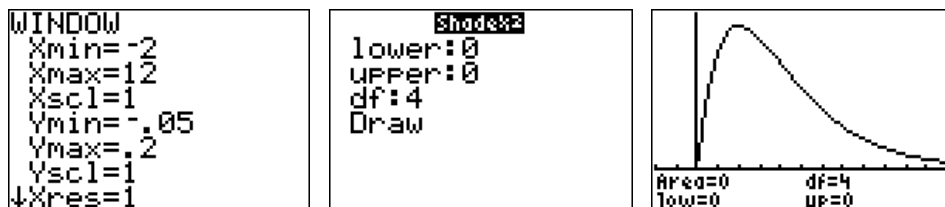
Als zowel de populatie van “punten op aardrijkskunde” als de populatie van “punten op biologie” kan beschreven worden door een kansmodel dat normaal verdeeld is, dan kan je vergelijken op basis van z-scores. In beide gevallen kom je immers terecht op de standaard normale verdeling.

Deze techniek van standaardiseren met z-scores kan je niet zomaar toepassen als je niet weet of de onderliggende populaties normaal verdeeld zijn.

Nota.

Naast populaties die normaal verdeeld zijn, zijn er ook heel veel andere populaties. Die beschrijf je met andere verdelingen zoals: t-verdelingen, chi-kwadraat verdelingen, F-verdelingen, enz.

Als voorbeeld kan je met de GRM eens een chi-kwadraat verdeling tekenen. Gebruik de aangegeven vensterinstellingen. Druk dan $\boxed{2\text{nd}}\boxed{[DISTR]}$, loop naar DRAW en druk $3:\text{Shade } \chi^2$ (. Vul in zoals aangegeven, ga op Draw staan en druk $\boxed{[ENTER]}$).



Je ziet dat de dichtheidsfunctie van deze chi-kwadraat verdeling helemaal niet lijkt op een symmetrische klokvormige curve.

2.2.1 Een t-verdeling en een normale verdeling.

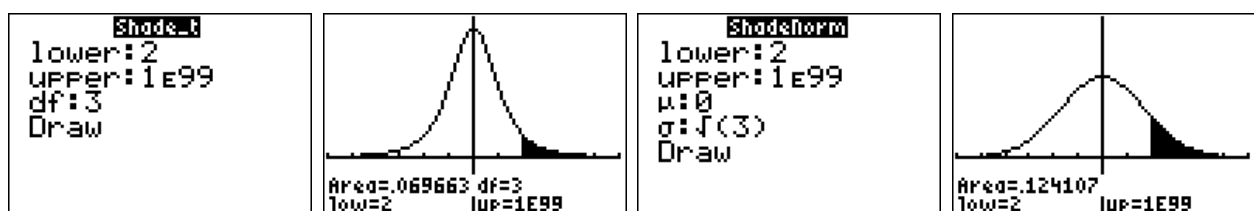
Onderstel dat een eerste populatie X verdeeld is volgens een t-verdeling met 3 vrijheidsgraden. Die verdeling heeft een gemiddelde $\mu_x = 0$ en een standaardafwijking $\sigma_x = \sqrt{3}$.

Onderstel dat een tweede populatie Y normaal verdeeld is met gemiddelde $\mu_y = 0$ en standaardafwijking $\sigma_y = \sqrt{3}$.

Neem nu in beide populaties de waarde 2. Vanuit het standpunt van z-scores is de waarde 2 even extreem in beide populaties want zowel voor X als voor Y geldt:

2 is gelijk aan het gemiddelde + **1.155 keer** de standaardafwijking: $2 = 0 + 1.155 \cdot \sqrt{3}$.

Maar de gestandaardiseerde vorm van die twee verdelingen is niet dezelfde en het gebied $[2; +\infty[$ bakent geen evenwaardig “top-gebied” af.



Bij de eerste populatie X hoort 2 tot de top 7%, bij de tweede populatie Y tot de top 12%.

2.2.2 De familie van de t-verdelingen.

Grafisch lijken t-verdelingen goed op normale verdelingen (met gemiddelde nul) maar toch hebben zij niet die typische normale eigenschap wanneer je bij t-verdelingen de standaardafwijking als nieuwe meetlat neemt.

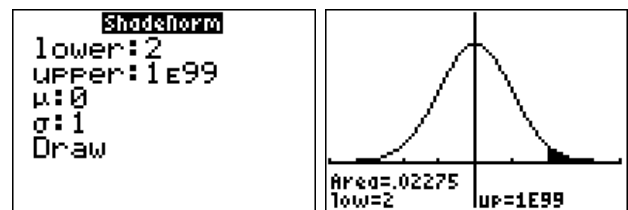
De familie van de normale verdelingen.

Dit weet je: bij elk normaal model is er dezelfde kans (ongeveer 2.3 %) om getallen uit te komen die minstens twee standaardafwijkingen groter zijn dan het gemiddelde.

In formulevorm: $P(X \geq \mu + 2\sigma) \cong 0.02275$

voor elke $X \sim N(\mu; \sigma)$.

Hiernaast zie je een voorbeeld met $\mu = 0$ en $\sigma = 1$ (de standaard normale verdeling).

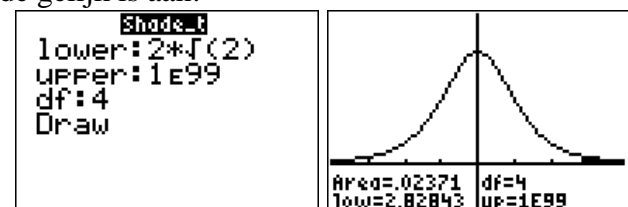


De familie van de t-verdelingen.

Als X verdeeld is volgens een t-verdeling met 4 vrijheidsgraden dan is het gemiddelde $\mu_x = 0$ en de standaardafwijking $\sigma_x = \sqrt{2}$ zodat de kans om getallen uit te komen die minstens twee standaardafwijkingen groter zijn dan het gemiddelde gelijk is aan:

$$P(X \geq \mu_x + 2\sigma_x) = P(X \geq 0 + 2\sqrt{2}) \cong 0.02371$$

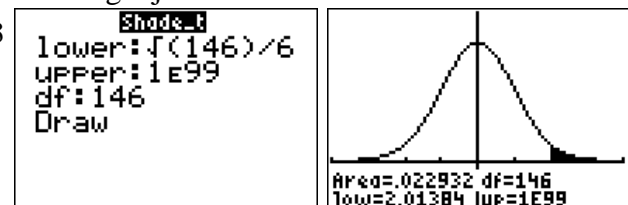
voor $X \sim t(4)$.



Als Y verdeeld is volgens een t-verdeling met 146 vrijheidsgraden dan is het gemiddelde $\mu_y = 0$ en de standaardafwijking $\sigma_y = \sqrt{146}/12$ zodat de kans om getallen uit te komen die minstens twee standaardafwijkingen groter zijn dan het gemiddelde gelijk is aan:

$$P(Y \geq \mu_y + 2\sigma_y) = P(Y \geq 0 + \sqrt{146}/6) \cong 0.02293$$

voor $Y \sim t(146)$.



Bij t-verdelingen is het dus NIET WAAR dat je bij elke t-verdeling dezelfde kans hebt om getallen uit te komen die minstens twee standaardafwijkingen groter zijn dan het gemiddelde.