



STATISTIEK VOOR HET SECUNDAIR ONDERWIJS

Verklarende statistiek

8. Betrouwbaarheidsintervallen voor gemiddelden

Werktekst voor de leerling

Prof. dr. Herman Callaert

Hans Bekaert
Cecile Goethals
Lies Provoost
Marc Vancaudenberg

Betrouwbaarheidsintervallen voor gemiddelden

1. Bouw verder op je kennis	2
2. Waarover gaat de onderzoeksvraag?	2
3. Verzamel je data	3
4. Bouwstenen voor een betrouwbaarheidsinterval.....	5
4.1. Hoe het werkte voor proporties	5
4.1.1. Een schatter voor de populatieproportie	5
4.1.2. Een model voor een interval: de theorie.....	6
4.1.3. Een model voor een interval: de praktijk	7
4.2. Hoe het werkt voor gemiddelden.....	9
4.2.1. Een schatter voor het populatiegemiddelde	9
4.2.2. Een model voor een interval: de theorie.....	10
4.2.3. Een model voor een interval: de praktijk	11
4.2.4. Wanneer mag je het praktijkmodel gebruiken?	12
4.2.5. Facultatief deeltje: de Student t-verdelingen	13
5. Betrouwbaarheid en precisie.....	15
5.1. De betrouwbaarheid veranderen	15
5.2. De foutenmarge	16
6. Hoe groot moet de steekproef zijn?	17
6.1. De theorie	17
6.2. De praktijk	18
6.2.1. Een benadering voor de <i>t</i> – waarde	18
6.2.2. Een benadering voor <i>s</i>	18

1. Bouw verder op je kennis

De belangrijkste begrippen die aan bod komen bij het opstellen van betrouwbaarheidsintervallen voor gemiddelden zijn dezelfde als deze die je ontmoet hebt bij het bestuderen van betrouwbaarheidsintervallen voor proporties.

Er zijn verschillende methoden om betrouwbaarheidsintervallen voor gemiddelden op te stellen. Die hangen af van de steekproefgrootte en van wat je over de populatie vooraf al weet. In deze tekst gaan wij die verschillende methoden niet behandelen. We beperken ons hier tot één situatie en kiezen daarbij voor een methode die goed werkt bij de overgrote meerderheid van realistische studies in de praktijk.

De verschillende stappen die nodig zijn om betrouwbaarheidsintervallen voor proporties op te stellen heb je vroeger reeds doorlopen. Die kennis ga je nu gebruiken om, zonder nieuwe theoretische afleidingen, betrouwbaarheidsintervallen voor gemiddelden te maken. Daarvoor moet je de tekst over betrouwbaarheidsintervallen voor proporties beheersen. Let daarbij vooral op de onderliggende **structuur** van de toen gebruikte procedures (die dezelfde zal blijven bij gemiddelden) en niet zozeer op de **concrete grootheden en getallen** (die er anders zullen uitzien bij gemiddelden).

2. Waarover gaat de onderzoeksvraag?

Voor alle kinderen die in 2003 in Vlaanderen geboren zijn heb je je vroeger afgevraagd wat de proportie jongens is bij al die geboorten. Dat is een vraag over **een kenmerk van de totale populatie** van al die geboorten. Wat de proportie jongens is, weet je niet omdat je de volledige populatie niet kent. Maar die **populatieproportie** bestaat wel. Het is een **vast** (maar ongekend) getal, genoteerd als π . Om over de populatieparameter π iets te weten te komen, trek je een steekproef en je kijkt welke proportie jongens je in je steekproef vindt. Die proportie noteer je als \hat{p} . Dat is je puntschatting voor de populatieproportie π . Die puntschatting geeft je een eerste idee over de proportie jongens in de totale populatie maar je weet ook dat een volgende steekproef uit diezelfde populatie je een andere \hat{p} zal opleveren. Dat is een probleem want welke \hat{p} is nu “de beste”? Om een idee te hebben over de nauwkeurigheid waarmee je iets over de echte π kan zeggen, kan je beter met een “interval” dan met een “punt” werken. Dat bestudeer je straks verder.

Voor alle kinderen die in 2003 in Vlaanderen geboren zijn, zou je je ook eens kunnen afvragen wat hun gemiddeld geboortegewicht is. Ook hier heb je dan een vraag over **een kenmerk van de totale populatie** van al die geboorten. Het gemiddelde van al die geboortegewichten ken je niet want je beschikt niet over de gewichten van al die baby's (het zijn er meer dan 60 000). Maar dat **populatiegemiddelde** bestaat wel. Het is een **vast** getal, genoteerd als μ . Om over μ iets te weten trek je een steekproef en bereken je het gemiddelde van alle gewichten die in jouw steekproef zijn terechtgekomen. Het gemiddelde dat jij gevonden hebt, noteer je als \bar{x} . Dat is jouw puntschatting voor μ . Maar ook hier levert een andere steekproef een andere \bar{x} op en heb je geen enkel idee welke \bar{x} “de beste” is. Daarom zal je overstappen op een interval.

Soms zorgt de onderzoeksvraag ervoor dat de **populatie** volgens een of ander kenmerk opgesplitst wordt in een **dichotomie** (= twee niet-overlappende delen) zoals meisje/jongen, jong/oud, voor/tegen, of algemeen: succes/mislukking. In zo'n geval heb je te maken met een **0-1 populatie** die je kan voorstellen door een Bernoulli kansverdeling met $P(\text{succes}) = \pi$. Om de ongekende populatieproportie π te bestuderen, moet je dan **methoden voor proporties** gebruiken.

Als de onderzoeksvraag gaat over een **populatie-eigenschap** die als **continu** kan behandeld worden (gewicht, lengte, tijdsduur,...) dan betekent dit dat je die populatie-eigenschap kan voorstellen door een continu kansmodel (zoals bijvoorbeeld door een normale verdeling). Als je bij een **continue populatie** iets wil weten over het populatiegemiddelde μ dan maak je gebruik van **methoden voor gemiddelden**.

3. Verzamel je data

Trek een steekproef van grootte $n = 70$ uit de databank van alle kinderen die in het jaar 2003 in Vlaanderen geboren zijn. Hoe je dat doet, staat hieronder kort beschreven.

Ga naar <http://www.uhasselt.be/lesmateriaal-statistiek>. Klik dan op "Link naar het lesmateriaal" en daarna op "Link naar de databank". Duid aan dat je enkel kinderen wil uit het jaar 2003 en dat je een steekproef van grootte $n = 70$ wil trekken. Trek dan de steekproef. Op dat ogenblik verschijnt ook een link: "Steekproefgegevens inbrengen in GRM. Handleiding". Die tekst kan je openen (of downloaden) en je hoeft daar enkel de eerste bladzijde te lezen om te leren hoe je snel de gegevens van je getrokken steekproef kan overzetten naar een Excel bestand en hoe je daaruit kolommen kan kopiëren naar lijsten in je GRM.

Voor het voorbeeld in deze tekst gingen we als volgt te werk:

DATABANK GEBORTEN IN VLAANDEREN

Kies de jaartallen waaruit je de steekproef wil trekken. Als je niets aangeeft dan worden alle jaartallen genomen.

1993 1994 1995 1996
 1997 1998 1999 2000
 2001 2002 2003 2004

Selecteer het geslacht (niets aangeven = alles): Jongen Meisje

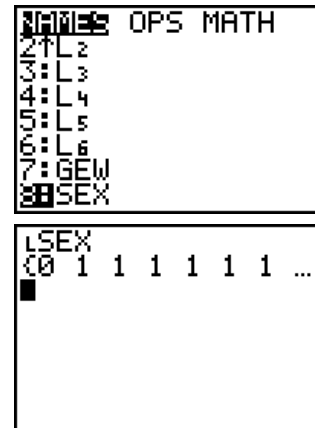
Hoe groot moet de steekproef zijn (n=?)

De gevonden resultaten werden in een Excel bestand overgebracht waarvan je hier een gedeelte ziet.

A	B	C	D	E	F
1	38	3050	0	25	2003
2	41	3110	1	24	2003
3	41	3610	1	26	2003
.....
.....
69	39	3280	1	21	2003
70	39	3250	0	32	2003

Daarna is de C-kolom gekopieerd naar een lijst in de GRM waarbij aan die lijst de naam GEW is gegeven. Tenslotte is de D-kolom naar een lijst in de GRM overgebracht met naam SEX.

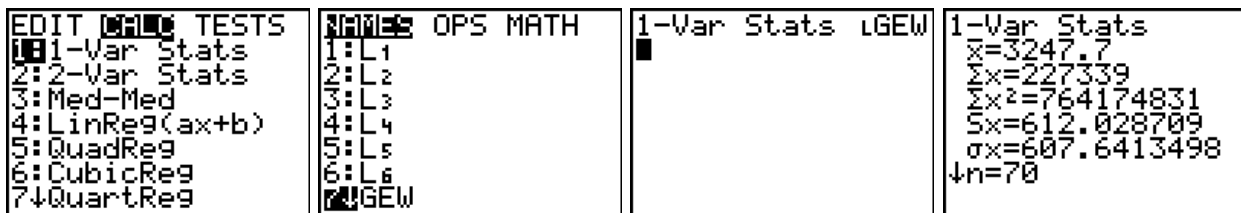
Als je in je GRM [2nd] [LIST] intikt, dan zie je de namen van de aanwezige lijsten. In dit voorbeeld volgen na de klassieke namen (van [L1] tot [L6]) de namen GEW en SEX. Als de cursor naast de naam van een lijst staat (bijvoorbeeld SEX) en je drukt twee keer [ENTER], dan zie je de eerste getallen die in die lijst staan. Met het pijltje \blacktriangleright kan je dan in die lijst verder scrollen en volgende getallen zien.



Doe nu hetzelfde zoals hierboven is beschreven maar werk met de data die jij in jouw steekproef hebt gevonden. Dat betekent ook dat je de volledige bespreking die nu volgt, moet uitvoeren met de steekproef die jij hebt getrokken.

In deze studie zal je grootheden nodig hebben die te maken hebben met het geslacht en met het gewicht. Daarbij zal je gegevens uit jouw steekproef moeten gebruiken. Heel wat klassieke grootheden kan je op een eenvoudige manier berekenen “voor later gebruik”. Daarvoor doe je het volgende.

Druk [STAT], loop naar CALC en druk op 1:1-Var Stats. Vervolledig dan het commando op je scherm als volgt: druk [2nd] [LIST], loop naar GEW en druk [ENTER]. Op je scherm staat nu het commando om statistische grootheden te berekenen voor de getallen in de lijst GEW.



Druk nu nog eens [ENTER] en schrijf de getallen op het scherm over (of maak een schermafdruck) zodat je die later kan gebruiken. Doe daarna hetzelfde voor de steekproefgegevens in de lijst SEX. Vat dan alle resultaten samen in een tabel zoals hieronder.

Berekeningen van statistische grootheden	
Geboortegewicht (in gram)	Geslacht jongen = 1 ; meisje = 0
GEW	SEX
<pre> 1-Var Stats x̄=3247.7 Sx=227339 Sx²=764174831 sx=612.028709 σx=607.6413498 n=70 </pre>	<pre> 1-Var Stats x̄=.5285714286 Sx=37 Sx²=37 sx=.5027872622 σx=.499183006 n=70 </pre>

Tabel 1

Opdracht 1

In tabel 1 staat twee keer de grootheid \bar{x} . Wat is de betekenis van deze grootheid bij GEW en wat is haar betekenis bij SEX? Geef een nauwkeurige formulering in woorden en voeg er wat uitleg aan toe.

4. Bouwstenen voor een betrouwbaarheidsinterval

4.1. Hoe het werkte voor proporties

In dit deeltje herhalen we de basisstappen die je gezet hebt bij het opstellen van een betrouwbaarheidsinterval voor een populatieproportie π . Het is de bedoeling dat je deze basisstappen heel goed begrijpt zodat je ze straks kan imiteren als het over gemiddelden zal gaan. Een gedetailleerde uitleg wordt hier niet meer gegeven. Als je extra uitleg nodig hebt, dan moet je even gaan kijken in de vroegere teksten over “Proporties” en over “Betrouwbaarheidsintervallen voor proporties”.

4.1.1. Een schatter voor de populatieproportie

Als je iets wil weten over de populatieproportie π (de proportie jongens in de totale populatie), dan trek je een steekproef uit die populatie. Bij een steekproefgrootte van bijvoorbeeld $n = 70$ zou het kunnen dat jij 37 jongens (= 37 successen of 37 enen) en 33 meisjes (= 33 mislukkingen of 33 nullen) vindt. Om een idee te krijgen over de **proportie** jongens π in de totale populatie bereken jij de proportie jongens die jij in je steekproef hebt gevonden. Dat is hier $\hat{p} = 37/70 \cong 0.529$. De proportie 0.529 is **jouw schatting** voor de waarde van π .

Als methode heb je een goede start genomen. Een vraag over een populatieproportie π (die je niet kent) beantwoord je met de waarde van een steekproefproportie \hat{p} . Maar steekproeven zijn aan het toeval onderhevig zodat verschillende waarden \hat{p} met verschillende kansen tot jou zullen komen als je meerdere keren een steekproef zou trekken. Hoe dat juist werkt wordt beschreven door het onderliggende kansmodel \hat{P} . Dat kansmodel noem je de steekproefproportie. Het is een **schatter** voor π .

4.1.2. Een model voor een interval: de theorie

De steekproefproportie \hat{P} heb je vroeger bestudeerd. Herinner je de volgende 3 eigenschappen.

1. Het gemiddelde van de steekproefproportie is gelijk aan de populatieproportie $E(\hat{P}) = \pi$.
2. De standaardfout van de steekproefproportie is gelijk aan de standaardafwijking van de populatie gedeeld door de wortel uit de steekproefgrootte $se(\hat{P}) = \frac{\sqrt{\pi(1-\pi)}}{\sqrt{n}}$.
3. Als de steekproef “groot genoeg” is dan gedraagt \hat{P} zich zoals een normaal verdeeld kansmodel, zodat $\hat{P} \sim N\left(\pi, \frac{\sqrt{\pi(1-\pi)}}{\sqrt{n}}\right)$.

Je weet dat alle normale kansmodellen met 95 % kans niet verder dan 1.96 standaardafwijkingen van hun gemiddelde vallen. Voor \hat{P} kan je deze eigenschap schrijven als:

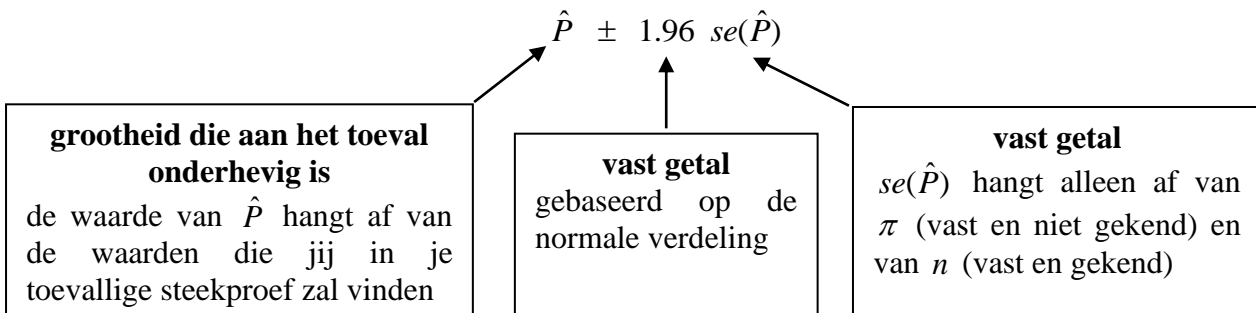
$$P(\pi - 1.96 se(\hat{P}) \leq \hat{P} \leq \pi + 1.96 se(\hat{P})) = 0.95$$

wat ook gelijk is aan:

$$P(\hat{P} - 1.96 se(\hat{P}) \leq \pi \leq \hat{P} + 1.96 se(\hat{P})) = 0.95.$$

Bemerk dat de bovenstaande formule een uitspraak doet over een model dat intervallen genereert die met 95 % kans de ongekende maar vaste populatieproportie π omsluiten.

Het kansmodel dat zegt op welke manier **intervallen** tot jou komen en met welke kans ze π omsluiten, ziet er dus als volgt uit:



Als je nu een steekproef van grootte $n = 70$ trekt en je vindt $\hat{p} = 37/70 \cong 0.529$, dan moet je jouw gevonden waarde alleen maar invullen in het kansmodel $\hat{P} \pm 1.96 se(\hat{P})$ om tot een 95 % betrouwbaarheidsinterval voor π te komen. Dat levert:

$$0.529 \pm 1.96 \text{ ???}$$

Inderdaad, $se(\hat{P})$ is een vast getal maar je kent het niet want de onbekende π komt erin voor. Hoe je deze moeilijkheid kan omzeilen, lees je hieronder.

4.1.3. Een model voor een interval: de praktijk

Als je in $se(\hat{P}) = \frac{\sqrt{\pi(1-\pi)}}{\sqrt{n}}$ de ongekende π vervangt door de schatting $\hat{p} = 37/70 \cong 0.529$ die jij in je steekproef hebt gevonden, dan ben je van het probleem verlost dat $se(\hat{P})$ niet te berekenen is.

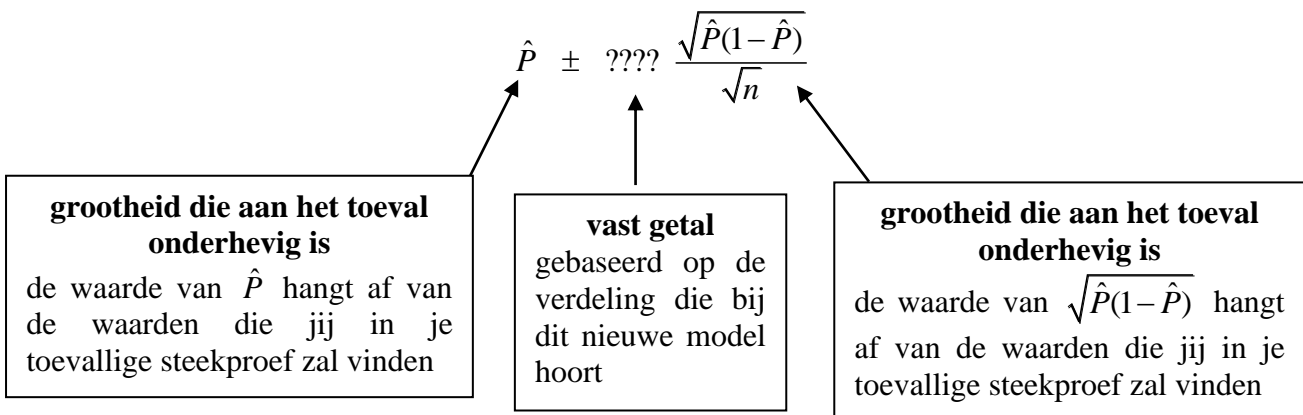
Je werkt dan gewoon met $\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} = \frac{\sqrt{0.529(1-0.529)}}{\sqrt{70}} = 0.060$.

Als je nu een volgende keer een steekproef trekt en je vindt 28 successen op 70 uitkomsten dan is $\hat{p} = 28/70 = 0.4$. Om dan de moeilijkheid van de ongekende $se(\hat{P}) = \frac{\sqrt{\pi(1-\pi)}}{\sqrt{n}}$ te omzeilen zal je

werken met $\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} = \frac{\sqrt{0.4(1-0.4)}}{\sqrt{70}} = 0.059$.

Welk getal je in de plaats van $se(\hat{P})$ moet invullen, hangt blijkbaar af van de resultaten in je toevallige steekproef. Dat in te vullen getal is dus zelf een uitkomst van een onderliggend model, namelijk $\frac{\sqrt{\hat{P}(1-\hat{P})}}{\sqrt{n}}$. In woorden zegt dit model: als je een steekproef zal trekken, dan moet je daarna het product berekenen van je “gevonden steekproefproportie” met “één min die steekproefproportie”. Uit dat resultaat trek je de vierkantswortel en tenslotte deel je nog eens door de wortel uit de steekproefgrootte.

Op deze manier heb je nu een nieuw kansmodel gebouwd voor het opstellen van betrouwbaarheidsintervallen. Het ziet er als volgt uit:



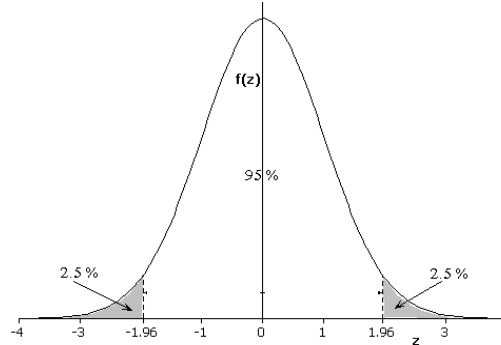
Je hebt vroeger geleerd dat je bij het bovenstaande kansmodel altijd kan werken met kritische punten van de normale verdeling. Als het criterium voor “voldoende grote steekproef” niet voldaan is dan kan je de +2/+4 regel gebruiken. Zo kan je altijd met het getal 1.96 werken om een 95 % betrouwbaarheidsinterval op te stellen. Als kansmodel krijg je dan:

$$\hat{P} \pm 1.96 \frac{\sqrt{\hat{P}(1-\hat{P})}}{\sqrt{n}}$$

of, in algemene notatie,

$$\hat{P} \pm (z\text{-waarde}) \frac{\sqrt{\hat{P}(1-\hat{P})}}{\sqrt{n}}$$

waarbij (*z-waarde*) verwijst naar het gepaste kritische punt uit de standaard normale verdeling. Voor een 95 % betrouwbaarheidsinterval is dit inderdaad het punt 1.96 zoals je op de onderstaande figuur kan aflezen.



In onze getrokken steekproef is $\hat{p} = 37/70 \cong 0.529$ zodat $\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} = \frac{\sqrt{0.529(1-0.529)}}{\sqrt{70}} = 0.060$.

Ons 95 % betrouwbaarheidsinterval is dus $0.529 \pm 1.96 \times 0.060$ of $[0.411 ; 0.646]$, wat ook

rechtstreeks met de GRM als volgt wordt gevonden.

Druk **[STAT]**, loop naar TESTS en dan naar A:1-PropZInt.. en **[ENTER]**. Vul dan het scherm in

<pre> EDIT CALC TESTS 7↑ZInterval... 8:TInterval... 9:2-SampZInt... 0:2-SampTInt... 1-PropZInt... 2-PropZInt... C↓X²-Test... </pre>	<pre> 1-PropZInt x:37 n:70 C-Level:.95 Calculate </pre>	<pre> 1-PropZInt (.41163,.64551) P=.5285714286 n=70 </pre>
---	---	--

zoals aangegeven. Ga daarna op Calculate staan en druk **[ENTER]**.

Opdracht 2

Maak nu met je eigen gegevens een 95 % betrouwbaarheidsinterval. Zeg in woorden waarvoor dit een betrouwbaarheidsinterval is en geef ook aan welke notatie je hier gebruikt. Heb jij hetzelfde betrouwbaarheidsinterval als je medeleerlingen? Hoe komt dat? Geef hiervoor een goede uitleg waarbij je naar “een model” verwijst.

Opdracht 3

Wat kan je zeggen over het door jou gevonden betrouwbaarheidsinterval in verband met de informatie die je wil te weten komen over π ? Kan je hier een kansuitspraak doen over π of over het door jou gevonden interval? Indien ja, welke uitspraak is dat dan? Indien neen, wat kan je dan wel zeggen over het verband tussen π en jouw betrouwbaarheidsinterval?

4.2. Hoe het werkt voor gemiddelden

Om bij een continue populatie een 95 % betrouwbaarheidsinterval voor het populatiegemiddelde μ op te stellen, kan je proberen om de bovenstaande stappen na te bootsen. Dat gaat als volgt.

4.2.1. Een schatter voor het populatiegemiddelde

Als je iets wil weten over het populatiegemiddelde μ (het gemiddeld geboortegewicht van al die baby's in de totale populatie), dan trek je een steekproef uit die populatie. Bij een steekproefgrootte van bijvoorbeeld $n = 70$ kan je met die 70 getrokken gewichten verschillende dingen doen. Maar als je iets wil weten over het **gemiddelde** gewicht van de baby's in de populatie, dan is het nogal logisch dat jij met de resultaten van je steekproef een gemiddelde uitreken, namelijk het gemiddelde van alle gewichten die jij in je steekproef hebt gevonden. Dat zou bijvoorbeeld gelijk kunnen zijn aan (in gram) $\bar{x} = \frac{1}{70} \sum_{i=1}^{70} x_i = 3247.7$. Het gemiddelde $\bar{x} = 3247.7$ is **jouw schatting** voor de waarde van μ .

Als methode ben je hier logisch tewerk gegaan. Een vraag over een populatiegemiddelde μ (dat je niet kent) beantwoord je met de waarde van een steekproefgemiddelde \bar{x} . Maar steekproeven zijn aan het toeval onderhevig zodat verschillende waarden \bar{x} met verschillende kansen tot jou zullen komen als je meerdere keren een steekproef zou trekken. Hoe al die \bar{x} -waarden juist tot jou komen wordt beschreven door het onderliggende kansmodel \bar{X} . Dat kansmodel noem je het steekproefgemiddelde. Het is een **schatter** voor het populatiegemiddelde μ .

4.2.2. Een model voor een interval: de theorie

Het steekproefgemiddelde \bar{X} heb je vroeger bestudeerd. Herinner je de volgende 3 eigenschappen.

1. Het gemiddelde van het steekproefgemiddelde is gelijk aan het populatiegemiddelde $E(\bar{X}) = \mu$.
2. De standaardfout van het steekproefgemiddelde is gelijk aan de standaardafwijking van de populatie gedeeld door de wortel uit de steekproefgrootte $se(\bar{X}) = \frac{\sigma}{\sqrt{n}}$.
3. Als de steekproef “groot genoeg” is dan gedraagt \bar{X} zich zoals een normaal verdeeld kansmodel, zodat $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Je weet dat alle normale kansmodellen met 95 % kans niet verder dan 1.96 standaardafwijkingen van hun gemiddelde vallen. Voor \bar{X} kan je deze eigenschap schrijven als:

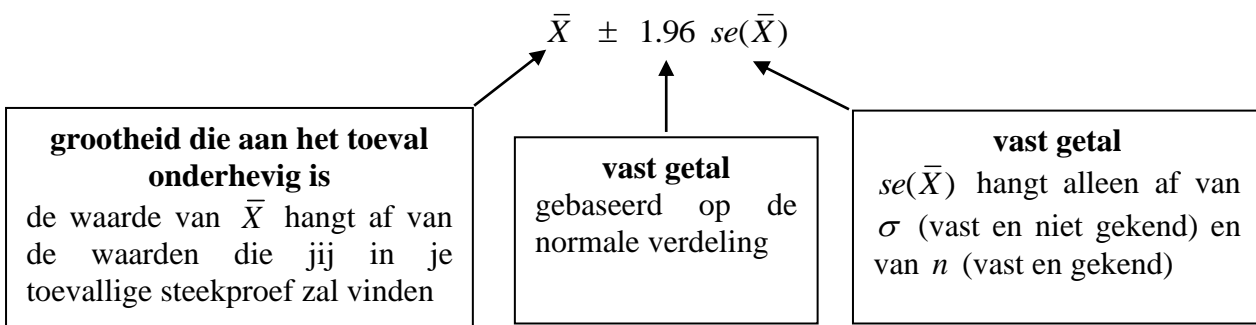
$$P(\mu - 1.96 se(\bar{X}) \leq \bar{X} \leq \mu + 1.96 se(\bar{X})) = 0.95$$

wat ook gelijk is aan:

$$P(\bar{X} - 1.96 se(\bar{X}) \leq \mu \leq \bar{X} + 1.96 se(\bar{X})) = 0.95.$$

Bemerk dat de bovenstaande formule een uitspraak doet over een model dat intervallen genereert die met 95 % kans het ongekende maar vaste populatiegemiddelde μ omsluiten.

Het kansmodel dat zegt op welke manier **intervallen** tot jou komen en met welke kans ze μ omsluiten, ziet er dus als volgt uit:



Als je nu een steekproef van grootte $n = 70$ trekt en je vindt $\bar{x} = 3247.7$, dan moet je jouw gevonden waarde alleen maar invullen in het kansmodel $\bar{X} \pm 1.96 se(\bar{X})$ om tot een 95 % betrouwbaarheidsinterval voor μ te komen. Dat levert:

$$3247.7 \pm 1.96 \text{ ???}$$

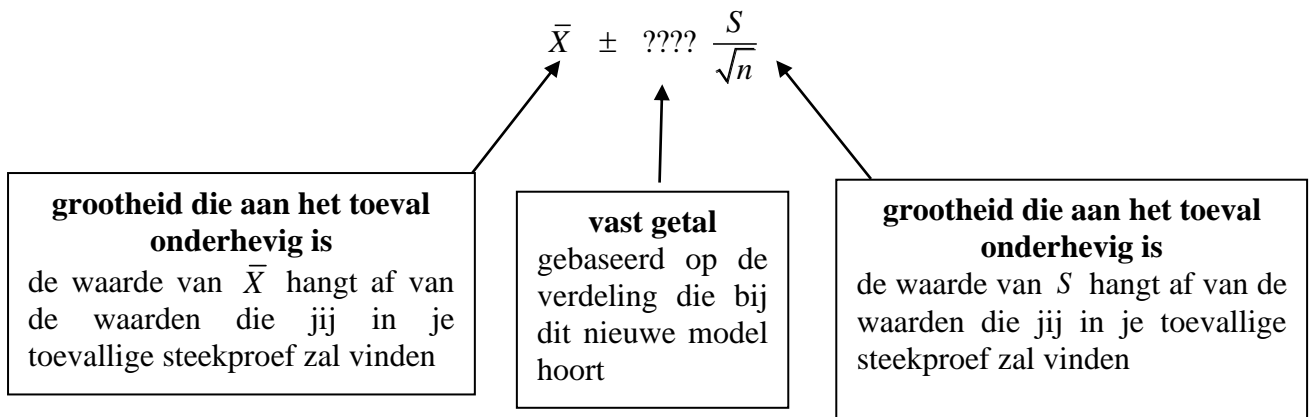
Inderdaad, $se(\bar{X})$ is een vast getal maar je kent het niet want de onbekende σ komt erin voor. Hoe je deze moeilijkheid kan omzeilen, lees je hieronder.

4.2.3. Een model voor een interval: de praktijk

In $se(\bar{X}) = \frac{\sigma}{\sqrt{n}}$ staat de ongekeerde standaardafwijking σ van de populatie. Die zou je kunnen vervangen door de standaardafwijking van de getallen die je in jouw steekproef hebt gevonden. Dat zou bijvoorbeeld $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{1}{69} \sum_{i=1}^{70} (x_i - \bar{x})^2} = 612.03$ kunnen zijn. Maar bij een volgende steekproef zou je wel eens $s = 546$ kunnen vinden en dan zal je de ongekeerde σ vervangen door 546.

Welk getal je in de plaats van $se(\bar{X})$ moet invullen, hangt blijkbaar af van de resultaten in je toevallige steekproef. Dat in te vullen getal is dus zelf een uitkomst van een onderliggend model, namelijk $\frac{S}{\sqrt{n}}$ waarbij $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$ de steekproefstandaardafwijking is.

Je hebt nu een nieuw kansmodel gebouwd voor het opstellen van betrouwbaarheidsintervallen. Het ziet er als volgt uit:

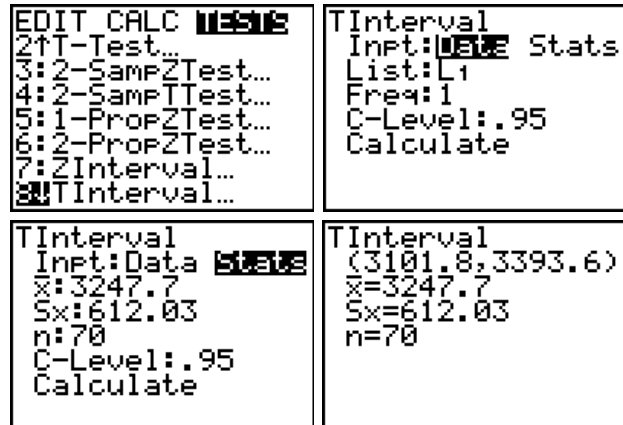


De verdeling die bij dit nieuwe model hoort is de t-verdeling. Als algemene formule krijg je dan

$$\bar{X} \pm (t\text{-waarde}) \frac{S}{\sqrt{n}}$$

De t-verdeling heb je nog niet bestudeerd maar zij lijkt goed op de standaard normale verdeling. Er zijn natuurlijk wel enkele verschillen. Iets meer informatie hierover kan je lezen in volgend puntje maar dat heb je allemaal niet nodig om je betrouwbaarheidsinterval op te stellen. Met je GRM gaat dit erg eenvoudig. Als je een steekproef van grootte $n = 70$ getrokken hebt en je hebt van die getallen reeds het gemiddelde $\bar{x} = 3247.7$ en de standaardafwijking $s = 612.03$ berekend, dan ga je als volgt te werk.

Druk **[STAT]**, loop naar TESTS, kies 8:TInterval en druk **[ENTER]**. Als je gegevens in een lijst staan (zoals [L1]) dan kies je bij Inpt: voor Data. Maar als je reeds de kengetallen \bar{x} en s berekend hebt, loop dan naar Stats, druk **[ENTER]** en vul de waarden in voor \bar{x} , voor s (op je GRM genoteerd als Sx) en voor n . Voor C-Level (Confidence level of betrouwbaarheidsniveau) neem je .95 (als je een 95 % betrouwbaarheidsinterval wil opstellen) en dan loop je naar Calculate en druk je **[ENTER]**. Na even wachten krijg je het gezochte interval. Hier is dat [3101.8 ; 3393.6], wat je ook kan schrijven als 3247.7 ± 145.9 .



Opdracht 4

Hierboven is getoond hoe je met de GRM een 95 % betrouwbaarheidsinterval opstelt. Volg de beschreven procedure maar werk met de resultaten die jij in jouw steekproef hebt gevonden. Zeg ook voor welke grootheid je nu een 95 % betrouwbaarheidsinterval hebt opgesteld en hoe deze grootheid genoteerd wordt.

4.2.4. Wanneer mag je het praktijkmodel gebruiken?

Wanneer je een populatie X hebt die grootheden beschrijft die je continu kan behandelen (zoals lengte, gewicht, tijdsduur,...) dan gebruik je het model $\bar{X} \pm (t\text{-waarde}) \frac{S}{\sqrt{n}}$ om voor het populatiegemiddelde μ betrouwbaarheidsintervallen op te stellen.

Om met een t -waarde te mogen werken, moet (in theorie) de onderliggende populatie X van lengten, gewichten,... te beschrijven zijn door een normaal kansmodel. Dat is niet altijd het geval, want je weet bijvoorbeeld dat geboortegewichten scheef naar links zijn. Gelukkig is het model $\bar{X} \pm (t\text{-waarde}) \frac{S}{\sqrt{n}}$ “robust”, wat wil zeggen dat je het nog altijd (als goede benadering) mag gebruiken, zelfs als de onderliggende populatie X niet normaal verdeeld is.

Wanneer je grote steekproeven hebt (dikwijls is $n \geq 30$ al voldoende), dan hoeft de populatie X helemaal niet normaal te zijn. Bij kleine steekproeven hoeft de populatie X ook niet perfect normaal te zijn, maar daar moet je er toch op letten dat ze niet te extreem scheef is of geen te extreme uitschieters bevat. En in ieder geval is het nodig dat de getallen waarmee je werkt afkomstig zijn van een goede steekproef (EAS).

Voorwaarden om het model $\bar{X} \pm (t\text{-waarde}) \frac{S}{\sqrt{n}}$ te mogen gebruiken bij het opstellen van betrouwbaarheidsintervallen voor een populatiegemiddelde μ .

- De getallen waarmee je werkt om jouw betrouwbaarheidsinterval op te stellen, zijn afkomstig van een “goede” steekproef (EAS) uit de populatie X .
- Als de steekproef klein is ($n < 30$) dan moet je redelijkerwijs kunnen onderstellen dat de populatie X niet te extreem afwijkt van een normaal kansmodel.

Opdracht 5

In opdracht 4 heb je met de GRM een betrouwbaarheidsinterval voor μ opgesteld. De GRM heeft daarbij het model $\bar{X} \pm (t\text{-waarde}) \frac{S}{\sqrt{n}}$ gebruikt. Waren de voorwaarden voldaan om met dat model te mogen werken? Verklaar.

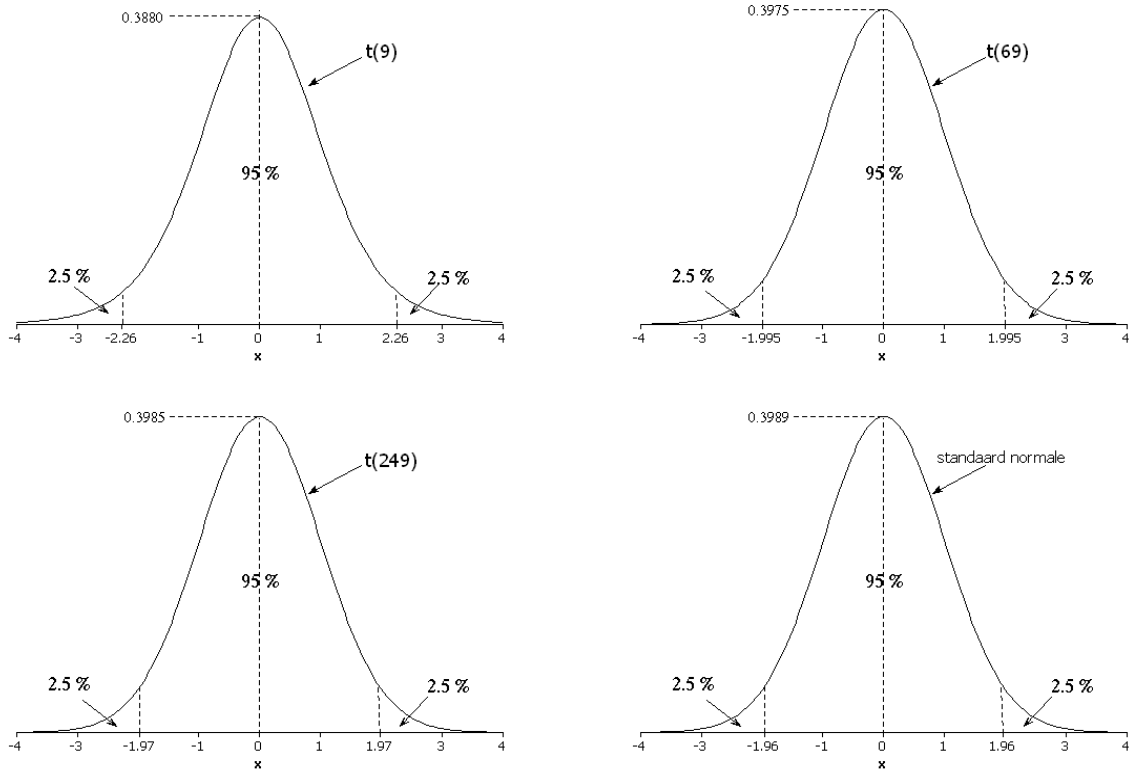
4.2.5. Facultatief deeltje: de Student t-verdelingen

William Gosset (1876 – 1937) is de statisticus die de t-verdeling ontdekte. Hij werkte in die tijd voor de Guinness brouwerij die aan haar medewerkers verbood om nieuwe vondsten te publiceren (want dat was fabrieksgeheim). Om elk verband met Guinness te verbergen publiceerde Gosset zijn resultaten onder het pseudoniem “Student”. Op die manier is de naam “Student t-verdeling” ontstaan.

Eigenlijk is er niet één t-verdeling, maar zijn er oneindig veel. In de wiskundige formule voor de dichtheidsfunctie van de t-verdeling staat een parameter k . Bij elke andere waarde van k hoort een andere functie en dus ook een andere t-verdeling. Welke waarde je voor de parameter k moet nemen, hangt af van de steekproefgrootte. Bij een steekproef van grootte 10 hoort een t-verdeling met parameterwaarde $k = 9$ en bij een steekproef van grootte 36 hoort een t-verdeling met parameterwaarde $k = 35$. Algemeen hoort bij een steekproef van grootte n een t-verdeling waarbij de parameterwaarde gelijk is aan $n - 1$. Die parameterwaarde krijgt ook de naam vrijheidsgraden zodat je bij een steekproef van grootte n een t-verdeling hebt met $n - 1$ vrijheidsgraden.

Elke t-verdeling ziet eruit zoals een standaard normale: klokvormig en symmetrisch rond nul, maar met staarten die “dikker” zijn. Naarmate het aantal vrijheidsgraden toeneemt (wat dus hoort bij grotere steekproeven) naderen de t-verdelingen naar de standaard normale verdeling.

Hieronder zie je voorbeelden van t-verdelingen met respectievelijk 9, 69 en 249 vrijheidsgraden. Zij horen bij steekproeven van grootte 10, 70 en 250. Je ziet telkens ook de kritische punten (*t-waarden*) die nodig zijn bij het opstellen van een 95 % betrouwbaarheidsinterval. De laatste figuur toont de standaard normale verdeling.



Bij de t-verdeling met 69 vrijheidsgraden is de *t-waarde* voor het opstellen van een 95 % betrouwbaarheidsinterval gelijk aan 1.995. Bij een steekproef van grootte $n = 70$ waarbij je $\bar{x} = 3247.7$ en $s = 612.03$ vindt, krijg je vanuit het model

$$\bar{X} \pm (t\text{-waarde}) \frac{S}{\sqrt{n}}$$

een 95 % betrouwbaarheidsinterval dat, voor deze steekproef, gelijk is aan

$$\bar{x} \pm (t\text{-waarde}) \frac{s}{\sqrt{n}} = 3247.7 \pm 1.995 \frac{612.03}{\sqrt{70}} = 3247.7 \pm 145.9$$

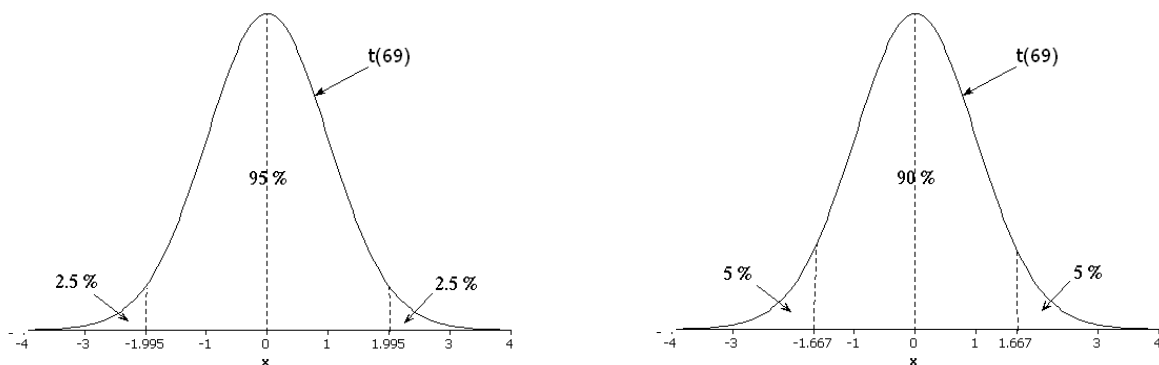
wat overeenkomt met het resultaat dat met de GRM is gevonden. Bemerkt dat je hier de *t-waarde* gebruikt hebt gebaseerd op een t-verdeling met $n - 1 = 69$ vrijheidsgraden.

5. Betrouwbaarheid en precisie

Wanneer je beslist om met een vooraf gekozen steekproefgrootte n te werken (bijvoorbeeld met $n = 70$), dan ligt nog niet alles vast. Tot nu toe heb je alleen **95 %** betrouwbaarheidsintervallen bestudeerd, maar je kan ook met een andere betrouwbaarheid werken.

5.1. De betrouwbaarheid veranderen

Werken met een andere betrouwbaarheid (bijvoorbeeld 90 %) komt erop neer dat je bij de gepaste t -verdeling ervoor zorgt dat je “rond nul” terechtkomt met kans 0.90 (in plaats van 0.95). Bij een t -verdeling met 69 vrijheidsgraden moet je dan werken met een t -waarde van 1.667 in plaats van 1.995. Dat zie je op de onderstaande figuren.



De t -waarden die je op de figuren kan aflezen, kan je eenvoudig uit je GRM halen. Juist zoals vroeger met `invNorm(` (werk je nu met `invT(` waarbij je aangeeft hoeveel kans er in de **linkerstaart** moet zitten (bijvoorbeeld 0.05). Daarna gebruik je de symmetrie om een even grote rechterstaart af te bakenen (waar je dus ook met kans 0.05 terechtkomt). In het midden kom je dan terecht met kans 0.90.

Druk `[2nd] [DISTR]`, loop naar `4:invT(` (en zeg dat je in de linkerstaart een kans van 0.05 wil bij een t -verdeling met 69 vrijheidsgraden. Tik dus

```

0:1: DRAW
1:normalPdf(
2:normalcdf(
3:invNorm(
4:invT(
5:tPdf(
6:tcdf(
7:χ²Pdf(
    
```

```

invT(0.05,69)
    
```

```

invT(0.05,69)
-1.667238489
    
```

`invT(0.05,69)` en druk `[ENTER]`. Na even wachten verschijnt het kritische punt -1.667 op het scherm. Dat betekent dat je bij een t -verdeling met 69 vrijheidsgraden met kans 0.05 terechtkomt in de linkerstaart $]-\infty; -1.667]$. Uit de symmetrie volgt dan dat je ook met kans 0.05 terechtkomt in de rechterstaart $[1.667; +\infty[$. In het midden, dus in het interval $[-1.667; 1.667]$, kom je terecht met kans 0.90.

Het model voor betrouwbaarheidsintervallen werkt met $\bar{X} \pm (t\text{-waarde}) \frac{S}{\sqrt{n}}$.

Als je in dit model de gevonden waarden invult, dan krijg je nu

$$\bar{x} \pm (t\text{-waarde}) \frac{s}{\sqrt{n}} = 3247.7 \pm 1.667 \frac{612.03}{\sqrt{70}} \cong 3247.7 \pm 122$$

zodat [3125.7 ; 3369.7] een 90 % betrouwbaarheidsinterval is voor het populatiegemiddelde μ .

Alle bovenstaande tussenstappen hoef je niet zelf te zetten. Je GRM geeft je rechtstreeks het 90 % betrouwbaarheidsinterval als volgt.

Druk **[STAT]**, loop naar TESTS, kies 8:TInterval en druk **[ENTER]**. Loop naar Stats, druk **[ENTER]** en vul de waarden \bar{x} , s en n in. Voor C-Level neem je .90 en dan loop je naar Calculate en druk je **[ENTER]**. Na even wachten krijg je het gezochte interval. Hier is dat [3125.7 ; 3369.7].

```
TInterval
Inet:Data Stats
x̄:3247.7
Sx:612.03
n:70
C-Level:.9
Calculate
```

```
TInterval
(3125.7,3369.7)
x̄=3247.7
Sx=612.03
n=70
```

Opdracht 6

Hierboven is getoond hoe je met de GRM een 90 % betrouwbaarheidsinterval opstelt. Volg de beschreven procedure maar werk met de resultaten die jij in jouw steekproef hebt gevonden. Voor welke grootheid heb je nu een 90 % betrouwbaarheidsinterval opgesteld?

5.2. De foutenmarge

De halve lengte van een betrouwbaarheidsinterval, dat is dus het getal dat na \pm staat, noem je de foutenmarge. Dat heb je vroeger geleerd.

Hierboven heb je gevonden dat:

- bij een betrouwbaarheid van 95 % het interval gelijk is aan

$$\bar{x} \pm (t\text{-waarde}) \frac{s}{\sqrt{n}} = 3247.7 \pm 1.995 \frac{612.03}{\sqrt{70}} = 3247.7 \pm 145.9$$

- bij een betrouwbaarheid van 90 % het interval gelijk is aan

$$\bar{x} \pm (t\text{-waarde}) \frac{s}{\sqrt{n}} = 3247.7 \pm 1.667 \frac{612.03}{\sqrt{70}} = 3247.7 \pm 122$$

Als je een grotere betrouwbaarheid eist (95 %), dan is de foutenmarge groter (145.9) en krijg je een breder en dus minder precies interval. Bij een kleinere betrouwbaarheid (90 %) vind je een kleinere foutenmarge (122) en dus een korter en nauwkeuriger interval. Deze eigenschap heb je vroeger ook

al ontmoet. Je moet er hier wel bij opmerken dat je intervallen gemaakt hebt op basis van **dezelfde** steekproefresultaten. Het enige wat er dan verandert is de *t-waarde*. Die wordt groter naarmate je met een grotere betrouwbaarheid werkt.

Wat je zopas gezien hebt, is algemeen waar. Een grotere betrouwbaarheid gaat samen met een grotere foutenmarge. Je hebt dan minder precieze informatie (een langer interval). Omgekeerd, als je bereid bent om een model te gebruiken dat met minder kans een “goed” interval genereert (bijvoorbeeld met kans 90 % in plaats van met kans 95 %), dan krijg je in ruil meer precieze informatie (een korter interval). Je moet dus zelf de voor- en nadelen afwegen wanneer je beslist met welk model je wil werken.

6. Hoe groot moet de steekproef zijn?

Vooraf bepalen hoe groot de te trekken steekproef moet zijn, is niet eenvoudig. We beperken ons hier tot enkele eenvoudige regels die je als eerste ruwe benadering kan gebruiken.

6.1. De theorie

Het model $\bar{X} \pm (t\text{-waarde}) \frac{S}{\sqrt{n}}$ is het kansmodel dat je gebruikt om betrouwbaarheidsintervallen voor een populatiegemiddelde μ te genereren. Als jij nu een steekproef trekt, dan kan je met de gevonden \bar{x} en s jouw 95 % betrouwbaarheidsinterval opstellen. Dat levert $\bar{x} \pm (t\text{-waarde}) \frac{s}{\sqrt{n}}$, wat een foutenmarge oplevert die gelijk is aan: $(t\text{-waarde}) \frac{s}{\sqrt{n}}$.

Je hebt dus:

$$\text{foutenmarge} = (t\text{-waarde}) \frac{s}{\sqrt{n}}$$

zodat

$$\sqrt{n} = (t\text{-waarde}) \frac{s}{\text{foutenmarge}}$$

of

$$n = (t\text{-waarde})^2 \frac{s^2}{\text{foutenmarge}^2}$$

Deze theoretische formule om bij een vooropgestelde foutenmarge en een vooropgestelde betrouwbaarheid de steekproefgrootte n te bepalen, helpt je niet veel want

- je bent aan het bepalen hoe groot de steekproef moet zijn en dus ken je n nog niet, zodat je ook het aantal vrijheidsgraden $n-1$ niet kent. Je weet niet bij welke t-verdeling je moet kijken om de gepaste *t-waarde* voor een 95 % betrouwbaarheidsinterval te vinden.
- in de formule komt ook s voor en dat is de standaardafwijking van de getallen in je steekproef. Maar je hebt je steekproef nog niet!

6.2. De praktijk

We bespreken hier een (ruwe) benadering voor het opstellen van 95 % betrouwbaarheidsintervallen.

6.2.1. Een benadering voor de t -waarde

Bij verschillende steekproefgrootten n horen verschillende vrijheidsgraden $n-1$ en dus ook verschillende t -verdelingen met verschillende t -waarden voor een 95 % betrouwbaarheidsinterval. In de onderstaande tabel zie je wat voorbeelden.

steekproefgrootte n	vrijheidsgraden $n-1$	t -waarde
10	9	2.262
20	19	2.093
30	29	2.045
40	39	2.023
50	49	2.010
60	59	2.001
70	69	1.995
80	79	1.990
90	89	1.987
100	99	1.984
200	199	1.972
300	299	1.968
.....	
∞	∞	1.96

Van zodra de steekproefgrootte niet te extreem klein is schommelt de t -waarde rond 2 en nadert naar 1.96 als de steekproef groter en groter wordt. Als je dus als benadering voor de t -waarde het getal 2 neemt, dan zit je meestal goed.

Als eerste benadering werk je dus met

$$n = 2^2 \frac{s^2}{\text{foutenmarge}^2} = 4 \frac{s^2}{\text{foutenmarge}^2}$$

Nu moet je nog iets weten over “wat je als standaardafwijking zou vinden als je een steekproef zou gaan trekken”.

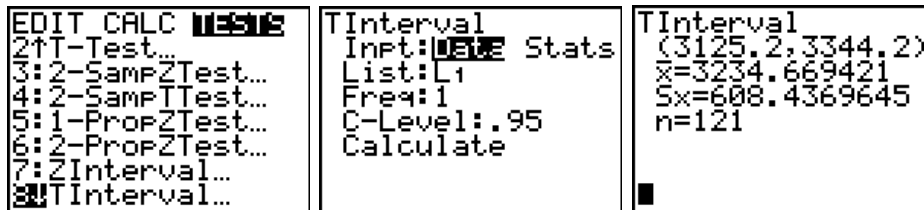
6.2.2. Een benadering voor s

Soms heb je informatie uit andere analoge studies. Zo kan je een idee krijgen over wat jij als standaardafwijking kan verwachten in de te trekken steekproef. Dergelijke informatie kan je dan gebruiken in je formule.

Als je uit vroegere studies over geboortegewichten weet dat men toen standaardafwijkingen vond die in de buurt van 550 gram lagen, dan kan je die informatie gebruiken. Om dan voor het gemiddelde geboortegewicht μ een 95 % betrouwbaarheidsinterval op te stellen met een foutenmarge van 100 gram, moet je (bij benadering) een steekproef trekken van grootte:

$$n = 4 \frac{550^2}{100^2} = 121$$

Als controle kan je eens een steekproef van grootte 121 trekken uit de geboortegewichten van 2003 en controleren of jouw resultaten ongeveer overeenkomen met je verwachtingen. Wij hebben dat ook eens gedaan. Wij hebben eerst een steekproef van grootte 121 getrokken (eigenlijk een steekproef van grootte 125 waarbij we de laatste 4 gewichten hebben weggelaten) en dan de getallen overgebracht in de lijst [L1] van de GRM. Dan hebben we een 95 % betrouwbaarheidsinterval als volgt opgesteld.



De data in onze toevallige steekproef hebben een standaardafwijking van $s = 608.44$, wat meer is dan de vermoede 550. Het 95 % betrouwbaarheidsinterval is $[3125.2; 3344.2] = 3234.7 \pm 109.5$ zodat we hier een foutenmarge hebben van 109.5, wat iets meer is dan de vooropgestelde 100.

Opdracht 7

Imiteer de bovenstaande werkwijze (met je GRM) en trek zelf eens een steekproef van grootte 121 uit die populatie. Vul daarna de onderstaande tabel in met resultaten van 5 medeleerlingen (die allemaal een steekproef van grootte 121 getrokken hebben). Welk verband zie je in de tabel tussen de standaardafwijkingen en de foutenmarges? Kan je dit verband in een formule uitdrukken? Je medeleerlingen hebben allemaal met dezelfde betrouwbaarheid (95 %) gewerkt en zij hebben allemaal een steekproef getrokken die even groot was ($n=121$) en toch krijgen zij verschillende foutenmarges. Hoe verklaar je dat?

	gevonden standaardafwijking s	berekende foutenmarge
Leerling 1		
Leerling 2		
Leerling 3		
Leerling 4		
Leerling 5		

Een andere manier om s te benaderen is als volgt. Als de populatie normaal verdeeld zou zijn dan vallen zo goed als alle gegevens in het gebied $\mu \pm 3\sigma$ want voor elke normaal verdeelde X geldt dat

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = P(-3 \leq \frac{X - \mu}{\sigma} \leq +3) = 0.997 \text{ [normalcdf(-3,3,0,1)]}.$$

Doe nu een verstandige gok voor het gebied waarbinnen je de overgrote meerderheid van je resultaten verwacht en onderstel dat de s in je steekproef niet te veel zal verschillen van de σ in de populatie. Stel dan de lengte van het verwachte gebied gelijk aan $6s$ en los hieruit s op.

Bij de geboortegewichten verwacht je er niet veel onder 1 kilogram en ook niet veel boven 5 kilogram. Dat betekent dat waarschijnlijk de overgrote meerderheid van de gewichten tussen 1000 en 5000 ligt. Stel dan $5000 - 1000 = 4000$ gelijk aan $6s$ zodat $s \cong 667$. Volgens deze benadering heb je dan een steekproef nodig van grootte

$$n = 4 \frac{667^2}{100^2} \cong 178$$

Als controle hebben we ook hier eens een steekproef van grootte 178 getrokken en de resultaten in [L1] geplaatst.

<pre> EDIT CALC TESTS 2:T-Test... 3:2-SampZTest... 4:2-SampTTest... 5:1-PropZTest... 6:2-PropZTest... 7:ZInterval... 8:TInterval... </pre>	<pre> TInterval Inpt: DATA Stats List: L1 Freq: 1 C-Level: .95 Calculate </pre>	<pre> TInterval (3164.1, 3352.9) x=3258.505618 Sx=638.0798725 n=178 </pre>
---	--	--

Het 95 % betrouwbaarheidsinterval voor μ is $[3164.1; 3352.9] = 3258.5 \pm 94.4$ We hebben hier een foutenmarge van 94.4 wat een beetje kleiner is dan de vooropgestelde 100.

Opdracht 8

Hierboven is 3258.5 ± 94.4 gevonden als betrouwbaarheidsinterval. Gebruik dit resultaat om aan leerlingen van een andere klas, die niet veel weten over statistiek, kort en duidelijk uit te leggen wat je gevonden hebt. Let daarbij goed op de betekenis van een betrouwbaarheidsinterval.