

Statistiek in business voor gevorderden

Statistiek in business voor gevorderden

Rob Erven

Tweede druk

2014
Boom Lemma uitgevers
Den Haag

Inhoud

Inleiding 12

1	Werken met een gemiddelde	15
1.1	Case Mercedes	16
1.2	Betrouwbaarheidsinterval één gemiddelde	17
1.3	Minimale steekproefgrootte één gemiddelde	22
1.4	Betrouwbaarheidsinterval één gemiddelde beperkt eindige populatie	24
1.5	Minimale steekproefgrootte één gemiddelde beperkt eindige populatie	26
	Trefwoorden	26
	Samenvatting	27
	Kennisopgaven	27
	Toegepaste opgaven	28
	Competentieprikkel	30

2	Toetsen op één gemiddelde	33
2.1	Case Philips	34
2.2	Het formuleren van hypothesen	34
2.3	De toetsingsgrootte	35
2.4	Het kritieke gebied	36
2.5	De conclusie	37
2.6	Een andere benadering	38
2.7	Een- en tweezijdig toetsen	40
2.8	Fout van de eerste en tweede soort	41
2.9	De p-waarde	43
	Trefwoorden	46
	Samenvatting	46
	Kennisopgaven	46
	Toegepaste opgaven	47
	Competentieprikkel	49

- 3 Toetsen op twee gemiddelden 53**
 - 3.1 Case kantoorkosten Amsterdam en Brussel 53
 - 3.2 Toetsen op twee gemiddelden: onafhankelijke steekproeven 54
 - 3.3 Gepaarde toets op twee gemiddelden 56
 - Trefwoorden 60
 - Samenvatting 60
 - Kennisopgaven 60
 - Toegepaste opgaven 61
 - Competentieprikkels 63

- 4 Kleine aantallen 67**
 - 4.1 Case klm 68
 - 4.2 Betrouwbaarheidsintervalgemiddelde bij kleine aantallen 69
 - 4.3 Toets op één gemiddelde en kleine aantallen 71
 - 4.4 Toets op twee gemiddelden en kleine aantallen 72
 - 4.5 Gepaarde toets gemiddelde bij kleine aantallen 75
 - Trefwoorden 78
 - Samenvatting 78
 - Kennisopgaven 78
 - Toegepaste opgaven 79
 - Competentieprikkels 82

- 5 Toetsen op standaarddeviaties 85**
 - 5.1 Case Nike 85
 - 5.2 Toets op één standaarddeviatie 86
 - 5.3 Toets op twee standaarddeviaties 89
 - Trefwoorden 92
 - Samenvatting 92
 - Kennisopgaven 92
 - Toegepaste opgaven 92
 - Competentieprikkels 95

- 6 Werken met een percentage 97**
 - 6.1 Case prijzenoorlog 98
 - 6.2 Betrouwbaarheidsinterval één percentage z-benadering 99
 - 6.3 Betrouwbaarheidsinterval één percentage Poisson-benadering 101
 - 6.4 Minimale steekproefgrootte één percentage 103
 - 6.5 Betrouwbaarheidsinterval één percentage beperkt eindige populatie 104

6.6	Minimale steekproefgrootte één percentage beperkt eindige populatie	105
	Trefwoorden	107
	Samenvatting	107
	Kennisopgaven	108
	Toegepaste opgaven	108
	Competentieprikkel	111
7	Toetsen op percentages	115
7.1	Case accountants falen	116
7.2	Toets op één percentage z-benadering	117
7.3	Toets op één percentage Poisson-benadering	119
7.4	Toets op gelijkheid van twee percentages	122
7.5	Toets op verondersteld verschil van twee percentages	125
7.6	Toets op meer dan twee percentages	127
	Trefwoorden	130
	Samenvatting	130
	Kennisopgaven	131
	Toegepaste opgaven	131
	Competentieprikkel	136
8	Analyse van een kruistabel	139
8.1	Case Esprit	139
8.2	Kruistabel en percenteren	140
8.3	De chikwadraattoets	142
	Trefwoorden	145
	Samenvatting	145
	Kennisopgaven	145
	Toegepaste opgaven	146
	Competentieprikkel	149
9	Variantieanalyse	151
9.1	Case de beste verpakking	151
9.3	De Tukey-toets	159
	Trefwoorden	160
	Samenvatting	160
	Kennisopgaven	161
	Toegepaste opgaven	161
	Competentieprikkel	163

10	Lineaire regressie	167
10.1	Case ab Seafood	167
10.2	Het spreidingsdiagram	168
10.3	Regressievergelijking	169
10.4	Voorspellingsinterval	173
10.5	Toets op correlatie	175
	Trefwoorden	177
	Samenvatting	178
	Kennisopgaven	178
	Toegepaste opgaven	178
	Competentieprikkels	181
11	Niet-parametrische toetsen	185
11.1	Case B&C Products	185
11.2	De tekentoets	187
11.3	De Wilcoxon-toets	191
11.4	De Mann-Whitney-toets	195
11.5	De Kruskal-Wallis-toets	201
11.6	De rangcorrelatiecoëfficiënt van Spearman	204
	Trefwoorden	207
	Samenvatting	207
	Kennisopgaven	208
	Toegepaste opgaven	208
	Competentieprikkels	212
12	Het keuzeproces	215
12.1	Case iedereen tv-manager	215
12.2	Het keuzeproces	217
	Gemiddelde	219
	Standaarddeviatie	220
	Percentage	220
	Correlatie	220
	Niet-parametrische toetsen	221
	Trefwoorden	223
	Samenvatting	223
	Toegepaste opgaven	223

Bijlagen 229

- 1 Tabel standaardnormale verdeling 229
- 2 Tabel t-verdeling 231
- 3 Tabel chikwadraatverdeling 233
- 4 Tabel F-verdeling 235
- 5 Tabel binomiale verdeling 236
- 6 Tabel Poisson-verdeling 240
- 7 Tabel betrouwbaarheidsinterval voor μ met de Poisson-verdeling 242
- 8 Tabel Wilcoxon-toets 243
- 9 Tabel Mann-Whitney-toets 244
- 10 De Poisson-verdeling 245

Formuleblad 247**Register 256****Over de auteur 259**

Inleiding

In het *Basisboek statistiek in business* heb je gezien dat statistiek wordt toegepast in diverse gebieden binnen het zakendoen. Bijvoorbeeld in de marketing, accountancycontrole, logistiek en communicatie, maar ook bij bedrijfseconomische onderwerpen en internationaal onderzoek voor exportmogelijkheden. In *Statistiek in business voor gevorderden* wordt deze lijn doorgetrokken, zodat dit boek voor alle economische opleidingen in het hoger onderwijs geschikt is. Elk hoofdstuk begint met een praktijkcase, die wordt opgelost met de behandelde theorie. Tussen de theorie staan korte opdrachten waarmee je kunt controleren of je de stof hebt begrepen. Aan het eind van het hoofdstuk staan trefwoorden, een samenvatting, een aantal kennisopgaven, toegepaste opgaven en een competentieprikkel.

Achter in het boek staan de diverse tabellen, een korte uitleg van de Poisson-verdeling en een formuleblad met een overzicht van alle gebruikte formules.

Voor degene die nog meer wil oefenen, staan er op de website bij dit boek, www.sibvoorgevorderden.nl, multiplechoicevragen met antwoorden. Bovendien kun je op deze site een aantal korte instructies vinden over de toepassing van de grafische rekenmachine TI-83 en kun je een powerpointpresentatie van elk hoofdstuk downloaden. Daarnaast vind je er een SPSS-trainer, waarin uitgebreid aandacht wordt besteed aan het verkrijgen van vaardigheden in SPSS. Alle statistische technieken die worden behandeld, passeren hierbij de revue.

De volgende competenties worden bereikt met het bestuderen van dit boek:

- Je bent in staat om een analyse uit te voeren bij één of meer gemiddelden.
- Je kunt een analyse uitvoeren bij één of twee standaarddeviaties.
- Je kunt een analyse uitvoeren bij één of meerdere percentages.
- Je kunt het mogelijke lineaire verband tussen twee economische variabelen onderzoeken.
- Je bent in staat om diverse niet-parametrische toetsen uit te voeren.
- Je kunt de juiste keuze maken bij de oplossing van een statistisch vraagstuk in een businesscontext.

Deze competenties leiden tot de onderwerpen in dit boek. Deze onderwerpen komen terug in het schema aan het eind van hoofdstuk 12. In dit schema staan vijf belangrijke trefwoorden:

- gemiddelde;
- standaarddeviatie;
- percentage;
- correlatie;
- niet-parametrische toetsen.

Achter deze trefwoorden staan alle statistische technieken. Aan de hand van het schema wordt uitgelegd wanneer je welke techniek kunt gebruiken.

Gemiddelde	Betrouwbaarheidsinterval grote n, zeer grote populatie	§ 1.2
	Betrouwbaarheidsinterval grote n, beperkt eindige populatie	§ 1.4
	Minimale steekproefgrootte zeer grote populatie	§ 1.3
	Minimale steekproefgrootte beperkt eindige populatie	§ 1.5
Standaarddeviatie		
Percentage		
Correlatie		
Niet-parametrische toetsen		



Werken met een gemiddelde

Veel bedrijven en ondernemingen laten regelmatig onderzoeken wat hun mogelijkheden zijn op de markt. Dikwijls gebeurt dit door het afnemen van een enquête. Nadat de enquête is afgenomen, worden de resultaten verwerkt in tabellen en grafieken. Ook worden diverse centrummaten en spreidingsmaten berekend, waarvan de bekendste het gemiddelde is. Bij een gemiddelde kun je de vraag stellen hoe nauwkeurig deze waarde is. Hiertoe bestuderen we het begrip betrouwbaarheidsinterval voor een gemiddelde.

Verder staan we stil bij de vraag hoeveel enquêtes je minimaal moet afnemen om een statistisch verantwoorde uitspraak te doen over een gemiddelde. Tot slot besteden we aandacht aan de invloed van de grootte van de populatie op de berekeningen van het betrouwbaarheidsinterval en de minimale steekproefgrootte.

Leerdoelen

Kennis:

- Je kent de begrippen populatie en steekproef.
- Je kent de begrippen populatiegemiddelde en steekproefgemiddelde.
- Je kent de begrippen populatiestandaarddeviatie en steekproefstandaarddeviatie.
- Je weet wat de foutmarge betekent.
- Je kent de begrippen betrouwbaarheid en onbetrouwbaarheid.
- Je weet wat aselekt is.

Vaardigheid:

- Je kunt een tweezijdig en eenzijdig betrouwbaarheidsinterval voor een gemiddelde bepalen bij een groot steekproefaantal en een zeer grote populatie.
- Je bent in staat om een tweezijdig en eenzijdig betrouwbaarheidsinterval voor een gemiddelde te bepalen bij een groot steekproefaantal en een beperkt eindige populatie.

- Je kunt de minimale steekproefgrootte uitrekenen voor een gemiddelde bij een zeer grote populatie.
- Je bent in staat om de minimale steekproefgrootte uit te rekenen voor een gemiddelde bij een beperkt eindige populatie.

1.1 Case Mercedes

Frank Willemsen loopt stage bij een Mercedesgarage. Een van de opdrachten die hij uitvoert, is een tevredenheidsonderzoek onder de huidige klanten. In overleg met de directeur stelt hij een enquête op. In deze enquête komen tal van aspecten aan bod. Een van deze aspecten is de vraag hoe groot het bestede bedrag is geweest voor de aanschaf van een Mercedes bij de garage. Hierbij gaat het zowel om een gloednieuwe Mercedes als om een occasion. Deze vraag is van belang om een mogelijke relatie te leggen met andere vragen uit de enquête. De betreffende vraag uit de enquête luidt: hoeveel euro hebt u betaald voor de auto?

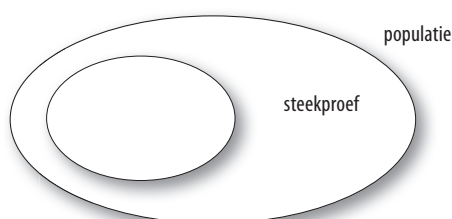


Het totale klantenbestand van de afgelopen vijf jaar bestaat uit 2.000 personen. Frank besluit om een steekproef te nemen.

Hoeveel enquêtes moet Frank afnemen om een statistisch verantwoorde uitspraak te doen over het gemiddelde bedrag dat iemand bij deze garage besteedt aan een Mercedes?

1.2 Betrouwbaarheidsinterval één gemiddelde

Bij marktonderzoeken zoals de in paragraaf 1.1 beschreven case worden in de analysefase tal van grafieken gemaakt en kengetallen berekend. Het meest gehanteerde kengetal bij een interval- of ratiomeetniveau is ongetwijfeld het gemiddelde. Als je alle 2.000 klanten zou ondervragen, dan vind je het *populatie-gemiddelde*. Maar Frank gaat slechts een steekproef verrichten en vindt dan het *steekproefgemiddelde*.



Figuur 1.1 Populatie en steekproef

Ter illustratie: als je alle klanten ondervraagt, zou je een gemiddelde besteding aan een Mercedes kunnen vinden van € 25.000. Maar in de steekproef zou je bijvoorbeeld € 24.800 kunnen vinden. Je ziet dat de twee gemiddelden verschillend zijn.

Bij het gemiddelde bepaal je altijd de standaarddeviatie. Als je alle klanten ondervraagt, dan bereken je de *populatiestandaarddeviatie*. Maar als je een steekproef neemt van de klanten, dan bereken je de *steekproefstandaarddeviatie*. Ook deze twee kengetallen zijn meestal door toeval verschillend.

Om deze kengetallen aan te duiden wordt in de statistiek de volgende notatie gehanteerd: voor een populatiegemiddelde wordt het Griekse symbool μ gebruikt en voor de bijbehorende populatiestandaarddeviatie het symbool σ . Deze kengetallen zijn in de praktijk vaak onbekend, omdat niet de totale populatie wordt ondervraagd vanwege tijd en kosten. Er wordt een steekproef genomen en er worden kengetallen op basis van de steekproef berekend. De corresponderende steekproefkengetallen staan in tabel 1.1.

Tabel 1.1 Populatie- en steekproefkengetallen

Kengetal	Populatie	Steekproef
gemiddelde	μ	$\bar{x}$
standaarddeviatie	σ	s

Om het populatiegemiddelde te bepalen tel je alle getallen bij elkaar op en deel je door het *populatieaantal* N . Bij de berekening van het steekproefgemiddelde tel je alle waarden die je hebt op en deel je door het *steekproefaantal* n . Kenmerkend verschil tussen de populatiestandaarddeviatie en de steekproefstandaarddeviatie is dat je bij de populatiestandaarddeviatie deelt door N en bij de steekproefstandaarddeviatie door $n - 1$.

De berekeningen van alle kengetallen zijn uitgebreid aan bod geweest in hoofdstuk 1 van het *Basisboek statistiek in business*. Voor het vervolg gaan we ervan uit dat je in staat bent om met je (grafische) rekenmachine deze kengetallen te berekenen.

In de praktijk ken je het populatiegemiddelde niet. Brandende vraag is dan hoever je er met de schatting aan de hand van het steekproefgemiddelde naast zit. Het antwoord hierop geeft de *foutmarge*.

Definitie De *foutmarge* geeft aan hoever je er met een zekere waarschijnlijkheid maximaal naast zit met de gegeven schatting.

Bij een gemiddelde wordt deze foutmarge aangeduid met een waarde. Zo zou in de case de foutmarge bijvoorbeeld € 300 kunnen zijn. Als we een steekproefgemiddelde hebben gevonden van € 24.800, dan betekent een foutmarge van € 300 dat een mogelijke ondergrens € 24.500 is en een mogelijke bovengrens € 25.100, waarmee we hopen het populatiegemiddelde te ‘vangen’.

Opvallend gegeven in de definitie van de foutmarge is het woord waarschijnlijkheid. Om dit nader toe te lichten introduceren we het begrip *betrouwbaarheid*.

Definitie De *betrouwbaarheid* geeft aan in hoeveel procent van alle gevallen het populatiekengetal correct geschat wordt door middel van het steekproefkengetal plus of min de foutmarge.

Gangbare waarden voor de betrouwbaarheid zijn 90%, 95% en 99%, waarbij 95% het meest wordt gebruikt.

Tegenover de betrouwbaarheid staat de *onbetrouwbaarheid*. Dit is het complement van de betrouwbaarheid en geeft aan hoe vaak je ernaast zit met je steekproefgemiddelde plus of min de foutmarge. De notatie voor de onbetrouwbaarheid is α . De waarden die veel toegepast worden, zijn 10%, 5% en 1%.

Uitspraken die verricht worden met de betrouwbaarheid zijn alleen van toepassing als de steekproef *aselect* is.

Aselect betekent dat elk element in een populatie een even grote kans heeft om in een steekproef terecht te komen. Definitie

In de case is hieraan voldaan: er is een klantenbestand aanwezig waaruit je willekeurig klanten kunt halen. Iedere klant heeft een even grote kans om in de steekproef terecht te komen.

We zagen dat je voor de gemiddelde besteding aan een Mercedes een ondergrens kunt vinden van € 24.500 en een bovengrens van € 25.100. We koppelen het begrip betrouwbaarheid aan de ondergrens en bovengrens. Een veel gehanteerde uitspraak is: 'Met 95% betrouwbaarheid kun je concluderen dat het populatiegemiddelde omvat wordt in het interval van € 24.500 tot € 25.100.'

Het is dus erg aannemelijk dat je het populatiegemiddelde correct schat, maar er is een kleine kans, namelijk 5%, dat dit niet correct is. Dit wordt veroorzaakt door het feit dat je een steekproef hebt genomen. Als je pech hebt, zitten in de steekproef per toeval veel lage waarden of hoge waarden, zodat het steekproefgemiddelde nogal afwijkt van het populatiegemiddelde. Door de onbetrouwbaarheid laag in te stellen, op bijvoorbeeld 5%, loop je geen groot risico dat jou dit overkomt.

Na alle overwegingen kijken we naar een *betrouwbaarheidsinterval* voor een gemiddelde.

Een betrouwbaarheidsinterval voor een gemiddelde vertelt welke getallen waarschijnlijk zijn voor het populatiegemiddelde. Definitie

We vinden het betrouwbaarheidsinterval voor een gemiddelde door het steekproefgemiddelde te nemen plus of min de foutmarge. De formule van het betrouwbaarheidsinterval voor een gemiddelde luidt:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

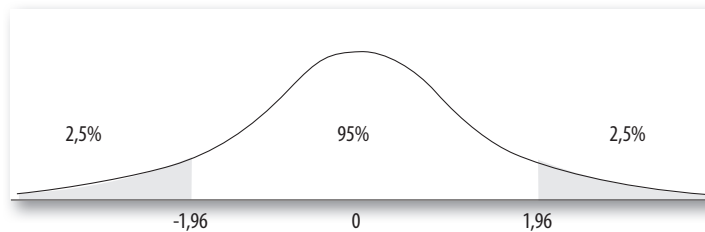
De formule is als volgt samengesteld:

- Allereerst zien we het steekproefgemiddelde $\bar{x}$.
- Daarna komt een z-waarde. Deze hebben we eerder gezien in het basisboek en is een waarde uit de standaardnormale verdeling.
- Dan volgen de index $\alpha/2$, de steekproefstandaarddeviatie s en de wortel van het steekproefaantal n .

Na het verkrijgen van de steekproefgegevens kun je het steekproefgemiddelde, de steekproefstandaarddeviatie en het steekproefaantal bepalen. Blijft over de

vraag hoe je aan de z-waarde komt en wat je moet doen met de index $\alpha/2$. Veel gebruikt worden 95%-betrouwbaarheidsintervallen voor een gemiddelde. Als tussenberekening van deze intervallen hebben we een z-waarde nodig:

- De betrouwbaarheid is 95%, dus de onbetrouwbaarheid $\alpha = 5\%$.
- Deze moet je delen door 2, dan is $\alpha \div 2 = 5\% \div 2 = 2,5\%$.
- Koppel deze $2,5\% = 0,025$ in de standaardnormale verdeling aan de rechteroppervlakte in figuur 1.2.



Figuur 1.2 Tabelwaarde z-verdeling

- Uitgaande van de standaardnormale verdeling kijk je in bijlage 1 bij een oppervlakte van 0,025 en zoek je hierbij de z-waarde op. Terugzoeken levert op $z = 1,96$. Deze waarde gebruiken we om een betrouwbaarheidsinterval te construeren.

Als voorbeeld kijken we naar een onderzoek onder 500 personen. In dit onderzoek is onder andere gevraagd naar de wijnconsumptie per kwartaal. Het blijkt dat het steekproefgemiddelde van wijnconsumptie in een kwartaal 12 liter is met een steekproefstandaarddeviatie van 2 liter. We maken een 95%-betrouwbaarheidsinterval voor de gemiddelde wijnconsumptie per kwartaal.

We verzamelen alle informatie. De formule luidt:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

- Het steekproefgemiddelde is $\bar{x} = 12$ liter wijn per kwartaal.
- De z-waarde, corresponderend met 95% betrouwbaarheid, is $z = 1,96$.
- De steekproefstandaarddeviatie is $s = 2$ liter wijn per kwartaal.
- Het steekproefaantal is $n = 500$ personen.

Als we al deze waarden invullen vinden we:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 12 \pm 1,96 \cdot \frac{2}{\sqrt{500}} = 12 \pm 0,18$$

Het 95%-betrouwbaarheidsinterval loopt van 11,82 liter tot aan 12,18 liter. De gemiddelde wijnconsumptie per kwartaal ligt met 95% betrouwbaarheid tussen 11,82 liter en 12,18 liter.

Het gevonden betrouwbaarheidsinterval is een *tweezijdig betrouwbaarheidsinterval*. Er is sprake van een bovengrens en een ondergrens.

Opdracht 1

Bepaal het 90%-betrouwbaarheidsinterval voor de gemiddelde wijnconsumptie per kwartaal.

Er zijn ook *eenzijdige betrouwbaarheidsintervallen*, namelijk rechtseenzijdige en linkseenzijdige betrouwbaarheidsintervallen. Bij een *linkseenzijdig betrouwbaarheidsinterval* is er alleen een ondergrens, maar geen bovengrens. De formule hiervan luidt:

$$\bar{x} - z_{\alpha} \cdot \frac{s}{\sqrt{n}}$$

Kenmerkend aspect is dat bij een eenzijdig betrouwbaarheidsinterval de onbetrouwbaarheid niet door twee wordt gedeeld. Dit gegeven zorgt voor een andere z-waarde, namelijk $z = 1,645$. Als je een 95% linkseenzijdig betrouwbaarheidsinterval bepaalt in het voorbeeld van de gemiddelde wijnconsumptie per jaar, dan vind je:

$$\bar{x} - z_{\alpha} \cdot \frac{s}{\sqrt{n}} = 12 - 1,645 \cdot \frac{2}{\sqrt{500}} = 12 - 0,15 = 11,85$$

Met 95% betrouwbaarheid kun je beweren dat de gemiddelde wijnconsumptie per kwartaal minimaal 11,85 liter is.

Op soortgelijke wijze vind je een rechtseenzijdig betrouwbaarheidsinterval. De formule hiervan is:

$$\bar{x} + z_{\alpha} \cdot \frac{s}{\sqrt{n}}$$

Toegepast op het voorbeeld met 95% betrouwbaarheid vind je:

$$\bar{x} + z_{\alpha} \cdot \frac{s}{\sqrt{n}} = 12 + 1,645 \cdot \frac{2}{\sqrt{500}} = 12 + 0,15 = 12,15$$

Nu kun je zeggen dat met 95% betrouwbaarheid de gemiddelde wijnconsumptie per kwartaal maximaal 12,15 liter per persoon is.

De formules voor het betrouwbaarheidsinterval zijn van toepassing op het moment dat de steekproefgrootte minimaal $n = 200$ is. We noemen een steekproef vanaf $n = 200$ een grote steekproef. Hiertegenover staat een steekproef met een n die kleiner is dan 200, die we een kleine steekproef noemen. Verder dient het populatieaantal bij de toepassing van de formule zeer groot te zijn (zie paragraaf 1.4).

1.3 Minimale steekproefgrootte één gemiddelde

We kijken eens goed naar de formule voor een betrouwbaarheidsinterval van een gemiddelde en concentreren ons op de foutmarge:

$$z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

We zien in de formule dat hoe groter de n wordt, hoe kleiner de foutmarge. Klaarblijkelijk bestaat er een relatie tussen de foutmarge en de n . Je zult in overleg moeten gaan met de opdrachtgever en een keuze moeten maken voor de foutmarge en de betrouwbaarheid als je de *minimale steekproefgrootte* wilt bepalen.

Definitie De minimale steekproefgrootte is de minimale waarde van de steekproefgrootte waarbij voldaan is aan zorgvuldig geformuleerde statistische eisen.

De betreffende statistische eisen volgen uit de overweging dat de foutmarge niet groter kan zijn dan een weloverwogen bovengrens F . Er volgt dat

$$z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \leq F$$

en hieruit is wiskundig de formule van de minimale steekproefgrootte af te leiden. Deze formule blijkt dan gelijk te zijn aan:

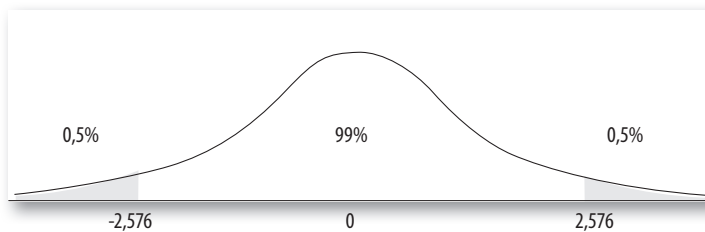
$$n \geq \frac{z_{\alpha/2}^2 \cdot s^2}{F^2}$$

We lichten een en ander toe aan de hand van een voorbeeld. Bij een grote verzekeringsmaatschappij vraagt men zich op de schadeafdeling af hoelang men gemiddeld bezig is met het behandelen van een schadeformulier. Om hierachter te komen zou eigenlijk een enorm archief moeten worden doorgespit. Dit is ondoenlijk en men besluit tot het nemen van een steekproef. De vraag is nu hoeveel formulieren er minimaal moeten worden onderzocht.

We nemen de formule van de minimale steekproefgrootte van een gemiddelde bij de hand:

$$n \geq \frac{z_{\alpha/2}^2 \cdot s^2}{F^2}$$

- De verzekeringsmaatschappij wil een hoge mate van betrouwbaarheid en stelt deze op 99%. Bij een 99%-betrouwbaarheid hoort een z-waarde van 2,58. Immers, de onbetrouwbaarheid is 1%. Deze deel je door 2 en wordt 0,5%. Terugzoeken in bijlage 1 levert bij een rechteoppervlakte van 0,5% = 0,005 de z-waarde 2,58 op. Eigenlijk is de waarde $z = 2,576$, die je vindt als je de z-waarde met een grafische rekenmachine bepaalt.



Figuur 1.3 Tabelwaarde z-verdeling

- De tweede waarde die je dient in te vullen, is de steekproefstandaarddeviatie s . Voorafgaand aan het onderzoek is deze uiteraard onbekend. Een manier om deze standaarddeviatie globaal te bepalen, is door een klein aantal formulieren al door te werken. Op basis van de verkregen tijden kun je dan een steekproefstandaarddeviatie geven over deze kleine aantallen. Stel, je vindt een standaarddeviatie die gelijk is aan 7 minuten. Dus $s = 7$ minuten.
- Tot slot moet je met de opdrachtgever de maximale foutmarge F vaststellen. De gemiddelde behandelingstijd moet je vrij precies bepalen. Reden hiervan is dat je deze gemiddelde tijd moet vermenigvuldigen met de te verwachten aantallen schadeformulieren op jaarbasis. Als de opdrachtgever een onnauwkeurige schatting krijgt voor de behandeling van een gemiddeld schadeformulier, dan kan dat ernstige consequenties hebben voor de tijdsbesteding van heel veel formulieren. F moet daarom klein zijn en wordt vastgesteld op 0,5 minuut.

Samengevat:

- de z-waarde $z = 2,576$;
- de steekproefstandaarddeviatie $s = 7$ minuten;
- de foutmarge $F = 0,5$ minuut.

Deze informatie vullen we in de formule in:

$$n \geq \frac{z_{\alpha/2}^2 \cdot s^2}{F^2} = \frac{2,576^2 \cdot 7^2}{0,5^2} = 1.300,61$$

Er moeten in totaliteit 1.301 schadeformulieren worden onderzocht.

Opdracht 2

Reken eens uit hoe groot het minimumaantal schadeformulieren dient te zijn als gewerkt wordt met 95% betrouwbaarheid, een steekproefstandaarddeviatie van 10 minuten en een foutmarge van 0,5 minuut.

1.4 Betrouwbaarheidsinterval één gemiddelde beperkt eindige populatie

De tot nu toe gepresenteerde formules zijn van toepassing op een (zeer) grote populatie. Bij marktonderzoeken komt het echter regelmatig voor dat je een niet zo grote populatie moet onderzoeken. Je noemt een niet zo grote populatie voor de eenvoud een *beperkt eindige populatie*. Als voorbeeld kun je denken aan ondervraging van bedrijven in een bepaalde branche in een regio. Op dat moment veranderen de formules. Je moet dan een *eindigheidscorrectie* toepassen.

De foutmarge in de behandelde betrouwbaarheidsintervallen moet je dan vermenigvuldigen met:

$$\sqrt{\frac{N-n}{N-1}}$$

De formule van het tweezijdig betrouwbaarheidsinterval voor een gemiddelde wordt dan:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

In het algemeen pas je deze correctiefactor toe als het steekproefaantal groter is dan 10% van het populatieaantal. Verder moet N niet heel erg klein zijn. Is dit wel het geval, dan kun je beter een integraal onderzoek uitvoeren. Vaak wordt als vuistregel een N van 500 in dit verband als grens genoemd.

Aan de hand van een voorbeeld kijken we naar de toepassing van het betrouwbaarheidsinterval. Het hardware- en softwarebureau Dutch Compusteel heeft in totaliteit 1.000 klanten. De directie besluit een onderzoek onder de klanten

te laten uitvoeren. Al gauw blijkt dat het onderzoek onder alle klanten zeer veel tijd zal gaan kosten. De bedrijfsonderzoeker stelt voor om een steekproef te nemen. De directie bestudeert de ingediende onderzoeksopzet en besluit om hiertoe toestemming te geven. Een van de aspecten die onderzocht moeten worden, is de hoeveelheid tijd die klanten per week besteden aan het gebruik van het financieel softwarepakket geleverd door Dutch Compusteel.

Uit de steekproef van 300 klanten komt het steekproefgemiddelde van 6 uur per week en een steekproefstandaarddeviatie van 3 uur per week. In de analyse maakt de onderzoeker gebruik van een 95%-betrouwbaarheidsinterval voor het gemiddelde.

Samengevat zien we:

- het steekproefgemiddeld $\bar{x} = 6$ uur per week;
- de steekproefstandaarddeviatie $s = 3$ uur per week;
- $N = 1.000$ klanten;
- het steekproefaantal $n = 300$;
- de onbetrouwbaarheid $\alpha = 5\%$ met bijbehorende z -waarde $z = 1,96$.

We vullen dit alles in de betreffende formule in:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = 6 \pm 1,96 \cdot \frac{3}{\sqrt{300}} \cdot \sqrt{\frac{1.000-300}{1.000-1}} = 6 \pm 0,28$$

Dat betekent dat de gemiddelde besteding per week varieert van 5,72 uren tot 6,28 uren. Vertaald naar alle klanten betekent dit een besteding tussen de $1.000 \times 5,72 = 5.720$ uren per week en $1.000 \times 6,28 = 6.280$ uren per week. Deze uren worden onder andere gerelateerd aan het aantal uren dat Dutch Compusteel kwijt is aan de telefonische helpdesk voor dit pakket per week. Zo kan het bedrijf anticiperen op veranderingen in het klantenbestand in relatie met het klantenbeheer.

Opdracht 3

Bepaal het 95%-betrouwbaarheidsinterval voor de gemiddelde besteding per week voor Dutch Compusteel als blijkt dat het totaal aantal klanten 1.200 is en het steekproefaantal 350.

De formules voor het rechtseenzijdig en linkseenzijdig betrouwbaarheidsinterval bij een beperkt eindige populatie bepaal je op dezelfde manier als de formule voor het tweezijdig betrouwbaarheidsinterval, namelijk door de foutmarge te vermenigvuldigen met de correctiefactor.