



Basisboek Statistiek met Excel

**Handleiding voor het
verwerken en analyseren
van en rapporteren over
(onderzoeks)gegevens**



Noordhoff

Ben Baarda & René van Vianen

3^e druk

Basisboek Statistiek met Excel

Handleiding voor het verwerken en
analyseren van en rapporteren over
(onderzoeks)gegevens

Ben Baarda

René van Vianen

Derde druk

Noordhoff Uitgevers Groningen / Utrecht

Ontwerp omslag: Shootmedia, Groningen

Omslagillustratie: Dreamstime

Eventuele op- en aanmerkingen over deze of andere uitgaven kunt u richten aan:
Noordhoff Uitgevers bv, Afdeling Hoger Onderwijs, Antwoordnummer 13, 9700 VB
Groningen of via het contactformulier op www.mijnnoordhoff.nl.

De informatie in deze uitgave is uitsluitend bedoeld als algemene informatie. Aan
deze informatie kunt u geen rechten of aansprakelijkheid van de auteur(s), redactie of
uitgever ontlenen.

0 / 19



© 2019 Noordhoff Uitgevers bv, Groningen/Utrecht, Nederland.

Deze uitgave is beschermd op grond van het auteursrecht. Wanneer u (her)gebruik wilt maken van de informatie in deze uitgave, dient u vooraf schriftelijke toestemming te verkrijgen van Noordhoff Uitgevers bv. Meer informatie over collectieve regelingen voor het onderwijs is te vinden op www.onderwijsauteursrecht.nl.

This publication is protected by copyright. Prior written permission of Noordhoff Uitgevers bv is required to (re)use the information in this publication.

ISBN (ebook) 978-90-01-57534-2

ISBN 978-90-01-57533-5

NUR 916

Woord vooraf bij de derde druk

In dit boek *Basisboek Statistiek met Excel* vind je aanwijzingen voor het verwerken en analyseren van onderzoeksgegevens met behulp van Excel. Doordat Excel standaard op vrijwel alle computers staat, kun je bijna overal je databestanden analyseren. Je hebt daar dus geen dure statistische software voor nodig, tenzij je zeer geavanceerde statistische technieken wilt gebruiken. Als je de invoegtoepassing 'Analysis ToolPak' van Excel installeert (zie de uitleg in paragraaf 1.6), beschik je over de meest gangbare statistische analyseprocedures; ook over multivariate technieken als multivariate variantieanalyse.

In dit boek gebruiken wij de versie van 2019. Wil je meer weten over Excel, kijk dan eens op de Wikipedia-pagina over Excel (http://en.wikipedia.org/wiki/Microsoft_Excel). Daar vind je niet alleen info over wat er mogelijk is met Excel, maar ook de geschiedenis.

Met dit boek willen we je leren al doende vertrouwd te raken met de statistische procedures in Excel, en ook met statistiek. We nodigen je uit om aan de hand van een databestand over de relatie tussen geld, geluk en gezondheid aan de slag te gaan.

Na een algemene inleiding en kennismaking met de basisregels van Excel kun je direct beginnen met het invoeren en analyseren van gegevens in de computer. Om verantwoord statistische procedures met Excel te kunnen uitvoeren is niet alleen kennis van Excel, maar ook statistische kennis nodig. We bespreken in de eerste hoofdstukken statistische basisbegrippen als meetniveau, normaalverdeling, kans, significantie, een- en tweezijdige toetsing en effectgrootte. In de laatste hoofdstukken bespreken we technieken die je gebruikt om gegevens te beschrijven, om verschillen te toetsen en om samenhang na te gaan.

We leggen zo veel mogelijk in woorden en met voorbeelden de essentie uit van iedere techniek die we bespreken, zonder daar diep wiskundig op in te gaan. Ook geven we aan wanneer je de techniek wel kunt gebruiken en wanneer niet. Uiteraard wordt uitgelegd hoe je die techniek in Excel uitvoert.

Soms maken we ook uitstapjes naar internet, omdat sommige statistische procedures makkelijker via interactieve internetprogramma's zijn uit te voeren. Wel bereiden we het databestand altijd voor in Excel. Via schermafdrucken en de instructies die we daarbij geven, laten wij letterlijk zien wat je moet doen. Voor iedere besproken techniek laten we een voorbeeld van de uitvoer zien. Aan de hand hiervan leggen we uit hoe je het moet lezen en wat het betekent. Tevens geven we aan hoe je over het resultaat kunt rapporteren.



De databestanden waarop we de analyses uitvoeren, zijn te vinden op internet via www.basisboekstatistiekmetexcel.noordhoff.nl. Deze databestanden zijn nieuw, er is onder andere een gezondheidsvariabele toegevoegd. De bestanden wijken sterk af van de bestanden behorende bij de eerdere uitgaven van dit boek. Ga rechtstreeks naar de website via de volgende QR-code:

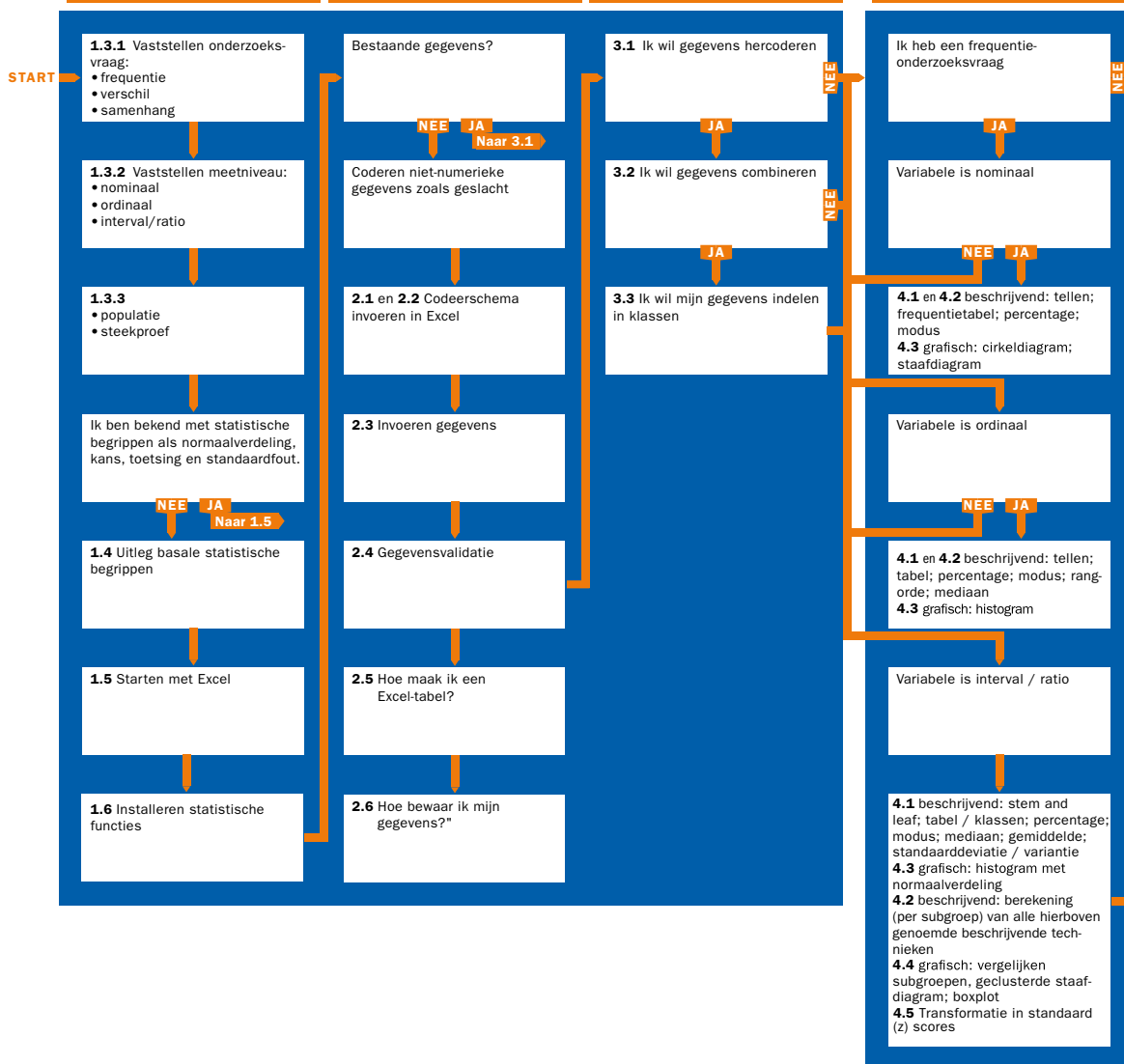


Verder wordt in deze nieuwe editie de Excel-tabelfunctie besproken. Dit maakt niet alleen het databestand overzichtelijker, maar je kunt er ook snel wijzigingen, zoals het combineren van gegevens, mee doorvoeren en het geeft direct statistische maten als het gemiddelde en de standaarddeviatie. Ook worden in twee bijlagen tips gegeven over Excelgebruik in het algemeen, zoals de betekenis van de verschillende cursoren.

Het *Basisboek Statistiek met Excel* komt optimaal tot zijn recht in samenhang met het *Basisboek Methoden en Technieken* of *Dit is onderzoek!* Aanvullend kan ook het *Statistiekkwartetspel* gebruikt worden, waarin op een speelse wijze de belangrijkste statistische begrippen en toetsen worden uitgelegd. *Basisboek Statistiek met Excel* is uiteraard ook los daarvan goed te gebruiken.

Najaar 2019
Ben Baarda
René van Vianen

1 Voorbereiding op Excel	2 Gegevens invoeren en bewaren	3 Het veranderen en combineren van gegevens	4 Analyse bij frequentie-onderzoeksvraag
------------------------------------	--	---	--



5 Analyse bij verschil- onderzoeksvraag

Ik heb een verschil-
onderzoeksvraag

JA

Splitsingsvariabele is nominaal;
testvariabele is nominaal

NEE JA

Kruistabel
Populatie (4.1)
• Steekproef: chikwadraat (5.1)

Splitsingsvariabele is nominaal;
testvariabele is ordinaal

NEE JA

Verskil in gemiddelde
rangorde:
• Populatie (4.4)
• Steekproef gekoppeld:
– twee: Wilcoxon (5.2.3)
– meer: Friedman (5.2.4)
• Steekproef niet-gekoppeld:
– twee: Mann-Whitney (5.2.1)
– meer: Kruskal-Wallis (5.2.2)

Splitsingsvariabele is nominaal;
testvariabele is interval/ratio

5.2.1 Verskil in gemiddelde:
• Populatie (4.4)
• Steekproef gekoppeld:
– twee: t-test (5.3.2)
– meer: MANOVA (niet in
dit boek)
• Steekproef niet-gekoppeld:
– twee: t-test (5.3.1)
– meer: variantieanalyse
(5.3.3)

6 Analyse bij samenhang- onderzoeksvraag

Ik heb een samenhang-
onderzoeksvraag

JA

Variabelen zijn nominaal
(ga uit van de variabelen
met het laagste meetniveau)

NEE JA

6.1 Cramér's V
Bij steekproef: significantie
vaststellen met chikwadraat
(5.1)

Variabelen zijn ordinaal
(ga uit van de variabelen
met het laagste meetniveau)

NEE JA

6.2 Spearman
Bij steekproef: significantie
vaststellen (6.3)

Variabelen zijn interval / ratio
(ga uit van de variabelen
met het laagste meetniveau)

Twee variabelen
• beschrijvend:
– Pearson (6.3)
bij steekproef significantie
vaststellen (6.3.2)
– enkelvoudige regressie
(6.3.3)
• grafisch (6.3.1):
– scattergram
– regressielijn

Meer dan twee variabelen
• voorspellend:
– multipele correlatie /
regressie (6.4)

Verslag maken

Bijlage

Berekenen van homogeniteit

I Overzicht van de verschillende
cursorvormen en de
gebruiksmogelijkheden

JA

II Handige tips en formules in Excel

III Berekenen van homogeniteit

Inhoud

Effectief studeren 12

1 Hoe bereid ik mij voor op statistiek met Excel? 15

- 1.1 Fasen van de onderzoekscyclus 16
 - 1.2 Geld, geluk en gezondheid; toelichting op de gebruikte casus 17
 - 1.3 Hoe analyseer ik mijn data? Een gebruikswijzer 18
 - 1.4 Enkele algemene statistische begrippen 22
 - 1.5 Hoe werkt Excel? 25
 - 1.6 Installeren van de statistische functies van Excel 27
- [Samenvatting 29](#)

2 Hoe breng ik mijn gegevens in de computer? 31

- 2.1 Hoe maak ik een codeerschema of codeboek? 32
 - 2.2 Hoe voer ik mijn codeboek in Excel in? 34
 - 2.3 Hoe voer ik mijn gegevens in? De datamatrix 35
 - 2.4 Hoe valideer ik mijn gegevens? 39
 - 2.5 Hoe zet ik mijn datamatrix om in een Excel-tabel? 40
 - 2.6 Hoe bewaar ik mijn ingevoerde gegevens? 44
- [Samenvatting 46](#)

3 Hoe verander en combineer ik gegevens? 49

- 3.1 Hoe moet ik mijn gegevens hercoderen? 50
 - 3.2 Hoe kan ik mijn gegevens combineren? 55
 - 3.3 Hoe kan ik een variabele indelen in klassen? 60
- [Samenvatting 62](#)

4 Hoe analyseer ik mijn gegevens bij een frequentieonderzoeksvraag? 65

- 4.1 Maken van een frequentieverdeling in de vorm van een tabel 66
 - 4.2 Hoe bereken ik samenvattende statistische maten? 74
 - 4.3 Hoe geef ik frequenties grafisch weer? 82
 - 4.4 Hoe vergelijk ik frequenties van (sub)groepen? 87
 - 4.5 Hoe maak ik scores vergelijkbaar? 94
- [Samenvatting 96](#)

5 Hoe analyseer ik mijn gegevens bij een verschilonderzoeksvraag? 99

- 5.1 Verschil bij een nominale test- en splitsingsvariabele? 100
 - 5.2 Verschilvragen bij ordinale testvariabelen en nominale splitsingsvariabelen 109
 - 5.3 Verschilvraag bij interval- en ratiotestvariabele en nominale splitsingsvariabele? 126
- [Samenvatting 141](#)

- 6 Hoe analyseer ik mijn gegevens bij een samenhangonderzoeksvraag? 143**
- 6.1 Samenhang tussen twee nominale variabelen? Cramer's V 144
- 6.2 Samenhang bij ordinale variabelen: Spearman's rangcorrelatie 146
- 6.3 Samenhang bij interval- of ratiovariabelen 151
- 6.4 Samenhang van twee of meer variabelen die gemeten zijn op interval- of rationiveau: multipele regressie 162
[Samenvatting 166](#)

Bijlage I Overzicht van de verschillende cursorvormen en de gebruiksmogelijkheden 167

Bijlage II Handige tips en formules in Excel 169

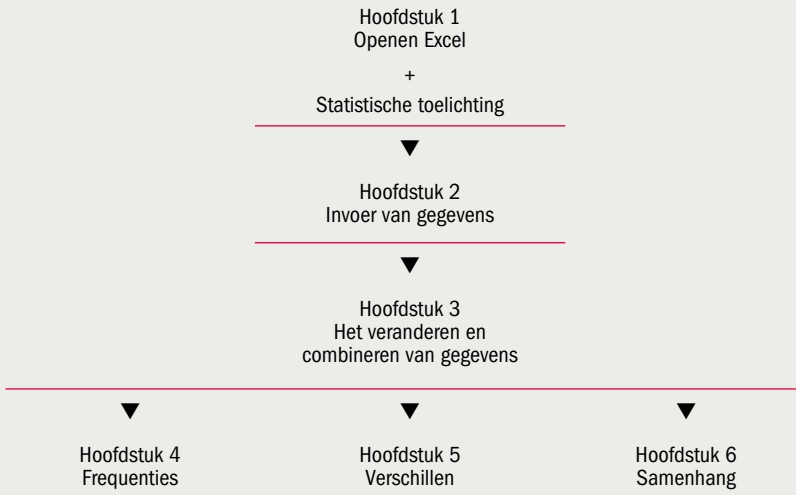
Bijlage III Berekenen van homogeniteit 171

Over de auteurs 178

Register 179

Effectief studeren

Opbouw boek



Bijlage I Overzicht van verschillende cursorvormen en de gebruiksmogelijkheden

Bijlage II Handige tips en formules in Excel

Bijlage III Berekenen van homogeniteit

Opbouw hoofdstuk

Voorkennis

We gaan ervan uit dat:

- je bekend bent met de verschillende meetniveaus (1.3.2)
- je Excel kunt starten (1.5)
- je een databestand kunt oproepen (2.5)
- je weet wat continue en discontinue of discrete variabelen zijn (1.3.2)
- je weet wat een normaalverdeling is (1.4)

Dit zijn de vragen die in het hoofdstuk worden behandeld.

Dit is wat je moet weten als je aan het hoofdstuk begint.



Hoe bereid ik mij voor op statistiek met Excel?

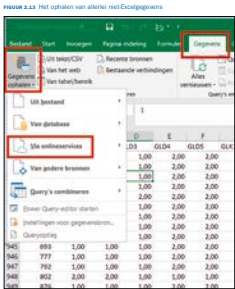
- 1.1 Fasen van de onderzoekscyclus
- 1.2 Geldt gelijk en ongelijkheids toelichting op de gebruikte cases
- 1.3 Hoe analyseer ik mijn data? Een analysekwestie
- 1.4 Fall of algemeen statistische begrippen
- 1.5 Hoe werkt Excel?
- 1.6 Initialiseren van de statistische functies van Excel

- Frequentie 19
- Variabel 19
- Samenhang 19
- Nominaal meetniveau 19
- Ordinaal meetniveau 19
- Intervalniveau 20
- Nulpunt 20
- Raaienniveau 20
- Continue variabelen 21
- Discrete variabelen 21
- Bevatende statistiek 21
- Inductieve/inductieve statistiek 21
- Nominaalmeetniveau 21
- Standaardfout 21
- Kans 21
- Significantie 24
- Een of tweevoudig toetsen 24
- Referentie 24
- Efficiëntie 25
- Vrijheidsgraden 25
- Templeta 26

Dit zijn de begrippen die in een hoofdstuk worden behandeld.

Dit is een voorbeeld van een schermafdruk om uit te leggen hoe je Excel moet gebruiken.

In de nieuwste versie van Excel is het ook mogelijk een met de functie 'Gegevens opslaan' en 'Transformeren bestaande gegevens' om te gebruiken in artikelen die je op internet vindt, of Facebook-gegevens te importeren (figuur 2.13). Wil je hier meer informatie over kijk dan op <https://support.office.com>



Margewoorden om de belangrijkste begrippen te benadrukken.

Dit is een voorbeeld van een samenvatting die aan het eind van ieder hoofdstuk staat.

Opbouw paragraaf

- Wat is ...?
- Wanneer gebruik ik ...?
- Hoe voer ik ... uit?
- Hoe lees ik de uitvoer van ...?
- Hoe rapporteer ik over ...?

HOE VERANDER EN COMBINER IK GEGEVENS? 57

3.2 Hoe kan ik mijn gegevens combineren?

In sommige gevallen is het handig een nieuwe variabele te maken door gegevens te combineren. Het gaat hier om de kans dat antwoorden op de vijf vragen over geld te combineren tot een totaalvoor. Om dit te kunnen doen, moeten de antwoorden op deze vragen wel in dezelfde richting gepoold zijn. De antwoorden op de volgende vragen 1 tot en met 3 moeten dan omgepoold zijn.

Een voorwaarde voor het combineren van gegevens is dat de vragen wel min of meer hetzelfde moeten. Dit kan je statistisch controleren door gebruik te maken van een t-testanalyse. Hoe dit moet, staat ingevoerd in bijlage III. Uit de analyse in bijlage III blijkt dat je niet alle vragen over geld en geldsomma kan opvoelen. Het blijkt dat sommige vragen niet hetzelfde moeten. Voor een t-testanalyse op grote die representatie dat een beperkt aantal behoudt een de respondenten moet vrijant heeft, voldoet het met de variabelen R_GLD1, GLD1 en GLD2 op je dit niet te combineren.

Voor een betrouwbaar t-testanalyse voor de variabele geldt volstaat het opstellen van de variabelen G.L.K.1, G.L.K.2 en R_G.L.K.1. Een hogere score op getal staat dan voor meer geld.

Opmerkt is het verstandig om in Excel overzicht te houden. Dit heeft ook met het werken met de formules in Excel te maken. We zetten met Knippen en Plakken specifiek de antwoorden van de variabelen die we gebruiken, dus R_GLD1, G.L.D.1, G.L.D.2, G.L.K.1, G.L.K.2 en R_G.L.K.1, op een nieuw werkblad (figuur 3.5). Kopieer vervolgens 'het' 'R_G.L.K.1' en 'G.L.K.1'.

FIGUUR 3.5 Het herschikken van de variabelen



© Noordhoff Uitgevers bv

Somervatting

- Het prepareren en analyseren van onderzoekgegevens is een van de laatste fasen in de onderzoekscyclus. De etafelaf, de rapportage, is hier voor een groot deel op gebaseerd.
- In paragraaf 1.2 lichten wij de cases toe, die wij gebruiken om het uitvoeren van statistische analyse uit te leggen (cases 1.1).
- In paragraaf 1.3 geven wij aan dat je voordat je gaat analyseren, een aantal vragen moet beantwoorden. Gaat het bijvoorbeeld in je onderzoeksvraag om een frequentie, een verschil, of om een samenhang? Wat is het meetniveau van je variabelen? Gaat het om een meetniveau of om een populatie? Afhankelijk van het antwoord op deze vragen kan je meer behulp van het bibliotek een keuze maken voor een statistische analysetechniek.
- In paragraaf 1.4 leggen wij enkele gangbare statistische begrippen uit. Het gaat daarbij om begrippen als standaardfout, betrouwbaarheid, kans, significantie, een- of tweevoudig toetsen, efficiëntie en vrijheidsgraden.
- In paragraaf 1.5 geven wij in het kort aan hoe Excel werkt.
- Voor de statistische analyse is het belangrijk dat je de statistische functies in Excel initialiseert. Hoe je dat doet, wordt uitgelegd in paragraaf 1.6.

© Noordhoff Uitgevers bv



Voorkennis

Er wordt geen specifieke voorkennis vereist. Wel wordt aanbevolen om naast dit boek het *Basisboek Methoden en Technieken* (zesde herziene druk, 2017) of *Dit is onderzoek!* (derde herziene druk, 2019) te gebruiken.

1

Hoe bereid ik mij voor op statistiek met Excel?

- 1.1 Fasen van de onderzoekscyclus**
- 1.2 Geld, geluk en gezondheid; toelichting op de gebruikte casus**
- 1.3 Hoe analyseer ik mijn data? Een gebruikswijzer**
- 1.4 Enkele algemene statistische begrippen**
- 1.5 Hoe werkt Excel?**
- 1.6 Installeren van de statistische functies van Excel**

Frequentie 19	Inductieve/inferentiële statistiek 21
Verschil 19	Normaalverdeling 22
Samenhang 19	Standaardfout 23
Nominaal meetniveau 19	Kans 23
Ordinaal meetniveau 19	Significantie 24
Intervalniveau 20	Een- of tweezijdig toetsen 24
Nulpunt 20	Relevantie 24
Rationiveau 20	Effectgrootte 25
Continue variabelen 21	Vrijheidsgraden 25
Discrete variabelen 21	Templates 26
Beschrijvende statistiek 21	

1.1 Fasen van de onderzoekscyclus

Als je onderzoek doet, is het kiezen van de juiste statistische techniek om de verzamelde gegevens te analyseren een belangrijke schakel in de lange keten van beslissingen die je moet nemen. Het uiteindelijke doel is het beantwoorden van de onderzoeksvraag of -vragen.

Om de plaats van de data-analyse in de onderzoekscyclus als geheel voor ogen te houden, zetten we de gebruikelijke fasen van een onderzoek hier op een rij. Iedere fase van de onderzoekscyclus is in de vorm van een vraag opgenomen:

- 1 Wat is/zijn mijn onderzoeksvra(a)g(en) en wat is de doelstelling van mijn onderzoek?
- 2 Hoe zoek ik informatie (onder meer literatuurstudie)?
- 3 Wat voor type onderzoek ga ik doen?
- 4 Hoe ziet mijn onderzoeksontwerp eruit?
- 5 Betrek ik de populatie in mijn onderzoek, of trek ik een steekproef?
- 6 Welke dataverzamelmethode ga ik gebruiken?
- 7 *Hoe prepareer ik mijn data voor de analyse?*
- 8 *Hoe analyseer ik mijn data?*
- 9 *Hoe rapporteer en evalueer ik mijn onderzoek?*

In dit boek staan de fasen 7 en 8 en een deel van 9 centraal: preparatie, analyse en beschrijving van de onderzoeksgegevens die met Excel worden geanalyseerd. Aan de hand van een onderzoek naar de relatie tussen geld, geluk en gezondheid (zie casus 1.1), beantwoorden we stap voor stap de volgende vragen:

- Hoe moet je onderzoeksgegevens prepareren, zodat je ze kunt invoeren (hoofdstuk 2)?
- Hoe kun je gegevens met Excel aanpassen en veranderen? Voordat je aan de analyse begint, moet je vaak eerst van bepaalde gegevens de waarden ompolen of hercoderen, of de waarden van gegevens combineren tot een nieuwe score (hoofdstuk 3).
- Hoe kies je de juiste analysetechniek? Om de juiste analysetechniek te kunnen kiezen moet je eerst vaststellen wat het karakter van je onderzoeksvraag is (subparagraaf 1.3.1). Gaat het om frequenties (hoofdstuk 4), verschillen (hoofdstuk 5) of om samenhangen (hoofdstuk 6)? Vervolgens moet je nagaan wat het meetniveau van je gegevens is (subparagraaf 1.3.2). Tot slot moet je vaststellen of het in je onderzoek om een steekproef of een populatie gaat. Je kunt dan aan de hand van het blokschema vaststellen welke analysetechniek het beste bij jouw onderzoeksvraag past. Je vindt het schema voorin dit boek.
- Hoe moet je de analyse met Excel uitvoeren en hoe moet je de resultaten interpreteren? Dit geven we telkens aan bij iedere statistische techniek die we behandelen. We bespreken hoe je verslag kunt doen van de resultaten.

In dit hoofdstuk behandelen we verder een aantal belangrijke statistische begrippen als normaalverdeling, significantie, effectgrootte, kans en standaardfout (paragraaf 1.4). In paragraaf 1.5 leggen we uit hoe je Excel kunt starten en in de laatste paragraaf (1.6) tonen we je hoe je de 'Analysis Tool-Pak' invoegtoepassing installeert. Hierdoor kun je allerlei statistische analysetechnieken, zoals de variantieanalyse, gebruiken.

1.2 Geld, geluk en gezondheid; toelichting op de gebruikte casus

Voordat we ingaan op het gebruik van Excel, introduceren we eerst het fictieve voorbeeldonderzoek Geld, geluk en gezondheid (zie casus 1.1). De vragen uit casus 1.1 zijn voorgelegd aan een representatieve steekproef van 490 Nederlandse vrouwen en 500 Nederlandse mannen van 25 tot en met 60 jaar. Er is bewust voor deze leeftijdsgrenzen gekozen. Veel jonge mensen studeren nog en hebben daardoor geen vast inkomen. Bij mensen boven de 60 is er vaak al sprake van (gedeeltelijke) uittreding uit het arbeidsproces, waardoor zij in een andere financiële situatie komen. De data van dit onderzoek vind je op de website www.basisboekstatistiekmetexcel.noordhoff.nl onder de naam 'data_geld_geluk_gezondheid'.



CASUS 1.1

Voorbeeldonderzoek Geld, geluk en gezondheid

Maakt geld gelukkig en gezond?

Een onderzoeker wil weten of er een verband bestaat tussen geld, geluk en gezondheid. Hij vraagt zich af of geld gelukkig en gezond maakt. Zijn centrale onderzoeksvraag luidt dan ook: 'Is er een positief verband tussen de mate waarin iemand over geld beschikt en de mate waarin hij/zij zich gelukkig en gezond voelt?' Het begrip 'geld' heeft de onderzoeker als volgt gedefinieerd: geld is de hoeveelheid financiële middelen waarover iemand kan beschikken. Deze definitie is met opzet ruim gekozen. Hierdoor worden niet alleen het inkomen en het vermogen, maar ook andere financiële bronnen waar iemand over kan beschikken in het onderzoek betrokken. Geluk wordt als volgt gedefinieerd: geluk is de mate waarin iemand tevreden is met het leven dat hij/zij leidt. Gezondheid wordt gemeten met de vraag hoe gezond iemand zichzelf voelt. De onderzoeker heeft deze drie begrippen geoperationaliseerd in een vragenlijst. Zowel voor het meten van het begrip geld, als voor het begrip geluk heeft hij vijf uitspraken of items gemaakt, gezondheid wordt gemeten met één vraag.

Geld is gemeten met de items:

1	Ik ben in het bezit van een auto	ja/nee
2	Ik heb een koopwoning/flat	ja/nee
3	Ik bezit een tablet-pc	ja/nee
4	Ik krijg zorgtoeslag	ja/nee
5	Ik krijg huursubsidie	ja/nee

Geluk is gemeten met de volgende vragen:

1 Als ik mijn leven over zou mogen doen zou ik het ... op dezelfde manier doen.

- absoluut niet
 niet
 ten dele wel/niet
 wel
 absoluut

Geluk is gemeten met de volgende vragen:

2 De meeste anderen hebben het beter dan ik.

absoluut niet niet ten dele wel/niet wel absoluut

3 Ik heb het ... naar mijn zin.

absoluut niet niet ten dele wel/niet wel absoluut

4 Het leven is zwaar.

absoluut niet niet ten dele wel/niet wel absoluut

5 Ik voel mij eenzaam.

absoluut niet niet ten dele wel/niet wel absoluut

Gezondheid is gemeten met de vraag:

Hoe is uw gezondheid?

Zeer goed Goed Voldoende Slecht Zeer slecht

Omdat geluk niet alleen van financiële middelen afhangt, maar ook van andere zaken, vraagt de onderzoeker tevens naar een aantal gemakkelijk te meten kenmerken, zoals geslacht, leeftijd, opleidingsniveau en leefsituatie van de personen die hij enquêteert.

Geslacht man vrouw

Leeftijd in jaren ...

Leefsituatie alleen
 met partner
 met partner en kinderen

Hoogste afgeronde opleiding lager (beroeps)onderwijs (lagere school, lager technisch onderwijs, lager voortgezet onderwijs)
 middelbaar (beroeps)onderwijs (middelbaar technisch onderwijs, middelbaar voortgezet onderwijs, mavo, mulo, vmbo enz.)
 hoger (beroeps)onderwijs (universiteit, hbo, vwo, havo)

1.3 Hoe analyseer ik mijn data? Een gebruikswijzer

Voor de keuze van een statistische analysetechniek zijn de antwoorden op de volgende vragen van belang:

- 1 Gaat het in de vraagstelling om frequenties (hoe vaak / in welke mate zijn mensen gelukkig), om een verschil (zijn vrouwen gelukkiger dan mannen?) of om een samenhang (Is er een relatie tussen leeftijd en gezondheid)? Of gaat het om een combinatie daarvan?

- 2 Wat is het meetniveau (nominaal, ordinaal of interval-/rationiveau) van de gegevens die je hebt verzameld?
- 3 Gaat het om een steekproef, of om een populatie?

Het blokschema (te vinden voorin dit boek) is ontworpen aan de hand van deze vragen. In de volgende subparagrafen gaan we daar nader op in.

1.3.1 Om wat voor specifieke onderzoeksvragen gaat het in mijn onderzoek?

Bij het beantwoorden van de vraag welke statistische techniek je kunt gebruiken, vormt de *onderzoeksvraag* of *onderzoeksvragen* het uitgangspunt. Het onderzoek omvat altijd één of meer specifieke onderzoeksvragen waarop een antwoord moet worden gegeven. Globaal zijn er drie soorten onderzoeksvragen te onderscheiden:

- 1 Vragen waarbij het erom gaat hoe vaak of in welke mate iets voorkomt. Een voorbeeld daarvan is: 'In welke mate zijn Nederlanders gelukkig?' of 'Hoeveel procent van de Nederlanders is in het bezit van een auto?' Frequentie
- 2 Vragen waarbij het gaat om een *verschil*. Bijvoorbeeld: 'Zijn vrouwen gelukkiger dan mannen?' Verschil
- 3 Vragen waarbij het gaat om een *samenhang*. Bijvoorbeeld: 'Is er een samenhang tussen leeftijd en gezondheid?' Samenhang

In hoofdstuk 4 geven we een voorbeeld van de analyse van gegevens bij frequentievraagstellingen. De analyse van gegevens bij de verschillende vragen behandelen we in hoofdstuk 5. In hoofdstuk 6 behandelen we voorbeelden van data-analyse bij een samenhangonderzoeksvraag.

1.3.2 Wat is het meetniveau van mijn gegevens?

Als je hebt bepaald om wat voor type onderzoeksvragen het in je onderzoek gaat (zie de eerste kolom in het blokschema), dan moet je vervolgens nagaan op welk meetniveau de variabele(n) is/zijn gemeten. Zie daarvoor in het genoemde blokschema de cellen onder frequentie, verschil of samenhang. Per onderzoeksvraag moet je aangeven wat het meetniveau van de betreffende variabelen is. Bij de onderzoeksvraag naar het verschil tussen vrouwen en mannen in de mate waarin ze zich gelukkig voelen is het meetniveau van de betreffende variabelen namelijk anders (en lager) dan het meetniveau van de variabelen in bijvoorbeeld de onderzoeksvraag naar de samenhang tussen leeftijd en gezondheid. Bij de variabele *geslacht* zijn er maar twee categorieën of waarden, namelijk 'man' en 'vrouw'. Hierbij is alleen sprake van een verschil. Een man is anders dan een vrouw, maar niet meer of minder. Hetzelfde geldt voor *leefsituatie*: iemand is gehuwd, samenwonend of alleenstaand. Bij dit type antwoordmogelijkheden gaat het om een *nominaal meetniveau*. Je kunt alleen zeggen hoeveel vrouwen of mannen een auto hebben, maar niet dat iemand meer 'mans' of meer 'vrouws' is. Je bent man of vrouw. Je kunt niet een beetje man of veel vrouw zijn. Nominaal meetniveau

Dat kan wel bij gegevens die zijn gemeten op respectievelijk ordinaal, interval- of rationiveau. Bij gegevens op *ordinaal meetniveau* is er wél sprake van meer of minder, maar het verschil tussen de categorieën is niet in een getal uit te drukken. Bij de variabele *opleidingsniveau* bijvoorbeeld is er duidelijk sprake van meer en minder. De havo is hoger dan het vmbo, maar er is niet aan te geven hoeveel hoger. Dat geldt ook voor de medailleverdeling op een kampioenschap. Het is duidelijk dat bijvoorbeeld een 100-meter-loper die goud wint, sneller heeft Ordinaal meetniveau

gelopen dan een loper die zilver heeft gewonnen. Het feit dat hij goud heeft gewonnen geeft aan dat hij sneller heeft gelopen, maar niet hoeveel sneller. Bij interval- en ratiomeetniveau is dat verschil tussen categorieën in termen van meer of minder wel in een getal uit te drukken. Temperatuur is daar een goed voorbeeld van. Het verschil tussen 5 en 10 graden Celsius is even groot als het verschil tussen 25 en 30 graden. Bij het *intervalniveau* is er alleen geen sprake van een natuurlijk *nulpunt*, zoals wel het geval is bij gegevens die op *ratiomeetniveau* zijn gemeten, bijvoorbeeld gewicht en lengte. Nul graden Celsius is geen natuurlijk nulpunt. Het natuurlijke nulpunt voor de temperatuur is -273 graden Celsius, wat overeenkomt met nul graden Kelvin. Wanneer je de temperatuur weergeeft in graden Kelvin, is er wel sprake van een ratiometing, want hier is sprake van een natuurlijk nulpunt. Figuur 1.1 is een illustratie van verschillende meetniveaus.

FIGUUR 1.1 Verschillende meetniveaus

Meetniveaus in de praktijk



Ik ben Iwan:

- professioneel gewichtheffer (nominaal)
- en vrijgezel (nominaal)
- ben tweede geworden op wereldkampioenschap (ordinaal)
- mijn IQ is 110 (interval)
- en ik weeg 140 kilo (ratio)

Het meetniveau heeft gevolgen voor de rekenkundige bewerkingen die je mag uitvoeren.

Bij temperatuur gemeten in graden Celsius mag je niet zeggen dat 20 graden tweemaal zoveel is als 10 graden. Bij een meting in graden Kelvin mag je wel zeggen dat 20 graden tweemaal zoveel is als 10 graden en bij een gewichtsmeting dat 20 kilo tweemaal zo zwaar is als 10 kilo.

Het voorgaande wordt schematisch weergegeven in tabel 1.1.

TABEL 1.1 Overzicht van meetniveaus, hun rekenkundige consequenties en voorbeelden

Meetniveau	Rekenkundige consequentie	Voorbeeld
Nominaal	Tellen, percentages (alleen onderscheid)	Geslacht
Ordinaal	Tellen, percentages en hoger/lager (onderscheid en ordening)	Opleidingsniveau
Interval	Tellen, hoger/lager, waarbij verschillen in eenheden zijn uit te drukken, gemiddelde, spreiding (onderscheid en ordening)	Intelligentie
Ratio	Tellen, hoger/lager, waarbij verschillen in eenheden zijn uit te drukken, gemiddelde, spreiding en het berekenen van verhoudingen (onderscheid en ordening)	Leeftijd

We onderscheiden verder continue en discontinue of discrete variabelen. Bij *continue variabelen* kun je je een lijn voorstellen waarop waarden een aaneengesloten rij punten vormen: een continuüm. Tussen twee punten liggen altijd nog (oneindig veel) andere mogelijke waarden. Voorbeelden van continue variabelen zijn de lengte en het gewicht van een persoon. Variabelen die alleen hele waarden kunnen aannemen, noemen we *discrete variabelen*, zoals het aantal auto's dat iemand bezit, of het aantal kinderen in een gezin.

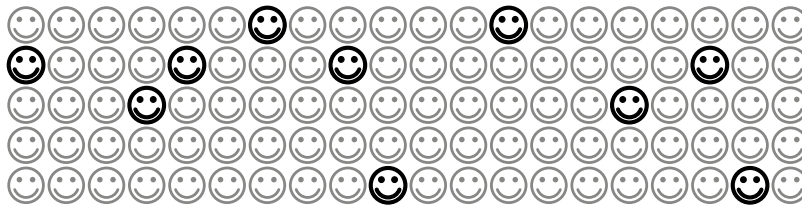
Continue variabelen

Discrete variabelen

1.3.3 Gaat het om een populatie of om een steekproef?

Er zijn twee vormen van statistiek: de beschrijvende en de inductieve statistiek. *Beschrijvende statistiek* gebruik je, wanneer je onderzoek doet bij een *populatie*. Er is sprake van een populatie wanneer alle eenheden waarover je uitspraken wilt doen, in je onderzoek worden betrokken. Dus wanneer je bijvoorbeeld alle werknemers van een bedrijf enquêteert om hun arbeidsatisfactie vast te stellen. Om kosten te besparen kun je ook een deel van de werknemers enquêteren (*steekproef*), die je aselekt uit het totale bestand van werknemers trekt (zie figuur 1.2).

Beschrijvende statistiek

FIGUUR 1.2 Voorbeeld van een populatie en een aselechte steekproef

N=100 hele populatie van werknemers

n=10 aselekt getrokken steekproef uit populatie van werknemers

Bij een steekproef is het wel de opzet dat je uitspraken doet over de totale populatie van werknemers. In dit geval moet je gebruikmaken van de *inductieve of inferentiële statistiek*. Je wilt op grond van een speciaal geval (een steekproef) algemene uitspraken doen (over de populatie). Raadpleeg een methode- of statistiekboek, zoals het *Basisboek Methoden en Technieken*, voor meer informatie. Ook op Wikipedia vind je uitgebreide informatie over het trekken van steekproeven: [http://en.wikipedia.org/wiki/Sampling_\(statistics\)](http://en.wikipedia.org/wiki/Sampling_(statistics)).

Inductieve/inferentiële statistiek

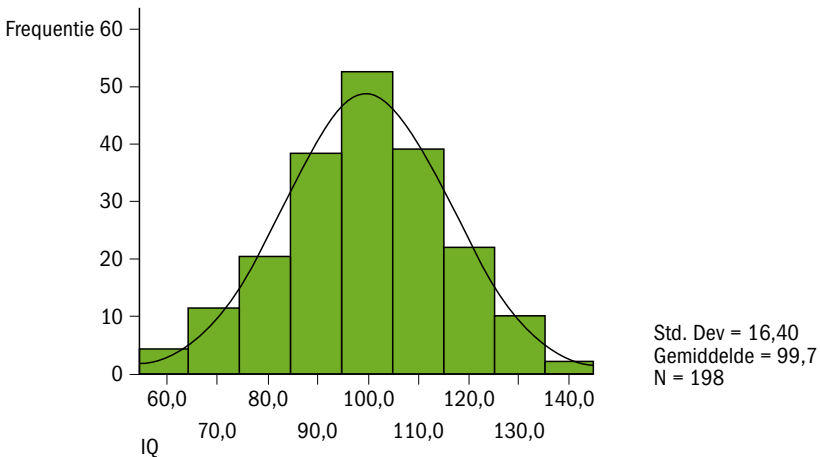
Voordat je aan de analyse van je gegevens begint, moet je jezelf de vraag stellen over welke eenheden (wie of wat) je uitspraken wilt doen. Wanneer dat alleen de personen of zaken zijn die in je onderzoek zijn betrokken, dan is er sprake van een *populatieonderzoek*. Wil je ook uitspraken doen over de personen of zaken die niet betrokken zijn bij je onderzoek, maar als het ware gerepresenteerd worden door de onderzoekseenheden die je geselecteerd hebt, dan gaat het om een *steekproefonderzoek*. In paragraaf 1.4 behandelen we in het kort enkele statistische termen die je steeds tegenkomt bij het toetsen of de resultaten die je in een steekproef vindt, op toeval berusten of, met een bepaalde marge aan onzekerheid, kunnen worden generaliseerd naar de populatie waaruit de steekproef is getrokken.

1.4 Enkele algemene statistische begrippen

Het doel van de beschrijvende statistiek is om op een inzichtelijke en overzichtelijke manier je gegevens te presenteren. Als je van bijna duizend werknemers de arbeidssatisfactie hebt vastgesteld, heeft het weinig zin al die gegevens los te presenteren. Meestal vat je ze samen in bijvoorbeeld een histogram (subparagraaf 4.1.4) of in de vorm van percentages of een gemiddelde (paragraaf 4.2). Je beschrijft je gegevens in een gereduceerde en daarvoor overzichtelijke vorm. Als je gegevens grafisch weergeeft, is het resultaat nogal eens de zogenoemde *normaalverdeling*. In figuur 1.3 zie je daar een (fictief) voorbeeld van.

Normaalverdeling

FIGUUR 1.3 Het intelligentieniveau van de 198 werknemers van de firma 'Arbeid'



In figuur 1.3. zie je de verdeling van de scores op een intelligentietest die is afgenomen bij 198 werknemers van de firma 'Arbeid'. Die verdeling neemt nagenoeg de vorm van een normaalverdeling aan. Ter vergelijking is de normaalverdeling ingetekend. Deze verdeling wordt ook wel de verdeling van *Gauss* of *Gausskromme* genoemd. Het karakteristieke kenmerk van de normaalverdeling is de vorm van een klok en de symmetrie van de linkerhelft met de rechterhelft.

Wanneer de 198 werknemers een aselechte steekproef vormen uit het totale bestand van werknemers van de firma 'Arbeid' ($N = 2213$), moeten we gebruikmaken van de inductieve of inferentiële statistiek. De vraag is dan in welke mate het gevonden gemiddelde IQ van 99,7 representatief is voor de totale populatie van werknemers. Met andere woorden, wat is de kans op een gemiddeld IQ van 99,7 in de populatie, in het geval je inderdaad de gehele populatie in je onderzoek zou kunnen betrekken? Die kans is uiteraard niet zo groot. Want het gemiddelde dat je in de steekproef hebt gevonden, is afhankelijk van de toevallige samenstelling van deze steekproef en zal bij ieder steekproeftrekking een beetje anders zijn.

Wanneer we opnieuw een aselechte steekproef trekken en nog een keer enzovoort, dan zal het gemiddelde IQ waarschijnlijk steeds iets hoger of mogelijk iets lager zijn. De gevonden waarden zullen iets van elkaar afwijken, maar waarschijnlijk ook weer niet erg veel.

In Excel kun je de standaardfout laten berekenen (subparagraaf 4.2.2). De *standaardfout* geeft aan in hoeverre het gevonden steekproefgemiddelde een *betrouwbare* schatting is van het populatiegemiddelde. De standaardfout is groter naarmate het verschil in IQ binnen de groep groter en de steekproef kleiner is. De standaardfout wordt dus bepaald door de steekproefgrootte en de homogeniteit van de steekproef. Op basis van de standaardfout kun je bijvoorbeeld met minstens 95% zekerheid aangeven dat het populatiegemiddelde ongeveer ligt tussen het gevonden steekproefgemiddelde minus tweemaal de standaardfout en het gemiddelde plus tweemaal de standaardfout. Bij een steekproefgrootte van 10 en een spreiding van 20, is het 95% betrouwbaarheidsinterval bij een gevonden gemiddeld IQ van 100 tussen de 85,69 en 114,31 (tabel 1.2), wat betekent dat je niet veel kunt zeggen over de gemiddelde intelligentie in de populatie. Wanneer de steekproefgrootte 1000 is, wordt het betrouwbaarheidsinterval een stuk kleiner, namelijk 98,76 – 101,24. Nu kun je wel zinnige uitspraken doen over de gemiddelde intelligentie van de populatie.

Standaardfout

TABEL 1.2 Voorbeeld invloed spreiding en steekproefgrootte op het betrouwbaarheidsinterval om een gemiddelde te schatten in de populatie op basis van een gevonden steekproefgemiddelde van 100

	Kleine steekproef, grote onderlinge verschillen	Kleine Steekproef kleine onderlinge verschillen	Grote steekproef, grote onderlinge verschillen	Grote steekproef, kleine onderlinge verschillen
Steekproefgemiddelde	100	100	100	100
Steekproefgrootte	10	10	1000	1000
Spreiding	20	5	20	5
Standaardfout	6,32	1,58	0,63	0,16
95% betrouwbaarheidsinterval	85,69 – 114,31	96,42 – 103,58	98,76 – 101,24	99,69 – 100,31

Bron: <https://www.graphpad.com/quickcalcs/Ctmean1.cfm>

Het begrip *zekerheid* of *kans* speelt een belangrijke rol in de inductieve statistiek. Ook wanneer je de gemiddelden van twee steekproeven vergelijkt, is het de vraag wat de kans is dat je een gevonden verschil in gemiddelden te-

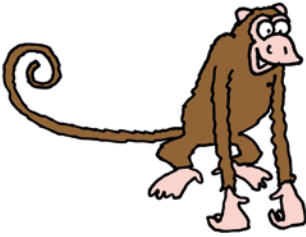
Kans

rugvindt in de populatie. Stel dat de steekproef van werknemers van de firma 'Arbeid' uit 99 vrouwen en 99 mannen bestaat. Je vindt dat het gemiddelde IQ van de vrouwen 102,2 is en dat van de mannen 98,2. Kun je dan stellen dat de vrouwelijke werknemers van de firma 'Arbeid' gemiddeld intelligenter zijn dan de mannelijke werknemers? Of dat verschil 'significant' is, kun je toetsen. In hoofdstuk 5 leggen we uit hoe je dat doet voor verschillen en in hoofdstuk 6 hoe je dat doet bij samenhangen. Wanneer spreek je nu van *significantie*? Men houdt over het algemeen de regel aan dat er van significantie sprake is, als de overschrijdingskans kleiner is dan 5% of bij grotere steekproeven (> 1000) kleiner is dan 1%. Vaak staat er in de Excel-uitdraai ook bij of er *een- of tweezijdig* getoetst is. Je toetst eenzijdig (*one-tailed*) wanneer je een hypothese of verwachting hebt geformuleerd. Als je een theorie hebt op grond waarvan je kunt verwachten dat de vrouwelijke werknemers intelligenter zijn, dan kun je eenzijdig toetsen. Heb je echter geen idee of er sprake is van een verschil en zeker niet van de richting van dat verschil, dan toets je tweezijdig (*two-tailed*).

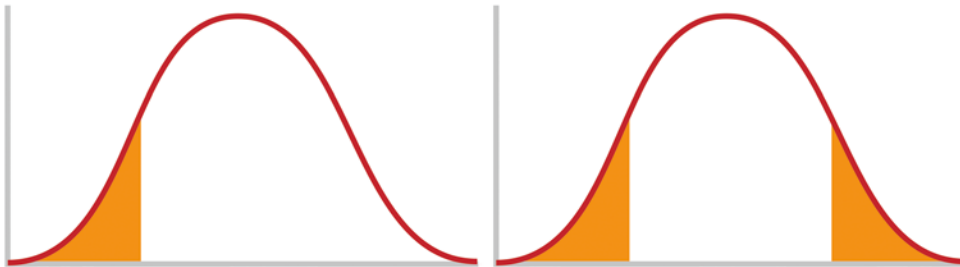
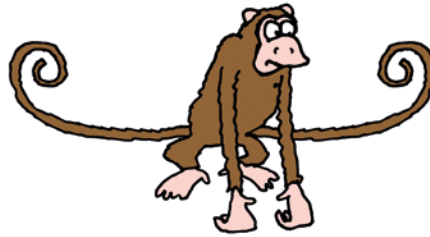
Significantie

Een- of tweezijdig toetsen

One-tailed



Two-tailed



Het bepalen van de significantie is gebaseerd op enkele kenmerken van de steekproef. Dat zijn vaak de omvang en de homogeniteit van de steekproef. Naarmate de steekproef groter is, is de kans op toeval uiteraard kleiner. En naarmate de verschillen op een variabele binnen groepen kleiner zijn (homogene groepen), is de kans dat de verschillen tussen groepen op toeval berusten, eveneens kleiner.

Als een verschil significant is, wil dat niet automatisch zeggen dat het ook *relevant* is. Je ziet in het voorbeeld van figuur 1.3 dat het verschil in gemiddeld IQ vier IQ punten in het voordeel van de vrouwen is. Dat verschil is inderdaad significant, de kans dat het op toeval berust is twee procent, dus inderdaad kleiner dan vijf procent. In hoofdstuk 5 leggen we uit hoe je dat

Relevantie

berekent. Het verschil van vier IQ-punten is dan wel significant, maar niet erg relevant. Het zegt erg weinig over de verschillen tussen mannelijke en vrouwelijke werknemers. Je ziet tegenwoordig in onderzoeksverslagen dat naast de significantie steeds vaker een *effectgrootte* wordt vermeld. De gebruikelijke maat voor effectgrootte is *Cohen's d*. De Cohen's d in het voorbeeld is 0,17, wat een te verwaarlozen effect is: pas als d groter is dan 0,20 wordt er van een klein effect gesproken. Het geslacht verklaart hier namelijk nog geen één procent van de verschillen in IQ. Hoe je dit uitrekent, leggen we uit in subparagraaf 5.3.1.

Effectgrootte

1

In de Excel-uitvoer kom je ook vaak de term *vrijheidsgraden* ($df = \text{degrees of freedom}$) tegen. Het aantal vrijheidsgraden geeft de mate aan waarin scores kunnen variëren. Als je van twee getallen er maar een kent (namelijk 36) en je weet dat het gemiddelde 40 is, dan moet het andere getal 44 zijn. Je hebt hier één vrijheidsgraad. Als je namelijk het ene getal weet, weet je het andere ook. Bij veel toetsen, zoals de t-toets (zie paragraaf 5.3), is het aantal vrijheidsgraden het aantal steekproefelementen minus 1. Bij een kruistabel (zie paragraaf 5.1) is het aantal vrijheidsgraden het aantal rijen minus 1, vermenigvuldigd met het aantal kolommen minus 1. Voor een 2×2 -tabel is het aantal vrijheidsgraden dus 1. Als de randtotalen van een 2×2 -kruistabel bekend zijn en je weet ook een van de celfrequenties (paragraaf 5.1), dan kun je de andere celfrequenties berekenen. Of een waargenomen verschil of samenhang in een steekproef significant is hangt dus mede af van het aantal vrijheidsgraden. Dit aantal vrijheidsgraden is, met uitzondering van kruistabellen, vaak afhankelijk van de grootte van de steekproef.

Vrijheidsgraden

1.5 Hoe werkt Excel?

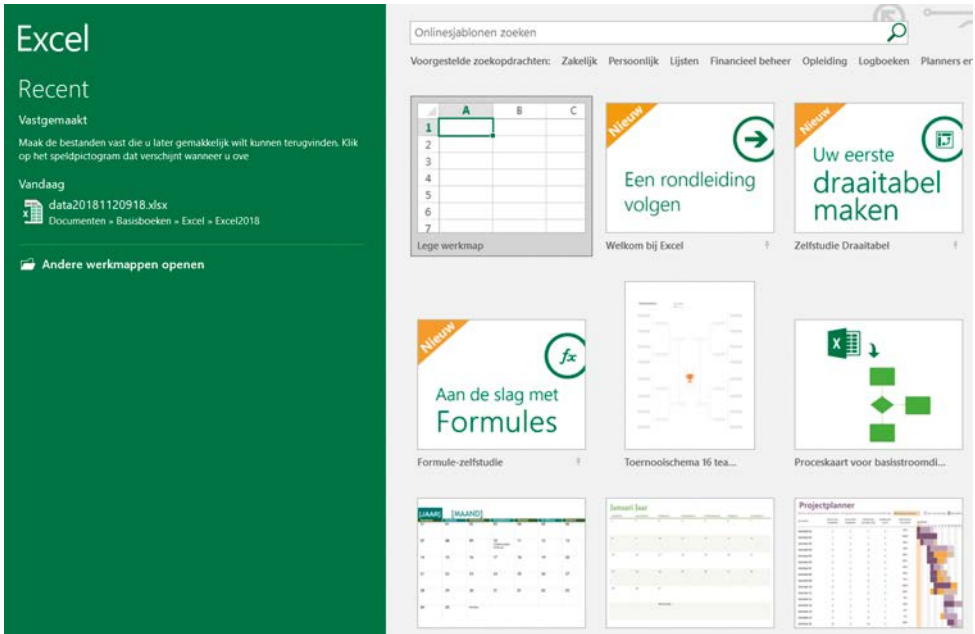
Zoals in paragraaf 1.1 al aangegeven, is Excel niet alleen een middel om data op te slaan en te ordenen, maar biedt het programma ook veel mogelijkheden voor statistische analyse. Als Excel is geïnstalleerd vind je, als het goed is, het volgende icoon op je scherm: zie figuur 1.4.

FIGUUR 1.4 Het Microsoft Excel-icoon



Je 'opent' Excel door op het Microsoft Office Excel-icoon te dubbelklikken. Als je het icoon niet kunt vinden, ga je naar de startknop linksonder en vervolgens naar Programma's. Je krijgt dan een overzicht van de programma's die op je computer staan. Als Microsoft Office in het rijtje voorkomt, klik je daarop en zie je dat Excel een onderdeel is van Microsoft Office. Je kunt Excel starten door er op te dubbelklikken. Je ziet dan het openingsscherm zoals dat in figuur 1.5 is weergegeven.

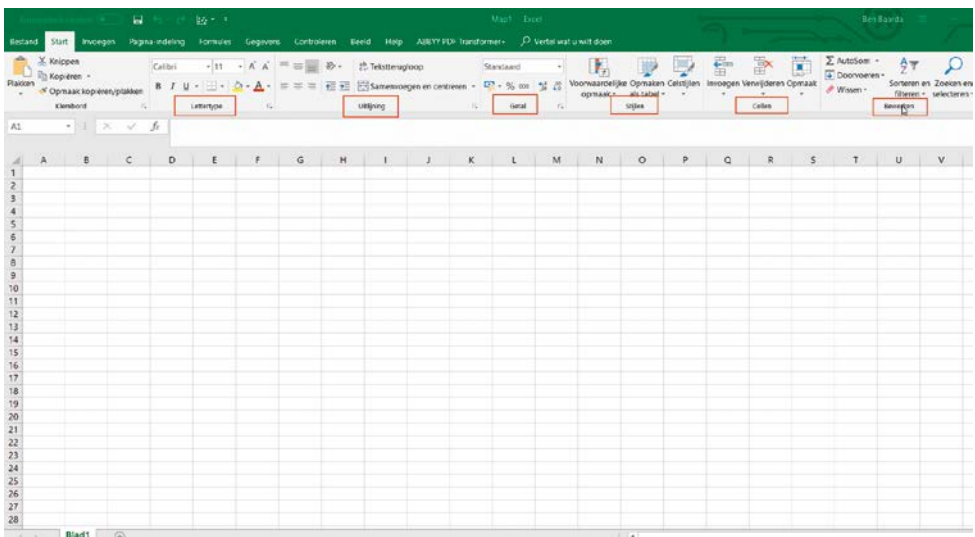
FIGUUR 1.5 Het openingscherm van Excel



Links in figuur 1.5 vind je de bestanden waarmee je het laatst hebt gewerkt. Je opent ze door erop te klikken. Verder zie je dat Excel allerlei voorbeelden van Excel-opzetten (*templates*) klaar heeft staan. Er zijn bijvoorbeeld templates om een project te plannen of om een toernooischema te maken en nog veel meer. We beginnen met een lege werkmap: klik daar dus op. Je komt dan in het lege data-invoerscherm, dat je in figuur 1.6 vindt.

Templates

FIGUUR 1.6 Het lege datascherm



Het data-invoerscherm kent de volgende belangrijke menu's:

- *Lettertype*: hiermee kies je voor een lettertype, grootte, stijl, kleur en dergelijke.
- *Uitlijning*: hiermee kun je de informatie in de cel uitlijnen, bijvoorbeeld links en boven, of juist in het midden van de cel.
- *Getal*: dit staat standaard ingesteld op 'standaard', maar je kunt ook kiezen voor getal, valuta, datum en percentage, al naargelang het karakter van je gegevens. Voor de meeste statistische analyses is het handig als de *getalinstelling* wordt gebruikt.
- *Stijlen*: hier kun je kiezen voor een specifieke opmaak van de cel. Het is soms handig om onderscheid te maken tussen gegevens die zijn ingevoerd en gegevens die je berekend of bewerkt hebt. Door te kiezen voor twee verschillende celstijlen, zie je direct het verschil.
- *Cellen*: je kunt hiermee cellen invoegen, verwijderen of opmaken. In een cel kun je bijvoorbeeld rekenkundige formules zetten. Je kunt zo twee cellen bij elkaar optellen en het resultaat daarvan komt in een derde cel. Het werken met formules komt uitgebreid aan bod (hoofdstuk 3). In bijlage II staan de te gebruiken commando's voor de formules uitgewerkt.
- *Bewerken*: je kunt daar, zoals de titel al aangeeft, gegevens zoeken, sorteren, bewerken, zoals het uitrekenen van een totaalscore en gegevens verwijderen. We komen op de meeste functies in de latere hoofdstukken terug.

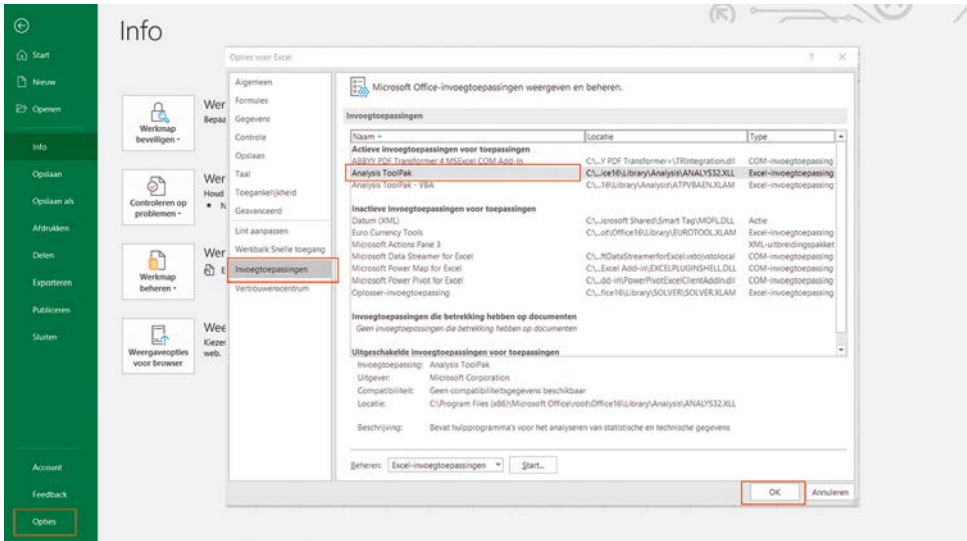
Een aantal van de specifieke Excel-functies, zoals de formulefunctie, zal in dit boek uitgebreid aan bod komen. Een aantal algemene functies, zoals knippen, plakken, kopiëren, kolommen aanpassen en ceileigenschappen veranderen, bespreken wij in hoofdstuk 2. Ben je erg onhandig met Excel, dan verdient het misschien aanbeveling om eerst een korte cursus Excel op internet te volgen. Je kunt dan bijvoorbeeld naar de website <http://www.gratiscursus.be/> gaan.

1.6 Installeren van de statistische functies van Excel

Wanneer je ooit gekozen hebt voor de standaardconfiguratie van het pakket Office, dan worden de statistische mogelijkheden van Excel niet direct geïnstalleerd. Zij zijn wel onderdeel van Excel. Deze statistische functies heb je later nodig wanneer we gaan indelen in klassen, samenhangen gaan analyseren enzovoort.

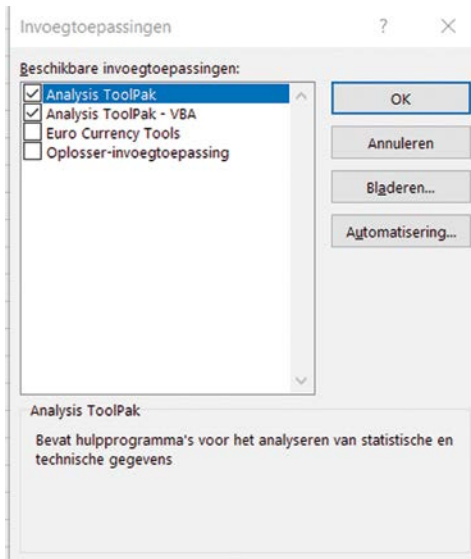
Onder de knop Bestand vind je het commando Opties. Klik hierop en daarna in het menu dat verschijnt op Invoegtoepassingen (figuur 1.7).

FIGUUR 1.7 Het invoegtoepassingenmenu in Excel



Klik dan op Start onderaan bij het beheren van de Excel-invoegtoepassingen. Vink daarna de invoegtoepassingen 'Analysis ToolPak' en 'Analysis ToolPak VBA' aan (figuur 1.8).

FIGUUR 1.8 Het installeren van de Analysis ToolPak



Excel installeert vervolgens de benodigde statistische functies. In het tabblad Gegevens is nu een nieuw tabblad toegevoegd: Gegevensanalyse. Statistiek met Excel kan beginnen!

Samenvatting

-
- ▶ Het prepareren en analyseren van onderzoeksgegevens is een van de laatste fasen in de onderzoekscyclus. De eindfase, de rapportage, is hier voor een groot deel op gebaseerd.
 - ▶ In paragraaf 1.2 lichten wij de casus toe, die wij gebruiken om het uitvoeren van statistische analyse uit te leggen (casus 1.1).
 - ▶ In paragraaf 1.3 geven wij aan dat je voordat je gaat analyseren, een aantal vragen moet beantwoorden. Gaat het bijvoorbeeld in je onderzoeksvraag om een frequentie, een verschil, of om een samenhang? Wat is het meetniveau van je variabelen? Gaat het om een steekproef of om een populatie? Afhankelijk van het antwoord op deze vragen kun je met behulp van het blokschema een keuze maken voor een statistische analysetechniek.
 - ▶ In paragraaf 1.4 leggen wij enkele gangbare statistische begrippen uit. Het gaat daarbij om begrippen als standaardfout, betrouwbaarheid, kans, significantie, een- en tweezijdig toetsen, effectgrootte en vrijheidsgraden.
 - ▶ In paragraaf 1.5 geven wij in het kort aan hoe Excel werkt.
 - ▶ Voor de statistische analyses is het belangrijk dat je de statistische functies in Excel installeert. Hoe je dat doet, wordt uitgelegd in paragraaf 1.6.
-