

René van Vianen

Ben Baarda



---

# Basisboek Statistiek met Excel

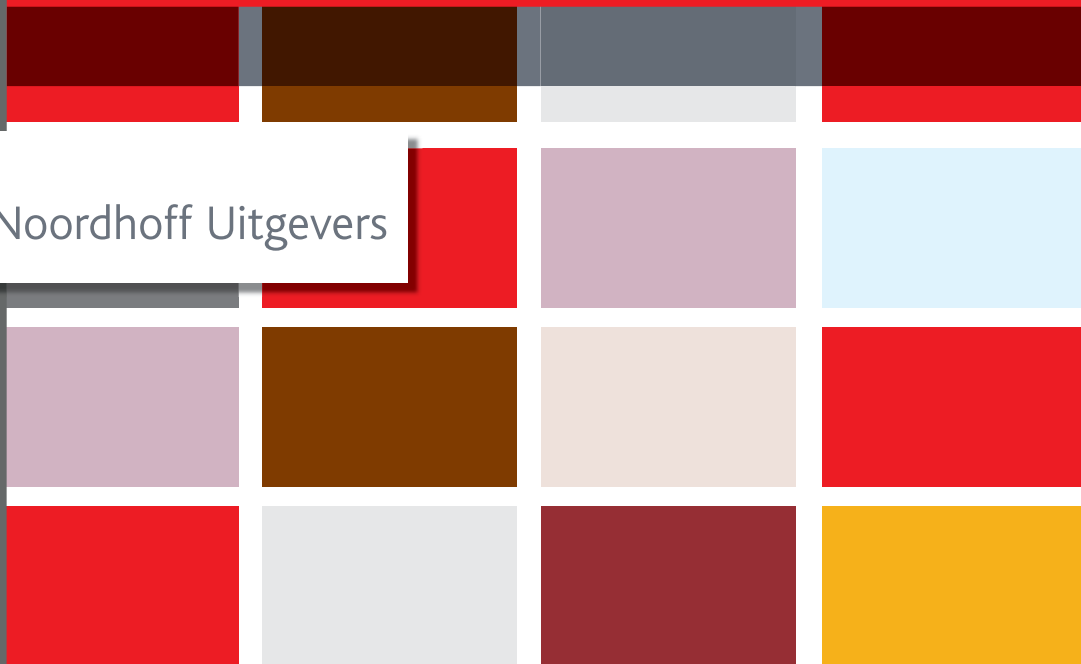
Handleiding voor het verwerken en analyseren  
van en rapporteren over (onderzoeks)gegevens

---

Tweede druk



Noordhoff Uitgevers





# Basisboek Statistiek met Excel

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

**Ben Baarda**

**René van Vianen**

---

Tweede druk

Noordhoff Uitgevers Groningen / Houten

*Ontwerp omslag:* Rocket Industries

*Omslagillustratie:* Getty Images, Amsterdam

*Cartoons:* Auke Herrema, Haarlemmerliede, p. 20 en 24

Eventuele op- en aanmerkingen over deze of andere uitgaven kunt u richten aan:  
Noordhoff Uitgevers bv, Afdeling Hoger Onderwijs, Antwoordnummer 13, 9700 VB  
Groningen, e-mail: [info@noordhoff.nl](mailto:info@noordhoff.nl)

Met betrekking tot sommige teksten en/of illustratiemateriaal is het de uitgever, ondanks zorgvuldige inspanningen daartoe, niet gelukt eventuele rechthebbende(n) te achterhalen. Mocht u van mening zijn (auteurs)rechten te kunnen doen gelden op teksten en/of illustratiemateriaal in deze uitgave dan verzoeken wij u contact op te nemen met de uitgever.

0 / 15



© 2015 Baarda & Van Vianen, vertegenwoordigd door Noordhoff Uitgevers bv Groningen/Houten, The Netherlands.

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever. Voor zover het maken van reprografische verveelvoudigingen uit deze uitgave is toegestaan op grond van artikel 16h Auteurswet 1912 dient men de daarvoor verschuldigde vergoedingen te voldoen aan Stichting Reprorecht (postbus 3060, 2130 KB Hoofddorp, [www.reprorecht.nl](http://www.reprorecht.nl)). Voor het overnemen van gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (artikel 16 Auteurswet 1912) kan men zich wenden tot Stichting PRO (Stichting Publicatie- en Reproductierechten Organisatie, postbus 3060, 2130 KB Hoofddorp, [www.stichting-pro.nl](http://www.stichting-pro.nl)).

*All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.*

ISBN-(ebook) 978-90-01-86120-9

ISBN-978-90-01-86119-3

NUR 916

# Woord vooraf bij de tweede 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 Excel-versie van 2013. Wil je meer weten over Excel, kijk dan eens op de Wikipedia-pagina over Excel ([http://en.wikipedia.org/wiki/Microsoft\\_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 en geluk 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 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.



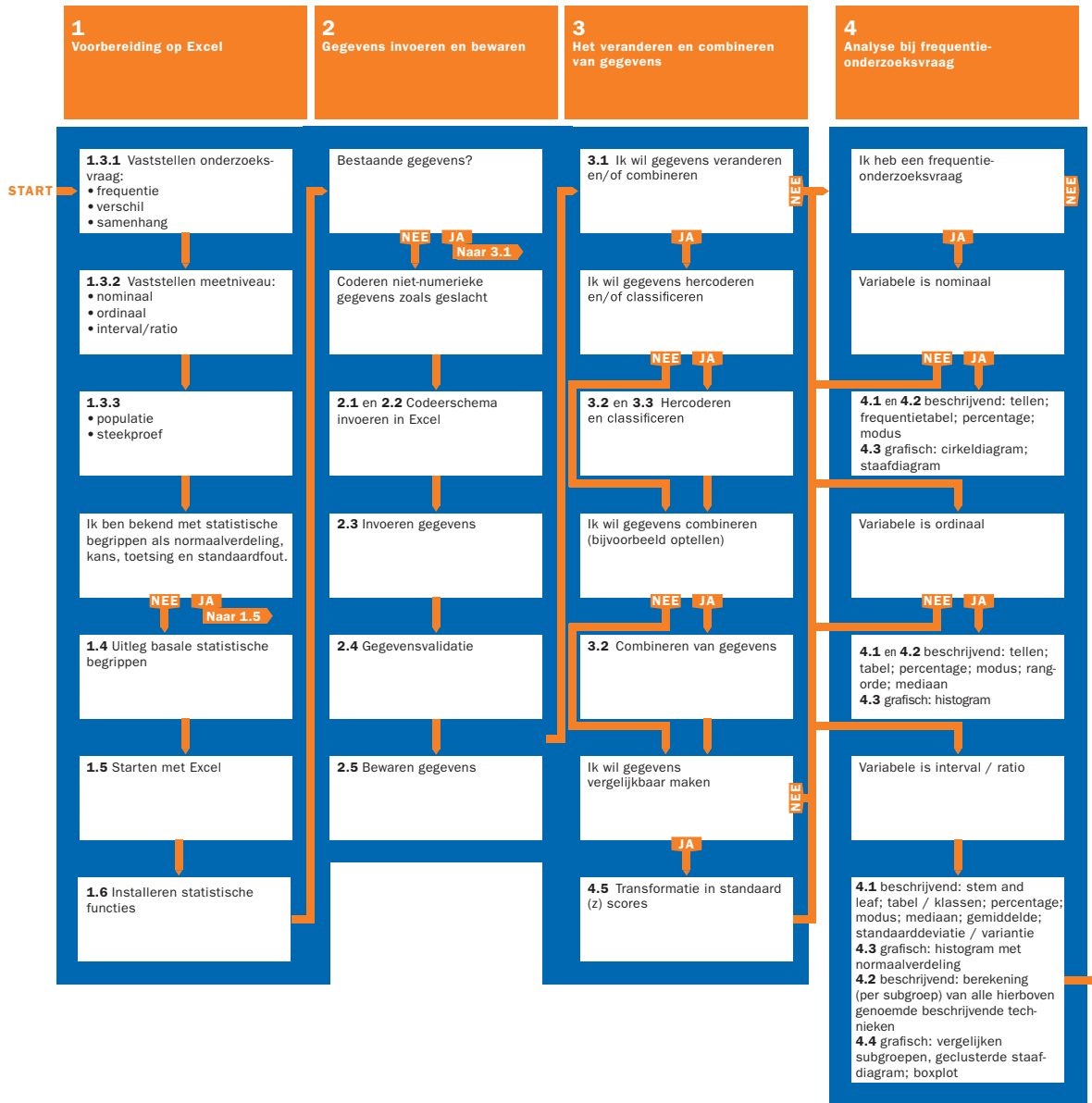
De databestanden waarop we de analyses uitvoeren, zijn te vinden op de site die hoort bij dit boek [www.basisboekstatistiekmetexcel.noordhoff.nl](http://www.basisboekstatistiekmetexcel.noordhoff.nl). Op de deze site vind je nog meer materiaal, zoals een begrippentrainer, ondersteunende video's en de figuren, meestal screendumps uit dit boek.

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.

De huidige druk van dit boek is gebaseerd op de Excel-versie van 2013. Dat betekent vooral dat er nieuwe schermen zijn opgenomen. Maar ook dat de tekst is aangepast. Excel 2013 is ten opzichte van de versie van 2007 in de vorige druk namelijk nogal aangepast. Er zijn functies toegevoegd die het gebruik van Excel handiger maken. Dit geldt bijvoorbeeld voor de 'Quick Analysis'-functie, waarmee direct vanuit het rekenblad, tabellen en grafieken gemaakt kunnen worden voor een geselecteerd deel van databestand. Voor een overzicht van alle wijzigingen, ga naar <http://office.microsoft.com/nl-nl/excel-help/nieuw-in-excel-2013-HA102809308.aspx>

Januari 2015  
Ben Baarda  
René van Vianen



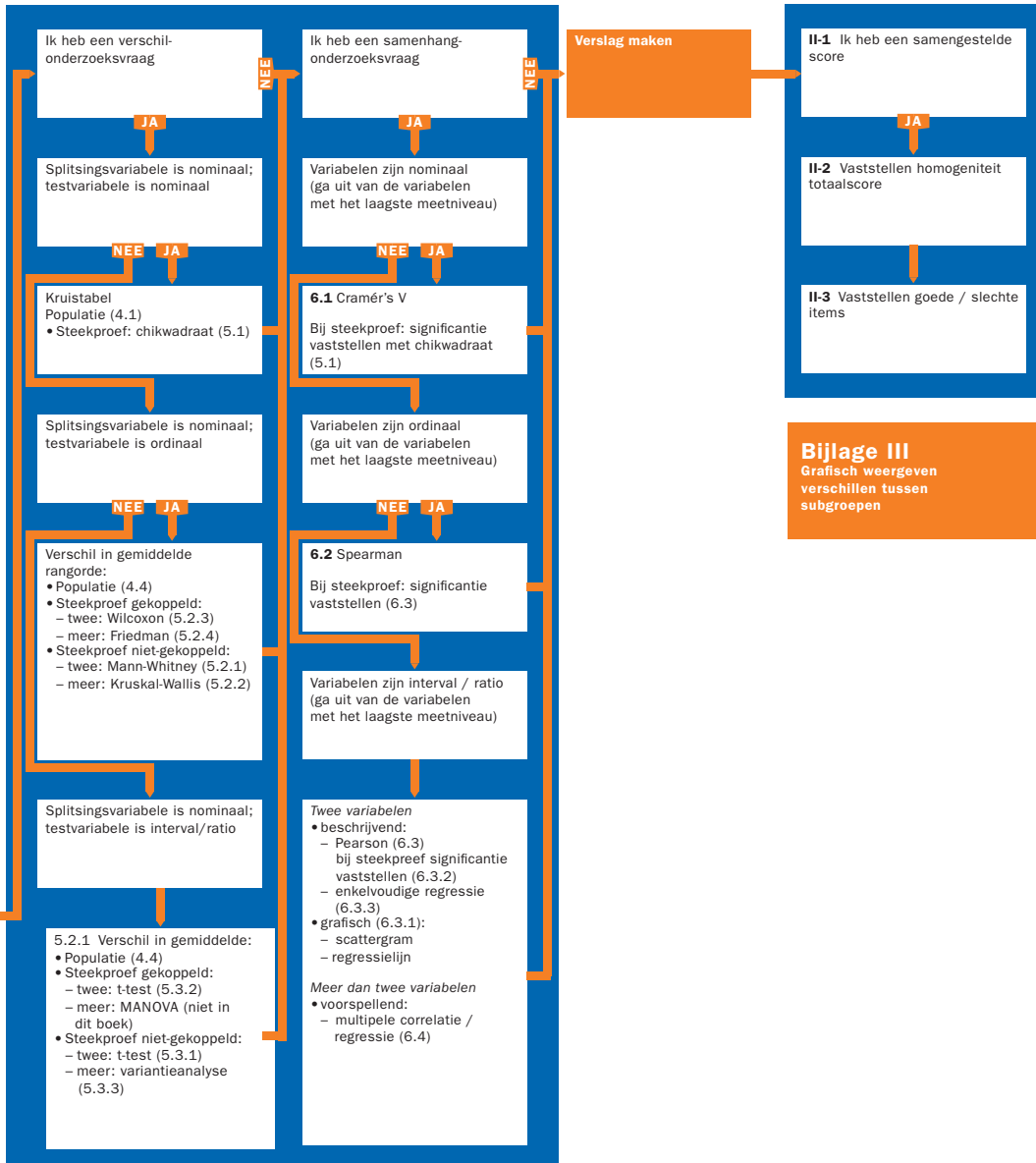




## 5 Analyse bij verschil-onderzoeksvraag

## 6 Analyse bij samenhang-onderzoeksvraag

## Bijlage II 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 en geluk; toelichting op de gebruikte casus 17
  - 1.3 Hoe analyseer ik mijn data? Een gebruikswijzer 19
  - 1.4 Enkele algemene statistische begrippen 22
  - 1.5 Hoe werkt Excel? 25
  - 1.6 Installeren van de statistische functies van Excel 28
- [Samenvatting 30](#)

## 2 Hoe breng ik mijn gegevens in de computer? 33

- 2.1 Hoe maak ik een codeerschema of codeboek? 34
  - 2.2 Hoe voer ik mijn codeboek in Excel in? 36
  - 2.3 Hoe voer ik mijn gegevens in? De datamatrix 39
  - 2.4 Hoe valideer ik mijn gegevens? 43
  - 2.5 Hoe bewaar ik mijn ingevoerde gegevens? 46
- [Samenvatting 47](#)

## 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 63](#)

## 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? 73
  - 4.3 Hoe geef ik frequenties grafisch weer? 81
  - 4.4 Hoe vergelijk ik frequenties van (sub)groepen? 87
  - 4.5 Hoe maak ik scores vergelijkbaar? 93
- [Samenvatting 95](#)

## 5 Hoe analyseer ik mijn gegevens bij een verschilonderzoeksvraag? 97

- 5.1 Verschil bij een nominale test- en splitsingsvariabele 98
  - 5.2 Verschilvragen bij ordinale testvariabelen en nominale splitsingsvariabelen 105
  - 5.3 Verschilvraag bij interval- en ratiotestvariabele en nominale splitsingsvariabele 121
- [Samenvatting 135](#)

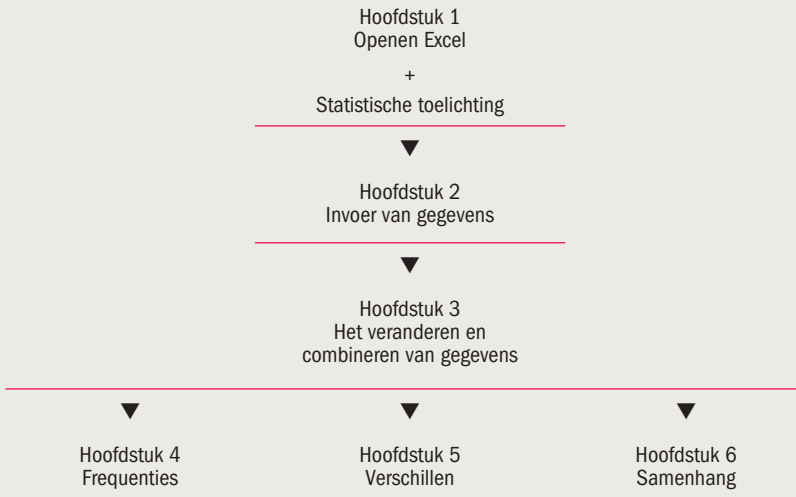


<b>6</b>	<b>Hoe analyseer ik mijn gegevens bij een samenhangonderzoeksvraag?</b>	<b>137</b>
6.1	Samenhang tussen twee nominale variabelen: Cramér's V	138
6.2	Samenhang bij ordinale variabelen: Spearman's rangcorrelatie	140
6.3	Samenhang bij interval- of ratiovariabelen	145
6.4	Samenhang van twee of meer variabelen die gemeten zijn op interval- of rationiveau: multiële regressie	156
	<a href="#">Samenvatting</a>	160
	<b>Bijlage I Handige tips en formules in Excel</b>	<b>161</b>
	<b>Bijlage II Berekenen van homogeniteit</b>	<b>163</b>
	<b>Bijlage III Grafisch weergeven verschillen tussen subgroepen</b>	<b>171</b>
	<b>Register</b>	<b>178</b>
	<b>Over de auteurs</b>	<b>180</b>



# Effectief studeren

## Opbouw boek



**Bijlage I** Handige tips en formules in Excel

**Bijlage II** Berekenen van homogeniteit

**Bijlage III** Grafisch weergeven verschillen tussen subgroepen

## 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 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.



15

# Hoe bereid ik mij voor op statistiek met Excel?

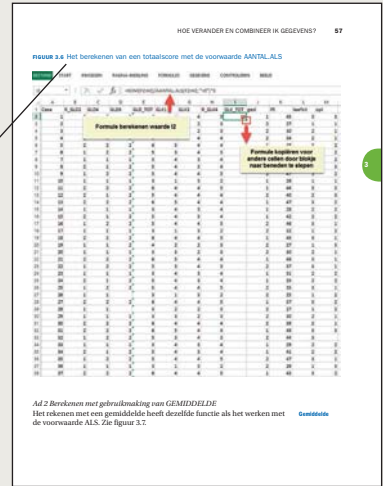
- 1.1 Fasen van de onderzoeksproces
- 1.2 Geldt en geldt, toelichting op de gebruikte casus
- 1.3 Hoe analyseer ik mijn data? Een gebruiksvoorbeeld
- 1.4 Enkele algemene statistische begrippen
- 1.5 Hoe werkt Excel?
- 1.6 Instellingen van de statistische functies van Excel

---

- Frequentie 19
- Verschild 19
- Samenvatting 19
- Nominiaal meetniveau 20
- Ordinaal meetniveau 20
- Intervalniveau 21
- Nulpunt 21
- Raamniveau 21
- Continue variabelen 21
- Discrete variabelen 21
- Beoefenende statistiek 21
- Inductieve/inductieve statistiek 22
- Normaalverdeling 22
- Standaardfout 23
- Kans 23
- Significantie 23
- Kans of toevalligheid toetsen 24
- Raamniveau 24
- Efficiëntie 24
- Vrijheidsgraden 24

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.



20

# HOE BEREID IK MIJ VOOR OP STATISTIEK MET EXCEL?

goed wilt, vooraf heeft gepland dan een lager die zeker heeft gemeten. Het is dat hij goed heeft gemeten geeft aan dat hij vooraf heeft gepland, maar dat is niet het doel.

Bij interval- of raamniveau is dat verschil tussen categorieën in termen van meer of minder wel in een gradé 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 45 en 50 graden. Bij het intervalniveau is er alleen geen sprake van een natuurlijk nulpunt, zoals wel het geval is bij gegevens die op raamniveau zijn gemeten, bijvoorbeeld gewicht en lengte.

Bij raamniveau is de temperatuur voorgesteld in graden Kelvin, in de taal van de natuurwetenschap, want hier is sprake van een natuurlijk nulpunt. Dit heeft gevolgen voor de rekenkundige bewerkingen die je mag uitvoeren. Bij raamniveau 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 geschatting dat 20 kilo tweemaal zo zwaar is als 10 kilo.

Het voorgaande wordt schematisch weergegeven in tabel 1.1.

Meetniveau	Rekenkundige bewerkingen	Voorbeeld
Nominiaal	Soort, percentages (alleen meetniveau)	Geslacht
Ordinaal	Soort, percentages en hoge/lage (ordinaal en nulpunt)	Opleidingsniveau
Interval	Soort, hoge/lage, aantal verschillen in eenheden zijn uit te drukken, gemiddelde, spreiding (continuïteit en nulpunt)	Intelligentie
Raam	Soort, hoge/lage, aantal verschillen in eenheden zijn uit te drukken, gemiddelde, spreiding en het berekenen van verhoudingen (continuïteit en nulpunt)	Lengte

De onderscheidingen tussen continue en discrete variabelen. Bij continue variabelen kun je je een lijn voorstellen waarop waarden en waarnemingen (rij punten) voortvloeien, een voorstelling. Tussen twee punten liggen altijd nog (onduidelijk veel) andere mogelijke waarden. Voorbeelden van continue variabelen zijn de lengte van een persoon, leeftijd en intelligentie. Variabelen die alleen hele waarden kunnen aannemen, noemen we discrete variabelen, zoals het aantal auto's dat iemand heeft of het aantal kinderen in een gezin.

**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 gebruikt je wanneer je onderzoek doet bij een populatie. Er is sprake van een populatie wanneer alle eenheden waarvan je informatie wilt hebben, in je onderzoek worden betrokken. Dit wanneer je bijvoorbeeld alle werknemers van een bedrijf enquêteert om hun arbeidsvoldoening te meten. Om kennis te bezorgen, kun je ook een deel van de werknemers enquêteren (steekproef), die je selecteert uit het totale bestand van werknemers (zie figuren 1.1).

Dit zijn margewoorden om de belangrijkste begrippen te benadrukken.

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

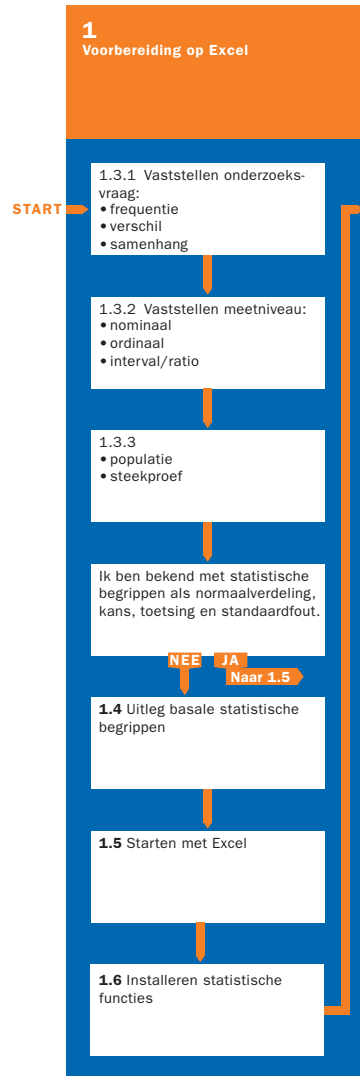
### Opbouw paragraaf

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

25

# Samenvatting

- In een databestand staan vaak veel variabelen, met allerlei verschillende scores van verschillende respondenten. Het is handig om die overzichtelijk te kunnen presenteren, bijvoorbeeld door tabellen te maken waarin gegevens worden samengevat. In paragraaf 4.1 laten we zien hoe je overvloedige kant gegevens met behulp van de functie Datasjabl. De scores op een variabele kunnen ook aangegeven worden met verschillende beschrijvende statistische maten, zoals het gemiddelde, de mediaan en de modus, het minimum en het maximum.
- Op welke wijze je een overzicht krijgt van de samenvattende statistische maten, behandelen we in paragraaf 4.2.
- Vaak worden grafieken gebruikt om onderzoeksgegevens te presenteren. In paragraaf 4.3 werken we uit hoe je de gegevens overzichtelijk en grafisch kunt weergeven.
- Verder willen we vaak onderscheid maken tussen (sub)groepen, bijvoorbeeld het verschil in leeftijd tussen mannen en vrouwen. Het is dit onderscheid maakt, zowel in tabellen als in grafieken, behandelen we in paragraaf 4.4. Een van de grafische mogelijkheden is de gebruik, waarin een aantal samenvattende statistische maten worden vergeleken.
- Voor een onderlinge vergelijking van respondenten moeten de scores getransformeerd worden naar een min/max-vergelijkbare schaal. Daarmee wordt meestal de z-schaal gebruikt. In paragraaf 4.5 leggen wij uit hoe je scores onderling vergelijkbaar maakt.



## Voorkennis

Er wordt geen specifieke voorkennis vereist. Wel wordt aanbevolen om naast dit boek het *Basisboek Methoden en Technieken* (vijfde herziene druk, 2012) of *Dit is onderzoek!* (tweede herziene druk, 2014) te gebruiken.



# 1

## Hoe bereid ik mij voor op statistiek met Excel?

- 1.1 Fasen van de onderzoekscyclus
- 1.2 Geld en geluk; 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

Verschil 19

Samenhang 19

Nominaal meetniveau 20

Ordinaal meetniveau 20

Intervalniveau 21

Nulpunt 21

Rationiveau 21

Continue variabelen 21

Discrete variabelen 21

Beschrijvende statistiek 21

Inductieve/inferentiële statistiek 22

Normaalverdeling 22

Standaardfout 23

Kans 23

Significantie 23

Een- of tweezijdig toetsen 24

Relevantie 24

Effectgrootte 24

Vrijheidsgraden 24

## 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 mijn 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 en geluk (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 en geluk; toelichting op de gebruikte casus

Voordat we ingaan op het gebruik van Excel, introduceren we eerst het voorbeeldonderzoek Geld en geluk (zie casus 1.1). De vragen uit casus 1.1 zijn voorgelegd aan een representatieve steekproef van 500 Nederlandse mannen en 500 Nederlandse vrouwen van 25 tot en met 55 jaar. Er is bewust voor deze leeftijdsgrenzen gekozen. Veel jonge mensen studeren nog en hebben daardoor geen vast inkomen. Bij mensen boven de 55 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](http://www.basisboekstatistiekmetexcel.noordhoff.nl) onder de naam 'Data1'.



### CASUS 1.1

## Voorbeeldonderzoek Geld en geluk

Een onderzoeker wil weten of er een verband bestaat tussen geld en geluk. Hij vraagt zich af of geld gelukkig 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 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. De onderzoeker heeft deze twee 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.

*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. |                          |                          |                          |                          |
|   | <input type="checkbox"/>  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
|   | absoluut niet   | niet                     | ten dele wel/niet        | wel                      | absoluut                 |
| 2 | De meeste anderen hebben het beter dan ik.                                    |                          |                          |                          |                          |
|   | <input type="checkbox"/>  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
|   | absoluut niet   | niet                     | ten dele wel/niet        | wel                      | absoluut                 |

*Geluk is gemeten met de volgende vragen:*

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

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.

<i>Geslacht</i>	<input type="checkbox"/> man	<input type="checkbox"/> vrouw
<i>Leeftijd in jaren</i>	...	
<i>Leefsituatie</i>	<input type="checkbox"/> alleen <input type="checkbox"/> met partner <input type="checkbox"/> met partner en kinderen	
<i>Hoogste afgeronde opleiding</i>	<input type="checkbox"/> lager (beroeps)onderwijs (lagere school, lager technisch onderwijs, lager voortgezet onderwijs) <input type="checkbox"/> middelbaar (beroeps)onderwijs (middelbaar technisch onderwijs, middelbaar voortgezet onderwijs, mavo, mulo, vmbo enz.) <input type="checkbox"/> hoger (beroeps)onderwijs (universiteit, hbo, vwo, havo)	

De onderzoeker heeft de vraagstelling uitgewerkt in een aantal specifieke onderzoeksvragen:

- 1 Hoeveel Nederlanders zijn er in het bezit van respectievelijk een auto, een koopwoning en een tablet-pc?
- 2 Hoeveel Nederlanders krijgen zorgtoeslag?
- 3 Hoeveel Nederlanders ontvangen huurtoeslag?
- 4 Hoe tevreden zijn Nederlanders over het leven dat zij leiden?
- 5 Hoe eenzaam voelen Nederlanders zich?
- 6 Zijn er verschillen tussen mannen en vrouwen wat betreft de volgende situatie?
  - het hebben van een auto, een koopwoning en een tablet-pc
  - het ontvangen van zorgtoeslag
  - het ontvangen van huurtoeslag

- 7 Is er tussen de volgende categorieën een verschil in tevredenheid over het leven dat men leidt?
  - mensen met en zonder partner
  - mensen met en zonder kinderen
- 8 Bestaat er een verschil in eenzaamheid tussen mannen en vrouwen?
- 9 Is er tussen mannen en vrouwen een verschil in tevredenheid over het leven dat men leidt?
- 10 Is er een relatie tussen de mate van tevredenheid over het leven en de leeftijd?
- 11 Bestaat er een relatie tussen geld en geluk?

## 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), om een verschil, of om een samenhang? 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 mannen gelukkiger dan vrouwen?' Verschil
- 3 Vragen waarbij het gaat om een *samenhang*. Bijvoorbeeld: 'Is er een samenhang tussen geld en geluk?' Samenhang

Het is duidelijk dat de voorbeeldonderzoeksvragen 1 tot en met 5 (casus 1.1 in paragraaf 1.2) zijn te karakteriseren als frequentieonderzoeksvragen. Het gaat bijvoorbeeld bij onderzoeksvraag 1 om het aantal Nederlanders dat onder andere een auto bezit. In hoofdstuk 4 geven we een voorbeeld van de analyse van gegevens bij dit type vraagstelling. De onderzoeksvragen 6 tot en met 9 zijn verschilonderzoeksvragen. De analyse van gegevens bij dit type onderzoeksvraag behandelen we in hoofdstuk 5. De onderzoeksvragen 10 en 11 zijn samenhangonderzoeksvragen. In hoofdstuk 6 behandelen we voorbeelden van data-analyse bij een samenhangonderzoeksvraag.

### 1.3.2 Wat is het meetniveau van mijn gegevens?

#### 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)

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 onderzoeksvraag 9 (het verschil tussen mannen en vrouwen 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 onderzoeksvraag 11 (de samenhang tussen geld en geluk). 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 mannen of vrouwen een auto hebben, maar niet dat iemand meer 'man' of meer 'vrouw' is. Je bent man of vrouw. Je kunt niet een beetje man of veel vrouw zijn. 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

Nominaal  
meetniveau

Ordinaal  
meetniveau

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 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 45 en 50 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. Dit 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 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 van een persoon, leeftijd en intelligentie. 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.

### 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 1.1).

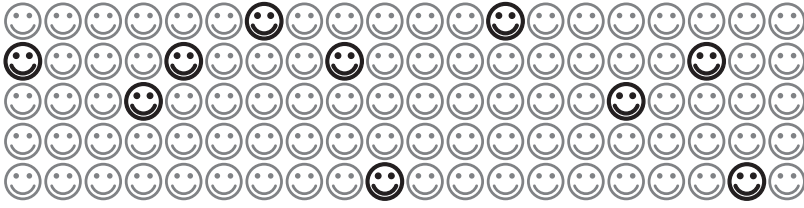
Intervalniveau  
Nulpunt  
Ratiomeetniveau

Continue  
variabelen

Discrete  
variabelen

Beschrijvende  
statistiek

FIGUUR 1.1 Voorbeeld van een populatie en een aselechte steekproef



$N=100$  hele populatie van werknemers

$n=10$  aselekt getrokken steekproef uit populatie van werknemers

**Inductieve/  
inferentiële  
statistiek**

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 methoden- of statistiekboek 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)).

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

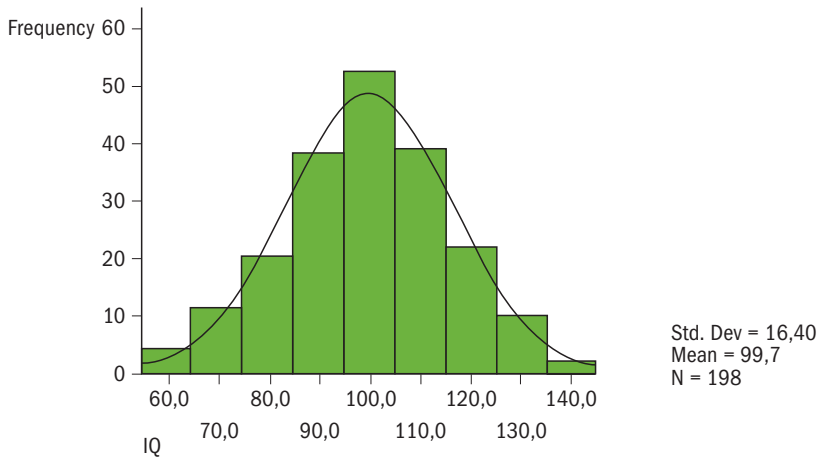
Zoals in paragraaf 1.3 is aangegeven, zijn er twee vormen van statistiek: de beschrijvende en de inductieve of inferentiële statistiek. Het doel van de beschrijvende statistiek is om op een inzichtelijke en overzichtelijke manier je gegevens te presenteren. Als je van bijna duizend werknemers de arbeidssatisfactie hebt vastgesteld, heeft het weinig zin al die gegevens los te presenteren. Meestal vat je ze samen in bijvoorbeeld een histogram (paragraaf 3.3 en subparagraaf 4.1.4) of in de vorm van percentages of een gemiddelde (paragraaf 4.2). Je beschrijft je gegevens in een gereduceerde en daardoor overzichtelijke vorm. Als je gegevens grafisch weergeeft, is het resultaat nogal eens de zogenaomde *normaalverdeling*. In figuur 1.2 zie je daar een (fictief) voorbeeld van.

**Normaalverdeling**

In figuur 1.2 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.



FIGUUR 1.2 Het intelligentieniveau van de 198 werknemers van de firma 'Arbeid'



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 kunt 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.

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 wel 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 ligt tussen het gevonden steekproefgemiddelde minus tweemaal de standaardfout en het gemiddelde plus tweemaal de standaardfout.

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 terugvindt 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

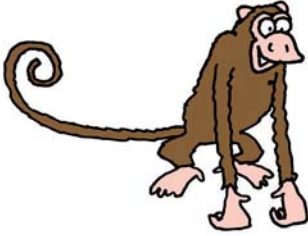
Standaardfout

Kans

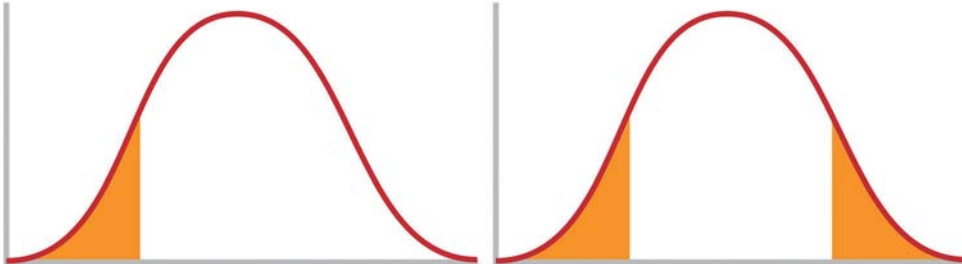
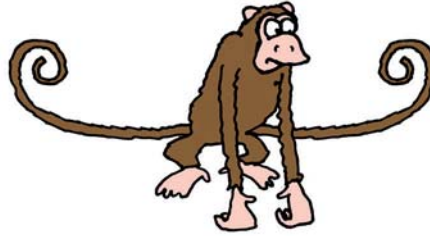
Significantie

als de overschrijdingskans kleiner is dan 5% of bij grotere steekproeven (> 1.000) kleiner is dan 1%. Vaak staat er in de Excel-uitdraai 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*).

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.

## Relevantie

Als een verschil significant is, wil dat niet automatisch zeggen dat het ook *relevant* is. Je ziet in het voorbeeld van figuur 1.2 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 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

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  $2 \times 2$ -tabel is het aantal vrijheidsgraden dus 1. Als de randtotalen van een  $2 \times 2$ -kruistabel bekend zijn en je weet ook een van de celfrequenties (paragraaf 5.1), dan kun je de andere celfrequenties berekenen. Vrijheidsgraden zijn van belang als je op basis van een steekproef een schatting wilt maken van bijvoorbeeld het gemiddelde van de populatie. 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.

## 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 2013 is geïnstalleerd vind je, als het goed is, het volgende icoon op je scherm: zie figuur 1.3.

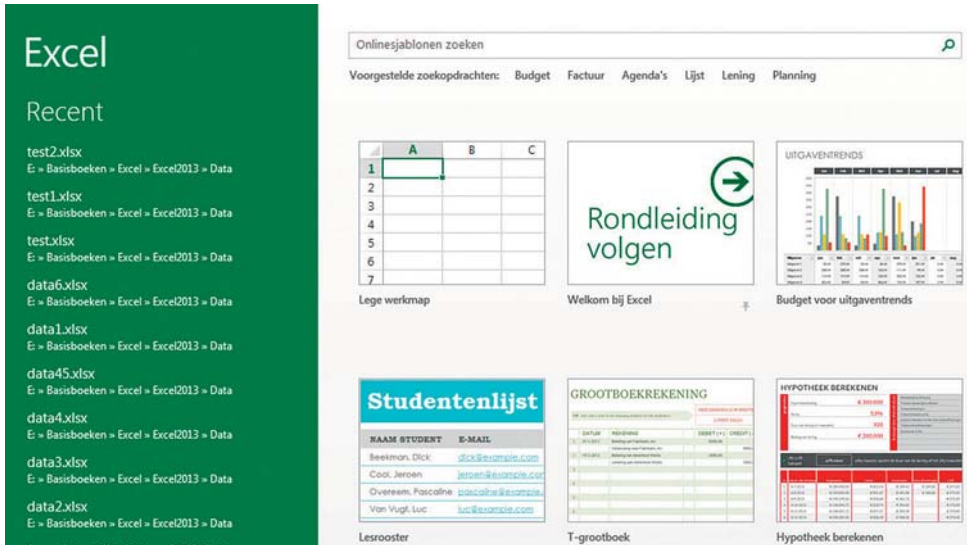
FIGUUR 1.3 Het Microsoft Excel 2013-icoon



Bron: Wikimedia Commons

Je opent Excel door op het Microsoft Office Excel 2013-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 erop te dubbelklikken. Je ziet dan het openingsscherm zoals dat in figuur 1.4 is weergegeven.

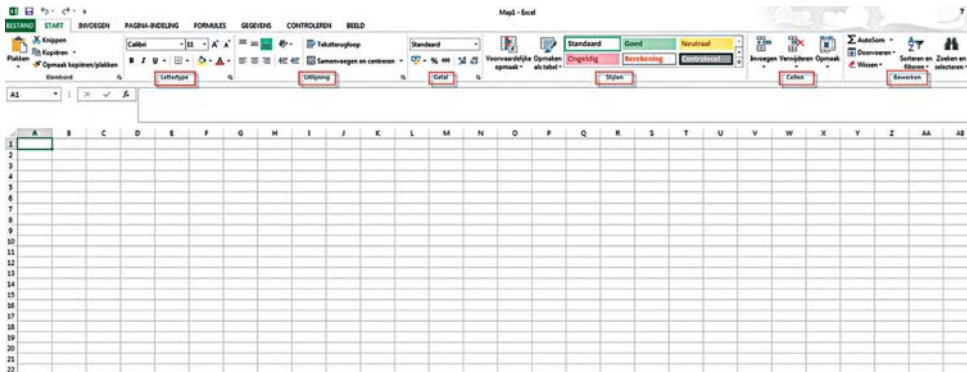
FIGUUR 1.4 Het openingsscherm van Excel 2013



Links in figuur 1.4 vind je de bestanden waarmee je het laatst hebt gewerkt; door erop te klikken, open je ze. Verder zie je dat Excel al allerlei voorbeelden van Excel-opzetten (*templates*) klaar heeft staan. Er zijn templates om bijvoorbeeld je uitgaven bij te houden of om lesroosters bij te houden en nog veel meer. Wij zullen beginnen met een lege werkmap: klik daar dus op. Je komt dan in het lege data-invoerscherm, zoals je dat in figuur 1.5 vindt.

## Templates

FIGUUR 1.5 Het lege data-invoerscherm van Excel 2013



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. Ook kun je hier cellen

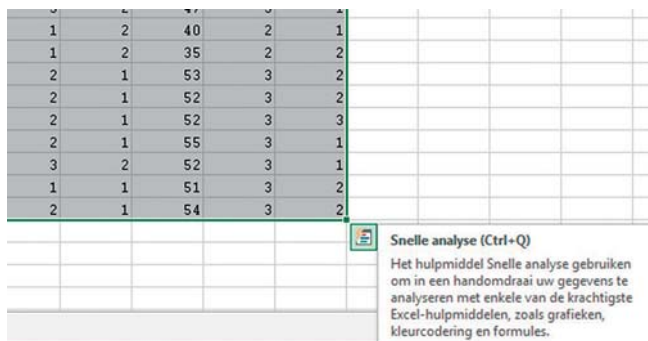
samenvoegen en met de tekstuitloopknop kun je ervoor zorgen dat een lange tekst wordt afgebroken, zodat deze in de cel past.

- *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 of kleuren, zie je direct het verschil. Je kunt bijvoorbeeld de cellen met de berekende scores oranje maken, door op de oranje 'berekening' -knop te klikken. Je moet dan wel eerst de betreffende kolom markeren, door op de betreffende letter bovenaan de kolom te klikken.
- *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 1 staat een aantal gebruikte 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, komt in dit boek uitgebreid aan bod. 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.gratis cursus.be/excel\\_cursus/](http://www.gratis cursus.be/excel_cursus/) gaan.

Een aantal van de besproken functies vind je terug in de 'Snelle analyse' -knop die in Excel 2013 is toegevoegd. Wanneer je gegevens markeert, verschijnt de 'Snelle analyse' -knop rechts onder het gemarkeerde deel, zie figuur 1.6. Door het gebruik van deze functie kun je sneller werken. Je hoeft namelijk niet eerst naar de verschillende menu's te gaan om bijvoorbeeld de opmaak te veranderen, of tabellen te maken. In de loop van dit boek zullen de verschillende functies van deze 'Snelle analyse' -functie aan bod komen.

**FIGUUR 1.6** De nieuwe 'Snelle analyse'-functie in Excel 2013



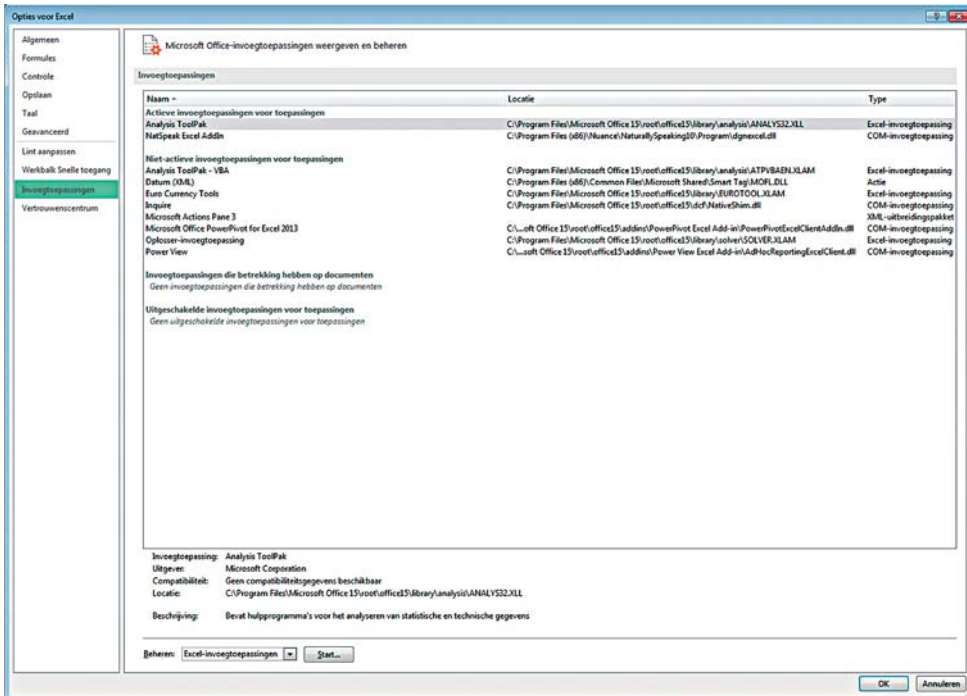


## 1.6 Installeren van de statistische functies van Excel

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.

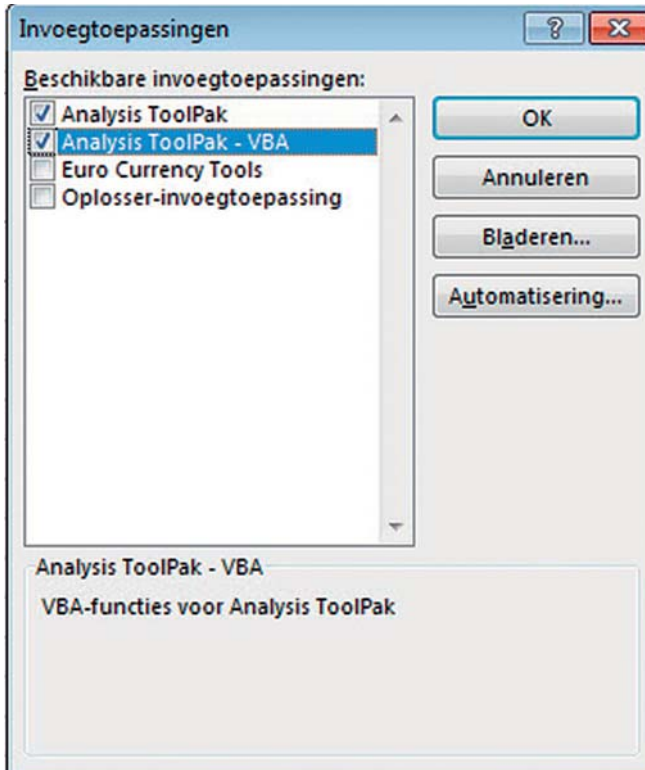
Onder de **BESTAND**knop, linksboven in het datascherm vind je het commando **Opties**. Klik hierop en daarna in het menu dat verschijnt op **Invoegtoepassingen** (figuur 1.7).

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.