

Basisboek Statistiek met Excel



Noordhoff

Ben Baarda & René van Vianen

4^e editie

Basisboek Statistiek met Excel

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

Ben Baarda
René van Vianen

Ontwerp omslag: Shootmedia
Omslagillustratie: Getty Images

Eventuele open 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.



© 2024 Noordhoff Uitgevers B.V. Alle rechten voorbehouden. Tekst- en datamining niet toegestaan.

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) 978-90-01-03898-4
ISBN 978-90-01-03897-7
NUR 916

Woord vooraf bij de vierde editie

In deze nieuwe versie van het *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 1.5), dan beschik je over de meest gangbare statistische analyseprocedures; ook over multivariate technieken, zoals de multivariate variantieanalyse.

Voor dit boek werken wij met de meest recente versie van Excel (365-dagenabonnement). Dit boek gaat niet in op de basisvaardigheden die je nodig hebt voor Excel. Hierover vind je meer op de website van Microsoft of in boeken die gaan over basisvaardigheden voor Excel. Ben je erg onhandig met Excel, dan verdient het misschien aanbeveling om eerst een korte cursus Excel op internet te volgen. Ga hiervoor bijvoorbeeld naar de website Gratis Cursus (www.gratiscursus.be/cursus/microsoft-excel).

In deze nieuwe versie van de handleiding zijn enkele verbeteringen aangebracht:

- Alle instructies voor het verwerken en analyseren van gegevens met behulp van Excel worden stapsgewijs uitgelegd, met de stappen links in beeld en de schermafbeeldingen van de instructie rechts.
- Bij de meeste instructies verwijzen QR-codes naar instructiefilmpjes, waarbij de stappen nogmaals uitgelegd worden.
- Elk hoofdstuk bevat leerdoelen voor het betreffende hoofdstuk. Elke paragraaf wordt afgesloten met een opdracht. Bij de meeste opdrachten wordt een koppeling gelegd naar de data van de student (of docent).
- In hoofdstuk 2 wordt aandacht besteed aan het exporteren van gegevens van een online vragenlijst naar Excel en het kunnen omzetten van tekstuele antwoorden naar waarden (getallen).
- De hoofdstukken 4, 5 en 6 (waarin de analyse van de data centraal staat) hebben een inleiding gekregen met een overzicht waarin je kunt zien welke technieken je kunt gebruiken bij welke meetniveaus.
- Het kunnen werken met een Excel-tabel (hoofdstuk 1) en het kunnen werken met draaitabellen (hoofdstuk 4) heeft meer uitleg en nadruk gekregen.
- Een nieuwe manier voor het indelen van een variabele met veel waarden in klassen is toegevoegd (hoofdstuk 3).

In de vorige editie van het boek is gewerkt aan de hand van een fictief voorbeeldonderzoek 'Geld, geluk en gezondheid'. Dit voorbeeld is gehandhaafd voor deze versie.

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 nodig, maar ook statistische kennis. 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 gemakkelijker via interactieve online programma's zijn uit te voeren. Wel bereiden we het databestand altijd voor in Excel. Via scherm-afdrukken en de instructies die we daarbij geven, laten we 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. Ook geven we aan hoe je over het resultaat kunt rapporteren.

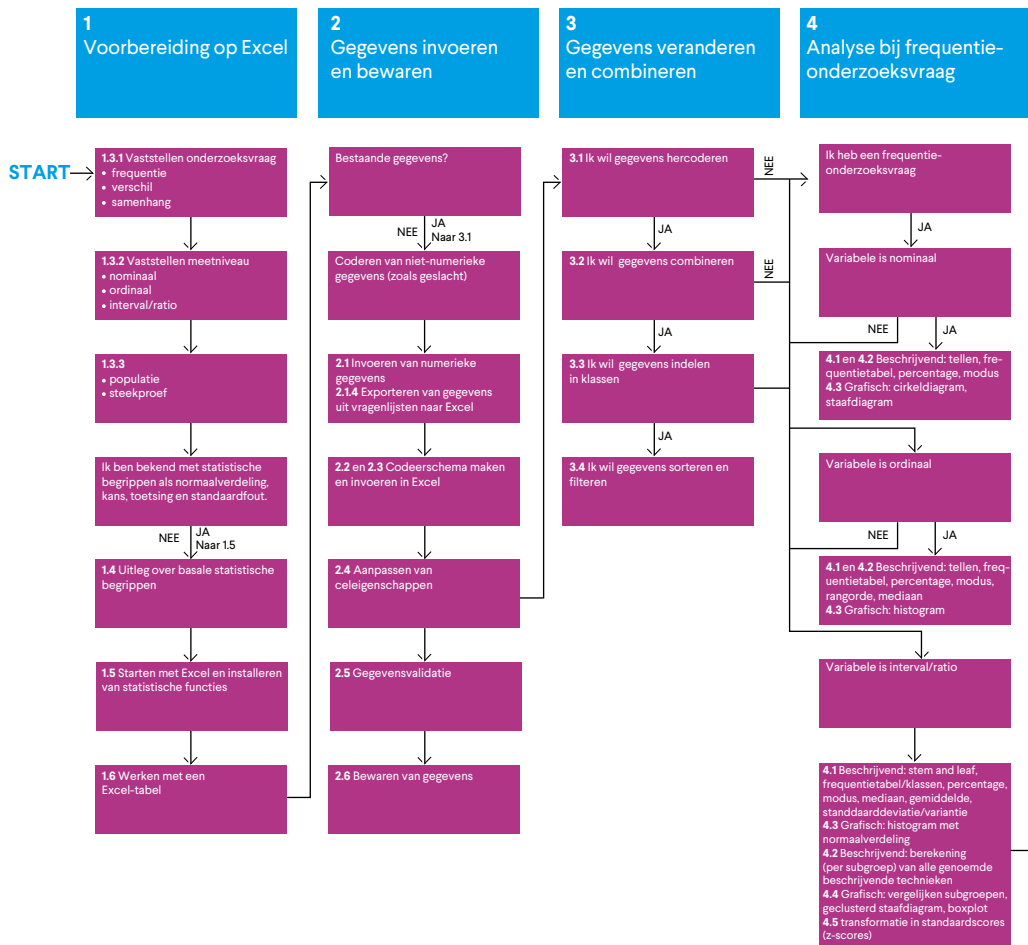


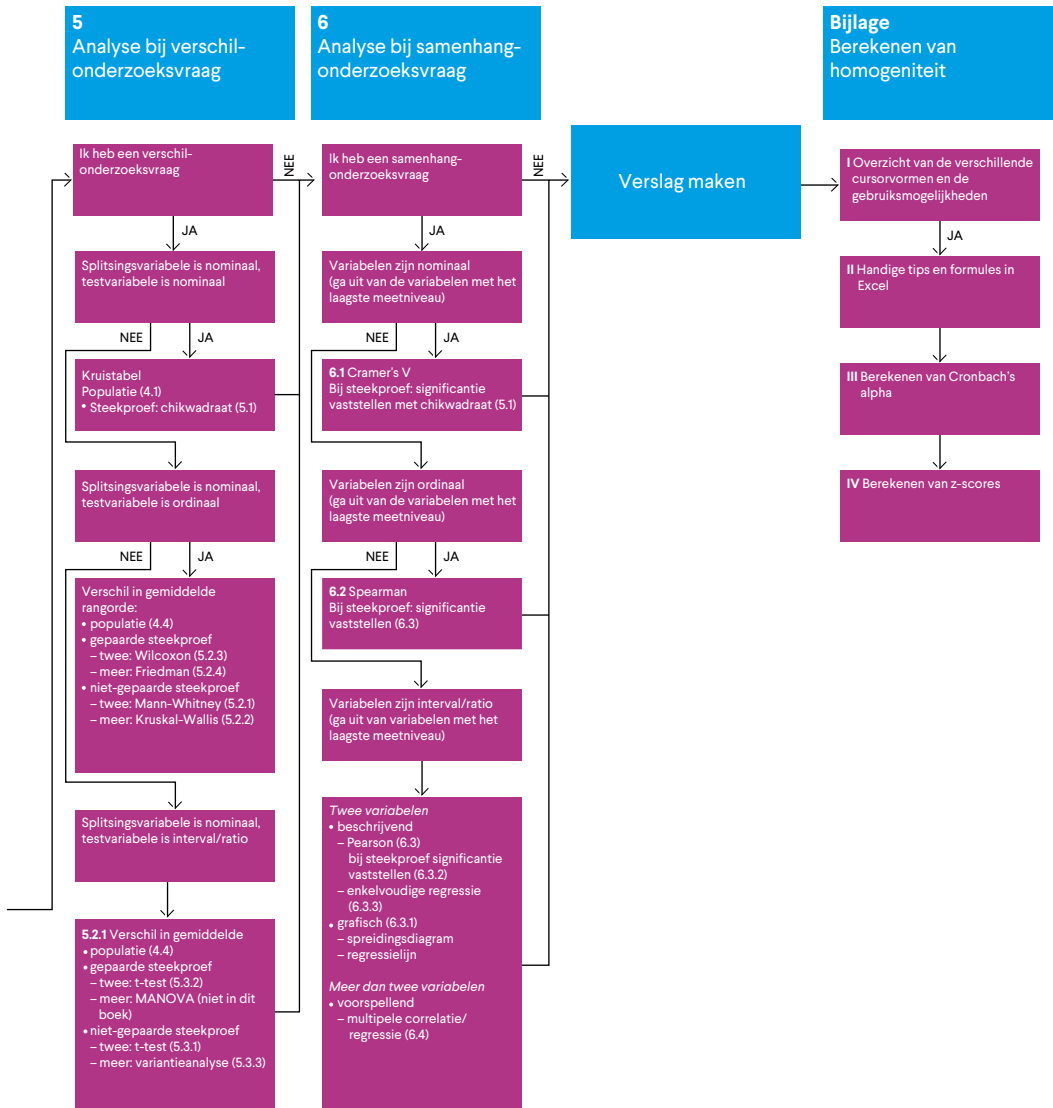
De databestanden waarop we de analyses uitvoeren zijn te vinden op studiemeister.noordhoff.nl. Deze databestanden zijn nieuw; er is onder andere een gezondheidsvariabele toegevoegd.

In de bijlagen geven we tips gegeven over het gebruik van Excel in het algemeen, zoals de betekenis van de verschillende cursoren. Daarnaast zijn er twee bijlagen waarin het berekenen van homogeniteit (Cronbach's alpha en z-scores) met behulp van Excel uitgelegd wordt.

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.

Voorjaar 2024
Ben Baarda
René van Vianen





Inhoud

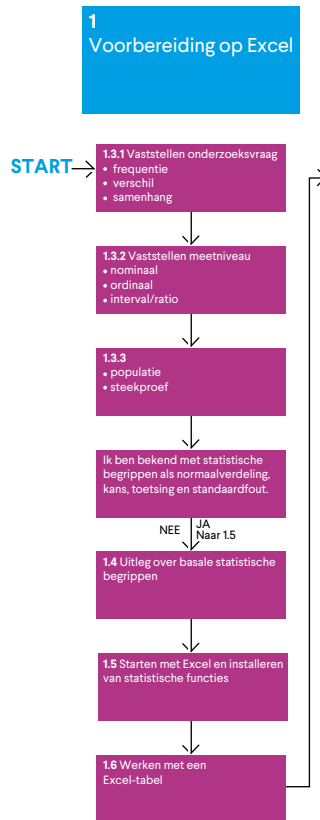
- 1** **Hoe bereid ik mij voor op statistiek met Excel?** 13
 - 1.1 Analyseren van gegevens in de onderzoekscyclus 14
 - 1.2 Geld, geluk en gezondheid: toelichting op de gebruikte casus 15
 - 1.3 Hoe analyseer ik mijn data? Een gebruikswijzer 19
 - 1.4 Enkele algemene statistische begrippen 23
 - 1.5 Hoe werkt Excel en hoe installeer ik statistische functies? 27
 - 1.6 Hoe werk ik met tabellen in Excel? 32

- 2** **Hoe breng ik mijn gegevens in de computer?** 41
 - 2.1 Hoe voer ik mijn gegevens in? De datamatrix 42
 - 2.2 Hoe maak ik een codeerschema of codeboek? 48
 - 2.3 Hoe voer ik mijn codeboek in Excel in? 50
 - 2.4 Hoe pas ik de data of celeigenschappen aan? 53
 - 2.5 Hoe valideer ik mijn gegevens? 58
 - 2.6 Hoe bewaar ik mijn ingevoerde gegevens? 61

- 3** **Hoe verander en combineer ik gegevens?** 67
 - 3.1 Hoe moet ik gegevens hercoderen? 68
 - 3.2 Hoe kan ik gegevens combineren? 77
 - 3.3 Hoe kan ik een variabele indelen in klassen? 88
 - 3.4 Hoe kan ik gegevens sorteren en filteren? 92

- 4** **Hoe analyseer ik mijn gegevens bij een frequentie-onderzoeksvraag?** 101
 - 4.1 Hoe beschrijf ik gegevens bij frequenties? 102
 - 4.2 Hoe werk ik met de functie Draaitabel? 104
 - 4.3 Hoe maak ik een frequentieverdeling in de vorm van een tabel? 108
 - 4.4 Hoe bereken ik samenvattende statistische maten? 118
 - 4.5 Hoe geef ik frequenties grafisch weer? 127
 - 4.6 Hoe vergelijk ik frequenties van (sub)groepen? 134

5	Hoe analyseer ik verschillen?	<i>147</i>
5.1	Het toetsen van verschillen	<i>148</i>
5.2	Verskil bij nominale test- en splitsingsvariabelen	<i>152</i>
5.3	Verschilvragen bij ordinale testvariabelen en nominale splitsingsvariabelen	<i>164</i>
5.4	Verschilvraag bij interval- en ratiotestvariabelen en nominale splitsingsvariabelen	<i>182</i>
6	Hoe bereken ik samenhang?	<i>207</i>
6.1	Het toetsen van samenhang	<i>208</i>
6.2	Samenhang tussen twee nominale variabelen: Cramer's V	<i>210</i>
6.3	Samenhang bij ordinale variabelen: Spearman's rangcorrelatie	<i>213</i>
6.4	Samenhang bij interval- of ratiovariabelen: Pearson's productmomentcorrelatie	<i>218</i>
6.5	Samenhang van twee of meer variabelen gemeten op interval- of rationiveau: multipele regressie	<i>230</i>
	Geraadpleegde bronnen	<i>236</i>
	Bijlage I Overzicht van enkele cursorvormen en gebruiksmogelijkheden	<i>237</i>
	Bijlage II Handige tips en formules in Excel	<i>240</i>
	Bijlage III Berekenen van Cronbach's alpha in Excel	<i>242</i>
	Bijlage IV Berekenen van z-scores in Excel	<i>250</i>
	Register	<i>253</i>
	Over de auteurs	<i>256</i>



Voorkennis

Er wordt geen specifieke voorkennis vereist. Wel wordt aanbevolen om naast dit boek het *Basisboek Methoden en Technieken* (zevende herziene editie, 2021) of *Dit is onderzoek!* (vierde herziene editie, 2024) te gebruiken.



Hoe bereid ik mij voor op statistiek met Excel?

Leerdoelen of -uitkomsten:

- Je weet in welke fase van jouw onderzoek je Excel kunt inzetten met als uiteindelijk doel het analyseren van gegevens.
- Je weet het verschil tussen de verschillende meetniveaus en je kunt de consequenties van de verschillende meetniveaus voor je analyse benoemen.
- Je weet wat een normaalverdeling is en je kent een aantal belangrijke statistische basisbegrippen, zoals standaardfout, zekerheid of kans, significantie en relevantie, effectgrootte en vrijheidsgraden.
- Je weet op welke wijze Excel werkt en hoe je gebruikmaakt van de statistische functies.
- Je kunt tabellen omzetten in draaitabellen (met als doel gegevens te analyseren).

1

- 1.1 Analyseren van gegevens in de onderzoekscyclus 14
- 1.2 Geld, geluk en gezondheid: toelichting op de gebruikte casus 15
- 1.3 Hoe analyseer ik mijn data? Een gebruikswijzer 19
- 1.4 Enkele algemene statistische begrippen 23
- 1.5 Hoe werkt Excel en hoe installeer ik statistische functies? 27
- 1.6 Hoe werk ik met tabellen in Excel? 32

1.1 Analyseren van gegevens in de onderzoekscyclus

Om antwoord te geven op je deelvragen heb je voor je onderzoek kwantitatieve data verzameld door vragenlijsten af te nemen, observatieformulieren te gebruiken, bestaande data zoals kassabestanden te benutten, dossieranalyses te doen of uitslagen van toetsen of tentamens in te voeren. Het resultaat van de dataverzameling tijdens een onderzoek is meestal een berg gegevens. Dit zijn allemaal data die je kunt gebruiken voor het analyseren van onderzoeksgegevens.

De volgende stap is het kiezen van de juiste statistische techniek om de verzamelde gegevens te analyseren. 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?*
- 10 Welke conclusies kan ik op basis van het onderzoek trekken?
- 11 Welke aanbevelingen kan ik formuleren?

In dit boek staan de fasen 7 en 8 en een deel van fase 9 centraal: preparatie, analyse en beschrijving van de onderzoeksgegevens die met Excel worden geanalyseerd. Deze analyse levert de basis voor de conclusies. Aan de hand van een onderzoek naar de relatie tussen geld, geluk en gezondheid (zie casus 1.1) krijg je stap voor stap antwoord op 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), om 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 om een populatie gaat. Je kunt dan aan de hand van het blok-schema vaststellen welke analysetechniek het best bij jouw onderzoeksvraag past. Je vindt het schema voor in dit boek.

- Hoe moet je de analyse met Excel uitvoeren en hoe moet je de resultaten interpreteren? Bij iedere statistische techniek is aangegeven hoe je dit doet. Ook zie je hoe je verslag kunt doen van de resultaten.

In dit hoofdstuk leer je een aantal belangrijke statistische begrippen, zoals normaalverdeling, significantie, effectgrootte, kans en standaardfout (paragraaf 1.4). Paragraaf 1.5 laat zien hoe je Excel kunt starten en hoe je de invoegtoepassing Analysis ToolPak installeert. Hierdoor kun je allerlei statistische analysetechnieken (zoals de variantieanalyse) gebruiken. De laatste paragraaf (1.6) gaat over de manier waarop je binnen Excel werkt met tabellen en draaitabellen om de gegevens te kunnen ordenen en analyseren.

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

Bij de meeste paragrafen (en analysetechnieken) maak je gebruik van 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 studiemeister.noordhoff.nl onder de naam 'data_geld_geluk_gezondheid'.



PRAKTIJKVOORBEELD 1.1 Voorbeeldonderzoek 'Geld, geluk en gezondheid'

Ja, geld maakt gelukkig

Geld. In de aanloop naar de verkiezingen gaat het er vaak over. De meeste mensen willen er graag wat meer van. Maar ben je ook tevreden als je uiteindelijk meer euro's hebt? Maakt geld gelukkig?

Alle tegeltjeswijsheden ten spijt, de grote lijn in het onderzoek luidt: ja, geld maakt gelukkig.

Meer inkomen maakt mensen gelukkiger. Meer vermogen en minder schulden doen dat ook. Belangrijke kanttekening: grote blijdschap over een loonsverhoging of een geldprijs houdt niet lang aan. Maar onze gevoelens van geluk en welbevinden worden wel degelijk beïnvloed door onze financiële situatie op de langere termijn.

Extra blij

De relatie tussen geluk en geld is best complex. Een voorbeeld: mensen die al heel gelukkig zijn, reageren anders op een stijging van hun inkomsten dan ongelukkige mensen. Dat blijkt uit recent onderzoek van psychologen Matthew Killingsworth, Daniel Kahneman en Barbara Mellers onder tienduizenden Amerikanen.

Ongelukkige mensen worden wel wat gelukkiger van meer inkomen. Maar de toename van hun geluk vlakt af bij een bruto jaarinkomen tussen de 60.000 en 90.000 dollar. En komt tot stilstand bij 100.000. Opmerkelijk genoeg blijken de meest gelukkige mensen juist blij met elke extra dollar. Ook die boven de 100.000.

(Nuttig weetje: volgens de OESO was in 2022 het gemiddelde bruto-jaarinkomen in de VS 77.463 dollar. In Nederland was dat gerekend in dollars: 63.225.)

Zelfverdiend maakt gelukkiger

Allerlei factoren bepalen hoe blij je wordt van geld. Ik noem er twee.

- Wonen in een rijk land maakt mensen gelukkiger. Geluksonderzoeker Edward Diener en zijn collega's stelden vast dat een hoger nationaal inkomen leidt tot een positievere evaluatie van je leven. Dat geldt zowel voor mensen met een hoger als met een lager individueel inkomen.
- Het maakt uit hoe je aan je geld komt. Experimenteel onderzoek van de Chinese psychologen Bihui Jin en Jing Li laat zien dat studenten gelukkiger worden wanneer ze geld uitgeven dat ze zelf hebben verdiend dan wanneer ze geld uitgeven dat ze van hun ouders krijgen. Volgens de onderzoekers komt zelfverdiend geld tegemoet aan twee belangrijke menselijke behoeften: autonomie en competentie.

Echt geluk?

Zit écht geluk niet in heel andere dingen dan geld? Ons antwoord op die vraag blijkt deels afhankelijk van ons inkomen.

Onderzoeker Rhia Catapano en haar collega's ploegden door de data van ruim een half miljoen mensen in meer dan honderd landen. Hun conclusie: mensen die meer te besteden hebben, zijn doorgaans iets gelukkiger.

Maar mensen met minder geld ontlenen hun geluk in grotere mate aan een gevoel van betekenis: de overtuiging dat hun leven een doel, waarde en richting heeft. En dat leidt dan weer tot meer welzijn op de lange termijn. Ben je rijk en eigenlijk een beetje jaloers op dat gevoel van betekenis? Dan helpt het, volgens de onderzoekers, om geld te geven aan goede doelen en dingen te kopen voor anderen.

Grappig. Echt, duurzaam geluk blijkt te koop, zolang je je geld niet aan jezelf besteedt.

Bron: NRC, rubriek 'Zo werkt dat', Ben Tichelaar,
2 november 2023

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 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 waarover iemand kan beschikken in het onderzoek betrokken.

Geluk wordt als volgt gedefinieerd: geluk is de mate waarin iemand tevreden is met het leven dat hij leidt. Gezondheid wordt gemeten met de vraag hoe gezond iemand zichzelf voelt.

De onderzoeker heeft deze begrippen meetbaar gemaakt door ze om te zetten in vragen. Dit wordt operationaliseren genoemd. Zowel voor het meten van het begrip geld als voor het meten van het begrip geluk heeft hij vijf uitspraken (of items) gemaakt. Het begrip gezondheid wordt gemeten met één vraag.

Geld is gemeten met de volgende items:

1	Ik kan een nieuwe elektrische auto aanschaffen.	ja/nee
2	Ik woon in een koopwoning.	ja/nee
3	Ik bezit meer dan € 50.000 aan eigen vermogen.	ja/nee
4	Ik krijg zorgtoeslag.	ja/nee
5	Ik heb geld om op vakantie te gaan.	ja/nee

Geluk is gemeten met de volgende vragen:

- 1 Als ik mijn leven zou mogen overdoen, zou ik het op dezelfde manier doen.

absoluut niet niet ten dele wel/niet wel absoluut

- 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 volgende 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 ook naar een aantal gemakkelijk te meten kenmerken, zoals geslacht, leeftijd, opleidingsniveau en leefsituatie van de personen die hij enquêteert.

- Geslacht*
- vrouw
 man
 non-binair
- Leeftijd in jaren*
- ...
- Leefsituatie*
- alleen
 alleen met kind(eren)
 met partner
 met partner en kind(eren)
- Hoogste afgeronde opleiding*
- lager (beroeps)onderwijs (lagere school, lager technisch onderwijs, lager voortgezet onderwijs)
 middelbaar (beroeps)onderwijs (middelbaar technisch onderwijs, middelbaar voortgezet onderwijs, mavo, mbo, vmbo enz.)
 hoger (beroeps)onderwijs (universiteit, hbo, vwo, havo)

OPDRACHT

- Met welke concepten (zoals geld, geluk of gezondheid) werk jij binnen jouw onderzoeksvragen?
- Op welke wijze heb je deze concepten geoperationaliseerd (in een of meer vragen)?
- Welke gegevens heb je (aanvullend) verzameld (zoals geslacht en leeftijd) om eventuele uitspraken te kunnen doen tussen deze kenmerken en uit je onderzoek afkomstige concepten?

1.3 Hoe analyseer ik mijn data? Een gebruikswijzer

Voor de keuze van een statistische analysetechniek zijn de antwoorden op de volgende drie 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 (dat je voor in dit boek vindt) is ontworpen aan de hand van deze vragen. In de volgende subparagrafen gaan we hierop 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 vormen de *onderzoeksvragen* het uitgangspunt.

Het onderzoek omvat altijd een of meer specifieke onderzoeksvragen waarop een antwoord moet worden gegeven. Globaal zijn er drie soorten onderzoeksvragen te onderscheiden:

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

Hoofdstuk 4 behandelt een voorbeeld van de analyse van gegevens bij frequentievraagstellingen. De analyse van gegevens bij de verschillende vragen wordt behandeld in hoofdstuk 5. In hoofdstuk 6 vind je voorbeelden van data-analyse bij een samenhangonderzoeksvraag.

OPDRACHT

- a Met welke specifieke onderzoeksvragen heb jij te maken?
- b Gaat het om vragen over frequenties, verschillen of samenhangen?

1.3.2 Wat is het meetniveau van mijn gegevens?



Als je hebt bepaald om welk 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 of zijn gemeten. Zie daarvoor in het genoemde blokschema de cellen onder frequentie, verschil of samenhang, of scan de QR-code voor een uitleg.

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 drie categorieën of waarden, namelijk ‘man’, ‘vrouw’ en ‘non-binair’. Hierbij is alleen sprake van een verschil. Een man is anders dan een vrouw of een non-binair persoon, 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 deze gegevens tellen (hoeveel respondenten wonen samen, zijn gehuwd of zijn alleenstaand), maar je kunt geen onderscheid maken: er is geen ordening mogelijk tussen samenwonend, gehuwd of alleenstaand. Het een is niet meer of minder dan het ander.

Nominaal
meetniveau

Ordenen is wel mogelijk bij gegevens die zijn gemeten op respectievelijk ordinaal, interval- of rationiveau. Bij gegevens op *ordinaal meetniveau* is 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-meterloper die goud wint sneller heeft gelopen dan een looper die zilver heeft gewonnen. Het feit dat hij goud heeft gewonnen geeft aan dat hij sneller heeft gelopen, maar niet hoeveel sneller.

Ordinaal
meetniveau

Bij interval- en ratiomeetniveau is het verschil tussen categorieën in termen van meer of minder wél in een getal uit te drukken. Temperatuur is een goed voorbeeld. 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 *rationiveau* zijn gemeten, bijvoorbeeld gewicht en lengte.

Intervalniveau
Nulpunt
Rationiveau

Nul graden Celsius is geen natuurlijk nulpunt. Het natuurlijke nulpunt voor de temperatuur is -273 graden Celsius, wat overeenkomt met 0 graden Kelvin. Wanneer je de temperatuur weergeeft in graden Kelvin is er wél sprake van een ratiometing, want er is sprake van een natuurlijk nulpunt. Figuur 1.1 is een illustratie van verschillende meetniveaus.

Figuur 1.1
Verschillende
meetniveaus



Ik ben Iwan:

- *professioneel gewichtheffer (nominaal)*
- *en vrijgezel (nominaal)*
- *ben tweede geworden op het 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 die wordt 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 wél 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 consequenties	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

Continue
variabelen

Discrete
variabelen

We onderscheiden verder continue en discontinue of discrete variabelen. Bij *continue variabelen* kun 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 noem je *discrete variabelen*, zoals het aantal auto's dat iemand bezit of het aantal kinderen in een gezin.

OPDRACHT

- Vul tabel 1.1 in met jouw gegevens. Welke variabelen heb jij gemeten (in de laatste kolom) en welk meetniveau hebben jouw variabelen?
- Check deze meetniveaus aan de hand van de wijze waarop jij de variabelen geoperationaliseerd hebt.

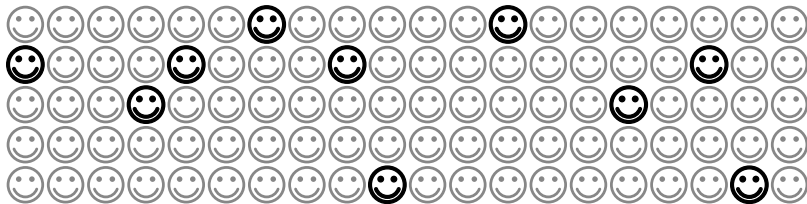
1.3.3 Gaat het om een populatie of om een steekproef?

Beschrijvende
statistiek

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. Dit is het geval wanneer je bijvoorbeeld alle werknemers van een bedrijf enquêteert om hun arbeidssatisfactie vast te stellen. Om kosten te besparen kun je ook een deel van de werknemers enquêteren (*steekproef*), die je aselect uit het totale bestand van werknemers trekt (zie figuur 1.2).

Figuur 1.2

Voorbeeld van een populatie en een aselecte steekproef



$N = 100$ hele populatie van werknemers
 $n = 10$ aselect getrokken steekproef uit populatie van werknemers

Inductieve/
inferentiële
statistiek

Bij een steekproef is het de opzet dat je uitspraken doet over de totale populatie. Bij verkiezingsonderzoek wordt bijvoorbeeld door een onderzoeksbureau niet aan alle stemgerechtigden, maar aan personen in een steekproef de vraag voorgelegd op wie ze gaan stemmen. In dit geval moet je gebruikmaken van de *inductieve of inferentiële statistiek*. Een steekproef is namelijk enigszins afhankelijk van toeval. Dit hangt onder andere af van de grootte van de steekproef. Je wilt op grond van een speciaal geval (een steekproef) algemene uitspraken doen (over de

populatie). Raadpleeg een methoden- of statistiekboek – zoals het *Basisboek Methoden en Technieken* – voor meer informatie. Ook op Wikipedia vind je uitgebreide informatie over het trekken van verschillende soorten steekproeven ([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 die als het ware gerepresenteerd worden door de onderzoekseenheden die je geselecteerd hebt, dan gaat het om een *steekproefonderzoek*. Paragraaf 1.4 behandelt 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 dat ze met een bepaalde marge aan onzekerheid kunnen worden gegeneraliseerd naar de populatie waaruit de steekproef is getrokken.

OPDRACHT

- a Heb jij je gegevens verzameld bij een populatie of heb je een steekproef getrokken?
- b Is er bij jouw onderzoek dan sprake van inductieve of beschrijvende statistiek?

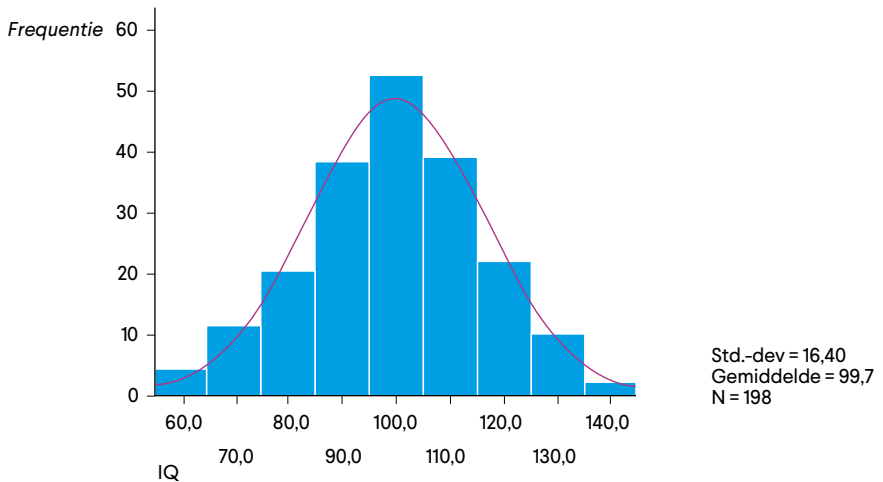
1.4 Enkele algemene statistische begrippen

Deze paragraaf gaat in op een aantal belangrijke begrippen die in de statistiek een rol spelen. Het gaat om de begrippen normaalverdeling, standaardfout, zekerheid of kans, significantie, relevantie, effectgrootte en vrijheidsgraden. 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 om 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 daardoor overzichtelijke vorm. Als je gegevens grafisch weergeeft, is het resultaat nogal eens de zogenoemde *normaalverdeling*. In figuur 1.3 zie je daarvan een (fictief) voorbeeld.

Normaalverdeling

Figuur 1.3

Het intelligentie-niveau 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 *de verdeling van Gauss* of *Gausskromme* genoemd. De karakteristieke kenmerken van de normaalverdeling zijn 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 ($N = 2213$), moet je 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. Het zal daarom bij elke steekproeftrekking iets anders zijn.

Wanneer je opnieuw een aselechte steekproef trekt en nog een keer (enzo voort), dan zal het gemiddelde IQ waarschijnlijk steeds iets hoger of mogelijk iets lager zijn. De gevonden waarden zullen iets van elkaar afwijken, maar waarschijnlijk niet veel.

Standaardfout

In Excel kun je de standaardfout laten berekenen (subparagraaf 4.2.2). De *standaardfout* geeft aan in hoeverre het gevonden steekproefgemiddelde een *betrouwbare* schatting van het populatiegemiddelde is. 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 minimaal 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 (zie tabel 1.2), wat betekent dat je niet veel kunt zeggen over de gemiddelde intelligentie in de populatie. Wanneer de steekproefgrootte 1.000 is, wordt het betrouwbaarheidsinterval een stuk kleiner, namelijk 98,76 - 101,24. Nu kun je wél zinnige uitspraken doen over de gemiddelde intelligentie van de populatie.

TABEL 1.2 Voorbeeld invloed spreiding en steekproefgrootte op het betrouwbaarheidsinterval

	Kleine steekproef, grote onderlinge verschillen	Kleine steekproef, kleine onderlinge verschillen	Grote steekproef, grote onderlinge verschillen	Grote steekproef, kleine onderlinge verschillen
Steekproef-gemiddelde	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

Je kunt het gemiddelde, de spreiding, de standaardfout en de betrouwbaarheidsintervallen berekenen door de data in te voeren in de tool GraphPad (www.graphpad.com/quickcalcs/CI1mean1.cfm).

Kans

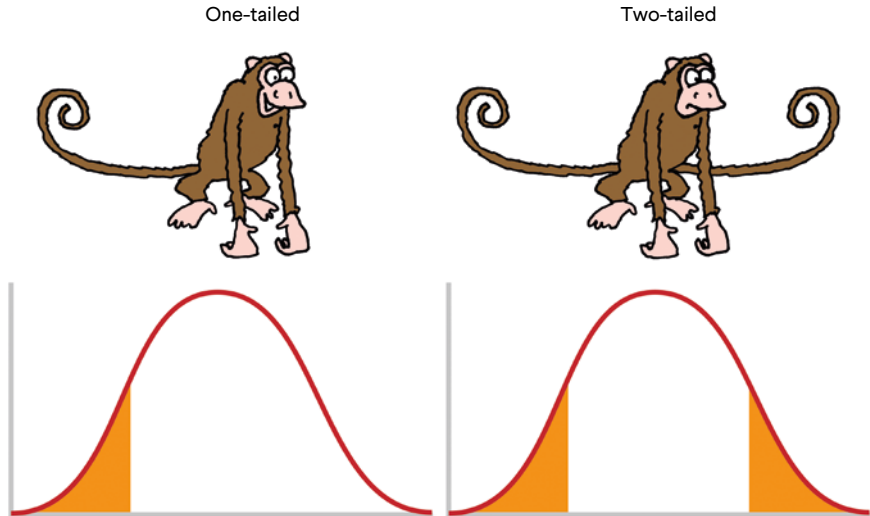
Het begrip *zekerheid* of *kans* speelt een belangrijke rol in de inductieve statistiek. Ook wanneer je de gemiddelden van twee steekproeven vergelijkt, is het de vraag wat de kans is dat je een gevonden verschil in gemiddelden terugvindt in de populatie. Stel dat de steekproef van werknemers van de 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 lees je 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*)

Significantie

Een- of tweezijdig toetsen

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*).



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, dan 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 2%, dus inderdaad kleiner dan 5%. In hoofdstuk 5 zie je hoe je dat berekent. Het verschil van vier IQ-punten is dan wel significant, maar niet erg relevant. Het zegt weinig over de verschillen tussen mannelijke en vrouwelijke werknemers. Je ziet tegenwoordig in onderzoeksverslagen steeds vaker dat naast de significantie een *effectgrootte* wordt vermeld. De gebruikelijke maat voor effectgrootte is *Cohen's d*. In het voorbeeld is *Cohen's d* 0,17 en dat is een te verwaarlozen effect. Pas als *d* groter is dan 0,20 wordt er van een klein effect gesproken. Het geslacht verklaart hier namelijk nog geen 1% van de verschillen in IQ. Uitleg over hoe je dat uitrekent vind je in subparagraaf 5.3.1.

Effectgrootte

Vrijheidsgraden

In de Excel-uitdraai kom je vaak de term *vrijheidsgraden* (*df = degrees of freedom*) tegen. Het aantal vrijheidsgraden geeft de mate aan waarin scores kunnen variëren. Als je van twee getallen er maar één 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, dan 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.

OPDRACHT

- a Voer een (beperkt) deel van je data voor een van de variabelen van minimaal intervalniveau (zoals leeftijd) in GraphPad in (www.graphpad.com/quickcalcs/CImean1.cfm).
- b Wat is in deze uitdraai het gemiddelde?
- c Wat is de spreiding?
- d Wat is de standaardfout?
- e Welke betrouwbaarheidsintervallen zijn er?
- f Hoe interpreteer je deze data aan de hand van het voorgaande? Wat valt jou op?

1.5 Hoe werkt Excel en hoe installeer ik statistische functies?

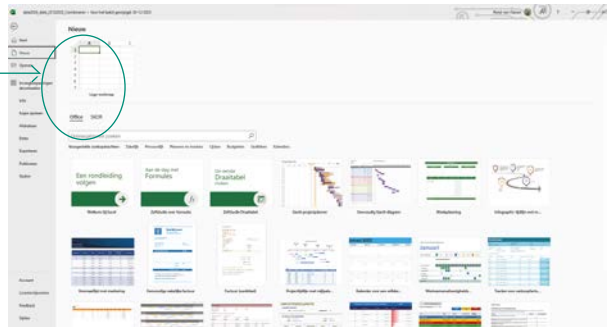
Zoals in paragraaf 1.1 werd aangegeven is Excel niet alleen een middel om data op te slaan en te ordenen. Het programma biedt ook veel mogelijkheden voor statistische analyse. Microsoft Office Excel werkt voor de meeste gebruikers via Microsoft Office 365. Bestanden kun je opslaan in jouw Cloud. Zorg in ieder geval voor een zorgvuldig versiebeheer van je bestanden, omdat gegevens overschreven kunnen worden wanneer je in de cloud werkt, en je de eerdere data dan 'kwijtraakt'. Open je een bestand en ga je erin werken, dan geef je het dus direct een volgend versienummer of een nieuwe datum mee.

Je start Excel door te dubbelklikken op het Excel-icoon. Je komt dan in een startmenu waarin Excel eerder geraadpleegde bestanden weergeeft (zie figuur 1.4). Dat is handig als je eenmaal bestanden hebt aangemaakt. In ons geval gaan we naar Nieuw om een nieuw databestand aan te maken. Verder zie je dat Excel allerlei voorbeelden van (*templates*) heeft klaarstaan. Er zijn bijvoorbeeld templates om een project te plannen of om een toernooischema te maken.

Figuur 1.4 Het openingsscherm van Excel

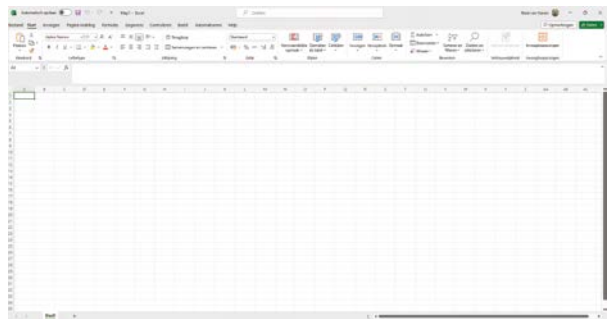
Stap 1

In Excel kom je onder **Nieuw** in een scherm met sjablonen en rondleidingen. Ook is er de mogelijkheid om een nieuwe lege werkmap van Excel te openen.



Stap 2

Klik op de lege werkmap. Je komt dan in een leeg data-invoerscherm.



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 menu is ingesteld op 'standaard', maar je kunt ook kiezen voor getal, valuta, datum of percentage, al naargelang het karakter van je gegevens. Voor de meeste statistische analyses is het handig om de *getalinstelling* te gebruiken.
- *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 hebt berekend of bewerkt. Wanneer je kiest voor twee verschillende celstijlen zie je direct het verschil.
- *Cellen*. Hiermee kun je cellen invoegen, verwijderen of opmaken. In een cel kun je bijvoorbeeld rekenkundige formules plaatsen. Je kunt zo twee cellen bij elkaar optellen en het resultaat daarvan komt dan in een derde cel. Het werken met formules komt in dit boek uitgebreid aan bod (zie hoofdstuk 3). In bijlage II zijn de te gebruiken commando's voor de formules uitgewerkt.
- *Bewerken*. In dit menu kun je gegevens zoeken, sorteren, bewerken (zoals een totaalscore uitrekenen) en verwijderen. Op de meeste functies komen we terug in latere hoofdstukken.

Enkele specifieke Excel-functies – zoals de functie Formules – komen in dit boek uitgebreid aan bod. Verschillende algemene functies – zoals knippen, plakken, kopiëren, kolommen aanpassen en ceileigenschappen veranderen – worden in hoofdstuk 2 behandeld.

Installeren van de statistische functies

De statistische mogelijkheden van Excel worden niet direct geïnstalleerd, maar zijn wel onderdeel van Excel. Deze statistische functies heb je later nodig wanneer je gaat indelen in klassen, samenhangen gaat analyseren, enzovoort. Het installeren van de statistische functies doe je zoals in figuur 1.5.

Figuur 1.5 Installeren van de statistische functies

Stap 1

Onder de knop **Bestand** vind je het commando **Opties**.

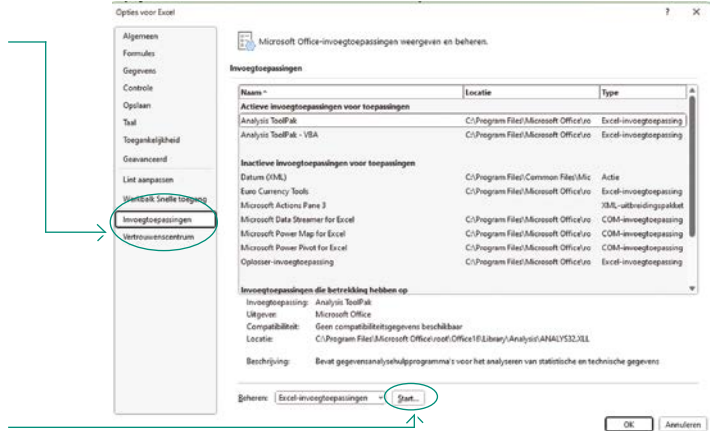
Misschien moet je in jouw menu eerst op **Meer** klikken om bij deze opties te komen.



1

Stap 2

Klik op **Opties**. Klik daarna in het menu dat verschijnt op **Invoegtoepassingen**.

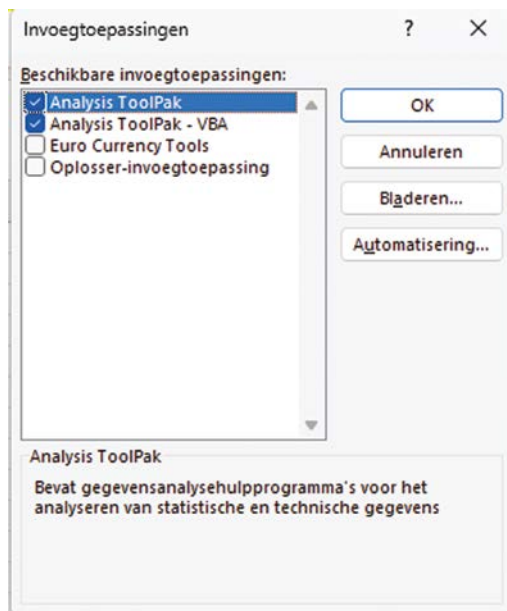


Stap 3

Klik onder aan het scherm bij **Beheren** (van de invoegtoepassingen) op **Start**.

Stap 4

Vink de invoegtoepassingen **Analysis ToolPak** en **Analysis ToolPak - VBA** aan.



Stap 5

Klik op **OK**.

Resultaat

Onder het menu **Gegevens** vind je rechts bovenaan een nieuwe functie, namelijk **Gegevensanalyse**. Deze functie was er in eerdere versies niet.



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

Over punten en komma's

We kunnen niet voor alle analyses gebruikmaken van Excel. Voor sommige analyses (vooral die in hoofdstuk 5) maken we gebruik van internettools, waaronder VassarStats (vassarstats.net). Op Statpages (statpages.info) vind je een overzicht van die programma's.

Er kan een probleem ontstaan bij het werken met een Nederlandse versie van Excel en het in combinatie met Engelstalige programma's zoals VassarStats. In Nederland gebruiken we een komma als decimaalteken, terwijl in Engelstalige gebieden – dus ook in de meeste online programma's – een punt als decimaalteken gebruikt wordt.

Een getal als 13.000 wordt bij knippen en plakken naar VassarStats gezien als het getal dertienduizend. Dit speelt alleen als je in Excel werkt met decimalen achter de gehele getallen. Het gaat hierbij om ordinale gegevens en gehele getallen. Als je geen decimalen hebt ingesteld (wat bij gehele getallen en ordinale gegevens ook niet gebruikelijk is), dan is het dus geen probleem om de gegevens te knippen en kopiëren naar VassarStats. Controleer dit.

Wil je dit probleem sowieso niet tegenkomen in berekeningen, dan is het verstandig om de decimaalkomma's in Excel te vervangen door punten (zie figuur 1.6).

Figuur 1.6 Omzetten van scheidingstekens

Stap 1

Kies binnen Excel voor het menu **Bestand**. Kies vervolgens onderaan voor **Opties**.

Kies daarna voor **Geavanceerd**. Zet het vinkje bij **Scheidingstekens van systeem gebruiken** uit.



In hoofdstuk 5 maak je gebruik van online programma's en kopieer je de data naar een internettool. Het is dan noodzakelijk om te controleren of je op de juiste wijze werkt met de scheidingstekens. Voor de eerste hoofdstukken heb je deze optie niet nodig.

OPDRACHT

Doorloop de stappen uit figuur 1.5 en controleer of de functie Gegevensanalyse aan staat (in het menu Gegevens).

1.6 Hoe werk ik met tabellen in Excel?

Een van de belangrijkste functies om data te kunnen analyseren en beschrijven is werken met de functie Tabel. Deze functie maakt namelijk verschillende eigenschappen mogelijk. Een voordeel is dat de rijen en de kolommen aan elkaar gekoppeld worden. Sorteren heeft dan geen invloed op de volgorde van de andere gegevens. Dat is niet het geval als de gegevens *niet* in een tabel staan. Het sorteren van gegevens van hoog naar laag voor de ene variabele leidt er dan toe dat de gegevens van de respondenten niet meer onderling gekoppeld zijn. Let hier dus op; ga pas sorteren als de gegevens in een tabel staan.

Excel-tabel

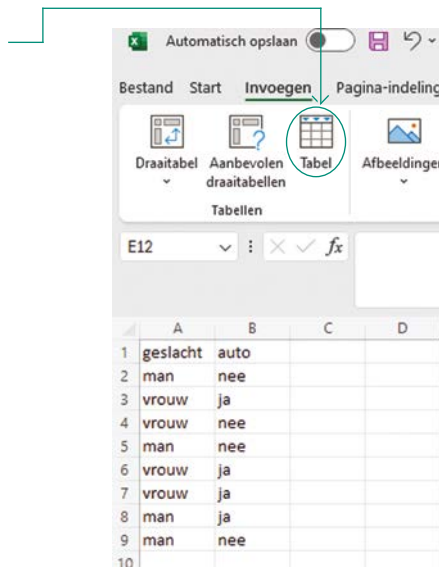
Een tweede voordeel van het omzetten van een databestand in een *Excel-tabel* is dat bewerkingen en analyses dan automatisch voor alle cellen in een kolom worden uitgevoerd. Wanneer in de kolom *leeftijd* het geboortjaar staat en je rekt dit om naar leeftijd in de eerste cel van de kolom, dan doet de Excel-tabel dat automatisch ook voor de andere cellen in de kolom leeftijd. Ook het berekenen van totalen, gemiddelden en dergelijke gaat zo een stuk gemakkelijker. Wij komen hierop terug bij het bespreken van bijvoorbeeld het gemiddelde (in hoofdstuk 4).

Om een voorbeeld te geven werken we met een eenvoudige datamatrix, waarin gegevens over het geslacht van een beperkt aantal respondenten gekoppeld zijn aan het beschikken over een auto. Het installeren van een Excel-tabel doe je zoals in figuur 1.7.

Figuur 1.7 Installeren van een Excel-tabel

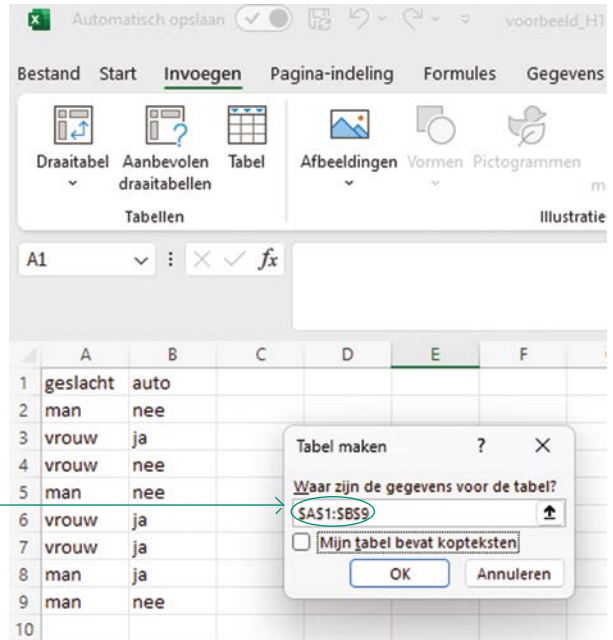
Stap 1

Onder het menu **Invoegen** vind je drie functies. Kies voor **Tabel**.



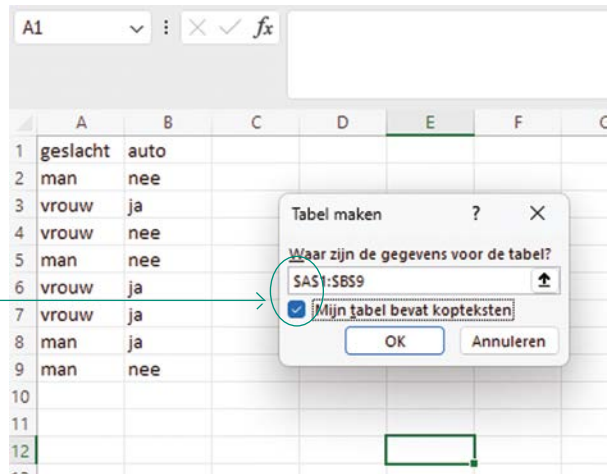
Stap 2

Er verschijnt een toepassingsmenu. Selecteer hierin met je muis de gegevens waarvoor je de functie **Tabel** wilt gebruiken. In dit geval gaat het om de cellen van A1:B9. Je kunt dit ook handmatig invoeren.

**Stap 3**

Vink aan dat de tabel (rij 1) in dit geval de namen van de variabelen bevat (**Mijn tabel bevat kopteksten**).

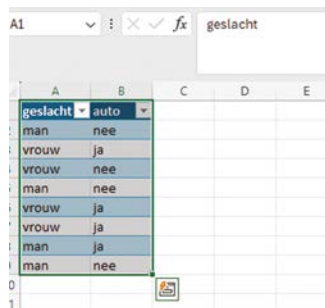
De vraag of de eerste rij de namen van de variabelen (kopteksten) bevat komt regelmatig terug binnen Excel. Let daar dus op.

**Stap 4**

Klik op **OK**.

Resultaat

Je ziet dat er een tabel is ontstaan waarbij de gegevens gekoppeld zijn. De driehoekjes bij de variabelen kun je gebruiken om gegevens te sorteren.



Je ziet in figuur 1.7 dat de kolomtitels gemarkeerd zijn. Ook blijven de kolomtitels in beeld als je naar beneden scrolt. Verder zie je naast de namen van de variabelen een omgekeerd zwart driehoekje. Als je daarop klikt, kom je in een menu dat de mogelijkheid biedt om je gegevens te sorteren, zoals bij leeftijd van laag naar hoog. Je kunt hiermee ook gegevens selecteren, bijvoorbeeld als je alleen de gegevens van de vrouwen wilt hebben.

Een voordeel van deze Excel-tabel is dat bij het creëren van nieuwe variabelen en het daarbij gebruiken van formules Excel de gewenste berekening direct doorvoert in de cellen (zie hoofdstuk 3). Je hoeft dan geen formules meer te kopiëren en te plakken. De Excel-tabel is een *voorwaarde* voor het maken van draaitabellen waarbij je overzichten kunt maken (zie hoofdstuk 4).

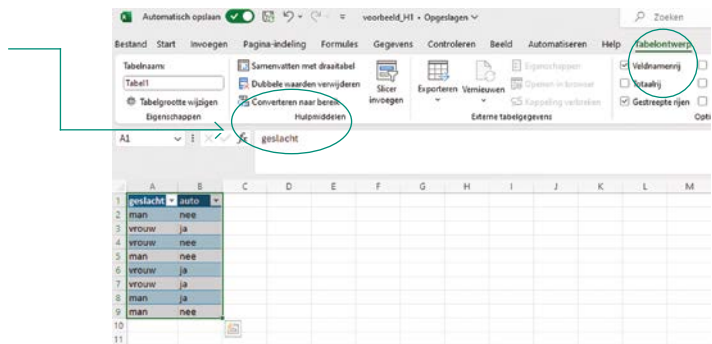
Opheffen van de Excel-tabel

Wil je aanvullende gegevens invoeren of is je databestand nog niet compleet, dan kan het handig zijn om de functie Tabel tijdelijk op te heffen. Dit doe je door in het menu Tabelontwerp te kiezen voor Converteren naar bereik (zie figuur 1.8).

Figuur 1.8 Opheffen van de Excel-tabel

Stap 1

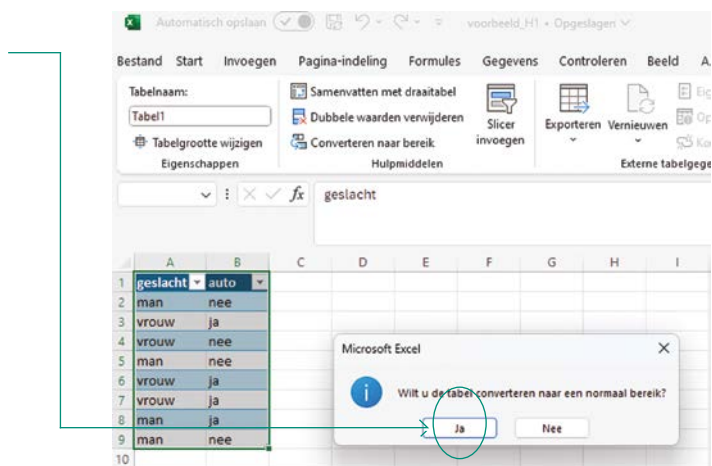
Kies in het menu **Tabelontwerp** voor **Converteren naar bereik**.



Stap 2

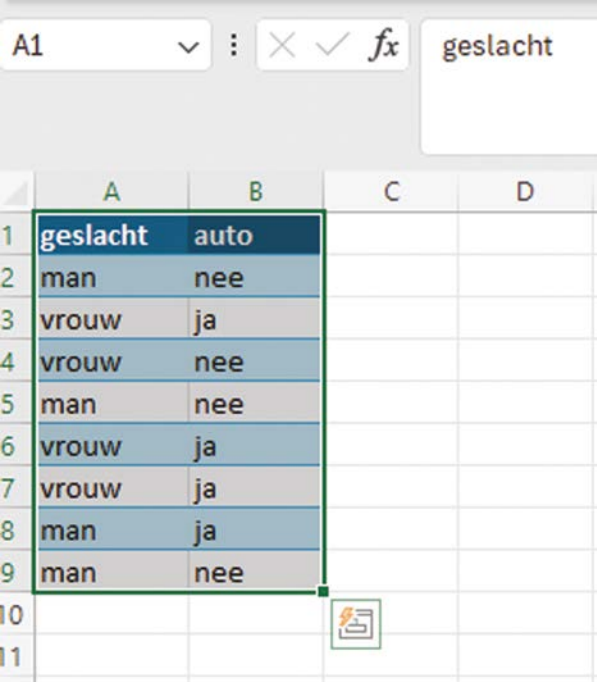
Er verschijnt een toepasingsmenu met de vraag 'Wilt u de tabel converteren naar een normaal bereik?'

Klik op **JA**.



Resultaat

Je ziet de gegevens waarin de eerdere opmaak is blijven staan, maar de gegevens zijn niet meer aan elkaar gekoppeld. Het is nu een 'normale' datamatrix. De driehoekjes bij de variabelen *geslacht* en *auto* zijn verdwenen.



	A	B	C	D
1	geslacht	auto		
2	man	nee		
3	vrouw	ja		
4	vrouw	nee		
5	man	nee		
6	vrouw	ja		
7	vrouw	ja		
8	man	ja		
9	man	nee		
10				
11				

OPDRACHT

Oefen met het maken van een tabel en het opheffen van de tabel. In de volgende hoofdstukken ga je veelvuldig met deze functie aan de slag om gegevens te sorteren, berekeningen te maken of overzichten te maken voor de analyse via de functie Draaitabel.

Samenvatting

- Zet Excel in voor het analyseren van gegevens. Installeer daarvoor de functie Gegevensanalyse via Analysis Toolpack.
- Maak een tabel met daarin de verschillende variabelen en het meetniveau van de variabelen. Dit heeft rekenkundige consequenties voor de analyse van de gegevens.
- Zorg ervoor dat je kennis hebt van belangrijke begrippen, zoals de verdelingen van gegevens (normaal versus scheef) en begrippen als spreiding van de gegevens, standaardfout, kans en zekerheid, significantie, relevantie, effectgrootte en vrijheidsgraden. Deze begrippen spelen een belangrijke rol bij het komen tot en het interpreteren van uitkomsten uit de verzamelde data.
- Werk met de functie Tabel in Excel. Gegevens blijven dan gekoppeld aan de respondentnummers, en bewerkingen zoals formules zijn eenvoudiger in te voeren.