

KG-publicatie nr. 18

Eenvoudige statistische
berekeningen met behulp van
Excel

Huub Everaert en Arie van Peet

KG-18

KG-publicatie nr. 18

Eenvoudige statistische berekeningen met behulp van Excel

Huub Everaert en Arie van Peet

H. Everaert en A. van Peet

Eenvoudige statistische berekeningen met behulp van Excel.

Correspondentie over deze KG-publicatie kunt u sturen naar:

E-mail: huub.everaert@hu.nl

KG-publicaties bevatten interne notities, verslagen en prepublicaties bestemd voor spreiding in kleine kring. De verantwoordelijkheid voor de inhoud van een KG-publicatie berust bij de auteur(s). Uit KG-publicaties mag alleen geciteerd worden met toestemming van de auteurs.

-
- nr.1 Everaert, H.A. en J.C. van der Wolf (2005).
Behaviorally Challenging Students and Teacher Stress.
-
- nr.2 Wolf, J.C. van der & J.M.F. Touw (2005).
Onderzoek naar zorg in het curriculum van de Theo Thijssen Academie.
-
- nr.3 Doorn, E.C. van (2005).
De Gedragingenlijst voor Leraren (Tweede onderzoeksrapport).
-
- nr.4 Wolf, J.C. van der (2005).
Probleemouders en de school: een onderontwikkeld terrein.
-
- nr.5 Enthoven, M. (2005).
The contribution of the school environment to youths' resilience: A Dutch middle-adolescent perspective.
-
- nr.6 Enthoven, M.; A.C. Bouwer; J.C. Van der Wolf & A. van Peet (2005).
Recognizing Resilience: Development and Validation of an Instrument to Recognize Resilience in Dutch Middle-Adolescents.
-
- nr.7 Velderman, H & H.A. Everaert (2005).
Time-out or switch? (Paper presented at the ECER conference On 9 September 2005, University College Dublin).
-
- nr.8 Touw, J.M.F., J.T.E. van Beukering & H.A. Everaert (2005).
Teachers' Personal Constructs on Problem Behaviour (Paper presented at the annual meeting of the European Educational Research Association (EERA), Dublin, Ireland, September 7-10).
-
- nr.9 Doorn, E.C. van (2005).
Levend leren: daar ga ik voor!
-
- nr.10 J.T.E. van Beukering, J.M.F. Touw & H. Everaert (2005).
Teachers' personal constructs on problem behaviour: towards professional development
& Kos, P. (2005). Personal constructs on (problem) pupils: a teacher's view.
2 Papers presented at the International Practitioner Research Conference & Collaborative Action Research Network Conference (PRAR 2005), Utrecht, The Netherlands, November 4-6, 2005.
-
- nr. 11 Everaert, H. & A. van Peet (2006). Kwalitatief en kwantitatief onderzoek.
-
- nr. 12 Everaert, H.A. & J.C. van der Wolf (2006).
A Comparison of Stress and Burnout between Dutch General and Special Education Teachers. Paper presented at the annual meeting of the American Educational Research Association (AERA), San Francisco, USA, April 7-11, 2006.
-
- nr. 13 Everaert, H.A. & J.C. van der Wolf (2006). Gender Perceptions of Challenging Student Behavior and Teacher Stress.
-
- nr. 14 Peet, A.A.J. van (2006). Schaalconstructie.
-

nr. 15 Dellevoet, S.M.E., Beukering, J.T.E. van, Everaert, H.A. & Touw, J.M.F. (2006).
Evaluatie van de methode Under Construction op Instituut Theo Thijssen.

nr. 16 Peet, A.A.J. van (2006). Q-sort. Een rangordening.

nr. 17 Everaert, H.A. (2007). Measuring challenging student behavior. An overview of
methodological properties and decisions.

nr. 18 Everaert, H.A. & Peet, A.A.J. van (2007). Eenvoudige statistische berekeningen met
behulp van Excel.

Inhoudsopgave

1	<i>Inleiding</i>	7
2	<i>De grote rekenmachine</i>	9
2.1	Optellen en delen	9
2.2	Referentieel en absoluut verwijzen	16
2.3	Functies en haakjes	21
2.4	Een paar rekenvoorbeelden.....	23
3	<i>De grote grafische rekenmachine</i>	26
3.1	Frequentietabellen.....	26
3.2	Grafieken.....	32
4	<i>Kruistabellen</i>	41
5.	<i>Rekenschema's</i>	50
5.1	Gemiddelde, variantie en standaarddeviatie	50
5.2	Covariantie en correlatiecoëfficiënt.....	58
5.3	Regressiecoëfficiënten	62
6	<i>Nawoord</i>	69
	<i>Bijlage 1 Literatuur</i>	70

1 Inleiding

In de loop der tijd hebben wij op zowel universiteits- als hogeschoolniveau studenten en collega's zien worstelen en soms ook verdrinken in het computerprogramma SPSS. Daar zijn vele oorzaken voor aan te wijzen. Dat gaan we hier niet doen. Wel willen we hier wijzen op een wijdverbreid misverstand. In tegenstelling tot wat velen denken: SPSS is niet synoniem met methodologie of statistiek. SPSS is een computerprogramma in BASIC toegesneden op het maken van statistische analyses en er zijn veel meer van dat soort programma's op de markt. Denk bijvoorbeeld aan Stata, S-plus, R, SAS, EQS, Lisrel, MlwiN, Mplus en HLM. In onze ogen zijn dat allemaal goede statistiekprogramma's en met een aantal van deze programma's werken we binnen de kenniskring 'Gedragsproblemen in de onderwijspraktijk' bijna dagelijks én ook nog eens met veel plezier.

In het document Koers 2012 van de Hogeschool Utrecht wordt gesproken over de verdere ontwikkeling van de hogeschool tot een professionele kennisorganisatie. Het gaat daarbij om kennis die bruikbaar is, voortkomt uit maatschappelijke behoeften, gericht is op de professionalisering van de beroepspraktijk en voldoet aan internationaal gangbare methodische vereisten. Dit onderschrijven wij van harte. Er zijn talloze doorwrochte kwalitatieve en kwantitatieve methoden voorhanden om de klas en de school te bestuderen. Er zijn ook veel computerprogramma's die daarbij een rol kunnen spelen, ware het niet dat voor het overgrote deel van onze studenten en een deel van collega's deze programma's te omvangrijk of te zwaar zijn. Behalve de specialistische kennis die nodig is om bovengenoemde software zinvol te gebruiken, zijn deze programma's in (privé)aanschaf relatief duur. Het is bovendien de vraag hoeveel afgestudeerde studenten aan de Faculteit Educatie ooit nog met SPSS in hun beroepspraktijk zullen werken, terwijl het toch handig zou zijn om zelfstandig analyses uit te kunnen voeren. Kortom, we zijn op zoek naar het juiste gereedschap om de gewenste kennisontwikkeling en kenniscirculatie in onze beroepspraktijk te verdiepen.

Wij hebben dan ook het plan opgevat om studenten en collega's te wijzen op de statistische mogelijkheden die het programma Excel biedt. Nagenoeg iedereen op de hogeschool heeft weliswaar Excel op de PC staan, maar niet iedereen gebruikt dit programma even efficiënt. De reken capaciteit blijft grotendeels onbenut. Een tweede en tevens didactisch voordeel van Excel is de mogelijkheid om de link tussen afzonderlijke getallen en statistische kengetallen inzichtelijk te maken. We zijn van mening dat dit het begrip van statistiek ten goede kan komen.

Alle in dit KG uitgewerkte voorbeelden zijn gemaakt in Microsoft Office Excel 2003. De KG-publicatie 'Eenvoudige statistische berekeningen met behulp van Excel' is geschreven voor studenten en docenten die statistiekonderwijs volgen of hebben gevolgd. Statistische begrippen worden als zodanig ook niet verder uitgewerkt. Er zijn veel goede Nederlandstalige boeken op de markt die hierin voorzien. Het gaat er ons om dat voor heel veel statistiek geen nieuwe of aparte software hoeft te worden aangeschaft en bestudeerd, maar dat u meteen in Excel aan de slag kunt.

De KG-publicatie 'Eenvoudige statistische berekeningen met behulp van Excel' is een eerste poging om het nog te verschijnen onderzoeksboek '*Klassenonderzoek? Dat doe je zó!*' met voorbeelden in Excel te onderbouwen. Ook de opgaven in het bijbehorende opgavenboek zullen allemaal in Microsoft Office Excel 2003 (of een eventueel nieuwere versie) nader worden uitgewerkt. U begrijpt dat we bijzonder geïnteresseerd zijn in uw didactische suggesties en opmerkingen met betrekking tot deze KG-publicatie.

2 De grote rekenmachine

In dit hoofdstuk worden de bouwstenen aangereikt om te kunnen rekenen in Excel. In eerste instantie doen we dit door eenvoudigweg naar cellen te verwijzen en de inhoud van deze cellen bij elkaar op te tellen. Gaandeweg worden de berekeningen wat gecompliceerder. U maakt tevens kennis met een van de belangrijkste principes van Excel: referentieel verwijzen naar cellen. Dit gebeurt standaard bij het kopiëren en plakken van cellen. Ook de tegenhanger van referentieel verwijzen wordt besproken: absoluut verwijzen naar cellen. Twee belangrijke rekenfuncties (sommen en middelen) worden op het laatst geïntroduceerd. We staan stil bij het gebruik van ronde haakjes en het hoofdstuk sluit af met een aantal rekenvoorbeelden die ontleend zijn aan het binnenkort te verschijnen boek *Klassenonderzoek? Dat doe je zó!*.

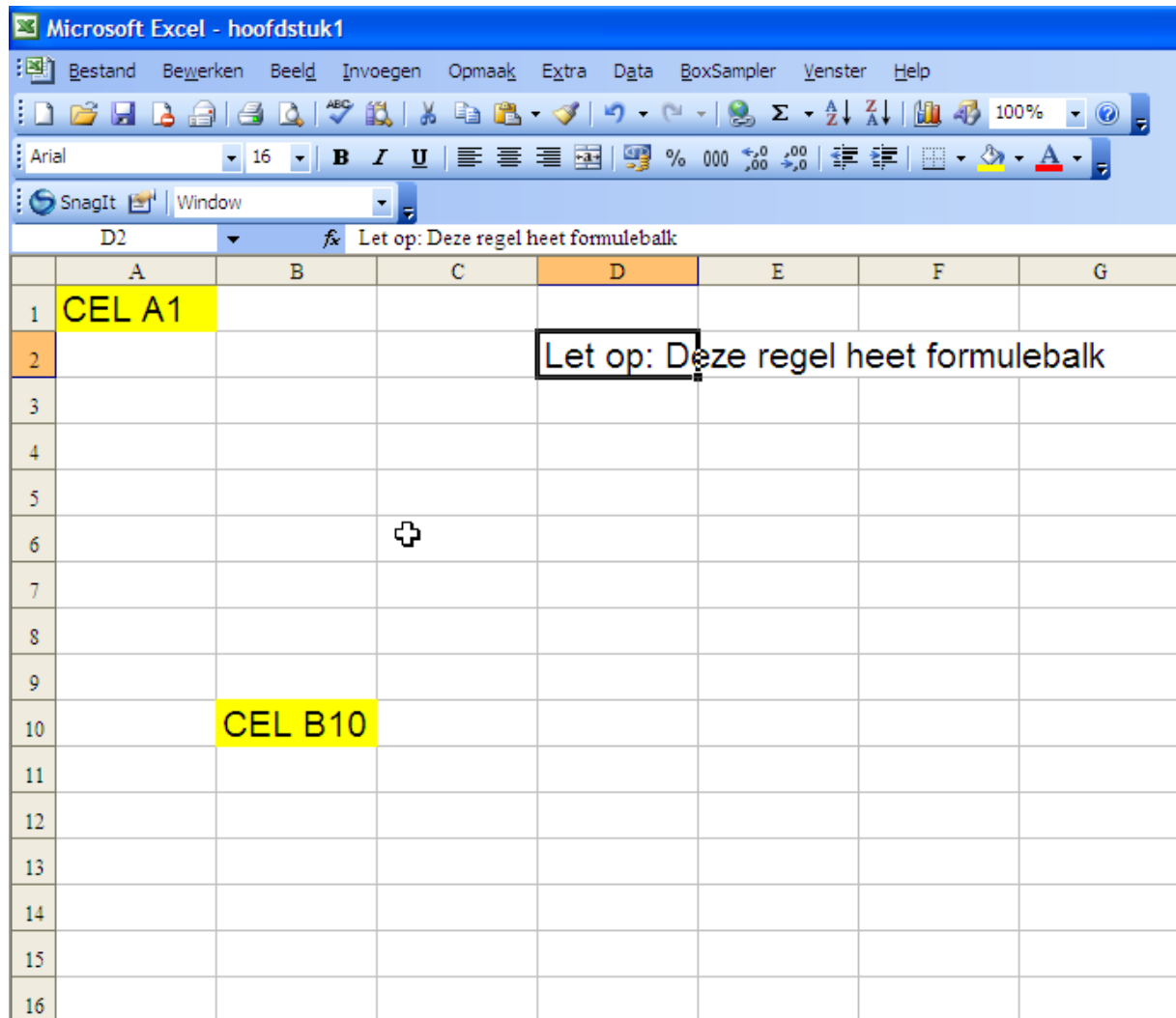
2.1 Optellen en delen

Wie het programma Excel opent, krijgt een raster van verticale en horizontale lijnen te zien op blad 1. De ruimte tussen de verticale lijnen worden kolommen genoemd en aangeduid met letters. De eerste kolom heet kolom A, de 26^e kolom noemen we Z en de 27^e kolom heet AA. De laatste kolom op uw blad wordt aangeduid met de letters IV. Per blad heeft u standaard 256 kolommen tot uw beschikking.

De ruimte tussen de horizontale lijnen noemen we in het vervolg rijen en worden genummerd van 1 tot en met 65536. Dat zal voor de meeste van ons ruim voldoende blijken te zijn. U kunt natuurlijk ook altijd op een ander blad dan blad 1 gaan werken. De naam van het blad waarop u werkt, is links onderaan op het scherm weergegeven. In dit geval is dat dus '**Blad1**'.

Het vakje dat als het ware op het snijpunt ligt van een rij en een kolom noemen we een cel. Cel A1 is de naam van het bovenste hokje in kolom A op rij 1. Cel B10 betreft het tiende vakje in kolom B. Beide cellen zijn aangeduid met de kleur geel in Figuur 2.1. U kunt de celinhoud kleuren door met de cursor in de cel te gaan staan en vervolgens op de werkbalk Opmaak de opvulleur geel te kiezen. U kunt ook eerst de kleur kiezen in de werkbalk Opmaak en vervolgens de cellen A1 en B10 selecteren. In de loop van deze KG-publicatie zult u ondervinden dat in Excel veel wegen naar Rome leiden.

In de cel D2 is de tekst 'Deze regel heet formulebalk' getypt. U vindt deze tekst ook in de formulebalk zelf terug. Links van de formulebalk is de plaats van de cursor aangegeven. Dat is dus cel D2. Dit alles is samengevat in Figuur 2.1.

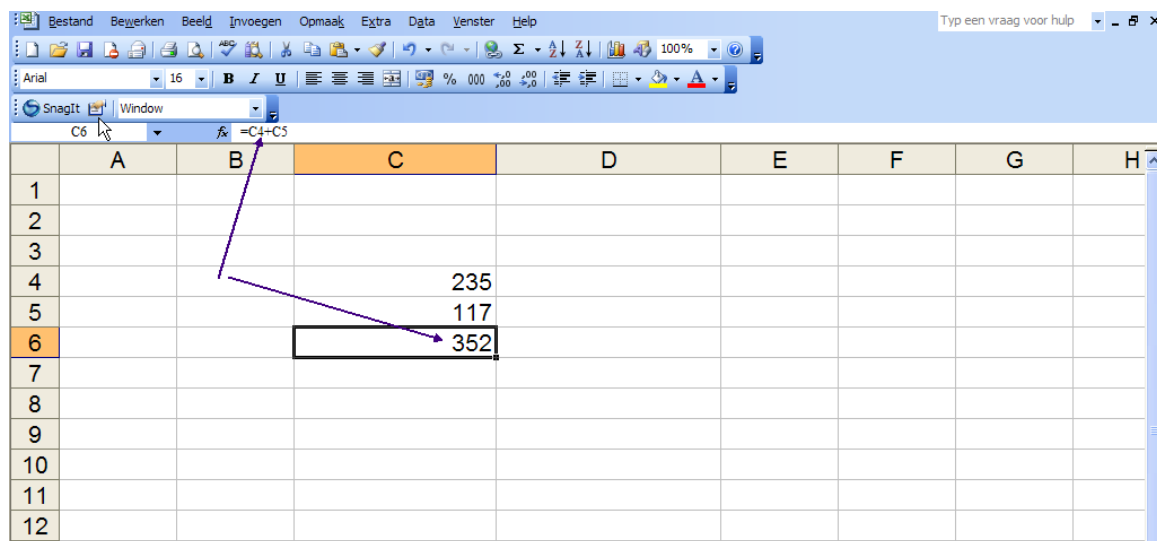


Figuur 2.1. De velden A1, B10, D2 op Blad1 met de aanduiding formulebalk.

In dit hoofdstuk gaan we Excel gebruiken als een uit de kluiten gewassen rekenmachine. Stel, u wilt weten hoeveel $235 + 117$ is. Beide getallen typt u in Excel onder elkaar in 2 verschillende cellen. Bijvoorbeeld in cel C4 typt u '235' en in cel C5 het getal '117'. Vervolgens gaat u met de cursor in C6 staan en typt u '=C4+C5', gevolgd door een harde return. Let op, we gebruiken twee apostrofjes om aan te geven wat u precies in de cel typt. De apostrofjes zelf typt u dus *niet* in.¹ In

¹ Als u toch de apostrofjes heeft ingetypt, herkent Excel de ingetypte karakters niet meer als een getal. De celinhoud wordt dan als tekst geïnterpreteerd door het computerprogramma. Het isgelijkteken moet u beslist intypen. Als u dit niet doet, verschijnt de tekst 'C4+C5'.

cel C6 verschijnt dan het getal '352'. In de formulebalk zelf ziet u echter niet het getal staan, maar de formule die u zojuist hebt ingetypt, namelijk '=C4+C5'. Eigenlijk gaf u niet de opdracht om 235 en 117 bij elkaar op te tellen, maar om de inhoud van cel C4 en de inhoud van cel C5 bij elkaar op te tellen (zie Figuur 2.2). Als u nu in C4 de inhoud '235' verandert in '2350', dan verschijnt automatisch in C6 het getal '2467' ($=2350+117$).



Figuur 2.2. Tel de inhoud van de cellen C4 en C5 bij elkaar op in cel C6.

Behalve optellen kunt u natuurlijk ook aftrekken, vermenigvuldigen en delen. De operatoren die u hiervoor gebruikt zijn achtereenvolgens: + (optellen), - (aftrekken), * (vermenigvuldigen) en / (delen). Dus als u nu 235×117 wilt berekenen, gaat u met uw cursor in C6 staan en typt u '=C4*C5'. Dat blijkt 27495 te zijn. U kunt ook denken, ik wil in veld C6 de som van de cellen C4 en C5 en in veld C7 het product van deze twee cellen. U typt daartoe in cel C6 '=C4+C5' en in cel C7 '=C4*C5'.

Behalve de som en het product van de getallen 235 en 117, wilt u ook de som en het product van de getallen 543 en 125 berekenen. U typt nu in cel D4 '543' en in cel D5 '125'. Vervolgens selecteert u de cellen C6 en C7 met uw cursor en kopieert de inhoud met de toetscombinatie <Ctrl>+<C>. Daarna selecteert u de cellen D6 en D7 gevolgd door <Ctrl>+<V>. De getallen 668 en 67875 verschijnen in de cellen D6 en D7. Het getal 668 is inderdaad de som van de getallen 543 en 125. Als u nu met de cursor in cel D6 gaat staan, dan ziet u in de formulebalk niet staan

'=C4+C5' (hetgeen u misschien wel dacht te kopiëren én te plakken), maar '=D4+D5' (zie Figuur 2.3).

	A	B	C	D	E	F
2						
3						
4			235	543		
5			117	125		
6			352	668		
7			41184	83500		
8						
9						
10						
11						
12						

Figuur 2.3. Referentieel kopiëren en plakken van cellen.

In Figuur 2.3 hebben we te maken met een zogenaamde referentiële verwijzing. Dat is de standaardmethode kopiëren en plakken in Excel. Het is niet alleen de standaard manier van werken, het is ook de kracht van Excel. U heeft zojuist misschien niet zozeer de inhoud van cel C6 gekopieerd, maar eigenlijk de verwijzing naar de som van de cellen C4 en C5. Als u nu drie plekje's naar rechts was gegaan en vervolgens twee omhoog, dan gaat ook uw gekopieerde celverwijzing mee naar drie plekje's rechts en twee omhoog. Dus onder uw kopieerknop verandert de verwijzing in de som van de cellen F2 en F3. Dat is de betekenis van referentieel. Dit is niet alleen een belangrijk principe in Excel, het is vooral ook een handig principe. U kunt natuurlijk ook absoluut verwijzen. Dit doet u door de celverwijzing als het ware vast te zetten. Dit kunt u doen met behulp van het teken \$. We komen hier zo meteen op terug.

Met behulp van de basisbegrippen rij, kolom, cel, formulebalk, operatoren en referentiële verwijzing kunt u naar hartenlust rekenen in Excel. Hieronder volgt een voorbeeld van een docent die de somscores van de toetsscores van 12 leerlingen wil uitrekenen op drie verschillende toetsen (zie Figuur 2.4). Aangezien de Figuren 2.1 tot en met 2.3 wel erg veel ruimte in beslag nemen, worden de resterende Figuren in deze KG-publicatie anders weergegeven. U ziet in deze

figuren niet altijd meer de namen van rijen en kolommen in de kantlijnen als zodanig genoemd. In plaats daarvan wordt steeds gewerkt vanuit een gele (of in zwartwit print, grijze) cel. In Figuur 2.4 is de cel G10 als referentiepunt gekozen. Zo kunt u meteen oefenen in het denken in celverwijzingen.

	toets 1	toets 2	toets 3	CEL G10
leerling 1	4	7	6	
leerling 2	6	8	6	
leerling 3	7	7	6	
leerling 4	8	8	8	
leerling 5	6	8	8	
leerling 6	5	8	7	
leerling 7	6	6	5	
leerling 8	7	7	6	
leerling 9	6	7	9	
leerling 10	6	6	9	
leerling 11	5	8	7	
leerling 12	9	8	10	

Figuur 2.4. Input voor het bereken somscores van twaalf individuele leerlingen over drie toetsen.

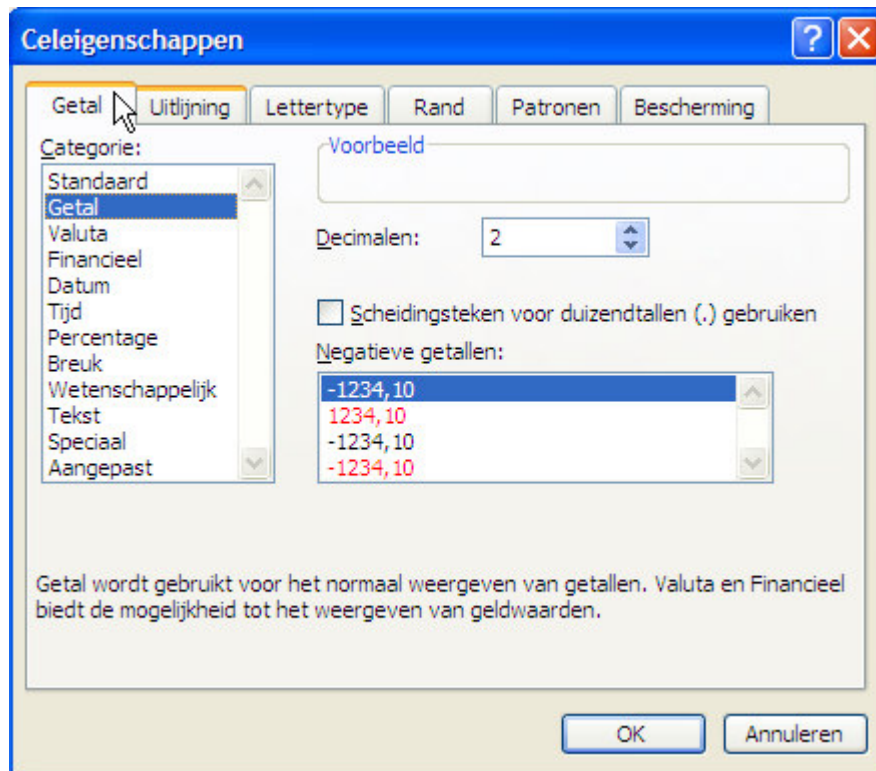
Stel u heeft als docent van elke leerling drie toetsen in Excel opgeslagen en u wilt de somscores op deze toetsen van deze leerlingen uitrekenen (zie Figuur 2.4). U gaat dan in cel G11 staan met uw cursor en typt daarin '=D11+E11+F11' gevolgd door een harde return. Het getal 17 verschijnt in cel G11. U selecteert daarna cel G11 gevolgd door <Ctrl>+<C> en vervolgens selecteert u de cellen G12 tot en met G22 gevolgd door <Ctrl>+<V>. In elke cel verschijnt nu van iedere leerling de som van deze drie toetsen. U bent natuurlijk niet zozeer geïnteresseerd in de som van deze toetsen als wel de gemiddelde score die de leerlingen behaald hebben. Iedere leerling heeft drie toetsen gemaakt, dus voor het gemiddelde delen we somscore door 3. Dat doen we eerst weer

even handmatig in cel H11. U typt daarin '=G11/3'. Immers, in cel G11 staat de som op de drie toetsen van leerling 1. Vervolgens kopieert u cel H11 en plakt deze in de cellen H12 tot en met H22. Dat levert u het volgende plaatje op (Figuur 2.5).

	toets 1	toets 2	toets 3	CEL G10	
leerling 1	4	7	6	17	5,666667
leerling 2	6	8	6	20	6,666667
leerling 3	7	9	6	22	7,333333
leerling 4	8	8	8	24	8
leerling 5	6	8	8	22	7,333333
leerling 6	5	8	7	20	6,666667
leerling 7	6	6	5	17	5,666667
leerling 8	7	7	6	20	6,666667
leerling 9	6	7	9	22	7,333333
leerling 10	6	6	9	21	7
leerling 11	5	8	7	20	6,666667
leerling 12	9	8	10	27	9

Figuur 2.5. Berekenen van gemiddelde score op drie toetsen.

Of u de handelingen juist heeft uitgevoerd, kunt u zien door naar leerling 4 (drie keer een 8, dus het gemiddelde moet 8 zijn) en leerling 12 ($[(9+8+10)/3=9]$), te kijken. Gemiddelde scores van 7,33333 of 6,66667 op basis van drie toetsen hebben niet veel zin. Het suggereert zeer nauwkeurige metingen en bovendien gaat u niet op een ouderavond vertellen dat een kind een rapportcijfer heeft op 6 decimalen nauwkeurig. Er moet afgerond worden en in dit voorbeeld doen we dat op één decimaal nauwkeurig, dus op één cijfer achter de komma. Dit kan gemakkelijk in Excel. U selecteert daartoe de cellen H11 tot en met H22, en klikt in de menubalk op Opmaak. Er verschijnt een dialoogscherm met een aantal tabbladen; u kiest in dit geval het eerste tabblad genaamd Getal, vervolgens selecteert u uit de keuzelijst het woord Getal (zie Figuur 2.6).



Figuur 2.6. Dialoogscherm voor het afronden van getallen.

In het vakje Decimalen vervangt u het cijfer 2 door 1, wederom gevolgd door een harde return. U kunt ook op de knop OK klikken. Alle getallen worden nu keurig op een decimaal afgerond. Voor later is het handig dat u eenvoudig kunt zien wat de getallen in dit werkblad voorstellen. In cel G10 typt u dan 'Somscore' en in cel H10 'Toetsgemiddelde'. Het getal 9 in cel E13 van Figuur 2.5 is helaas niet goed ingevuld. Dat moet een 7 zijn. U vervangt de 9 door een 7. Het gemiddelde wordt automatisch opnieuw berekend door Excel (zie Figuur 2.7).

		+				
					CEL G9	
		toets 1	toets 2	toets 3	Somscore	Toetsgemiddelde
leerling 1	4	7	6	17	5,7	
leerling 2	6	8	6	20	6,7	
leerling 3	7	7	6	20	6,7	
leerling 4	8	8	8	24	8,0	
leerling 5	6	8	8	22	7,3	
leerling 6	5	8	7	20	6,7	
leerling 7	6	6	5	17	5,7	
leerling 8	7	7	6	20	6,7	
leerling 9	6	7	9	22	7,3	
leerling 10	6	6	9	21	7,0	
leerling 11	5	8	7	20	6,7	
leerling 12	9	8	10	27	9,0	

Figuur 2.7. Berekening toetsgemiddelde na het corrigeren van cel E13.

Het kan zijn dat niet alles even recht en mooi gecentreerd staat zoals in Figuur 2.7. Ook de referentie cel is een cel naar boven verschoven. U kunt dit zelf aanpassen door de juiste opties te kiezen: of u klikt in de menubalk op Opmaak, of u werkt vanuit de Opmaak werkbalk.

2.2 Referentieel en absoluut verwijzen

We hebben zojuist betoogd dat de standaard manier van verwijzen in Excel referentieel is. In deze paragraaf gaan we dit nog eens illustreren door het invoegen van een kolom. Essentieel hierbij is de plaats waar de kolom in het werkblad wordt ingevoegd. De plaats waar de nieuwe kolom wordt ingevoegd, bepaalt namelijk of en hoe eventuele referentiële verwijzingen op uw werkblad mee veranderen. Dezelfde spelregels gelden ook bij het invoegen van nieuwe rijen. Hieronder zullen we dit op twee manieren voor het voetlicht brengen. Beide voorbeelden sluiten we af met de tegenhanger van referentieel verwijzen: absoluut verwijzen.

Het blijkt dat de docent inmiddels een vierde toets heeft afgenomen. Na een avondje nakijken staan de resultaten in Figuur 2.8.

		toets 4	
	leerling 1	6	
	leerling 2	6	
	leerling 3	5	
	leerling 4	4	
	leerling 5	5	
	leerling 6	6	
	leerling 7	4	
	leerling 8	5	
	leerling 9	6	
	leerling 10	8	
	leerling 11	7	
	leerling 12	7	

Figuur 2.8. Toetsresultaten van de vierde toets.

Na het nakijken van deze toets moet u dit nog even in uw rekenschema in Excel zetten. U kunt daarbij als volgt te werk gaan (Let wel op de juiste volgorde van de leerlingen!). U neemt de tabel voor u en selecteert kolom G in zijn geheel. Vervolgens klikt u in het menu op Invoegen gevolgd door Kolommen. Er verschijnt een nieuwe kolom G. De oude kolom G verandert automatisch in kolom H en kolom H verandert in kolom I. In de lege kolom G typt u de resultaten van de vierde toets. Vervolgens moeten de formules nog aangepast worden.

De somscore bestaat nu uit 4 getallen bij elkaar opgeteld. In cel H11 typt u nu `=D11+E11+F11+G11`. Het gemiddelde in kolom I moet ook aangepast worden. U moet niet langer door 3, maar door 4 delen. Dus de tekst `=H11/3` in cel I11 vervangt u door `=H11/4`. Vervolgens selecteert u zowel de beide cellen H11 en I11 gevolgd door `<Ctrl>+<C>`. Daarna

selecteert u in een keer het blokje cellen H12 tot en met I22 gevolgd door <Ctrl>+<V> waarna alle berekeningen automatisch opnieuw worden uitgevoerd. Het resultaat staat in Figuur 2.9.

8							
9	+					CEL G9	
10		toets 1	toets 2	toets 3	toets 4	Somscore	Toetsgemiddelde
11	leerling 1	4	7	6	6	23	5,8
12	leerling 2	6	8	6	6	26	6,5
13	leerling 3	7	7	6	5	25	6,3
14	leerling 4	8	8	8	4	28	7,0
15	leerling 5	6	8	8	5	27	6,8
16	leerling 6	5	8	7	6	26	6,5
17	leerling 7	6	6	5	4	21	5,3
18	leerling 8	7	7	6	5	25	6,3
19	leerling 9	6	7	9	6	28	7,0
20	leerling 10	6	6	9	8	29	7,3
21	leerling 11	5	8	7	7	27	6,8
22	leerling 12	9	8	10	7	34	8,5

Figuur 2.9. Toetsresultaten na het invoegen van de vierde toets.

In plaats van kolom G, had u ook kolom F kunnen selecteren. U doet dit door in Figuur 2. 7 kolom F te selecteren, gevolgd door in de menubalk op Invoegen en vervolgens op Kolommen te klikken. Het frappante is nu dat de inhoud van de cellen in de oude kolommen G en H (nu dus H en I) zich automatisch aanpassen. Stond in cel G11 van Figuur 2.7 nog '=D11+E11+ F11', na het invoegen van een nieuwe kolom F staat nu in cel H11 '=D11+E11+G11'. De verwijzing naar de desbetreffende cellen heeft zich automatisch aangepast aan de door u ingevoegde kolom. Ga zelf na wat er met de inhoud in de oude kolom H en de nieuwe kolom I gebeurt.

Verwijzingen naar cellen worden automatisch aangepast bij kopiëren en plakken en/of het invoegen van nieuwe rijen en kolommen. Het is ook mogelijk om een verwijzing vast te zetten. Dit wordt ook bevroren genoemd. U doet dit door een dollar teken (\$) voor zowel de rij- als kolomaanduiding van de cel te plaatsen. Stel in de cel A4 staat het getal '4'. Vervolgens typt u in cel B4 '= \$A\$4'. Kopieer nu cel B4 en welke cel u ook kiest om naar toe te kopiëren, de gekopieerde cel verwijst altijd naar cel A4. Probeer het maar. In Figuur 2.10 is dit principe weergegeven.

	A	B	C	D	E	F
1	1	1	=A1	CEL D1		
2	2	2	=A2			
3	3	3	=A3			
4	4	4	=\$A\$4			
5	5	5	=A5			
6	6	6	=A6			
7	7	7	=A7		4	
8	8	4	=\$A\$4		4	
9					4	
10					4	
11						

Figuur 2.10. Illustratie van referentieel en absoluut verwijzen.

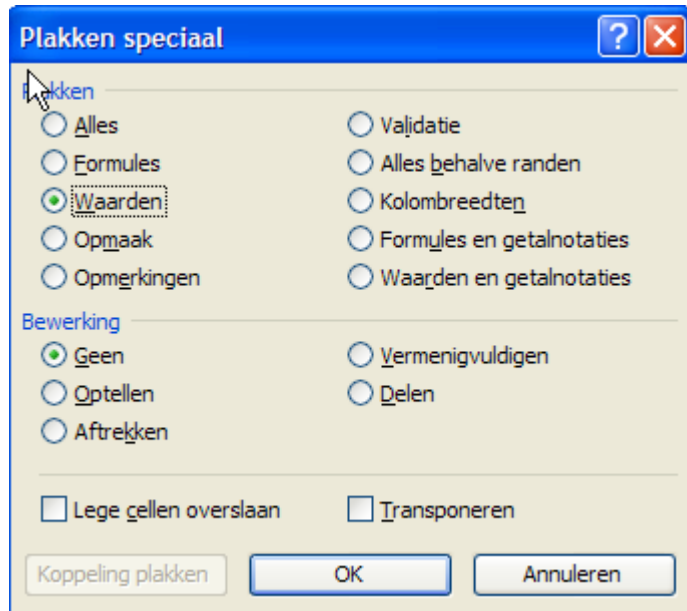
In de cellen A1 tot en met A8 van Figuur 2.10 zijn de getallen 1 tot en met 8 weergegeven. In de cellen van kolom C is de inhoud van de cellen van kolom B tekstueel aangegeven. De eerste vier cellen van kolom B zijn gekopieerd en vervolgens geplakt in de cellen B5 tot en met B8. U ziet dat de cellen B5, B6, en B7 automatisch de waarden van A5, A6 en A7 uitlezen. Precies zoals we dit eerder omschreven hebben. U ziet ook dat de inhoud van de cel B8 niet naar A8, maar naar A4 verwijst. Ook hebben we nog eens cel B4 apart gekopieerd en vervolgens geplakt in de cellen E7 tot en met E10. Deze cellen geven allemaal het getal 4 (de waarde van cel A4) terug. Dat komt door het gebruik van (beide) dollartekens. U kunt nu naar hartenlust de waarde van cel A4 laten variëren. Zelf hebben we in Figuur 2.11 de '4' door '123' vervangen. U ziet het resultaat in Figuur 2.11.

	A	B	C	D	E
1	1	1	=A1	CEL D1	
2	2	2	=A2		
3	3	3	=A3		
4	123	123	=\$A\$4		
5	5	5	=A5		
6	6	6	=A6		
7	7	7	=A7		123
8	8	123	=\$A\$4		123
9					123
10					123
11					

Figuur 2.11. Idem als Figuur 2.10 nadat we in cel A4 het getal '4' door '123' vervangen hebben.

U kunt zowel een gehele rij als een gehele kolom onafhankelijk van elkaar bevriezen. Als u de rij wilt bevriezen dan plaatst u een dollarteken voor het cijfer. Als u een kolom wilt vastzetten dan plaats u een dollarteken voor de letter. Probeer het maar eens.

Nog een laatste tip: referentieel verwijzen betekent dat u steeds alert moet zijn, waar u wat allemaal verandert in een werkblad. Het is ook mogelijk om alle formules en referentiële verwijzingen uit te schakelen en enkel de uitkomsten te bewaren. U selecteert daartoe met <Ctrl>+<A> het gehele werkblad en kopieert daarna alles met <Ctrl> + <C>. Vervolgens kiest u in de menubalk Bewerken, Plakken Speciaal en schakelt u in het geopende dialoogscherf Plakken speciaal de radiobutton Waarde in (zie Figuur 2.12).



Figuur 2.12. Dialoogscherf Plakken speciaal.

Als u Waarden heeft geselecteerd en vervolgens klikt op de knop OK (zie Figuur 2.12), dan krijgt u gewoon de resultaten van de opgevraagde formules te zien. De formules als ook de verwijzingen naar gebruikte cellen bent u kwijt. U houdt alleen getallen over.

2.3 Functies en haakjes

Tot nu toe hebben we laten zien hoe u met een beetje typewerk en veel kopiëren en plakken, eenvoudige berekeningen kunt uitvoeren. U kunt de werkzaamheden in Excel verregaand vereenvoudigen door gebruik te maken van functies.

U hebt zojuist steeds berekeningen gemaakt met behulp van directe verwijzingen naar cellen en het gebruik van operatoren. Zo berekende u hierboven bijvoorbeeld een som in cel G11 door '= $D11+E11+F11$ ' te typen. Excel beschikt echter ook over (heel veel) functies. Functies worden vaak ook met de benaming formules aangeduid. De formulebalk heet ook niet voor niets formulebalk. De formule voor een som kunt u uitvoeren door in veld G11 het volgende te typen: '= $SOM(D11:F11)$ '.

Een functie bestaat uit twee delen. Het eerste is de **naam** van de functie en meestal geeft die naam ook aan wat de functie gaat doen. $SOM()$ betekent niets anders optellen. Behalve de naam

kent de functie ook een **argument**. Zoals de naam aangeeft wat er moet gebeuren, zo geeft het argument geeft aan met welke cellen iets moet gebeuren. Een argument kan uit meerdere delen bestaan, bijvoorbeeld een rij getallen.

Er zijn verschillende manieren om een rij getallen in Excel op te geven. De puntkomma (;) gebruikt u bij afzonderlijk getallen. Bijvoorbeeld als '=SOM (1;2;3;7;8)' typt, dan geeft Excel u het cijfer 21 terug.² De getallen 1, 2, 3, 7 en 8 worden het argument van deze functie genoemd. Als u een reeks getallen op wilt geven, bijvoorbeeld 1 tot en met 5, dan kunt u de dubbele punt (:) tussen het eerste en laatste getal gebruiken: '=SOM(1:5)'.

De functie wordt uitgevoerd door het isgelijktteken (=) te gebruiken. Als u in cel G11 'SOM(D11:F11)' typt, dan leest Excel dit als een tekst. Door te beginnen met het isgelijktteken maakt u het programma duidelijk dat u niet in de tekst, maar in de uitkomst van de functie of formule geïnteresseerd bent '=SOM(D11:F11)'.

In plaats van cellen of getallen te typen kunt u ook de cellen selecteren als u eenmaal de functie heeft aangeroepen in Excel. Kortom, in plaats van C4 zelf te typen, kunt u ook cel C4 met de cursor selecteren in een formule. Automatisch worden kleuren gebruikt om de cellen aan te geven waarop de formule van toepassing is.

Als er een formule SOM() bestaat, dan is er ook vast wel een voor het berekenen van het gemiddelde. Dat klopt. Die kunt u oproepen door in een cel '=GEMIDDELDE()' te typen. Tussen de ronde haken zet u dan weer de getallen waarvan u het gemiddelde wilt berekenen. Zoals gezegd, er zijn enorm veel functies in Excel. U kunt dit zien door op de fx in de formulebalk te klikken. Hiermee opent u het dialoogscherf Functie invoegen.³

Een functie gaat altijd gepaard met ronde haakjes. U kunt echter in Excel zelf ook haakjes gebruiken voor het uitvoeren van berekeningen. Dit doen we om de volgorde waarin de berekeningen moeten worden uitgevoerd, nader aan te geven. Een open haakje wordt altijd ergens gevolgd door een gesloten haakje. Eerst worden de berekeningen uitgewerkt die binnen

² De verschillende getallen scheidt u van elkaar met een puntkomma (;). Dat is dus heel iets anders dan een komma. Wie per ongeluk '=SOM(1;2;3;4;5)' heeft ingetypt krijgt geen 21 maar 18,3 als uitkomst.

³ Berekeningen in de kolommen F en G van de Figuren 2.8 en 2.9 hadden we dus ook met de functies voor sommen ('=SOM(D11:F11)') en gemiddelden ('=GEMIDDELDE(D11:F11)') kunnen berekenen.

de haakjes staan. Denk bij het gebruik van haakjes altijd aan Meneer Van Dalen wacht op antwoord: dus eerst vermenigvuldigen, dan delen, vervolgens optellen en tot slot aftrekken.

Typ nu in cel A4 het getal 30 en in cel B4 het getal 70. In de cellen C6 t/m C11 typt u achtereenvolgens onderstaande formules zoals deze in Figuur 2.13 tekstueel zijn weergegeven in kolom D. Vergelijk de onderlinge resultaten nauwkeurig en ga bij uzelf de volgorde na waarin de berekeningen zijn uitgevoerd.

	A	B	C	D	E
1					
2					
3		CEL B3			
4	30	70	10,00	'=(A4 + B4)/(7+3)'	
5			43,00	'=A4 + B4/7+3'	
6			42,86	'=(A4 + B4)/7*3'	
7			44,86	'=(A4 + B4)/7*3 + 2'	
8			44,86	'=((A4 + B4)/7*3) + 2'	
9			71,43	'=(A4 + B4)/7*(3 + 2)'	
10					
11					

Figuur 2.13. Zowel binnen als buiten twee haakjes: Meneer Van Dalen wacht op antwoord.

2.4 Een paar rekenvoorbeelden

Vraag 1: Een weiland is 40 meter lang en 30 meter breed. We gaan dubbel prikkeldraad om dat weiland spannen. Hoeveel meter prikkeldraad hebben we nodig?

Antwoord 1: Eerst stellen we de omtrek van het weiland vast. Dus twee keer de breedte plus twee keer de lengte. Vervolgens moeten we daar twee keer omheen met de draad. De gevolgde stappen zijn aangegeven in Figuur 2.14

4	+		
5			
6	Gegevens	Cel B6	
7	30		
8	40		
9	70	optellen cellen A7 + A8	
10	140	Cel A9 maal 2 voor het bepalen van omtrek	
11	280	Cel A10 maal 2 voor het gebruik van dubbele draad	
12			

Figuur 2.14. Rekenschema voor het bepalen van het benodigde aantal meters prikkeldraad.

Het prikkeldraadvraagstuk kunt u natuurlijk ook gewoon met behulp van een formule oplossen (Excel is een rekenmachine!). U typt in een willekeurige cel daartoe `'=SOM(30;40)*2*2'` of `'=(30+40)*2*2'` en in de cel verschijnt het getal 280.

Vraag 2: Bereken de gemiddelde lengte in centimeters van de volgende vier jongens: 170, 170, 155 en 185.

Antwoord 2: `'=GEMIDDELDE(170;170;155;185)'` geeft 170 cm terug.

Ga bij jezelf na waarom `'=GEMIDDELDE(2*170;155;185)'` niet 170, maar 266,67 oplevert en dus fout is.

Vraag 3: Van een groep van 80 mannen is de gemiddelde lengte gelijk aan 1,84 en van een groep van 40 vrouwen is de gemiddelde lengte gelijk aan 1,74. Bereken de gemiddelde lengte van de gehele groep.

Antwoord 3: Het belangrijkste is om eerst de informatie die wordt gegeven in de vraag gestructureerd samen te vatten. (a) We hebben twee groepen mensen met een gemiddelde lengte per groep. We kennen het groepsgemiddelde en we weten uit hoeveel mensen een groep bestaat. (b) Dat betekent dus dat we eerst de totale lengte van alle mannen samen en vervolgens ook van alle vrouwen samen kunnen vaststellen. (c) Aangezien ook bekend is hoeveel mannen en vrouwen er in totaal zijn, dienen we alleen nog maar de totale lengte van de twee groepen bij elkaar op te tellen en (d) te delen door alle mannen en vrouwen samen. Het leuke van Excel is nu dat we elke gemaakte stap in een cel kunnen vastleggen.

Het doorlopen van de stappen (a) tot en met (d) is weer gegeven in Figuur 2.15. De gele cel E23 geeft weer aan waar we ons in het werkblad bevinden om deze som op te lossen.

	C	D	E	F
21	+			
22				
23			cel E23	
24		(a) aantal mannen	80	
25		(a) gemiddelde lengte man	1,84	
26		(b) totale lengte mannen	147,2	(b) vermenigvuldigen van cel E24 met E25
27				
28		(a) vrouwen	40	
29		(a) gemiddelde lengte vrouw	1,74	
30		(b) totale lengte vrouwen	69,6	(b) vermenigvuldigen van cel E28 met E29
31				
32		(c) Totale lengte van mannen en vrouwen	216,8	(c) optellen van de cellen E26 en E30
33		(c) Totaal aantal mannen en vrouwen samen	120	(c) optellen van de cellen E24 en E28
34				
35		(d) Gemiddelde lengte gehele groep	<u>1,81</u>	(d) deel cel E32 door cel E33
36				

Figuur 2.15. Berekenen van een gewogen gemiddelde.

De gemiddelde lengte voor de gehele groep mensen is 1,81 m. Met een duur woord heet dit het gewogen gemiddelde. We wegen niet de kilo's van de mannen en de vrouwen, maar wel van hoeveel mannen en vrouwen we informatie hebben. Het belang van het aantal mannen en het aantal vrouwen hebben we teruggewogen bij het vaststellen van het overall gemiddelde.

3 De grote grafische rekenmachine

In het vorige hoofdstuk hebben we Excel leren gebruiken als een rekenmachine. In dit hoofdstuk laten we allereerst zien hoe onderzoekers met eenvoudige rekenregels zelf frequentietabellen kunnen maken. In het tweede deel van dit hoofdstuk gaan we een relatieve frequentietabel grafisch weergeven.

3.1 Frequentietabellen

In hoofdstuk 3 van het boek *Klassenonderzoek? Dat doe je zó!* worden in het kort de resultaten van een tevredenheidsonderzoek besproken. Twintig individuen hebben acht vragen gescoord op een schaal van 1 tot en met 6. De somscore op deze acht vragen is weergegeven in Figuur 3.1 en varieert dus per definitie ergens vanaf 8 tot en met 48.

	A	B	
4			
5	A5		
6	Tevredenheidsscore	Frequentie (f)	
7	29	1	
8	33	2	
9	34	1	
10	35	1	
11	36	1	
12	38	3	
13	39	4	
14	40	3	
15	42	2	
16	43	1	
17	44	1	
18	Totaal	20	

Figuur 3.1. Tevredenheidsscore (N=20).

Er bestaan allerlei afspraken over het maken van tabellen en wat er nu juist wel of juist niet in moet staan. Lezers die hulp nodig hebben bij het maken van tabellen wordt aangeraden de website van het Centraal Bureau voor de Statistiek te bezoeken. Daar zijn vuistregels en veel, heel veel voorbeelden te vinden. Een van de belangrijkste criteria bij de presentatie van gegevens is informatie. Tabellen moeten informatief en gemakkelijk te lezen zijn. Figuur 3.1 is niet gemakkelijk te lezen en ook nog eens niet informatief. Hieronder gaan we dit rijtje getallen in Figuur 2.1 zodanig bewerken dat de tabel zelf ook echt informatie geeft. Het spreekt natuurlijk voor zich dat de tekst in cel A5 niet opgenomen hoeft te worden, het dient slechts om ons referentiepunt te markeren.

In de eerste plaats willen we alle gegevens precies in het midden van de cellen geplaatst hebben. Dit doen we door eerst de cellen A6 tot en met B18 te selecteren en vervolgens op Centreren in de Opmaak werkbalk te klikken. Het totaal in cel B18 is berekend met de formule ' $=\text{SOM}(\text{B7}:\text{B17})$ '. Behalve het intypen van deze tekst, kan men ook de cellen B7 tot en met B17 selecteren en vervolgens op Σ (AutoSom) in de Standaard werkbalk klikken. De formule ' $=\text{SOM}(\text{B7}:\text{B17})$ ' wordt dan automatisch door Excel weggeschreven in de cel onder het door u geselecteerde gebied. Dus in B17.

De interpretatie van rechte tellingen wordt al een stuk eenvoudiger als we de aantallen uitdrukken in proporties of procenten. Dat is de eerste stap.

- Typ daartoe allereerst het woord 'Proportie (p)' in cel C6. Men moet altijd in een oogopslag kunnen zien wat men waar heeft weggeschreven!
- Ga met cursor naar cel C7 en typ de formule voor het uitrekenen van een proportie: ' $=\text{B7}/\text{B18}$ '.
- Het kopiëren van de inhoud van cel C7 naar de cel C8 resulteert in ' $=\text{B8}/\text{B19}$ '. Aangezien cel B19 leeg is (er staat weliswaar niets in, nul dus) resulteert de gekopieerde formule in een foutmelding: '#DEEL/0!'. En zoals bekend: delen door 0 kan niet.
- Dit kunnen we op twee manieren oplossen. Of we typen in cel B19 het getal 20. Maar dat is natuurlijk een flut oplossing (*excusez le mot*). De echte oplossing is om de verwijzing naar cel B18 te bevriezen. Dit doen we door in cel C7 de tekst ' $=\text{B7}/\text{B18}$ ' te vervangen door ' $=\text{B7}/\$B\18 '. Uiteindelijk gaan we deze tekst kopiëren en plakken in de cellen C8 tot en met C17.
- Controleer in veld C18 of het totaal in de cellen C7 tot en met C17 inderdaad gelijk aan 1 is.
- Zorg er ook voor dat de proporties allemaal op twee decimalen zijn afgerond. Al kan men zich terecht afvragen of in dit voorbeeld twee decimalen niet net iets teveel van het goede is.

	A	B	C	
4				
5	A5			
6	Tevredenheidsscore	Frequentie (f)	Proportie (p)	
7	29	1	0,05	
8	33	2	0,10	
9	34	1	0,05	
10	35	1	0,05	
11	36	1	0,05	
12	38	3	0,15	
13	39	4	0,20	
14	40	3	0,15	
15	42	2	0,10	
16	43	1	0,05	
17	44	1	0,05	
18	Totaal	20	1	

Figuur 3.2. De tevredenheidsscore na het berekenen van proporties.

Vervolgens wordt de cumulatieve frequentieverdeling gepresenteerd. Ook dit laten we de computer dat voor ons uitrekenen.

- Typ in cel D6 de tekst 'Cumulatieve frequentie (cf)'. Een cumulatieve frequentie verdeling is gedefinieerd als de frequentie van die waarde plus de frequenties van alle waarden die kleiner zijn. Dat kunnen we in Excel; het is gewoon optellen, alleen moeten we even goed opletten hoe en waar we kopiëren.
- Typ in cel D7 '=B7'. De cumulatieve frequentie van de laagste waarde is per definitie gelijk aan aantal keren dat we tevredenheidsscore 29 scoren.
- De cumulatieve frequentie van tevredenheidsscore 30 is het aantal keren dat scores 29 en 30 gescoord zijn: dus 1+2. We moeten dus de waarden in de cellen D7 en B8 bij elkaar optellen in cel D8. Typ daartoe '=D7+B8' in cel D7.

- Kopieer de inhoud van cel D8 en plak dat in de cellen D9 tot en met D17. Klaar! Het resultaat staat in Figuur 3.3.

	A	B	C	D	
4					
5	A5 +				
6	Tevredenheidsscore	Frequentie (f)	Proportie (p)	Cumulatieve frequentie (cf)	
7	29	1	5%	1	
8	33	2	10%	3	
9	34	1	5%	4	
10	35	1	5%	5	
11	36	1	5%	6	
12	38	3	15%	9	
13	39	4	20%	13	
14	40	3	15%	16	
15	42	2	10%	18	
16	43	1	5%	19	
17	44	1	5%	20	
18	Totaal	20	1		

Figuur 3.3. Tevredenheidsscore na het berekenen van de kolom cumulatieve frequentie.

Wel altijd verifiëren dat in de laatste cel (in dit geval dus cel D17) ook inderdaad het aantal respondenten waarop de tabel gebaseerd is (N=20) staat.

Er moet nu nog een laatste kolom aangemaakt worden en dat is het 'Cumulatief percentage (cum %)'.

- Typ allereerst de tekst 'Cumulatief percentage (cum%)' in cel E6. U kunt deze lange tekst in één cel zichtbaar maken op verschillende manieren: (a) bredere kolommen, (b) kleiner lettertype (font), maar ook (c) door op Opmaak in de menubalk te klikken en vervolgens in tabblad Uitlijning van het dialoogschermb Cel eigenschappen het voor Terugloop af te vinken. Daarna moet u vermoedelijk nog de hoogte van rij 6 even aanpassen.

- Selecteer en kopieer nu de cellen D7 tot en met D17 (dus de formules voor proporties) en plak dit hele blok in de cellen E7 tot en met E17 met behulp van <Ctrl>+<V>. In de cellen E7 tot en met E17 verschijnen nu de cumulatieve proporties. In cel E17 staat het cijfer 1. Proporties zijn niets anders dan percentages. Het zijn dezelfde getallen; alleen de eenheid waarin het getal wordt uitgedrukt is anders. Proporties omzetten in percentages is eigenlijk een kwestie van opmaak.

- Selecteer de cellen C7 tot en met C17 gevolgd door op Opmaak in de menubalk te klikken. Daarna klikken we achtereenvolgens op Celeigenschappen en kies op het tabblad Getal Percentage. We willen geen decimalen, dus de 2 vervangen we in dialoog scherm door 0.

Het resultaat van deze exercitie staat in Figuur 3.4

	A	B	C	D	E
4					
5	A5				
6	Tevredenheidsscore	Frequentie (f)	Proportie (p)	Cumulatieve frequentie (cf)	Cumulatief percentage (cum%)
7	29	1	0,05	1	5%
8	33	2	0,10	3	15%
9	34	1	0,05	4	20%
10	35	1	0,05	5	25%
11	36	1	0,05	6	30%
12	38	3	0,15	9	45%
13	39	4	0,20	13	65%
14	40	3	0,15	16	80%
15	42	2	0,10	18	90%
16	43	1	0,05	19	95%
17	44	1	0,05	20	100%
18	Totaal	20	1		

Figuur 3.4. Vervolg absolute en relatieve (cumulatieve) frequentietabel Tevredenheidsscore (N=20).

Nog een opmerking is hier op zijn plaats. Het is niet helder om in een en dezelfde tabel zowel proporties als percentages weer te geven. Dus de proporties moet men eigenlijk ook als percentages weergeven. Dat is gebeurd in Figuur 3.5.

De basis van de door ons gemaakte frequentietabel is het feit dat alles draait om de uitkomst in cel B18: het totaal aantal respondenten. De waarde in die cel is de som van cellen in kolom B, terwijl we de verwijzingen naar cel B8 hebben bevroren. Het handige van deze manier van uitschrijven en uitwerken is dat als we in de cellen B6 ten met B7 iets veranderen, dit meteen in de tabel zelf wordt doorgerekend. Wat ook de invoer van de getallen in kolom B is, we krijgen altijd de juiste frequentieverdeling. Stel, tevredenheidsscore 39 blijkt geen 4, maar 14 keer gescoord. U verandert de 4 in cel B13 in 14 en de gevolgen zijn meteen zichtbaar in de Figuur. U hoeft verder niets te doen. Vergelijk Figuur 3.5 daartoe met Figuur 3.4.

	A	B	C	D	E
4					
5	A5				
6	Tevredenheidsscore	Frequentie (f)	Proportie (p)	Cumulatieve frequentie (cf)	Cumulatief percentage (cum%)
7	29	1	3%	1	3%
8	33	2	7%	3	10%
9	34	1	3%	4	13%
10	35	1	3%	5	17%
11	36	1	3%	6	20%
12	38	3	10%	9	30%
13	39	14	47%	23	77%
14	40	3	10%	26	87%
15	42	2	7%	28	93%
16	43	1	3%	29	97%
17	44	1	3%	30	100%
18	Totaal	30	1		

Figuur 3.5. Absolute en relatieve (cumulatieve) frequentietabel Tevredenheidsscore (N=30).

Nog een kleine toelichting op Figuur 3.5. De richting van de pijlen geeft de betekenis aan van een cumulatief percentage. U ziet dus dat 13% (groene pijlpunt in cel E9) van de totale populatie 34 of minder punten heeft gescoord op Tevredenheid. Let ook op het feit dat Excel voor u afrondt: 87% (cel E14) + 7% (cel C15) = 93% (cel E15). U had hier toch echt 94% verwacht. Kennelijk staat in cel E15 een getal dat kleiner is dan 93,5%. U kunt dit zichtbaar maken door de opmaak van kolom E te veranderen.

3.2 Grafieken

De basis voor een grafiek bestaat bijna altijd uit een tabel. Om de grafische mogelijkheden van Excel te benutten dienen we over een tabel met inhoud te beschikken. De religieuze achtergrond van onze 20 hierboven genoemde leerlingen is als volgt. Vijf leerlingen zijn 'christelijk'; vier leerlingen geven op 'moslim' te zijn en nog eens elf leerlingen vermelden 'geen godsdienst'. We gaan hieronder deze gegevens in een taartdiagram zichtbaar maken in Excel. Een taartdiagram is niets anders dan het uittekenen van proporties of procenten. Zojuist hebben we dat gedaan met betrekking tot de tevredenheidsscore. Met een beetje handig knippen, plakken en typen hebben we het tabelletje religie binnen de kortste keren voor elkaar.

Kopieer Figuur 3.4 in Excel naar Blad2.

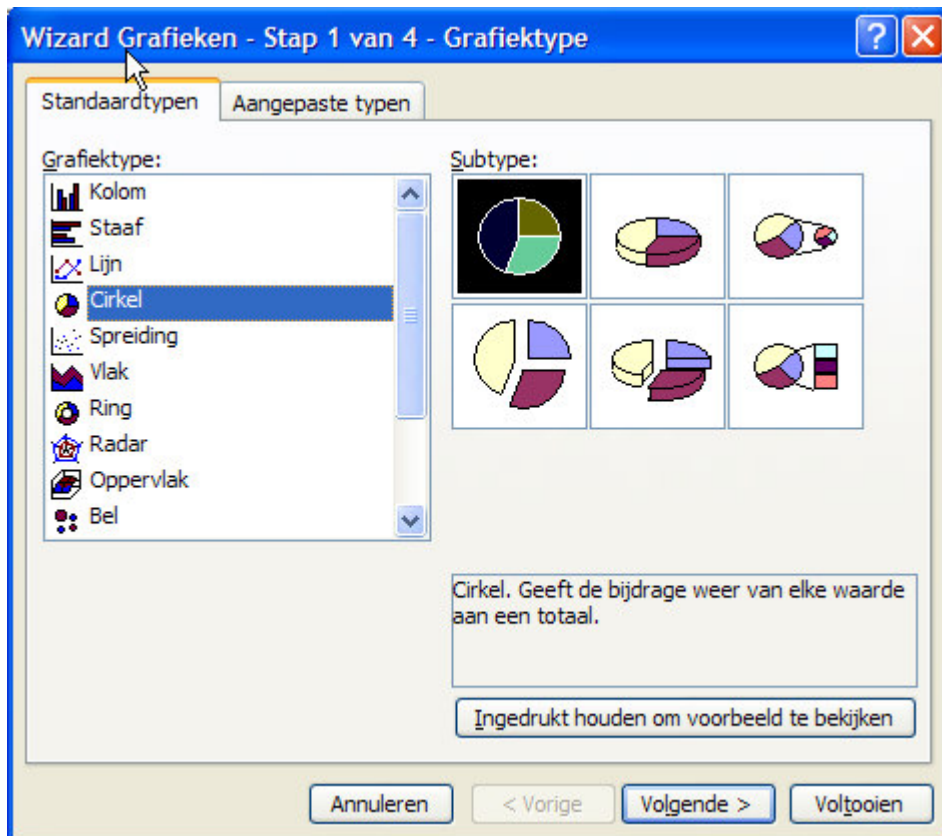
- Vervang 'Tevredenheidsscore' in cel A6 door 'Religieuze achtergrond'.
- Vervang in de cellen A7, A8 en A9 de scores 29, 33 en 34 door 'christelijk', 'moslim, en 'geen godsdienst'.
- Voer de juiste frequenties in in de cellen B7, B8 en B9.
- Selecteer de rijen 10 tot en met 17 en verwijder die.
- Selecteer de kolommen D en E en verwijder die.
- Selecteer de cellen C7 tot en met C9 en verander de weergave van de proporties in percentages. Pas ook de kolomtitel aan: vervang 'proportie (p)' door 'relatieve frequentie'.

	A	B	C	I
1				
2				
3				
4				
5	A5			
6	Religieuze achtergrond	frequentie (f)	relatieve frequentie	
7	christelijk	5	25%	
8	moslim	4	20%	
9	geen godsdienst	11	55%	
10	Totaal	20	100%	
11				

Figuur 3.6. Absolute en relatieve frequentietabel van religie (N=20).

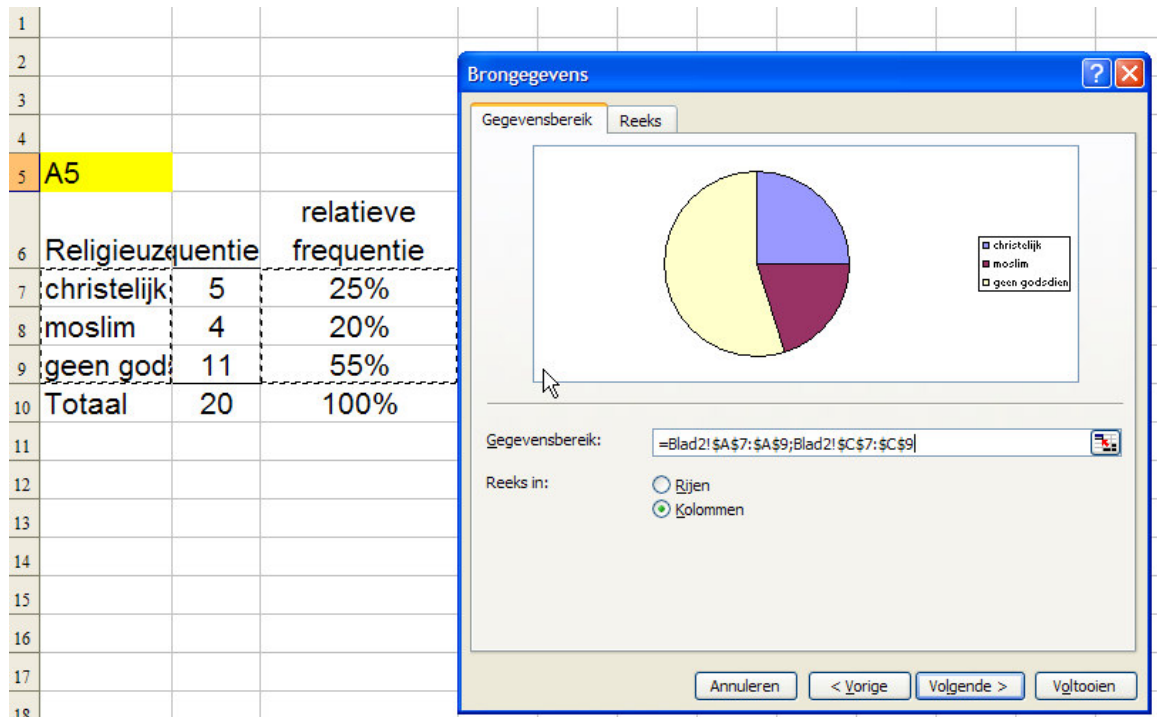
De relatieve frequenties in Figuur 3.6 gaan we hieronder in een taartdiagram (ook wel cirkeldiagram genoemd) weergeven.

- Klik op het icoontje Wizard grafieken  van de Standard werkbalk. Selecteer Cirkel. Het resultaat staat in Figuur 3.7.



Figuur 3.7. Dialoogscherm na aanklikken van de Wizard grafieken.

- Door op Volgende te klikken wordt het tweede dialoogscherm geopend.
- Ga met de cursor in de regel Gegevensbereik, druk de <Ctrl> toets in en selecteer de cellen A7, A8, A9, C7, C8 en C9. Laat na de selectie de <Ctrl>toets los. U ziet dan Figuur 3.8, het dialoogscherm Brongegevens.

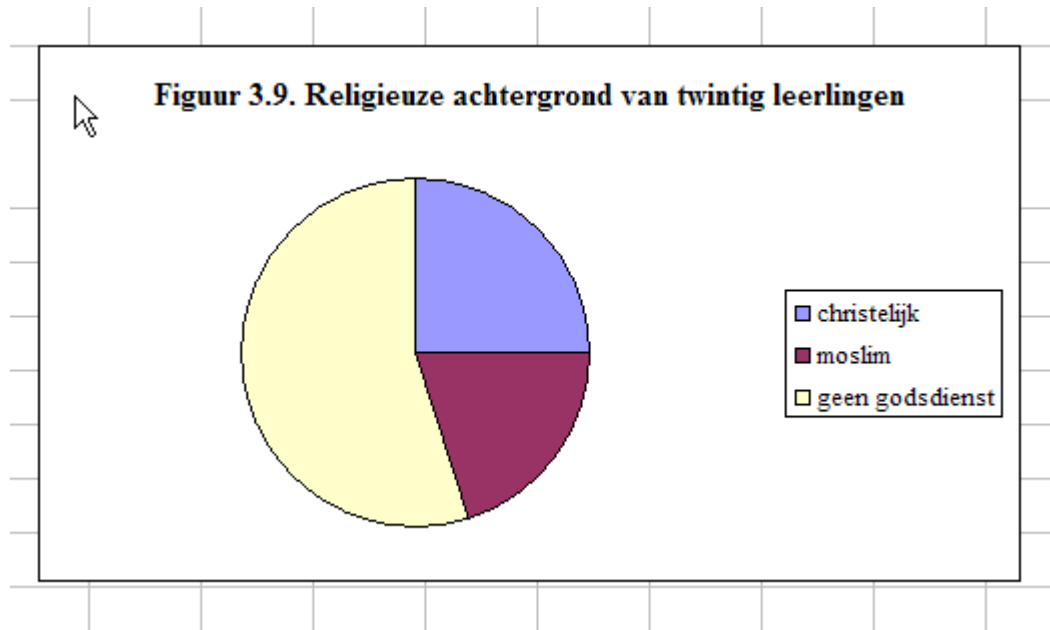


Figuur 3.8. Dialoogschermb Brongegevens.

. Merk op dat de cellen A7:A9 en C7:C9 zijn omkaderd met een stippeltjeslijn en dat de brongegevens in de regel Gegevensbereik als volgt zijn opgeschreven '=Blad2!\$A\$7:\$A\$9;Blad2!\$C\$7:\$C\$9'. Excel geeft zelf aan dat u niet meer op Blad1 aan het werk bent, maar op Blad2. De radiobutton bij Kolommen is geselecteerd. Klik nu op de knop Volgende in het dialoogschermb Brongegevens.

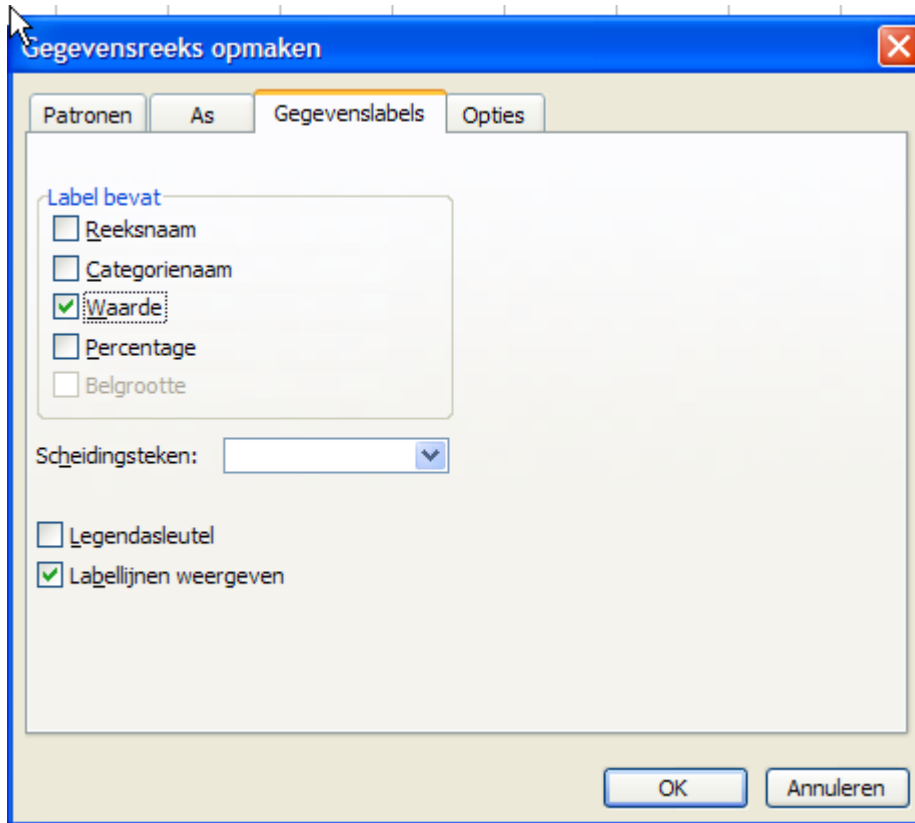
- Typ in het vierde en laatste scherm (niet afgebeeld) 'Figuur 3.9. Religieuze achtergrond van twintig leerlingen' bij Titels, Grafiektitel.

- Sluit de Wizard af door op Voltooien te klikken.



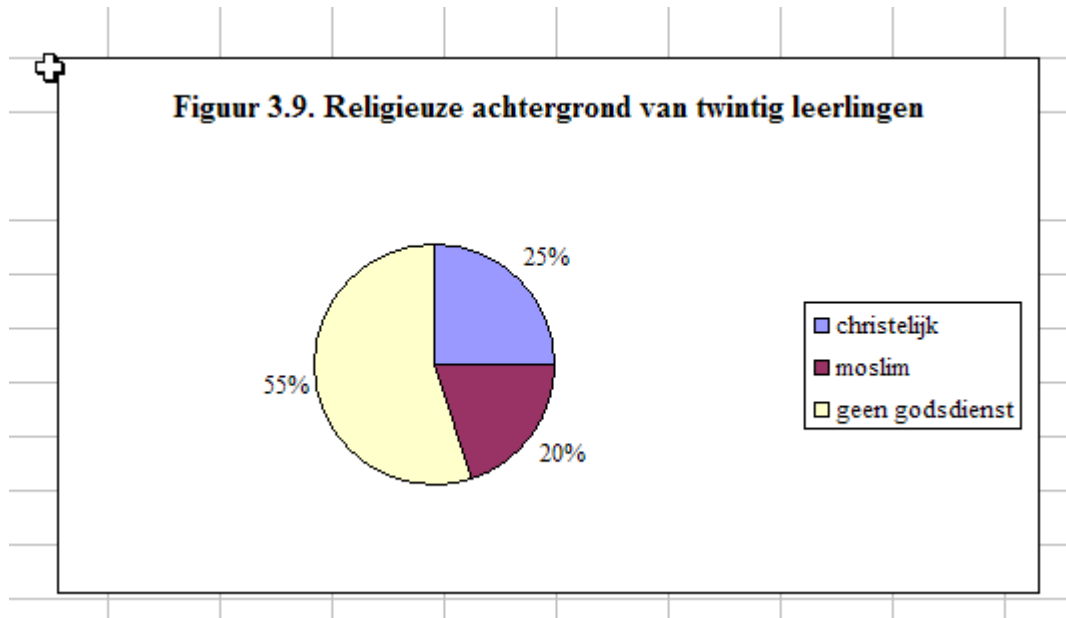
Figuur 3.9. Religieuze achtergrond (N=20).

Bij na klaar. We gaan eerst nog de juiste percentages in de grafiek zetten. Dubbelklik daartoe op de het grafische plaatje en het Dialoogscherf Gegevensreeks opmaken wordt geopend, kies tabblad Gegevenslabels en vink Waarde aan (zie Figuur 3.10).



Figuur 3.10. Tabblad van Gegevenslabels van Dialogoscherm Gegevensreeks opmaken.

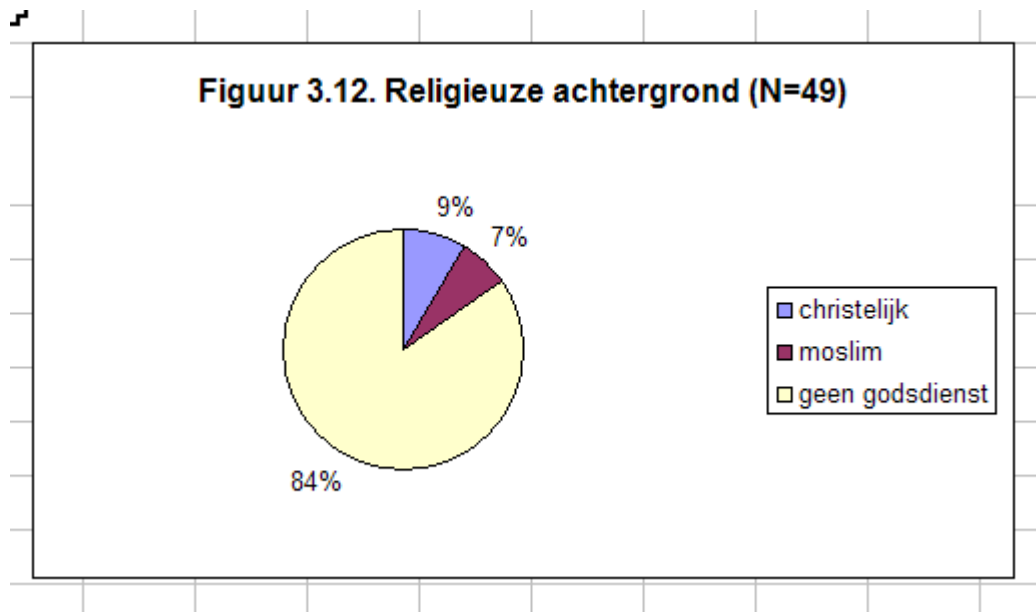
Het resultaat van deze handelingen kunt u terugvinden Figuur 3.11.



Figuur 3.11. Religieuze achtergrond (N=20).

En nu kunnen we, net als in het geval van de tevredenheidsscores weer de absolute scores op religie gaan veranderen in de tabel. Stel we hebben niet elf, maar veertig leerlingen met geen godsdienst. Het enige wat we nu moeten veranderen is het getal 11 in cel B9. Uiteraard passen we ook de titel aan, in cel B10 staat het totale aantal van 49 leerlingen vermeld, dat corrigeren we ook in de titel van de grafiek (Niet vergeten!).

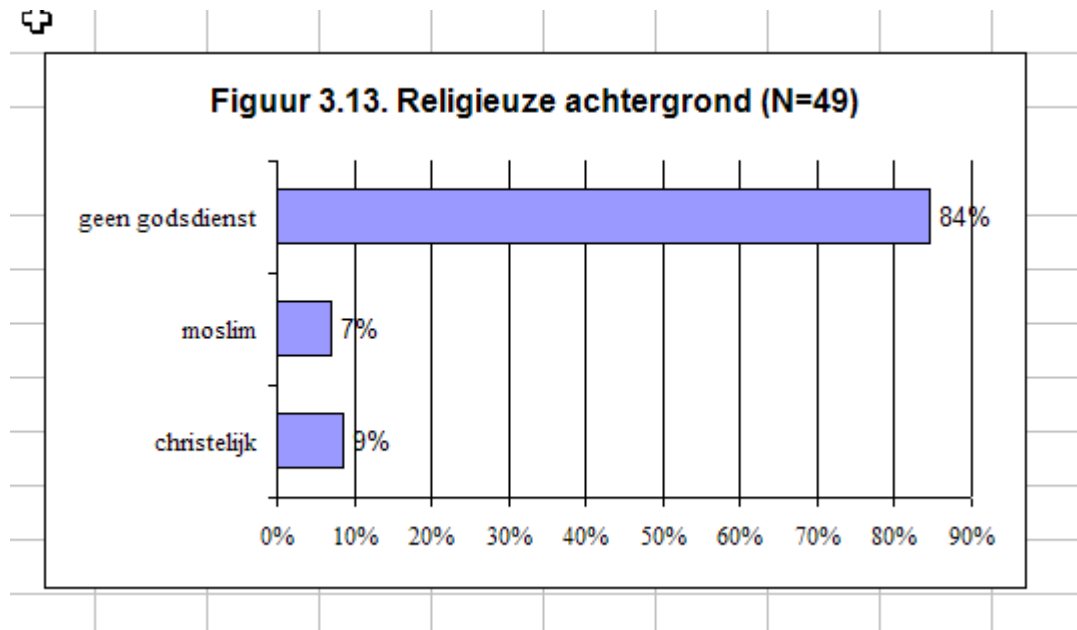
Selecteer de titel door er op te dubbelklikken. Je merkt dat je vanzelf in de tekst komt. Vervang '20' door '49'. Ook het lettertype passen we nog even aan. Dubbelklik in de legenda; het dialoogschermd Legenda opmaken verschijnt en klik op Tabblad Lettertype en selecteer Arial bij Lettertype. Het resultaat ziet u in Figuur 3.12.



Figuur 3.12.

Wie liever een staafdiagram ziet, gaat als volgt te werk. Selecteer de grafiek door er ergens een keer op te klikken, druk <Ctrl> toets in, selecteer in de balk weer de Wizard grafieken, en vervang Cirkel door Staafdiagram.

Nog even dit. Als u in het Grafiekvenster werkt, dan verschijnt het woord Grafiek in de menubalk in plaats van het woord Data. Grafiek verandert weer in Data in uw menubalk zodra u het Grafiekvenster verlaat. U kunt dit het gemakkelijkste zien door beurtelings ergens op een cel en het grafiekvenster te klikken en hou daarbij de menubalk in het oog.



Figuur 3.13. Staafdiagram van religie (N=49).

Selecteer het woord 'Reeks 1' en druk op Delete. En de variaties zijn dus echt eindeloos. Of toch maar weer een cirkeldiagram van de oorspronkelijk 20 leerlingen? Dat lukt na verloop van tijd binnen de 10 seconden. Oefenen dus!

Nog een laatste opmerking.

4 Kruistabellen

In de vorige hoofdstukken is de religieuze achtergrond van 20 leerlingen ter sprake gekomen. Het betreft 11 meisjes en 9 jongens. In feite hebben we dan te maken met twee frequentieverdelingen: een voor religie en een voor geslacht. Vragen als 'hoeveel procent van de meisjes is niet-religieus?', of 'hoeveel procent van alle islamitische leerlingen is een jongen?', willen we gemakkelijk kunnen beantwoorden. Dit doen we door de beide frequentieverdelingen in een kruistabel samen te vatten. We combineren de beide frequentieverdelingen met elkaar. De basis voor een kruistabel bestaat in dit geval uit 21 rijen (zie Figuur 4.1). Cel A1 met de tekst LeerlingId fungeert als referentie en in de eerste rij zijn de namen van de variabelen weggeschreven (cel B1 Geslacht, cel C1 Religie en cel D1 tel). De scores op geslacht staan in kolom B; scores op religie staan in kolom C. Elke leerling heeft dus zijn of haar eigen rij.

	A	B	C	D	
1	LeerlingId	Geslacht	Religie	tel	
2	1	jongen	christelijk	1	
3	2	meisje	niet-religieus	1	
4	3	jongen	moslim	1	
5	4	jongen	niet-religieus	1	
6	5	jongen	christelijk	1	
7	6	meisje	moslim	1	
8	7	meisje	niet-religieus	1	
9	8	meisje	christelijk	1	
10	9	meisje	niet-religieus	1	
11	10	meisje	christelijk	1	
12	11	meisje	niet-religieus	1	
13	12	jongen	niet-religieus	1	
14	13	meisje	niet-religieus	1	
15	14	jongen	niet-religieus	1	
16	15	meisje	christelijk	1	
17	16	jongen	niet-religieus	1	
18	17	meisje	niet-religieus	1	
19	18	jongen	niet-religieus	1	
20	19	jongen	moslim	1	
21	20	meisje	moslim	1	

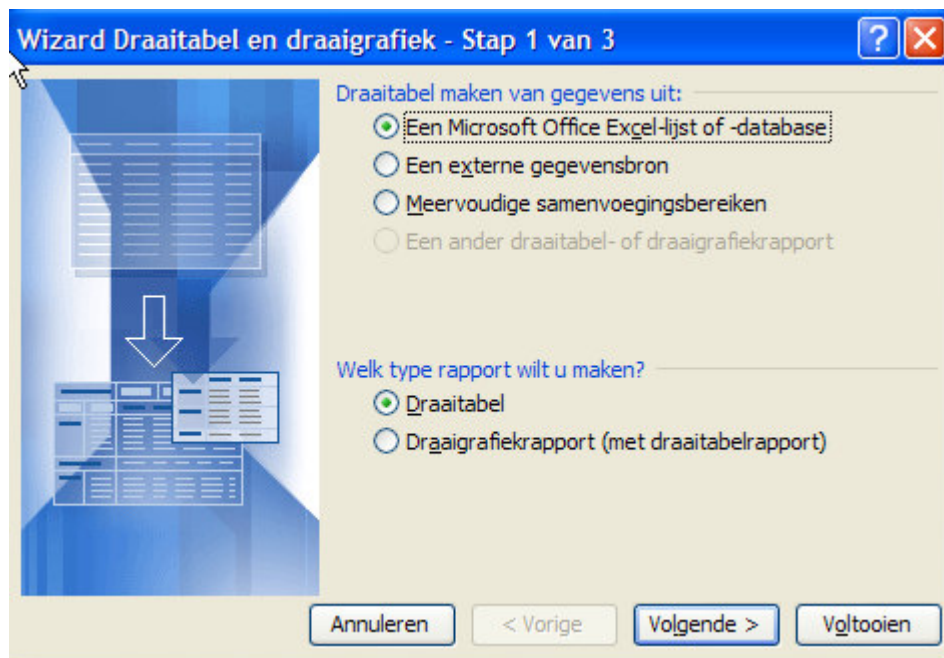
Figuur 4.1. Overzicht van 20 leerlingen.

Dus leerling 1 is een christelijke jongen, leerling 2 is een niet-religieus meisje, etc. Laten we eenvoudig beginnen: hoeveel christelijk jongens zijn er? Dat is een kwestie van tellen en het antwoord is twee. Als we zo alle mogelijkheden (zes in totaal) langs moeten gaan dan gebeuren er twee dingen. Er worden fouten gemaakt en we worden dit heel snel beu. Excel gaat dit voor ons doen.

De oplettende lezer heeft al natuurlijk al opgemerkt gezien dat in kolom D een serie 1-en is opgenomen. Strikt genomen is deze kolom niet nodig. Deze kolom geeft steeds aan hoeveel

leerlingen ik per rij heb, namelijk één. Het aanmaken van kolom D is een handigheidje om met een Wizard een kruistabel te maken en om te illustreren wat er precies gebeurt.

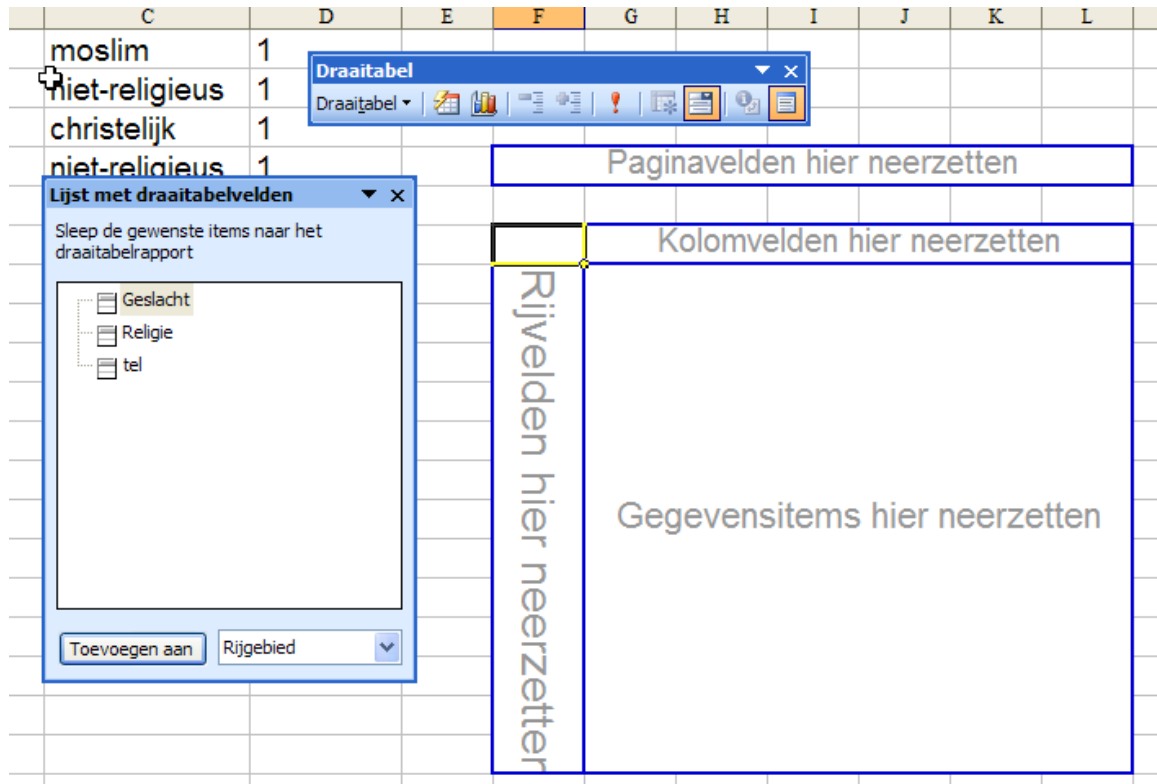
- Ga met de cursor in een ongebruikte cel staan, bijvoorbeeld F12. Klik in de menubalk op Data en selecteer Draaitabel en Draaigrafiek rapport ... (zie Figuur 4.2).



Figuur 4.2.

- Verifieer dat zowel de radiobuttons Een Microsoft Office Excel-lijst of -database als ook de radiobutton Draaitabel geselecteerd zijn.
- Klik op de knop Volgende en selecteer de cellen B1 tot en met D21. In het dialoogscherf Wizard Draaitabel en draaigrafiek - Stap 2 van 3 is dit aangegeven als Bereik: 'Blad1!\$B\$1:\$D\$21'. Klik weer op volgende.
- Nu verschijnt het derde en laatste dialoogscherf: Wizard Draaitabel en draaigrafiek - Stap 3 van 3 waarin de zojuist geselecteerde cel is aangegeven (in ons geval F12). Klik op Voltooien.

Helaas is er niets voltooid nadat u heeft geklikt op Voltooien. Integendeel, het scherm dat u voor u heeft, is zelfs tamelijk gecompliceerd en weergegeven in Figuur 4.3.



Figuur 4.3. Na het klikken op de knop Voltooid van de Wizard Draaitabel en draaigrafiek Stap 3 van 3.

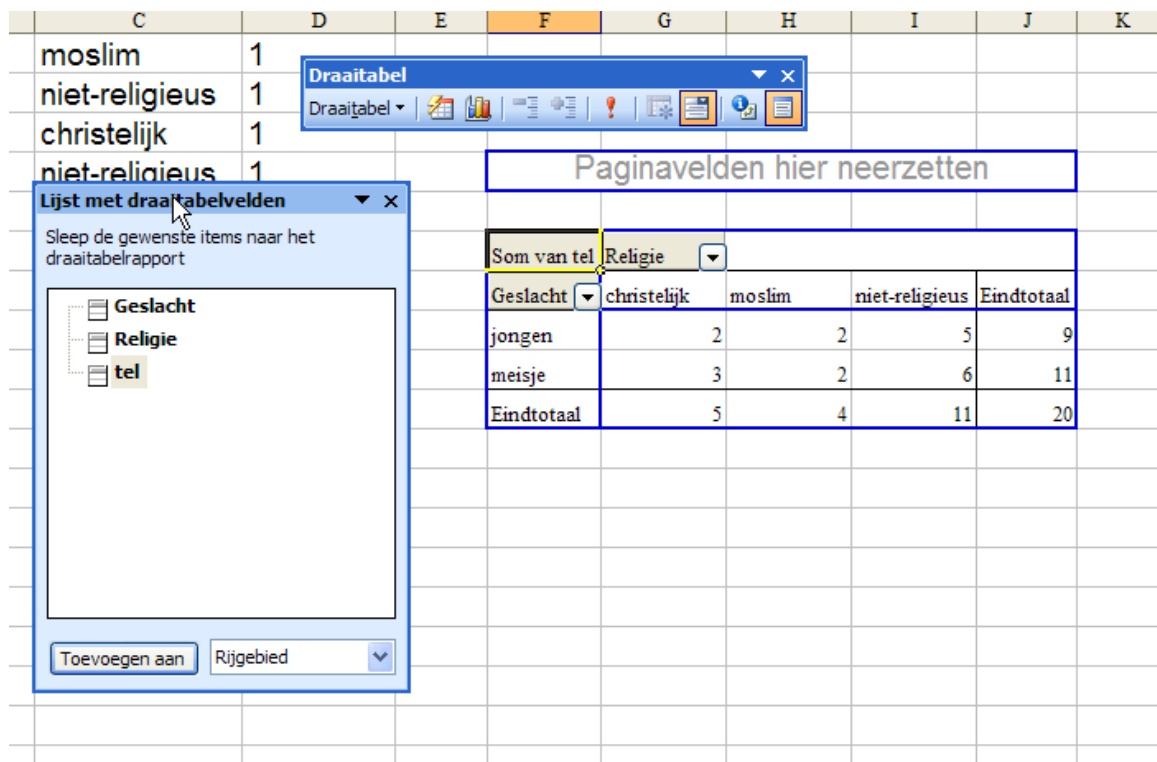
We gaan Figuur 4.3 even uitspitten. Er zijn twee vensters die er op dit moment toedoen.

Links staat het venster Lijst met draaitabelvelden. De drie velden die daarin staan zijn bekend: Geslacht, Religie en tel. Rechts staat een grijsvierkant met daarin verschillende teksten. Bovenaan lezen we 'Kolomvelden hier neerzetten'. In de kantlijn staat 'Rijvelden hier neerzetten'. In het midden staat 'Gegevensitems hier neerzetten'. Dat zijn er ook drie.

We gaan nu dus de namen Geslacht, Religie en tel van het linkervenster verplaatsen naar het rechtervenster. Dat kan door middel van selecteren en slepen, of door gebruik te maken van de knop toevoegen. We gaan in dit voorbeeld selecteren en slepen.

- Selecteer in linker venster Geslacht. Dit slepen we letterlijk op de tekst Rijvelden hier neerzetten.
- Vervolgens selecteren we in het linkervenster Religie en dat slepen we naar Kolomvelden hier neerzetten.
- Tot slot grijpen we tel in zijn nekvel en die plaatsen we in Gegevensitems hier neerzetten.

Wat we nu feitelijk gedaan hebben, komt neer op het volgende. Geslacht staat in de rijen. Religie is over de kolommen verdeeld. En in de cellen worden de 1-tjes bij elkaar opgeteld.



Figuur 4.4. De Wizard is nu echt klaar!⁴

Vraag: Is de kruistabel in Figuur 4.3 correct en wat zien we daar eigenlijk in? Er zijn een aantal getallen die ons inmiddels bekend moeten voorkomen. We hebben 20 leerlingen (klopt, staat rechtsonder). In de laatste kolom staan 11 meisjes en 9 jongens (klopt ook). De onderste rij laat de getallen 5, 4 en 11 zien. Dat is ook bekend, het is de frequentieverdeling van religie. Kortom, eindtotaal en randtotalen zijn correct weergegeven. En tot slot hebben we aan het begin van dit hoofdstuk vastgesteld dat twee jongens christelijk zijn. En dat klopt ook! Dan zal de rest ook wel goed zijn. Altijd een paar dingen controleren!

De kolom D1 in Figuur 4.1 is in zijn geheel gevuld met het cijfer 1. Nu is ook duidelijk waarom. In de rechter bovenhoek van Figuur 4.3 staat de tekst 'Som van tel'. De som van deze serie enen

⁴ Het veld 'Paginavelden hier neerzetten' kunt u gebruiken voor driedimensionale kruistabellen. Bijvoorbeeld als u behalve religie en geslacht ook nog een derde variabele zou willen gebruiken.

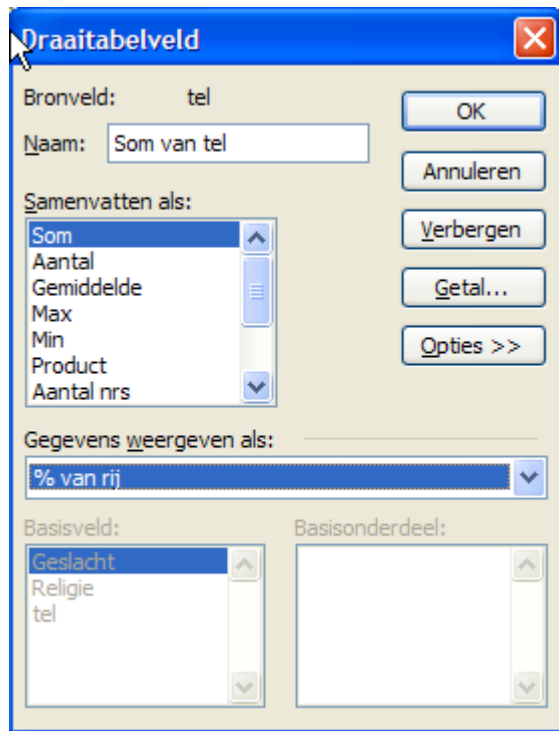
hebben we gebruikt om het aantal records aan te geven. Dit kunnen we laten zien door bijvoorbeeld het cijfer 1 in cel D20 te vervangen door een 3, en het cijfer 4 in cel D21 door een 3. We hebben nu dus 25 leerlingen. Maar er verandert nu iets in de kruistabel? Op het eerste gezicht in ieder geval niet.

In het vorige hoofdstuk werden dergelijke veranderingen meteen in een grafiek opgenomen. Dat is hier niet het geval en er is een extra handeling nodig. We klikken nu de werkbalk Draaitabel aan. In Figuur 4.4 staat deze werkbalk boven de tekst Paginavelden hier neerzetten. Met het aanklikken van de kruistabel licht de werkbalk Draaigrafiek op. Klik nu op het uitroepteken in de werkbalk en Excel telt het aantal records opnieuw. Dat wil zeggen: de laatste twee records worden nu 3 en 4 keer meegeteld (zie Figuur 4.5).

christelijk	1						
niet-religieus	1		Som van tel	Religie			
niet-religieus	1		Geslacht	christelijk	moslim	niet-religieus	Eindtotaal
niet-religieus	1		jongen	2	4	5	11
niet-religieus	1		meisje	3	5	6	14
christelijk	1		Eindtotaal	5	9	11	25
niet-religieus	1						
niet-religieus	1						
niet-religieus	1						
moslim	3						
moslim	4						

Figuur 4.5. De verschillen met de kruistabel in Figuur 4.4 zijn vet en dik onderstreept.

Evenals in Figuur 4.4, zijn in Figuur 4.5 drie velden in de bovenhoek grijs gekleurd. Als de kruistabel eenmaal is gemaakt, kunnen we door op deze velden te klikken de kruistabel nog op allerlei manieren aanpassen. Door te dubbelklikken op het veld Som van tel wordt het dialoogschermdraaitabelveld geopend. Klik nu eerst op de knop Opties en vervolgens kunnen we het uitrolmenu 'Gegevens weergeven als' gebruiken. In Figuur 4.6 hebben we al % van rij' geselecteerd.



Figuur 4.6. Na klikken op uitrolmenu Gegevens weergegeven als gevolgd door de keuze % van rij

In het geopende uitrolmenu kiezen we % van rij. Aangezien in de laatste rijen nog steeds de getallen 3 en 4 staan, wijzigen we deze ook weer in twee keer een 1, gevolgd door het aanklikken van het uitroepteken. Standaard – althans in onze default instelling van Excel - worden deze percentages met twee decimalen weergegeven. Dat is veel teveel, er zijn maar 20 leerlingen. Dit passen we aan door eerst de cellen in de kruistabel te selecteren en achtereenvolgens in menubalk op Opmaak te klikken, gevolgd door tabblad Celeigenschappen en vervolgens Percentage. We veranderen Decimalen in 0. Het resultaat van al deze handelingen is te zien in Figuur 4.7.

+	Som van tel	Religie			
	Geslacht	christelijk	moslim	niet-religieus	Eindtotaal
	jongen	22%	22%	56%	100%
	meisje	27%	18%	55%	100%
	Eindtotaal	25%	20%	55%	100%

Figuur 4.7. Rijpercentages.

In feite staat in Figuur 4.7 het antwoord op de vraag 'hoeveel procent van de meisjes niet-religieus is'. Dat blijkt 55% te zijn. De tweede vraag die we onszelf aan het begin van dit hoofdstuk gesteld hebben, hoeveel procent van alle islamitische leerlingen is een jongen?, kunnen we niet meteen terug vinden in Figuur 4.6. Daartoe hadden we in het in het zojuist geopende uitrolmenu niet '% van rij', maar '% van kolom' moeten selecteren. Als we dat gedaan hebben, krijgen we Figuur 4.8.

+	Som van tel	Religie			
	Geslacht	christelijk	moslim	niet-religieus	Eindtotaal
	jongen	40%	50%	45%	45%
	meisje	60%	50%	55%	55%
	Eindtotaal	100%	100%	100%	100%

Figuur 4.8. Kolompercentages.

Het antwoord is dus 50%. Maar dat hadden we al gezien in Figuur 4.4. Toch?

Al met al lijkt het een hoop werk om een kruistabel te maken. Toch is het na een paar keer oefenen in een oogwenk gebeurd. We hebben zelfs kolom D met al die enen, helemaal niet nodig. Dus:

- Verwijder de Figuur 4.8 door kolommen E tot en met K te selecteren, klik in menubalk op Bewerken gevolgd door Verwijderen.

- Ga met cursor in een ongebruikte cel staan, bijvoorbeeld F12. Klik in de menubalk op Data en selecteer Draaitabel en Draaigrafiek rapport ...
- Klik op de knop Volgende en selecteer de cellen B1 tot en met C21.
- Klik in Wizard Draaitabel en draaigrafiek - Stap 3 van 3 op Voltooien.
- Plaats nu dialoogscherf Lijst met draaivelden Geslacht in Rijkgebied door op knop Toevoegen te klikken. Deze keer slepen we niet!
- Selecteer in Religie, kies Kolomgebied en klik weer op knop Toevoegen.
- Selecteer nog een keer Religie, kies nu Gegevensgebied en klik weer op Toevoegen.
- KLAAR!

Aantal van Religie	Religie			
Geslacht	christelijk	moslim	niet-religieus	Eindtotaal
jongen	2	2	5	9
meisje	3	2	6	11
Eindtotaal	5	4	11	20

Figuur 4.9. Dezelfde kruistabel, maar nu met de Wizard gemaakt zonder de extra kolom D.

In tegenstelling tot de voorgaande kruistabellen staat in de linker bovenhoek niet langer 'Som van tel', maar 'Aantal van religie'. Een laatste opdracht: dubbelklik op de linkerbovenhoek in Figuur 4.9 en verander tekst in 'Aantal'. We hadden dezelfde kruistabel gekregen als we in de laatste serie commando's niet Religie maar Geslacht in het Gegevensgebied hadden geplaatst. In de linkerbovenhoek had dan 'Aantal van Geslacht' gestaan.

5. Rekenschema's

Veel gebruikte statistische kengetallen zoals modus, mediaan, variantie, standaarddeviatie, correlaties en regressieparameters kan men eenvoudig met behulp van de daartoe in Excel bestaande functies gewoon opvragen. Wat op het eerste gezicht misschien minder voor zich spreekt, is dat wij in dit hoofdstuk toch blijven werken met ouderwets aandoende rekenschema's om deze kengetallen zelf uit te rekenen. De desbetreffende functies komen weliswaar aan bod, maar het zwaartepunt ligt op zelf met rijen, kolommen en cellen aan de slag te gaan. De door ons gepresenteerde rekenschema's verschaffen inzicht in de relatie tussen afzonderlijke cijfers en de daarop gebaseerde kengetallen. Het rekenschema in de paragrafen 5.1 en 5.2 is gebaseerd op de definitieformules voor het berekenen van het gemiddelde, variantie, standaarddeviatie, covariantie en een correlatiecoëfficiënt. Het rekenschema voor het berekenen van regressieparameters in paragraaf 5.3 is gebaseerd op zogenaamde berekeningsformules. Hoewel deze berekeningsformules er ingewikkelder uitzien, zijn ze gemakkelijker in een rekenschema uit te schrijven.

5.1 Gemiddelde, variantie en standaarddeviatie

In hoofdstuk 2 hebben we de toetsresultaten van 12 leerlingen besproken. We hebben toen het gemiddelde over drie toetsen van iedere leerling uitgerekend. We gaan nu het gemiddelde en de standaarddeviatie van de eerste toets bepalen. De onderstaande formules gaan we omzetten in berekeningen over rijen en kolommen in Excel (de letter n geeft in onderstaande formules het aantal waarnemingen aan):

$$\text{Gemiddelde van } x: \bar{x} = \frac{\sum x}{n}$$

$$\text{Variantie van } x: s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

$$\text{Standaarddeviatie van } x: s = \sqrt{s^2}$$

Het is bij het zelf opstellen van een rekenschema van groot belang om altijd een paar cellen te reserveren om op te schrijven waar u wat heeft uitgerekend. Bovendien is het aan te raden om her en der cellen te onderstrepen als u over kolommen optelt. In Figuur 5.1 zijn allereerst de eerste toetsresultaten van twaalf leerlingen weggeschreven.

	A	B	C
1	CE+ A1		toets 1
2		leerling 1	4
3		leerling 2	6
4		leerling 3	7
5		leerling 4	8
6		leerling 5	6
7		leerling 6	5
8		leerling 7	6
9		leerling 8	7
10		leerling 9	6
11		leerling 10	6
12		leerling 11	5
13		leerling 12	9

Figuur 5.1. Het basismateriaal voor het maken van een rekenschema voor het bepalen van het toetsgemiddelde en de standaarddeviatie.

- Neem Figuur 5.1 over in Excel.
- Selecteer rij 2 in zijn geheel en plaats onder alle cellen in deze rij een dunne streep door in de werkbalk Randopmaak het juiste blokje te kiezen.
- Herhaal dit voor rij 14, nu kiezen we voor een dikke streep onder deze cellen via Randopmaak.
- Typ in cel C2 de letter 'x'. In plaats van 'toets 1' gaan we in het vervolg werken met de letter x.
- Typ in cel D15 'Gemiddelde' en maak dit woord vet door op vet in de werkbalk Opmaak te klikken.
- De formule aan het begin van deze paragraaf voor het berekenen van het gemiddelde van x zegt dat we eerst alle waarden van x bij elkaar moeten optellen ($\sum x$) om vervolgens te delen door het aantal waarnemingen. In Excel kunnen we dit uitschrijven met behulp van de functies SOM() en AANTAL(). Dus typ in cel C15 '=SOM(C3:C14)/AANTAL(C3:C14)'. De functie AANTAL() geeft hier het aantal cellen terug dat een getal bevat.

- Typ in cel D2 'x-gem(x)'. Hier staat dus neem de waarde van x (= toets 1) en trek daar de gemiddelde score van alle leerlingen van af. Het gemiddelde van x hebben we zojuist berekend in cel C15.

- Om het bovenstaande voor elke leerling uit te rekenen typen we in cel D3 '=C3-\$C\$15'. Het resultaat in cel D3 is -2,25. We hebben nu dus het begin gemaakt met het uitschrijven van de definitieformule voor de variantie ($x - \bar{x}$).

- Kopieer D3 en plak dit over de cellen D4 tot en met D14. Nu hebben we voor alle leerlingen uitgewerkt hoeveel ze afwijken van het gemiddelde ($x - \bar{x}$). Sommige kinderen zitten boven het gemiddelde en andere kinderen zitten onder het gemiddelde (minteken).

- Sommeer in cel D15 de waarden van de cellen D3 tot en met D14 met '=SOM(D3:D14)'. Het resultaat is 0. Als dit onverhoopt niet het getal 0 oplevert, dan heeft u een fout gemaakt. Deze laatste stap is volstrekt overbodig volgens onze formule, maar wel heel handig om te kijken of er geen fouten zijn gemaakt. Het resultaat van alle bovenstaande handelingen levert Figuur 5.2 op.

	A	B	C	D
1	CEL A1		toets 1	
2			X	x-gem(x)
3		leerling 1	4	-2,25
4		leerling 2	6	-0,25
5		leerling 3	7	0,75
6		leerling 4	8	1,75
7		leerling 5	6	-0,25
8		leerling 6	5	-1,25
9		leerling 7	6	-0,25
10		leerling 8	7	0,75
11		leerling 9	6	-0,25
12		leerling 10	6	-0,25
13		leerling 11	5	-1,25
14		leerling 12	9	2,75
15		Gemiddelde	6,25	0,00

Figuur 5.2. Rekenschema na vaststelling hoeveel iedere leerling van groepsgemiddelde afwijkt.

Ga de betekenis van de getallen in kolom D na. Leerling 1 scoorde een 4 (zie cel C4) op toets 1. In cel D4 kunt u zien dat deze leerling onder het gemiddelde scoorde (minteken) en wel 2,75 punten. Leerling 12 scoorde precies 2,75 boven het gemiddelde. Immers, 2,75 (cel C14) + 6,25 (cel D14) = 9 (cel D15).

We gaan nu verder met het berekenen van de variantie. Volgens de formule moet er achtereenvolgens nog gekwadeerd worden $(x - \bar{x})^2$, opgeteld ($\sum (x - \bar{x})^2$) en gedeeld door $n-1$.

- Typ in cel E2 de tekst '(x-gem(x))²'. Superscripten maken in Excel gaat op precies dezelfde wijze als in Word met behulp van Opmaak in het menu.

- Typ in cel E3 '=D3*D3'. Voor leerling 3 heeft u hiermee dus $(x - \bar{x})^2$ berekent.⁵

- Kopieer inhoud E3 naar de cellen E4 tot en met E14. Nu doen we $(x - \bar{x})^2$ voor de overige elf leerlingen.

- Sommeer over de cellen E3 tot en met E14 door de cellen te selecteren en te klikken op Autosom () in de werkbalk. In cel E15 staat '=SOM(E3:E14)' en verschijnt het getal 20,25.

Kortom, de uitkomst van $\sum (x - \bar{x})^2$ verschijnt in cel E15.

- Er moet nu enkel nog gedeeld worden door $n-1$ en de formule voor het berekenen van de variantie van toets 1 is in het rekenschema weggeschreven. Ga in cel E15 staan en verplaats daarna de cursor naar de Formulebalk. Typ nu '=SOM(E3:E14)/(AANTAL(C3:C14)-1)'. Kijk goed naar de haakjes in de formule. Er zijn twee soorten haakjes: de haakjes die bij de functies SOM() en AANTAL() horen én een paar vette haakjes om aan te geven waardoor er gedeeld moet worden: namelijk AANTAL()-1. Het resultaat in cel E15=1,84.

- Typ in cel D16 de tekst 'Standaarddeviatie'. Maak deze cel vet.

- Typ in cel E16 '=Wortel(E15)'. Uit de eerder gegeven definitieformule wisten we al dat de standaarddeviatie gelijk is aan de wortel uit de variantie. Bedenk dat we in kolom E eerder hebben gekwadeerd, nu doen we het omgekeerde.

We zijn klaar en het resultaat staat in Figuur 5.3. Zorg voor een rustige opmaak en laat vooral het aantal decimalen niet alle kanten opspringen.

	A	B	C	D	E
1	CEL A1		toets 1		
2			X	x-gem(x)	$(x-gem(x))^2$
3		leerling 1	4	-2,25	5,06
4		leerling 2	6	-0,25	0,06
5		leerling 3	7	0,75	0,56
6		leerling 4	8	1,75	3,06
7		leerling 5	6	-0,25	0,06
8		leerling 6	5	-1,25	1,56
9		leerling 7	6	-0,25	0,06
10		leerling 8	7	0,75	0,56
11		leerling 9	6	-0,25	0,06
12		leerling 10	6	-0,25	0,06
13		leerling 11	5	-1,25	1,56
14		leerling 12	9 ⁺	2,75	7,56
15		Gemiddelde	6,25	0,00	1,84
16			Standaarddeviatie		1,36

Figuur 5.3. Afgerond rekenschema voor het bepalen van het toetsgemiddelde en de standaarddeviatie.

Het gemiddelde van de twaalf leerlingen op toets 1 is 6,25, de variantie bedraagt 1,84 en de standaarddeviatie is 1,36. Zoals gezegd, Figuur 5.3 geeft het volledige rekenschema weer voor het berekenen van gemiddelde, variantie en standaarddeviatie. Het leuke van Excel is nu dat we op eenvoudige wijze inzichtelijk kunnen maken wat er gebeurt met deze drie kengetallen als we de individuele toetsresultaten veranderen. Voer de volgende opdracht uit en probeer voor jezelf te verklaren waarom de kengetallen veranderen: Typ in cel C3 een 8, kijk naar het schema, typ in C4 een 8, kijk naar het schema. Herhaal dit stapsgewijs tot en met C14.

⁵ U had hier ook '=MACHT(D3;2)' kunnen typen. Dus cel D3 tot de macht 2.

We keren nog een keer terug naar de cijfers in Figuur 5.3. Stel er is een extra leerling in de klas gekomen. Aangezien deze leerling op de vorige school ook toets 1 heeft afgelegd, wilt u de score van dit kind invoegen. We doen dit tussen de leerlingen 7 en 8. De naam van de leerling is leerling 13.

- Selecteer rij 10 in zijn geheel.
- Klik in menubalk op Invoegen gevolgd door Rij.
- Typ 'Leerling 13' in cel B10.
- Leerling 13 heeft een 9 behaald op toets 1. Typ dit in cel C10.

Het resultaat staat in Figuur 5.4.

	A	B	C	D	E
1	CEL A1		toets 1		
2			x	x-gem(x)	$(x-gem(x))^2$
3		leerling 1	4	-2,46	6,06
4		leerling 2	6	-0,46	0,21
5		leerling 3	7	0,54	0,29
6		leerling 4	8	1,54	2,37
7		leerling 5	6	-0,46	0,21
8		leerling 6	5	-1,46	2,14
9		leerling 7	6	-0,46	0,21
10		leerling 13	9	2,54	6,44
11		leerling 8	7	0,54	0,29
12		leerling 9	6	-0,46	0,21
13		leerling 10	6	-0,46	0,21
14		leerling 11	5	-1,46	2,14
15		leerling 12	9	2,54	6,44
16		Gemiddelde	6,46	0,00	2,27
17			Standaarddeviatie		1,51

Figuur 5.4. Rekenschema na invoeren van een extra leerling.

Figuur 5.4 laat zien dat ten opzichte van het Figuur 5.3 het gemiddelde iets omhoog is gegaan, evenals de variantie en standaarddeviatie. Waarom eigenlijk?

Het is zeer verstandig om na het invoegen van een rij (of een kolom) altijd te verifiëren of de formules ook automatisch zijn aangepast.

- In cel C16 staat nu '=SOM(C3:C15)/AANTAL(C3:C15)' Dus C14 is vervangen door C15.
- In cel D16 staat nu '=SOM(D3:D15)', Ook hier is C14 vervangen door C15. Maar de uitkomst als zodanig is natuurlijk weer gelijk aan 0!
- In de cellen D4 tot en D15 is de verwijzing naar '\$C\$15' automatisch vervangen door '\$C\$16'. De dollartekens inclusief.
- In cel E16 staat nu '=SOM(E3:E15)/(AANTAL(C3:C15)-1)' en weer is cel C14 vervangen door C15.
- In cel E17 staat nu '=WORTEL(E16)'.

We hebben bij het uitwerken van het rekenschema twee nieuwe functies gebruikt, namelijk AANTAL() en WORTEL(). Aangezien Excel als een grote rekenmachine te beschouwen is, mag het niet al te veel verbazing opwekken dat worteltrekken tot de mogelijkheden behoort. Om na te gaan of we echt alles goed berekend hebben, typt u zelf de formules in voor gemiddelde, variantie en standaarddeviatie zoals weergegeven in Figuur 5.5.

12		leerling 9	6	-0,46	0,21	
13		leerling 10	6	-0,46	0,21	
14		leerling 11	5	-1,46	2,14	
15		leerling 12	9	2,54	6,44	
16		Gemiddelde	6,46	0,00	2,27	
17			Standaarddeviatie	1,51		
18						
19	+	6,46	=GEMIDDELDE(C3:C15)			
20		2,27	=VAR(C3:C15)			
21		1,51	=STDEV(C3:C15)			
22						

Figuur 5.5. De eigenlijke formules voor gemiddelde, variantie en standaarddeviatie.

Twee andere, vaak gebruikte centrummaten zijn de modus en de mediaan. U kunt deze opvragen met '=MODUS()' en '=MEDIAAN()'. Aan de hand van de laatste functie demonstreren we het dialoogschermd Functie invoegen. Klik daartoe op het pijltje naast autosom Σ en het dialoogschermd Functie invoegen wordt geopend. Vervolgens selecteren we Statistisch in het uitrolmenu van Of selecteer een categorie. Daarna hebben we met behulp van de schuifbalk van Selecteer een functie de mediaan op gezocht (zie Figuur 5.6).



Figuur 5.6. Dialoogschermd Functie invoegen nadat we de functie MEDIAAN geselecteerd hebben.

Met het klikken op OK in Figuur 5.6 wordt Figuur 5.7 geopend. Geef de cellen C3 tot en met C15 op als **Getal1**, dus '(C3:C15)', en het resultaat is 6. Kortom, de mediaan is 6. Sluit het scherm Functieargumenten door op OK te klikken.



Figuur 5.7. Dialoogscherm Functieargumenten.

5.2 Covariantie en correlatiecoëfficiënt

In deze paragraaf gaan we het in paragraaf 5.1 gemaakte rekenschema verder uitbreiden met de covariantie en correlatiecoëfficiënt tussen x (toets 1) en y (toets 5). De formules voor beide kengetallen zijn als volgt:

$$\text{Covariantie van } x \text{ en } y: s_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{n - 1}$$

$$\text{Correlatie tussen } x \text{ en } y: r_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{n - 1} \cdot \frac{1}{s_x * s_y}$$

De correlatie tussen x en y is gelijk aan de covariantie gedeeld door het product van de standaarddeviatie van x (hier aangeduid als s_x) en de standaarddeviatie van y (hier aangeduid als s_y). Ga dat even na voor uzelf.

Bij het berekenen van de covariantie nemen we Figuur 5.3 als uitgangspunt. De toetsuitslagen op de 5^e toets (in het vervolg ook aangeduid met de letter y) zijn in kolom F van Figuur 5.8 opgenomen.

	A	B	C	D	E	F	
+							
1		CEL B1	toets 1			toets 5	
2			x	x-gem(x)	$(x-gem(x))^2$	y	
3		leerling 1	4	-2,25	5,06	5	
4		leerling 2	6	-0,25	0,06	6	
5		leerling 3	7	0,75	0,56	8	
6		leerling 4	8	1,75	3,06	8	
7		leerling 5	6	-0,25	0,06	7	
8		leerling 6	5	-1,25	1,56	5	
9		leerling 7	6	-0,25	0,06	9	
10		leerling 8	7	0,75	0,56	8	
11		leerling 9	6	-0,25	0,06	7	
12		leerling 10	6	-0,25	0,06	6	
13		leerling 11	5	-1,25	1,56	4	
14		leerling 12	9	2,75	7,56	8	
15		Gemiddelde	6,25	0,00	1,84		
16			Standaarddeviatie		1,36		

Figuur 5.8. Startpositie voor het uitwerken van een rekenschema voor de covariantie en de correlatiecoëfficiënt.

Met wat knippen en plakken kunt u in een oogwenk het gemiddelde en de standaarddeviatie van y (= toets 5) bepalen.

- Kopieer cel C15 naar F15. Het gemiddelde van y is 6,75.
- Kopieer de cellen D3 en E3 naar G3 en H3. Aangezien de bevroren cel \$C\$15 is gekopieerd, dient u in cel G3 de verwijzing naar cel \$C\$15 te vervangen door \$F\$15. Dus G3 is '=F3-\$F\$15'.
- Kopieer de cellen G3 en H3, selecteer daarna het gebied G4:H14, gevolgd door <Ctrl>+<V>.

- Selecteer de cellen C15:E16, gevolgd door <Ctrl>+<C>
- Selecteer de cellen F15: H16, gevolgd door <Ctrl>+<V>. Cel C15 had u overigens al naar F15 gekopieerd. Het resultaat van dit kopiëren en plakken ziet u in Figuur 5.9.

	A	B	C	D	E	F	G	H
1		CEL B1	toets 1			toets 5		
2			x	x-gem(x)	(x-gem(x)) ²	y	y-gem(y)	(y-gem(y)) ²
3		leerling 1	4	-2,25	5,06	5	-1,75	3,06
4		leerling 2	6	-0,25	0,06	6	-0,75	0,56
5		leerling 3	7	0,75	0,56	8	1,25	1,56
6		leerling 4	8	1,75	3,06	8	1,25	1,56
7		leerling 5	6	-0,25	0,06	7	0,25	0,06
8		leerling 6	5	-1,25	1,56	5	-1,75	3,06
9		leerling 7	6	-0,25	0,06	9	2,25	5,06
10		leerling 8	7	0,75	0,56	8	1,25	1,56
11		leerling 9	6	-0,25	0,06	7	0,25	0,06
12		leerling 10	6	-0,25	0,06	6	-0,75	0,56
13		leerling 11	5	-1,25	1,56	4	-2,75	7,56
14		leerling 12	9	2,75	7,56	8	1,25	1,56
15		Gemiddelde	6,25	0,00	1,84	6,75	0,00	2,39
16			Standaarddeviatie	1,36	Standaarddeviatie	1,54		

Figuur 5.9. Rekenschema voor het berekenen van de covariantie en de correlatiecoëfficiënt tussen x en y nadat van beide variabelen het gemiddelde, de variantie en de standaarddeviatie zijn berekend.

In Figuur 5.9 is dus twee keer hetzelfde rekenschema uitgevoerd: een keer voor x en een keer voor y. We willen echter weten of de scores op x (toets 1) en y (toets 5) met elkaar samenhangen en hoe groot die samenhang is. Daartoe gaan we nog eens twee kolommen extra aanmaken. We beginnen met kolom I.

- Typ in de 2^e cel van kolom I de tekst '(x-gem(x))*(y-gem(y))'. Ten einde alle kolommen ongeveer even breed te houden, is cel I2 samengevoegd met cel I1. Onderzoek zelf de mogelijkheden daartoe met behulp van het dialoogscherm Celeigenschappen dat u kunt openen door op Opmaak in de menubalk te klikken.

- In cel I3 typt u '=D3*G3'. Kopieer cel I3, selecteer I4:I14, gevolgd door <Ctrl>+<V> . Dus in de cellen van kolom I staat nu dat deel van de definitieformule van de covariantie dat achter het sommatieteken staat: $(x - \bar{x})(y - \bar{y})$.

- In cel I15 berekent u de eigenlijke covariantie door op te tellen en te delen. In cel C15 typt u '=SOM(I3:I14)/(AANTAL(I3:I14)-1)'. Met het gebruik van de functie SOM() heeft u in cel I15 het sommatieteken (Σ) van de formule voor de covariantie uitgevoerd. Door ook nog eens te delen in deze cel door (AANTAL()-1) is de formule voor de covariantie in cel I15 volledig gedefinieerd. Alle zojuist uitgevoerde stappen zijn weergegeven in Figuur 5.10. De uitkomst in cel I15 bedraagt 1,52.

- Typ in cel H17 de tekst 'Correlatie' en maak deze vet.

- Typ in cel G17 '=I15/(E16*H16)'. De correlatie tussen x en y is bepaald door de covariantie gedeeld door de beide standaarddeviaties.

De resultaten van uw handelingen staan in Figuur 5.10.

	A	B	C	D	E	F	G	H	I
1		CEL B1	toets 1			toets 5			
2			x	x-gem(x)	(x-gem(x)) ²	y	y-gem(y)	(y-gem(y)) ²	(x-gem(x)) *(y-gem(y))
3		leerling 1	4	-2,25	5,06	5	-1,75	3,06	3,94
4		leerling 2	6	-0,25	0,06	6	-0,75	0,56	0,19
5		leerling 3	7	0,75	0,56	8	1,25	1,56	0,94
6		leerling 4	8	1,75	3,06	8	1,25	1,56	2,19
7		leerling 5	6	-0,25	0,06	7	0,25	0,06	-0,06
8		leerling 6	5	-1,25	1,56	5	-1,75	3,06	2,19
9		leerling 7	6	-0,25	0,06	9	2,25	5,06	-0,56
10		leerling 8	7	0,75	0,56	8	1,25	1,56	0,94
11		leerling 9	6	-0,25	0,06	7	0,25	0,06	-0,06
12		leerling 10	6	-0,25	0,06	6	-0,75	0,56	0,19
13		leerling 11	5	-1,25	1,56	4	-2,75	7,56	3,44
14		leerling 12	9	2,75	7,56	8	1,25	1,56	3,44
15		Gemiddelde	6,25	0,00	1,84	6,75	0,00	2,39	1,52
16			Standaarddeviatie		1,36	Standaarddeviatie		1,54	
17					CEL E17			Correlatie	0,73

Figuur 5.10. Afgerond rekenschema voor het bepalen van de covariantie en correlatie coëfficiënt tussen twee variabelen x en y.

De covariantie tussen x en y bedraagt 1,52. De correlatiecoëfficiënt is 0,73. Zoals bekend, de correlatie is altijd een getal tussen -1 en 1. U kunt uw correlatieberekening controleren door gebruik te maken van de functie '=CORRELATIE(C3:C14;F3:F14)'.⁶

Nu u beschikt over dit rekenschema, wordt u aangemoedigd allerlei veranderingen in de waarden van x en y door te voeren. Geniet ervan hoe uw kengetallen veranderen en ga vooral na waarom dat zo is. Een paar voorbeelden:

- Kopieer de waarden van x in y.
- Kopieer de waarden van x in y en sorteer in y de toetsuitslagen op volgorde van klein naar groot.
- Kopieer de waarden van x in y en sorteer in x de toetsuitslagen op volgorde van groot naar klein en in y de toetsuitslagen op volgorde van klein naar groot.
- Maak of x of y in zijn geheel negatief (al is dit moeilijk voorstelbaar bij toetsuitslagen).
- Typ in x overal het getal 1.
- Typ in x overal het getal 1 en in y overal 10.
- Bedenk zelf allerlei veranderingen in de waarden van x en y.

5.3 Regressiecoëfficiënten

Bij het berekenen van gemiddelden, varianties, standaarddeviaties, covariantie en correlatiecoëfficiënt zijn we uitgegaan van een rekenschema gebaseerd op de definitieformules. Het is ook mogelijk om deze kengetallen uit te rekenen met kwadraatsommen. Het in deze paragraaf gebruikte rekenschema is gebaseerd op de berekeningsformules. De berekeningsformules zijn afgeleid uit de definitieformules. In het computerloze verleden was dit bijzonder handig. Een hedendaags voordeel is dat het eenvoudig is om op de hand regressiecoëfficiënten te berekenen.

De onderstaande berekeningsformules gaan we in Excel uitwerken.

⁶-Helaas kunt u de covariantie niet zondermeer controleren door de functie '=COVARIANTIE(H3:H14;I3:I14)' op te geven. Excel deelt hier namelijk door n en niet door n-1. U krijgt wel de juiste uitkomst voor de covariantie door te vermenigvuldigen met 12/11. U typt in cel H14 dus '=COVARIANTIE(H3:H14;I3:I14) *(12/11)'. In deze KG-publicatie delen we bij het berekenen van de (co)variantie steeds door n-1. We geven hiermee impliciet te kennen dat we onze twaalf leerlingen als een steekproef uit een grotere populatie beschouwen.

Gemiddelde van x: $\bar{x} = \frac{\sum x}{n}$

Variantie van x: $s^2 = \frac{\sum (x - \bar{x})^2}{n - 1} = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n - 1}$

Covariantie van x en y: $s_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{n - 1} = \frac{\sum xy - \frac{(\sum x \sum y)}{n}}{n - 1}$

Stel $y = a + bx$, dan geldt voor het berekenen van

de richtingscoëfficiënt b: $b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$

de intercept a: $a = \bar{y} - b\bar{x}$

De bovenstaande formules kunnen in het begin lastig zijn om te lezen. De remedie is eenvoudig: lees het nog een keer. Het is opvallend dat er veel sommatietekens en kwadraten in zijn verwerkt. Om deze berekeningsformules in Excel uit te werken, is het overzichtelijk om voor de kolomtotalen (dat is precies wat er met de Σ bedoeld wordt) een aparte rij te reserveren. In Figuur 5.11 is het gehele rekenschema uitgewerkt.

	A	B	C	D	E	F
1	CEL A1	toets 1		toets 5		
2		x	x²	y	y²	xy
3	leerling 1	4	16	5	25	20
4	leerling 2	6	36	6	36	36
5	leerling 3	7	49	8	64	56
6	leerling 4	8	64	8	64	64
7	leerling 5	6	36	7	49	42
8	leerling 6	5	25	5	25	25
9	leerling 7	6	36	9	81	54
10	leerling 8	7	49	8	64	56
11	leerling 9	6	36	7	49	42
12	leerling 10	6	36	6	36	36
13	leerling 11	5	25	4	16	20
14	leerling 12	9	81	8	64	72
15	Totaal	75	489	81	573	523
16						
17	Aantal leerlingen	12				CEL F17
18	Gemiddelde	6,25		6,75		
19	(Co)variantie $\oplus$	1,84		2,39		1,52
20	Richtingscoëfficiënt b					0,83
21	Intercept a					1,58

Figuur 5.11 Rekenschema gebaseerd op berekeningsformules voor het bepalen van gemiddelden, (co)variantie en regressiecoëfficiënten.

Maak Figuur 5.11 na aan de hand van de onderstaande toelichting op de verschillende cellen. Hieronder volgt in het kort wat in de diverse cellen is uitgerekend. Ga voor uzelf steeds na naar welke cellen wordt verwezen en hoe het getal in de betreffende cel tot stand gekomen is.

Eerst worden de kwadraten en producten van x en y uitgewerkt.

cel C3 '=B3*B3', ..., cel C14 '=B14*B14'

cel E3 '=D3*D3', ..., cel E14 '=D14*D14'

cel F3 '=B3*D3', ..., cel F14 '=B14*D14'

In rij 15 zijn alle kolomtotalen (Σ) weggeschreven:

cel B15 '=SOM(B3:B14)'; in B15 staat dus $\sum x$.

cel C15 '=SOM(C3:C14)'; in C15 staat dus $\sum x^2$.

cel D15 '=SOM(D3:D14)'; in D15 staat dus $\sum y$.

cel E15 '=SOM(E3:E14)'; in F15 staat dus $\sum y^2$.

cel F15 '=SOM(F3:F14)'; u raadt het al, in G15 staat $\sum xy$.

Het aantal leerlingen zetten we gemakshalve in een aparte cel: B17 '=AANTAL(B3:B14)'.

Bij het uitwerken van de eigenlijke formules kan steeds naar de uitkomst in B17 worden verwezen. Dat vinden we net iets overzichtelijker dan steeds de functie AANTAL() in te schakelen.

Berekeningen voor x (toets 1):

Gemiddelde van x: cel B18 '=SOM(B3:B14)/B17'

Variantie van x: cel B19 '=(C15-(B15*B15)/B17)/(B17-1)'

Berekeningen voor y (toets 5):

Gemiddelde van y: cel D18 '=SOM(D3:D14)/B17'

Variantie van y: cel D19 '=(E15-(D15*D15)/B17)/(B17-1)'

Berekeningen voor de samenhang tussen x en y:

Covariantie van x en y: cel F19 '=(F15-(B15*D15)/B17)/(B17-1)'

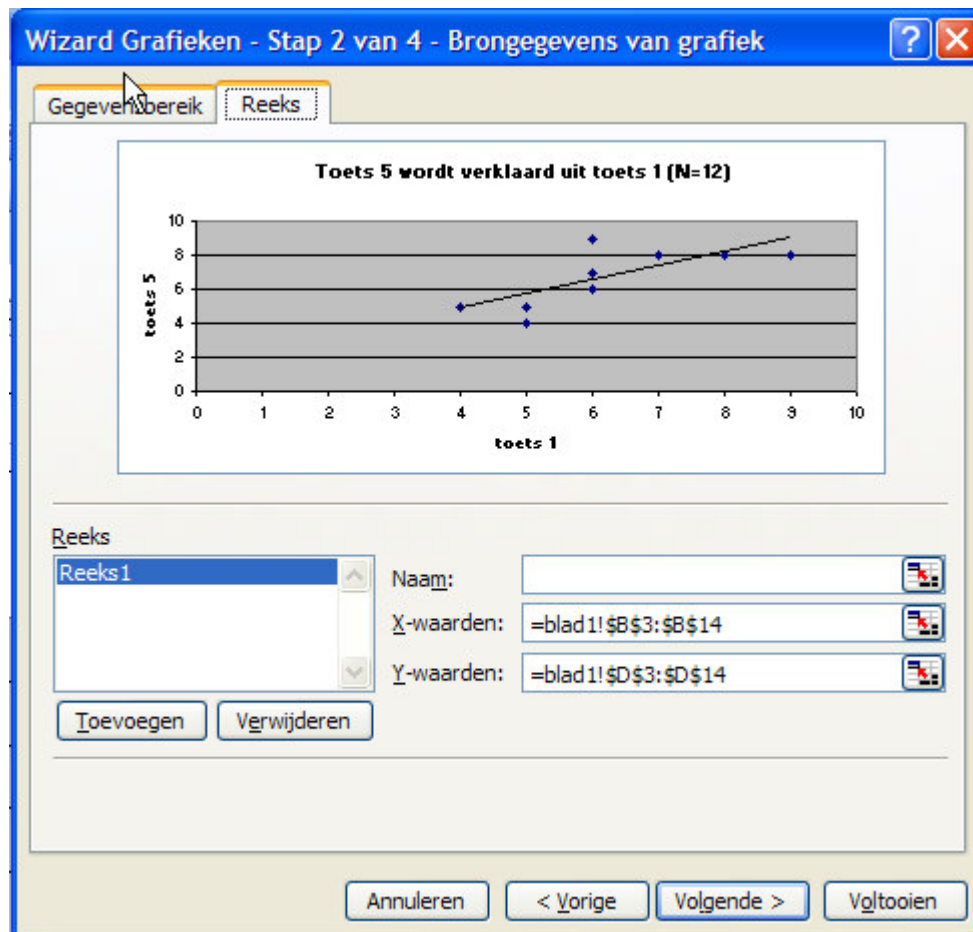
Richtingscoëfficiënt van lijn: cel F20 '=(B17*F15-(B15*D15))/(B17*C15 -B15*B15)'

Intercept van de rechte lijn: cel F21 '=D18-F20*B18'.

U kunt de regressielijn die u in Figuur 5.11 heeft berekend ($\hat{y} = 1,58 + 0,83x$), ook in een grafiek verwerken. U maakt eerst een scatterplot van de toetsuitslagen en daarna laat u Excel een rechte lijn tekenen.

- Ga in een lege cel staat, bijvoorbeeld I6
- Open het dialoogscherf Wizard Grafieken - Stap 1 van 4 – Grafiektype en kies Spreiding in tabblad Standaardtypen. Klik op Volgende
- Selecteer de cellen B3:B14 en de cellen D3:D14 in één keer met behulp van het indrukken van de <Ctrl> toets in het tabblad Gegevensbereik. Dat zijn dus de toetsresultaten op toets 1 en toets 5.

- Klik op tabblad Reeks, u ziet dan Figuur 5.12.



Figuur 5.12. Dialogscherm Wizard Grafieken - Stap 1 van 4 – Brongegevens van grafiek.

- Verifieer in figuur 5.12 nadrukkelijk dat toets 1 ook als X-waarde in de grafiek is opgenomen. U ziet dit door de tekst in Figuur 5.12 achter X-waarden te lezen. Hier wordt verwezen naar de cellen B3:B14 op blad1!
- Verifieer even nadrukkelijk dat toets 5 ook als Y-waarde in de grafiek is opgenomen. We hebben namelijk de resultaten van toets 5 (y) uit de resultaten op toets 1 (x) verklaard en dus niet de resultaten van toets 1 uit de resultaten van toets 5.
- Klik na verificatie op de knop Volgende.
- Typ als titel: 'Toets 5 wordt verklaard uit toets 1 (N=12)'.
- Typ bij Waardeas (X) 'toets 1'

- Typ bij Waardeas (y) 'toets 5'
- Sluit de Wizard met een klik op Voltooien.
- Selecteer in de grafiek Reeks 1 gevolgd door Delete.
- Klik in de menubalk op Grafiek, gevolgd door Trendlijn toevoegen ...⁷
- Kies in het dialoogscherf Trendlijn het tabblad Type (Type trend/regressie) en klik op het hokje Lineair (Zie Figuur 5.13).

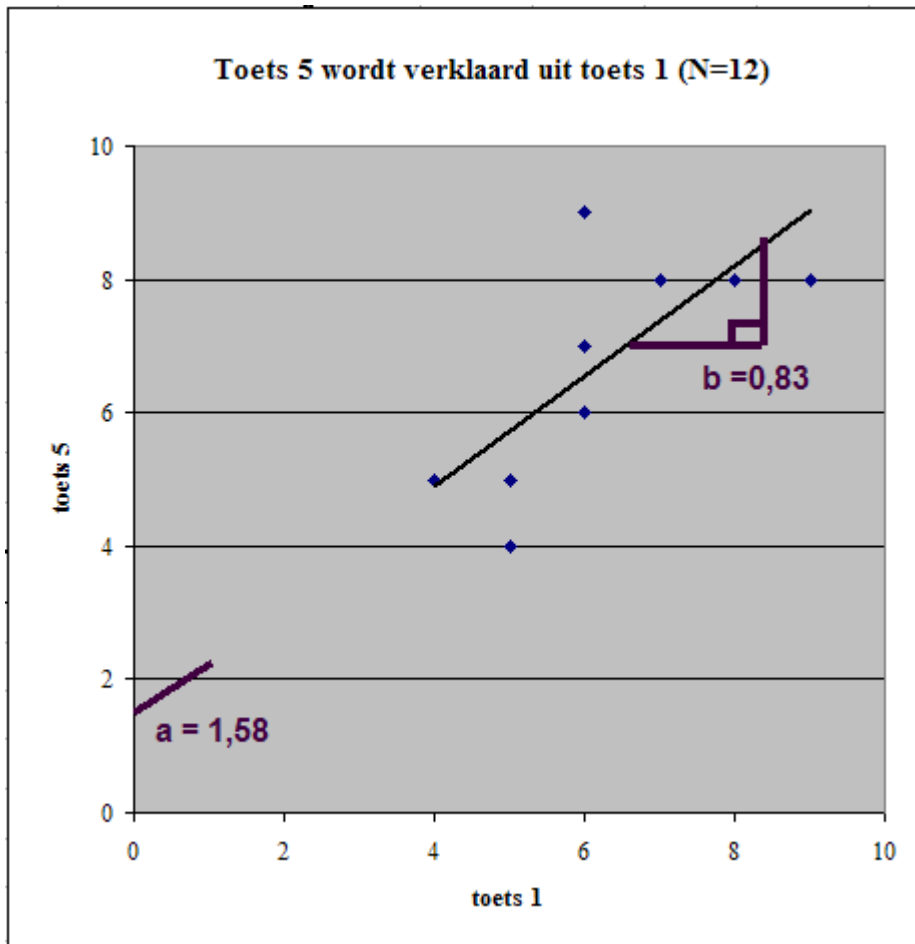


Figuur 5.13. Tabblad Type van dialoogscherf Trendlijn.

- Klik op knop OK.

De onderstaande grafiek inclusief uw regressielijn $\hat{y} = 1,58 + 0,83x$, verschijnt. We hebben er zelf de waarden van intercept a en richtingscoëfficiënt b in getekend. Zorg ervoor dat de beide assen even lang zijn. U moet Figuur 5.14 enkel nog interpreteren en het werk zit erop.

⁷ Zoals in hoofdstuk 3 al is opgemerkt ziet u Grafiek in de menubalk verschijnen in plaats van Data als u uw Grafiekvenster heeft geselecteerd.



Figuur 5.14.

Als u nu heel erg twijfelt aan de juistheid van uw grafiek, verander dan een toetsresultaat van toets 1 of toets 5 *in extremis* en kijk wat er gebeurt. Verander daartoe bijvoorbeeld de score op toets 1 van de 6^e leerling van 5 in 50. U ziet precies meteen waarom Excel in dit opzicht zo verdienstelijk kan zijn.

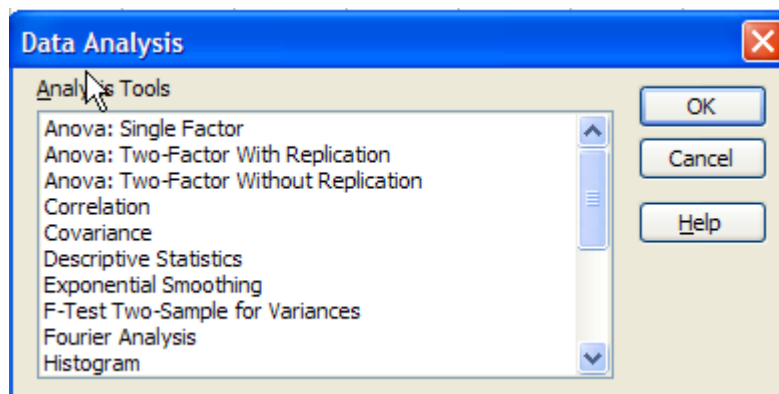
Nog een laatste opmerking over Figuur 5.14. De regressielijn (en dus de waarden voor a en b) in Figuur 5.14 wordt in Excel gegenereerd op basis van de commando's die in het Grafiekvenster en met behulp van Grafiek in de menubalk zijn gegeven. Hiervoor worden van het rekenschema in Figuur 5.11 enkel en alleen de cellen B3:B14 en D3:D14 gebruikt en dus niet uw berekeningen. Maak dus ook gebruik van uw timmermansoog bij het vergelijken van Figuur 5.11 met Figuur 5.14.

6 Nawoord

We hebben in deze KG-publicatie nog niet eens een fractie, en eigenlijk nog veel minder dan dat, van de verschillende mogelijkheden in Excel besproken. Wel hebben we voor een andere invalshoek gekozen dan doorgaans gebruikelijk is in zowel de onderzoekspraktijk als ook in gangbare Excel leerboeken. Lezers worden aangemoedigd om eindeloos 'met de hand' formules te schrijven en te gebruiken. Het is leuk, het is leerzaam, het is relatief gemakkelijk en u heeft het gevoel zelf baas te blijven over wat u uitrekent. Dat is vaak inzichtelijker dan werken met grote statistische software pakketten, zeker als men over relatief weinig statistische kennis beschikt. Toch zou deze KG-publicatie niet volledig zijn als we u niet zouden wijzen op een aparte, door uzelf in Excel te installeren statistiekmodule.

Als de statistiekmodule Gegevensanalyse niet beschikbaar is, kunt u het Analysis ToolPak laden door op Invoegtoepassingen in het menu Extra te klikken. Vervolgens schakelt in de lijst Beschikbare invoegtoepassingen het selectievakje Analysis ToolPak in en klikt u op OK.

Om de Gegevensanalyse te kunnen gebruiken, klikt u op Extra in de Menubalk en vervolgens selecteert u Data Analysis U kunt met behulp van dit dialoogscherf verschillende standaard statistische analyses uitvoeren (zie Figuur 6.1).



Figuur 6.1. Het dialoogscherf Data Analysis in Excel.

Tot slot wensen we u veel inzicht, maar vooral ook plezier bij het berekenen en interpreteren van statistische kengetallen.

Bijlage 1 Literatuur

Hogeschool Utrecht en Zunderop Beleidsadvies en Management (2007). *Koers 2012. HU kiest richting*. Hogeschool Utrecht.