

8 Statistiek met de computer

Voorkennis Werken met Excel

Bladzijde 124

1 *

Bladzijde 125

- 2 a Gebruik de functie SOM.
 $Y2 = \text{SOM}(B2:X2)$ geeft 82 punten.
- b $Y3 = \text{SOM}(B3:X3)$ geeft 61 punten voor Joris.
Lisa had de meeste punten, Kim de minste.
- c $Z2 = Y3 / \text{SYS}2 * 9 + 1$ geeft 7,69512.
- d $Z2 = \text{AFRONDEN}(Y3 / \text{SYS}2 * 9 + 1; 1)$ geeft 7,7 voor Joris.
- e $B27 = \text{GEMIDDELDE}(B3:B25)$
 $B28 = \text{STDEV.P}(B3:B25)$
 $B29 = \text{MIN}(B3:B25)$
 $B30 = \text{KWARTIEL}(B3:B25; 1)$
 $B31 = \text{MEDIAAN}(B3:B25)$
 $B32 = \text{KWARTIEL}(B3:B25; 3)$
 $B33 = \text{MAX}(B3:B25)$
Met de vulgreep bereken je in één keer de overige waarden.
- f $B34 = B32 - B30$ en gebruik de vulgreep.
Bij de vragen 14 en 15 is de kwartielafstand 2,5.
 $B35 = B27 - B31$ en gebruik de vulgreep.
Bij de vragen 2, 3 en 20 is het gemiddelde gelijk aan de mediaan.

Bladzijde 126

- 3 a $J3 = \text{AFRONDEN}((B3 * \$B\$1 + C3 * \$C\$1 + D3 * \$D\$1 + \dots + I3 * \$I\$1) / \text{SOM}(\$B\$1:\$I\$1); 1)$
De eindcijfers zijn achtereenvolgens 6,7 5,7 7,2 3,8 6,9 5,7 6,8 5,2 4,9 4,9 5,4 6,4 4,8 7,9 6,3 6,0 6,3 5,3 7,0 8,1 en 6,6.
- b $J25 = \text{GEMIDDELDE}(J3:J23)$ en weergave instellen op twee decimalen geeft 6,09.
- c $I1 = 6$ geeft 6,10.
Het scheelt 0,01 punt.

4 *

Bladzijde 127

5 *

8.1 Werken met datasets

Bladzijde 129

- 1 a De dataset bevat 11 variabelen.
- b Kiezen van 'Aantal' in cel A23239 van de totalenrij geeft 23237.
De dataset bevat van 23237 dagen gegevens.
- c *
- d Sorteer de kolom *MaximaleTemperatuur(°C)* met 'Sorteren van hoog naar laag'.
De warmste dag was 19 juli 2006.
- e Filter de kolom *GemiddeldeTemperatuur(°C)* op 'Is kleiner dan of gelijk aan' 0.
Kiezen van 'Aantal' in cel D23239 van de totalenrij geeft 1561.
Op 1561 dagen was het gemiddeld 0 °C of kouder.
- f Filter uit vraag e verwijderen.
Filter de kolom *MaximaleTemperatuur(°C)* op 'Is groter dan of gelijk aan' 30.
Kiezen van 'Gemiddelde' in cel F23239 van de totalenrij geeft 31,512...
De gemiddelde temperatuur tijdens de tropische dagen was 31,5 °C.

- g Filter uit vraag f verwijderen.
 Filter de kolom *Datum* op 1965.
 Kiezen van 'Som' in cel J23239 van de totalenrij geeft 1151,9 mm neerslag.
 Sorteert de kolom *Neerslagduur(uren)* met 'Sorteren van hoog naar laag'.
 Op 18 december regende het in dat jaar het langst.

- 2 a *
- b Sorteert de kolom *Leeftijd* met 'Sorteren van hoog naar laag' en daarna met 'Sorteren van laag naar hoog'.
 De jongste atleet was 13 en de oudste was 71.
- c Filter de kolom *Leeftijd* op 'Is kleiner dan of gelijk aan' 16.
 Kiezen van 'Aantal' in cel C10386 van de totalenrij geeft 100.
 Dus $\frac{100}{10384} \cdot 100\% \approx 1,0\%$ van de atleten was 16 jaar of jonger.
- d Filter uit vraag c verwijderen.
 Filter de kolom *Land* op Netherlands.
 Kiezen van 'Gemiddelde' in cel C10386 van de totalenrij geeft 27,19...
 De gemiddelde leeftijd van de Nederlandse atleten was 27 jaar.
- e Filter de kolom *Land* op Netherlands.
 Filter de kolom *Sport* op Taekwondo.
 De enige Nederlandse Taekwondo-deelnemer komt uit in de klasse tot 80 kg.
 Hij weegt dus minder dan 80 kg.
- f Filters uit vraag e verwijderen.
 Filter de kolom *Sport* op Triathlon.
 Kiezen van 'Gemiddelde' in cel E10386 van de totalenrij geeft 61,22... kg.
 Filter de kolom *Sport* op Rowing.
 Aflezen in cel E10386 van de totalenrij geeft 80,91... kg.
 Het scheelt $80,91... - 61,22... \approx 20$ kg.
- g Filter uit vraag f verwijderen.
 Filter de kolom *Land* op Netherlands en filter de kolom *Geslacht* op V.
 Kiezen van 'Gemiddelde' in cel D10386 van de totalenrij geeft 173,92... cm.
 Filter de kolom *Land* op alles behalve Netherlands.
 Aflezen in cel D10386 van de totalenrij geeft 170,17... cm.
 Het verschil is dus $173,92... - 170,17... \approx 4$ cm.
 De journalist heeft dus gelijk.
 Filter de kolom *Geslacht* op M en voer dezelfde handelingen uit als hierboven bij de vrouwen.
 Het verschil bij de mannen is $185,92... - 182,31... \approx 4$ cm.
 Bij de mannen geldt dus hetzelfde.

Bladzijde 130

- 3 a *
- b Kiezen van 'Gemiddelde' in cel C273 van de totalenrij.
 Filter de kolom *Groep* op D geeft 177,47... cm.
 Filter de kolom *Groep* op A geeft 173,5 cm.
 Het verschil is $177,47... - 173,5 \approx 4$ cm.
- c Filter de kolom *Groep* steeds op een andere groep en kies 'Min' en 'Max' in cel C273 van de totalenrij.
 Bereken voor elke groep de spreidingsbreedte en maak een overzicht. Je krijgt

Groep	Minimum (cm)	Maximum (cm)	Spreidingsbreedte (cm)
A	152	188	$188 - 152 = 36$
B	150	199	$199 - 150 = 49$
C	137	198	$198 - 137 = 61$
D	152	198	$198 - 152 = 46$
E	152	201	$201 - 152 = 49$

- Dus de spreidingsbreedte is bij groep C het grootst en bij groep A het kleinst.
- d Kiezen van 'Aantal' in cel G273 van de totalenrij.
 Kolom *UrenSlaap* filteren op 'Kleiner dan' 8 geeft 194 kortslapers.
 Het percentage kortslapers is $\frac{194}{271} \cdot 100\% \approx 71,6\%$.

- e Kies 'Aantal' in cel G273 van de totalenrij.
 Kolom *Groep* filteren op A geeft 88 studenten in groep A.
 Kolom *UrenSlaap* filteren op 'Kleiner dan' 8 geeft 62 kortslapers in groep A.
 Het percentage kortslapers in groep A is $\frac{62}{88} \cdot 100\% \approx 70,5\%$.
 Herhalen van deze handelingen voor de groepen B, C, D en E geeft de overige percentages.
 Het percentage kortslapers in groep B is $\frac{58}{82} \cdot 100\% \approx 70,7\%$.
 Het percentage kortslapers in groep C is $\frac{32}{42} \cdot 100\% \approx 76,2\%$.
 Het percentage kortslapers in groep D is $\frac{36}{48} \cdot 100\% = 75\%$.
 Het percentage kortslapers in groep E is $\frac{6}{11} \cdot 100\% \approx 54,5\%$.
 In groep C is het percentage kortslapers het grootst.
- f Kiezen van 'Aantal' in cel I273 van de totalenrij en de kolom *LR-handig-student* filteren op L geeft 31 linkshandigen.
 Filteren van de kolom *UrenSlaap* op 'Kleiner dan' 8 geeft 20 linkshandige kortslapers.
 Het percentage kortslapers onder linkshandigen is $\frac{20}{31} \cdot 100\% \approx 64,5\%$.
 Filteren van de kolom *LR-handig-student* op R geeft 240 rechtshandigen.
 Filteren van de kolom *UrenSlaap* op 'Kleiner dan' 8 geeft 174 linkshandige kortslapers.
 Het percentage kortslapers onder linkshandigen is $\frac{174}{240} \cdot 100\% = 72,5\%$.
 Er zijn dus relatief minder kortslapers onder de linkshandige studenten.
 Niels heeft gelijk.

- 4 Deze aanpak werkt.

Bladzijde 131

- 5 Maak een nieuwe kolom door in cel N1 *HockeyDameBenelux* te typen.
 Filter de kolom *Land* op Belgium, Luxembourg en Netherlands. Filter de kolom *Geslacht* op V en filter de kolom *Sport* op Hockey.
 Zet in de eerste cel van de nieuwe kolom Ja en gebruik de vulgreep.
 Verwijder alle filters en filter dan de kolom *HockeyDameBenelux* op Lege cellen.
 Zet in de eerste cel van de kolom Nee en gebruik de vulgreep.
 Verwijder alle filters.

Bladzijde 132

- 6 Werk door met het bestand uit de vorige opgave.
- a Maak een nieuwe kolom door in cel O1 *AantalMedailles* te typen.
 $O2=SOM([@Goud];[@Zilver];[@Brons])$
- b Kiezen van 'Aantal' in cel O10386 van de totalenrij en de kolom *AantalMedailles* op 2, 3, 4 en 5 filteren geeft 36 atleten met meer dan één medaille.
 Filteren van de kolom *Goud* op 1 en 2 geeft 28 atleten met meer dan één medaille, waarvan ten minste één gouden.
- c Verwijderen van alle filters en de kolom *Medailles* sorteren met 'Sorteren van hoog naar laag' geeft Ryan Lichte met 5 medailles.
- 7 Werk door met het bestand uit de vorige opgave.
- a Filter in de kolommen *Lengte* en *Gewicht* de Lege cellen eruit.
 Kiezen van 'Aantal' in cel D10386 van de totalenrij geeft 9038 atleten.
- b Voeg een kolom toe door in cel P1 *BMI* te typen.
 $P2=[@[Gewicht(kg)]/([@[Lengte(cm)]]/100)^2$
 De kolom *BMI* sorteren met 'Sorteren van hoog naar laag' geeft Ricardo Blas Jr.
- c Voeg een kolom toe door in cel Q1 *OvergewichtBMI* te typen.
 Filter de kolom *BMI* op 'Is groter dan of gelijk aan' 25.
 Typ in de eerste cel van de nieuwe kolom Ja en gebruik de vulgreep.
 Filter de kolom *OvergewichtBMI* op Lege cellen en zet in de eerste cel van deze kolom Nee en gebruik de vulgreep.

- d Kies van 'Aantal' in cel Q10386 van de totalenrij en de kolom *OvergewichtBMI* filteren op Ja geeft 1668 atleten met overgewicht.
- e Voeg een kolom toe door in cel R1 *NBMI* te typen.
 $R2=1,3*[@[Gewicht(kg)]/([@[Lengte(cm)]/100)^2,5]$
 De kolom *NBMI* sorteren met 'Sorteren van laag naar hoog' geeft Joanna Lochowska.
- f Voeg een kolom toe door in cel S1 *OvergewichtNBMI* te typen.
 Filter de kolom *NBMI* op 'Is groter dan of gelijk aan' 25.
 Typ in de eerste cel van de nieuwe kolom Ja en gebruik de vulgreep.
 Filter de kolom *OvergewichtNBMI* op Lege cellen en zet in de eerste cel van deze kolom Nee en gebruik de vulgreep.
 Kies van 'Aantal' in cel S10386 van de totalenrij en filteren van *OvergewichtBMI* op Ja en *OvergewichtNBMI* op Nee geeft 57 atleten die volgens het BMI wel, maar volgens het NBMI geen overgewicht hebben.
 Kies van 'Aantal' in cel S10386 van de totalenrij en filteren van *OvergewichtBMI* op Nee en *OvergewichtNBMI* op Ja geeft 521 atleten die volgens het BMI niet, maar volgens het NBMI wel overgewicht hebben.
 Dus het geldt voor $57 + 521 = 578$ atleten.
- g Voeg een kolom toe door in cel T1 *BMI-NBMI* te typen.
 Sorteer deze kolom met 'Sorteren van hoog naar laag' of 'Sorteren van laag naar hoog', of gebruik 'Min' en 'Max' in de totalenrij. Je krijgt als kleinste getal -3,52.. en als grootste getal 3,64...
 Het grootste verschil is dus bij de atleet die hoort bij 3,64...
 Dit is Carl Myerscough. Deze atleet is 2,08 m lang en weegt 160 kg.

Bladzijde 133

- 8 Zet de gegevens om in een Exceltabel en maak een totalenrij aan.
 Voeg een nieuwe kolom toe door in cel L1 *Gevoelstemperatuur* te typen.
 Filter de kolom *GemiddeldeWindsnelheid(m/s)* op 'Is kleiner dan of gelijk aan' 1,3.
 $B709 = C709$ (en gebruik indien nodig de vulgreep).
 Filter de kolom *GemiddeldeWindsnelheid(m/s)* op 'Groter dan' 1,3.
 $B2=13,12+0,6215*C2-11,37*(3,6*B2)^{0,16}+0,3965*C2*(3,6*B2)^{0,16}$ (en gebruik indien nodig de vulgreep).
 Sorteren van de kolom *Gevoelstemperatuur* op 'Sorteren van laag naar hoog' geeft 1 februari 1956.

8.2 Draaitabellen en draaigrafieken

Bladzijde 134

- 9 a Zet de gegevens om in een Exceltabel en maak een totalenrij aan.
 Kies 'Aantal' in cel E801 van de totalenrij.
 Filteren van kolom *Geslacht* op M geeft 402 jongens en filteren op V geeft 396 meisjes.
- b Kies 'Aantal' in cel C801 van de totalenrij.
 Filteren van kolom *Regio* op Midwest geeft 170, filteren op Northwest geeft 111, filteren op South geeft 294 en filteren op West geeft 223 jongeren.
- c Filter per Regio op M en op V.

		geslacht		
		M	V	
regio	Midwest	95	75	170
	Northeast	52	59	111
	South	148	146	294
	West	107	116	223
		402	396	798

10 a •

Aantal van ID Kolomlabels					
Rijlabels	Midwest	Northeast	South	West	Eindtotaal
12	20	20	35	37	112
13	24	13	52	43	132
14	23	17	55	30	125
15	26	21	47	41	135
16	42	21	53	38	154
17	35	19	52	34	140
Eindtotaal	170	111	294	223	798

Aantal van ID Kolomlabels					
Rijlabels	Rural	Suburban	Urban	(leeg)	Eindtotaal
Midwest	32	82	51	5	170
Northeast	8	69	33	1	111
South	38	152	101	3	294
West	23	106	90	4	223
Eindtotaal	101	409	275	13	798

b • Maak een draaitabel met als rij-variabele *HeeftMobieleTelefoon* en als kolom-variabele *HeeftTablet* en zorg bij Waarden voor Aantal van *ID*.

Aantal van ID Kolomlabels			
Rijlabels	ja	nee	Eindtotaal
ja	170	466	636
nee	31	131	162
Eindtotaal	201	597	798

NB 170 van de onderzochte personen hebben zowel een mobiele telefoon als een tablet.

• Maak een draaitabel met als rij-variabele *Leeftijd* en als kolom-variabele *Geslacht* en zorg bij Waarden voor Aantal van *ID*.

Aantal van ID Kolomlabels			
Rijlabels	M	V	Eindtotaal
12	60	52	112
13	60	72	132
14	63	62	125
15	76	59	135
16	78	76	154
17	65	75	140
Eindtotaal	402	396	798

NB Bij dit onderzoek waren 76 jongens van 15 jaar betrokken.

Bladzijde 135

11 • Maak een draaitabel met als rij-variabelen *Geslacht* en *Groep* en zorg bij Waarden voor Min van *Afstand*, Gemiddelde van *Afstand* en Max van *Afstand*. Stel de getalnotatie van Gemiddelde van *Afstand* in op twee decimalen.

Rijlabels	Min van Afstand	Gemiddelde van Afstand2	Max van Afstand3
M	71,19	78,38	88,34
A	72,29	78,42	84,47
B	71,19	78,33	88,34
V	47,65	58,53	66,19
A	47,65	58,68	66,19
B	50,01	58,40	65,92
Eindtotaal	47,65	68,95	88,34

12 • Er zijn 12 studenten waarbij zowel de variabele *DeeltFotosVanZichzelf* als de variabele *DeeltVideosVanZichzelf* de waarde #Null! hebben. Van deze studenten is niet bekend of ze al dan niet foto's en/of video's van zichzelf delen.

13 a • Maak een draaitabel met *Leeftijd* als rij-variabele, *Geslacht* als kolom-variabele en zorg bij Waarden voor Gemiddelde van *FB_vrienden*. Stel de getalnotatie voor de hele tabel in op gehelen.

Gemiddelde van FB_Vrienden Kolomlabels			
Rijlabels	M	V	Eindtotaal
12	182	185	183
13	271	393	337
14	370	470	425
15	367	492	423
16	409	445	426
17	491	594	546
Eindtotaal	374	467	420

Naarmate de jongeren ouder zijn hebben ze, met uitzondering van de 15-jarigen, meer Facebookvrienden. Voor alle leeftijden geldt dat de meisjes meer Facebookvrienden hebben dan de jongens.

Susanne heeft dus gelijk.

- b** Maak een Exceltabel en filter de kolom *FB_Vrienden* op #Null!.
 Typ in de eerste cel van de kolom 0 en gebruik de vulgreep.
 Maak een draaitabel met als rij-variabele *Woonomgeving*, als kolom-variabele *Geslacht* en zorg bij Waarden voor Gemiddelde van *FB_Vrienden*.
 Stel de getalnotatie voor de hele tabel in op gehelen.

Gemiddelde van FB_Vrienden		Kolomlabels		
Rijlabels	M	V	Eindtotaal	
Rural	324	307	315	
Suburban	281	348	313	
Urban	242	356	301	
(leeg)	114	442	265	
Eindtotaal	271	347	308	

Michiel concludeert dat het gemiddeld gezien voor de jongens en meisjes samen niet veel verschil maakt. Hij merkt wel op dat jongens op het platteland gemiddeld meer Facebookvrienden hebben dan jongens in de stad, terwijl dat bij meisjes juist omgekeerd is.

- 14 a** Maak een draaitabel met als rij-variabele *Lengte* en kies bij Waarden voor Aantal van *Lengte*.
 Er komen veel lengtes voor en de meeste lengtes hebben een lage frequentie. Dat levert een weinig overzichtelijke frequentietabel op.
- b** Maak een nieuwe kolom aan door *LengteKlasse* te typen in cel H1.
 Filter *Lengte* op 'Is groter dan of gelijk aan' 150 en 'Is kleiner dan of gelijk aan' 159.
 Typ in de eerste cel van de nieuwe kolom 150-159 en gebruik de vulgreep.
 Herhaal deze handeling voor de overige lengteklassen.
- c** Maak een draaitabel met als rij-variabele *LengteKlasse* en kies bij Waarden voor Aantal van *LeerlingID*.

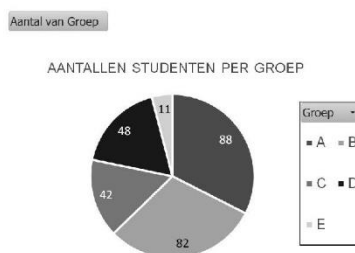
Rijlabels	Aantal van LeerlingID
> 189	8
150 - 159	9
160 - 169	44
170 - 179	54
180 - 189	31
Eindtotaal	146

- d** Een histogram.

Bladzijde 136

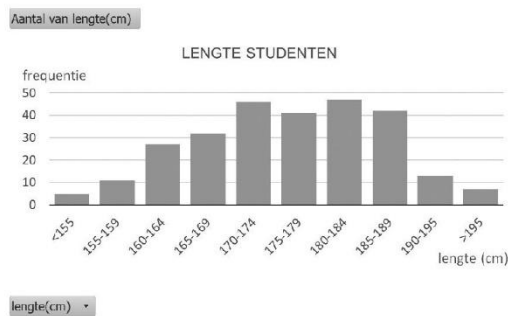
- 15 a** Maak een draaitabel met als rij-variabele *Groep* en kies bij Waarden voor Aantal van *Groep*.

Rijlabels	Aantal van Groep
A	88
B	82
C	42
D	48
E	11
Eindtotaal	271



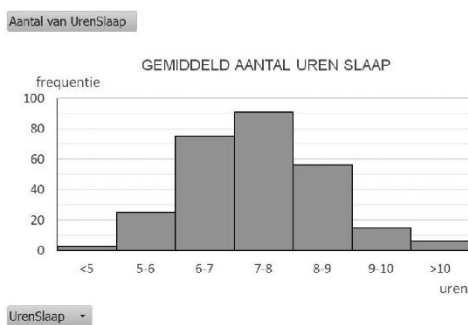
- b Maak een draaitabel met als rij-variabele *Lengte* en kies bij Waarden voor Aantal van *Lengte*. Kies bij *Lengte* voor groeperen: Beginnen bij 150, Eindigen bij 195 en Op 5.

Rijlabels	Aantal van lengte(cm)
<155	5
155-159	11
160-164	27
165-169	32
170-174	46
175-179	41
180-184	47
185-189	42
190-195	13
>195	7
Eindtotaal	271



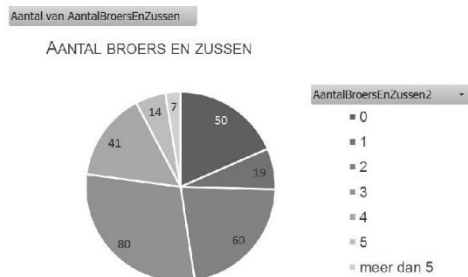
- c Maak een draaitabel met als rij-variabele *UrenSlaap* en kies bij Waarden voor Aantal van *UrenSlaap*. Kies bij *UrenSlaap* voor groeperen: Beginnen bij 5, Eindigen bij 10 en Op 1.

Rijlabels	Aantal van UrenSlaap
<5	3
5-6	25
6-7	75
7-8	91
8-9	56
9-10	15
>10	6
Eindtotaal	271



- d Maak een draaitabel met als rij-variabele *AantalBroersEnZussen* en kies bij Waarden voor Aantal van *AantalBroersEnZussen*. Kies bij *AantalBroersEnZussen* voor groeperen van 6 en 7 broers en zussen.

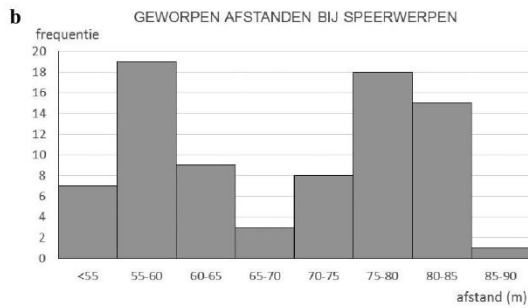
Rijlabels	Aantal van AantalBroersEnZussen
0	50
1	19
2	60
3	80
4	41
5	14
meer dan 5	7
Eindtotaal	271



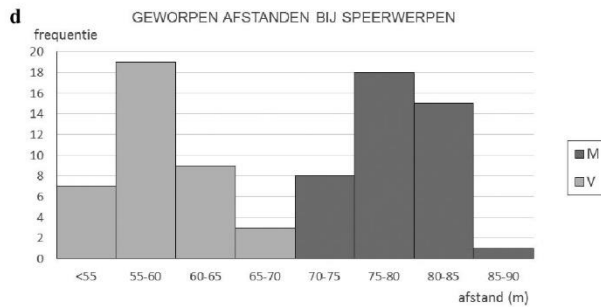
Bladzijde 137

- 16 a Maak een draaitabel met als rij-variabele *Afstand* en kies bij Waarden voor Aantal van *Afstand*. Kies bij *Afstand* voor groeperen: Beginnen bij 55, Eindigen bij 90 en Op 5.

Rijlabels	Aantal van Afstand
<55	7
55-60	19
60-65	9
65-70	3
70-75	8
75-80	18
80-85	15
85-90	1
Eindtotaal	80



c Je hebt te maken met een meertoppige verdeling.



Mannen gooien verder dan vrouwen. Dit veroorzaakt de twee toppen.

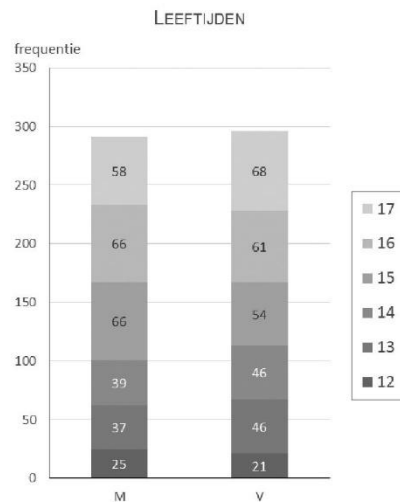
e Voeg bijvoorbeeld *Groep* toe als as-variabele boven *Afstand*. Of neem achtereenvolgens als as-variabelen *Geslacht*, *Groep* en *Afstand*.

f Maak een draaitabel met als rij-variabele *Groep* en zorg bij waarden voor Gemiddelde van *Afstand*.

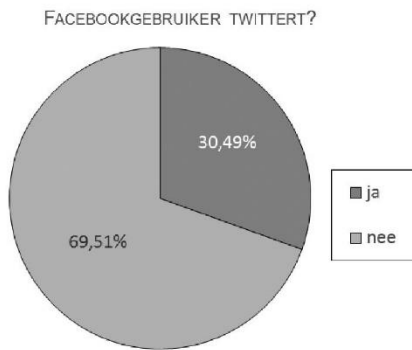
Rijlabels	Gemiddelde van Afstand
A	69,537
B	68,36575
Eindtotaal	68,951375

Groep A gooit inderdaad gemiddeld iets verder dan groep B. De prestatie van een deelnemer is bij speerwerpen echter niet afhankelijk van de tegenstanders. Het lijkt dus onzin te denken dat je verder kunt gooien als je in groep A zit. Het zou echter wel zo kunnen zijn dat bijvoorbeeld de weersomstandigheden voor groep A gunstiger zijn geweest, waardoor er in deze groep gemiddeld verder werd gegooit.

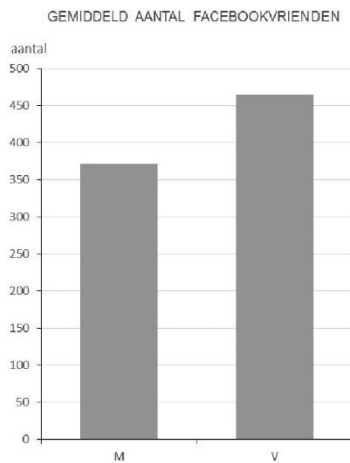
17 a Maak een draaitabel met als rij-variabele *Geslacht*, als kolom-variabele *Leeftijd* en kies bij Waarden voor Aantal van *Leeftijd*. Maak de draaigrafiek.



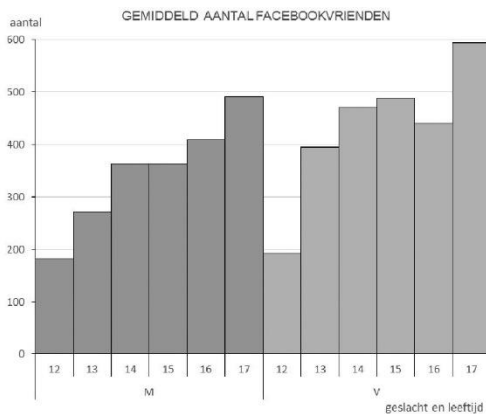
- b Maak een draaitabel met als rij-variabele *Twittert* en kies bij Waarden voor Aantal van *Twittert*. Maak de draaigrafiek.



- c Maak een draaitabel met als rij-variabele *Geslacht* en kies bij Waarden voor Gemiddelde van *FB_Vrienden*. Maak de draaigrafiek.



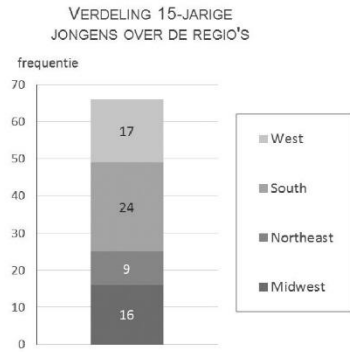
Voeg *Leeftijd* als kolom-variabele of legenda-variabele toe.



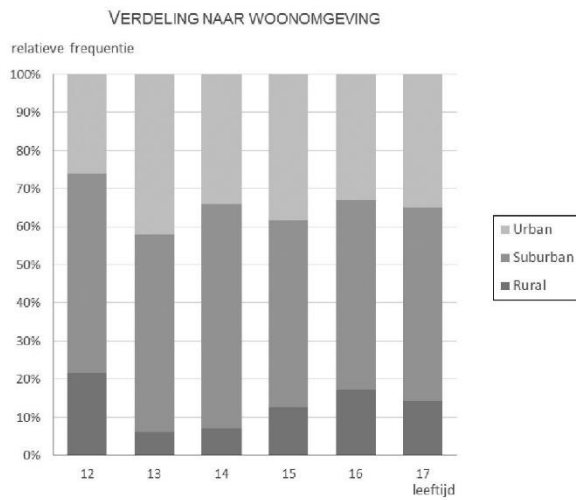
- d Simone concludeert dat meisjes gemiddeld meer Facebookvrienden hebben dan jongens en dat voor zowel jongens als meisjes geldt dat naar mate ze ouder zijn, ze gemiddeld meer Facebookvrienden hebben (op de groep 16-jarige meisjes na).

Bladzijde 138

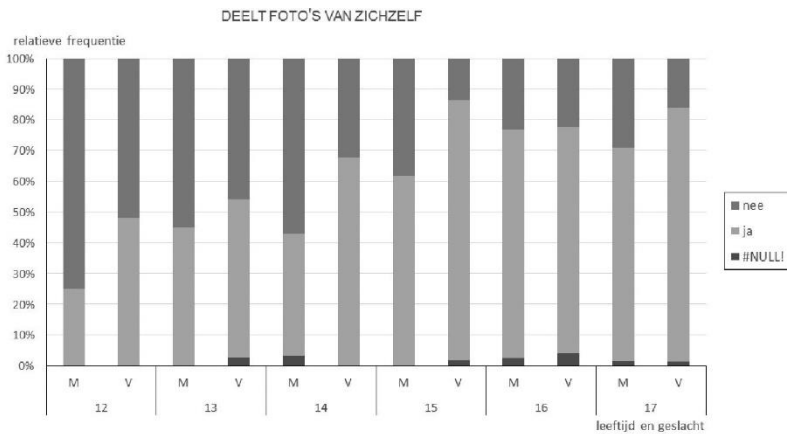
- 18 a Maak een draaitabel met als rij-variabelen *Regio* en *Woonomgeving*, als kolom-variabele *Leeftijd* en kies bij Waarden voor Aantal van *ID*. Maak de draaigrafiek.
 b Maak een draaitabel met als kolom-variabele *Regio* en kies bij Waarden voor Aantal van *ID*. Stel *Leeftijd* en *Geslacht* in als filter-variabelen en filter op 15 en M. Maak de draaigrafiek.



- c Maak een draaitabel met als rij-variabele *Leeftijd*, als kolom-variabele *Woonomgeving* en kies bij Waarden voor Aantal van *Woonomgeving*. Maak de draaigrafiek.



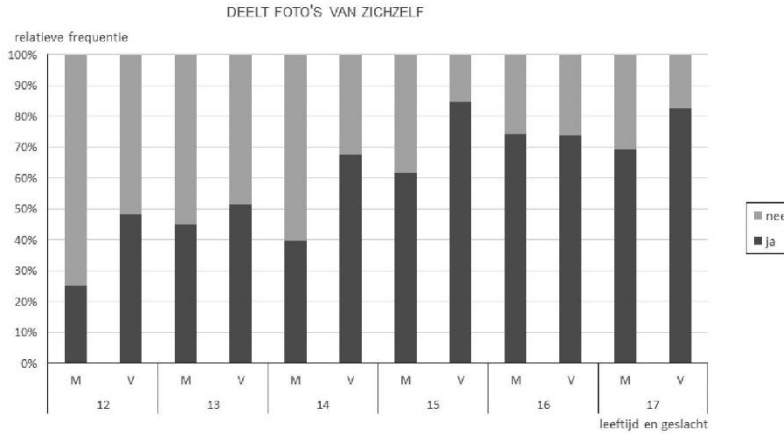
- 19 a Maak een draaitabel met als rij-variabelen *Leeftijd* en *Geslacht*, als kolom-variabele *DeeltFotosVanZichzelf* en kies bij Waarden voor Aantal van *DeeltFotosVanZichzelf*.



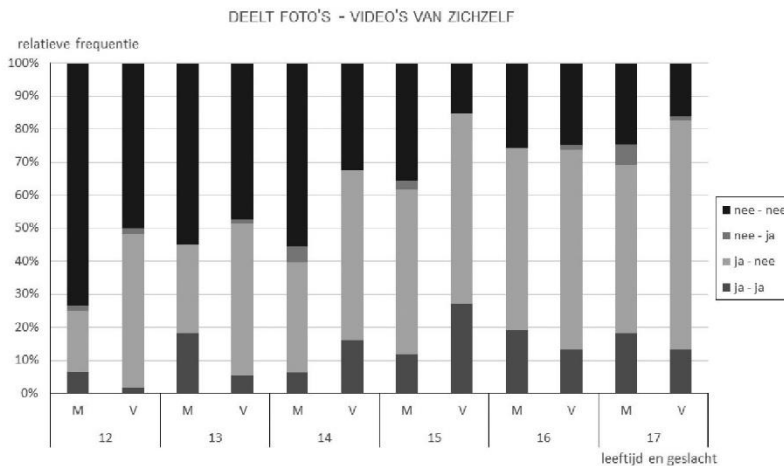
- b Niels kijkt naar het aantal in de kolom Nee. Hij berekent $\frac{17}{76} \cdot 100\% \approx 22\%$.
 Dayal rekent ook de meisjes bij #Null! mee. Zij berekent $\frac{20}{76} \cdot 100\% \approx 26\%$.

Dayal heeft gelijk. Meisjes die geen sociale netwerken gebruiken delen er dus ook geen foto's van zichzelf op.

- c Maak een Exceltabel en filter de kolom *DeeltFotosVanZichzelf* op #Null!. Typ in de eerste cel van deze kolom nee en gebruik de vulgreep. Vernieuw de draaitabel en maak de figuur opnieuw.



- d Filter in de Exceltabel de kolom *DeeltVideosVanZichzelf* op #Null!. Typ in de eerste cel van deze kolom nee en gebruik de vulgreep. Vernieuw de draaitabel en voeg de variabele *DeeltVideosVanZichzelf* toe als legenda-variabele (of als kolom-variabele) en maak de figuur opnieuw.

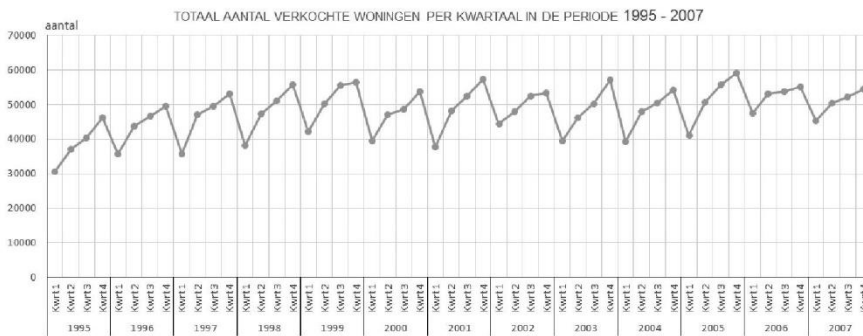


Bladzijde 139

- 20 a *
b Nee, er zijn zoveel datapunten dat de grafiek onoverzichtelijk wordt.

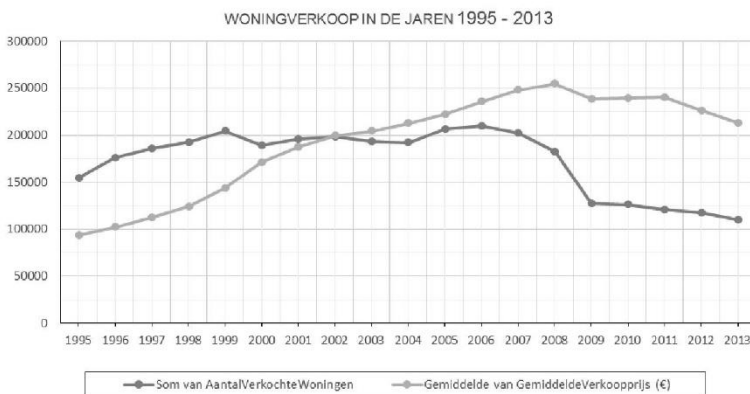
Bladzijde 140

- 21 a Maak van de gegevens een Exceltabel. Maak een draaitabel met als rij-variabele *Periode* en als kolom-variabele *AantalVerkochteWoningen* en kies bij Waarden voor Som van *AantalVerkochteWoningen*. Kies bij *Periode* voor groeperen: selecteer zowel Jaren als Kwartalen. Filter de variabele *Periode* met een Datumfilter op Tussen 1-1-1995 en 1-1-2008 of filter op 1995 tot en met 2007 door alle andere categorieën uit te vinken. Maak het lijndiagram bij de draaitabel.

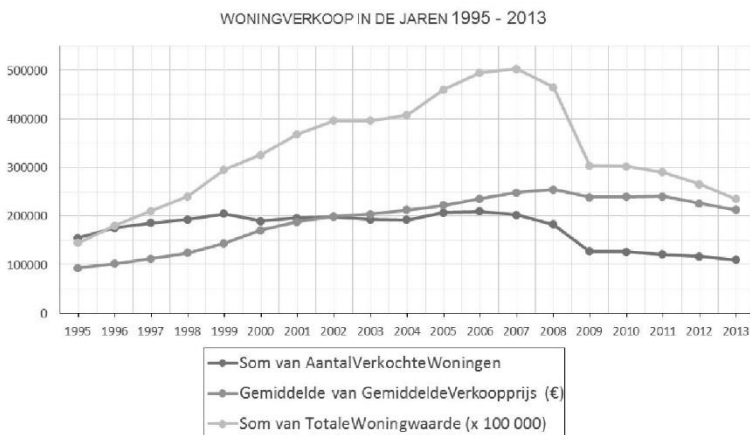


Steeds neemt de kwartaalverkoop binnen een jaar toe en steeds is het aantal verkochte woningen in het eerste kwartaal minder dan in het laatste kwartaal van het voorafgaande jaar.

- b** Voeg aan de draaitabel de kolom-variabele *GemiddeldeVerkoopprijs* toe en kies bij Waarden voor Gemiddelde van *GemiddeldeVerkoopprijs*.
 Pas de filter van de variabele *Periode* aan, zodat de jaren 1995 - 2013 getoond worden.
 Kies bij *Periode* voor groeperen: selecteer alleen Jaren.



- c** NB Het is ook mogelijk een diagram te maken met twee assen. Kies daartoe bij grafieken voor Keuzelijst.
 Maak een Exceltabel en voeg een kolom toe met de *TotaleWoningwaarde* ($\times 100\,000$).
 Typ in de eerste lege cel van deze kolom de formule
 $=[@AantalVerkochteWoningen]*[@GemiddeldeVerkoopprijs (\text{€})]/100\,000$.
 Vernieuw de draaitabel. Voeg aan de draaitabel de variabele *TotaleWoningwaarde* ($\times 100\,000$) toe als waarden-variabele en kies voor Som van *TotaleWoningwaarde* ($\times 100\,000$).



- 22 a** Maak bij de Exceltabel op het tabblad Juli 2014 de totalenrij aan. Kies Gemiddelde in de totalenrij bij de kolom *GemTemperatuur*. Het gemiddelde is 19,8 °C. Dit is het getal dat op de eerste rij puntjes moet staan.

Maak bij de gegevens in het tabblad Tijdvak 1981-2010 een draaitabel met als rij-variabele *Datum* en kies bij Waarden voor Gemiddelde van *GemTemperatuur*. Groepeer *Periode* op Maanden.

Lees bij juli af dat de gemiddelde temperatuur 17,9 °C is. Dit is het getal dat op de tweede rij puntjes moet staan.

Maak bij de gegevens van het tabblad Historie een draaitabel met als rij-variabele *Datum* en kies bij Waarden voor Gemiddelde van *GemTemperatuur*.

Kies bij Groeperen voor zowel Maanden als Jaren. Zet de rij-variabele *Datum* bovenaan bij de rij-variabelen.

Sorteer de waarden van Gemiddelde van *GemTemperatuur* bij de jaren binnen juli met 'Sorteren van hoog naar laag'. Merk op dat 2014 op plek 6 in de rij staat.

OF

Kies bij 'Waarden weergeven als' voor 'Positie van groot naar klein' bij Gemiddelde van *GemTemperatuur*. Lees af dat 2014 op de plek 6e plaats staat.

Dus op de derde rij puntjes moet 6 staan.

Voeg aan de draaitabel bij het tabblad Tijdvak 1981-2010, bij Waarden Gemiddelde van *Zonneschijnduur* toe en ook Som van *HoeveelheidNeerslag*.

Lees bij juli af dat het gemiddelde aantal uren zonneschijn 6,6 uur per dag is en dat de totale hoeveelheid neerslag in deze 30 jaren in deze maand 2437,75 mm is. Dat is gemiddeld 81,3 mm per jaar.

Ga naar het tabblad Juli 2014. Kies Gemiddelde in de totalenrij bij de kolom *Zonneschijnduur*.

De gemiddelde zonneschijnduur was 6,9 uur.

Kies Som in de totalenrij bij de kolom *HoeveelheidNeerslag*. De hoeveel neerslag was 137,5 mm.

In de laatste alinea komen op de puntjes achtereenvolgens: 6,9; 6,6; 137,5; 81,3.

- b** Maak een draaitabel bij het tabblad Tijdvak 1981-2010. Gebruik als rij-variabele *Datum* en als Waarden Gemiddelde van *GemTemperatuur*. Groepeer *Datum* op Maanden én Dagen.

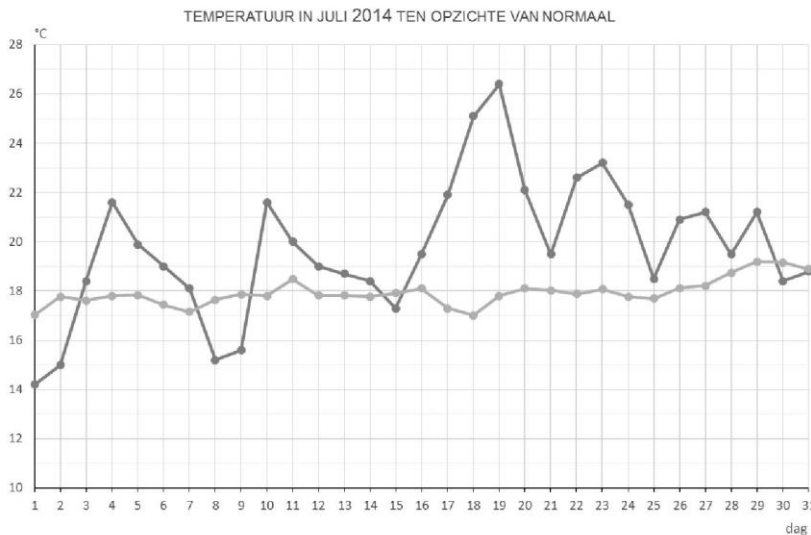
Zet op een nieuw tabblad in cel A1 'juli in 2014' en in cel B1 'normaal in juli'.

Zet in de kolommen A2 t/m A32 de gemiddelde dagtemperaturen van juli 2014. Kopieer deze van het tabblad Juli 2014.

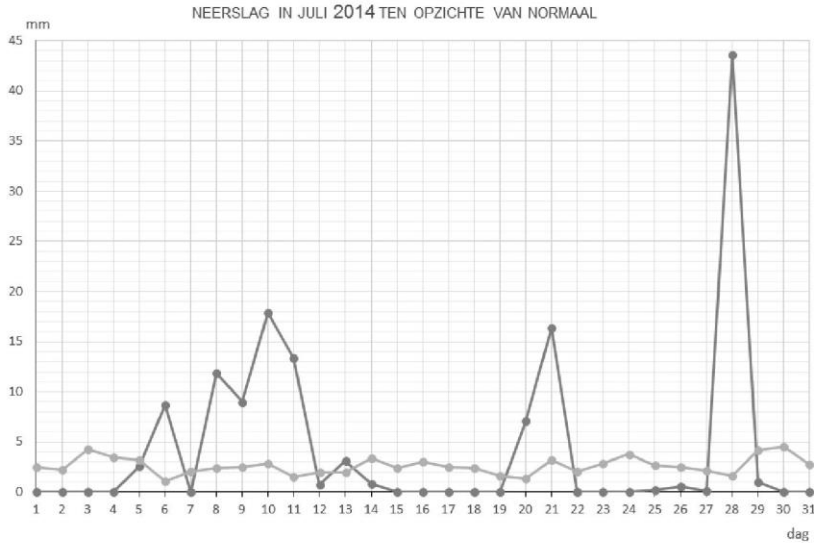
Zet in de kolommen B2 t/m B32 de gemiddelde dagtemperaturen van juli in de periode 1981-2010.

Kopieer deze uit de bij deze vraag gemaakte draaitabel.

Maak het lijndiagram.



c Ga te werk zoals bij vraag b.



Hoewel de eerste paar dagen van juli 2014 droog waren kwamen al snel de eerste buien. De tweede week van juli was erg nat en na weer een aantal droge dagen viel er op 20 en 21 juli een flinke hoeveelheid neerslag. De rest van de maand waren de meeste dagen droog behalve op 28 juli toen zeer zware buien voor grote hoeveelheden neerslag zorgden.

8.3 Data analyseren

Bladzijde 141

23 $\hat{p} = 0,10$ en $\sigma = \sqrt{\frac{0,1 \cdot 0,9}{1000}} = 0,0094\dots$

$\hat{p} - 2\sigma = 0,10 - 2 \cdot 0,0094\dots \approx 0,091$

$\hat{p} + 2\sigma = 0,10 + 2 \cdot 0,0094\dots \approx 0,109$

Het 95%-betrouwbaarheidsinterval is $[0,091; 0,109]$.

24 a $H5 = \text{WORTEL}(G5 \cdot (1 - G5) / 500)$

$I5 = G5 - 2 \cdot H5$

$J5 = G5 + 2 \cdot H5$

Gebruik vervolgens de vulgreep.

b Je ziet dat er veel overlap is in de intervallen. De breedte van de intervallen geeft een beeld van de nauwkeurigheid van het resultaat van de steekproef.

Die nauwkeurigheid is hier te laag om iets te kunnen zeggen.

Bladzijde 142

25 a Maak een draaitabel met als rij-variabele *Woning*, kies bij Waarden voor Aantal van *ID*. Kies in de draaitabel bij 'Waarden weergeven als' voor '% van kolomtotaal'.

De gevraagde steekproefproportie is dus 0,373.

b Gebruik de draaitabel van vraag a en vervang de rij-variabele *Woning* door *TypeWoning*.

De steekproefproportie van hoekwoningen is 0,1321 en de dataset bevat 117058 records, dus

$\sigma = \sqrt{\frac{0,1321 \cdot 0,8679}{117058}} = 0,00098\dots$

$\hat{p} - 2\sigma = 0,1321 - 2 \cdot 0,00098\dots \approx 0,130$

$\hat{p} + 2\sigma = 0,1321 + 2 \cdot 0,00098\dots \approx 0,134$

Het 95%-betrouwbaarheidsinterval is $[0,130; 0,134]$.

c Gebruik de draaitabel van vraag b en voeg *Woning* als kolom-variabele toe.

Je ziet dat onder de koopwoningen de steekproefproportie van vrijstaande woningen 0,2138 is.

Kies bij 'Waarden weergeven als' voor 'Geen berekening'. Je ziet dat er 73 398 koopwoningen in de dataset zitten.

$$\sigma = \sqrt{\frac{0,2138 \cdot 0,7862}{73\,398}} \approx 0,0015\dots$$

$$\hat{p} - 2\sigma = 0,2138 - 2 \cdot 0,0015\dots \approx 0,211$$

$$\hat{p} + 2\sigma = 0,2138 + 2 \cdot 0,0015\dots \approx 0,217$$

Het 95%-betrouwbaarheidsinterval is [0,211; 0,217].

Bladzijde 143

- 26 a Maak een draaitabel zonder rij- en kolomvariabelen. Kies bij Waarden voor Aantal van ID, Gemiddelde van WozWaarde en ook voor Stdevp van WozWaarde. Je krijgt

Aantal van ID	Gemiddelde van WozWaarde	Stdevp van WozWaarde
117058	251136,1738	153294,1554

Bereken in een lege cel, bijvoorbeeld cel E4, de betrouwbaarheid met de functie BETROUWBAARHEID en verwijzingen naar de juiste cellen.

E4=BETROUWBAARHEID(0,05;DRAAITABEL.OPHALEN("Stdevp van WozWaarde2";\$A\$3);DRAAITABEL.OPHALEN("Aantal van ID";\$A\$3))

Bereken vervolgens de ondergrens van het betrouwbaarheidsinterval in cel G3 met G3=DRAAITABEL.OPHALEN("Gemiddelde van WozWaarde";\$A\$3) - E4

en de bovengrens van het betrouwbaarheidsinterval Cel H3 met

H3=DRAAITABEL.OPHALEN("Gemiddelde van WozWaarde";\$A\$3) + E4

Afgerond op duizendtallen krijg je het 95%-betrouwbaarheidsinterval [250 000, 252 000].

- b Voeg aan de draaitabel van vraag a de rij-variabele Woning en daaronder de rij-variabele Woningtype toe.
Ga op dezelfde wijze te werk als bij a om het 95%-betrouwbaarheidsinterval te berekenen. Je krijgt [425 000, 432 000].
- c Verwijder uit de draaitabel van vraag b de variabelen Woning en Woningtype en voeg de rij-variabelen OverlastJongeren en OverlastVerkeer toe.
Je ziet: steekproefgrootte = 2300, $\bar{X} \approx 200\,357$ en $\sigma \approx 111\,999$
Ga op dezelfde wijze te werk als bij a om het 95%-betrouwbaarheidsinterval te berekenen. Gebruik daarbij in de functie BETROUWBAARHEID alfa = 0,20.
Afgerond op duizendtallen krijg je het 80%-betrouwbaarheidsinterval [197 000, 203 000].

- 27 a Maak een draaitabel met als rij-variabele TevredenheidWoningen kies bij Waarden voor Aantal van ID. De steekproeflengte is 117 058. Kies bij 'Waarden weergeven als' voor '% van kolomtotaal'.

$$\text{Er geldt } \hat{p} = 0,4440 \text{ en } \sigma = \sqrt{\frac{0,4440 \cdot 0,5560}{117\,058}} \approx 0,0014\dots$$

$$\hat{p} - \sigma = 0,4440 - 0,0014\dots \approx 0,443$$

$$\hat{p} + 2\sigma = 0,4440 + 0,0014\dots \approx 0,445$$

Het 68%-betrouwbaarheidsinterval is [0,443; 0,445].

- b Maak een draaitabel zoals hieronder.

Rijlabels	Gemiddelde van WozWaarde	Stdev van WozWaarde	Aantal van ID
1. Zeer ontevreden	147 960,08	65 420,23	952
2. Ontevreden	165 003,94	75 959,44	2716
3. Niet tevreden...	184 999,47	93 253,57	8520
4. Tevreden	224 515,49	117 628,45	51 976
5. Zeer tevreden	294 227,64	181 123,23	52 894
Eindtotaal	251 136,17	153 294,81	117 058

Je ziet: steekproefgrootte = 2300, $\bar{X} \approx 294\,227,64$ en $\sigma \approx 52\,894$.

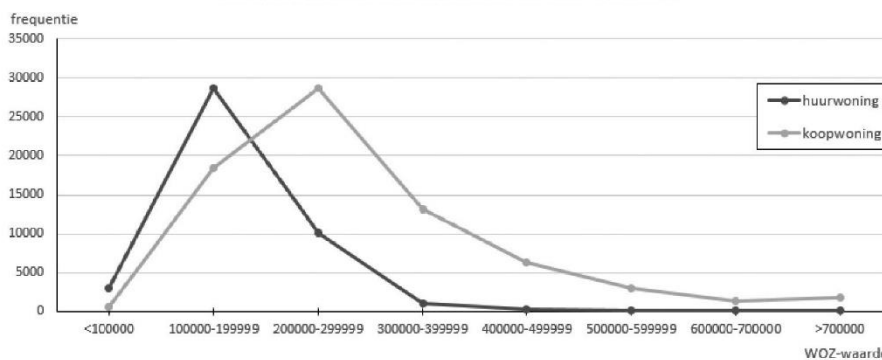
Gebruik BETROUWBAARHEID(0,1;181123,23;52 894) om de betrouwbaarheid te berekenen. En bereken de grenzen van het betrouwbaarheidsinterval. Afgerond op duizenden euro's vind je het 90%-betrouwbaarheidsinterval [293 000; 296 000].

- 28 a Maak een draaitabel met als rij-variabele *WozWaarde* en als kolom-variabele *Woning*. Kies bij Waarden voor Aantal van ID. Kies bij *WozWaarde* voor groeperen: Beginnen bij 100 000, Eindigen bij 700 000 en Op 100 000.

Aantal van ID	Kolomlabels		
Rijlabels	huurwoning	koopwoning	Eindtotaal
<100000	3020	624	3644
100000-199999	28 741	18 369	47 110
200000-299999	10 138	28 664	38 802
300000-399999	1088	13 191	14 279
400000-499999	347	6363	6710
500000-599999	152	3010	3162
600000-700000	74	1355	1429
>700000	100	1822	1922
Eindtotaal	43 660	73 398	117 058

Maak het lijndiagram.

VERDELING WOZ-WAARDEN BIJ HUUR- EN KOOPWONINGEN



Overeenkomst: Bij zowel huur- als koopwoningen heb je te maken met een rechts-scheve verdeling.
Verschil: koopwoningen hebben gemiddeld een hogere WOZ-waarde dan huurwoningen.

- b Laat Excel voor beide type woningen de gemiddelde WOZ-waarde berekenen. Gebruik bijvoorbeeld een draaitabel zoals hieronder.

Rijlabels	Gemiddelde van WozWaarde
huurwoning	176 552,52
koopwoning	295 501,45
Eindtotaal	251 136,17

Het verschil in gemiddelde WOZ-waarde is $295\,501,45 - 176\,552,52 = 118\,948,93$ euro.

- *
c Verreweg de meeste eengezinswoningen zijn koopwoningen terwijl verreweg de meeste meergezinswoningen huurwoningen zijn.
*
d Huishoudens die in een koopwoning wonen zijn over het algemeen meer tevreden over hun woning dan huishoudens die in een huurwoning wonen.
*
e *TypeWoning* is een nominale variabele.
WozWaarde is een ratio variabele.
TevredenheidWoning is een ordinale variabele.

Bladzijde 145

29 Bij de eengezinswoningen is $\frac{62382}{20002 + 62382} \times 100\% \approx 75,7\%$ koopwoning.

Bij de meergezinswoningen is dit $\frac{11016}{23658 + 11016} \times 100\% \approx 31,8\%$.

$PV = 75,7 - 31,8 = 43,9\%$.

De kruisproducten zijn $20002 \cdot 11016 = 220342032$ en $23658 \cdot 62382 = 1475833356$.

$OR = \frac{1475833356}{220342032} \approx 6,7$

Beide maten duiden op een groot verschil.

30

Aantal van ID	Kolomlabels			
Rijlabels	huurwoning	koopwoning	V_{cp}	
1. Zeer ontevreden	1,88%	0,18%	1,71%	
2. Ontevreden	6,67%	1,03%	5,64%	
3. Niet tevreden...	19,58%	4,96%	14,62%	
4. Tevreden	70,46%	45,51%	24,96%	←max. V_{cp}
5. Zeer tevreden	100,00%	100,00%	0,00%	
Eindtotaal				

Het verschil is middelmatig.

Bladzijde 146

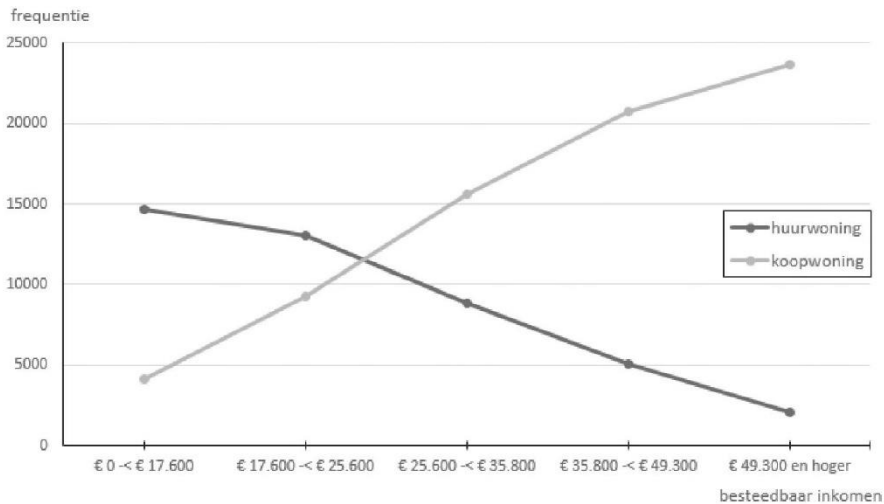
31 Maak een draaitabel met als rij-variabele *BesteedbaarInkomen* en als kolom-variabele *Woning* en kies bij Waarden Aantal van ID.

Laat Excel de aantallen weergeven als cumulatieve percentages en bereken het V_{cp} .

Aantal van ID	Kolomlabels			
Rijlabels	huurwoning	koopwoning	Eindtotaal	V_{cp}
€ 0 -< € 17.600	33,57%	5,67%	16,08%	27,9%
€ 17.600 -< € 25.600	63,37%	18,32%	35,12%	45,0% ←max. V_{cp}
€ 25.600 -< € 35.800	83,63%	39,62%	56,04%	44,0%
€ 35.800 -< € 49.300	95,17%	67,83%	78,03%	27,3%
€ 49.300 en hoger	100,00%	100,00%	100,00%	0,0%
Eindtotaal				

Huishoudens die wonen in een koopwoning hebben over het algemeen meer te besteden dan huishoudens die wonen in een huurwoning.

VERDELING BESTEEDBAAR INKOMEN VAN HUISHOUDENS UIT HUUR-EN KOOPWONINGEN



Het verschil is groot.

- 32 Maak een draaitabel met als rij-variabele *TevredenheidWoonomgeving* en als kolom-variabele *TypeWoning* en kies bij Waarden voor Aantal van ID. Groepeer de waarden '2-onder-1-kap', 'Etagewoning', 'Hoekwoning', 'Tussenwoning of overig', noem deze groep niet-vrijstaand. Vouw deze groep samen en kies bij 'Waarden weergeven als' voor '% van totaalkolom'. Bereken de V_{cp} 's.

Aantal van ID	Kolomlabels			Eindtotaal	
	niet-Vrijstaand	Vrijstaand			
Rijlabels		Vrijstaand			V_{cp}
1. Zeer ontevreden	1,35%	0,45%	1,22%	0,9%	
2. Ontevreden	6,00%	2,48%	5,50%	3,5%	
3. Niet tevreden ...	17,11%	8,30%	15,87%	8,8%	
4. Tevreden	70,17%	52,65%	67,69%	17,5%	←max. V_{cp}
5. Zeer tevreden	100,00%	100,00%	100,00%	0,0%	
Eindtotaal					

Huishoudens die wonen in een vrijstaand huis zijn over het algemeen iets meer tevreden over de woonomgeving dan huishoudens die niet in een vrijstaand huis wonen. Dit verschil in tevredenheid is middelmatig te noemen.

- 33 Gebruik een draaitabel met als rij-variabele *OverlastVerkeer* en als kolom-variabele *OverlastJongeren*. Groepeer bij beide variabelen de waarden '1. (Bijna) nooit' en 'Soms' in een groep met de naam niet-Vaak en vouw alle groepen samen. Kies bij Waarden voor Aantal van ID.

Aantal van ID	Kolomlabels		
	niet-Vaak	3. Vaak	Eindtotaal
Rijlabels			
niet-Vaak	98418	4792	103210
3. Vaak	11548	2300	13848
Eindtotaal	109966	7092	117058

De kruisproducten zijn $98\,418 \cdot 2\,300 = 226\,361\,400$ en $4\,792 \cdot 11\,548 = 55\,338\,016$.

$$OR = \frac{226\,361\,400}{55\,338\,016} \approx 4,1$$

Huishoudens die vaak overlast hebben van verkeer hebben ongeveer vier keer zo vaak overlast van jongeren als huishoudens die niet vaak overlast hebben van verkeer.

Je kunt het met de politicus eens zijn.

- 34 Maak een draaitabel met als kolom-variabele *Plaats*, waarbij de waarden 'G4' en 'G27' gegroepeerd zijn in de groep G31. Kies als rij-variabele *AanwezigheidBekladding* en kies bij Waarden voor Aantal van ID. Kies voor 'Waarden weergeven als % voorlopig totaal in' *AanwezigheidBekladding*. Bereken vervolgens de V_{cp} 's en bepaal de $max. V_{cp}$.

Aantal van ID	Kolomlabels			Eindtotaal	
	G31	overig			
Rijlabels					V_{cp}
1. (Bijna) nooit	68,80%	87,86%	82,20%	19,1%	←max. V_{cp}
2. Soms	93,69%	98,55%	97,11%	4,9%	
3. Vaak	100,00%	100,00%	100,00%	0,0%	
Eindtotaal					

In de G31-steden komt bekladding iets vaker voor dan in de overige woonplaatsen. Het verschil in aanwezigheid van bekladding is middelmatig te noemen.

Vervang de rij-variabele door de variabele *AanwezigheidRommel*. Kies voor 'Waarden weergeven als % voorlopig totaal in' *AanwezigheidRommel* en bepaal de $max. V_{cp}$.

Aantal van ID	Kolomlabels				
	G31	Overig	Eindtotaal		
Rijlabels				V_{cp}	
1. (Bijna) nooit	32,88%	49,58%	44,61%	16,7%	$\leftarrow \max.V_{cp}$
2. Soms	67,87%	83,95%	79,17%	16,1%	
3. Vaak	100,00%	100,00%	100,00%	0,0%	
Eindtotaal					

In de G31-steden komt iets meer rommel voor dan in de overige woonplaatsen. Het verschil in aanwezigheid van rommel is middelmatig te noemen.

Vervang de rij-variabele door de variabele *AanwezigheidPoep*. Kies voor 'Waarden weergeven als % voorlopig totaal in' *AanwezigheidPoep* en bepaal de $\max.V_{cp}$.

Aantal van ID	Kolomlabels				
	G31	Overig	Eindtotaal		
Rijlabels				V_{cp}	
1. (Bijna) nooit	34,54%	36,96%	36,25%	2,4%	
2. Soms	75,04%	70,85%	73,80%	4,2%	$\leftarrow \max.V_{cp}$
3. Vaak	100,00%	100,00%	100,00%	0,0%	
Eindtotaal					

Het verschil in de aanwezigheid van hondenpoep tussen de G31-steden en de overige woonplaatsen is klein. In de G31-steden komt nipt meer hondenpoep voor dan in de overige woonplaatsen.

Het grootste verschil tussen de G31-steden en de overige woonplaatsen is de aanwezigheid van bekladding.

8.4 Onderzoeken

Bladzijde 147

- 35 a *
- b Bij vraag 1: aantal uren besteed aan huiswerk en aantal uren besteed aan bijbaan.
 Bij vraag 2: aantal uren huiswerk bovenbouwleerlingen en aantal uren huiswerk onderbouwleerlingen.
 Bij vraag 3: aantal uren tv kijken.
 Bij vraag 4: aantal uren sporten meisjes en aantal uren sporten jongens.
 *

Bladzijde 148

- 36 Bijvoorbeeld
- Bestaat er een verschil tussen jongens en meisjes in de profielen die zij kiezen?
 - In welke mate is de keuze bepaald door de keuze van vrienden?
 - Is er rekening gehouden met het advies van de mentor?
 - Bestaat er een verband tussen de eindcijfers van de exacte vakken in de 3^e klas en de gekozen profielen?
- Hierbij horen bijvoorbeeld de variabelen
- geslacht (m/v)
 - invloed vrienden (1/2/3/4/5)
 - advies mentor (CM/EM/NG/NT)
 - eindcijfer natuurkunde (1-10)

Bladzijde 149

- 37 a Er wordt aangestuurd op een wenselijk antwoord.
 b Meer vragen in één en de vraag is suggestief.
 c Wat is 'veel'? Onduidelijke vraag.
 d Niet alle antwoordmogelijkheden zijn opgenomen.
- 38 a Hoe ervaart u de veiligheid in deze buurt ten opzichte van een jaar geleden?
 Vergeleken met een jaar geleden voel ik me in deze buurt nu
 I veiliger II even (on)veilig III minder veilig.

- b** Geef aan of u het eens bent met de stelling:
“vegetarisch eten is gezonder dan niet-vegetarisch eten”
I mee eens II mee oneens?
- c** Hoeveel uur heeft vorige week besteed aan sport?
I 0 tot 1 uur II 1 tot 2 uur III 2 tot 3 uur IV 3 tot 5 uur V 5 uur of meer
- d** Op hoeveel dagen heeft u vorige week de website Nu.nl bezocht?
I op 0 dagen II op 1 of 2 dagen III op 3 of 4 dagen V op 5 of meer dagen.

39 Bijvoorbeeld

- Hoeveel uur per week besteed je aan een bijbaan/bijbanen?
I 1 tot 3 uur II 3 tot 5 uur III meer dan 5 uur
- In welke klas zit je?
- Hoeveel uur per week besteed je aan sporten buiten schooltijd om?
I 1 tot 3 uur II 3 tot 5 uur III meer dan 5 uur
- Wat is je geslacht?
I man II vrouw