

Zelf statistiek oefenen



Photo by rawpixel on Unsplash

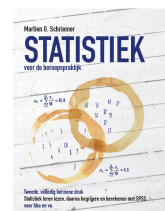
Oefening baat kunst

Antwoorden bij de oefenvragen.

Literatuur

Schriemer, M.G. (2017). *Statistiek voor de beroepspraktijk. Statistiek leren lezen, daarna begrijpen en berekenen met SPSS. Voor hbo en wo*. Tweede herziene druk. Haarlem: SVW

Schriemer, M.G. (2017). *Uitwerkingen*. Eerste druk. Haarlem: SVW



Zelf statistiek oefenen

Multiple Choice

Inleiding vraag 1 tot en met 7

Bekijk de volgende datamatrix (tabel 1.1) waarin klantgegevens van een fitnesscentrum zijn weergegeven.

Tabel 1.1. Afnamegedrag van klanten van een fitnesscentrum

klant	geslacht	maandbedrag (euro's)	betaalwijze	bezoeken
a	m	35	AG	19
b	m	40	CA	16
c	m	35	AI	21
d	v	40	AG	12
e	v	35	AI	9
f	m	40	AG	14
g	v	35	AI	15
h	v	40	AI	19
i	m	35	CA	19
j	m	35	AG	18

Voor een tiental klanten (a t/m j) is weergegeven: het geslacht, het maandbedrag dat zij in euro's betalen, de wijze waarop zij betalen (AG = accept giro, CA = cash, AI = automatisch incasso) en het aantal maal dat zij afgelopen maand de fitness club bezocht hebben.

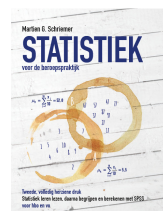
1. Wat is de mediaan van het aantal bezoeken (§2.5.2):

- A. 17 xx
- B. 20
- C. 18
- D. 19

Nummer	Klant	Bezoeken	Mediaan
1	e	9	
2	d	12	
3	f	14	
4	g	15	
5	b	16	
6	j	18	17 ← mediaan
7	a	19	
8	h	19	
9	i	19	
10	c	21	

2. Wat is de modus van het aantal bezoeken (§2.5.2):

- A. 19 xx komt drie keer voor en komt het meest voor
- B. 20
- C. 21
- D. 18



Zelf statistiek oefenen

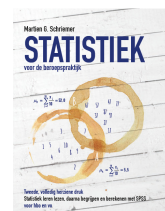
3. Wat is het meetniveau van de onderstaande variabelen? (meerdere antwoorden mogelijk). Zie tabel 4.1. (§2.2.2)

Tabel 4.1. Variabelen hebben betrekking op tabel 1.1.

	A. Nominaal	B. Ordinaal	C. Interval	D. Ratio
a. klant	X			
b. geslacht	X			
c. maandbedrag				X
d. betaalwijze	X			
e. bezoeken				X

4. De range van het aantal bezoeken (§2.4.3):

- A. 15
- B. 13
- C. 14
- D. 12 xx Want maximum - minimum = $21 - 9 = 12$



Zelf statistiek oefenen

5. De variantie van het aantal bezoeken (§2.4.5):

- A. 14,6
- B. 15,6
- C. 12,6 xx
- D. 13,6

Klant (i)	Bezoeken (x_i)	$x_i - \mu$	$(x_i - \mu)^2$
a	19	2,80	7,84
b	16	-0,20	0,04
c	21	4,80	23,04
d	12	-4,20	17,64
e	9	-7,20	51,84
f	14	-2,20	4,84
g	15	-1,20	1,44
h	19	2,80	7,84
i	19	2,80	7,84
j	18	1,80	3,24

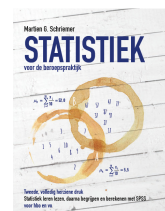
$$\text{Gemiddelde} = \mu = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{10} (162) = 16,20$$

$$\text{Kwadraatsom} = 125,60$$

$$\text{Variantie} = \sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n} = \frac{125,60}{10} = 12,56 = 12,6$$

6. Stel dat de uitkomst van de vorige vraag 16,4 zou zijn (dit is niet juist). Wat is in dat geval de standaarddeviatie van het aantal bezoeken? (§2.4.5)

- A. 7,05
- B. 5,05
- C. 4,05 xx Standaarddeviatie is de wortel van variantie: $\sigma = \sqrt{16,4} = 4,05$
- D. 6,05



Zelf statistiek oefenen

7. Een onderzoek onder 110 yuppen toont aan dat hun gemiddelde maandelijkse uitgaven voor mobiel internetten 45 euro bedraagt. De standaarddeviatie van deze steekproef is 7,50 euro. Wat is het 95% betrouwbaarheidsinterval? (§2.5.2)

- A. van 46,57 tot 49,19 euro
- B. van 44,57 tot 47,43 euro
- C. van 43,81 tot 46,19 euro
- D. van 45,81 tot 48,43 euro

Omdat het hier een betrouwbaarheidsinterval betreft van een steekproef, en niet van een populatie, hanteer het betrouwbaarheidsinterval waarbij de populatiestandaarddeviatie onbekend is: $BI = \bar{x} \pm t \cdot \frac{s}{\sqrt{n}}$. Dan:

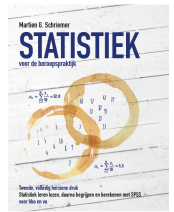
$$BI = 45 \pm 1,66 \cdot \frac{7,50}{\sqrt{110}} = 45 \pm 1,66 \cdot \frac{7,50}{\sqrt{110}} = 45 \pm 1,66 \cdot \frac{7,50}{10,49} = 45 \pm 1,66 \cdot \frac{7,50}{10,49} = 45 \pm 1,19 \text{ dus}$$

$$OG = 45 - 1,19 = 43,81 \text{ en } BG = 45 + 1,19 = 46,19$$

8. Welk antwoord is onjuist? (§1.1.1 t/m §1.1.4)

- A. Een database kan nooit meerdere 'keys' hebben xx dat kan dus wél.
- B. Een bestand kan nooit een database zijn
- C. Informatie bestaat nooit uit meetgegevens
- D. Variabelen zijn nooit meetgegevens

Zelf statistiek oefenen



Inleiding vraag 9

Er wordt onderzoek gedaan naar de prijs van B-merken doucheschuim. Er wordt een steekproef getrokken van 5 verschillende merken, zie tabel 9.1.

Tabel 9.1. Steekproefgegevens over B-merken doucheschuim, x_i is de prijs van merk i .

n	x_i	$(x_i - \mu)$	$(x_i - \mu)^2$
1	90	-23	529
2	140	27	729
3	125	12	144
4	115	2	4
5	95	-18	324

9. Wat is het 95% betrouwbaarheidsinterval voor de prijzen van deze 5 merken uit tabel 9.1? (§2.5.2)

- A. van 84,40 tot 122,60
- B. van 93,19 tot 132,81 xx
- C. van 104,43 tot 121,57
- D. van 114,40 tot 151,60

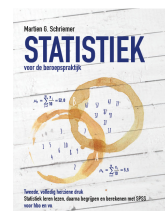
Het gemiddelde bedraagt $\mu = \sum_{i=1}^5 \frac{x_i}{n} = 113$.

De steekproefstandaarddeviatie is $s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{1730}{4}} = \sqrt{432,5} = 20,8$.

Omdat het hier een betrouwbaarheidsinterval betreft van een steekproef, en niet van een populatie, hanteer het betrouwbaarheidsinterval waarbij de populatiestandaarddeviatie onbekend is: $BI = \bar{x} \pm t \cdot \frac{s}{\sqrt{n}}$. Dan:

$$BI = 113 \pm 2,13 \cdot \frac{20,8}{\sqrt{5}} = 113 \pm 2,13 \cdot 9,30 \text{ dus } OG = 113 - 19,81 = 93,19 \text{ en}$$

$$BG = 113 + 19,81 = 132,81$$



Zelf statistiek oefenen

10. In een onderzoek is aan klanten gevraagd of zij tevreden zijn over de dienstverlening van een onderneming. Van de 299 mensen zijn er 239 tevreden. Het 95% betrouwbaarheidsinterval van tevreden klanten loopt (§2.5.2):

- A. van 65,3% tot 74,6%
- B. van 75,4% tot 84,4% xx
- C. van 85,3% tot 96,6%
- D. van 97,3% tot 92,6%

Het betreft hier een betrouwbaarheidsinterval met proporties.

$p = 239 / 299 = 79,9\%$. De steekproefstandaarddeviatie is

$$\sigma = \sqrt{\frac{p \cdot (1 - p)}{n}} = \sqrt{\frac{0,799 \cdot 0,201}{299}} = 0,02317 = 2,3\%$$

Hanteer het betrouwbaarheidsinterval voor proporties onbekend is: $BI = p \pm z \cdot s$. Dan:

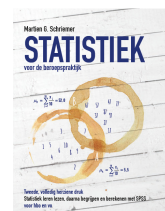
$BI = 79,9 \pm 1,96 \cdot 2,3$, dus $OG = 79,9 - 1,96 \cdot 2,3 = 75,4\%$ en $BG = 79,9 + 1,96 \cdot 2,3 = 84,4$

11. Stel dat bij de vorige vraag (vraag 13) geen 299 maar 399 mensen ondervraagd zijn. Wat gebeurt er dan met het betrouwbaarheidsinterval? Dat betrouwbaarheidsinterval (§2.5.2):

- A. wordt kleiner xx
- B. wordt groter
- C. blijft gelijk
- D. is niet te berekenen zonder additionele informatie

12. Als een betrouwbaarheidsinterval kleiner wordt dan betekent dit (§2.5.2):

- A. dat je uitspraken meer precies zijn xx
- B. dat je uitspraken nergens op slaan
- C. dat je onderzoeksgegevens niet goed zijn
- D. dat je uitspraken onnauwkeuriger worden.



Zelf statistiek oefenen

Inleiding vragen 13 tot en met 15

Een nieuwe Amerikaanse soapserie is aan de doelgroep voorgelegd. De resultaten zijn in onderstaande tabel 13.1 weergegeven waarbij de resultaten voor jongeren (14-24 jaar) en ouderen (25 en ouder) afzonderlijk weergegeven zijn.

Tabel 13.1. Wat vindt u van deze soapserie?

Antwoord	Jongeren	Ouderen	Totaal
Zeer goed/goed	145	135	280
Matig/slecht	35	65	100
Totaal	180	200	380

13. De nulhypothese is "de onderzoeker verwacht dat jongeren de Amerikaanse soapserie beter vinden dan ouderen". Wat is deze hypothese? (§4.8)

- A. is linkszijdig
- B. is tweezijdig
- C. is rechtszijdig xx
- D. is niet goed

De nulhypothese houdt in dat de waardering voor de soapserie van jongeren (x_{jong}) hoger is dan die van ouderen (x_{oud}) $\rightarrow x_{jong} > x_{oud}$. Dus rechtszijdig.

14. De berekende chi-kwadraat bedraagt (§5.2):

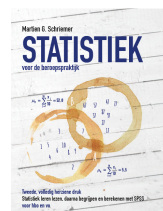
- A. 8,3
- B. 9,3
- C. 10,3
- D. 11,3

Berekening verwachte waarden: $E(ce_{ij}) = \frac{R_i \cdot K_j}{Totaal}$

Verwachte waarden	Jongeren	Ouderen	Totaal
Zeer goed/goed	132,63	147,37	280
Matig/slecht	47,37	52,63	100
Totaal	180	200	380

Cel	f	e	$f - e$	$(f - e)^2$	$\frac{(f - e)^2}{e}$
1,1	145	132,63	12,37	153,02	1,15
1,2	135	147,37	-12,37	153,02	1,04
2,1	35	47,37	-12,37	153,02	3,23
2,2	65	52,63	12,37	153,02	2,91

$$\chi^2 = \sum_{cel} \frac{(f - e)^2}{e} = 8,33$$



Zelf statistiek oefenen

15. De kritieke chi-waarde bij een betrouwbaarheid van 95% bedraagt (§5.2):

- A. 3,84 xx zie bijlage chikwadraatverdeling bij gvv=1.
- B. 4,81
- C. 7,97
- D. 12,1

Vragen 16 tot en met 18 gaan over SPSS.

16. In SPSS is het mogelijk records van een bestand te sorteren. Bij welke menuoptie kun je dat doen? (§1.9, opgave 4)

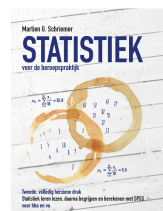
- A. File
- B. Data xx
- C. Analyse
- D. Graphs

17. In SPSS kun je een representativiteittoets uitvoeren om je onderzoeksgegevens te toetsen aan bijvoorbeeld CBS-gegevens. Met welke menuopties (handelingen in SPSS) doe je dat? (§5.9, opgave 2)

- A. In menuoptie Analyse > Nonparametric Tests > Legacy Dialog > kies optie Chisquare xx
- B. In menuoptie Analyse > Descriptive Statistics > Frequencies > kies optie Chisquare
- C. In menuoptie Analyse > Compare Means > Means > kies optie Chisquare
- D. In menuoptie Analyse > Descriptive Statistics > Crosstabs > kies optie Chisquare

18. In SPSS kun je een Chikwadraattoets uitvoeren om de samenhang tussen twee variabelen te toetsen. Met welke menuopties (handelingen in SPSS) doe je dat? (§5.9, opgave 3)

- A. In menuoptie Analyse > Nonparametric Tests > Legacy Dialog > kies optie Chisquare
- B. In menuoptie Analyse > Descriptive Statistics > Frequencies > kies optie Chisquare
- C. In menuoptie Analyse > Compare Means > Means > kies optie Chisquare
- D. In menuoptie Analyse > Descriptive Statistics > Crosstabs > kies optie Chisquare xx



Zelf statistiek oefenen

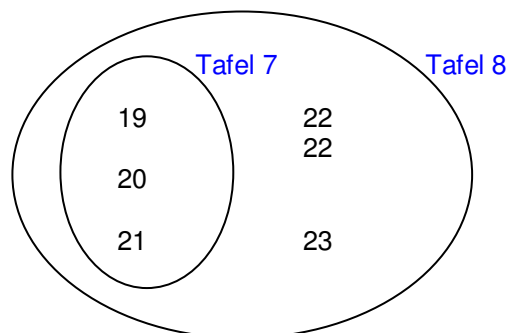
Inleiding vraag 19

Van Het Hikkende Heksje Leeuwarden zijn de gegevens verzameld van twee tafels. Tabel 19.1 geeft daar een weergave van.

Tabel 19.1. Gegevens van tafel 7 en 8.

Tafel	Naam	Leeftijd
7	Arie	19
7	Ben	20
7	Cor	21
8	Dirk	23
8	Eduard	21
8	Frans	19
8	Gerard	20
8	Henk	22
8	Ineke	22

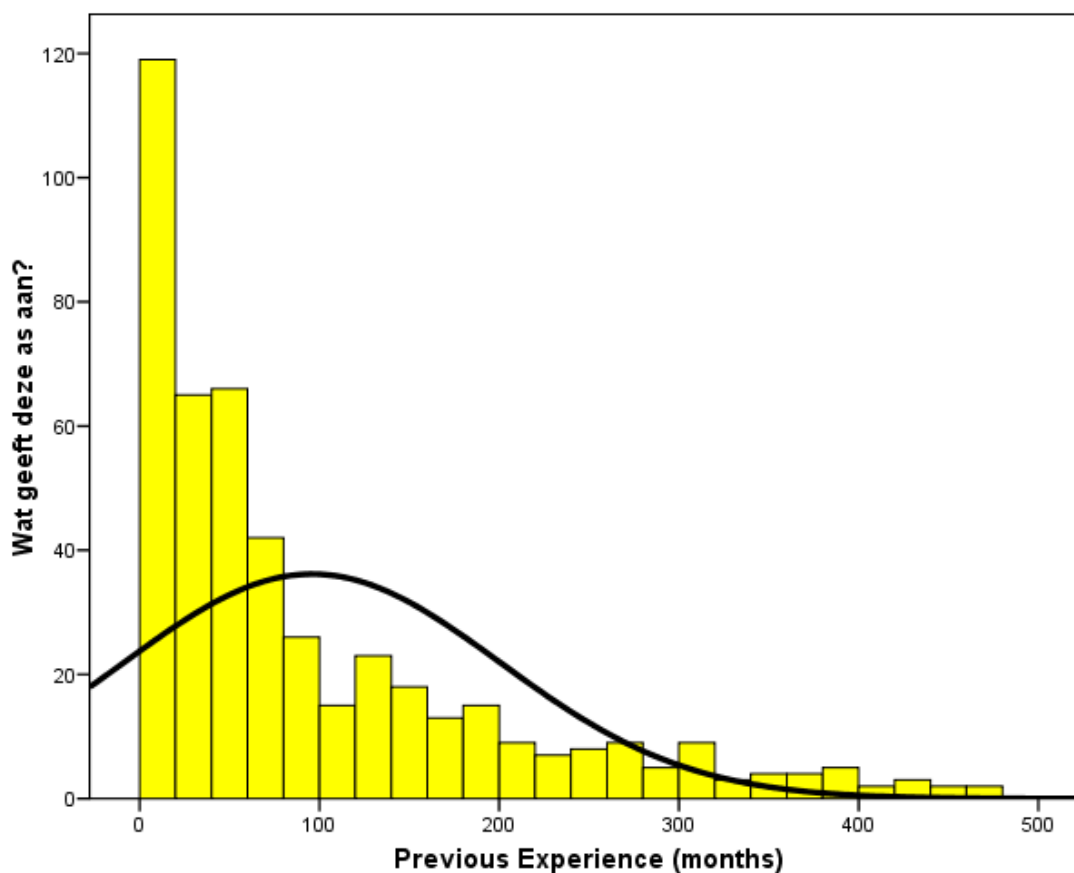
19. Teken het Venn-diagram ten aanzien van leeftijd dat verkregen wordt door $Tafel_7 \cap Tafel_8$. (§1.3)



Zelf statistiek oefenen

20. Zie figuur 20.1. Welke van de onderstaande stellingen is correct? (§2.2.1 en §2.8, opgave 6)

Figuur 20.1. Grafiek.



- A. De variabele "previous experience" is een discrete variabele en op de verticale as staan percentages afgedrukt.
- B. De variabele "previous experience" is een continue variabele en op de verticale as staan aantallen afgedrukt. XX want het betreft een histogram.
- C. De variabele "previous experience" is een continue variabele en op de verticale as staan percentages afgedrukt.
- D. De variabele "previous experience" is een discrete variabele en op de verticale as staan aantallen afgedrukt.

Zelf statistiek oefenen

Open vragen

Inleiding vraag 1 t/m 6.

Het grand café Het Hikkende Heksje 's Hertogenbosch heeft 8 tafels met elk vier stoelen. Stel dat er twee tafels volledig bezet zijn: tafel 3 en tafel 5 (bij het raam). De bestellingen van tafel drie zijn twee koffie (à € 1,75 per stuk) en twee cappuccino's (à € 2,10 per stuk) met één appelgebak (à € 2,85 per stuk). De bestellingen van tafel vijf betreffen drie koffie (à € 1,75 per stuk) en een thee (à € 1,25 per stuk) en vier appelgebak (à € 2,85 per stuk).

Antwoorden

- $T_3 = \{v_1, v_2, v_3, v_4, v_5\} = \{Koffie, Koffie, Cappuccino, Cappuccino, Appelgebak\}$ en
 $T_5 = \{w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8\} = \{Koffie, Koffie, Koffie, Thee, Appelgebak, Appelgebak, Appelgebak, Appelgebak\}$
- $T_3 \cap T_5 = \{Koffie, Koffie, Appelgebak\}$
- Zie tabel 3.1.

Tabel 3.1. Antwoord opgave c.

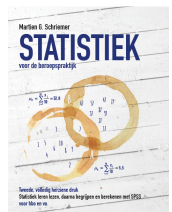
Nummer	Tafel	Bestelling	Prijs in euro
1	3	Koffie	1,75
2	3	Koffie	1,75
3	3	Cappuccino	2,10
4	3	Cappuccino	2,10
5	3	Appelgebak	2,85
6	5	Koffie	1,75
7	5	Koffie	1,75
8	5	Koffie	1,75
9	5	Thee	1,25
10	5	Appelgebak	2,85
11	5	Appelgebak	2,85
12	5	Appelgebak	2,85
13	5	Appelgebak	2,85

- Bereken de gemiddelde prijs per persoon voor tafel drie en voor tafel vijf. Welke gemiddelde is groter?

$$\mu_3 = \frac{1}{5} \sum_{i=1}^5 v_i = \frac{1}{5} (1,75 + 1,75 + 2,10 + 2,10 + 2,85) = 2,11$$

$$\mu_5 = \frac{1}{8} \sum_{i=1}^8 w_i = \frac{1}{8} (1,75 + 1,75 + 1,75 + 1,25 + 2,85 + 2,85 + 2,85 + 2,85) = 2,24$$

Dus $\mu_5 > \mu_3$.



Zelf statistiek oefenen

5. De berekening kan op twee manieren. De eerste manier is alle dertienbestellingen optellen en delen door het totaal:

$$\mu = \frac{1}{13} \sum_{i=1}^{13} t_i = \frac{1}{13} (1,75 + \dots + 2,85) = 2,19$$

Een andere manier is een gewogen gemiddelde, oftewel gebruik maken van de relatieve frequenties f_i^r met $t = 3$ of $t = 5$.

$$\mu = \frac{(f_3^r \cdot \mu_3 + f_5^r \cdot \mu_5)}{13} = \frac{5 \cdot 2,11 + 8 \cdot 2,24}{13} = 2,19$$

Uiteraard ligt het totaalgemiddelde μ tussen μ_3 en μ_5 in, oftewel in wiskundige notatie:

$$\mu_3 < \mu < \mu_5.$$

6. De drie personen van tafel twee bestellen samen: $3 \cdot (2,10) = 6,30$. Het totaalgemiddelde is: $\mu = 2,19$. Nu moet de vierde persoon zoveel bestellen dat geldt:

$$\mu_{\text{nieuw}}: \frac{14,85 + x_4}{4} = 6,30 + 0,50 \Rightarrow 14,85 + x_4 = 4 \cdot (6,80) \Rightarrow x_4 = 27,20 - 14,85 = 12,35$$

Dus de vierde persoon moet nogal veel bestellen in vergelijking met de anderen wil het totaalgemiddelde vijftig eurocent hoger te komen.