

Zelf statistiek oefenen



Photo by rawpixel on Unsplash

Oefening baat kunst

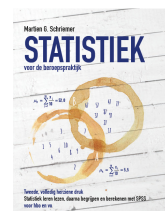
Nu zelf aan de slag. De vragen staan door elkaar. Er zijn multiplechoicevragen en open vragen. In de toekomst komen er vragen bij. Het is een greep van de mogelijke vragen die je kunt verwachten op een tentamen. Zorg ervoor dat je van te voren weet wat voor soort tentamen je krijgt en op welke stof (hoofdstukken, paragrafen) het tentamen betrekking heeft.

Bij elke vraag is verwezen naar de betreffende paragraaf uit het boek van Schriemer (2017) en de bijbehorende uitwerkingen van de opgaven ("Uitwerkingen"). Bij het maken van de opgaven zijn gegeven: Formuleblad en de chikwadraatverdeling, t-verdeling en standaardnormale verdeling (zie bijlagen van dit document)

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

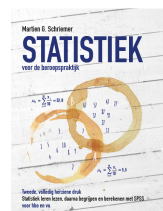
Voor een tiental klanten (a t/m j) is weergegeven: het geslacht, het maandbedrag dat zij in euro's betalen, de wijze waarop zij betalen (AG = accept giro, CA = cash, AI = automatisch incasso) en het aantal maal dat zij afgelopen maand de fitness club bezocht hebben.

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

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

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

- A. 19
- 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				
b. geslacht				
c. maandbedrag				
d. betaalwijze				
e. bezoeken				

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

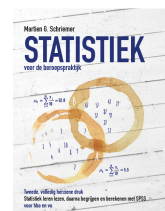
- A. 15
- B. 13
- C. 14
- D. 12

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

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

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
- 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

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

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

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

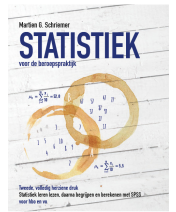
9. Wat is het 95% betrouwbaarheidsinterval voor de prijzen van deze 5 merken uit tabel 9.1? (§2.5.2)

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

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%
- C. van 85,3% tot 96,6%
- D. van 97,3% tot 92,6%

Zelf statistiek oefenen



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

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?

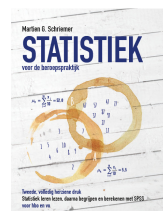
Antwoord	Jongeren	Ouderen	Totaal
Zeer goed/goed	145	135	280
Matig/slecht	35	65	100
Totaal	180	200	380

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
- D. is niet goed

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

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



Zelf statistiek oefenen

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

- A. 3,84
- 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)

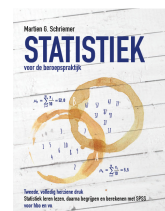
- A. File
- B. Data
- C. Analyse
- D. Graphs

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
- 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



Zelf statistiek oefenen

Inleiding vraag 19

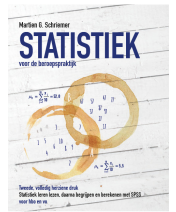
Van Het Hikkende Heksje Leeuwarden zijn de gegevens verzameld van twee tafels. Tabel 19.1 geeft daar een weergave van.

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

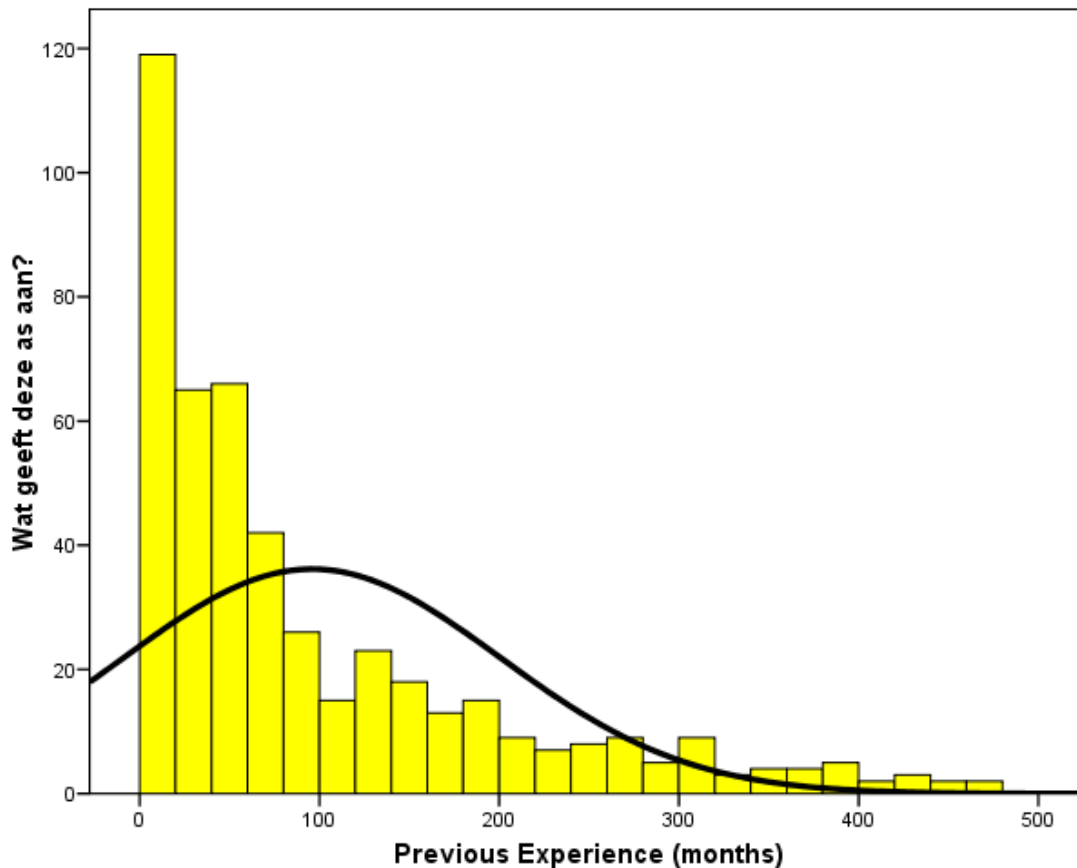
19. Teken het Venn-diagram ten aanzien van leeftijd dat verkregen wordt door $Tafel_7 \cap Tafel_8$. (§1.3)

Zelf statistiek oefenen

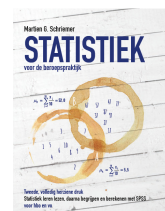


20. Zie figuur 20.1. Welke van de onderstaande stellingen is correct? (§2.2.1 en §2.8, opgave 6)

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.
- 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.



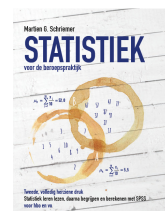
Zelf statistiek oefenen

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).

1. **Beschrijf de verzamelingen van tafel drie en tafel vijf met behulp van formele wiskundige notatie. (§§1.3)**
2. **Beschrijf vervolgens de deelverzameling met behulp van formele wiskundige notatie van de tafels drie en vijf. (§1.4)**
3. **Maak één tabel van alle bestellingen van de tafels drie en vijf zodat de bediening ze straks uit kan serveren. Zorg voor duidelijke kolomtitels. (§1.4)**
4. **Bereken de gemiddelde prijs per persoon voor tafel drie en voor tafel vijf. Welke gemiddelde is groter? (§1.5)**
5. **Bereken de totale gemiddelde bestelling per persoon (van tafel drie en vijf tezamen). (§1.5)**
6. **Stel dat er een viertal personen binnenkomt en gaat zitten aan tafel twee. Stel dat drie personen allemaal cappuccino (à € 2,10 per stuk) bestellen. Wat moet de bestelling (in euro's) van de vierde persoon zijn om het totaalgemiddelde per tafel van alle bestellingen daardoor een vijftig eurocent hoger komt te liggen? (§1.5)**



Zelf statistiek oefenen

Bijlagen

Formuleblad (1)

Omvang populatie N

Omvang steekproef n

Populatiegemiddelde $\mu = \frac{1}{N} \sum_{i=1}^N x_i$

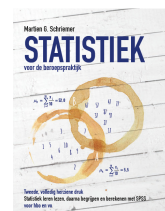
Steekproefgemiddelde $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

Variantie (populatie) $\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$

Standaarddeviatie (populatie) $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$

Steekproefvariantie $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

Steekproefstandaarddeviatie $s = \sqrt{s^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$



Zelf statistiek oefenen

Formuleblad (2)

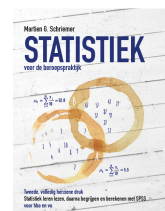
Chikwadraat
$$\chi^2 = \sum_{cel} \frac{(f_{cel} - e_{cel})^2}{e_{cel}}$$

Proportie
$$P(X_i = A) = \frac{N\pi}{N} = \pi$$

Standaarddeviatie bij proportie
$$\sigma = \sqrt{\frac{p \cdot q}{n}}$$

Schattingsfout bij proportie
$$\sigma_p = \sqrt{\frac{p \cdot q}{n}}$$

Schattingsfout bij metrische variabelen
$$\sigma_m = \frac{\sigma}{\sqrt{n}}$$



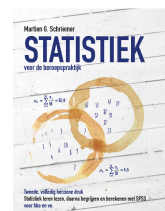
Zelf statistiek oefenen

Bijlage Chikwadraatverdeling

Onderstaande tabel geeft de kritieke waarde van chikwadraat aan. Verwerp H_1 oftewel accepteer H_0 als

$$\chi^2_{\alpha} < \chi^2_{\text{Kritiek}}$$

Graden van vrijheid	Betrouwbaarheidspercentage ($1 - \alpha$)					
	75,0%	80,0%	90,0%	95,0%	99,0%	99,9%
1	1,323	1,642	2,706	3,841	6,635	10,828
2	2,773	3,219	4,605	5,991	9,210	13,816
3	4,108	4,642	6,251	7,815	11,345	16,266
4	5,385	5,989	7,779	9,488	13,277	18,467
5	6,626	7,289	9,236	11,070	15,086	20,515
6	7,841	8,558	10,645	12,592	16,812	22,458
7	9,037	9,803	12,017	14,067	18,475	24,322
8	10,219	11,030	13,362	15,507	20,090	26,124
9	11,389	12,242	14,684	16,919	21,666	27,877
10	12,549	13,442	15,987	18,307	23,209	29,588
11	13,701	14,631	17,275	19,675	24,725	31,264
12	14,845	15,812	18,549	21,026	26,217	32,909
13	15,984	16,985	19,812	22,362	27,688	34,528
14	17,117	18,151	21,064	23,685	29,141	36,123
15	18,245	19,311	22,307	24,996	30,578	37,697
16	19,369	20,465	23,542	26,296	32,000	39,252
17	20,489	21,615	24,769	27,587	33,409	40,790
18	21,605	22,760	25,989	28,869	34,805	42,312
19	22,718	23,900	27,204	30,144	36,191	43,820
20	23,828	25,038	28,412	31,410	37,566	45,315
21	24,935	26,171	29,615	32,671	38,932	46,797
22	26,039	27,301	30,813	33,924	40,289	48,268
23	27,141	28,429	32,007	35,172	41,638	49,728
24	28,241	29,553	33,196	36,415	42,980	51,179
25	29,339	30,675	34,382	37,652	44,314	52,620
26	30,435	31,795	35,563	38,885	45,642	54,052
27	31,528	32,912	36,741	40,113	46,963	55,476
28	32,620	34,027	37,916	41,337	48,278	56,892
29	33,711	35,139	39,087	42,557	49,588	58,301
30	34,800	36,250	40,256	43,773	50,892	59,703
50	56,334	58,164	63,167	67,505	76,154	86,661
70	77,577	79,715	85,527	90,531	100,425	112,317
100	109,141	111,667	118,498	124,342	135,807	149,449



Zelf statistiek oefenen

Bijlage standaardnormale verdeling

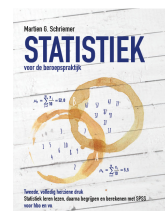
De tabel¹ III.1 geeft voor de standaardnormaal verdeelde variabele z_i voor verschillende waarden van z de overschrijdingskans $\frac{1}{2} \alpha = P(\mu < x_i < \mu + z_{\frac{1}{2}\alpha} \cdot \sigma)$.

Tabel III.1. Overschrijdingskansen voor diverse z -waarden.

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,10	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,20	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,30	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,40	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,50	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,60	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,70	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,80	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,90	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,00	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,10	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,20	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,30	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,40	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,50	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,60	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,70	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,80	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,90	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767

Vervolg tabel: volgende bladzijde

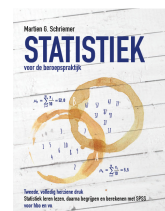
¹ De tabel is berekend m.b.v. SPSS, versie 15.0, via de functie 'CDF.NORMAL(zwaarde,0,1)'.



Zelf statistiek oefenen

Vervolg tabel III.1. Overschrijdingskansen voor diverse z-waarden.

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
2,00	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,10	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,20	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,30	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,40	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,50	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,60	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,70	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,80	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,90	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,00	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,10	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,20	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,30	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,40	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,50	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,60	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,70	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,80	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,90	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000



Zelf statistiek oefenen

Bijlage t-verdeling

De tabel² IV.1 geeft voor de t-verdeelde variabele t voor verschillende graden van vrijheid de kritieke t-waarde:

- Eenzijdig (zie tabel bij '1'): $1 - \alpha = P(t > T)$.
- Tweezijdig (zie tabel bij '2'): $1 - \alpha = P(-t < T < t)$.

Tabel IV.1. Kritieke t-waarde voor verschillende graden van vrijheid.

df	Alpha				df	Alpha			
1	0,1000	0,0500	0,0100	0,0010	1	0,1000	0,0500	0,0100	0,0010
2	0,0500	0,0250	0,0050	0,0005	2	0,0500	0,0250	0,0050	0,0005
1	3,08	6,31	31,82	318,31	20	1,33	1,72	2,53	3,55
2	1,89	2,92	6,96	22,33	21	1,32	1,72	2,52	3,53
3	1,64	2,35	4,54	10,21	22	1,32	1,72	2,51	3,50
4	1,53	2,13	3,75	7,17	23	1,32	1,71	2,50	3,48
5	1,48	2,02	3,36	5,89	24	1,32	1,71	2,49	3,47
6	1,44	1,94	3,14	5,21	25	1,32	1,71	2,49	3,45
7	1,41	1,89	3,00	4,79	26	1,31	1,71	2,48	3,43
8	1,40	1,86	2,90	4,50	27	1,31	1,70	2,47	3,42
9	1,38	1,83	2,82	4,30	28	1,31	1,70	2,47	3,41
10	1,37	1,81	2,76	4,14	29	1,31	1,70	2,46	3,40
11	1,36	1,80	2,72	4,02	30	1,31	1,70	2,46	3,39
12	1,36	1,78	2,68	3,93	40	1,30	1,68	2,42	3,31
13	1,35	1,77	2,65	3,85	50	1,30	1,68	2,40	3,26
14	1,35	1,76	2,62	3,79	70	1,29	1,67	2,38	3,21
15	1,34	1,75	2,60	3,73	100	1,29	1,66	2,36	3,17
16	1,34	1,75	2,58	3,69	200	1,29	1,65	2,35	3,13
17	1,33	1,74	2,57	3,65	∞	1,28	1,64	2,33	3,09
18	1,33	1,73	2,55	3,61					
19	1,33	1,73	2,54	3,58					

² De tabel is berekend m.b.v. SPSS, versie 15.0, via de functie 'IDF.T(alpha,df)'. Voor ∞ is 100000 gehanteerd.