INLEIDING TOT DE BIOMEDISCHE STATISTIEK

CD-ROM

Opgeloste oefeningen

Georges De Moor
Georges Van Maele

Inhoudsopgave

Oefeningen

Oef 1. Inleiding

Deze verzameling aan zorgvuldig uitgezochte opgaven met toepassingen uit de gezondheidswetenschappen is ingedeeld in een aantal hoofdstukken die grotendeels overeenstemmen met de volgorde zoals aangehouden in het theoretisch gedeelte van het boek.

Deze oefeningen moeten de student helpen de leerstof goed te beheersen en in de praktijk toe te passen. Voor de uitwerking van de opgaven volstaat het te beschikken over een zakrekenmachine die de basis wiskundige bewerkingen (machtsverheffing y^x, getal e en π, logaritme en trigonometrische functies) aankan.

Andere oefeningen kunnen beter met een statistisch pakket worden opgelost. Deze CD-ROM bevat tevens de oplossingen van bepaalde oefeningen in SPSS en R.

Oef 2. Probabiliteit en medische besliskunde

2.1 Opgooien van een regelmatig geldstuk

Bij het negen maal opgooien van een regelmatig geldstuk werd zeven maal kruis bekomen. Bereken de kans dat bij de tiende worp het resultaat munt is. [0.50]

2.2 Falen nieuw type rookdetector

Stel dat een nieuw type rookdetector in één op tweeduizend gevallen faalt bij brand. Wegens veiligheidsoverwegingen wordt besloten om in een laboratorium twee van deze detectoren te plaatsen.

a) Wat is de kans dat er geen rook gesignaleerd wordt terwijl er wel degelijk brand is in het laboratorium? [0.00000025]

b) Wat is de kans dat beide detectoren alarm slaan wanneer er inderdaad een brand is? [0.99900025]

c) Wat is de kans dat slechts één van de detectoren werkt in geval van brand? [0.0009995]

Oplossing

Stel: B = brand
A_1 = 1^e detector slaat alarm
A_2 = 2^e detector slaat alarm

a. Wat is de kans dat er geen rook gesignaleerd wordt terwijl er wel degelijk brand is in het laboratorium?

$$\mathbf{P(\overline{A}_1 \cap \overline{A}_2 \mid B) = P(\overline{A}_1 \mid B) \times P(\overline{A}_2 \mid B)}$$
$$= \frac{1}{2000} \times \frac{1}{2000} = 0.00000025 = 0.25 \times 10^{-6}$$

b. Wat is de kans dat beide detectoren alarm slaan wanneer het inderdaad brandt?

$$\mathbf{P(A_1 \cap A_2 \mid B) = P(A_1 \mid B) \times P(A_2 \mid B)}$$
$$= \left(1 - \frac{1}{2000}\right) \times \left(1 - \frac{1}{2000}\right)$$
$$= \frac{1999}{2000} \times \frac{1999}{2000} = \frac{3996001}{4000000}$$
$$= 0.99900025$$

c. Wat is de kans dat slechts één van de detectoren werkt in geval van brand?

$$\mathbf{P(A_1 \cap \overline{A}_2 \mid B) + P(\overline{A}_1 \cap A_2 \mid B)}$$

eerste term: $\mathbf{P(A_1 \mid B) \times P(\overline{A}_2 \mid B)} = \left(1 - \frac{1}{2000}\right) \times \frac{1}{2000}$

tweede term: $\mathbf{P(\overline{A}_1 \mid B) \times P(A_2 \mid B)}$ **is analoog**

$$\mathbf{P(A_1 \cap \overline{A}_2 \mid B) + P(\overline{A}_1 \cap A_2 \mid B)} = 2 \times \frac{1999}{2000} \times \frac{1}{2000} = \frac{1999}{2000000} = 0.0009995$$

2.3 Kans op bepaalde gebeurtenis

Bepaal voor elk van de volgende experimenten het universum Ω en bereken de kans van de gebeurtenis A, respectievelijk de gebeurtenis B:

a. een onvervormde dobbelsteen werpen
A = aantal ogen deelbaar door 2 of door 3 [0.67]
B = aantal ogen is minstens 2 [0.83]

b. een kaart trekken uit een volledig kaartspel
A = een aas of een heer trekken [0.15]
B = een aas of een schoppen trekken [0.31]

c. twee onvervormde dobbelstenen werpen
A = som der ogen is 7 of 8 [0.31]
B = absolute waarde van het verschil tussen het aantal ogen is 0, 1 of 2 [0.67]

d. twee personen kiezen elk lukraak een getal uit de verzameling {1, 2, 3, 4, 5}
A = ze kiezen allebei hetzelfde getal [0.20]
B = de absolute waarde van het verschil tussen de twee gekozen getallen bedraagt niet meer dan twee [0.76]

e. uit een vaas die drie rode ballen (R_1 R_2 R_3) en twee witte ballen (W_1 W_2) bevat een bal trekken, deze terugleggen, en vervolgens een tweede bal trekken
A = de eerste bal is rood [0.60]
B = beide ballen zijn rood [0.36]

f. een onvervormd muntstuk viermaal opgooien
A = minstens één kruis bekomen [0.94]
B = viermaal hetzelfde bekomen [0.12]

Oplossing

a. een onvervormde dobbelsteen werpen
A = aantal ogen deelbaar door 2 of door 3
B = aantal ogen is minstens 2

$$\Omega_A = \{1, \underline{2}, \underline{3}, \underline{4}, 5, \underline{6}\}$$

$$P(A) = \frac{4}{6} = \frac{2}{3} = 0.67$$

$$\text{optellings wet: } P(A) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{2}{3} = 0.67$$

$\Omega_B = \{1, \underline{2}, \underline{3}, \underline{4}, \underline{5}, \underline{6}\}$

$P(B) = \frac{5}{6} = 0.83$

negatie : $P(B) = 1 - P(\overline{B}) = 1 - \frac{1}{6} = \frac{5}{6} = 0.83$

b. een kaart trekken uit een volledig kaartspel

A = een aas of een heer trekken

B = een aas of een schoppen trekken

$\Omega_A = \{\underline{A}, 2, 3, \ldots, B, D, \underline{H}\}$

$P(A) = \frac{2}{13} = 0.15$

optellings wet : $P(aas) + P(heer) = \frac{1}{13} + \frac{1}{13}$

$$\begin{aligned}\Omega_B = \{&\spadesuit \underline{A, 2, 3, \ldots, B, D, H},\\ &\heartsuit \underline{A}, 2, 3, \ldots, B, D, H,\\ &\diamondsuit \underline{A}, 2, 3, \ldots, B, D, H,\\ &\clubsuit \underline{A}, 2, 3, \ldots, B, D, H\}\end{aligned}$$

$P(B) = \frac{16}{52} = 0.31$

optellings wet : $P(aas + schoppen) = \frac{1}{13} + \frac{1}{4} - \frac{1}{52} = \frac{4 + 13 - 1}{52} = \frac{16}{52} = 0.31$

c. twee onvervormde dobbelstenen werpen

A = som van de ogen is 7 of 8

B = absolute waarde van het verschil tussen het aantal ogen is 0, 1 of 2

$$\begin{aligned}\Omega_A = \{&(1,1)(1,2)(1,3)(1,4)(1,5)\underline{(1,6)}\\ &(2,1)(2,2)(2,3)(2,4)\underline{(2,5)}\underline{(2,6)}\\ &(3,1)(3,2)(3,3)\underline{(3,4)}\underline{(3,5)}(3,6)\\ &(4,1)(4,2)\underline{(4,3)}\underline{(4,4)}(4,5)(4,6)\\ &(5,1)\underline{(5,2)}\underline{(5,3)}(5,4)(5,5)(5,6)\\ &\underline{(6,1)}\underline{(6,2)}(6,3)(6,4)(6,5)(6,6)\}\end{aligned}$$

$P(A) = \frac{11}{36} = 0.31$

optellings wet: $P(A) = \frac{6}{36} + \frac{5}{36} = \frac{11}{36} = 0.31$

$$
\begin{aligned}
\Omega_B = \{ & \underline{(1,1)}\ \underline{(1,2)}\ \underline{(1,3)}\ (1,4)\ (1,5)\ (1,6) \\
& \underline{(2,1)}\ \underline{(2,2)}\ \underline{(2,3)}\ \underline{(2,4)}\ (2,5)\ (2,6) \\
& \underline{(3,1)}\ \underline{(3,2)}\ \underline{(3,3)}\ \underline{(3,4)}\ \underline{(3,5)}\ (3,6) \\
& (4,1)\ \underline{(4,2)}\ \underline{(4,3)}\ \underline{(4,4)}\ \underline{(4,5)}\ \underline{(4,6)} \\
& (5,1)\ (5,2)\ \underline{(5,3)}\ \underline{(5,4)}\ \underline{(5,5)}\ \underline{(5,6)} \\
& (6,1)\ (6,2)\ (6,3)\ \underline{(6,4)}\ \underline{(6,5)}\ \underline{(6,6)} \}
\end{aligned}
$$

$$P(B) = \frac{24}{36} = \frac{2}{3} = 0.67$$

$$\text{optellings wet : } P(B) = \frac{6}{36} + \frac{10}{36} + \frac{8}{36} = \frac{24}{36}$$

d. twee personen kiezen elk lukraak een getal uit de verzameling {1, 2, 3, 4, 5}

A = ze kiezen allebei hetzelfde getal
B = de absolute waarde van het verschil tussen de twee gekozen getallen bedraagt niet meer dan twee

$$
\begin{aligned}
\Omega_A = \{ & \underline{(1,1)}\ (1,2)\ (1,3)\ (1,4)\ (1,5) \\
& (2,1)\ \underline{(2,2)}\ (2,3)\ (2,4)\ (2,5) \\
& (3,1)\ (3,2)\ \underline{(3,3)}\ (3,4)\ (3,5) \\
& (4,1)\ (4,2)\ (4,3)\ \underline{(4,4)}\ (4,5) \\
& (5,1)\ (5,2)\ (5,3)\ (5,4)\ \underline{(5,5)} \}
\end{aligned}
$$

$$P(A) = \frac{5}{25} = \frac{1}{5} = 0.20$$

$$
\begin{aligned}
\Omega_B = \{ & \underline{(1,1)}\ \underline{(1,2)}\ \underline{(1,3)}\ (1,4)\ (1,5) \\
& \underline{(2,1)}\ \underline{(2,2)}\ \underline{(2,3)}\ \underline{(2,4)}\ (2,5) \\
& \underline{(3,1)}\ \underline{(3,2)}\ \underline{(3,3)}\ \underline{(3,4)}\ \underline{(3,5)} \\
& (4,1)\ \underline{(4,2)}\ \underline{(4,3)}\ \underline{(4,4)}\ \underline{(4,5)} \\
& (5,1)\ (5,2)\ \underline{(5,3)}\ \underline{(5,4)}\ \underline{(5,5)} \}
\end{aligned}
$$

$$P(B) = \frac{19}{25} = 0.76$$

e. uit een vaas die drie rode ballen (R_1 R_2 R_3) en twee witte ballen (W_1 W_2) bevat een bal trekken, deze terugleggen, en vervolgens een tweede bal trekken

A = de eerste bal is rood
B = beide ballen zijn rood

$$P(A) = \frac{3}{5} = 0.60$$

$$P(B) = P(A).P(A) = \frac{3}{5} \times \frac{3}{5} = \frac{9}{25} = 0.36$$

f. een onvervormd muntstuk viermaal opgooien
A = minstens één kruis bekomen
B = viermaal hetzelfde bekomen

$$\begin{aligned}\Omega = \{&(KKKK)\ (KKKM)\ (KKMK)\ (KKMM)\\ &(KMKK)\ (KMKM)\ (KMMK)\ (KMMM)\\ &(MKKK)\ (MKKM)\ (MKMK)\ (MKMM)\\ &(MMKK)\ (MMKM)\ (MMMK)\ (MMMM)\}\end{aligned}$$

$$P(A) = \frac{15}{16} = 0.94; \qquad P(A) = 1 - P(\overline{A}) = 1 - \frac{1}{16} = \frac{15}{16} = 0.94$$

$$P(B) = \frac{2}{16} = \frac{1}{8} = 0.12$$

2.4 Bloedafname

Verpleegster Nursina prikt een ader met een waarschijnlijkheid op succes van 0.80 en haar collega Zustrina prikt een ader met een kans 0.90. Wat is de kans dat indien beiden samen zouden prikken de ader minstens één keer succesvol wordt geprikt. [0.98]

Oplossing

N → Nursina prikt met succes
Z → Zustrina prikt met succes

$P(N) = 0.80$ $P(Z) = 0.90$

	algemene optellingswet
$P(N \text{ of } Z) = P(N) + P(Z) - P(N \text{ en } Z)$	niet mutueel exclusief
$= 0.80 + 0.90 - P(N).P(Z)$	onafhankelijk
$= 1.7 - 0.8 \times 0.9 = 1.7 - 0.72 = 0.98$	

2.5 Laboratoriumtest

Noem D het verschijnsel dat iemand een bepaalde ziekte heeft (D ~ disease) en S het verschijnsel dat het test resultaat positief is (S ~ symptoom).

Zij verder P(D) = 0.005.

a. Stel dat men beschikt over een laboratoriumtest zodanig dat P(S|D) = 0.95 en $P(\overline{S}|\overline{D}) = 0.95$. Wat is de kans dat de geteste persoon (patiënt) werkelijk de ziekte heeft als het testresultaat positief is? [0.087]

b. Zij verder $P(S|D) = P(\overline{S}|\overline{D}) = k$. Voor welke waarde van k is P(D|S) gelijk aan 0.95? [0.99974]

Oplossing

D → patiënt is ziek

S → testresultaat is positief

P(S|D) = 0.95; P(D) = 0.005; $P(\overline{S}|\overline{D}) = 0.95$

$$P(D|S) = \frac{P(S|D)\times P(D)}{P(S|D)\times P(D) + P(S|\overline{D})\times P(\overline{D})} = \frac{0.00475}{P(S)}$$

$$= \frac{0.95\times 0.005}{0.95\times 0.005 + (1-0.95)\times(1-0.005)}$$

$$\text{dus}\quad P(D|S) = \frac{0.00475}{0.0545} = 0.087$$

b. Zij verder $P(S|D) = P(\overline{S}|\overline{D}) = k$. Voor welke waarde van k is P(D|S) gelijk aan 0.95?

$$P(D|S) = \frac{P(S|D)\times P(D)}{P(S|D)\times P(D) + P(S|\overline{D})\times P(\overline{D})}$$

$$0.95 = \frac{k\times 0.005}{k\times 0.005 + (1-k)\times(1-0.005)}$$

$$0.005\times k = 0.95\times[k\times 0.005 + (1-k)\times(1-0.005)]$$

of

$$0.005\times k = 0.95\times(0.005\times k + 0.995 - 0.995\times k)$$

$$k = \frac{0.95\times 0.995}{0.005 - 0.95\times 0.005 + 0.95\times 0.995}$$

$$k = 0.99974$$

2.6 Kraaminrichting

In een kraaminrichting zijn er 1000 pasgeboren baby's waarvan 600 jongens. De kans dat een jongen meer dan 4000g weegt is 0.20. Voor een meisje is deze kans 0.10. Bereken de kans dat een pasgeborene meer dan 4 kg weegt. [0.16]

Oplossing

A → gewicht jongen > 4000g
B → gewicht meisje > 4000g

$P(A) = P(W^{+++}|J) = 0.20$
$P(B) = P(W^{+++}|M) = 0.10$

$$P(J) = \frac{600}{1000} = \frac{3}{5} = 0.60$$
$$P(M) = 1 - P(J) = 0.40$$

zie afleiding van de noemer in de regel van Bayes:

$$P(S \mid D) \times P(D) + P(S \mid \overline{D}) \times P(\overline{D}) = P(S) \quad \text{of}$$
$$P(W^{+++}) = P(W^{+++} \mid J) \times P(J) + P(W^{+++} \mid M) \times P(M)$$

$P(W^{+++}) = 0.60 \times 0.20 + 0.40 \times 0.10 = 0.16$

2.7 Wordt het een jongen of een meisje?

Indien gegeven is dat de eerstgeborene een meisje is, wat is de kans op een jongen bij de geboorte van een tweede kind? Uit studies is gebleken dat P(jongen) = 0.52 en P(meisje) = 0.48. [0.52]

Oplossing

Geslachtsverdeling:

jongens	51.2%
meisjes	48.8%

Ref. SPE - Jaarverslag 1997 (Studiecentrum voor Perinatale Epidemiologie)

Uiteraard blijft de kans bij elk nieuwgeborene gelijk aan wat ze is.

2.8 Verjaren op 29 februari

a. Wat is de kans om te verjaren op 29 februari? Hoeveel Belgen zijn dit?

[P = 0.00068, n = 6845]

b. Wat is de kans voor Belgen om zowel geboren te worden op 29 februari als ook te sterven op 29 februari? [P = 0.00000047, n = 5]

Uitgangspunten voor deze oefening:

a) er zijn 10 miljoen Belgen;

b) geboorten en sterfte zijn gelijkmatig over het jaar verspreid; populatie blijft enigszins onveranderd

c) enkel de periode tussen de eeuwwisselingen wordt in aanmerking genomen, waarin vierjaarlijkse periodes te onderscheiden zijn met drie jaren van 365 dagen en één jaar van 366 dagen (immers, de eeuwwisseling is geen schrikkeljaar)

Oplossing

G ➔ geboren worden op 29 februari
D ➔ overlijden op 29 februari

$$\text{a. aantal dagen} = (3 \times 365) + 366 = 1461$$

$$P(G) = \frac{1}{1461} = 0.0006845$$

Besluit: 6845 van de 10 miljoen Belgen verjaren op 29 februari

$$\text{b. } P(D) = P(G) = \frac{1}{1461}$$

$$P(G \text{ en } D) = P(G) \times P(D) = \frac{1}{1461^2}$$

$$= 0.00000047$$

Besluit: vijf van de 10 miljoen Belgen verjaren en sterven op 29 februari

2.9 Zwemmers aan de Belgische kust

In een studie wordt nagegaan of zwemmers aan de Belgische kust een verhoogd risico hebben op acute infecties.
Duizend zwemmers worden gevolgd. Tevens worden gezondheidsgegevens opgemeten bij 1000 niet-zwemmers aan dezelfde kust.

Tabel. Aantal zieken met acute infectie bij zwemmers en niet-zwemmers aan de Belgische kust

		zwemmen? ja	zwemmen? neen	
acute infectie?	**ja**	200	100	300
	neen	800	900	1700
		1000	1000	2000

a) Bereken de kans op infecties bij zwemmers en bij niet zwemmers. Vergelijk beide kansen door ze op elkaar te betrekken (= relatief risico) [0.10, 0.20, 2]
b) Bereken de odds-ratio [2.25]

Oplossing

$\mathbf{I} = \textbf{acute infectie} \qquad \bar{\mathbf{I}} = \textbf{geen infectie}$

$\mathbf{Z} = \textbf{zwemmer} \qquad \bar{\mathbf{Z}} = \textbf{geen zwemmer}$

$$\mathbf{P(I \mid Z)} = \frac{200}{1000} = 0.20 \qquad \mathbf{P(I \mid \bar{Z})} = \frac{100}{1000} = 0.10$$

$$\textbf{a.}\quad \mathbf{RR}_{\text{zwemmer/niet-zwemmer}} = \frac{0.20}{0.10} = 2$$

$$\textbf{b.}\quad \mathbf{OR} = \frac{\left[\dfrac{P(I)}{P(\bar{I})}\right]_{\text{zwemmers}}}{\left[\dfrac{P(I)}{P(\bar{I})}\right]_{\text{niet-zwemmers}}} = \frac{\dfrac{200}{800}}{\dfrac{100}{900}} = \frac{200 \times 900}{800 \times 100} = \frac{18}{8} = \frac{9}{4} = 2.25$$

2.10 Relatie tussen roken en hypertensie

In volgende tabel zijn enkele aantallen weergegeven omtrent de relatie tussen roken en hypertensie (hoge bloeddruk) bij een groep mannen en vrouwen:

	MANNEN		VROUWEN		TOTAAL
	roker	niet-roker	roker	niet-roker	
niet-hypertensief	30	55	24	60	169
hypertensief	10	5	6	10	31
totaal	40	60	30	70	200

Bereken:

1. de kans om iemand toevallig te selecteren in deze groep die man is, rookt en hypertensief is; [0.050]
2. de kans een man te selecteren, de kans een roker onder de mannen te selecteren, de kans om een hypertensie patiënt onder deze mannelijke rokers te selecteren; [0.50, 0.40, 0.25]
3. de odds op hypertensie bij rokers; [0.30]
4. de odds ratio voor de relatie tussen roken en hypertensie voor de totale groep en voor mannen en vrouwen afzonderlijk. [2.27, 3.67, 1.50]

Oplossing

Stel: M = man
R = roker
H = hypertensie

$$1.\ \mathbf{P(M\,en\,R\,en\,H)} = \frac{10}{200} = 0.050$$

$$2.\ \mathbf{P(M)} = \frac{40+60}{200} = 0.50$$

$$\mathbf{P(R\,|\,M)} = \frac{\mathbf{P(M\,en\,R)}}{\mathbf{P(M)}} = \frac{\frac{40}{200}}{\frac{100}{200}} = 0.40$$

$$\mathbf{P(H\,|\,M\,en\,R)} = \frac{\mathbf{P(H\,en\,M\,en\,R)}}{\mathbf{P(M\,en\,R)}} = \frac{\frac{10}{200}}{\frac{40}{200}} = 0.25$$

$$3.\ \text{Odds}(H|R) = \frac{P(H|R)}{1 - P(H|R)}$$

$$P(H|R) = \frac{P(H\,en\,R)}{P(R)} = \frac{\frac{16}{200}}{\frac{70}{200}} = \frac{16}{70}$$

$$\text{Odds}(H|R) = \frac{\frac{16}{70}}{1 - \frac{16}{70}} = \frac{16}{54} = 0.30$$

$$4.\ OR = \frac{\text{Odds}(H|R)}{\text{Odds}(H|\overline{R})} = \frac{\frac{16}{54}}{\frac{15}{115}} = 2.27$$

$$OR_{mannen} = \frac{\text{Odds}(H|M\,en\,R)}{\text{Odds}(H|M\,en\,\overline{R})} = \frac{\frac{10}{30}}{\frac{5}{55}} = 3.67$$

$$OR_{vrouwen} = \frac{\text{Odds}(H|\overline{M}\,en\,R)}{\text{Odds}(H|\overline{M}\,en\,\overline{R})} = \frac{\frac{6}{24}}{\frac{10}{60}} = 1.50$$

2.11 Pillenarsenaal tegen de vlakte

Een wat oudere patiënt heeft per ongeluk zijn ganse pillenarsenaal tegen de vlakte gegooid. Nadat hij zowat alles op de tast opnieuw heeft samengeraapt in twee dozen, is de inhoud van de dozen als volgt: doos A bevat 3 aspirines, 4 slaappillen en 5 tranquillizers; doos B bevat 5 aspirines, 6 slaappillen en 7 tranquillizers. ‘s Middags neemt hij lukraak uit elk doosje één pil:

a) wat is de kans dat beide geneesmiddelen slaappillen zijn? (gegarandeerd middagdutje!). [0.11]

b) wat is de waarschijnlijkheid dat beide geneesmiddelen van dezelfde categorie zijn? [0.34]

Oplossing

doos A: 3 aspirines, 4 slaappillen en 5 tranquillizers
doos B: 5 aspirines, 6 slaappillen en 7 tranquillizers

$$\textbf{a. P(beide slaappillen)} = \frac{4}{3+4+5} \times \frac{6}{5+6+7} = 0.11$$

$$\textbf{b. P(zelfde categorie)} = \textbf{som van de kansen}$$

$$= \frac{3}{12} \times \frac{5}{18} + \frac{4}{12} \times \frac{6}{18} + \frac{5}{12} \times \frac{7}{18} \quad \textbf{(a,s,t)}$$

$$= \frac{5}{72} + \frac{1}{9} + \frac{35}{216} = \frac{15+24+35}{216} = \frac{74}{216} = \frac{37}{108} = 0.342$$

$$\textbf{P} = 0.34$$

2.12 Operatiekwartier orthopedie

Het operatiekwartier van de afdeling orthopedie plant vier patiënten voor deze voormiddag. De orthopedische chirurg besluit te starten met de eenvoudigste ingreep en in klimmende volgorde van ernst de andere operaties aan te pakken. Wat is de kans dat de vier aangereden patiënten reeds toevallig in de goede volgorde staan? [0.042]

Oplossing $P = \frac{1}{4!} = \frac{1}{4 \times 3 \times 2 \times 1} = \frac{1}{24} = 0.042$

2.13 Aandeel op de beurs

Op de beurs stijgt of daalt de waarde van een aandeel dagelijks met één punt met respectievelijke kansen 1/3 en 2/3.

a. Wat is de kans dat een aandeel na vier dagen tot op zijn oorspronkelijke waarde terugvalt? [0.30]

b. Als na vier dagen het aandeel inderdaad terug zijn startwaarde bezit, wat is de kans dat het aandeel de eerste dag in waarde steeg? [0.50]

Oplossing

Stel: D = aandeel daalt

S = aandeel stijgt

N = aandeel heeft na 4 dagen terug zijn startwaarde

SSSS	DSSS
SSSD	DSSD
SSDS	DSDS
SSDD	DSDD
SDSS	DDSS
SDSD	DDSD
SDDS	DDDS
SDDD	DDDD

a. P(N) = P(SSDD) + P(SDSD) + P(SDDS) + P(DSSD) + P(DSDS) + P(DDSS)

mutueel exclusief

onafhankelijk

$$\mathbf{P(SSDD) = P(S)P(S)P(D)P(D) = \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3}}$$

$$\mathbf{Analoog:\quad P(N) = 6 \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{8}{27} = 0.30}$$

$$\mathbf{b.\quad P(S_1 \mid N) = \frac{P(S_1\, en\, N)}{P(N)}}$$

$$\mathbf{P(S_1\, en\, N) = P(SSDD) + P(SDSD) + P(SDDS) = 3 \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{4}{27}}$$

$$\mathbf{Derhalve:\quad P(S_1 \mid N) = \frac{\frac{4}{27}}{\frac{8}{27}} = 0.50}$$

2.14 Hazenlip

Algemeen is het voorkomen van baby's geboren met een "gespleten lip" 10 op 10000. De afwijking komt vaker voor bij jongens dan bij meisjes, stel 15/10000 bij jongens versus 5/10000 bij meisjes. Neem aan dat er tijdens de observatieperiode evenveel jongens als meisjes geboren worden. Wat is dan de kans dat een baby geboren met een gespleten lip een jongen is? En wat is de kans dat het een meisje is? [0.75, 0.25]

Oplossing

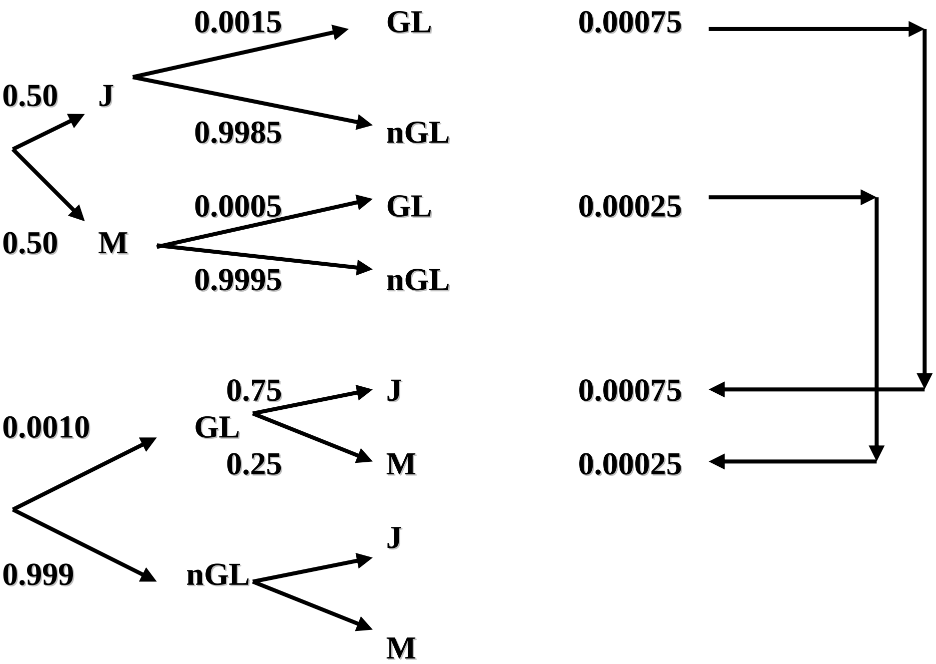
Stel: **GL = gespleten lip; J = jongen**

BAYES:

$$\mathbf{P(J \mid GL)} = \frac{\mathbf{P(J) \times P(GL \mid J)}}{\mathbf{P(GL)}} = \frac{\mathbf{0.5 \times 0.0015}}{\mathbf{0.001}} = \mathbf{0.75}$$

$$\Rightarrow \quad \mathbf{P(\bar{J} \mid GL) = 1 - 0.75 = 0.25}$$

KANSBOOM



2.15 Kans op ziek zijn

De kans op een bepaald symptoom is 0.20 in een bepaalde populatie. De kans op ziekte is 0.25. Tevens is gekend dat de kans op ziekte bij de groep die het symptoom vertoont 0.90 is. Wat is dan de kans op symptomen bij de zieken? Wat is de kans op symptomen bij niet-zieken? [0.72, 0.027]

Oplossing

Stel: S = symptoom
D = ziekte

P(S) = 0.20 P(D) = 0.25 P(D|S) = 0.90

$$\mathbf{P(S \mid D)} = \frac{\mathbf{P(S) \times P(D \mid S)}}{\mathbf{P(D)}} = \frac{\mathbf{0.90 \times 0.20}}{\mathbf{0.25}} = \mathbf{0.72}$$

$$\mathbf{P(S \mid \overline{D})} = \frac{\mathbf{P(S) \times P(\overline{D} \mid S)}}{\mathbf{P(\overline{D})}} = \frac{\mathbf{0.10 \times 0.20}}{\mathbf{0.75}} = \mathbf{0.027}$$

	ziek	**niet ziek**	**totaal**
	D	$\overline{D}$	
symptoom (S)	180	20	200
geen symptoom ($\overline{S}$)	70	730	800
	250	750	1000

$$\mathbf{P(S \mid D)} = \frac{\mathbf{180}}{\mathbf{250}} = \mathbf{0.72}$$

$$\mathbf{P(S \mid \overline{D})} = \frac{\mathbf{20}}{\mathbf{750}} = \mathbf{0.027}$$

Oef 3. Beschrijvende statistiek

3.1 Jobstress

Een arbeidsgeneeskundige dienst gaat bij 103 werknemers tussen 20 en 64 jaar aan de hand van een uitgebreide vragenlijst na welke mate van "jobstress" zij rapporteren. Gedurende het daaropvolgende jaar wordt het ziekteverzuim gevolgd bij de uiterste decielgroepen op de gemeten "jobstress-schaal". Volgende gegevens worden genoteerd:

- **ID** = identificatienummer
- **SEX** = geslachtsgroep (1 = man; 2 = vrouw)
- **AGE** = leeftijd (in jaren)
- $\mathbf{d_i}$ = decielgroep op de jobstress-schaal (voor i:1->10)
- **OBS** = totale observatieduur
- **XPER** = het aantal ziekteperiodes
- **AFW** = het totaal aantal afwezige kalenderdagen wegens ziekte
- **ABSPER** = het absenteïsme percentage = 100 * (AFW / OBS)

ID	SEX	AGE	d_i	OBS	XPER	AFW	ABSPER
001	1	45	10	365	1	14	3.8
003	1	28	1	366	0	0	0.0
007	2	39	10	200	1	25	12.5
008	1	56	1	155	0	0	0.0
012	1	41	1	365	1	7	1.9
017	2	61	10	280	2	21	7.5
025	1	37	10	365	5	30	8.2
028	1	40	1	366	1	7	1.9
038	2	24	1	366	0	0	0.0
045	1	36	10	365	7	200	54.8
052	1	58	10	365	2	30	8.2
053	1	28	1	210	3	21	10.0
060	1	24	1	365	2	14	3.8
065	2	24	10	366	2	20	5.5
067	2	32	1	100	0	0	0.0
078	1	51	10	365	1	7	1.9
082	1	46	10	365	2	14	3.8
090	1	34	1	308	3	10	3.2
097	2	38	1	300	5	57	19.0
100	2	40	1	365	0	0	0.0
103	1	50	10	365	1	3	0.8

a) Bereken het gemiddeld aantal ziekteperiodes. Welke opmerkingen kan je hierbij formuleren? [1.9]

b) Bepaal voor beide onderscheiden decielgroepen het gemiddelde absenteïsme percentage alsook de modus en de mediaan. Bereken tevens voor deze variabele de range, de interkwartiel range, de standaardafwijking en de variatiecoëfficiënt in beide groepen.

c) Schets een "box-and-whisker"-plot voor deze variabele.

d) Bereken de gemiddelde leeftijd in beide decielgroepen. [35.0, 44.7]
Formuleer in het kort uw besluit.

Oplossing (zie ook SPSS)

a. gemiddeld aantal ziekteperiodes (Zper):

$$\frac{\sum \text{ZPer}}{n} = \frac{39}{21} = 1.9$$

b. voor beide decielgroepen:
gemiddelde absenteïsme percentage en de mediaan.

Deciel 1 Gemiddelde

Id	Sex	Age	Dec	Obs	Zper	Dafw	Abs Per
3	1	28	1	366	0	0	0.0
8	1	56	1	155	0	0	0.0
12	1	41	1	365	1	7	1.9
28	1	40	1	366	1	7	1.9
38	2	24	1	366	0	0	0.0
53	1	28	1	210	3	21	10.0
60	1	24	1	365	2	14	3.8
67	2	32	1	100	0	0	0.0
90	1	34	1	308	3	10	3.2
97	2	38	1	300	5	57	19.0
100	2	40	1	365	0	0	0.0
		385				som =	39.8

n=11 gemiddelde = 3.62%

Deciel 10 Gemiddelde

Id	Sex	Age	Dec	Obs	Zper	Dafw	AbsPer
1	1	45	10	365	1	14	3.8
7	2	39	10	200	1	25	12.5
17	2	61	10	280	2	21	7.5
25	1	37	10	365	5	30	8.2
45	1	36	10	365	7	200	54.8
52	1	58	10	365	2	30	8.2
65	2	24	10	366	3	20	5.5
78	1	51	10	365	1	7	1.9
82	1	46	10	365	2	14	3.8
103	1	50	10	365	1	3	0.8
		447				som =	107.0

n=10 gemiddelde = 10.7%

Mediaan en kwartielen

Dec 1 Dec 10

1. Waarnemingen sorteren (n=11) (n=10)

2. Rangwaarde bepalen

			(n=11)	(n=10)
mediaan :	$\text{rang}\{\frac{n}{2}+0.5\}$	$\Rightarrow$	**rang 6**	**rang 5.5**
eerste kwartiel Q1:	$\text{rang}\{\frac{n}{4}+0.5\}$	$\Rightarrow$	**rang 3.25**	**rang 3**
derde kwartiel Q3:	$\text{rang}\{\frac{3n}{4}+0.5\}$	$\Rightarrow$	**rang 8.75**	**rang 8**

3. Afronding

		(n=11)	(n=10)
mediaan:	→	rang 6	rang 5.5
eerste kwartiel Q1:	→	rang 3	rang 3
derde kwartiel Q3:	→	rang 9	rang 8

	Dec 1	Dec 10		
	0.0	**0.8**		
	0.0	**1.9**		
Q1	**0.0 %**	**3.8 % Q1**		
	0.0	**3.8**		
	0.0	**5.5**	**6.5%**	**med**
med	**1.9 %**	**7.5**		
	1.9	**8.2**		
	3.2	**8.2 % Q3**		
Q3	**3.8 %**	**12.5**		
	10.0	**54.8**		
	19.0			
	n = 11	**n = 10**		

	Deciel 1	**Deciel 10**
Mediaan	1.9%	6.5%
Q1	0.0%	3.8%
Q3	3.8%	8.2%
Range (in %)	[0.0 – 19.0]	[0.8 – 54.8]

Bereken tevens voor AbsPer de range, de interkwartiel range, de standaardafwijking en de variatiecoëfficiënt in beide groepen.

Deciel 1	AbsPer	$AbsPer^2$	(AbsPer-Gemidd)	$(AbsP\text{-}Gem)^2$
	0.0	0.0	-3.62	13.09
	0.0	0.0	-3.62	13.09
	1.9	3.61	-1.72	2.95
	1.9	3.61	-1.72	2.95
	0.0	0.0	-3.62	13.09
	10.0	100.0	6.38	40.73
	3.8	14.44	.18	.03
	0.0	0.0	-3.62	13.09
	3.2	10.24	-0.42	0.17
	19.0	361.0	15.38	236.60
	0.0	0.0	-3.62	13.09
Som	39.8	492.9	0.0	348.9
Gemidd.	3.6%			34.89 (var)
SD		5.91%		

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

$$\mathbf{s} = \sqrt{\frac{348.90}{10}} = \mathbf{5.91\%}$$

$$\mathbf{s} = \sqrt{\frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}}$$

$$\mathbf{s} = \sqrt{\frac{492.9 - \frac{(39.8)^2}{11}}{10}} = \mathbf{5.91\%}$$

$$\mathbf{cv} = \frac{5.91}{3.6} \times 100 = \mathbf{164.17\%}$$

Bereken tevens voor AbsPer de range, de interkwartiel range, de standaardafwijking_en de variatiecoëfficiënt in beide groepen.

Deciel 10	AbsPer	$AbsPer^2$	(AbsPer-Gemidd)	$(AbsP\text{-}Gem)^2$	
	3.8	14.44	-6.9	47.61	
	12.5	156.25	1.8	3.24	
	7.5	56.25	-3.2	10.24	
	8.2	67.24	-2.5	6.25	
	54.8	3003.04	44.1	1944.81	
	8.2	67.24	-2.5	6.25	
	5.5	30.25	-5.2	27.04	
	1.9	3.61	-8.8	77.44	
	3.8	14.44	-6.9	47.61	
	.8	0.64	-9.9	98.01	
som	107.0	3413.4	0.0	2268.5	
gemidd.	10.7%			252.06	(var)
		SD 15.88%			

$$s = \sqrt{\frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}}$$

$$s = \sqrt{\frac{3413.4 - \frac{107.0^2}{10}}{9}} = 15.88\%$$

$$cv = \frac{15.88}{10.7} \times 100 = 148.4\%$$

Overzicht Resultaten Absenteïsme Percentage

	deciel 1	deciel 10
gemiddelde	3.6%	10.7%
st.deviatie	5.91%	15.88%
variatiecoëfficiënt	164.2%	148.4%
mediaan	1.9%	6.5%
interkwartiel range	3.8%	4.4%
range	19.0%	54.0%
modus	0%	3.8 en 8.2%

d. Gemiddelde leeftijd

deciel 1:	35.0 jaar
deciel 10:	44.7 jaar

e. Besluiten

- scheve distributie voor absenteïsme percentage
- vermoedelijk verschil tussen deciel1 en deciel10
- verschil mede verklaard door verschil in leeftijd? (confounding variable)

3.2 Acute leukemie

In een klinische studie die als doel heeft een nieuwe medicatie voor patiënten met acute leukemie te evalueren, werden volgende remissietijden (in weken) genoteerd:

4 22 5 72 8 4 11 2 16 46 8 2 1 5 11 12 8 1 12 3 15 23 8

a) Bereken het gemiddelde, de mediaan, de modus, de range, de steekproefvariantie, de standaardafwijking, de variatiecoëfficiënt en de kwartielen
b) Maak de frequentietabel op, teken een histogram en schets zowel een 'box-and-whisker' als een 'stem-and-leaf' plot voor deze gegevens.
c) Hoe verhoudt het gemiddelde zich t.o.v. de mediaan? Wat leert dit ons over de symmetrie van de distributie?

Oplossing (zie ook SPSS)

a. Gemiddelde:

na rangschikking:

1 1 2 2 3 4 4 5 5 8 8 8 8 11 11 12 12 15 16 22 23 46 72

n = 23
mediaan: 8w (rang 12) modus: 8 w
Q1: 4w (rang 6.25 $\longrightarrow$ 6)
Q3: 15w (rang 17.75 $\longrightarrow$ 18)

IQR: 15 - 4 = 11w [4w, 15w]
Range: 72 - 1 = 71w [1w, 72w]

$$\text{Var} = \frac{9681 - \frac{(299)^2}{23}}{22} = 263.36$$

$$\text{SD} = \sqrt{263.36} = 16.23 \text{ weken}$$

$$\text{CV} = \frac{16.23}{13.0} \text{x} 100 = 124.8\%$$

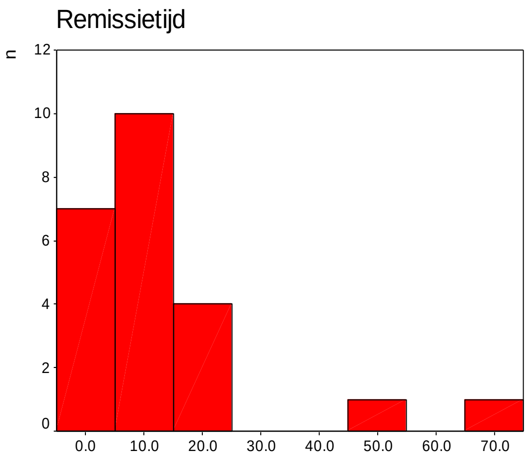
b. Maak de frequentietabel op, teken een histogram en schets een 'box-and-whisker' en ook een 'stem-and-leaf' plot.

x	f	%	Cum%
1	2	8.7	8.7
2	2	8.7	17.4
3	1	4.3	21.7
4	2	8.7	30.4
5	2	8.7	39.1
8	4	17.4	56.5
11	2	8.7	65.2
12	2	8.7	73.9
15	1	4.3	78.3
16	1	4.3	82.6
22	1	4.3	87.0
23	1	4.3	91.3
46	1	4.3	95.7
72	1	4.3	100.0

Stem & Leaf

0	1122344558888
1	112256
2	23
3	
4	6
5	
6	
7	2
	eenheid = 1w

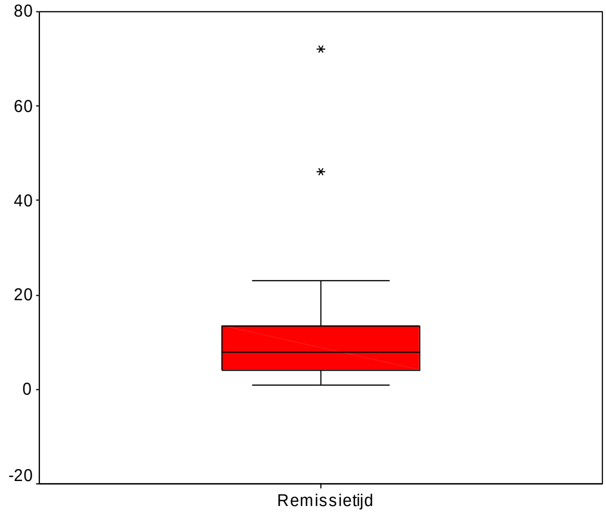
Maak de frequentietabel op, teken een histogram en schets een 'box-and-whisker' plot.



Bovenstaand histogram werd bekomen in SPSS (SPSS kiest voor 8 klassen met breedte 10; gezien de range is dat tenslotte ook de meest logische klassenindeling).

Maak de frequentietabel op, teken een histogram en schets een 'box-and-whisker' plot.

n = 23
laagste = 1w; hoogste = 72w
mediaan = 8w; Q1 = 4w; Q3 = 15w; IQR = 11w


80
60
40
20
0
-20
Remissietijd

3.3 Opname op psychiatrie

Onderstaande tabel geeft de opnameduur (in dagen) voor elf patiënten die voor diagnose en evaluatie opgenomen werden in een nieuw opgerichte psychiatrische afdeling:

Patiënt	Verblijf (dagen)
1	29
2	14
3	11
4	24
5	14
6	14
7	28
8	14
9	18
10	22
11	14

Bereken voor de bij deze patiënten bekomen steekproef de gemiddelde verblijfsduur, mediaan, modus, range, variantie, standaardafwijking, variatiecoëfficiënt en de kwartielen.

Oplossing (zie ook SPSS)

	V-duur	V-duur^2	
	29	841	
	14	196	
	11	121	
	24	576	
	14	196	
	14	196	
	28	784	
	14	196	
	18	324	
	22	484	
	14	196	
Som	202	4110	
Gemidd.	18.4	40.05	(var)
		6.33	(SD)

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{1}{11} 202 = 18.4 \quad \text{dagen}$$

$$s^2 = \frac{4110 - \frac{202^2}{11}}{10} = 40.06$$

$$s = \sqrt{40.06} = 6.33 \quad \text{dagen}$$

$$cv = \frac{6.33}{18.4} x100 = 34.5\%$$

rang	V-duur		
1	11		
2	14		
3	***14***	***Q1***	rang 3.25
4	14		
5	14		
6	***14***	***mediaan***	rang 6
7	18		
8	22		
9	***24***	***Q3***	rang 8.75
10	28		
11	29		

modus = 14d; range = 29 - 11 = 18d;
IQR = 24 - 14 = 10d

Mediaan:	14d
Q1	14d
Q3	24d
Range	[11 – 29d]

3.4 Dossiers huisartsenpraktijk

De volgende gegevens zijn afkomstig uit medische dossiers van 32 mannen tussen 45 en 65 jaar, die op toevalsbasis werden geselecteerd in een huisartsenpraktijk:

Roken	BMI	SBP	DBP
1	2	152	71
1	2	124	77
1	2	105	61
2	2	146	96
2	3	167	120
1	1	156	94
1	1	133	89
1	3	144	81
1	3	186	138
2	1	103	75
1	2	98	67
1	2	131	87
2	2	155	99
1	2	163	90
2	2	136	74
2	2	129	66
1	2	170	112
1	3	160	85
1	2	142	86
2	2	142	82
1	2	115	76
2	2	201	119
1	2	129	83
1	2	158	92
1	3	113	70
2	1	149	84
2	1	157	98
2	3	132	78
1	3	146	88
1	2	175	103
2	2	142	79
1	2	118	68

* Roken: 1 = ja / 2 = neen
* BMI (Body Mass Index) = gewicht in kg / (grootte in m)^2
  1 = minder dan 25 kg/m^2
  2 = 25 tot 30 kg/m^2
  3 = meer dan 30 kg/m^2
* SBP = Systolic Blood Pressure (mmHg)
  DBP = Diastolic Blood Pressure (mmHg)

a) Maak voor systolische en diastolische bloeddruk box-and-whisker plots. Zijn de verdelingen van deze variabelen scheef? Zijn er belangrijke uitschieters?
b) Schets een taartdiagram voor de BMI.
c) Maak een kruistabel tussen BMI en roken. Wat is je oordeel omtrent de relatie tussen BMI en roken?
d) Teken een schatterdiagram voor systolische en diastolische bloeddrukken.

Herhaal punt c, maar vervang de frequenties in de cellen van de kruistabel door de mediaan van zowel SBP als DBP voor elk van deze subgroepen. Zijn de bloeddrukken ergens verwant met BMI en/of roken?

Oplossing (zie ook SPSS)

a. Maak voor systolische en diastolische bloeddruk box-and-whisker plots. Zijn de verdelingen van deze variabelen scheef? Zijn er uitschieters?

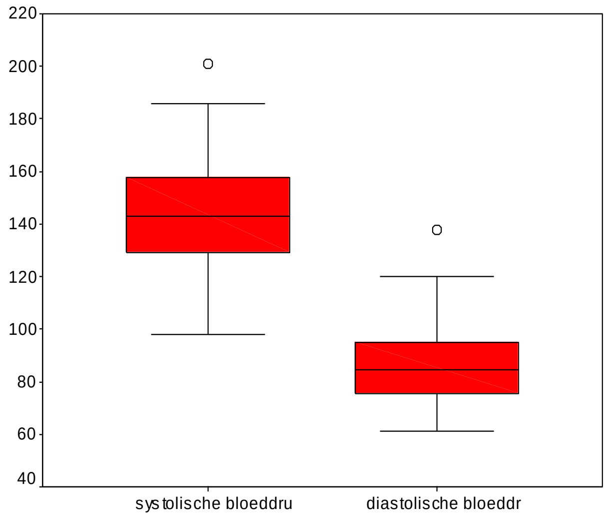
rangorde tabel (gepaardheid is verloren – voor een correlatietoepassing zou dit een moordend verkeerde handelswijze zijn!)

Rng	S	D
1	98	61
2	103	66
3	105	67
4	113	68
5	115	70
6	118	71
7	124	74
8	*129*	*75*
9	*129*	*76*
10	131	77
11	132	78
12	133	79
13	136	81
14	142	82
15	142	83
16	*142*	*84*
17	*144*	*85*
18	146	86
19	146	87
20	149	88
21	152	89
22	155	90
23	156	92
24	*157*	*94*
25	*158*	*96*
26	160	98
27	163	99
28	167	103
29	170	112
30	175	119
31	186	120
32	201	138

$$\textbf{Q1} \quad \frac{32}{4} + \frac{1}{2} = 8.5^e \text{ rang}$$

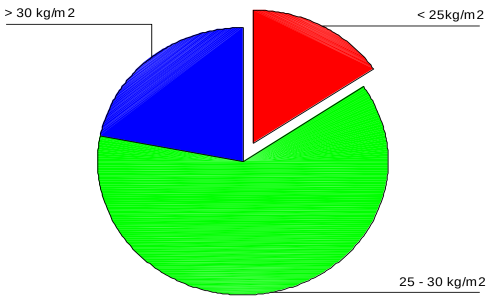
$$\textbf{Med} \quad \frac{32}{2} + \frac{1}{2} = 16.5^e \text{ rang}$$

$$\textbf{Q3} \quad \frac{3 \times 32}{4} + \frac{1}{2} = 24.5^e \text{ rang}$$



	SYS	DIAS (mmHg)
min	98	61
max	201	138
med	143	84.5
Q1	129	75.5
Q3	157.5	95

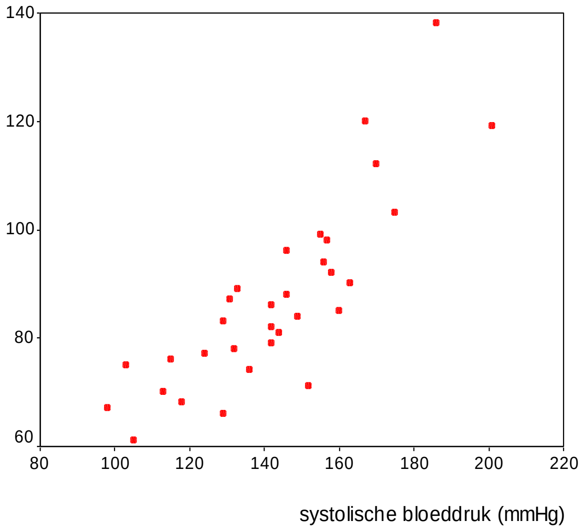
b.



c. Maak een kruistabel tussen BMI en roken. Wat is je oordeel over de relatie tussen BMI en roken?

BMI	Roken	Niet-Roken
1 (<25)	2	3
2 (25-30)	13	7
3 (>= 30)	5	2

d. Teken een scatterdiagram voor systolische en diastolische bloeddrukken



e. Herhaal punt c, maar vervang de frequenties in de cellen van de kruistabel door de mediaan van zowel SBP als DBP voor elk van deze subgroepen. Zijn de bloeddrukken ergens verwant met BMI en/of roken?

BMI			Roken	Niet-Roken
1	(<25)	S	144.5	149
		D	91.5	84
2	(25-30)	S	131	142
		D	83	82
3	(>= 30)	S	146	149
		D	85	99

Eventueel verwantschap is niet onmiddellijk uit deze tabellen af te lezen – meer geavanceerde statistische testen zullen hier uitsluitsel moeten geven.

Oef 4. Associatiematen

4.1 Bloeddrukmetingen

Bereken voor de gegevens van de vorige oefening (bloeddrukmetingen in een huisartsenpraktijk) de Pearson correlatiecoëfficiënt r tussen systolische en diastolische bloeddruk voor de deelgroep met BMI klasse gelijk aan 1. Vermeld bij het resultaat eveneens de covariantie tussen beide variabelen voor deze deelgroep.

[r = 0.87]

Oplossing (zie ook SPSS)

	Sys x	Dia y	x^2	y^2	xy
	156	94	24336	8836	14664
	133	89	17689	7921	11837
	103	75	10609	5625	7725
	149	84	22201	7056	12516
	157	98	24649	6904	15386
Σ	698	440	99484	39042	62128

n = 5

Covariantie

$$\mathrm{cov}(x,y) = s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

$$\mathrm{cov}(x,y) = s_{xy} = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{n-1}$$

$$\mathrm{cov}(x,y) = \frac{704}{4} = 176 \mathrm{mmHg}^2$$

Correlatie

$$r = \frac{62128 - \frac{698 \times 440}{5}}{\sqrt{99484 - \frac{(698)^2}{5}}\sqrt{39042 - \frac{(440)^2}{5}}}$$

$$r = \frac{62128 - 61424}{\sqrt{2043.2}\sqrt{322}} = \frac{704}{\sqrt{657910.4}} = \frac{704}{811.17} = 0.8679$$

r = 0.87

4.2 De komst van de ooievaar in de Vogezen

In een stadje in de Vogezen telde men in een bepaald jaar maandelijks het aantal ooievaars en het aantal geboorten. Dit levert volgende tabel op:

maand	ooievaars	geboorten
januari	2	8
februari	4	8
maart	10	12
april	16	14
mei	18	18
juni	18	16
juli	12	15
augustus	15	12
september	5	14
oktober	3	9
november	2	6
december	2	7

Bereken de Pearson correlatiecoëfficiënt r, [r = 0.87]
alsook de Spearman correlatiecoëfficiënt r_S. [r_s = 0.91]

Oplossing (zie ook SPSS)

maand	X	X^2	Y	Y^2	XY
jan	2	4	8	64	16
feb	4	16	8	64	32
mrt	10	100	12	144	120
apr	16	256	14	196	224
mei	18	324	18	324	324
jun	18	324	16	256	288
jul	12	144	15	225	180
aug	15	225	12	144	180
sep	5	25	14	196	70
okt	3	9	9	81	27
nov	2	4	6	36	12
dec	2	4	7	49	14
som	107	1435	139	1779	1487

$$r = \frac{\mathrm{cov}(x,y)}{\sqrt{\mathrm{var}(x)\,\mathrm{var}(y)}} = \frac{\mathrm{cov}(x,y)}{s_x s_y} = \frac{s_{xy}}{s_x s_y}$$

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad r \in [-1,+1]$$

rekenformule:

$$\mathbf{r} = \frac{\sum \mathbf{xy} - \frac{\left(\sum \mathbf{x}\right)\left(\sum \mathbf{y}\right)}{\mathbf{n}}}{\sqrt{\sum \mathbf{x}^2 - \frac{\left(\sum \mathbf{x}\right)^2}{\mathbf{n}}}\sqrt{\sum \mathbf{y}^2 - \frac{\left(\sum \mathbf{y}\right)^2}{\mathbf{n}}}}$$

$$\mathbf{r} = \frac{\mathbf{1487} - \frac{\mathbf{107 \times 139}}{\mathbf{12}}}{\sqrt{\mathbf{1435} - \frac{(\mathbf{107})^2}{\mathbf{12}}}\sqrt{\mathbf{1779} - \frac{(\mathbf{139})^2}{\mathbf{12}}}}$$

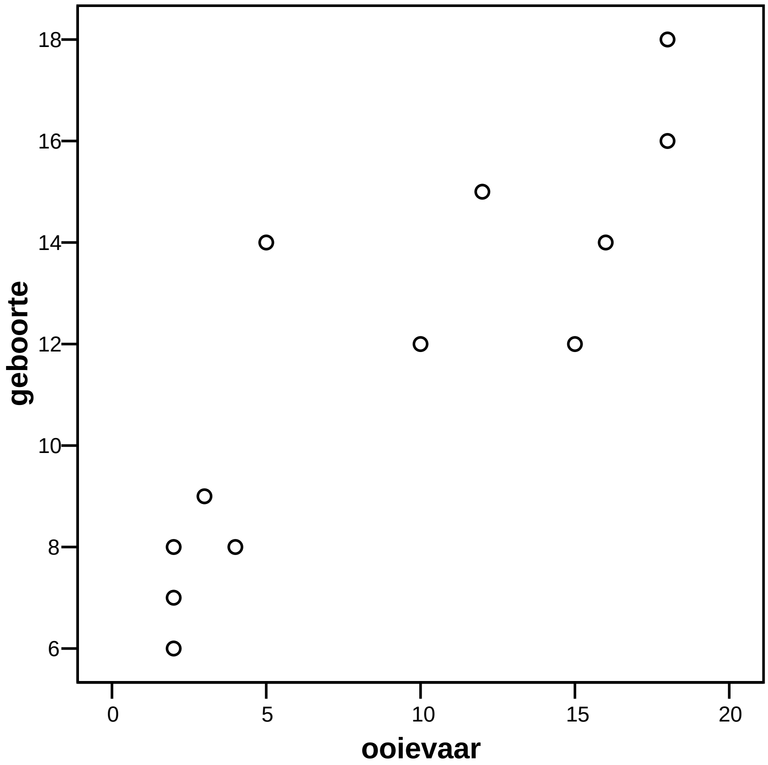
$$\mathbf{r} = \frac{\mathbf{1487 - 1239.42}}{\sqrt{\mathbf{480.92}}\sqrt{\mathbf{168.92}}} = \frac{\mathbf{247.58}}{\sqrt{\mathbf{81234.839}}}$$

$$= \frac{\mathbf{247.6}}{\mathbf{285.017}} = \mathbf{0.8687}$$

r = 0.87

Spearman rangcorrelatie r_S:

maand	ooie-vaars	rang	geboor-tes	rang	$\lvert d_i \rvert$	d_i^2
jan	2	2	8	3.5	1.5	2.25
feb	4	5	8	3.5	1.5	2.25
mrt	10	7	12	6.5	0.5	0.25
apr	16	10	14	8.5	1.5	2.25
mei	18	11.5	18	12	0.5	0.25
jun	18	11.5	16	11	0.5	0.25
jul	12	8	15	10	2	4
aug	15	9	12	6.5	2.5	6.25
sep	5	6	14	8.5	2.5	6.25
okt	3	4	9	5	1	1
nov	2	2	6	1	1	1
dec	2	2	7	2	0	0
					som:	26



$$r_S = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \qquad r_S \in [-1, +1]$$

$$r_S = 1 - \frac{6 \times 26}{12 \times 143} = 1 - \frac{156}{1716}$$

r_S = 1 - 0.0909 = 0.90909

r_S = 0.91

Oef 5. Verdelingsfuncties

5.1 Slaagkansen operatie

Stel dat een bepaalde operatie slaagt in 80% van alle gevallen. Bereken de kansverdeling op een geslaagde ingreep als vier patiënten zich aanbieden.

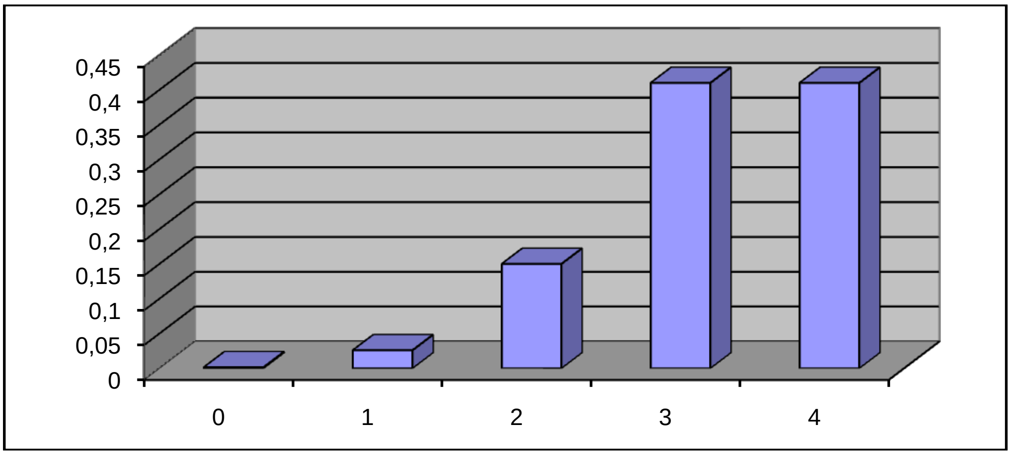
Oplossing

Kansenpatroon (vier patiënten)

N = niet geslaagd
S = geslaagd $2^4 = 16$

NNNN NNNS NNSN NNSS
NSNN NSNS NSSN NSSS
SNNN SNNS SNSN SNSS
SSNN SSNS SSSN *SSSS*

S	P	P
0	$(0.2)^4 = 0.0016$	0.0016
1	$4{\times}(0.2)^3(0.8) = 0.0256$	0.026
2	$6{\times}(0.2)^2(0.8)^2 = 0.1536$	0.15
3	$4{\times}(0.2)(0.8)^3 = 0.4096$	0.41
4	$(0.8)^4 = 0.4096$	0.41
		$\Sigma = 1$

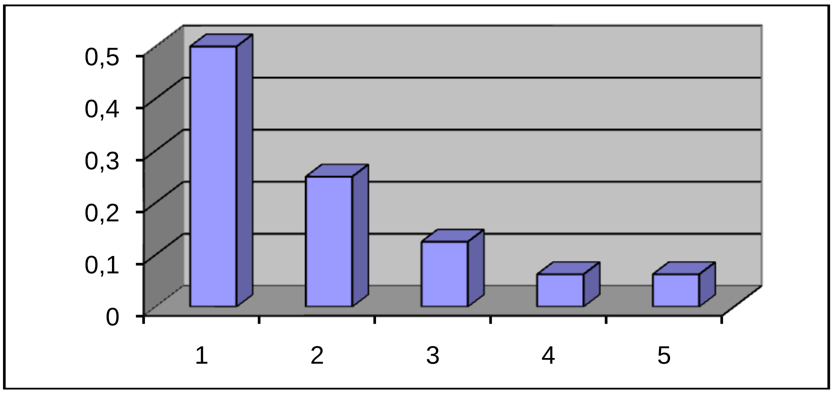


5.2 Met geld gooien

Een regelmatig muntstuk wordt herhaaldelijk opgegooid tot het kruis toont of tot het vijfmaal is opgeworpen. Noem T de variabele die aangeeft hoeveel worpen nodig zijn om ofwel een kruis te bekomen ofwel tot het maximaal aantal worpen te komen. Bereken de kansverdeling van T en de relatieve frequentieverdeling.

Oplossing

T	P
1	0.5
2	0.25
3	0.125
4	0.0625
5	0.0625
	$\Sigma = 1$



Oef 6. Normaalverdeling

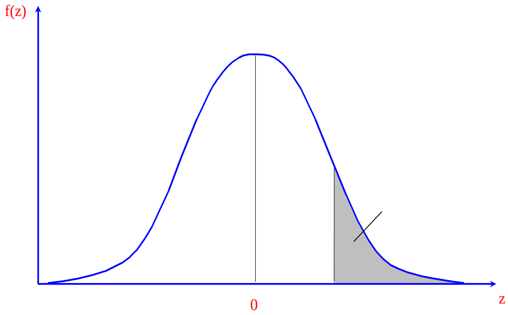
6.1 z-waarden

Met betrekking tot de standaardnormale verdeling, wat is de kans op het bekomen van een z-waarde van:

a)	tenminste 1.35	[0.088]
b)	tenminste -0.85	[0.80]
c)	minder dan 1.76	[0.96]
d)	tussen -0.85 en 1.24	[0.69]
e)	tussen -1.96 en 1.96	[0.95]

Oplossing

Excel [+normsdist()]
R [pnorm(z)]



a. $P(Z \geq 1.35) = 0.088$

b. $P(Z \geq -0.85) = 0.802$

c. $P(Z < 1.76) = 0.961$

d. $P(-0.85 < Z < 1.24)$
$= P(Z < 1.24) - P(Z < -0.85)$
$= P(Z < 1.24) - P(Z \geq 0.85)$
$= 0.8925 - 0.1977 = 0.695$

e. $P(-1.96 < Z < 1.96)$
$= P(Z < 1.96) - P(Z < -1.96)$
$= P(Z < 1.96) - P(Z \geq 1.96)$
$= 0.975 - 0.025 = 0.950$

Welke is de z-waarde zodanig dat de kans op het bekomen van een nog grotere z-waarde gelijk is aan:

a)	**0.05**	**[1.64]**
b)	**0.02**	**[2.05]**
c)	**0.10**	**[1.28]**

6.2 Studieresultaten

Stel dat de studieresultaten voor het vak 'medische statistiek' in de eerste kandidatuur normaalverdeeld zijn met een gemiddelde van 74 en een standaarddeviatie van 10 punten;

a) Stel dat een student 88 punten behaalde. Welk percentage van zijn studiegenoten heeft een nog beter resultaat? [8.1%]

b) Stel dat de lesgever het aantal onderscheidingen wil beperken tot niet meer dan 20%. Welk puntenaantal moet een student minimaal behalen om een onderscheiding toegekend te kunnen worden? [82.4]

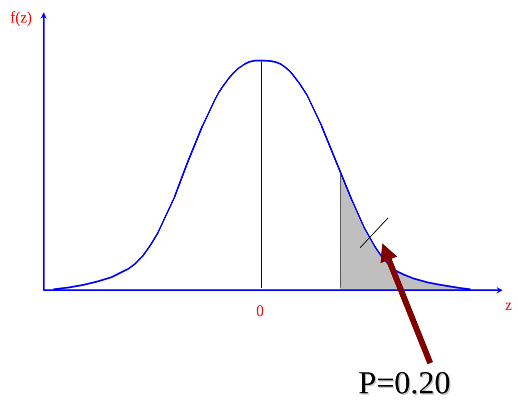
Oplossing

$\mu = 74$; $\sigma = 10$ ~ N(74,100)

➔ **N(0,1);** $Z = \frac{X - \mu}{\sigma}$

a.

$$P(X > 88) = P(Z > \frac{88 - 74}{10})$$
$$= P(Z > 1.4)$$
$$= 0.0808$$



Oplossing: 8.1 %

b. 20% ➔ z = 0.842

$Z = \frac{X - \mu}{\sigma}$ ➔ $x = z.\,\sigma + \mu$

$= 0.842 \times 10 + 74$
$= 8.42 + 74$
$= 82.42$

Oplossing: 82.4 punten

6.3 Diabetici

In de populatie van diabetici blijkt de gemiddelde polsslag 81 slagen per minuut te zijn met een standaarddeviatie van 3 slagen per minuut. In de veronderstelling dat de polsslag een normale verdeling volgt:

a) wat is de kans dat een willekeurig gekozen patiënt een polsslag heeft die lager is dan 84 slagen per minuut? [0.84]

b) Wat is de proportie patiënten die een polsslag hebben tussen 72 en 90 slagen per minuut? [99.7%]

c) Als men 100 patiënten selecteert om deel te nemen aan een klinische studie, van hoeveel kan men verwachten dat ze een polsslag hebben van meer dan 85 slagen per minuut? [9]

Oplossing

$\mu = 81$; $\sigma = 3$ ~ N(81,9)

➔ N(0,1); $Z = \dfrac{X - \mu}{\sigma}$

a.

$$P(X < 84) = P(Z < \frac{84 - 81}{3})$$
$$= P(Z < 1)$$
$$= 0.8413 = 0.84$$

b.

$$P(72 < X < 90) = P(X < 90) - P(X < 72)$$
$$= P(Z < \frac{90 - 81}{3}) - P(Z < \frac{72 - 81}{3})$$
$$= P(Z < 3) - P(Z < -3)$$
$$= P(Z < 3) - P(Z \geq 3)$$
$$= 0.9987 - 0.0013 = 0.9974$$

<u>Oplossing</u>: 99.7%

c.

$$P(X > 85) = P(Z > \frac{85 - 81}{3})$$
$$= P(Z > 1.3333) = 0.0918$$

<u>Oplossing</u>: 9 patiënten

6.4 IQ-score

In een normale populatie kinderen is de IQ-score normaalverdeeld met een gemiddelde van 100. Een kind met IQ-score boven 115 wordt 'begaafd' genoemd. Een kind is 'superbegaafd' als deze IQ-score minstens 140 bedraagt. Studies hebben uitgewezen dat ongeveer 1% van de populatie hieraan voldoet.

a) Wat is de standaardafwijking van deze distributie? [17.2]
b) Welk percentage kinderen in de populatie is begaafd? [19.2%]
c) Welk percentage kinderen heeft een IQ-score van 70 of lager? [4.1%]

Oplossing

$\mu = 100; \sigma = ?$

$P(X \geq 140) = 0.01$ $\qquad Z = \frac{X - \mu}{\sigma}$

a.

$$P(X \geq 140) = 0.01$$
$$P(Z \geq \frac{140 - 100}{\sigma}) = 0.01$$
$$P(Z \geq \frac{40}{\sigma}) = 0.01$$
$$\Rightarrow 2.326 = \frac{40}{\sigma}$$
$$\sigma = \frac{40}{2.326} = 17.1969 = 17.2$$

b.

$$P(X > 115) = P(Z > \frac{115 - 100}{17.2})$$
$$= P(Z > 0.8722) = 0.1922$$

<u>Oplossing</u>: 19.2%

c.

$$P(X \leq 70) = P(Z \leq \frac{70 - 100}{17.2})$$
$$= P(Z > 1.7445) = 0.0406$$

<u>Oplossing</u>: 4.1%

6.5 Bloedstalen

Een laboratorium bepaalt van 1000 bloedstalen het cholesterolgehalte. Dit blijkt normaalverdeeld te zijn met gemiddelde gelijk aan 218.1 mg/dl en standaardafwijking gelijk aan 39.80 mg/dl.
Het laboratorium neemt zich voor om een controlebepaling te voorzien voor afwijkende cholesterolwaarden.
Waarnemingen lager dan 150 mg/dl en hoger dan 300 mg/dl worden als afwijkend gedefinieerd. De kostprijs van een herbepaling voor "lage waarden" bij de eerste bepaling is 130 Fr, de kostprijs voor "hoge waarden" is 150 Fr.
Welk budget moet het laboratorium voor herbepaling voorzien? [8720 BEF]

Oplossing

$\mu = 218.1$; $\sigma = 39.80$ (mg/dl) $Z = \dfrac{X - \mu}{\sigma}$

$$\text{a. } P(X > 300 mg/dl) = P(Z > \frac{300 - 218.1}{39.8})$$
$$= P(Z > 2.058) = 0.0197 = 0.020$$
$$\text{b. } P(X < 150 mg/dl) = P(Z < \frac{150 - 218.1}{39.8})$$
$$= P(Z < -1.71) = P(Z \geq 1.71) = 0.044$$

<u>Berekening op 1000 stalen</u>:

# > 300mg/dl ➔ 2% van 1000 = 20
kostprijs = 20 × 5 euro = 100 euro

# < 150mg/dl ➔ 4.4% van 1000 = 44
kostprijs = 44 × 4 euro = 176 euro

Verwachte budget: 276 euro

6.6 Verkouden

Uit medisch onderzoek blijkt dat normale personen gemiddeld tweemaal per jaar een verkoudheid krijgen. Stel dat de tijd tussen het krijgen van verkoudheden normaalverdeeld is met een gemiddelde van 160 dagen en een standaarddeviatie van 40 dagen.

a) Wat zijn de kansen dat een persoon respectievelijk 200 dagen of meer en 365 dagen of meer geen verkoudheid krijgt? [0.16; ≅0]

b) Wat is de kans op een verkoudheid binnen de 80 dagen na een vorige verkoudheid? [0.023]

μ = 160 d; σ = 40 d

$$Z = \frac{X - \mu}{\sigma}$$

a. $$P(F \geq 200) = P(Z \geq \frac{200 - 160}{40}) = P(Z \geq 1)$$

$$= 0.1587 = 0.16$$

$$P(F \geq 365) = P(Z \geq \frac{365 - 160}{40}) = P(Z \geq 5.125)$$

$$= 0.0000...0.16 \cong 0$$

b. $$P(F < 80 \mid F) = P(Z < \frac{80 - 160}{40}) = P(Z < -2)$$

$$= P(Z \geq 2) = 0.0228 = 0.023$$

(onafhanke lijk)

Oef 7. Binomiaalverdeling

7.1 Psychiatrisch ziekenhuis

In een afdeling van een psychiatrisch ziekenhuis verblijven vijf patiënten.

a) Alle vijf patiënten staan op de wachtlijst voor de namiddagconsultatie bij de psychiater. Op hoeveel wijzen kan deze groep geordend worden? [120]

b) Tevens wil men een therapiegroep van drie patiënten uit deze groep van vijf samenstellen. Hoeveel dergelijke groepjes van drie patiënten kan men samenstellen? [10]

Oplossing

a. n = 5

$$P_5 = 5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

b.

$$\mathbf{C_n^r = \frac{n!}{r!(n-r)!} = \binom{n}{r}}$$

$$\mathbf{C_5^3 = \frac{5!}{3!(5-3)!} = \binom{5}{3} = 10}$$

7.2 Oudere diabetici

Veronderstel dat bij oudere diabetici (65+) de kans op ernstig gezichtsverlies 0.35 is.

Wat is de kans dat van de vijf oudere diabetici die een oogarts raadplegen er:

a) exact drie een ernstig gezichtsverlies vertonen? [0.18]

b) minstens drie een ernstig gezichtsverlies vertonen? [0.24]

Oplossing

$$\mathbf{P(X = r) = P(r) = \binom{n}{r}\pi^{r}(1-\pi)^{n-r}}$$

n = 5 π = 0.35

r	$\binom{n}{r}$	P	
0	1	$1 \times 0.35^0 \times 0.65^5 = 0.116$	
1	5	$5 \times 0.35^1 \times 0.65^4 = 0.312$	
2	10	$10 \times 0.35^2 \times 0.65^3 = 0.336$	
3	10	$10 \times 0.35^3 \times 0.65^2 = 0.181$	$\left.\right\}$ = 0.235
4	5	$5 \times 0.35^4 \times 0.65^1 = 0.049$	
5	1	$1 \times 0.35^5 \times 0.65^0 = 0.005$	

$\Sigma = 1.00$

a. $P(X = 3) = 0.18$

b. minstens 3 ⇔ 3 of meer

$P(X \geq 3) = 0.24$

7.3 Miskraam

Indien 12% van alle zwangerschappen eindigen in een miskraam, wat is dan de kans dat meer dan de helft van een groep van zes zwangere vrouwen een miskraam zullen hebben?

[0.0025]

Oplossing

$$\mathbf{P(X = r) = P(r) = \binom{n}{r}\pi^{r}(1-\pi)^{n-r}}$$

n = 6 π = 0.12

r	$\binom{n}{r}$	P	
0	1	$1 \times 0.12^{0} \times 0.88^{6} = 0.464$	
1	6	$6 \times 0.12^{1} \times 0.88^{5} = 0.380$	
2	15	$15 \times 0.12^{2} \times 0.88^{4} = 0.130$	
3	20	$20 \times 0.12^{3} \times 0.88^{3} = 0.024$	
4	15	$15 \times 0.12^{4} \times 0.88^{2} = 0.00241$	$\left.\right\}$ = 0.002543
5	6	$6 \times 0.12^{5} \times 0.88^{1} = 0.000131$	
6	1	$1 \times 0.12^{6} \times 0.88^{0} = 0.0000030$	
		Σ = 1.00	

meer dan de helft: 4 of meer

P(X ≥ 4) = 0.0025

7.4 Hypertensie

Stel dat geobserveerd werd dat in een populatie 25% van de leden een systolische bloeddruk boven de waarde 150 mmHg heeft.
Op een ziekenhuisafdeling met zes patiënten vindt men bij vier patiënten een bloeddruk die hoger is dan 150 mmHg. Vormt dit een toevalssteekproef uit bovenvermelde populatie?

[0.038, N]

Oplossing

$$\mathbf{P(X = r) = P(r) = \binom{n}{r} \pi^{r} (1 - \pi)^{n-r}}$$

n = 6 π = 0.25

r	$\binom{n}{r}$	P
0	1	$1 \times 0.25^0 \times 0.75^6 = 0.178$
1	6	$6 \times 0.25^1 \times 0.75^5 = 0.356$
2	15	$15 \times 0.25^2 \times 0.75^4 = 0.297$
3	20	$20 \times 0.25^3 \times 0.75^3 = 0.132$
4	15	$15 \times 0.25^4 \times 0.75^2 = 0.0330$
5	6	$6 \times 0.25^5 \times 0.75^1 = 0.00439$
6	1	$1 \times 0.25^6 \times 0.75^0 = 0.000244$

$$\left.\begin{array}{l} r=4: 0.0330 \\ r=5: 0.00439 \\ r=6: 0.000244 \end{array}\right\} = 0.03760$$

Σ = 1.00

4 patiënten met verhoogde bloeddruk nog aanvaardbaar? ➔ kans op 4 of meer

P(X ≥ 4) = 0.038 (geen toevalssteekproef)

7.5 Bedden op spoedopname

Een afdeling spoedopname heeft voldoende bedden om aan 97% van de dagelijkse belasting te voldoen. Stel dat de werkbelasting onafhankelijk is van dag tot dag.

a) Wat is de kans dat de afdeling overbezet (te weinig bedden) is op twee dagen in een enkele week? [0.016]

b) Op hoeveel dagen in het jaar kan men een dergelijke overbezetting verwachten? [11]

Oplossing

a.

$$\mathbf{P(X = r) = P(r) = \binom{n}{r} \pi^{r} (1 - \pi)^{n-r}}$$

n = 6 π = 0.25

r	$\binom{n}{r}$	P
0	1	0.808
1	7	0.175
2	21	0.0162
3	35	0.000837
4	35	0.0000259
5	21	0.0000005
6	7	0.0000000
7	1	0.0000000

Σ = 1.00

b. n = 7 π = 0.03

365 × 0.03 = 10.95

➔ 11 dagen / jaar

Poisson benadering:

λ = 7 × 0.03 = 0.21

$$\mathbf{P(X = 2) = e^{-\lambda} \frac{\lambda^{k}}{k!} = e^{-0.21} \frac{0.21^{2}}{2!} = 0.0179 = 0.018}$$

7.6 Geslachtsverdeling

De geslachtsverdeling is dusdanig dat er in verhouding iets meer jongens geboren worden: het geboortepercentage bedraagt 52%. Indien men op een moderniteit twaalf bedjes naast elkaar ziet gerangschikt, wat is de kans er hoogstens vier meisjes geteld worden? [0.23]

Oplossing

$$\mathbf{P(X = r) = P(r)} = \binom{\mathbf{n}}{\mathbf{r}} \pi^{\mathbf{r}} \mathbf{(1 - \pi)^{n-r}}$$

n = 12 π = 0.48

r	$\binom{n}{r}$	P
0	1	0.000391
1	12	0.004333
2	66	0.021982
3	220	0.067635
4	495	0.140474
5	792	0.207470
6	924	0.223429
7	…	0.176779

$\left.\begin{array}{l} 0.000391 \\ 0.004333 \\ 0.021982 \\ 0.067635 \\ 0.140474 \end{array}\right\} = 0.23481$

Σ = 1.00

hoogstens 4 ⇔ 4 of minder

P(X ≤ 4) = 0.23

7.7 Heupoperaties

Indien men vaststelt dat 20% van de heupoperaties (plaatsen van een heupprothese) een bijkomende ingreep vereisen binnen de eerste vijf jaren na de ingreep:

a) Hoe groot is de kans dat bij zes patiënten precies één patiënt een tweede ingreep zal moeten ondergaan? [0.39]
b) Stel de binomiale kansverdeling op in deze groep van zes patiënten.

Oplossing

$$\mathbf{P(X = r) = P(r) = \binom{n}{r}\pi^{r}(1-\pi)^{n-r}}$$

n = 6 π = 0.20

r	$\binom{n}{r}$	P
0	1	0.262
1	6	0.393
2	15	0.246
3	20	0.0819
4	15	0.01536
5	6	0.001536
6	1	0.000064

Σ = 1.00

a. P = 0.39

b. zie bovenstaande tabel

Oef 8. Poissonverdeling

8.1 Telefoonoproepen spoedgevallen

Het aantal telefoonoproepen die men op het dispatchcentrum van een spoedafdeling ontvangt volgt vaak een Poisson verdeling. Als men gemiddeld 120 oproepen per uur ontvangt wat is dan de kans dat:

a) hij tenminste één oproep ontvangt per minuut [0.86]
b) er minder dan een pauze van twee minuten is tussen twee oproepen? [0.98]

Oplossing

a. tenminste één oproep per minuut
gemiddeld 120 oproepen per uur
gemiddeld 2 oproepen per minuut

➔ $\lambda = 2$

$$\mathbf{P}(r \geq 1) = 1 - \mathbf{P}(r < 1) \qquad r = 0, 1, 2, \ldots$$
$$= 1 - \mathbf{P}(r = 0)$$
$$= 1 - e^{-\lambda} \frac{\lambda^0}{0!}$$
$$= 1 - e^{-2} = 0.86$$

b. *minder dan 2 minuten tussen 2 oproepen*
1 of meer oproepen in 2 minuten
gemiddeld 2 oproepen per minuut
gemiddeld 4 oproepen per 2 minuten

➔ $\lambda = 4$

$$\mathbf{P}(r \geq 1) = 1 - \mathbf{P}(r < 1) \qquad r = 0, 1, 2, \ldots$$
$$= 1 - \mathbf{P}(r = 0)$$
$$= 1 - e^{-\lambda} \frac{\lambda^0}{0!}$$
$$= 1 - e^{-4} = 0.98$$

8.2 Arbeidsongevallen

In een groot bedrijf (waar het aantal werknemers constant blijft) gebeuren gemiddeld vijf arbeidsongevallen per jaar. Wat is de kans dat in een bepaald jaar:

a) precies zeven ongevallen zich voordoen [0.10]
b) geen ongevallen gebeuren [0.0067]
c) minder dan vijf ongevallen zich voordoen? [0.44]

Oplossing

$\lambda = 5$

a. $P(r = 7) = e^{-5}\dfrac{5^7}{7!} = 0.10$

b. $P(r = 0) = e^{-5}\dfrac{5^0}{0!} = 0.0067$

c.
$$\begin{aligned}P(r < 5) &= e^{-5}\left(\frac{5^0}{0!} + \frac{5^1}{1!} + \frac{5^2}{2!} + \frac{5^3}{3!} + \frac{5^4}{4!}\right)\\ &= e^{-5}\left(1 + 5 + \frac{25}{2} + \frac{125}{6} + \frac{625}{24}\right)\\ &= 0.44\end{aligned}$$

8.3 Labo openbare hygiëne

In een labo van een departement openbare hygiëne werden honderden flesjes bier onderzocht op de aanwezigheid van schadelijke micro-organismen. Er bleek dat in 4% van de gevallen deze inderdaad werden waargenomen. Gebruik zowel de binomiale distributie als de benadering a.h.v. de Poisson-distributie om de kans te bepalen dat een student juist drie dergelijke flesjes in zijn krat (24 flesjes) bier vindt. [0.055]

Oplossing

a. Binomiaalverdeling

$n = 24$
$\pi = 0.04$

$$\mathbf{P(r = 3) = \binom{24}{3} 0.04^3 (1 - 0.04)^{21} = 0.055}$$

b. Poissonverdeling

$n = 24$
$\pi = 0.04$ $\lambda = n.\ \pi = 0.96$

$$\mathbf{P(r = 3) = e^{-0.96}\,\frac{0.96^3}{3!} = 0.056}$$

8.4 Ziekenhuisinfecties

Observatie op een grote ziekenhuisafdeling leert dat gemiddeld drie patiënten per week een ziekenhuisinfectie oplopen. Wat is de kans dat minstens vier patiënten een infectie oplopen tijdens een verblijf van twee weken? [0.85]

Oplossing

gemiddeld 3 patiënten per week

➔ gemiddeld 6 patiënten per 2 weken

➔ $\lambda = 6$

$$\mathbf{P}(\mathbf{r} \geq 4) = 1 - \mathbf{P}(\mathbf{r} < 4) \quad \mathbf{r} = 0, 1, 2, \ldots$$

$$= 1 - e^{-6}\left(\frac{6^0}{0!} + \frac{6^1}{1!} + \frac{6^2}{2!} + \frac{6^3}{3!}\right)$$

$$= 1 - e^{-6}(1 + 6 + 18 + 36) = 0.85$$

8.5 Incidentie zelfdoding

De incidentie van zelfmoord in een bepaald jaar wordt voor België geschat op 2.2 gevallen per 10.000 inwoners. In een bepaald dorp met 2.000 inwoners bleek nu echter dat er in dat jaar drie gevallen werden genoteerd. Wat is de kans dat er inderdaad drie of meer zelfmoorden worden gepleegd in dit dorp als wordt aangenomen dat er geen geografische verschillen in zelfmoordcijfers te verwachten zijn? [0.010]

Oplossing

2.2 gevallen per 10000 inwoners

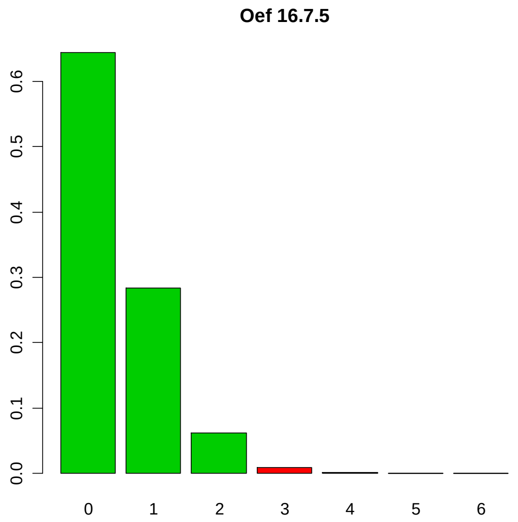
➔ 0.44 gevallen per 2000 inwoners

➔ $\lambda = 0.44$

$$\mathbf{P(r \geq 3) = 1 - P(r < 3) \qquad r = 0, 1, 2, \ldots}$$

$$\mathbf{= 1 - e^{-0.44}\left(\frac{0.44^0}{0!} + \frac{0.44^1}{1!} + \frac{0.44^2}{2!}\right)}$$

$$\mathbf{= 1 - e^{-0.44}(1 + 0.44 + 0.0968) = 0.010}$$



Oef 9. Normale benaderingen

1. **Normaalverdeling (Gauss)** (μ, σ^2)

$$P(X \le x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \int_{-\infty}^{x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \qquad (x \in R)$$

2. **Binomiale verdeling** (n, π)

$$P(X \le x) = \sum_{i=0}^{x} \binom{n}{i} \pi^i (1-\pi)^{n-i} (x \in N)$$

3. **Poisson verdeling** (λ)

$$P(X \le x) = \sum_{i=0}^{x} \frac{e^{-\lambda}\lambda^i}{i!} (x \in N)$$

NORMALE BENADERING VOOR DE BINOMIAALVERDELING

Beschouw: binomiale met parameters n en π

als: n is groot zodanig dat $n\pi \ge 5$
$n(1 - \pi) \ge 5$

dan: $B(n, \pi)$ benaderend normaalverdeeld
met parameters $N(n\pi, n\pi(1-\pi))$

NORMALE BENADERING VOOR DE POISSONVERDELING

Beschouw: Poisson met parameter λ

als: λ is groot zodanig dat $\lambda \ge 20$

dan: $P(\lambda)$ benaderend normaalverdeeld met parameters $N(\lambda, \lambda)$

NORMALE BENADERINGEN

Binomiale (n, π) $\xrightarrow[\pi < 0.10]{n > 20}$ Poisson (λ)

$\lambda = n.\pi$

$n\,\pi \geq 5$

$n\,(1 - \pi) \geq 5$

$N(n\pi, n\pi(1 - \pi))$

$\lambda \geq 20$

$N(\lambda, \lambda)$

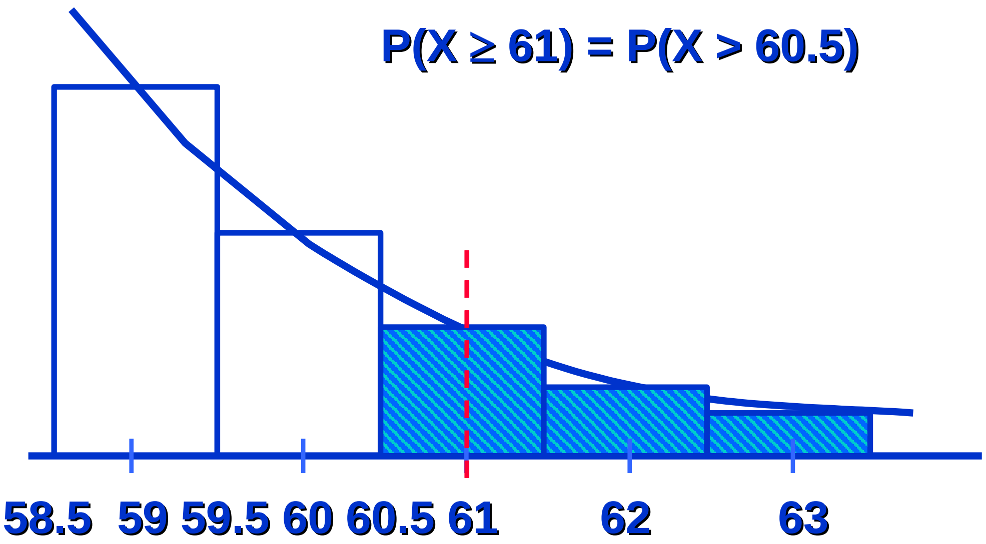
Normaalverdeling $N(\mu, \sigma^2)$

CONTINUITEITSCORRECTIE

discrete verdelingen ➡ continue verdeling

$P(X \geq x)$ vervangen door $P(X > x - 0.5)$

$P(X \leq x)$ vervangen door $P(X < x + 0.5)$



9.1 Zeeziek op een ferryboot

Waargenomen wordt dat 30% van de passagiers op een ferryboot zeeziek worden bij windkracht zeven. Bereken de kans dat bij de volgende overtocht met windkracht zeven er meer dan een kwart van de 200 passagiers zeeziek zijn. [0.93]

Oplossing

30% $\longrightarrow$ $\pi = 0.30$ n = 200

$n\pi = 200 \times 0.30 = 60$

$n(1 - \pi) = 200 \times 0.70 = 140$

P (X > 50 zieken)

$\approx N(60, 200 \times 0.30 \times 0.70)$

$\approx N(60, 42)$ $\sqrt{42} = 6.48$

$$P(X > 50) = P(X \geq 51)$$

$$= P(Z \geq \frac{50.5 - 60}{6.48})$$

$$= P(Z \geq -1.466) = 0.928$$

$$P = 0.93$$

9.2 Zakrekenmachine

Stel dat 40% van alle studenten vandaag een rekenmachine naar de les meebrachten. Als we een willekeurige groep van 50 studenten zouden selecteren, wat is de kans dat er 30 of meer studenten een rekenmachine kunnen voorleggen? [0.0031]

Oplossing

40% $\longrightarrow$ $\pi = 0.40$ $n = 50$

$n\pi = 50 \times 0.40 = 20$

$n(1 - \pi) = 50 \times 0.60 = 30$

P (X ≥ 30 studenten)

$\approx N(20, 50 \times 0.40 \times 0.60)$

$\approx N(20, 12)$ $\sqrt{12} = 3.464$

$$
\begin{aligned}
P(X \geq 30) &= P(X > 29) \\
&= P\left(Z > \frac{29.5 - 20}{3.464}\right) \\
&= P(Z > 2.7425) = 0.0031
\end{aligned}
$$

9.3 Hersenvliesontsteking

Eén van de risicofactoren voor het ontstaan van meningitis (hersenvliesontsteking) is de aanwezigheid van Haemophilus Influenzae type B. Bij kinderen onder de 5 jaar wordt het jaarlijks aantal gevallen van een dergelijke infectie geschat op 33 per 100000 kinderen. Stel dat in een bepaald jaar 'Kind & Gezin' deze ernstige infectie bij 350000 kinderen onder de 5 jaar opvolgt. Wat is de kans dat deze infectie bij meer dan 130 van deze kinderen wordt vastgesteld? [0.081]

Oplossing

33 gevallen per 100.000 kinderen

$\longrightarrow$ 115.5 gevallen per 350.000 kinderen

$\longrightarrow$ $\lambda = 115.5$ (> 20)

normale benadering met:

$\mu = \lambda = 115.5 \quad \sigma^2 = \lambda = 115.5 \qquad \sigma = 10.75$

$\approx N(115.5, 115.5)$

$$\begin{aligned}
P(X > 130) &= P(X \geq 131) \\
&= P\left(Z \geq \frac{130.5 - 115.5}{10.75}\right) \\
&= P(Z \geq 1.396) \\
&= 0.081
\end{aligned}$$

9.4 AIDS en HIV-infecties

De I.H.E.-uitgave "De Epidemiologie van AIDS en HIV-infectie in België. Toestand op 31 december 1995" (A. Sasse, H. Van Renterghem, J. Van der Heyden, nov. 1996, p.8) meldt dat tussen 1987 en 1993 er gemiddeld 71 nieuwe HIV-infecties per maand geregistreerd werden in België, ofwel 2 à 3 per dag. Sinds 1994 daalde dit gemiddelde tot 65 nieuwe diagnoses per maand.

> Aangenomen dat deze gemiddelde waarde zich stabiliseerde, wat is dan de kans dat in het laatste kwartaal van vorig jaar meer dan 220 nieuwe HIV-seropositieven geteld werden? [0.034]

Oplossing

HIV-seropositieven:

65 /maand $\longrightarrow$ 195 / kwartaal

Poisson: $\lambda = 195$ (> 20)

normale benadering met:

$\mu = \lambda = 195$ $\sigma^2 = \lambda = 195$

$\approx N(195, 195)$ $\sqrt{195} = 13.96$

$$\begin{aligned} P(X > 220) &= P(X \geq 221) \\ &= P\left(Z \geq \frac{220.5 - 195}{13.96}\right) \\ &= P(Z \geq 1.826) \\ &= 0.034 \end{aligned}$$

9.5 Heupprothesen

Geef de normale benadering van oefening 7.7.

Oplossing

52% $\longrightarrow$ $\pi = 0.48$ $n = 12$

$n\pi = 12 \times 0.48 = 5.76$

$n(1 - \pi) = 12 \times 0.52 = 6.24$

P (hoogstens 4 meisjes)

$\approx N(5.76, 12 \times 0.48 \times 0.52)$

$\approx N(5.76, 2.9952)$ $\sqrt{2.9952} = 1.73066$

$$\begin{aligned} P(X \leq 4) &= P\left(Z < \frac{4.5 - 5.76}{1.73066}\right) \\ &= P(Z < -0.7280) \\ &= 0.2327 \end{aligned}$$

P = 0.23

Oef 10. Centraal Limiet Theorema

10.1 Menselijke schedel

Stel dat de lengte van de menselijke schedel in een bepaalde populatie benaderend normaalverdeeld is met gemiddelde 185.6 mm en standaardafwijking 12.7 mm. Wat is de kans dat in een willekeurige steekproef van 10 personen uit deze populatie het gemiddelde groter dan 190 mm zal zijn? [0.14]

Oplossing

μ = 185.6 mm, σ = 12.70 mm, n = 10

$$? \ \ P(\overline{X}_{n=10} > 190)$$

$$\overline{X} \approx N(185.6, \frac{12.7^2}{10}) \approx N(185.6, 16.12)$$

$$\Rightarrow \ N(0,1); \ Z = \frac{\overline{X} - \mu}{\sigma}$$

$$P(\overline{X} > 190) = P(Z > \frac{190 - 185.6}{4.0161})$$

$$= P(Z > 1.09556) = 0.136$$

P = 0.14

10.2 IJzergehalte in serum

Als het gemiddelde en de standaardafwijking van ijzerwaarden in serum voor normale mannen respectievelijk 120 en 15 microgram per 100 ml bedragen, wat is dan de kans dat een willekeurige steekproef van 50 mannen een gemiddelde zal opleveren tussen 115 en 125 microgram per 100 ml? [0.98]

Oplossing

μ = 120 mg/100ml, σ = 15 mg/100ml, n = 50

$$? \ P(115 < \overline{X}_{n=50} < 125)$$

$$\overline{X} \approx N(120, \frac{15^2}{50}) \approx N(120, 4.5)$$

$$\Rightarrow N(0,1); \ Z = \frac{\overline{X} - \mu}{\sigma}$$

$$P(115 < \overline{X} < 125) = P(\frac{115-120}{2.1213} < Z < \frac{125-120}{2.1213})$$

$$= P(-2.3570 < Z < 2.3570) = 1 - 2 \times P(Z \geq 2.3570)$$

$$= 1 - 2 \times 0.0092 = 1 - 0.0184 = 0.9816$$

P = 0.98

10.3 Handicap

Uit een studie bleek dat gehandicapte personen, na het aanleren van een nieuwe techniek, een bepaalde taak konden uitvoeren na gemiddeld 25 seconden met een standaarddeviatie van 5 seconden. In de veronderstelling dat de tijdsduur van de taak normaalverdeeld is, wat is dan de kans dat de gemiddelde tijd van uitvoering bij een steekproef van 25 personen:

a) 26 seconden of meer bedraagt [0.16]
b) tussen 24 en 27 seconden bedraagt [0.82]
c) 26 seconden of minder bedraagt [0.84]
d) meer dan 22.5 seconden bedraagt? [0.99]

Oplossing

μ = 25 s σ = 5.0 s n = 25

a. $$P(\overline{X}_{n=25} \geq 26) = P\left(Z \geq \frac{26-25}{\frac{5}{\sqrt{25}}}\right) = P(Z \geq 1) = 0.1587$$

P = 0.16

b. $$P(24 < \overline{X}_{n=25} < 27) = P(-1 < Z < 2) = P(Z < 2) - P(Z < -1)$$
$$= P(Z < 2) - P(Z \geq 1) = 0.9772 - 0.1587 = 0.8185$$

P = 0.82

10.4 HDL-cholesterol

De gemiddelde HDL-cholesterolwaarde bij een populatie van vrouwen tussen 20 en 70 jaar is 61 mg/dl. De mediaan is 58 mg/dl en de standaarddeviatie is 14.3 mg/dl.
Analyse op het serum van bloedstalen, afgenomen bij 225 vrouwen, levert een gemiddelde HDL-cholesterol waarde van 63 mg/dl.
Wat is de kans dat bij een toevallige steekproef van die omvang minstens die meetwaarde voor het gemiddelde wordt geregistreerd? [0.018]

Oplossing

n = 225 μ = 61 mg/dl

σ = 14.3 mg/dl

$\sigma_{\bar{x}}$ is de standaardafwijking van het gemiddelde

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{14.3}{15} = 0.953$$

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \frac{63 - 61}{\frac{14.3}{15}} = \frac{2 \times 15}{14.3} = 2.098$$

P = 0.0179 <u>P = 0.018</u>

10.5 Vitale longcapaciteit

In een steekproef van 30 personen blijkt het gemiddeld longvolume (vitale longcapaciteit) 4.5 liter te bedragen. Hoeveel observaties zouden er nodig zijn om de precisie van dit gemiddelde te verdubbelen? [120]

Oplossing

$n = 30$ $\quad\quad x = 4.5$ l

$\sigma_{\bar{x}}$ is de standaardafwijking van het gemiddelde

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$ voor steekproef met grootte n

$\sigma^*_{\bar{x}}$ is de standaardafwijking van het gemiddelde

$$\sigma^*_{\bar{x}} = \frac{\sigma}{\sqrt{n^*}}$$ voor steekproef met grootte n*

Bepaal n* zodanig dat:

$$\sigma^*_{\bar{x}} = \frac{1}{2}\sigma_{\bar{x}} \quad \textbf{m.a.w.} \frac{\sigma}{\sqrt{n^*}} = \frac{1}{2}\frac{\sigma}{\sqrt{n}}$$

$\Rightarrow$ $n^* = 4n = 120$

10.6 Gemiddeld IQ

Wat is de kans dat het gemiddelde IQ van een groep van 35 studenten de waarde 106 overschrijdt?

(gegeven zijnde gemiddelde 100 en standaardafwijking 15) [0.0091]

Oplossing

$$\overline{\mathbf{Y}} \approx \mathbf{N}(\mu, \frac{\sigma^2}{\mathbf{n}}) \approx \mathbf{N}(100, \frac{15.0^2}{35}) \approx \mathbf{N}(100, 6.43)$$

$$\mathbf{P}(\overline{\mathbf{Y}} \geq 106) = \mathbf{P}(\mathbf{Z} \geq \frac{106-100}{2.54})$$

$$= \mathbf{P}(\mathbf{Z} \geq 2.37) = 0.0089$$

Voor een individuele student
(en indien IQ ~ N):

$$\mathbf{P}(\mathbf{Y} \geq 106) = \mathbf{P}(\mathbf{Z} \geq \frac{106-100}{15})$$

$$= \mathbf{P}(\mathbf{Z} \geq 0.40) = 0.34$$

Oef 11. Betrouwbaarheidsintervallen

Betrouwbaarheidsinterval rond steekproefproportie:

π = populatieparameter

$\mathbf{p} = \frac{\mathbf{r}}{\mathbf{n}}$ **= puntschatting van π**

standaardfout van p: $\mathbf{sep} = \sqrt{\frac{\mathbf{p(1-p)}}{\mathbf{n}}}$

Binomiale benaderend normaal verdeeld:

$$100(1-\alpha)\%\mathrm{BI} = \left[\mathbf{p} - \mathbf{z}_{1-\frac{\alpha}{2}}\sqrt{\frac{\mathbf{p(1-p)}}{\mathbf{n}}},\ \mathbf{p} + \mathbf{z}_{1-\frac{\alpha}{2}}\sqrt{\frac{\mathbf{p(1-p)}}{\mathbf{n}}}\right]$$

11.1 Melanoom

Onderzoekers willen een uitspraak doen over het voorkomen van een kwaadaardig melanoom bij Belgische vrouwen in de leeftijdsgroep 45 tot 54 jaar. In een willekeurige steekproef van 5000 vrouwen in deze leeftijdsgroep bleken er 28 vrouwen de ziekte te vertonen. Bereken een 95% betrouwbaarheidsinterval voor de ware ‘prevalentie’ (= relatief voorkomen) van deze ziekte. [0.35%, 0.77%]

Oplossing

Ware prevalentie p?
Puntschatting:

$$\mathbf{p} = \frac{\mathbf{r}}{\mathbf{n}} = \frac{\mathbf{28}}{\mathbf{5000}} = \mathbf{0.0056}$$

Intervalschatting:

95% BI: α = 0.05 $\longrightarrow$ $\mathbf{z}_{1-\frac{\alpha}{2}} = \mathbf{1.96}$

$$\mathbf{sep} = \sqrt{\frac{\mathbf{p(1-p)}}{\mathbf{n}}} = \sqrt{\frac{\mathbf{0.0056 \times 0.9944}}{\mathbf{5000}}} = \mathbf{0.00106}$$

95% BI: [0.0056 - 1.96×sep, 0.0056 + 1.96×sep] = [0.0035, 0.0077]

M.a.w. de ware prevalentie ligt met 95% betrouwbaarheid in het interval [0.35%, 0.77%]

11.2 Auditorium

Op de eerste drie rijen van het auditorium zitten vandaag.... studenten, waaronder... meisjes. Bereken een 90%-betrouwbaarheidsinterval voor het ware percentage meisjes in het volledig auditorium (in de veronderstelling dat meisjes willekeurig over het ganse auditorium verspreid zitten).

Oplossing

11.3 Verhoogde bloeddruk

Uit grote populatiestudies blijkt dat 15% van de Belgische mannen ouder dan 30 jaar een verhoogde bloeddruk heeft. Een onderzoeker wil nagaan of door de aanwezigheid van een chemisch bedrijf in de buurt van een grote stad, het drinkwater besmet is met bepaalde bloeddruk verhogende componenten. In een steekproef van 250 mannen uit de stad bleken er 60 inderdaad hypertensief. Bereken een 95%-betrouwbaarheidsinterval rond de proportie hypertensieve mannen. Welke conclusie kan de onderzoeker hieruit trekken?

[0.19; 0.29]

Oplossing

n = 250 hypertensieve mannen: r = 60

Puntschatting:

$$\mathbf{p} = \frac{\mathbf{r}}{\mathbf{n}} = \frac{\mathbf{60}}{\mathbf{250}} = \mathbf{0.24}$$

Intervalschatting:

$$\mathbf{sep} = \sqrt{\frac{\mathbf{p(1-p)}}{\mathbf{n}}} = \sqrt{\frac{\mathbf{0.24 \times 0.76}}{\mathbf{250}}} = \mathbf{0.0270}$$

95% BI: α = 0.05, $\longrightarrow$ $\mathbf{z}_{1-\frac{\alpha}{2}} = \mathbf{1.96}$

95% BI: [0.24 - 1.96×0.027, 0.24 + 1.96×0.027] = [0.187, 0.293]

Het 95% betrouwbaarheidsinterval bedraagt [0.19, 0.29]
Aangezien de populatiewaarde niet vervat zit in dit interval wijst dit op een significante toename van het aantal hypertensieve mannen!

Betrouwbaarheidsinterval 100(1- α)% rond het steekproefgemiddelde:

μ = populatiegemiddelde

$\overline{X}$ = puntschatting van μ

Wanneer σ gekend is:

Standaardfout van $\overline{X}$: $\mathbf{sem} = \dfrac{\sigma}{\sqrt{n}}$

$$100(1-\alpha)\%BI = \left[\bar{x} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$$

11.4 Ziekenhuisopnames

Een ziekenhuisdirecteur wenst een inschatting te maken van de gemiddelde leeftijd van de 1000 patiënten die opgenomen werden in het voorbije jaar. Hiertoe neemt hij een aselecte steekproef van 100 patiënten uit het patiëntenregister van dat jaar. Uit een voorgaand onderzoek was reeds bekend dat de standaarddeviatie van de populatie gelijk is aan 12 jaar. De steekproef levert een gemiddelde op gelijk aan 62 jaar. Bereken de 95% en 99% betrouwbaarheidsintervallen [59.6,64.4]; [58.9,65.1]

Oplossing

n = 100 σ = 12 j

Puntschatting: $\bar{x}$ = 62 j

$$\bar{X} \approx N(\mu, \frac{\sigma^2}{n}) \approx N(\mu, \frac{12^2}{100})$$

Intervalschatting:

95% BI: α = 0.05 → $z_{1-\frac{\alpha}{2}} = \mathbf{1.96}$

95% BI: [62 - 1.96 × 1.2, 62 + 1.96 × 1.2] = [59.6, 64.4]

99% BI: α = 0.01 → $z_{1-\frac{\alpha}{2}} = \mathbf{2.58}$

99% BI: [62 - 2.58 × 1.2, 62 + 2.58 × 1.2] = [58.9, 65.1]

11.5 Chronische vermoeidheid

Gevraagd wordt het 95% betrouwbaarheidsinterval te bepalen voor μ, zijnde de gemiddelde leeftijd voor vrouwen die lijden aan “chronische vermoeidheid” welke vervat zit onder de noemer “Chronic Fatigue Syndrome” (CFS, GP. Holmes, 1988). Op goede gronden wordt aangenomen dat de standaarddeviatie in de vrouwelijke CFS-populatie gelijk is aan 7.8 jaar. Bij 81 vrouwelijke patiënten is de gemiddelde leeftijd 32 jaar. [30.3, 33.7]

Oplossing

n = 81 σ = 7.8 j

Puntschatting: $\bar{x}$ = 32 j

$$\bar{\mathbf{X}} \approx \mathbf{N}(\mu, \frac{\sigma^2}{\mathbf{n}}) \approx \mathbf{N}(\mu, \frac{\mathbf{7.8}^2}{\mathbf{81}})$$

Intervalschatting:

95% BI: α = 0.05 → $\mathbf{z}_{1-\frac{\alpha}{2}}$ = **1.96**

95% BI: [32 - 1.96 × 0.8667, 32 + 1.96 × 0.8667] = [30.3, 33.7]

11.6 Slaapduur

Aangenomen wordt dat de standaardafwijking voor de slaapduur bij de volwassen populatie gelijk is aan 0.72 uur.
Bereken de vereiste steekproefomvang opdat de gemiddelde slaapduur in de populatie met 95% zekerheid kan geschat worden met een nauwkeurigheid van respectievelijk 10 minuten en 5 minuten. [72; 287]

Oplossing

σ = 0.72 u

$$100(1-\alpha)\%BI = \left[\bar{x} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$$

$$\bar{x} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

$$\left|\bar{x} - \mu\right| \le z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le a \Rightarrow a^2 \ge z^2\frac{\sigma^2}{n}$$

$$\Rightarrow n \ge z^2\frac{\sigma^2}{a^2}$$

$$|x| \le a \Leftrightarrow -a \le x \le a \Rightarrow$$

$$\left|\bar{x} - \mu\right| \le z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \Leftrightarrow -z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \bar{x} - \mu \le z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

$$-\bar{x} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le -\mu \le -\bar{x} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

$$\bar{x} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \ge \mu \ge \bar{x} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

$$\bar{x} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

10 minuten:

$$a = \frac{10}{60} = 0.1667$$

$$n \ge 1.96^2\frac{0.72^2}{0.1667^2} = 71.4$$

$n \ge 72$

5 minuten:

$$a = \frac{5}{60} = 0.08333$$

$$n \ge 1.96^2\frac{0.72^2}{0.08333^2} = 286.8$$

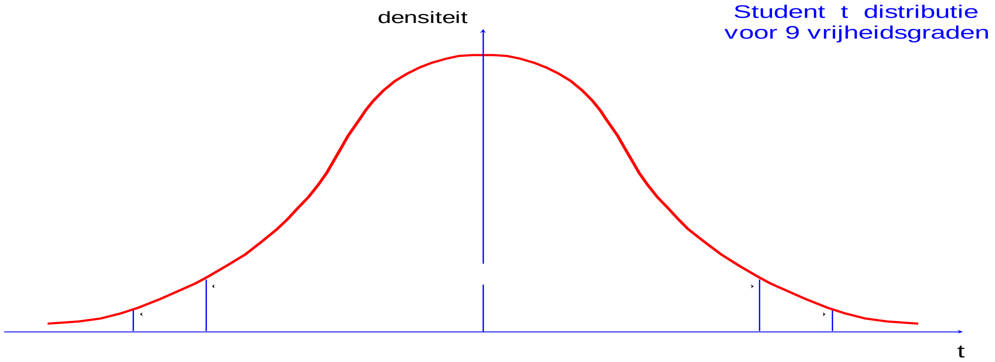
$n \ge 287$

Student's t-distributie

$$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \approx N(0,1) \quad (\approx z,\ s\tan daardnormde)$$

$$\frac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}} \approx t(n-1) \quad (Student's\ \ t\ \ met\ \ n-1\ \ vrijheidsgraden)$$

11.7 Student's t distributie



Hierboven staat de Student's t-distributie afgebeeld voor 9 vrijheidsgraden. Lees uit de Student's t-tabel (Appendix F.3) de waarden voor t af, waarvoor:

a) de gearceerde oppervlakte rechts = 0.05 [1.83]
b) de totaal gearceerde oppervlakte = 0.05 [2.26]
c) de totaal niet gearceerde oppervlakte = 0.99 [3.25]
d) de gearceerde oppervlakte links = 0.01 [-2.82]

Oplossing

de gearceerde oppervlakte rechts = 0.05 → t = 1.83

de totale gearceerde oppervlakte rechts = 0.05 → t = 2.26

de totale niet-gearceerde oppervlakte rechts = 0.99 → t = 3.25

de gearceerde oppervlakte links = 0.01 → t = -2.82

11.8 Betrouwbaarheidsgrenzen voor Student's t-distributie

De 95% betrouwbaarheidsgrenzen (tweezijdig) voor de normale verdeling worden gegeven door -1.96 en 1.96. Welke zijn de hiermee corresponderende waarden voor de t-distributie met respectievelijk:

a) $\nu = 9$ [2.26]
b) $\nu = 20$ [2.09]
c) $\nu = 30$ [2.04]
d) $\nu = 60$ [2.00]

Oplossing

a) $\nu = 9$ $t = 2.26$
b) $\nu = 20$ $t = 2.09$
c) $\nu = 30$ $t = 2.04$
d) $\nu = 60$ $t = 2.00$

Betrouwbaarheidsinterval 100(1-α)% rond het steekproefgemiddelde X

Wanneer σ niet gekend is:

$$100(1-\alpha)\%BI = \left[\bar{x} - t_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \bar{x} + t_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right]$$

Benadering voor grote n (wanneer n het bereik van de t-tabel te buiten gaat)

$$100(1-\alpha)\%BI = \left[\bar{x} - z_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right]$$

11.9 Fysische conditie van joggers

In een poging om de fysische conditie van joggers te bepalen werd in een willekeurige groep van 40 joggers de maximum volume zuurstof (VO_2) opname gemeten. Hierbij bleek het gemiddelde 47.5 ml/kg te bedragen en de standaardafwijking 4.8 ml/kg. Bereken een 95% betrouwbaarheidsinterval rond het gemiddelde. [46.0 , 49.0 ml/kg]

Oplossing

puntschatting: $\bar{x}$ = 47.5 SD = 4.8 (ml/kg) n = 40

intervalschatting: σ niet gekend

$$\text{sem} = \frac{s}{\sqrt{n}} = \frac{4.8}{\sqrt{40}} = \mathbf{0.759}$$

95% BI: α = 0.05; ν = 39 $\rightarrow t_{1-\frac{\alpha}{2}} = \mathbf{2.023}$

95% BI: [47.5 - 2.023 × 0.759, 47.5 + 2.023 × 0.759] = [46.0, 49.0]

Het ware populatiegemiddelde ligt met een 95% betrouwbaarheid in het interval:
[46.0 ml/kg, 49.0 ml/kg]

11.10 Aderverkalking

Een laboratorium bepaalt het HDL-cholesterol. “Hoge Dichtheid Lipoproteïnen” vormen de ‘goede’ fractie van het cholesterol in het bloed. Hoge HDL-waarden zijn gunstig met betrekking tot de ontwikkeling van aderverkalking. Het gemiddelde van de bepalingen bij een groep van 121 bloedstalen bij mannen is *50.1 mg/dl* met een standaarddeviatie gelijk aan *13.11 mg/dl*.

De gemiddelde waarde en de standaardafwijking voor 144 bloedstalen bij vrouwen is *64.5 mg/dl* en *15.88 mg/dl*. Bereken voor beide groepen het 95% betrouwbaarheidsinterval voor het populatiegemiddelde. [47.8, 52.4 mg/dl]
[61.9, 67.1 mg/dl]

Oplossing

puntschatting: $\bar{x}$ = 50.1 mg/dl n = 121 **MANNEN**

intervalschatting: σ niet gekend

$$\textbf{sem} = \frac{\mathbf{s}}{\sqrt{\mathbf{n}}} = \frac{\mathbf{13.11}}{\sqrt{\mathbf{121}}} = \mathbf{1.19}$$

95% BI: α = 0.05;
~ normale benadering $\rightarrow \mathbf{z}_{1-\frac{\alpha}{2}} = \mathbf{1.96}$

95% BI: [50.1 - 1.96 × 1.19, 50.1 + 1.96 ×1.19] = [47.9, 52.4]
Het ware populatiegemiddelde ligt met 95% betrouwbaarheid in het interval:
[47.9 mg/dl, 52.4 mg/dl]

puntschatting: $\bar{x}$ = 64.5 mg/dl n = 144 **VROUWEN**

intervalschatting: σ niet gekend

$$\textbf{sem} = \frac{\mathbf{s}}{\sqrt{\mathbf{n}}} = \frac{\mathbf{15.88}}{\sqrt{\mathbf{144}}} = \mathbf{1.32}$$

95% BI: α = 0.05;
~ normale benadering $\rightarrow \mathbf{z}_{1-\frac{\alpha}{2}} = \mathbf{1.96}$

95% BI: [64.5 - 1.96 × 1.32, 64.5 + 1.96 × 1.32] = [61.9, 67.1]

Het ware populatiegemiddelde ligt met 95% betrouwbaarheid in het interval:
[61.9 mg/dl, 67.1 mg/dl]

11.11 Systolische bloeddruk

Bij een groep patiënten worden volgende systolische bloeddrukken gemeten (in mmHg):
142, 136, 148, 152, 146, 142, 144, 136, 134, 138
Bereken het 95% betrouwbaarheidsinterval rond het steekproefgemiddelde.
[137.6,146.0 mmHg]

Oplossing

Bereken vooreerst de gemiddelde waarde en de standaarddeviatie voor deze steekproef.

puntschatting: $\bar{x}$ = 141.8 SD = 5.85 (mmHg) n = 10

intervalschatting: σ niet gekend

$$\mathbf{sem} = \frac{\mathbf{s}}{\sqrt{\mathbf{n}}} = \frac{\mathbf{5.85}}{\sqrt{\mathbf{10}}} = \mathbf{1.85}$$

95% BI: $\alpha = 0.05$; $\nu = 9$ $\rightarrow$ $\mathbf{t}_{1-\frac{\alpha}{2}} = \mathbf{2.26}$

95% BI: [141.8 - 2.26 × 1.85, 141.8+ 2.26 × 1.85] = [137.6, 146.0]

Het ware populatiegemiddelde ligt met een 95% betrouwbaarheid in het interval:
[137.6 mmHg, 146.0 mmHg]

In SPSS (dataset oef16.10.11.sav) **Analyse → Explore**

SPSS output **Descriptives**

			Statistic	Std. Error
systolische bloeddruk	Mean		141.80	1.849
	95% Confidence Interval for Mean	Lower Bound	137.62	
		Upper Bound	145.98	
	5% Trimmed Mean		141.67	
	Median		142.00	
	Variance		34.178	
	Std. Deviation		5.846	
	Minimum		134	
	Maximum		152	
	Range		18	
	Interquartile Range		11	
	Skewness		.323	.687
	Kurtosis		-.825	1.334

11.12 Energie-inname

Gespreid over tien dagen is bij elf gezonde vrouwen de waargenomen gemiddelde energie inname per dag (kJ) als volgt:

5260, 5470, 5640, 6180, 6390, 6515, 6805, 7515, 7515, 8230, 8770

Wat kan men besluiten betreffende hun energie-inname ten opzichte van het aangewezen niveau van 7725 kJ?
(bereken het 95% betrouwbaarheidsinterval)

[5986.4,7520.8 kJ; ongewoon laag gemiddelde]

Oplossing

Bereken vooreerst de gemiddelde waarde en de standaarddeviatie voor deze steekproef.

puntschatting: $\bar{x}$ = 6753.6 kJ SD = 1142.12 (kJ) n = 11

intervalschatting: σ niet gekend

$$\textbf{sem} = \frac{\textbf{s}}{\sqrt{\textbf{n}}} = \frac{\mathbf{1142.12}}{\sqrt{\mathbf{11}}} = \mathbf{344.363}$$

95% BI: $\alpha = 0.05$; $\nu = 10$ $\rightarrow$ $\mathbf{t}_{1-\frac{\alpha}{2}} = \mathbf{2.228}$

95% BI: [6754.6 - 2.228 $\times$ 344.363, 6754.6 + 2.228 $\times$ 344.363] = [5986.4, 7520.8]

De referentiewaarde 7725 kJ ligt niet in het bekomen interval

In SPSS (dataset oef16.10.12.sav) **Analyse ➔ Explore**

Descriptives

			Statistic	Std. Error
energieinname (kJ)	Mean		6753.64	344.363
	95% Confidence Interval for Mean	Lower Bound	5986.35	
		Upper Bound	7520.93	
	5% Trimmed Mean		6724.60	
	Median		6515.00	
	Variance		1304445.455	
	Std. Deviation		1142.123	
	Minimum		5260	
	Maximum		8770	
	Range		3510	
	Interquartile Range		1875	
	Skewness		.428	.661
	Kurtosis		-.793	1.279

11.13 Standaardfout

Indien de standaardfout een maat is voor de nauwkeurigheid van een steekproefgemiddelde, hoeveel observaties zijn er dan nodig om de precisie van een gemiddelde van 10 observaties te verdubbelen? [40]

Oplossing

Precisie van het gemiddelde voor n = 10

$$\mathbf{sem} = \frac{\mathbf{s}}{\sqrt{\mathbf{n}}} = \frac{\mathbf{s}}{\sqrt{10}}$$

Precisie verdubbelen $\Rightarrow$ sem halveren

$$\mathbf{sem^*} = \frac{\mathbf{sem}}{2} \quad \Rightarrow \quad \frac{\mathbf{s}}{\sqrt{\mathbf{n^*}}} = \frac{\frac{\mathbf{s}}{\sqrt{10}}}{2} = \frac{\mathbf{s}}{2\sqrt{10}}$$

$\Rightarrow$ n* = 40 observaties

Oef 12. BI voor een verschil in populatieparameters

BETROUWBAARHEIDSINTERVAL VOOR EEN VERSCHIL IN POPULATIEPROPORTIES

Beschouw twee verschillende, onafhankelijke populaties:

$$\Omega_A \rightarrow \pi_A,\ \text{geschat door } p_A = \frac{r_A}{n_A};\quad \Omega_B \rightarrow \pi_B,\ \text{geschat door } p_B = \frac{r_B}{n_B}$$

Vraag: $\pi_A = \pi_B$? of $\pi_A - \pi_B = 0$?

puntschatting van $\pi_A - \pi_B$: $p_A - p_B$

standaardfout van $p_A - p_B$: $SE(p_A - p_B) = \sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}$

$100(1-\alpha)\%$ BI voor $\pi_A - \pi_B$ is:

$$\left[p_A - p_B - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}},\ p_A - p_B + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}} \right]$$

Bevat dit betrouwbaarheidsinterval de 0-waarde?

BETROUWBAARHEIDSINTERVAL VOOR EEN VERSCHIL IN POPULATIEGEMIDDELDEN

Beschouw twee verschillende, onafhankelijke populaties:

$$\Omega_A \to \mu_A, \text{ geschat door } \overline{X}_A; \quad \Omega_B \to \mu_B, \text{ geschat door } \overline{X}_B$$

Vraag:

$$\mu_A = \mu_B \text{? of } \mu_A - \mu_B = 0 \text{?}$$

puntschatting van $\mu_A - \mu_B : \overline{X}_A - \overline{X}_B$

standaardfout van $\overline{X}_A - \overline{X}_B$ **:**

1. σ_A en σ_B gekend: $SE(\overline{X}_A - \overline{X}_B) = \sqrt{\dfrac{\sigma_A^2}{n_A} + \dfrac{\sigma_B^2}{n_B}}$

Bevat dit betrouwbaarheidsinterval de 0-waarde?

2. σ_A en σ_B niet gekend: $SE(\overline{X}_A - \overline{X}_B) = \sqrt{\dfrac{s_A^2}{n_A} + \dfrac{s_B^2}{n_B}}$

$100(1-\alpha)\%$ BI voor $\mu_A - \mu_B$ is:

$$\left[\overline{X}_A - \overline{X}_B - t_{1-\frac{\alpha}{2}}\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}, \overline{X}_A - \overline{X}_B + t_{1-\frac{\alpha}{2}}\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}\right]$$

$\nu = n_A + n_B - 2$ vrijheidsgraden

Bevat dit betrouwbaarheidsinterval de 0-waarde?

Gepoolde standaarddeviatie:

steekproeven $n_1, \overline{X}_1, s_1 \qquad n_2, \overline{X}_2, s_2$

wanneer $\sigma_1^2 \cong \sigma_2^2 \; (= \sigma^2)$

$$s_1^2 = \frac{\sum_{i=1}^{n_1}(x_i - \overline{x}_1)^2}{n_1 - 1}; \quad s_2^2 = \frac{\sum_{i=1}^{n_2}(x_i - \overline{x}_2)^2}{n_2 - 1}$$

gezamelijk $\Rightarrow$ alles in 1 pot

$$s^2 = \frac{\sum_{i=1}^{n_1}(x_i - \overline{x}_1)^2 + \sum_{i=1}^{n_2}(x_i - \overline{x}_2)^2}{(n_1 - 1) + (n_2 - 1)}$$

$$s = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

$$SE(\overline{X}_1 - \overline{X}_2) = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$100(1-\alpha)\%$ BI voor $\mu_A - \mu_B$ is:

$$\left[\overline{X}_A - \overline{X}_B - t_{1-\frac{\alpha}{2}} s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}},\ \overline{X}_A - \overline{X}_B + t_{1-\frac{\alpha}{2}} s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$$

$\nu = n_A + n_B - 2$ vrijheidsgraden

Bevat dit betrouwbaarheidsinterval de 0-waarde?

achterliggende redenering:

$$E(X_1 - X_2) = E(X_1) - E(X_2)$$

$$Var(X_1 - X_2) = Var(X_1) + Var(X_2) \qquad \text{app.40}$$

$$E(\overline{X}_1 - \overline{X}_2) = E(\overline{X}_1) - E(\overline{X}_2) \qquad (5.17)$$

$$Var(\overline{X}_1 - \overline{X}_2) = Var(\overline{X}_1) + Var(\overline{X}_2) \qquad \text{app.41}$$

$$= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \qquad (5.18)$$

$$SE(\overline{X}_1 - \overline{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

wanneer $\sigma_1^2 \cong \sigma_2^2 \ (= \sigma^2)$

$$SE(\overline{X}_1 - \overline{X}_2) = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

<u>BETROUWBAARHEIDSINTERVAL VOOR DE ODDS RATIO</u>

100(1-α)% BI odds ratio ln(OR) *Woolf*

$$\mathbf{SE(ln(OR)) = \sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}}$$

$$\mathbf{BI: \left[ln(OR) - z_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}},\ ln(OR) + z_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}} \right]}$$

<u>BETROUWBAARHEIDSINTERVAL VOOR HET RELATIEF RISICO</u>

100(1-α)% BI relatief risico ln(RR):

$$\mathbf{SE(ln(RR)) = \sqrt{\frac{\frac{b}{a}}{a+b}+\frac{\frac{d}{c}}{c+d}}}$$

$$\mathbf{BI: \left[ln(RR) - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\frac{b}{a}}{a+b}+\frac{\frac{d}{c}}{c+d}},\ ln(RR) + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\frac{b}{a}}{a+b}+\frac{\frac{d}{c}}{c+d}} \right]}$$

12.1 Orale contraceptie

Een onderzoeker bestudeert het effect van orale contraceptiva op hart- en vaatziekten bij vrouwen van 40 tot 45 jaar. In een groep van 5000 OC gebruiksters bleken er over een periode van drie jaar 13 hartinfarcten op te treden terwijl in de controle groep van 10000 niet-gebruiksters over diezelfde periode zeven gevallen werden genoteerd.
Bereken het 95% betrouwbaarheidsinterval rond het verschil in ware proporties tussen beide groepen. Wat kan de onderzoeker besluiten? [0.04, 0.34%]

Oplossing

$n_A = 5000$	$r_A = 13$	$p_A = 0.0026$
$n_B = 10000$	$r_B = 7$	$p_B = 0.0007$

Puntschatting: $p_A - p_B = 0.0019$

Intervalschatting:

$$\mathbf{SE}(p_A - p_B) = \sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}$$

$$= \sqrt{\frac{0.00259}{5000} + \frac{0.000700}{10000}} = \mathbf{0.000767}$$

95% BI: $\alpha = 0.05$ $\longrightarrow$ $z_{1-\frac{\alpha}{2}} = \mathbf{1.96}$

[0.0019 - 1.96 × 0.000767, 0.0019 + 1.96 × 0.000767] = [0.00040, 0.00340]

Besluit:

Het 95% BI rond het verschil in steekproefproporties bedraagt [0.04 %, 0.34 %].
Aangezien de nulwaarde, die men zou verwachten indien er geen verschil is tussen beide groepen, niet vervat ligt in dit interval kan men spreken van een significant verhoogd risico op infarct.

12.2 Pancreatitis

Sedert enkele jaren weet men dat roken een rol speelt bij het ontstaan van pancreatitis. In een zgn. ‘gevallen-controle studie’ (case-control study) werden volgende resultaten genoteerd:

	pancreatitis	**geen pancreatitis**
Rokers	38	81
Niet-rokers	15	136

Door het berekenen van 95%-betrouwbaarheidsintervallen op de verschillen, doe een uitspraak over:

- het voorkomen van pancreatitis bij rokers en niet-rokers [0.12, 0.32]
- het rookgedrag bij ‘gevallen’ en ‘controles’. [0.21, 0.48]

Oplossing

	pancreatitis	geen pancreatitis
rokers	38	81
niet-rokers	15	136

Door het berekenen van 95% betrouwbaarheidsintervallen op de verschillen, doe een uitspraak over:
- het voorkomen van pancreatitis bij rokers en niet-rokers,
- het rookgedrag bij ‘gevallen’ en ‘controles’.

	pancreatitis	geen pancreatitis	totaal
Rokers	38	81	119
Niet-rokers	15	136	151
Totaal	53	217	270

Case-control

Pancreatitis: $n_A = 53$ $r_A = 38$ $\longrightarrow$ $p_A = 0.717$

Geen-pancr: $n_B = 217$ $r_B = 81$ $\longrightarrow$ $p_B = 0.373$

Cohort

Rokers: $n_A = 119$ $r_A = 38$ $\longrightarrow$ $p_A = 0.319$

Niet-rokers: $n_B = 151$ $r_B = 15$ $\longrightarrow$ $p_B = 0.0993$

Case-control

Pancreatitis: $n_A = 53$ $r_A = 38$ $\longrightarrow$ $p_A = 0.717$

Geen-pancr: $n_B = 217$ $r_B = 81$ $\longrightarrow$ $p_B = 0.373$

Puntschatting: $p_A - p_B = 0.344$

Intervalschatting:

$$\mathbf{SE(p_A - p_B)} = \sqrt{\frac{\mathbf{0.717 \times 0.283}}{\mathbf{53}} + \frac{\mathbf{0.373 \times 0.627}}{\mathbf{217}}} = \mathbf{0.0700}$$

95% BI: $\alpha = 0.05$ $\longrightarrow$ $z_{1-\frac{\alpha}{2}} = \mathbf{1.96}$

95% BI: [0.344 - 1.96 × 0.070, 0.344 + 1.96 × 0.070] = [0.21, 0.48]
$\longrightarrow$ meer rokers bij pancr. groep

Cohort

Rokers: $n_A = 119$ $r_A = 38$ $p_A = 0.319$
Niet-rokers: $n_B = 151$ $r_B = 15$ $p_B = 0.0993$
Puntschatting: $p_A - p_B = 0.220$
Intervalschatting:

$$\mathbf{SE(p_A - p_B)} = \sqrt{\frac{\mathbf{0.319 \times 0.681}}{\mathbf{119}} + \frac{\mathbf{0.0993 \times 0.9007}}{\mathbf{151}}} = \mathbf{0.0492}$$

95% BI: α = 0.05 $\longrightarrow$ $z_{1-\frac{\alpha}{2}} = \mathbf{1.96}$

95% BI: [0.22 - 1.96 × 0.0492, 0.22 + 1.96 × 0.0492] = [0.12, 0.32]
$\longrightarrow$ verhoogde kans op pancr.

Uitgewerkt in de vorm van odds-ratio's:

	pancreatitis	geen pancreatitis	totaal
Rokers	38	81	119
Niet-rokers	15	136	151
Totaal	53	217	270

$$\mathbf{OR} = \frac{\frac{\mathbf{38}}{\mathbf{81}}}{\frac{\mathbf{15}}{\mathbf{136}}} = \frac{\mathbf{38 \times 136}}{\mathbf{81 \times 15}} = \mathbf{4.25}$$

$$\mathbf{ln(OR)} = \mathbf{1.448}$$

$$\mathbf{SE(ln(OR))} = \sqrt{\frac{\mathbf{1}}{\mathbf{a}} + \frac{\mathbf{1}}{\mathbf{b}} + \frac{\mathbf{1}}{\mathbf{c}} + \frac{\mathbf{1}}{\mathbf{d}}} = \mathbf{0.336}$$

Odds ratio:

95% BI (ln(OR))
= [1.448 - 1.96 × 0.336, 1.448 - 1.96 × 0.336] = [0.7898, 2.1057]

inverse van ln is exp

95% BI(OR) = [2.20, 8.21]

Besluit: OR = 4.25 met 95% BI van [2.20, 8.21]

12.3 Zuurstofopname

In een poging om de fysische conditie van joggers na te gaan, werd in een steekproef van 25 joggers het maximum volume zuurstofopname (VO_2max) gemeten. Het gemiddelde in deze groep werd bepaald als 47.5 ml/kg met een steekproefvariantie van 23.04 (ml/kg)^2. In een controle groep van 26 niet-joggers bleek het gemiddelde 37.5 ml/kg met een steekproefvariantie van 26.01 (ml/kg)^2. Kan men met 95% zekerheid zeggen dat joggers een grotere zuurstofopname hebben? [7.2, 12.8 ml/kg]

Oplossing

joggers:	$n_A = 25$	$\overline{X}_A = 47.5$	$SD_A = 4.8$
niet-joggers:	$n_B = 26$	$\overline{X}_B = 37.5$	$SD_B = 5.1$

Puntschatting: $\overline{\mathbf{X}}_{\mathbf{A}} - \overline{\mathbf{X}}_{\mathbf{B}} = \mathbf{10.0}$

Intervalschatting:

$$\mathbf{SE}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) = \sqrt{\frac{(\mathbf{n}_1 - 1)\mathbf{s}_1^2 + (\mathbf{n}_2 - 1)\mathbf{s}_2^2}{\mathbf{n}_1 + \mathbf{n}_2 - 2}} \times \sqrt{\frac{1}{\mathbf{n}_1} + \frac{1}{\mathbf{n}_2}}$$

$$= \sqrt{\frac{\mathbf{24 \times 23.04 + 25 \times 26.01}}{\mathbf{49}}} \times \sqrt{\frac{\mathbf{1}}{\mathbf{25}} + \frac{\mathbf{1}}{\mathbf{26}}} = \mathbf{1.39}$$

95% BI: $\alpha = 0.05$; $\nu = 25 + 26 - 2 = 49$; $\longrightarrow \mathbf{t}_{1-\frac{\alpha}{2}} = \mathbf{2.01}$

95% BI: [10.0 - 2.01 × 1.39, 10.0 + 2.01 × 1.39] = [7.2, 12.8] (ml/kg)

$\longrightarrow$ grotere zuurstofopname

12.4 Kunstvezelbedrijf

Een arbeidsgeneesheer in een kunstvezelbedrijf wil nagaan of het blootstellen aan CS_2 aanleiding geeft tot een verhoogde diastolische bloeddruk. In zijn studie vindt hij dat de 240 aan CS_2 blootgestelde arbeiders een gemiddelde bloeddruk van 83.2 mmHg hebben met een standaarddeviatie van 14.80 mmHg. Voor de 125 niet-blootgestelden bleken deze waarden echter respectievelijk 72.5 mmHg en 13.50 mmHg te bedragen.
Bereken een 90% betrouwbaarheidsinterval voor het verschil in bloeddruk tussen beide groepen.

[7.6, 13.8 mmHg]

Oplossing

CS_2:	$n_A = 240$	$= 83.2$	$SD_A = 14.80$
niet-CS_2:	$n_B = 125$	$\overline{X}_B = 72.5$	$SD_B = 13.50$

Puntschatting: $\overline{\mathbf{X}}_\mathbf{A} - \overline{\mathbf{X}}_\mathbf{B} = \mathbf{10.7}$

Intervalschatting:

$$\mathbf{SE(\overline{X}_1 - \overline{X}_2)} = \sqrt{\frac{\mathbf{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}}{\mathbf{n_1 + n_2 - 2}}} \times \sqrt{\frac{\mathbf{1}}{\mathbf{n_1}} + \frac{\mathbf{1}}{\mathbf{n_2}}}$$

$$= \sqrt{\frac{\mathbf{239 \times 219.04 + 124 \times 182.25}}{\mathbf{363}}} \times \sqrt{\frac{\mathbf{1}}{\mathbf{240}} + \frac{\mathbf{1}}{\mathbf{125}}} = \mathbf{1.58}$$

90% BI: α = 0.10; ν = 240 + 125 - 2 = 363; $\longrightarrow \mathbf{t}_{\mathbf{1-\frac{\alpha}{2}}} \approx \mathbf{z}_{\mathbf{1-\frac{\alpha}{2}}} = \mathbf{1.645}$

90% BI: [10.7 - 1.645 × 1.58, 10.7 + 1.645 × 1.58] = [8.1, 13.3] (mmHg)

$\longrightarrow$ verschil in bloeddruk

12.5 Screening voor borstkanker

In een grootstad wordt een screeningsprogramma naar borstkanker uitgevoerd. Alle vrouwen tussen 40 en 69 jaar worden aangeschreven met de vraag zich te laten onderzoeken. Twintig radiologen verlenen hierbij hun medewerking aan een ‘dubbele lezing’ van het mammogram, waarbij elk mammogram nogmaals door een andere collega van de groep beoordeeld wordt. Van de 24 kankers die zij het afgelopen halfjaar ontdekten is de gemiddelde diameter 17 mm met een standaardafwijking van 5.2 mm. De overige radiologen uit de grootstad, die niet deelnamen aan de ‘dubbele lezing’, kwamen het afgelopen halfjaar 16 borstkankers op het spoor bij routine-onderzoeken. Hier bleek de gemiddelde diameter van het gezwel 20 mm met een standaardafwijking van 6.5 mm. Bereken een 95% betrouwbaarheidsinterval voor het verschil tussen de populatie-gemiddelden van beide groepen. [-6.8, 0.8 mm]

Oplossing

$n_A = 24$	$= 17$	$SD_A = 5.2$ (mm)
$n_B = 16$	$\overline{X}_B = 20$	$SD_B = 6.5$ (mm)

Puntschatting: $\mathbf{\overline{X}_A - \overline{X}_B = -3}$

$$\mathbf{SE(\overline{X}_1 - \overline{X}_2) = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \sqrt{\frac{\mathbf{23 \times 27.04 + 15 \times 42.25}}{\mathbf{38}}} \times \sqrt{\frac{\mathbf{1}}{\mathbf{24}} + \frac{\mathbf{1}}{\mathbf{16}}} = \mathbf{1.86}$$

Intervalschatting:

95% BI: $\alpha = 0.05$; $\nu = 24 + 16 - 2 = 38$; $\longrightarrow \mathbf{t_{1-\frac{\alpha}{2}} = 2.02}$

95% BI: [-3 - 2.02 × 1.86, -3 + 2.02 × 1.86] = [-6.8, 0.8] (mm)

$\longrightarrow$ geen “statistisch” verschil

12.6 Spermastalen

Van 75 mannen geboren in 1972 wordt een spermastaal onderzocht. Op grond van de vorm, voorkomen en beweeglijkheid van het sperma worden 18 stalen als “onvruchtbaar” aangemerkt. Vijftien jaar eerder bleek dat van de spermastalen genomen bij 50 mannen uit de geboortecohorte 1957 er 7 als “onvruchtbaar” gekwalificeerd werden.

a) Bereken het 95% betrouwbaarheidsinterval voor het verschil in populatie-proporties; [-0.04, 0.24]

b) Bereken nogmaals dit betrouwbaarheidsinterval, maar dan voor driemaal zo grote steekproeven. De proporties “onvruchtbare stalen” blijven constant.[0.02, 0.18]

Formuleer telkens uw conclusies.

[b ondersteunt hypothese van reëel verschil, a niet]

Oplossing

a. Infertiliteit: $n_A = 75$ $\quad r_A = 18 \longrightarrow p_A = 0.24$

$n_B = 50$ $\quad r_B = 7 \longrightarrow p_B = 0.14$

Puntschatting: $p_A - p_B = 0.10$

$$\mathbf{SE(p_A - p_B)} = \sqrt{\frac{\mathbf{0.24 \times 0.76}}{\mathbf{75}} + \frac{\mathbf{0.14 \times 0.86}}{\mathbf{50}}} = \mathbf{0.070}$$

Intervalschatting:

95% BI: $\alpha = 0.05 \longrightarrow z_{1-\frac{\alpha}{2}} = \mathbf{1.96}$

95% BI: [0.10 - 1.96 × 0.07, 0.10 + 1.96 × 0.07] = [-0.04, 0.24]

$\longrightarrow$ geen “statistisch” verschil

b. (n x 3) $\quad n_A = 225$ $\quad r_A = 54 \longrightarrow p_A = 0.24$

$n_B = 150$ $\quad r_B = 21 \longrightarrow p_B = 0.14$

Puntschatting: $p_A - p_B = 0.10$

$$\mathbf{SE(p_A - p_B)} = \sqrt{\frac{\mathbf{0.24 \times 0.76}}{\mathbf{225}} + \frac{\mathbf{0.14 \times 0.86}}{\mathbf{150}}} = \mathbf{0.040}$$

Intervalschatting:

95% BI: $\alpha = 0.05 \longrightarrow z_{1-\frac{\alpha}{2}} = \mathbf{1.96}$

95% BI: [0.10 - 1.96 × 0.04, 0.10 + 1.96 × 0.04] = [0.02, 0.18]

$\longrightarrow$ wel verschil

12.7 Onafhankelijke steekproeven

Beschouw twee onafhankelijke steekproeven getrokken uit eenzelfde normale verdeling N(μ,σ).

a. Bereken de kans dat beide steekproefgemiddelden hoogstens σ verschillen bij een steekproefgrootte n = 15. [P = 0.99]

b. Indien men vooropstelt dat de kans dat beide steekproefgemiddelden hoogstens σ verschillen gelijk is aan P = 0.95, hoe groot moet n dan minimaal gekozen worden? [n ≥ 8]

Oplossing

a.

$$n_A = n_B = 15$$

$$SE(\overline{X}_A - \overline{X}_B) = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = \sqrt{\frac{2\sigma^2}{15}} = \sigma\sqrt{\frac{2}{15}}$$

$$z = \frac{\overline{X}_A - \overline{X}_B}{SE(\overline{X}_A - \overline{X}_B)} = \frac{\overline{X}_A - \overline{X}_B}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}}$$

$$z = \frac{\sigma}{\sigma\sqrt{\frac{2}{15}}} = \sqrt{\frac{15}{2}}$$

z = 2.74; P = 1 - 0.0061 = 0.9939

Besluit: P = 0.99

b. P = 0.95 z = 1.96

$$\frac{\overline{X}_A - \overline{X}_B}{SE(\overline{X}_A - \overline{X}_B)} = 1.96$$

$$\text{of} \quad \frac{\sigma}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{n}}} = 1.96$$

$$\text{of} \quad \frac{\sigma}{\sigma\sqrt{\frac{2}{n}}} = 1.96 \quad \text{of} \quad \sqrt{\frac{n}{2}} = 1.96$$

M.a.w. n ≥ 2 × 3.84 of n ≥ 7.68

Besluit: n ≥ 8

12.8 Behandeling van hypertensie

Deel 1. In een klinische studie van een nieuwe soort medicatie tegen hypertensie worden in twee verschillende groepen van vijf personen de bloeddrukken (in mmHg) vergeleken. De ene groep krijgt de nieuwe medicatie; de andere blijft de oude medicatie innemen. Bereken een 95% betrouwbaarheidsinterval rond het gemiddeld verschil in bloeddrukken tussen beide groepen.

Oude medicatie	Nieuwe medicatie (mmHg)
135	120
125	110
145	125
140	115
155	130

[5.9, 34.1 mmHg]

Oplossing

$n_A = 5$ $\quad \overline{X}_A = 140.0$ $\quad S^2{}_A = 125.00$ $\quad SD = 11.18$ $\quad$ mmHg

$n_B = 5$ $\quad = 120.0$ $\quad S^2{}_B = 62.50$ $\quad SD = 7.90$ $\quad$ mmHg

Puntschatting: $\mathbf{\overline{X}_A - \overline{X}_B = -20}$

$$\mathbf{SE(\overline{X}_1 - \overline{X}_2) = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \times \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$= \sqrt{\frac{\mathbf{4\times125+4\times62.5}}{\mathbf{8}}} \times \sqrt{\frac{\mathbf{1}}{\mathbf{5}}+\frac{\mathbf{1}}{\mathbf{5}}} = \mathbf{6.12}$$

Intervalschatting:

95% BI: $\alpha = 0.05$; $\nu = 5 + 5 - 2 = 8$; $\longrightarrow$ $\mathbf{t_{1-\frac{\alpha}{2}} = 2.31}$

95% BI: [20 - 2.31 × 6.12, 20 + 2.31 × 6.12] = [5.9, 34.1] (mmHg)

Independent Samples Test

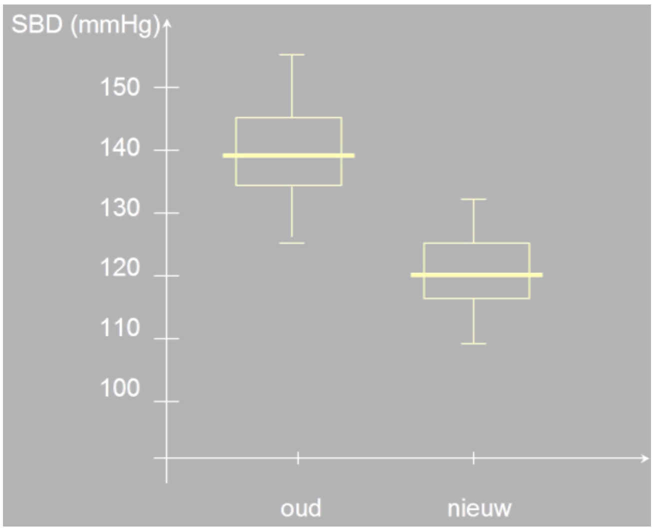
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
systolische bloeddruk (mmHg)	Equal variances assumed	.320	.587	3.266	8	.011	20.000	6.124	**5.879**	**34.121**

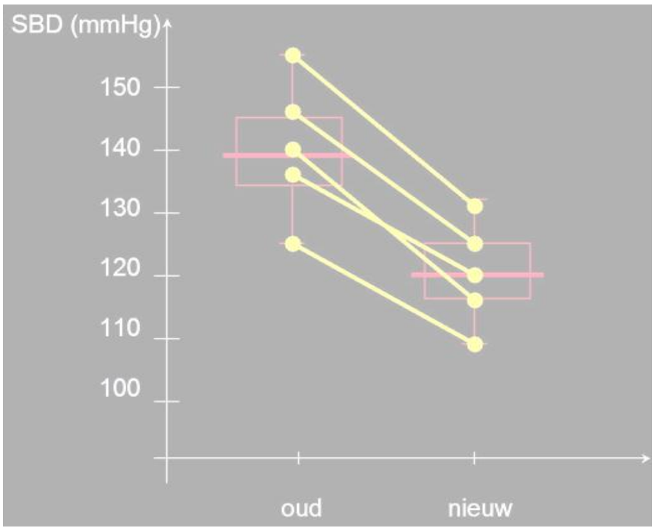
Deel 2. In een klinische cross-over studie van een nieuwe soort medicatie tegen hypertensie worden in een groep van vijf personen de bloeddrukken (in mmHg) vergeleken vóór en tijdens het gebruik van de nieuwe medicatie.
Bereken een 95% betrouwbaarheidsinterval rond het gemiddeld verschil in bloeddrukken te wijten aan het gebruik van de nieuwe medicatie.

Oude medicatie	Nieuwe medicatie (mmHg)
135	120
125	110
145	125
140	115
155	130

[13.8, 26.2 mmHg]

Oplossing





	Vroeger	Nieuw	Verschil (mmHg)
1	135	120	15
2	125	110	15
3	145	125	20
4	140	115	25
5	155	130	25

Puntschatting: $\overline{\mathbf{X}}_{\Delta} = \mathbf{20}$
Intervalschatting:

$$\mathbf{SE(\overline{X}_{\Delta})} = \frac{\mathbf{s}}{\sqrt{\mathbf{n}}} = \frac{\mathbf{5.0}}{\sqrt{\mathbf{5}}} = \mathbf{2.24}$$

95% BI: $\alpha = 0.05$; $\nu = n - 1 = 4$; $\longrightarrow \mathbf{t}_{1-\frac{\alpha}{2}} = \mathbf{2.78}$

95% BI: [20 - 2.78 × 2.24, 20 + 2.78 × 2.24] = [13.8, 26.2] (mmHg)

Oef 13. Hypothesetest in het geval van één steekproef

<u>HYPOTHESETESTEN</u>

- **voorbeelden**
- **formulering van de nulhypothese**
- **aangewende methode**
- **uitvoering van de test**
- **interpretatie en besluit**
- **grafische voorstelling**
- **bijzondere voorwaarden**

<u>TEST VOOR EEN POPULATIEPROPORTIE</u>

H_0: $\pi = \pi 0$
H_1: eenzijdig: $\pi < \pi 0$ of $\pi > \pi 0$
tweezijdig: $\pi \neq \pi 0$

Steekproef: $p = \frac{r}{n}$

Bereken:

$$z = \frac{p - \pi_0}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$$

Yates correctie

$$z = \frac{|p - \pi_0| - \frac{1}{2n}}{\sqrt{\frac{p(1-p)}{n}}}$$

z toets

Tabel F.2
Is $P \leq \alpha$?

H0 verwerpen

<u>TEST VOOR EEN POPULATIEGEMIDDELDE</u>

H_0: $\mu = \mu 0$
H_1: eenzijdig: $\mu < \mu 0$ of $\mu > \mu 0$
tweezijdig: $\mu \neq \mu 0$

Steekproef: x en s (steekproefgrootte n)

Bereken:

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \approx N(0,1)$$

z toets

Tabel F.2
Is P ≤ α?

→ **H0 verwerpen**

<u>TEST VOOR EEN POPULATIEGEMIDDELDE</u>

H_0: $\mu = \mu 0$
H_1: eenzijdig: $\mu < \mu 0$ of $\mu > \mu 0$
tweezijdig: $\mu \neq \mu 0$

Steekproef: x en s (steekproefgrootte n)

Bereken:

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \approx t_{v=n-1}$$

t toets

Tabel F.3
Is P ≤ α?

→ **H0 verwerpen (voor n voldoende groot → z-distributie)**

13.1 Vermageringsmiddel

In een studie over de werking van een nieuw soort vermageringsmiddel bleken er 28 van de 40 proefpersonen op het einde van de studieperiode inderdaad een lager gewicht te hebben. Als men aanneemt dat het menselijk gewicht op een willekeurige wijze fluctueert, kan men dan uit de resultaten van deze studie afleiden dat het middel niet werkt? (significantiedrempel $\alpha = 0.05$). [P = 0.0058]

Oplossing

H_0: $\pi = \pi_0$ ($\pi_0 = 0.50$)
H_1: tweezijdig: $\pi \neq \pi_0$ ($\pi \neq 0.50$)
$\alpha = 0.05$

Steekproef: $\mathbf{p} = \frac{\mathbf{r}}{\mathbf{n}} = \frac{\mathbf{28}}{\mathbf{40}} = \mathbf{0.70}$

Bereken:

$$\mathbf{z} = \frac{|\mathbf{p} - \pi_0| - \frac{1}{2\mathbf{n}}}{\sqrt{\frac{\mathbf{p}(1-\mathbf{p})}{\mathbf{n}}}} \approx \mathbf{N(0,1)} \qquad \textbf{CLT}$$

$$\mathbf{z} = \frac{|0.70 - 0.50| - \frac{1}{80}}{\sqrt{\frac{0.70 \times 0.30}{40}}} = \mathbf{2.588}$$

zonder de correctie van Yates:

$$\mathbf{z} = \frac{|\mathbf{p} - \pi_0|}{\sqrt{\frac{\mathbf{p}(1-\mathbf{p})}{\mathbf{n}}}} \approx \mathbf{N(0,1)} \qquad \textbf{CLT}$$

$$\mathbf{z} = \frac{|0.70 - 0.50|}{\sqrt{\frac{0.70 \times 0.30}{40}}} = \mathbf{2.760}$$

z-tabel: tabelwaarde $z_{tabel} = 1.96$

Is $P \leq \alpha$? $\longrightarrow$ <u>Besluit</u>: H_0 verwerpen (P = 0.0058)

13.2 Drinkwaterpolutie

Uit grote populatiestudies blijkt dat 15% van de Belgische mannen ouder dan 30 jaar een verhoogde bloeddruk heeft. Een onderzoeker wil nagaan of door de aanwezigheid van een chemisch bedrijf in de buurt van een grote stad, het drinkwater besmet is met bepaalde bloeddrukverhogende componenten. In een steekproef van 250 mannen uit de stad bleken er 60 inderdaad hypertensief. Kan de onderzoeker besluiten dat het resultaat van zijn studie binnen de lijn van de toevalsverwachting ligt en er dus niet echt sprake is van een besmetting? (Neem aan dat de kans op type I fout gelijk is aan $\alpha = 0.05$). [$P < 0.001$]

Oplossing

H_0: $\pi = \pi_0$ ($\pi_0 = 0.15$)
H_1: tweezijdig: $\pi \neq \pi_0$ ($\pi \neq 0.15$)
$\alpha = 0.05$

Steekproef: $\mathbf{p} = \frac{\mathbf{r}}{\mathbf{n}} = \frac{\mathbf{60}}{\mathbf{250}} = \mathbf{0.24}$

Bereken:

$$\mathbf{z} = \frac{|\mathbf{p} - \pi_0| - \frac{1}{2\mathbf{n}}}{\sqrt{\frac{\mathbf{p}(1-\mathbf{p})}{\mathbf{n}}}} \approx \mathbf{N(0,1)} \qquad \mathbf{CLT}$$

$$\mathbf{z} = \frac{|0.24 - 0.15| - \frac{1}{500}}{\sqrt{\frac{0.24 \times 0.76}{250}}} = \mathbf{3.258}$$

z-tabel

Is $P \leq \alpha$? $\longrightarrow$ <u>Besluit</u>: H_0 verwerpen ($P = 0.0012$)

Zonder Yates correctie: $z = 3.3320$; $P < 0.001$

13.3 Medicijn tegen astma-aanvallen

Aan een beoordelingscommissie wordt advies gevraagd met betrekking tot het verlenen van een licentie voor een nieuw medicijn tegen astma-aanvallen. De commissie hanteert hierbij twee stelregels: (1) het medicijn moet in minstens 99 % van de gevallen werkzaam zijn; (2) bij hoogstens één op vijfhonderd patiënten kunnen er ernstige bijwerkingen getolereerd worden. Als controle wordt het medicijn getest bij een steekproef van 200 patiënten. Bij vijf patiënten heeft het middel geen effect, bij één patiënt noteert men ernstige bijwerkingen. De werkzaamheid wordt getest met een α-drempel van 0.01, de bijwerkingen met $\alpha = 0.10$. Wat zal de commissie besluiten? [P = 0.26; P = 0.92]

Oplossing

H_0: $\pi = \pi_0$ ($\pi_0 = 0.01$) (1) <u>WERKZAAMHEID</u>

H_1: tweezijdig: $\pi \neq \pi_0$ ($\pi \neq 0.01$)

$\alpha = 0.01$

Steekproef: $\mathbf{p} = \frac{\mathbf{r}}{\mathbf{n}} = \frac{\mathbf{5}}{\mathbf{200}} = \mathbf{0.025}$

Bereken: (standaardnormale z) P = 0.258

$$\mathbf{z} = \frac{|\mathbf{0.025 - 0.01}| - \frac{\mathbf{1}}{\mathbf{400}}}{\sqrt{\frac{\mathbf{0.025 \times 0.975}}{\mathbf{200}}}} = \mathbf{1.132}$$

Is P ≤ α? ⟶ <u>Besluit</u>: H_0 NIET verwerpen (P = 0.26)

H_0: $\pi = \pi_0$ ($\pi_0 = 0.002$) (2) <u>BIJWERKINGEN</u>

H_1: tweezijdig: $\pi \neq \pi_0$ ($\pi \neq 0.002$)

$\alpha = 0.10$

Steekproef: $\mathbf{p} = \frac{\mathbf{r}}{\mathbf{n}} = \frac{\mathbf{1}}{\mathbf{200}} = \mathbf{0.005}$

$$\mathbf{z} = \frac{|\mathbf{0.005 - 0.002}| - \frac{\mathbf{1}}{\mathbf{400}}}{\sqrt{\frac{\mathbf{0.005 \times 0.995}}{\mathbf{200}}}} = \mathbf{0.10}$$

Bereken: (standaardnormale z) P = 0.92

Is P ≤ α? ⟶ <u>Besluit</u>: H_0 NIET verwerpen

⟶ licentie verlenen

13.4 Bloedgroepbepalingen

Bloedgroepbepalingen bij een aselecte steekproef van 200 West-Europese zigeuners geeft aan dat 60 onder hen de in West-Europa eerder zeldzame bloedgroep B hebben (in WE ca. 10%). Kan men aannemen dat het voorkomen van bloedgroep B bij deze WE zigeuners overeenstemt met de frequentie van 35% bij de Hindoes in Rajastan (N. Indië)? (A.E. Mourant, 1958)

Toets tweezijdig met kans op type I-fout $\alpha = 0.05$. [P = 0.143]

Oplossing

H_0: $\pi = \pi_0$ ($\pi_0 = 0.35$)

H_1: tweezijdig: $\pi \neq \pi_0$ ($\pi \neq 0.35$)

$\alpha = 0.05$

Steekproef: $\mathbf{p} = \frac{\mathbf{r}}{\mathbf{n}} = \frac{\mathbf{60}}{\mathbf{200}} = \mathbf{0.30}$

$$\mathbf{z} = \frac{\left|\mathbf{0.30 - 0.35}\right| - \frac{\mathbf{1}}{\mathbf{400}}}{\sqrt{\frac{\mathbf{0.30 \times 0.70}}{\mathbf{200}}}} = \mathbf{1.466}$$

Bereken: (standaardnormale z) P = 0.143

Is $P \leq \alpha$? $\longrightarrow$ <u>Besluit</u>: H_0 NIET verwerpen

13.5 Leven boven de 2000m

Een onderzoeker wil nagaan of personen die op grote hoogte leven (boven de 2000 meter) een gewijzigde serum hemoglobine concentratie vertonen. In een steekproef van 20 gezonde mannen uit een dorp uit de Alpen vindt hij een gemiddelde hemoglobine waarde van 15.3 g/dl met een standaarddeviatie van 1.17 g/dl. Is het gemiddelde van deze groep mannen inderdaad significant (op 5% niveau) verschillend van de referentiewaarde van 14.7 g/dl voor mannen die leven ter hoogte van de zeespiegel? [P < 0.05]

Oplossing

H_0: $\mu = \mu_0$ (μ_0 = 14.7 g/dl)
H_1: tweezijdig: $\mu \neq \mu_0$ ($\mu \neq$ 14.7g/dl)
$\alpha = 0.05$

Steekproef: $\bar{x} = 15.3$ s = 1.17 n = 20

$$\mathbf{t} = \frac{\bar{\mathbf{x}} - \mu_0}{\frac{\mathbf{s}}{\sqrt{\mathbf{n}}}} = \frac{\mathbf{15.3 - 14.7}}{\frac{\mathbf{1.17}}{\sqrt{\mathbf{20}}}} = \frac{\mathbf{0.60}}{\mathbf{0.262}} = \mathbf{2.29}$$

Bereken:

Aantal vrijheidsgraden: ν = n - 1 = 19

t-tabel = 2.093

Is $P \leq \alpha$? $\longrightarrow$ Besluit: H_0 verwerpen (P < 0.05)

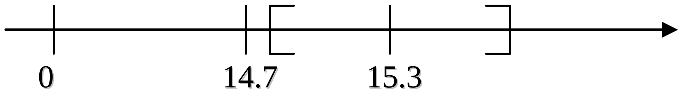
Benadering d.m.v. betrouwbaarheidsinterval:

$$\mathbf{100(1-\alpha)\%\ BI\ voor\ \mu\ is}$$

$$\left[\bar{\mathbf{x}} - \mathbf{t}_{1-\frac{\alpha}{2}} \frac{\mathbf{s}}{\sqrt{\mathbf{n}}} ; \bar{\mathbf{x}} + \mathbf{t}_{1-\frac{\alpha}{2}} \frac{\mathbf{s}}{\sqrt{\mathbf{n}}}\right]$$

95% BI: [15.3 - 2.093*0.262, 15.3 + 2.093×0.262] = [14.75, 15.85] (g/dl)

Referentie waarde = 14.7 g/dl



13.6 Ondervoeding

In een bepaalde leeftijdsklasse hebben gezonde jongens een gemiddeld lichaamsgewicht van 52.0 kg. Een epidemioloog wil het vermoeden verifiëren dat in een bepaalde buurt van een grote stad, kinderen van die leeftijd ondervoed zijn. In zijn studie van 25 dergelijke kinderen uit deze buurt vindt hij een gemiddeld gewicht van 47.2 kg met een standaarddeviatie van 6.5 kg. Is er voldoende evidentie om het vermoeden te ontkrachten? (5% significantiedrempel). [P < 0.001]

Oplossing

H_0: $\mu = \mu_0$ (μ_0 = 52.0kg)
H_1: éénzijdig: $\mu < \mu_0$ (μ < 52.0kg)
H_1: tweezijdig: $\mu \neq \mu_0$ ($\mu \neq$ 52.0kg)
α = 0.05

Steekproef: $\bar{x} = 47.2$ s = 6.5 n = 25

Bereken:

$$\mathbf{t} = \frac{\bar{\mathbf{x}} - \mu_0}{\frac{\mathbf{s}}{\sqrt{\mathbf{n}}}} = \frac{\mathbf{47.2 - 52}}{\frac{\mathbf{6.5}}{\sqrt{\mathbf{25}}}} = \frac{\mathbf{-4.8}}{\mathbf{1.3}} = \mathbf{-3.69}$$

Aantal vrijheidsgraden: ν = n - 1 = 24

t-tabel:

Is P ≤ α? ⟶ <u>Besluit</u>: H_0 verwerpen (P < 0.001)

13.7 REM-fase

Uit de literatuur is gekend dat de gemiddelde slaapduur van de verkwikkende droomslaap (REM-fase) bij volwassenen gelijk is aan twee uren. Bij een steekproef van 20 personen die slaapmiddelen innemen, is de gemiddelde duur 105 minuten met een standaardafwijking van een halfuur.
Wijkt dit gemiddelde significant af van de vooropgestelde waarde bij een tweezijdige toets? (α = 0.05). [P < 0.025]

Oplossing

H_0: $\mu = \mu_0$ (μ_0 = 2u = 120m)
H_1: éénzijdig: $\mu < \mu_0$ (μ < 120m)
$\alpha = 0.05$

Steekproef: $\bar{x} = 105$ s = 30 n = 20

Bereken:

$$\mathbf{t} = \frac{\bar{\mathbf{x}} - \mu_0}{\frac{\mathbf{s}}{\sqrt{\mathbf{n}}}} = \frac{\mathbf{105 - 120}}{\frac{\mathbf{30}}{\sqrt{\mathbf{20}}}} = \frac{\mathbf{-15}}{\mathbf{6.71}} = \mathbf{-2.24}$$

Aantal vrijheidsgraden: $\nu = n - 1 = 19$

t- tabel:

Is P ≤ α? ⟶ <u>Besluit</u>: H_0 verwerpen (P < 0.025)

Opm.: links-éénzijdig komt overeen met significantieniveau van α = 0.1; de kritische waarde is bij 19 vrijheidsgraden gelijk aan 1.729. Omdat de berekende t-waarde ook significant blijkt te zijn voor het éénzijdig 0.05 niveau, noteert men P < 0.025.

13.8 Levensverwachting bij mannen

De levensverwachting bij mannen in een bepaald land is gekend als zijnde 71.9 jaar. Onderzoekers willen nu nagaan of voor een bepaalde regio in dit land, waarvan gekend is dat er sterke milieuvervuiling is door zware nijverheid, de gemiddelde levensverwachting lager ligt. In de steekproef van 10000 mannen uit deze regio bleek de gemiddelde levensverwachting 71.7 jaar te zijn met een standaardafwijking van 7.0 jaar. Test eenzijdig bij een 5% significantiedrempel of het vermoeden van de onderzoekers gegrond is.

[P = 0.0021]

Oplossing

H_0: $\mu = \mu_0$ ($\mu_0 = 71.9$j)

H_1: tweezijdig: $\mu \neq \mu_0$ ($\mu \neq 71.9$j)

$\alpha = 0.05$

Steekproef: $\bar{x} = 71.7$ s = 7.0 n = 10000

Bereken:

$$\mathbf{t} = \frac{\bar{\mathbf{x}} - \mu_0}{\frac{\mathbf{s}}{\sqrt{\mathbf{n}}}} = \frac{\mathbf{71.7 - 71.9}}{\frac{\mathbf{7}}{\sqrt{\mathbf{10000}}}} = \mathbf{-2.86}$$

Aantal vrijheidsgraden (= 10000 – 1): overgang naar z-distributie

standaardnormale z

Is P ≤ α? ⟶ <u>Besluit</u>: H_0 verwerpen (P = 0.0042)

13.9 Incidentie van de Creutzfeldt-Jacob ziekte

Veronderstel een jaarlijkse incidentie van de Creutzfeldt-Jacob ziekte (CJ, tegenhanger van BSE bij de mens) gelijk aan 1/15 miljoen bij de bevolking onder de 45 jaar. In 1995 werden in de UK, onder de 30 miljoen inwoners jonger dan 45 jaar, tien personen getroffen door CJ. Is hier sprake van een alarmerende stijging van de ziekte?
Toets tweezijdig bij $\alpha = 0.01$. [P = 0.002]

Oplossing

H_0: $\lambda = \lambda_0$ ($\lambda = 2$) (/ 30 miljoen)
H_1: tweezijdig: $\lambda \neq 2$

$\alpha = 0.01$ $r = 10$

Steekproevenverdeling: Poisson

Overschrijdingskans:

$$P(r \geq 10) = 1 - P(r < 10)$$

$$P(r < 10) = P(r \leq 9) = e^{-2} \sum_{r=0}^{9} \frac{2^k}{k!}$$

$$= 0.135 \times 7.389 = 0.998$$

$$P(r \geq 10) = 1 - P(r < 10) = 0.002$$

$\longrightarrow$ <u>Besluit</u>: H_0 verwerpen

13.10 HIV-infecties

De aantallen nieuw gediagnosticeerde gevallen van HIV geïnfecteerden tijdens de voorbije jaren zijn als volgt (bron: Tropisch Instituut / UIA):

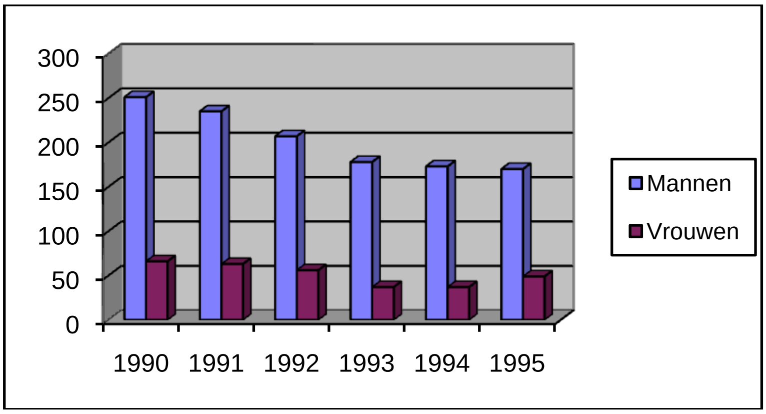
Aantallen	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**
Mannen	251	235	207	178	173	170
Vrouwen	66	63	56	37	37	49

Ga na of de verdeling van deze frequenties beantwoordt aan een uniforme distributie.
Toets tweezijdig, $\alpha = 0.05$.

[voor beide groepen kan H_0 verworpen worden: $P < 0.01$]

Oplossing

Staafdiagram



Mannen

```
Obs   Exp   (O-E)    (O-E)²    χ² term
________________________________
 251  202.3   48.7  2371.69  11.72
 235  202.3   32.7  1069.29   5.29
 207  202.3    4.7    22.09   0.11
 178  202.3  -24.3   590.49   2.92
 173  202.3  -29.3   858.49   4.24
 170  202.3  -32.3  1043.29   5.16
1214    (som)                χ² = 29.44
  202.3 (gemiddelde)       ν = 6 - 1 = 5
```

$\chi^2_{tab} = 11.070$ H_0 verwerpen ($P < 0.001$)

Vrouwen

Obs	Exp	(O-E)	$(O-E)^2$	χ^2 term
66	51.3	14.7	216.09	4.21
63	51.3	11.7	136.89	2.67
56	51.3	4.7	22.09	0.43
37	51.3	-14.3	204.49	3.99
37	51.3	-14.3	204.49	3.99
49	51.3	-2.3	5.29	0.10

308 (som) $\chi^2 = 15.39$

51.3 (gemiddelde) $\nu = 6 - 1 = 5$

$\chi^2_{tab} = 11.070$ H_0 verwerpen ($P < 0.01$)

Oef 14. Vergelijken van proporties bij twee of meerdere groepen – χ^2-testen

14.1 Inwoners uit Ibadan (Nigeria) en het sikkelcelgen

In een steekproef van 200), vindt men dat 20% drager is van het sikkelcelgen. Dit gen is verantwoordelijk voor de abnormale sikkelvorm van de rode bloedlichaampjes, hetgeen kan leiden tot bloedarmoede. Bij een steekproef van 250 negers uit New Orleans (USA), wier voorouders voor drie generaties uit Nigeria migreerden, is dit percentage 9%.
Kan men stellen dat de relatieve frequentie van het sikkelgen met meer dan 5% daalde na migratie naar een gebied waar geen malaria voorkomt, veroorzaakt door de parasiet Plasmodium falciparum? Toets tweezijdig met $\alpha = 0.05$. [P = 0.078]

Oplossing

H_0: $\pi_1 - \pi_2 = 0.05$
H_1: tweezijdig: $\pi_1 - \pi_2 \neq 0.05$
$\alpha = 0.05$

Steekproef: $p_1 = 0.20$ (Nig); $p_2 = 0.09$ (US)

Bereken:

$$\mathbf{z} = \frac{(\mathbf{p}_1 - \mathbf{p}_2) - \delta}{\sqrt{\frac{\mathbf{p}_1(1-\mathbf{p}_1)}{\mathbf{n}_1} + \frac{\mathbf{p}_2(1-\mathbf{p}_2)}{\mathbf{n}_2}}}$$

$$= \frac{\mathbf{(0.20 - 0.09) - 0.05}}{\sqrt{\frac{\mathbf{0.09 \times 091}}{\mathbf{250}} + \frac{\mathbf{0.20 \times 0.80}}{\mathbf{200}}}} = \frac{\mathbf{0.06}}{\mathbf{0.034}} = \mathbf{1.76}$$

Overschrijdingskans (P-waarde)
tweezijdig: P = 0.078

Is $P < \alpha$? $\longrightarrow$ H_0 niet verwerpen

14.2 Migraine en acupunctuur

Een beter kan verholpen worden met een acupunctuurbehandeling dan met medicatie. Immers, van 100 migraine-patiënten die de medicatie innemen, zijn er 50 die na verloop van tijd minder klachten hebben, terwijl van de 100 patiënten die een acupunctuurbehandeling volgen dit er 60 zijn. Toets tweezijdig met $\alpha = 0.05$. [P > 0.05]

Oplossing

H_0: $\pi_1 - \pi_2 = 0$
H_1: tweezijdig: $\pi_1 \neq \pi_2$
$\alpha = 0.05$

Steekproef: $n_1 = n_2 = 100$; $p_1 = 0.50$ (med); $p_2 = 0.60$ (acu)
Bereken:

$$\mathbf{z} = \frac{\mathbf{0.60 - 0.50}}{\sqrt{\frac{\mathbf{0.50 \times 050}}{\mathbf{100}} + \frac{\mathbf{0.60 \times 0.40}}{\mathbf{100}}}} = \frac{\mathbf{0.10}}{\mathbf{0.07}} = \mathbf{1.43}$$

Overschrijdingskans (P-waarde)
tweezijdig: $P(Z \geq 1.43) = 0.153$

Is $P < \alpha$? $\longrightarrow$ H_0 niet verwerpen

	klachten verminderen JA		NEEN	totaal
medicatie	50		50	100
		55	45	
acupunct.	60		40	100
totaal	110		90	200

$\alpha = 0.05$ $\chi^2_{tab} = 3.841$ (2z)

$$\chi^2_{ber} = \frac{(50-55)^2}{55} + \frac{(60-55)^2}{55} + \frac{(50-45)^2}{45} + \frac{(40-45)^2}{45} = \mathbf{2.0202}$$

$$\chi^2_{ber} = \frac{\mathbf{N(ad - bc)^2}}{\mathbf{(a+b)(a+c)(b+d)(c+d)}}$$

$$= \frac{\mathbf{200(2000 - 3000)^2}}{\mathbf{110 \times 90 \times 100 \times 100}} = \frac{\mathbf{200}}{\mathbf{99}} = \mathbf{2.0202}$$

H_0 niet verwerpen
$P > 0.05$

14.3 Schizofrenie

In een onderzoek wordt gepeild naar de genetische basis van schizofrenie. Hierbij wordt nagegaan of het risico op deze psychische stoornis afhangt van de mate van genetische verwantschap met een schizofrene persoon. Volgende tabel geeft de bevindingen weer bij 100 personen:

Tabel. Percentage schizofrene personen volgens graad van verwantschap met een schizofrene persoon, opgegroeid in hetzelfde gezin.

	halfbroer - of zus	***broer of zus***	***identieke tweeling***
niet schizofreen	90.0	80.0	50.0
wel schizofreen	10.0	20.0	50.0
Totaal	100.0 (30)	100.0 (50)	100.0 (20)

Is er een verband tussen risico op schizofrenie en mate van genetische verwantschap met een schizofrene persoon? (neem $\alpha = 0.05$, tweezijdig) [P < 0.01]

Oplossing

	halfbroer - of zus	***broer of zus***	***identieke tweeling***	
niet schizofreen	27	40	10	77
	23.1	38.5	15.4	
wel schizofreen	3	10	10	23
	6.9	11.5	4.6	
Totaal	30	50	20	100

```
O     E          O-E       (O-E)²/E
_________________________________
27   23.1        3.9        0.658
40   38.5        1.5        0.058
10   15.4       -5.4        1.894
 3    6.9       -3.9        2.204
10   11.5       -1.5        0.196
10    4.6        5.4        6.339
                 0         11.349
```

$\chi^2_{ber} = 11.349$ met $\nu = (2 - 1)(3 - 1) = 2$

$\alpha = 0.05$ $\chi^2_{tab} = 5.991$ $\longrightarrow$ H_0 verwerpen P<0.01

14.4 Gezondheid van delinquente jongeren

Onderstaande gegevens zijn afkomstig van een studie waarbij de gezondheid van een groep delinquenten (jongens) vergeleken wordt met een controle groep.
Er wordt onder meer een oogonderzoek uitgevoerd en men noteert in beide groepen voor die jongens met gezichtsproblemen of er al dan niet een bril wordt gedragen.
Bij een totaal van negen delinquenten is er slechts eentje die een bril draagt. Bij de controle groep zijn dit er vijf op een totaal van zeven met gezichtsstoornissen.
Test de nulhypothese dat bij delinquenten de noodzaak een bril te dragen evengoed wordt aangevoeld ($\alpha = 0.05$). [$P_{1zijdig} = 0.024$, $P < 0.05$]

Oplossing

Fisher's exact test

	JD	controle	totaal
bril	**1 0**	**5 6**	**6**
geen bril	**8 9**	**2 1**	**10**
totaal	**9**	**7**	**16**

$$P_{(1)} = \frac{6!10!9!7!}{16!1!5!8!2!} = \frac{9 \times 3}{8 \times 13 \times 11} = 0.0236$$

$$P_{(0)} = \frac{6!10!9!7!}{16!0!6!9!1!} = \frac{1}{8 \times 13 \times 11} = 0.00087$$

$P_{1zijdig}$ = <u>0.024</u>47 ($P < 0.05$)

14.5 Vlaamse artsenparen

Bij zestig Vlaamse artsenparen (echtgenoten hebben elk een afzonderlijke artsenpraktijk) werd nagegaan of er een associatie kan worden aangetoond tussen de grootte van beide praktijken.
Er wordt een onderverdeling gemaakt tussen de praktijken met meer of minder dan 100 patiënten per week.
De waargenomen frequentieverdeling bleek er als volgt uit te zien:

Artsenparen		man		
		<100	>100	
vrouw	<100	11	24	35
	>100	7	18	25
		18	42	60

Toets tweezijdig met α = 0.05. [P < 0.001]

Oplossing

McNemar test:

$$\chi^2 = \frac{(|24-7|-1)^2}{24+7} = \frac{16^2}{31} = \mathbf{8.26} \qquad \nu = \mathbf{1}$$

Zonder Yates' correctie: χ^2 = 9.32
De kritische χ^2 waarde is gelijk aan 3.841
Voor een tweezijdige toets bij een significantieniveau van α = 0.05 kan men de nulhypothese verwerpen; P < 0.01
Waargenomen verhoudingen van artsen met drukke praktijk:

bij de vrouw 25 op 60 of 41.7%
bij de man 42 op 60 of 70.0%
verschil: 28.3%
95% BI voor dit verschil: [11.6%, 45.0%]

Oef 15. Vergelijken van twee of meerdere groepen

15.1 Bloeddrukken bij diabetici

In een klinische studie worden volgende bloeddrukken (in mmHg) gemeten in een groep van negen diabetici en een groep van acht controles:

Diabetici:	128	138	114	135	141	120	130	120	140
Controles:	130	120	112	125	110	122	112	118	

Gebruik de Student t-test om uit te maken of deze diabetici gemiddeld dezelfde bloeddruk vertonen t.o.v. de controles ($\alpha = 0.05$). [P < 0.02]

Oplossing

H_0: $\mu_{diabetici} = \mu_{controles}$
H_1: $\mu_{diabetici} \neq \mu_{controles}$ ongepaarde Student's t-test

$n_1 = 9$ $\bar{x}_1 = 129.6$ $s_1 = 9.80$ mmHg

$n_2 = 8$ $\bar{x}_2 = 118.6$ $s_2 = 7.03$

$\bar{x}_1 - \bar{x}_2 = 129.6 - 118.6 = 11.0$

$$\mathbf{s} = \sqrt{\frac{(\mathbf{n_1}-1)\mathbf{s}_1^2 + (\mathbf{n_2}-1)\mathbf{s}_2^2}{\mathbf{n_1}+\mathbf{n_2}-2}} = \mathbf{8.62}$$

$$\mathbf{SE}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = \mathbf{s}\sqrt{\frac{1}{\mathbf{n_1}} + \frac{1}{\mathbf{n_2}}} = \mathbf{s}\sqrt{\frac{1}{9} + \frac{1}{8}} = \mathbf{4.19}$$

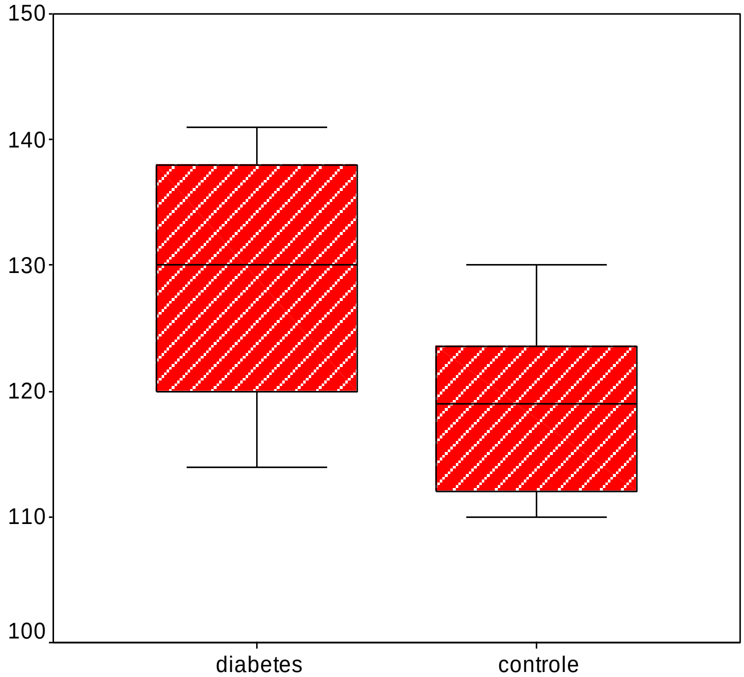
$$\mathbf{t} = \frac{\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2}{\mathbf{SE}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)} = \frac{\mathbf{11.0}}{\mathbf{4.19}} = \mathbf{2.63}$$

$\nu = n_1 + n_2 - 2 = 15$ (kritische waarde: t = 2.131) P < 0.02

<u>Besluit</u>: H_0 verwerpen; diabetici hebben verhoogde bloeddruk

<u>Opmerking</u>:

95% betrouwbaarheidsinterval voor het verschil tussen beide groepen:
[11.0 - 2.131 × 4.19, 11.0 + 2.131 × 4.19]= [2.1, 19.9]


150
140
130
120
110
100
diabetes
controle

15.2 Fysieke conditie van joggers

In een poging om de fysieke conditie van joggers na te gaan, werd in een steekproef van 25 joggers het maximum volume zuurstofopname (VO_2) gemeten. Het gemiddelde in deze groep werd bepaald als 47.5 ml/kg met een steekproefvariantie van 23.04 $(ml/kg)^2$. In een controle groep van 26 niet-joggers bleek het gemiddelde 37.5 ml/kg met een steekproefvariantie van 26.01$(ml/kg)^2$. Kan men beweren dat joggers een grotere zuurstofopname hebben?

(toets tweezijdig a.h.v. de Student t-toets, met $\alpha = 0.05$). [P < 0.001]

Oplossing

H_0: $\mu_{joggers} = \mu_{niet\text{-}joggers}$

H_1: $\mu_{joggers} \neq \mu_{niet\text{-}joggers}$ ongepaarde Student’s t-test

$n_1 = 25$ $\bar{X}_1 = 47.5$ ml/kg $S^2_1 = 23.04$

$n_2 = 26$ $\bar{X}_2 = 37.5$ $S^2_2 = 26.01$

$\bar{X}_1 - \bar{X}_2 = 47.5 - 37.5 = 10.0$

$$\mathbf{s} = \sqrt{\frac{(\mathbf{n}_1 - 1)\mathbf{s}_1^2 + (\mathbf{n}_2 - 1)\mathbf{s}_2^2}{\mathbf{n}_1 + \mathbf{n}_2 - 2}} = \mathbf{4.955}$$

$$\mathbf{SE}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = \mathbf{s}\sqrt{\frac{\mathbf{1}}{\mathbf{n}_1} + \frac{\mathbf{1}}{\mathbf{n}_2}} = \sqrt{\frac{\mathbf{1}}{\mathbf{9}} + \frac{\mathbf{1}}{\mathbf{8}}} = \mathbf{1.388}$$

$$\mathbf{t} = \frac{\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2}{\mathbf{SE}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)} = \frac{\mathbf{10.0}}{\mathbf{1.39}} = \mathbf{7.20}$$

$\nu = n_1 + n_2 - 2 = 49$ (krit.waarde: t = 2.01)

P < 0.001

<u>Besluit</u>: H_0 verwerpen; joggers hebben verhoogde VO_2

<u>Opmerking</u>:

95% CI voor het verschil tussen beide groepen:

[10.0 - 2.01 × 1.39, 10.0 + 2.01 × 1.39 = [7.2, 12.8] (ml/kg)

15.3 Bijwerkingen

In een klinische studie met acht patiënten worden bijwerkingen van een bepaald medicijn bestudeerd. Volgende tabel geeft de polsslag weer van de acht personen vóór en na de toediening van het medicijn.
Kan men besluiten a.h.v. de Student t-toets (tweezijdig met α = 0.05) dat de medicatie een verhoogde polsslag als bijwerking heeft?

Patiënt	Voor	Na
1	58	66
2	65	69
3	68	75
4	70	68
5	66	73
6	75	75
7	62	68
8	72	69

[H_0 kan niet verworpen worden]

Oplossing

Patiënt	voor	na	d
1	58	66	8
2	65	69	4
3	68	75	7
4	70	68	-2
5	66	73	7
6	75	75	0
7	62	68	6
8	72	69	-3

H_0: $\mu_d = 0$
H_1: $\mu_d \neq 0$ gepaarde Student's t-test

$$\overline{d} = \frac{27}{8} = 3.375; \; s_d = 4.406; \; SE(\overline{d}) = \frac{4.406}{\sqrt{8}} = 1.558$$

$$t = \frac{3.375}{1.558} = 2.17; \; \nu = n - 1 = 7$$

$$t_{1-\frac{\alpha}{2}} = 2.365 \;\Rightarrow\; P > 0.05$$

nulhypothese niet verwerpen
geen significant verhoogde pols

15.4 Schoolarts

Een schoolarts wil nagaan of haar aanbevelingen aan kinderen om minder zware boekentassen te torsen enig effect sorteerden. Meer in het bijzonder heeft de arts de elf kinderen uit het vijfde leerjaar, de schooljuf en de betrokken ouders hieromtrent nadere informatie verstrekt. Voor en na de actie zijn volgende gewichten genoteerd:

Gewicht van boekentas, in kg		
Kind	**Voor**	**Na**
1	7.0	7.0
2	8.5	7.5
3	9.0	9.0
4	9.5	8.5
5	9.5	10.0
6	10.0	8.0
7	10.5	9.5
8	10.5	7.0
9	11.5	10.0
10	12.0	9.5
11	12.0	13.0

Kan de schoolarts tevreden zijn over haar actie?
Test met significantieniveau $\alpha = 0.05$ onder aanname dat de verschillen tussen eerste en tweede meting normaal verdeeld zijn. [$P < 0.05$]

Oplossing

Kind	Voor	Na	d
1	7.0	7.0	0
2	8.5	7.5	1
3	9.0	9.0	0
4	9.5	8.5	1
5	9.5	10.0	-0.5
6	10.0	8.0	2
7	10.5	9.5	1
8	10.5	7.0	3.5
9	11.5	10.0	1.5
10	12.0	9.5	2.5
11	12.0	13.0	-1

H_0: $\mu_d = 0$

H_1: $\mu_d \neq 0$ gepaarde Student's t-test

$$\bar{d} = \frac{11}{11} = 1\,\text{kg};\ s_d = \sqrt{\frac{18}{10}} = 1.34\,\text{kg};\ SE(\bar{d}) = \frac{1.34}{\sqrt{11}} = 0.405$$

$$t = \frac{1}{0.405} = 2.47;\ \nu = 11 - 1 = 10$$

$$t_{1-\frac{\alpha}{2}} = 2.228 \Rightarrow P < 0.05$$

nulhypothese verwerpen;
actie had goed gevolg

Oef 16. Niet-parametrische statistische testen

16.1 Wiegedood

In een studie omtrent het verschijnsel wiegendood (het plotse en onverwachte overlijden van pasgeboren kinderen gedurende de eerste levensweken), werd de duur van de tweede ontsluitingsfase (uitgedrukt in minuten) in een aantal gevallen van wiegedood vergeleken met een groep controle kinderen.
De tijdsduur voor beide groepen is als volgt:

wiegedood: 40, 30, <6, 9, 6, 6, 11, 28, 16, 9 minuten
controles: 12, 24, 14, 6, 74, 115*, 10, 100, 8, 25, 27 minuten
* geëindigd in een heelkundige ingreep

Ga na of beide groepen van elkaar verschillen (significantieniveau $\alpha = 0.05$)

[H_0 kan niet verworpen worden]

Oplossing

WD	CON	rangen WD	rangen CON
<6		1	
6		3	
6		3	
	6		3
	8		5
9		6.5	
9		6.5	
	10		8
11		9	
	12		10
	14		11
16		12	
	24		13
	25		14
	27		15
28		16	
30		17	
40		18	
	74		19
	100		20
	115		21
	rangsom :	92	139

wiegedood: $n_1 = 10$ (mediaan = 10)
controles: $n_2 = 11$ (mediaan = 24)

$R_1 = 92$
$R_2 = 139$
$N = n_1 + n_2 = 21$
$R = (21 \times 22)/2 = 231$

TABEL: $n_1, n_2 \Longrightarrow (10, 11)$
(81 - 139)
$R_1 = 92$ (> 81, < 139)

Besluit: H_0 niet verwerpen

16.2 Vitamine C

De vitamine C uitscheiding (mg/dl) bij een groep studenten die verblijven in een home wordt vergeleken met deze van een groep die dagelijks huiswaarts keert.
De waarden zijn als volgt:

Home: 22, 38, 63, 70, 28, 121, 37, 37, 53, 27, 27, 14, 9, 28, 20, 16, 7, 37, 34

Thuis: 54, 48, 48, 47, 83, 372, 98, 255, 22, 163, 205, 89, 50

Ga na of beide groepen van elkaar verschillen (significantieniveau $\alpha = 0.05$).

[P < 0.001]

Oplossing

```
home   |  thuis       rangen
-------+--------------------------
7      |        1
9      |        2
14     |        3
16     |        4
20     |        5
22     |        6.5
   22  |              6.5
27     |        8.5
27     |        8.5
28     |        10.5
28     |        10.5
34     |        12
37     |        14
37     |        14
37     |        14
38     |        16
   47  |              17
   48  |              18.5
   48  |              18.5
   50  |              20
53     |        21
   54  |              22
63     |        23
70     |        24
   83  |              25
   89  |              26
   98  |              27
121    |              28    ...
-------+--------------------------
   rangsom:     225.5   302.5
```

home: $n_2 = 19$ (mediaan = 28)

thuis: $n_1 = 13$ (mediaan = 83)

$R_2 = 225.5$, $R_1 = 302.5$

$N = n_1 + n_2 = 32$; $R = (32 \times 33)/2 = 528$

TABEL: voor $n_1 = 13$ en $n_2 = 19$ niet beschikbaar

$$\overline{X}_T = \frac{n_1(n_1 + n_2 + 1)}{2} = \frac{13(32+1)}{2} = 214.5$$

$$S_T = \sqrt{\frac{n_2 \overline{X}_T}{6}} = \sqrt{\frac{19}{6} * 214.5} = \sqrt{679.25} = 26.06$$

$$z = \frac{T - \overline{X}_T}{S_T} = \frac{302.5 - 214.5}{26.06} = 3.38$$

<u>Besluit</u>: H_0 verwerpen (P < 0.001)

16.3 Serotoninegehalte in de hersenen

De krant bericht over een wetenschappelijk onderzoek waaruit zou blijken dat een te laag serotoninegehalte in de hersenen in agressie resulteert. Naar verluidt pleiten sommige onderzoekers ervoor om kinderen preventief te screenen op potentieel gewelddadig gedrag o.m. door analyse van de serotoninespiegels.

Stel dat men volgende waarnemingen heeft bekomen:

Kind	**serotonine-gehalte** ***(μ mol / liter bloedplaatjes)***	**agressief gedrag**
1	254	N
2	298	N
3	176	J
4	220	J
5	637	N
6	280	J
7	200	J
8	533	N
9	220	N
10	300	J
11	450	N
12	390	N
13	562	N
14	320	N
15	330	J
16	85	J
17	225	J
18	176	N

1. Bereken in elk der groepen het gemiddelde, de variantie en de standaarddeviatie. Vergelijk beide steekproefvarianties. Welke conclusie kan je hieraan verbinden met betrekking tot de keuze van een statistische test?
2. Ga na of beide steekproeven afkomstig zijn uit eenzelfde populatie. Stel de *kans* op een *type I fout* gelijk aan *0.05*. Rapporteer uw resultaten en conclusies. [$P \leq 0.05$]

Oplossing

kind	serot	agressie (rangen)	
1	254	8	
2	298	10	
3	176		2.5
4	220		5.5
5	637	18	
6	280		9
7	200		4
8	533	16	
9	220	5.5	
10	300		11
11	450	15	
12	390	14	
13	562	17	
14	320	12	
15	330		13
16	85		1
17	225		7
18	176	2.5	

agressief: $n_1 = 8$
niet-agressief: $n_2 = 10$

$R_1 = 53$
$R_2 = 118$

$N = n_1 + n_2 = 18$

$R = (18 \times 19)/2 = 171$

TABEL: n_1, n_2 $\Longrightarrow$ (8, 10)

$\alpha = 0.05$ (53 - 99)
$R_1 = 53$

<u>Besluit</u>: H_0 verwerpen ($P \leq 0.05$)

16.4 Meten van de lichaamslengte

Twee standaardprocedures voor het bepalen van de lichaamslengte (cm) bij kinderen worden met elkaar vergeleken.
De eerste methode (A) bestaat er in een zachte tractie uit te oefenen ter hoogte van het mastoïd been, terwijl bij de tweede methode (B) de kinderen gemeten worden terwijl ze gewoon in staande houding blijven.

Gegevensmatrix:

KIND	METHODE A	METHODE B
1	108.4	108.2
2	109.0	109.0
3	111.0	110.0
4	110.3	110.5
5	109.8	108.6
6	113.1	111.9
7	115.2	114.8
8	112.5	112.6
9	106.3	106.3
10	109.9	109.7
11	108.8	108.6
12	110.7	110.0
13	112.0	112.4
14	111.6	111.7
15	109.9	108.9
16	106.7	106.0
17	113.5	113.3
18	110.0	109.9
19	111.2	110.3
20	109.7	109.3

Ga na of beide methodes van elkaar verschillen (significantieniveau $\alpha = 0.05$).

[$P < 0.01$]

Oplossing

KIND	A	B	d_i (× 10)		rang	
1	108.4	108.2	2		6	
2	109.0	109.0	0			
3	111.0	110.0	10		15.5	
4	110.3	110.5		-2		6
5	109.8	108.6	12		17.5	
6	113.1	111.9	12		17.5	
7	115.2	114.8	4		10	
8	112.5	112.6		-1		2
9	106.3	106.3	0			
10	109.9	109.7	2		6	
11	108.8	108.6	2		6	
12	110.7	110.0	7		12.5	
13	112.0	112.4		-4		10
14	111.6	111.7		-1		2
15	109.9	108.9	10		15.5	
16	106.7	106.0	7		12.5	
17	113.5	113.3	2		6	
18	110.0	109.9	1		2	
19	111.2	110.3	9		14	
20	109.7	109.3	4		10	

$T^+ = 151$ $T^- = 20$

{ $n = 18$ $T = 18 \times 19/2 = 171$ }

Tabel: aantal niet-nul paren = 18

voor $\alpha = 0.05$: interval 40 - 131

H_0 wordt verworpen ($P < 0.01$)

16.5 Bijwerkingen

(deze oefening kwam reeds aan bod bij de gepaarde Student's t-test)

In een klinische studie bij acht patiënten worden de bijwerkingen van een bepaald medicijn bestudeerd. Volgende tabel geeft de polsslag (slagen per minuut) weer van de acht personen vóór en na de toediening van het medicijn. Kan men besluiten (significantieniveau $\alpha = 0.05$) dat de medicatie een gewijzigde polsslag als bijwerking heeft?

Patiënt	Voor	Na
1	58	66
2	65	69
3	68	75
4	70	68
5	66	73
6	75	75
7	62	68
8	72	69

[H_0 kan niet verworpen worden]

Oplossing

Patiënt	Voor	Na	d_i		rangen	
1	58	66	8		7	
2	65	69	4		3	
3	68	75	7		5.5	
4	70	68		-2		1
5	66	73	7		5.5	
6	75	75	0		-	
7	62	68	6		4	
8	72	69		-3		2
	rangsommen (n = 7)				25	3

$T^+ = 25$ $\qquad T^- = 3$

{ n = 7 $\qquad T = 7 \times 8/2 = 28$ }

Tabel: aantal niet-nul paren = 7
voor $\alpha = 0.05$: interval 2 - 26

H_0 kan niet verworpen worden ($P \leq 0.1$)
voor $\alpha = 0.1$: interval 3 - 25

Alternatieve benadering: **gepaarde Student t-test**

H_0: $\mu_d = 0$
H_1: $\mu_d \neq 0$ gepaarde Student's t-test

$$\bar{D} = \frac{27}{8} = 3.4 \text{ b/m}; \quad S_D = \sqrt{\frac{227 - \frac{729}{8}}{7}} = 4.40 \text{ b/m}$$

$$t = \frac{3.4}{\frac{4.40}{\sqrt{8}}} = 2.186; \quad \nu = 8 - 1 = 7$$

$$t_{1-\frac{\alpha}{2}} = 2.365 \quad \Rightarrow \quad P < 0.1 \ (NS)$$

nulhypothese kan niet verworpen worden

16.6 Glucose gehalte in bloed

Het glucose gehalte in het bloed (mg/kg) van konijnen, geprikt voor en twee uren na de toediening van een analgeticum dient te worden vergeleken.

Gegevenstabel:

Konijn	voor	na
1	125	206
2	124	134
3	158	204
4	123	123
5	119	115
6	122	116
7	89	171
8	111	212
9	135	134
10	129	129
11	138	177
12	122	136
13	127	137
14	127	117
15	137	127
16	120	140
17	118	153
18	126	147
19	134	131
20	134	139

Ga na of er een verschil is in glucosespiegel na de toediening van het analgeticum (significantieniveau $\alpha = 0.05$). [$P < 0.01$]

Oplossing

Konijn	voor	na	d_i	rang	
1	125	206	81	16	
2	124	134	10	7.5	
3	158	204	46	15	
4	123	123	0		
5	119	115	-4		3
6	122	116	-6		5
7	89	171	82	17	
8	111	212	101	18	
9	135	134	-1		1
10	129	129	0		
11	138	177	39	14	
12	122	136	14	10	
13	127	137	10	7.5	
14	127	117	-10		7.5
15	137	127	-10		7.5
16	120	140	20	11	
17	118	153	35	13	
18	126	147	21	12	
19	134	131	-3		2
20	134	139	5	4	

$T^- = 26$ $T^+ = 145$

($n = 18$; $T = 18 \times 19/2 = 171$)

<u>Besluit</u>: de nulhypothese kan verworpen worden ($P < 0.01$)

16.7 Alcoholconsumptie

Een onderzoeker wil nagaan of het drinken van alcohol resulteert in een verhoogde cholesterolconcentratie in het bloed. Hiertoe voert hij een cholesterolbepaling uit in drie groepen van elk vijf proefpersonen met verschillend alcoholverbruik:

	cholesterolconcentratie (mg/dl)				
niet-drinkers	225	250	235	225	220
matige drinkers	230	245	220	215	215
zware drinkers	250	240	250	280	265

Ga na of de drie groepen onderling verschillen (significantieniveau $\alpha = 0.05$).

[$P < 0.02$]

Oplossing

groep	cholesterol	rang		
M	215		1.5	
M	215		1.5	
N	220	3.5		
M	220		3.5	
N	225	5.5		
N	225	5.5		
M	230		7	
N	235	8		
Z	240			9
M	245		10	
N	250	12		
Z	250			12
Z	250			12
Z	265			14
Z	280			15
rangsommen:		34.5	23.5	62

Totale rangsom: 120

$n = n_1 + n_2 + n_3 = 5 + 5 + 5 = 15$

$R_{tot} = n(n + 1)/2 = 15 \times 16/2 = 120$

$$\mathbf{H} = \frac{12}{15 \times 16}\left(\frac{34.5^2}{5} + \frac{23.5^2}{5} + \frac{62^2}{5}\right) - 3 \times 16$$

$$= \frac{12}{240} \times \frac{1}{5}(1190.25 + 552.25 + 3844) - 48 = 7.865$$

3 groepen ⟹ 2 vrijheidsgraden
($\nu = 2$)

kritische χ^2 waarde ($\alpha = 0.05$): 5.991

<u>Besluit</u>: de nulhypothese kan verworpen worden (P < 0.02)

16.8 Verband tussen dieet en serum-cholesterolgehalte

In een studie naar de relatie tussen dieet en serum-cholesterolgehalte werden zestien mannen, die elk een verschillend dieet kregen, aselect in drie groepen verdeeld. In onderstaande tabel worden de cholesterolwaarden weergegeven die na het volgen van het dieet bekomen werden.

Dieet 1	**dieet 2**	**dieet 3**
5.2	5.4	6.1
5.4	5.7	6.4
5.5	5.8	7.1
5.6	6.1	6.8
5.9	7.0	7.3
6.1		

Kan men besluiten (significantieniveau $\alpha = 0.05$) dat het dieet verband houdt met de cholesterolgehaltes? [$P < 0.02$]

Oplossing

```
    dieetgroep      cholest            rang
_________________________________________________
     1         5.2            1
     1         5.4            2.5
        2          5.4                    2.5
     1         5.5            4
     1         5.6            5
        2          5.7                     6
        2          5.8                     7
     1         5.9            8
     1         6.1            10
        2          6.1                    10
           3                6.1                    10
           3                6.4                    12
           3                6.8                    13
        2         7.0                      14
           3                7.1                    15
           3                7.3                    16
_________________________________________________
          rangsommen:                  30.5  39.5  66
```

Totale rangsom: 136

$n = n_1 + n_2 + n_3 = 6 + 5 + 5 = 16$

$R_{tot} = n(n + 1)/2 = 16 \times 17/2 = 136$

$$\mathbf{H} = \frac{\mathbf{12}}{\mathbf{16 \times 17}}\left(\frac{\mathbf{30.5^2}}{\mathbf{6}} + \frac{\mathbf{39.5^2}}{\mathbf{5}} + \frac{\mathbf{66^2}}{\mathbf{5}}\right) - \mathbf{3 * 17}$$

$$= \frac{\mathbf{12}}{\mathbf{272}}\left(\frac{\mathbf{930.25}}{\mathbf{6}} + \frac{\mathbf{1560.25}}{\mathbf{5}} + \frac{\mathbf{4356}}{\mathbf{5}}\right) - \mathbf{5}$$

$$= \frac{\mathbf{12}}{\mathbf{272}}(\mathbf{155.0} + \mathbf{312.0} + \mathbf{871.2}) - \mathbf{51}$$

$$= \mathbf{59.04} - \mathbf{51} = \mathbf{8.04}$$

3 groepen $\Longrightarrow$ 2 vrijheidsgraden ($\nu = 2$)

kritische χ^2 waarde ($\alpha = 0.05$): 5.991

Besluit: de nulhypothese kan verworpen worden ($P < 0.02$)

Oef 17. Correlatie – lineaire regressie

Correlatie (Pearson)

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}}$$

$$r = \frac{\sum xy - \frac{\left(\sum x\right)\left(\sum y\right)}{n}}{\sqrt{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}\sqrt{\sum y^2 - \frac{\left(\sum y\right)^2}{n}}}$$

H_0: r = 0
H_1: r ≠ 0

Significantie: tabel F.6

als n > 60, dan volgt $\mathbf{r}\sqrt{\frac{\mathbf{n-2}}{\mathbf{1-r^2}}}$

t distributie met ν = n - 2

Correlatie (Spearman r_S)

$$r_S = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

H_0: r_S = 0
H_1: r_S ≠ 0

Significantie: tabel F.7

als n > 30, dan volgt $r_S\sqrt{\frac{n-2}{1-r_S^2}}$

t distributie met ν = n - 2

17.1 Digoxine

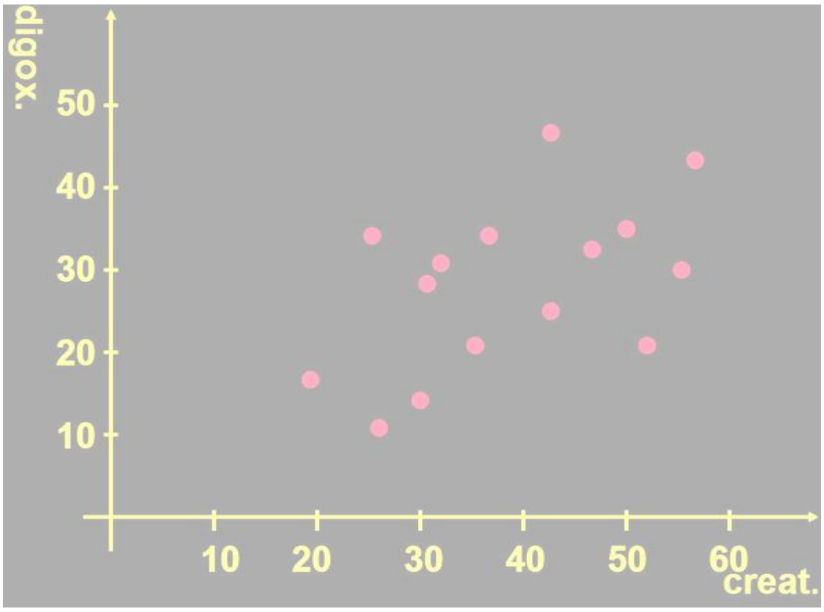
Digoxine is een geneesmiddel dat grotendeels onveranderd in de urine terug wordt afgescheiden. Bereken de correlatie (Pearson r en Spearman r_S) tussen de renale digoxine klaring en de creatinine klaring voor de bijgaande gegevenstabel afkomstig van vijftien patiënten die behandeld werden met dit geneesmiddel voor congestief hartfalen.

Patiënt	**Creatinine klaring**	**Digoxine klaring**
	ml/min/1.73m^2	
1	19.5	17.5
2	24.7	34.8
3	26.5	11.4
4	31.1	29.3
5	31.3	13.9
6	31.8	31.6
7	34.1	20.7
8	36.6	34.1
9	42.4	25.0
10	42.8	47.4
11	44.2	31.8
12	49.7	36.1
13	51.3	22.7
14	55.0	30.7
15	55.9	42.5

[$r = 0.50$; $r_S=0.48$]

[H_0 kan niet verworpen worden: geen significante correlatie]

Oplossing



CRE	DIG	$crea^2$	$digo^2$	crea*digo
19.5	17.5	380.3	306.3	341.3
24.7	34.8	610.1	1211.0	859.6
26.5	11.4	702.3	130.0	302.1
31.1	29.3	967.2	858.5	911.2
31.3	13.9	979.7	193.2	435.1
31.8	31.6	1011.2	998.6	1004.9
34.1	20.7	1162.8	428.5	705.9
36.6	34.1	1339.6	1162.8	1248.1
42.4	25.0	1797.8	625.0	1060.0
42.8	47.4	1831.8	2246.8	2028.7
44.2	31.8	1953.6	1011.2	1405.6
49.7	36.1	2470.1	1303.2	1794.2
51.3	22.7	2631.7	515.3	1164.5
55.0	30.7	3025.0	942.5	1688.5
55.9	42.5	3124.8	1806.3	2375.8
576.9	429.5	23987.9	13739.1	17325.2

r = 0.501

$$\mathbf{r} = \frac{\sum \mathbf{xy} - \frac{\left(\sum \mathbf{x}\right)\left(\sum \mathbf{y}\right)}{\mathbf{n}}}{\sqrt{\sum \mathbf{x}^2 - \frac{\left(\sum \mathbf{x}\right)^2}{\mathbf{n}}}\sqrt{\sum \mathbf{y}^2 - \frac{\left(\sum \mathbf{y}\right)^2}{\mathbf{n}}}}$$

$$\mathbf{r} = \frac{\mathbf{17325.2} - \frac{\mathbf{576.9} \times \mathbf{429.5}}{\mathbf{15}}}{\sqrt{\mathbf{23987.9} - \frac{(\mathbf{576.9})^2}{\mathbf{15}}}\sqrt{\mathbf{13739.1} - \frac{(\mathbf{429.5})^2}{\mathbf{15}}}}$$

r = 0.50

Voor $\alpha = 0.05$ en n = 15 kan H_0 niet verworpen worden.

Spearman correlatie

CREA	DIGO	$rang_C$	$rang_D$	d^2
19.5	17.5	1	3	4
24.7	34.8	2	12	100
26.5	11.4	3	1	4
31.1	29.3	4	7	9
31.3	13.9	5	2	9
31.8	31.6	6	9	9
34.1	20.7	7	4	9
36.6	34.1	8	11	9
42.4	25.0	9	6	9
42.8	47.4	10	15	25
44.2	31.8	11	10	1
49.7	36.1	12	13	1
51.3	22.7	13	5	64
55.0	30.7	14	8	36
55.9	42.5	15	14	1
			som	290

$$\mathbf{r_S = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}}$$

$$\mathbf{r_S = 1 - \frac{6 \times 290}{15 \times 224} = 1 - \frac{1740}{3360} = 1 - 0.518 = 0.482}$$

$r_S = 0.48$ $\quad$ $P = 0.054$

Voor $\alpha = 0.05$ en $n = 15$ kan H_0 niet verworpen worden.

17.2 Ruimtevaartgeneeskunde

Een jonge arts, gespecialiseerd in de ruimtevaartgeneeskunde, wil de relatie tussen lengte en gewicht van zijn medereizigers op Mars bestuderen. Op de rode planeet (1/2° aardstraal, 1/10° aardmassa) is de zwaartekrachtversnelling gelijk aan 3.9 ms^{-2}. Dit is 2/5 van de gravitatie-versnelling op aarde (g_A = 9.8 ms^{-2}), waardoor ieders gewicht op Mars overeenkomt met 2/5 van zijn of haar aards gewicht (W = mg).

Voor de 10 medepassagiers worden volgende lengtes en gewichten genoteerd:

	X	**Y**
Passagier	**lengte (cm)**	**gewicht op Mars (kg)**
1	155	18
2	160	22
3	163	26
4	165	24
5	168	23
6	170	27
7	170	23
8	175	28
9	180	28
10	185	37

Hieruit laat zich berekenen:

	lengte (cm)	**gewicht (kg)**
gemiddelde	169.1 cm	25.6 kg
standaarddeviatie	9.10 cm	5.06 kg
correlatie r (Pearson)	0.89	

a) Bereken hieruit de regressierechten van y op x en x op y. [y = -57.8 + 0.49 x]
[x = 128.3 + 1.59 y]
b) Voorspel iemands gewicht op Mars als deze persoon 170 cm groot is. [26 kg]
c) Bepaal de lengte als de persoon op de Aarde 100 kg weegt. [192 cm]
d) Bepaal de determinatiecoëfficiënt R^2. [0.79]