

Statistiek

Oefenboek

J. Demanet, J. Vlegels, L. Van den Broeck, L. Mertens, A. Van Pottelberge,
L. Van Eycken en J. Lievens

Alle rechten voorbehouden. Behoudens de uitdrukkelijk bij wet bepaalde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand of openbaar gemaakt, op welke wijze ook, zonder de voorafgaande en schriftelijke toestemming van de auteurs.

Beste student,

De oefeningen die je in dit boek terugvindt vormen, naast de theorielessen, het tweede luik van het vak 'Statistiek'. Ze zullen je helpen om de theoretische inzichten in de praktijk om te zetten.

Het eerste deel van dit oefenboek omvat de oefeningen die je thuis dient te maken als voorbereiding op de oefenlessen (A-oefeningen). Je kan deze ook terugvinden én oplossen op Minerva, waar je telkens gedetailleerde elektronische feedback op je antwoord krijgt. Eventueel bijkomende vragen kunnen in het begin van de les gesteld worden.

In het tweede deel van dit boek staan de oefeningen die in de les aan bod komen centraal (B-oefeningen). Ze worden daar in detail overlopen en gebruikt om de leerstof verder in te oefenen. Tijdens de les krijg je zo vaak mogelijk tijd om deze oefeningen zelf op te lossen. Het kan nuttig zijn om de oefeningen op voorhand al eens te bekijken.

Het derde deel van dit oefenboek bevat oefeningen die niet behandeld worden in de les (C-oefeningen). Deze oefeningen dien je zelfstandig thuis te maken om de leerstof verder in te studeren en het examen voor te bereiden. De oplossingen van deze C-oefeningen zijn als bijlage in dit boek opgenomen. Een selectie van deze oefeningen kan je ook terugvinden op Minerva, voorzien van uitgebreide elektronische feedback (C*-oefeningen). Deze C*-oefeningen dien je telkens zelfstandig na de oefenles te maken.

Veel succes!

DEEL A: VOORBEREIDENDE OEFENINGEN

1.	Univariate beschrijvende statistiek.....	2
1.1.	Basisconcepten.....	2
1.2.	Frequentieverdelingen	3
1.3.	Maten van positie	4
1.4.	Maten van spreiding	5
1.5.	Vorm van een verdeling	7
1.6.	Transformaties van variabelen	8
1.7.	Dichtheidskrommen.....	9
2.	Bivariate beschrijvende statistiek.....	9
2.0.	Studie van samenhangen en effecten.....	9
2.1.	Relaties tussen categorische variabelen	10
2.2.	Relaties tussen metrische variabelen.....	12
2.3.	De limieten van de beschrijvende bivariate statistiek	14
3.	Inductieve statistiek	15
3.1.	Studie van de systematiek van het toeval: kansrekenen.....	15
3.2.	Stochastische variabelen	15
3.3.	Studie van steekproefvariabiliteit.....	15
3.4.	Basistools voor de inductieve statistiek	16
3.5.	Inferentie voor verwachtingen.....	17
3.6.	Inferentie voor fracties	18
3.7.	Inferentie voor kruistabellen.....	19

DEEL B: BEGELEIDE OEFENINGEN

1.	Univariate beschrijvende statistiek.....	21
1.1.	Basisconcepten.....	21
1.2.	Frequentieverdelingen	23
1.3.	Maten van positie	24
1.4.	Maten van spreiding	28

1.5.	Vorm van een verdeling	30
1.6.	Transformaties van variabelen	31
1.7.	Dichtheidskrommen	31
2.	Bivariate beschrijvende statistiek	33
2.0.	Studie van samenhangen en effecten	33
2.1.	Relaties tussen categorische variabelen	33
2.2.	Relaties tussen metrische variabelen	36
2.3.	De limieten van de beschrijvende statistiek	40
2.	Inductieve statistiek	42
3.1.	Studie van de systematiek van het toeval: kansrekenen	42
3.2.	Stochastische variabelen	42
3.3.	Studie van steekproefvariabiliteit	42
3.4.	Basistools van de inductieve statistiek	42
3.5.	Inferentie voor verwachtingen	43
3.6.	Inferentie voor fracties	45
3.7.	Inferentie voor kruistabellen	47
DEEL C: ZELFSTUDIE		
1.	Univariate beschrijvende statistiek	49
1.1.	Basisconcepten	49
1.2.	Frequentieverdelingen	53
1.3.	Maten van positie	54
1.4.	Maten van spreiding	58
1.5.	Vorm van een verdeling	63
1.6.	Transformaties van een variabele	66
1.7.	Dichtheidskrommen	66
3.	Bivariate beschrijvende statistiek	70
2.0.	Studie van samenhangen en effecten	70
2.1.	Relaties tussen categorische variabelen	70
2.2.	Relaties tussen metrische variabelen	75
2.3.	De limieten van de beschrijvende statistiek	86

3.	Inductieve statistiek	91
3.1.	De studie van de systematiek van het toeval: kansrekenen	91
3.2.	Stochastische variabelen	91
3.3.	Studie van steekproefvariabiliteit	91
3.4.	Basistools van de inductieve statistiek.....	91
3.5.	Inferentie voor verwachtingen.....	91
3.6.	Inferentie voor fracties	97
3.7.	Inferentie voor kruistabellen.....	100
3.8.	Overzichtsoefeningen	101
DEEL D: BIJLAGEN		
	Bijlage 1: z-tabel.....	107
	Bijlage 2: t-tabel.....	109
	Bijlage 3: Chi-kwadraattabel	110
	Bijlage 4: Formularium.....	111
	Bijlage 5: Oplossingen C-oefeningen	113
	1. Univariate beschrijvende statistiek	113
	2. Bivariate beschrijvende statistiek	128
	3. Inductieve statistiek	142

DEEL A:
VOORBEREIDENDE OEFENINGEN

1. Univariante beschrijvende statistiek

1.1. Basisconcepten

1

Greenpeace is bezorgd over het walvissenbestand in de Atlantische Oceaan. Om na te gaan hoe de toestand op dit moment is, neemt men een representatieve steekproef uit alle walvissen die daar leven en gaat men na tot welk continent het deel van de Oceaan hoort waaruit de walvis in kwestie stamt en hoe oud de walvis is. Men is ook geïnteresseerd in het aantal kwetsuren dat elke walvis opgelopen heeft.

- Wat is de populatie in bovenstaand onderzoek?
- Wat is de statistische eenheid in bovenstaand onderzoek?
- Wat is het meetniveau van de variabele "leeftijd"?
- Wat is het meetniveau van de variabele "continent"?
- Wat is het meetniveau van de variabele "soort walvis"?

2

Steeds meer jongeren grijpen naar wapens om zich veilig te voelen op school. Dat blijkt uit een enquête gehouden bij 239 leerlingen van verschillende Antwerpse scholen. 17 procent van de ondervraagde jongens en 7 procent van de meisjes tussen 14 en 17 jaar komt er openlijk voor uit al eens een wapen mee naar school te nemen. Jongens hebben meestal een mes op zak, meisjes een busje pepperspray.

- Wat is de populatie in bovenstaand onderzoek?
- Wat is de statistische eenheid in bovenstaand onderzoek?
- Wat zijn de variabelen in bovenstaand onderzoek? Welke mogelijke waarden hebben deze variabelen?
- Op welk meetniveau zijn deze variabelen gemeten?

1.2. Frequentieverdelingen

1

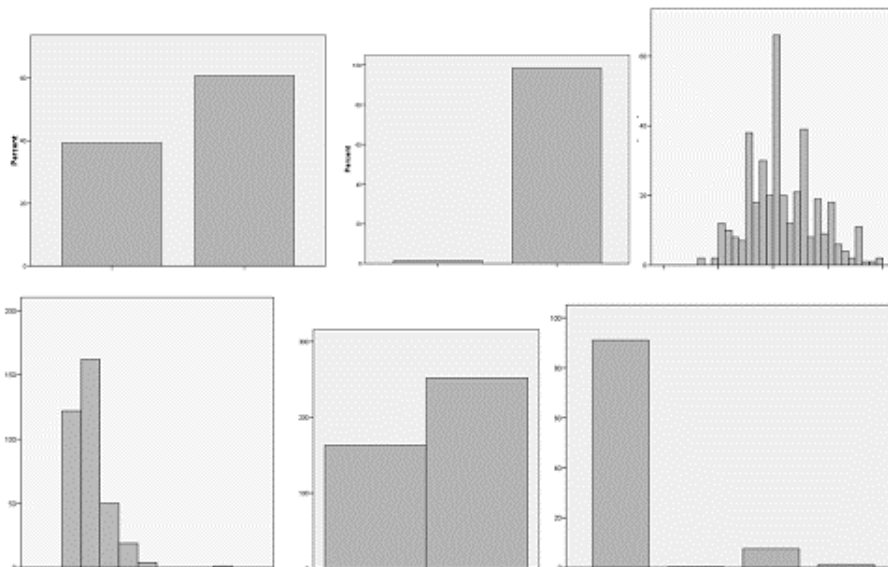
Aan de studenten uit de faculteit PSW werd gevraagd hoe vaak ze harddrugs gebruiken. Onderstaande tabel geeft de absolute frequenties weer.

x_i	f_i	Cum f_i	p_i	Cum p_i
Nooit	276			
Zelden	8			
Af en toe	4			
Dikwijls	2			
Zeer vaak	1			

- Bepaal de populatieomvang.
- Vul de tabel aan met cumulatieve frequenties, proporties en cumulatieve proporties.
- Hoeveel studenten nemen dikwijls harddrugs? Hoeveel procent is dat?
- Hoeveel procent van de studenten neemt zelden of nooit harddrugs?
- Welke proportie van de studenten neemt minstens dikwijls harddrugs?

2

Verbind de verschillende vragen (gesteld aan studenten aan faculteit PSW) met de juiste grafieken. Leg uit waarom.



- a) Wat is je geslacht?
- b) Hoeveel broers heb je?
- c) Heb je een functiebeperking?
- d) In welke onderwijsvorm zat je in het zesde middelbaar?
- e) Welke score behaalde je op het einde van het zesde middelbaar?

1.3. Maten van positie

1

Herneem de opgave van oefening A.1.2.1.

- a) Bepaal de modus.
- b) Bepaal de mediaan.

2

Tijdens de Gentse Feesten hebben de organisatoren van de Duveltent op het Sint-Baafsplein aan 30 random gekozen bezoekers gevraagd hoeveel Duvels ze op een gemiddelde feestavond consumeren. Onderstaande tabel geeft de resultaten weer.

Respondent	Aantal duvels
1	1
2	5
3	8
4	4
5	6
6	9
7	1
8	4
9	3
10	0
11	1
12	2
13	8
14	4
15	5

Respondent	Aantal duvels
16	4
17	3
18	2
19	0
20	7
21	8
22	1
23	8
24	5
25	6
26	1
27	2
28	4
29	8
30	0

- a) Stel een frequentietabel op (zonder klassen)
- b) Bepaal het rekenkundig gemiddelde
- c) Maakte je deze berekening van het gemiddelde op basis van de datamatrix of op basis van de frequentietabel? Waarom?

1.4. Maten van spreiding

1

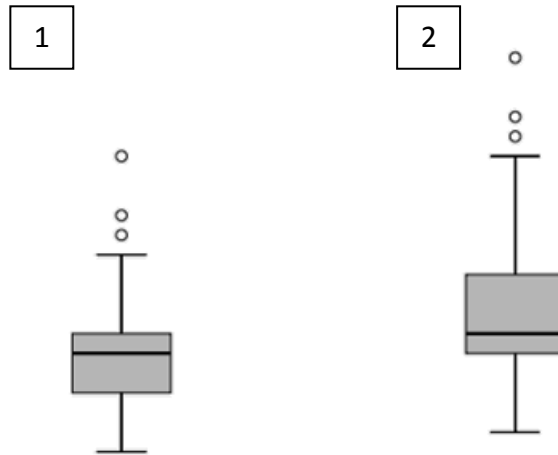
In een Vlaams onderzoek naar seksualiteit bij jongeren werd gepeild naar hun eerste seksuele ervaring. Onderstaande frequentietabel geeft aan hoe oud de verschillende respondenten waren bij hun eerste keer seks.

Leeftijd	f_i
10	4
11	4
12	7
13	15
14	29
15	50
16	77
17	54
18	47
19	19
20	10
21	5
22	1
23	1
24	1
25	2

- a) Bereken de relevante maten van centraliteit
- b) Welke maat van centraliteit die je berekende is het minst resistent voor uitschieters?
- c) Bereken het eerste en derde kwartiel
- d) Kan je hier de interkwartielafstand berekenen? Zo ja, bepaal op basis hiervan wat de uitschieters zijn.

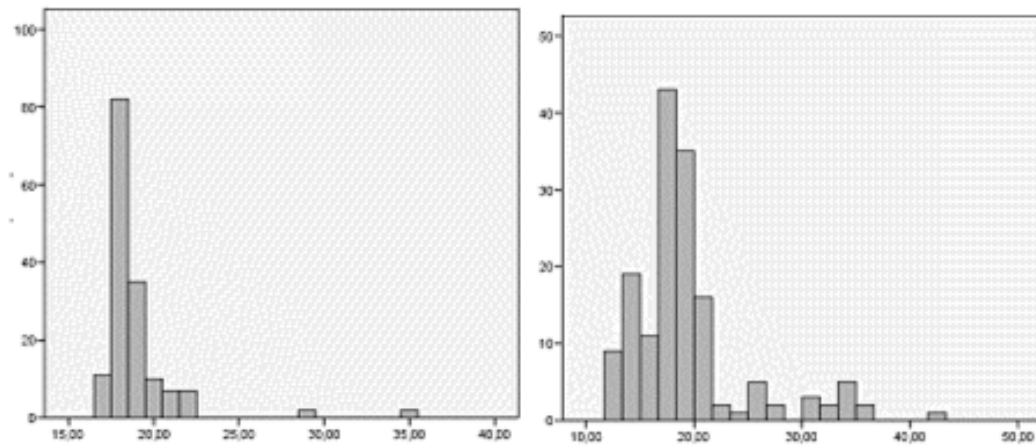
2

Boxplot 1 geeft een visuele weergave van de verdeling van de lonen van de 10 best betaalde voetballers. In boxplot 2 zie je de verdeling voor speler 11-20 in de ranking. Bij welke groep is de spreiding het grootst?



3

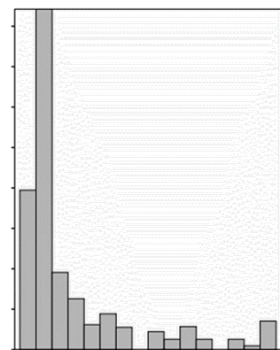
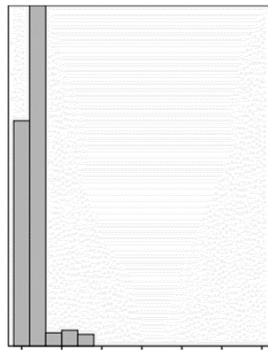
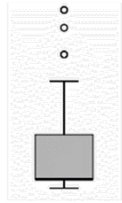
Bepaal met het blote oog waar de standaardafwijking het grootst zal zijn.



1.5. Vorm van een verdeling

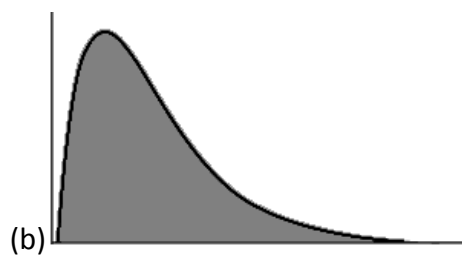
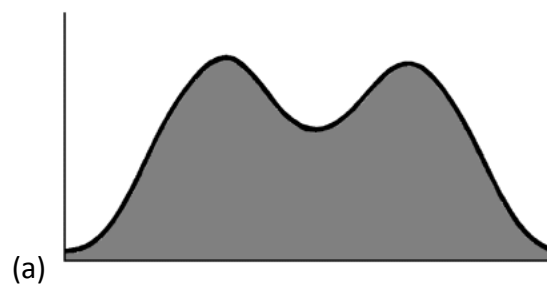
1

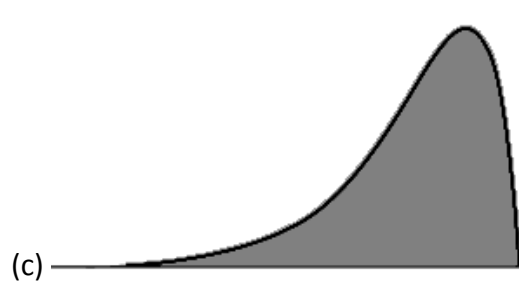
Welke boxplot hoort bij welk histogram?



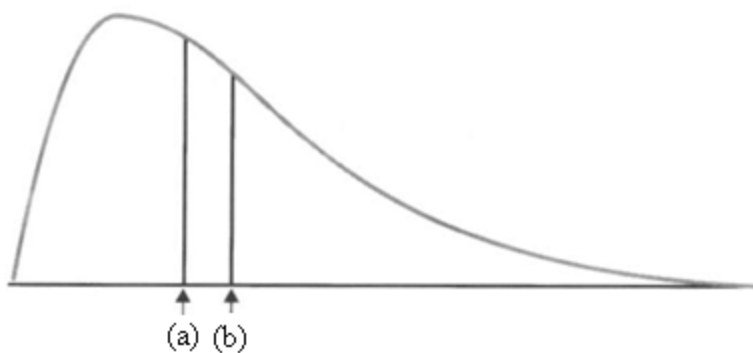
2

De examencommissie stelt vast dat de resultaten op het vak statistiek rechts a-symmetrisch verdeeld zijn. Welke van onderstaande figuren past hierbij?



**3**

Is punt (a) in onderstaande figuur de mediaan of het gemiddelde? Waarom?



1.6. Transformaties van variabelen

1

Het gemiddelde examenresultaat op het examen statistiek bedraagt 9,3 op 20 met een variantie van 2,4.

- Stel dat er een fout in het examen zat, en de studenten elk 2 punten extra verdienen. Wat gebeurt er met het gemiddelde? En met de variantie?
- Stel dat de bekendmaking van de examenresultaten op 100 gebeurt. We moeten alle aangepaste scores dus vermenigvuldigen met een factor van 5. Wat gebeurt er nu met het gemiddelde? En met de variantie?

1.7. Dichtheidskrommen

1

Het aantal dagen dat het in België per jaar bewolkt is, is normaal verdeeld volgens de volgende verdeling: $N(266, 16)$

- a) Wat is de gemiddelde verwachting?
- b) Wat is de modus?
- c) Wat is de mediaan?
- d) Wat is de standaardafwijking?
- e) Wat is de standaardnormaalverdeling die bij deze verdeling hoort?
- f) Stel dat het in een bepaald jaar 240 dagen bewolkt was. Bereken de bijhorende z-score en interpreteer.
- g) Bereken het percentage jaren met 240 of minder bewolkte dagen.
- h) Bereken het percentage jaren met meer dan 240 bewolkte dagen.
- i) Standaardiseer een concrete waarneming van 270 bewolkte dagen.
- j) Bereken het percentage jaren met 270 of minder bewolkte dagen.
- k) Bereken het percentage jaren met meer dan 270 bewolkte dagen.
- l) In welk percentage van alle jaren is het tussen 240 en 270 dagen bewolkt?

2. Bivariate beschrijvende statistiek

2.0. Studie van samenhangen en effecten

1

Is de relatie tussen volgende variabelen symmetrisch of asymmetrisch? Welke is de afhankelijke en welke de onafhankelijke variabele?

- a) Ernst van het gepleegde delict en aantal jaren gevangenisstraf
- b) Aantal jaren gevangenisstraf en kans op recidivisme na vrijlating
- c) Opleidingsniveau moeder en opleidingsniveau vader
- d) Opleidingsniveau leerling en latere beroepskansen
- e) Score op het examen sociologie en score op het examen statistiek

2.1. Relaties tussen categorische variabelen

1

Volgende kruistabel geeft, naargelang opleidingsniveau, aan in welke mate men het eens is met de uitspraak "Men moet steeds zijn eigen belang nastreven en zich niet te veel van anderen aantrekken".

	Diploma			
	Geen/LO	LSO	HSO	Hogeschool/Univ
Helemaal oneens	10	26	42	72
Oneens	68	132	193	186
Noch eens, noch oneens	39	66	81	67
Eens	88	62	74	33
Helemaal eens	22	11	14	0

- Stel de marginale verdelingen op.
- Hoeveel procent van de totale steekproef is het "oneens" met de uitspraak?
- Welk percentage van de hoogst opgeleide groep is het "helemaal oneens" of "oneens" met de uitspraak?
- Bereken het percentageverschil tussen de hoogst opgeleide groep en de laagst opgeleide groep wat betreft het "eens" zijn.
- Hoeveel procent van diegenen die het "helemaal eens" zijn behoren tot de categorie "LSO" of lager?
- Veronderstellen we hier een asymmetrische of een symmetrische relatie?

2

Na het Belgisch kampioenschap vinkenzetten stelt de voorzitter van de landelijke vinkeniersbond een kruistabel op met het aantal reglementair gezongen 'ring-ting-ting suskewieten' per vinkensoort in de half uur durende wedstrijd.

Vinkensoort	Aantal 'ring-ting-ting suskewieten'				Totaal (n)
	0 tot 40	41 tot 80	81 tot 120	121 tot 160	
Boomgaardvink	6		10	37	63
Oosterse vink		34	7	5	51
Groenling	10	17	8		38
Sijs	17	22	20	8	67

- a) Vul de ontbrekende frequenties aan
- b) Bereken de rij- en kolompercentages
- c) Hoeveel procent van de deelnemers heeft 81 tot 120 'ring-ting-ting suskewieten' gezongen?

3

In het jaarrapport van de federale politie vinden we nog andere resultaten terug die een belangrijk geslachtsverschil aan de kaak stelt. Hieronder vind je een kruistabel die nagaat of er een geslachtsverschil bestaat in het ooit gebruikt hebben van softdrugs.

	Geslacht	
	Man	Vrouw
Drugs gebruikt		
Ooit gebruikt	410	472
Nooit gebruikt	623	1263

- a) Bepaal voor mannen de odds op 'ooit gebruikt' t.o.v. 'nooit gebruikt'
- b) Bepaal voor vrouwen de odds op 'ooit gebruikt' t.o.v. 'nooit gebruikt'
- c) Bepaal nu de oddsratio mannen ten opzichte van vrouwen én de oddsratio vrouwen ten opzichte van mannen + interpreteer

4

Farid werkt voor het steunpunt Gelijke Kansen en onderzoekt of er een significant verband bestaat tussen actief zijn op de arbeidsmarkt en een partner hebben. Hij verzamelt onderstaande gegevens.

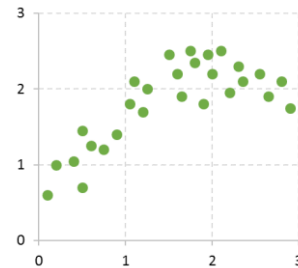
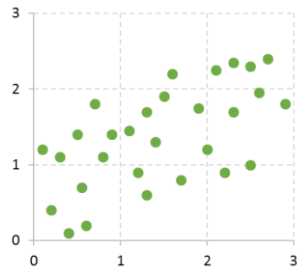
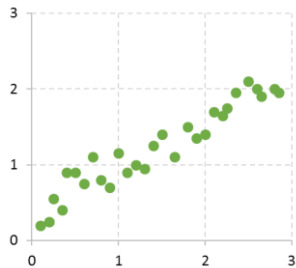
		Een partner hebben	
		Ja	Nee
Actief op de arbeidsmarkt	Ja	70	21
	Nee	12	68

- a) Welke tabel zou Farid uitkomen mocht er geen verband bestaan tussen een partner hebben en de activiteitsgraad op de arbeidsmarkt?
- b) Bereken de associatiemaat chi-kwadraat.

2.2. Relaties tussen metrische variabelen

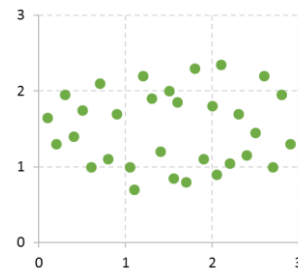
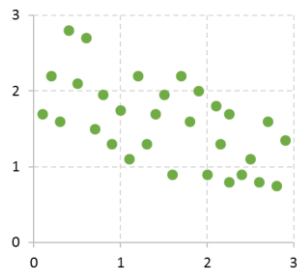
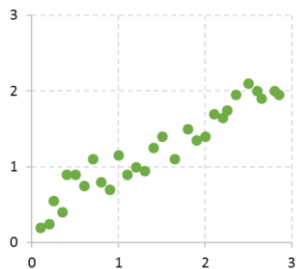
1

Bekijk de drie onderstaande scatterplots. Geef voor elke scatterplot aan of de data een lineaire samenhang vertoont.



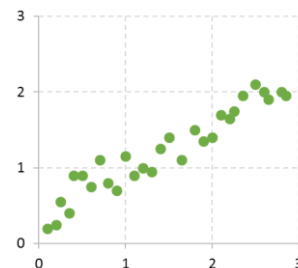
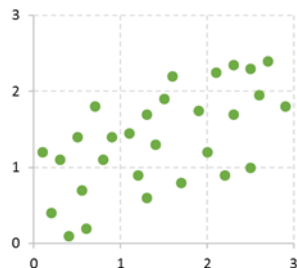
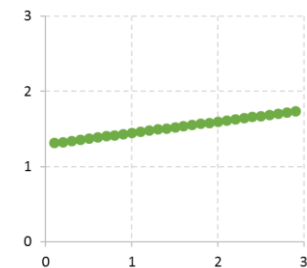
2

Beschouw de drie onderstaande scatterplots. Geef voor elke scatterplot aan of er sprake is van een positieve of negatieve lineaire samenhang. Baseer je hiervoor enkel op de grafische voorstelling; er zijn geen berekeningen nodig.



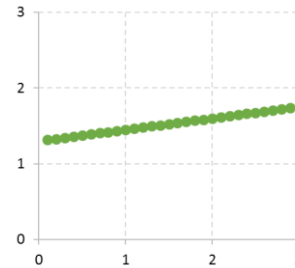
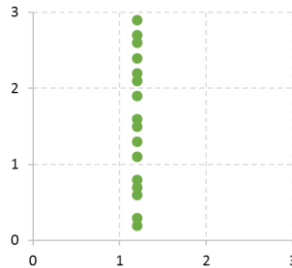
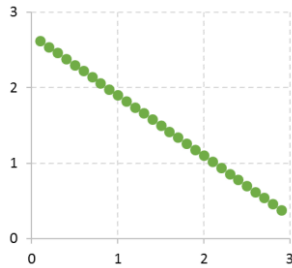
3

Beschouw de drie onderstaande scatterplots. Rangschik hen in functie van de grootte van de correlatie: de scatterplot met de kleinste correlatiecoëfficiënt eerst; de grootste laatst. Baseer je hiervoor enkel op de grafische voorstelling; er zijn geen berekeningen nodig.



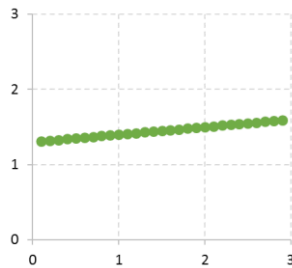
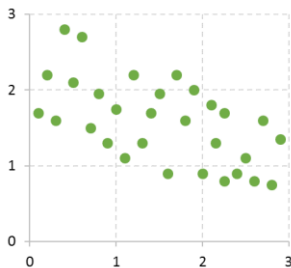
4

Bekijk de drie onderstaande scatterplots. Duid voor elke scatterplot aan of het om een zwak of sterk effect gaat.



5

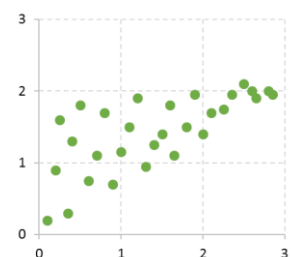
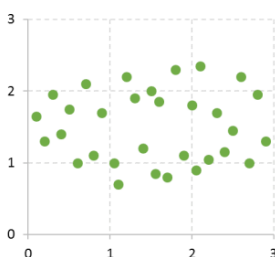
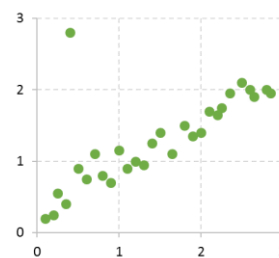
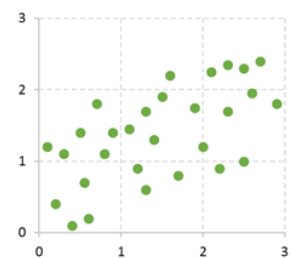
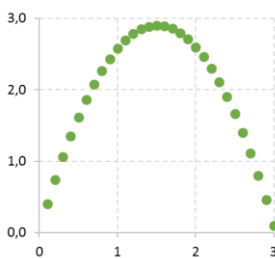
Beschouw de twee onderstaande scatterplots. Duid aan of onderstaande stellingen juist of fout zijn en geef aan waarom.



- a) Scatterplot 1 toont een sterker verband dan scatterplot 2.
- b) Scatterplot 1 toont een sterker effect dan scatterplot 2.

6

Bekijk de vier onderstaande scatterplots. Bepaal voor elke scatterplot of we een lineaire enkelvoudige regressieanalyse kunnen toepassen.



2.3. De limieten van de beschrijvende bivariate statistiek

1

Teken het correcte pijltjesdiagram (interactie-effect, gemeenschappelijke oorzaak, verstrengeling of onrechtstreeks verband) bij volgende onderzoeksresultaten...Indien je twijfelt tussen meerdere diagrammen, teken dan alle mogelijkheden.

- a) Bij onderzoek van kinderen op een basisschool, in de leeftijd van 6 tot 11 jaar, werd een sterk verband gevonden tussen de leesvaardigheid en de schoenmaat. Leg dit verband uit adhv een verborgen variabele
- b) Studenten uit eerste kandidatuur die vaker oefeningen bijwoonden voor een bepaald vak, blijken op het einde van het academiejaar hogere punten te halen. Voor een deel blijkt dit positief verband echter verklaard te kunnen worden door de hogere motivatie van deze studenten: meer gemotiveerd zijn zorgt er niet alleen voor dat men meer studeert, maar ook dat men vaker naar de oefeningen komt.
- c) Een onderzoek heeft aangetoond dat vrouwen die bij de productie van computerchips werken een abnormaal hoog aantal miskramen hebben. De vakbond beweert dat de blootstelling aan bij de productie gebruikte chemicaliën de miskramen veroorzaakt. Verduidelijk deze relaties in een diagram.
- d) Er is een negatieve correlatie tussen de verkoop van ijsjes en het aantal gevallen van griep in een bepaalde week van het jaar. Het is echter onwaarschijnlijk dat ijs griep voorkomt. Teken een diagram om deze relatie te verklaren.
- e) Er blijkt een positief verband te bestaan tussen het inkomen en het aantal boeken dat mensen op hun boekenrek hebben staan. Niet alleen heeft men meer geld ter beschikking om boeken te kopen, maar ook het opleidingsniveau blijkt hier een zeer belangrijke rol te spelen: hoger opgeleiden lezen en verdienen doorgaans ook meer dan lager opgeleiden.
- f) Wat wordt het pijltjesdiagram indien je er in vorige opgave van zou uitgaan dat inkomen op zich geen enkel effect heeft op het aantal boeken dat men koopt? Dat het effect enkel en alleen aan het opleidingsniveau te wijten is?
- g) Eerstgeboren jongens zouden meer creatief zijn dan jongens die later in de rij komen terwijl eerstgeboren meisjes net minder creatief zouden zijn dan meisjes die als tweede, derde, ... kind in een gezin werden geboren.

3. Inductieve statistiek

3.1. Studie van de systematiek van het toeval: kansrekenen

3.2. Stochastische variabelen

3.3. Studie van steekproefvariabiliteit

1

De rector van onze Alma Mater laat onderzoeken wat de impact is van de toelatingsexamens bij de Faculteit Geneeskunde. De onderzoekers ontdekken dat de scores op deze examens slechts grofweg normaal verdeeld zijn, volgens $N(20,8; 4,8)$.

- Hoe groot is de kans dat een enkele, uit de gehele examenpopulatie willekeurig gekozen leerling, een score behaalt van 23 of hoger?
- Neem nu een EAS van 25 leerlingen die het toelatingsexamen aflegden. Wat is het gemiddelde en de standaardafwijking van de gemiddelde steekproefscore?
- Wat is de kans dat deze gemiddelde steekproefscore hoger is dan 23?
- Welk van je schattingen is het meest correct? Die uit a) of die uit c)? Waarom is dat zo?

2

Lilian, een vrouw van 42, is bang dat ze diabetes heeft en gaat naar de dokter om zich te laten onderzoeken. De dokter geeft haar een suikerhoudend drankje en een uurtje later trekt hij bloed om het glucosegehalte te meten. Dit glucosegehalte varieert natuurlijk (dus als de patiënt geen diabetes heeft) volgens $N(125, 10)$ en diabetes wordt vastgesteld wanneer dit glucosegehalte hoger is dan 140.

- Hoe groot is de kans, bij één enkele glucosebepaling, dat Lilian te horen krijgt dat ze diabetes heeft, terwijl ze dat in principe niet heeft?
- Hoe groot is diezelfde kans wanneer de dokter Lilian op vier willekeurige dagen laat komen, telkens een meting doet en het gemiddelde glucosegehalte met de drempelwaarde vergelijkt?

3.4. Basistools voor de inductieve statistiek

1

Uit de studentenbarometer, afgenomen bij 3709 studenten, blijkt dat de gemiddelde score op het einde van het 6^{de} middelbaar 72 op 100 bedraagt. De standaardafwijking in de populatie bedraagt 8.

- a) Wat is de steekproefgrootte?
- b) Wat is de standaardafwijking van de steekproevenverdeling?
- c) Het 99%-betrouwbaarheidsinterval bedraagt [71,66; 72,34]. Interpreteer dit interval.
- d) Welke kritieke z-score hebben we nodig voor een 95%-betrouwbaarheidsinterval? Interpreteer.
- e) Bereken nu zelf een 95%-betrouwbaarheidsinterval.
- f) Vergelijk deelvraag e met deelvraag c: wat gebeurt er met het interval als de betrouwbaarheid van de schatter kleiner is?
- g) Stel dat we niet de tijd en de middelen hebben om de studentenbarometer bij 3709 studenten te doen, maar slechts een steekproef kunnen trekken van $n = 100$. We bekomen hetzelfde gemiddelde. Zou het 95%-betrouwbaarheidsinterval dan groter of kleiner zijn dan hetgeen bekomen in deelvraag e?

2

Stel de nulhypothese en de alternatieve hypothese op. Interpreteer.

- a) Het censusbureau heeft berekend dat de Belg gemiddeld 1500 euro spendeert aan reizen. We verwachten dat gepensioneerden een ander uitgavenpatroon vertonen. Je wenst dit te testen en neemt een EAS van 150 gepensioneerden.
- b) De gemiddelde tijd om een marathon te lopen bedraagt ongeveer 183 minuten. Men veronderstelt dat rapmuziek een positief effect heeft en test dit a.d.h.v. een steekproef van 12 hardlopers die naar rapmuziek luisteren. Uit de resultaten blijkt dat ze de marathon in gemiddeld 171 minuten afleggen.
- c) Koekjesfabriek LA boekte de voorbije twee jaar een gemiddelde winst van 7000 euro per maand. Dit jaar sluiten ze af met een gemiddelde van 7500 euro per maand, met een standaardafwijking van 175. Men vermoedt daarom dat dit jaar een beter boekjaar was dan de voorbije twee jaar, maar is dit effectief zo?

- d) Professionele gamers van over heel de wereld verdienen gemiddeld 50.000 euro per jaar. Een onderzoeksteam van de faculteit PSW wou graag weten of Vlaamse gamers ook zoveel verdienen met het streamen van hun spel. Uit een EAS bij 30 Vlaamse gamers blijkt dat zij gemiddeld 40.000 euro verdienen. Gaat het hier over een significant verschil?

3.5. Inferentie voor verwachtingen

1

Voor Opel zijn het moeilijke tijden. Ze moeten beslissen welke fabriek in Europa zijn deuren moet sluiten. Ze berekenen dat de gemiddelde tijd om een auto te maken in de Europese fabrieken 283 minuten bedraagt. Een steekproef van 20 auto's toont aan dat de fabriek van Opel Antwerpen voor het maken van een auto gemiddeld 287 minuten nodig heeft, met een standaardafwijking van 8,111.

- Kan je hier de z-procedure gebruiken? Waarom wel/niet?
- Geef de steekproevenverdeling.
- Bepaal het aantal vrijheidsgraden van de steekproevenverdeling.
- Stel dat het Europese gemiddelde niet gekend is. Bouw een 80%-betrouwbaarheidsinterval.
- Verschilt dit steekproefresultaat significant van het Europese gemiddelde?
- Indien 8,111 minuten de standaardafwijking van de populatie is in plaats van de standaardafwijking van de steekproef, wat gebeurt er dan met het interval en de overschrijdingskans?

2

Tourbaas Christian Prudhomme verwacht dat EPO een positieve invloed heeft op de prestaties van renners. Om dit te testen, werd de gemiddelde rittijd van de tijdrit Marseille naar de Mont Ventoux van renners die positief testten op EPO vergeleken met de gemiddelde rittijd van renners die negatief testten. Het gaat telkens om een EAS. De 137 renners die negatief testten hadden een gemiddelde rittijd van 169 minuten, met een standaardafwijking van 3,97 minuten, terwijl de 31 renners die positief testten een gemiddelde rittijd haalden van 167 minuten, met een standaardafwijking van 2,40 minuten.

- Welk soort significantietoets moeten we hier toepassen? Inferentie voor ...
- Kan je hier de z-procedure gebruiken? Waarom wel/niet?
- Wat is de verwachting van de steekproevenverdeling?
- Wat is de standaardafwijking van de steekproevenverdeling?

- e) Bepaal het aantal vrijheidsgraden van de steekproevenverdeling.
- f) Stel de hypothesen op.
- g) Verschillen deze groepen significant van elkaar?

3.6. Inferentie voor fracties

1

In de gehele populatie van Vlamingen bedraagt het stemmenpercentage voor de Open VLD 12%. In een verkiezingspeiling bij 1500 Oost-Vlamingen geven 128 Vlaamse kiezers aan dat ze van plan zijn om te stemmen op Open VLD.

- a) Welk soort significantietoets moeten we hier toepassen? Inferentie voor ...
- b) Gebruik je hier de z-procedure of de t-procedure? Waarom?
- c) Bepaal de steekproeffractie en interpreteer.
- d) Bepaal de standaardafwijking voor een betrouwbaarheidsinterval.
- e) Stel dat het Vlaamse percentage niet gekend is. Bouw een 90%-betrouwbaarheidsinterval op.
- f) Bepaal de steekproevenverdeling voor een significantietoets.
- g) Verschillen de Oost-Vlamingen significant van de hele Vlaamse bevolking op het 5%-niveau?

2

De deurwaarder die toezicht moet houden op 'Euromillions' vermoedt dat Belgen bevoorrecht worden in het spelletje. Hij weet dat ongeveer 3 % van de Europeanen die al eens meededen aan het spelletje, ook effectief iets won. Hij voert een EAS uit onder de Belgische deelnemers. De resultaten geven aan dat 534 van de 14365 Belgen in de steekproef ooit eens iets gewonnen hebben.

- a) Welk soort significantietoets moeten we hier toepassen? Inferentie voor ...
- b) Gebruik je hier de z-procedure of de t-procedure? Waarom?
- c) Bepaal de steekproeffractie.
- d) Bepaal de standaardafwijking voor een betrouwbaarheidsinterval.
- e) Stel dat het Europese percentage niet gekend is. Bouw een 85%-betrouwbaarheidsinterval.
- f) Bepaal de steekproevenverdeling voor een significantietoets.
- g) Verschillen de Belgische deelnemers significant van alle Europeanen?

3.7. Inferentie voor kruistabellen

1

Er wordt al een maand geprotesteerd tegen de toekenning van een milieuvergunning aan een multinational voor de uitbreiding van diens filiaal in waardevol bosrijk gebied. Nu raakt echter bekend dat er een verband bestaat tussen het al dan niet uitreiken van een milieuvergunning en het al dan niet hebben van een vriendschappelijke relatie met de minister van Milieu ($\chi^2 = 9,12$). De minister wordt in de media aan de schandpaal genageld en overal wordt opgeroepen tot ontslag. Maar is alle commotie terecht? Gaat het hier effectief over een significant verband?

- a) Bepaal het aantal vrijheidsgraden.
- b) Voer de significantietoets uit. Moet de minister-president zijn minister de laan uitsturen?

DEEL B: BEGELEIDE OEFENINGEN

1. Univariate beschrijvende statistiek

1.1. Basisconcepten

1

Op welk meetniveau werden volgende variabelen telkens gemeten? Geef bij metrische of kwantitatieve variabelen ook aan of het om een discrete of continue variabele gaat.

a)

Studentenbarometer 2008-2009

33. /50. Alcohol II

1. In welke mate ga je akkoord dat drinken:

	Helemaal akkoord	Akkoord	Neutraal	Niet akkoord	Helemaal niet akkoord
1. Je ontspant na de schooluren	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. De schoolstress doet dalen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Helpt schoolproblemen te vergeten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. De werkdruk doet afnemen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Helpt relaxen en ontspannen tijdens het weekend	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. Je helpt op te beuren als je depressief of verdrietig bent	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

2. Gedurende de voorbije 7 dagen, hoeveel glazen bier, wijn en sterke drank heb je gedronken?

1. Bier

2. Wijn

3. Sterke drank

b)

Studentenbarometer 2008-2009

2. /50. Demografische achtergrond

1. Geslacht

mannelijk

vrouwelijk

2. Ik ben geboren in het jaar...

3. Hoeveel broers en zussen heb je (inclusief adoptie- en stiefkinderen)?

1. Aantal broers

2. Aantal zussen

4. Heb je een functiebeperking?

Ja, ik heb een functiebeperking

Nee, ik heb geen functiebeperking

c)

Studentenbarometer 2008-2009**16. /50. Woonsituatie****1. Waar verblijf je tijdens de week?**

- Op een studentenhome van de universiteit
- In een kamer in een private woning
- In een studio of appartement
- Ik huur een huis of appartement samen met anderen
- Thuis

2. Wat is het postnummer van de plaats waar je gedomicileerd bent?

d)

Studentenbarometer 2008-2009**4. /50. Middelbare studies****1. In welke onderwijsvorm zat je in het zesde middelbaar?**

- algemeen secundair onderwijs (ASO)
- kunstsecundair onderwijs (KSO)
- technisch secundair onderwijs (TSO)
- beroepssecundair onderwijs (BSO)

2. Welk totaalpercentage behaalde je op het einde van het zesde middelbaar?

e)

Levensovertuiging	Frequentie
Gelovig en praktiserend katholiek	1
Gelovig en kerkgaand	22
Protestants	2
Iemand die twijfelt maar toch min of meer christelijk	28
Islamitisch religieus	11
Vrijzinnig	27
Iemand die zich niet voor religie interesseert	17
Niet gelovig	43
Anders	7

2

Als voorbereiding op een vergadering beschouwt een Europees parlamentslid de nationale criminaliteitscijfers van de lidstaten van de Europese Unie (deze variabele wordt gemeten als het aantal vastgestelde criminele delicten per 1000 inwoners). Welke uitspraken zijn juist en welke zijn fout?

- a) De bestudeerde populatie bestaat uit de landen die deel uitmaken van de Europese Unie
- b) De bestudeerde populatie bestaat uit inwoners van landen die deel uitmaken van de Europese Unie
- c) De variabele 'criminaliteitsgraad' is een kwalitatieve of categorische variabele
- d) Er wordt gewerkt met een steekproef van 1000 inwoners uit ieder Europees land

1.2. Frequentieverdelingen**1**

De feestpraeses van een studentenvereniging in de Gentse Overpoort wil een nieuwe bestelling bier plaatsen. De secretaris van de vereniging maant hem echter aan om niet teveel in te slaan: het zou een overstock opleveren. De feestpraeses verzamelt onmiddellijk gegevens in van een steekproef, die hij trok uit de populatie schachten die de vereniging vervoegden. De tabel met resultaten vind je hieronder terug.

- a) Is deze tabel een datamatrix of een frequentietabel?
- b) Wat is de variatiebreedte?
- c) Stel een frequentietabel op met 5 gelijke klassen.
- d) Stel een frequentietabel op met 4 klassen. Kunnen we klassen maken met gelijke klassenbreedte? Waarom (niet)?
- e) Vervolledig de tabel van de in 4 klassen ingedeelde data met cumulatieve frequenties, proporties, cumulatieve proporties, percentages en cumulatieve percentages.

Schacht	Glazen bier
1	5
2	6
3	11
4	3
5	2
6	8
7	6
8	24
9	2
10	5
11	9
12	8
13	10
14	11
15	2
16	3
17	16
18	14
19	3
20	0

1.3. Maten van positie

1

Tijdens de Gentse Feesten hebben de organisatoren van de Duveltent op het Sint-Baafsplein aan 30 random gekozen bezoekers gevraagd hoeveel Duvels ze op een gemiddelde feestavond consumeren. Onderstaande tabel geeft de resultaten weer.

- Verdeel de data in 4 exclusieve klassen.
- Wat is telkens het klassenmidden?
- Bereken het rekenkundig gemiddelde van deze data op basis van je klassenindeling.
- Hoe kan je het verschil tussen dit gemiddelde en de berekening zonder klassen (zie oefening A.1.3.1) verklaren? Welk gemiddelde correspondeert met de realiteit?

Respondent	Aantal duvels
1	1
2	5
3	8
4	4
5	6
6	9
7	1
8	4
9	3
10	0
11	1
12	2
13	8
14	4
15	5

Respondent	Aantal duvels
16	4
17	3
18	2
19	0
20	7
21	8
22	1
23	8
24	5
25	6
26	1
27	2
28	4
29	8
30	0

2

Een grote koffieketen heeft vorig jaar een nieuwe vestiging geopend in België. Nu wil de keten nagaan of het voordelig is nog een paar nieuwe vestigingen te openen. Ze vraagt de gegevens op van een willekeurig aantal gekozen dagen, waarbij expliciet de vraag gesteld wordt het totale aantal verkochte koffies op die dag op te lijsten, daarbij onderscheid makend tussen espresso, koffie verkeerd en cappuccino. De medewerkers boksen volgende tabel in elkaar, waarin, per koffiesoort, het aantal dagen weergegeven wordt waarop een bepaald aantal koffies verkocht wordt.

- Is dit een datamatrix of een frequentietabel?
- Bereken alle maten van centraliteit voor de drie groepen koffies.
- Wat is hier de meest relevante maat van centraliteit?
- Bereken voor elk van de drie groepen koffies de kwartielen en de interkwartielafstand. Zijn er outliers?

Aantal verkochte koffies	Aantal dagen		
	Espresso	Koffie verkeerd	Cappuccino
20	1	0	2
23	6	8	1
24	4	10	3
29	8	12	4
30	3	6	4
31	9	4	6
32	1	1	9
33	5	9	8
35	6	2	6
49	7	4	3
51	4	6	2
52	6	2	1
54	8	1	0
58	10	0	0
59	7	0	0
60	5	0	0
61	1	0	0

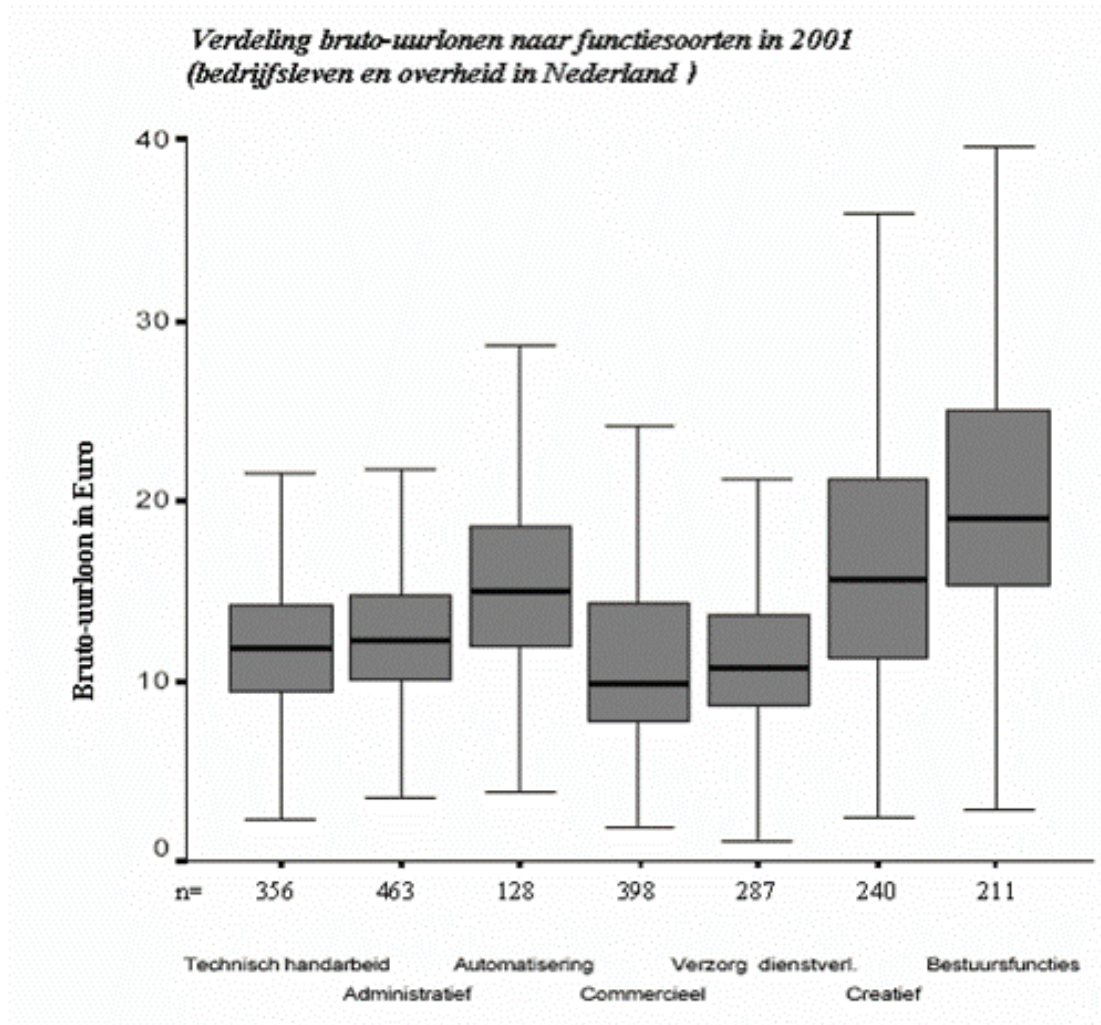
3

Onderstaande tabel geeft de resultaten weer van de studenten aan de faculteit PSW voor de vraag "Hoeveel dagen in de week ga je uit?" Stel de boxplot op voor deze tabel.

Waarde	f_i	Percentage
1	191	68,2
2	58	20,7
3	16	5,7
4	8	2,9
5	1	0,4
6	2	0,7
7	4	1,4

4

In onderstaande figuur worden zeven boxplots gegeven die informatie bevatten over het bruto uurloon van werknemers in Nederland in 2001, opgesplitst volgens het soort beroep dat ze uitoefenen. Per subgroep (per functiesoort) wordt onderaan de boxplot vermeld hoeveel personen tot deze categorie behoren. Bepaal aan de hand van deze figuur welke beweringen correct zijn.



- Binnen de categorie 'Verzorging en Dienstverlening' verdient ten minste 50% van alle personen meer dan 10 Euro per uur
- Ten minste 32 personen uit de categorie 'Automatisering' verdienen 15 Euro per uur of meer
- In de categorie 'Administratief' verdient niemand van de 25% laagst scorende personen minder dan 10 Euro per uur
- Ten minste 180 personen uit de categorie 'Creatief' verdienen 10 Euro per uur of meer
- In vergelijking met het percentage personen tussen het derde kwartiel en de mediaan binnen de categorie 'Commercieel' is het percentage mensen met een waarde tussen het derde kwartiel en de mediaan kleiner binnen de categorie 'Administratief'

5

In 2008 stelde het tijdschrift Forbes een lijst op met de 20 best betaalde voetballers ter wereld. Hieruit bleek dat de spelers die meer verdienen dan het derde kwartiel samen gemiddeld 21 360 000 euro verdienen. De spelers die minder binnenrijden dan het derde kwartiel, verdienen gemiddeld 10 880 000 euro. Bereken het gemiddelde "loon" voor de 20 best betaalde voetballers.

1.4. Maten van spreiding

1

Onderstaande tabel bevat de lonen van de 9 best betaalde voetballers ter wereld. Bereken alle maten van positie en spreiding. Interpreteer.

	Speler	Loon (in miljoen euro)
1	Lionel Messi	65
2	Cristiano Ronaldo	54
3	Neymar	36,5
4	Thiago Silva	27,5
5	Robin van Persie	25,6
6	Gareth Bale	23,8
7	Zlatan Ibrahimovic	21,5
8	Sergio Aguëro	21,5
9	Robert Lewandowski	20,2

2

Regi heeft sinds kort een Facebook account aangemaakt. Aangezien hij op zoek is naar een nieuw lief, heeft hij tijdens de eerste vijftien dagen bijgehouden hoeveel meisjes hem een vriendschapsverzoek verstuurden. Het resultaat hiervan vind je in onderstaande tabel.

- Bereken alle relevante maten van positie
- Bereken alle relevante maten van spreiding

Dag	Aantal vrouwelijke vriendschapsverzoeken
1	1
2	3
3	4
4	6
5	8
6	9
7	13
8	8
9	9
10	4
11	6
12	4
13	1
14	0
15	1

3

Bekijk de gegevens in de tabel. Bereken de best passende spreidingsmaat voor deze verdeling. Duid het correcte antwoord aan.

Aantal kamers in huis	Aantal gezinnen
8	54
6	69
9	30
7	50
10	14
5	10

- a) De best passende spreidingsmaat bedraagt 1,31
- b) De best passende spreidingsmaat bedraagt 1,24
- c) De best passende spreidingsmaat bedraagt 59
- d) We kunnen hier geen spreidingsmaat berekenen door een gebrek aan gegevens

1.5. Vorm van een verdeling

1

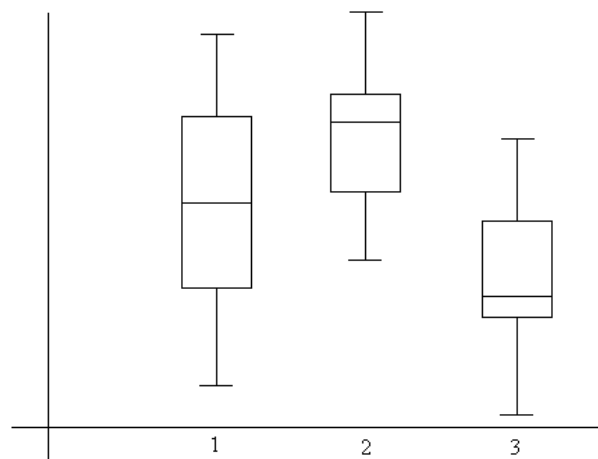
Aan 100 criminologiestudenten werd gevraagd hoe vaak ze softdrugs hebben gebruikt de afgelopen maand. Onderstaande frequentietabel geeft de resultaten weer.

- Geef deze verdeling weer in een staafdiagram.
- Is deze verdeling symmetrisch, rechts asymmetrisch of links asymmetrisch?

Softdrugs gebruik afgelopen maand	Frequentie
Geen	24
1 tot 2 keer	47
3 tot 5 keer	15
6 tot 10 keer	7
11 tot 20 keer	5
21 tot 30 keer	2

2

Bekijk de drie onderstaande boxplots:



- Welke boxplot heeft de grootste spreiding?
- Wat is de vorm van de verdelingen van de drie boxplots?

1.6. Transformaties van variabelen

1

De Belgische voetbalbond heeft een overzicht gemaakt van het aantal geïncasseerde doelpunten van de rode duivels bij de laatste 10 gespeelde matches. In onderstaande tabel worden deze waarden weergegeven.

2	3	2	4	2
0	3	2	2	0

- Bereken de variantie en de standaardafwijking van het aantal geïncasseerde goals van de Rode Duivels
- Veronderstel dat het Nederlandse elftal de laatste tien matches exact dubbel zoveel goals heeft geïncasseerd dan de Rode Duivels, wat is hun variantie dan?
- Stel dat de Rode Duivels in een extra oefenmatch tegen het Braziliaanse elftal 10 goals moeten incasseren. Welk gevolg heeft dit voor de standaardafwijking? Wordt ze groter of kleiner?

1.7. Dichtheidskrommen

1

Het gewicht van Vlaamse pasgeboren baby's is normaal verdeeld met een verwachting van 2,9 kg en een standaardafwijking die 0,3 kg bedraagt.

- Wat is de mediaan van het gewicht van pasgeboren baby's in Vlaanderen? En de modus?
- Hoeveel procent van de pasgeboren baby's in Vlaanderen weegt 2,3 kg of minder?
- Hoeveel procent van de Vlaamse baby's weegt bij de geboorte meer dan 3,7 kg?
- Hoeveel procent van de Vlaamse baby's weegt bij geboorte tussen 3,1 en 3,5 kg?

2

Kees Rijstboom, een medisch socioloog, heeft afgelopen zomer een onderzoek gedaan naar het geluidsniveau op alle Belgische festivals. Hij heeft vastgesteld dat het aantal decibels waaraan festivalgangers worden blootgesteld perfect normaal verdeeld is volgens $N(105, 7)$.

- a) Hoeveel procent van de festivalbezoekers wordt blootgesteld aan maximum 75 decibel, de veilige norm?
- b) Gehoorschade treedt op vanaf 85db. Hoeveel procent van de festivalbezoekers wordt blootgesteld aan schadelijke waarden?
- c) Kees stelt dat 92,4% van de festivalgangers de pijngrens niet overschrijdt. Hoeveel decibel bedraagt deze grens?
- d) Kees stelt dat 0,20% van de festivalgangers blijvende gehoorschade oplopen. Aan hoeveel decibel werden zij minstens blootgesteld?

3

Jean is brouwer en gaat dit jaar naar alle zomerfestivals. Hij wil weten hoe goed zijn brouwsel naar binnen gaat. Hij trekt daarom een steekproef uit alle bezoekers van alle festivals in Vlaanderen, en vraagt hen telkens hoeveel pintjes men op die dag gedronken heeft. Jean komt tot de vaststelling dat het aantal pintjes dat festivalgangers die dag drinken, perfect normaal verdeeld is, met een gemiddelde van 9 en een standaardafwijking van 3,36.

- a) Hoeveel procent van de festivalgangers drinkt maximum 1 pintje op de dag dat ze ondervraagd worden?
- b) Zware alcoholintoxicatie begint op te treden vanaf 15 pintjes. Bij hoeveel procent van de festivalgangers is dat het geval?
- c) Hoeveel bedraagt het eerste kwartiel van de verdeling? En het derde kwartiel?

4

Het gewicht van eieren is normaal verdeeld met een gemiddelde van 50 gram en een standaarddeviatie van 5 gram. De eieren worden onderverdeeld in drie categorieën. 'Klein' indien het gewicht lager is dan 46 gram, 'medium' indien het gewicht tussen 46 en 56 gram ligt en 'groot' indien ze meer dan 56 gram wegen. Op 'kleine' eieren is er geen winst: de verkoopprijs komt overeen met de kosten. Op 'medium' eieren wordt een winst gemaakt van 3 eurocent en op 'grote' eieren wordt een winst gemaakt van 5 eurocent per stuk. Bereken de gemiddelde winst per ei.

5

Flink Poyd, een Vlaamse charmezanger, is op weg om het zwaar te maken in de Verenigde Staten. Het kleine platenlabel Joshua Records weet de flamboyante zanger te strikken met de volgende regeling. Als Flink optreedt voor minder dan 2000 mensen, krijgt hij \$ 1000 per show. Treedt hij op voor 2000 tot 5000 mensen, dan ontvangt hij \$ 2000. Lukt het hem om voor meer dan 5000 mensen op te treden, dan krijgt hij maar liefst \$ 3000. Flink weet dat het aantal mensen dat naar zijn shows komt kijken, perfect normaal verdeeld is volgens $N(3100, 850)$.

- a) Op hoeveel procent van zijn shows komt er meer dan 2500 mensen kijken?
- b) Hoeveel mensen moeten er naar een optreden komen zodat het bij de 5% best verkopende shows hoort?
- c) Hoeveel geld verdient Flink Poyd volgens zijn contract gemiddeld per show?

2. Bivariate beschrijvende statistiek

2.0. Studie van samenhangen en effecten

1

Geef voor elke variabele een voorbeeld waar ze optreedt als onafhankelijke variabele én één waar ze optreedt als afhankelijke.

- a) Verkiezingsuitslag van de SP.a
- b) Inkomen van je ouders
- c) Proportie immigranten in de gevangenissen
- d) Islamofobie
- e) Geslacht
- f) Het al dan niet hebben van criminele vrienden

2.1. Relaties tussen categorische variabelen

1

Communicatiewetenschappers van de Universiteit Gent vragen zich af welke onbedoelde boodschappen reclameblokken zoal overbrengen. In het kader van het onderzoek werd aan respondenten gevraagd: "In welke mate vindt u dat reclame vrouwen en mannen op een stereotype wijze voorstelt?" Hieronder vind je de resultaten.

	Helemaal akkoord	Akkoord	Ik twijfel	Niet akkoord	Helemaal niet akkoord
Man	6	23	101	355	538
Vrouw	3	16	102	462	1130

- Bereken zowel de marginale verdeling als de rij- en kolompercentages
- Hoeveel procent van de mannen is "Helemaal akkoord" of "Akkoord"?
- Hoeveel procent van de totale steekproef twijfelt?
- Bepaal voor elke categorie het percentageverschil

2

De federale politie publiceert haar jaarlijks rapport over criminaliteit in België. Eén van de gepubliceerde tabellen vind je hieronder terug. De resultaten gaan na of er een geslachtsverschil is tussen mannen en vrouwen wat betreft het plegen van geweld tijdens een misdrijf.

	Man	Vrouw
Geweld	3266	210
Geen geweld	7491	1550

- Bereken zowel de marginale verdeling als de rij- en kolompercentages
- Hoeveel % van de mannen gebruikt geweld?
- Hoeveel procent van de totale steekproef gebruikt geweld?
- Bepaal het meest relevante percentageverschil en interpreteer.

3

De Vlaams minister voor Onderwijs bekijkt het jaarlijkse rapport van zijn departement. Een aspect van het rapport belicht het verschil tussen de onderscheiden onderwijstypes in het Vlaamse secundair onderwijs (ASO, KSO, TSO en BSO). Hieronder vind je de kruistabel die de verdeling van jongens en meisjes over die verschillende onderwijstypes weergeeft.

Studiejaar	Geslacht		Totaal
	Jongens	Meisjes	
ASO	2559	2977	5536
KSO	126	198	324
TSO	1843	1533	3376
BSO	1217	1390	2607
Totaal	5745	6098	11843

- a) Bereken de oddsratio voor meisjes tegenover jongens (ASO vs TSO)
- b) Bereken de oddsratio van meisjes tegenover jongens (TSO vs ASO)
- c) Bereken chi-kwadraat
- d) Bereken Cramer's V

4

De Verenigde Naties houden de ontwikkeling van derdewereldlanden nauwlettend in het oog. Met dat doel publiceert het tweejaarlijks een uitgebreide vergelijking tussen ontwikkelingslanden en ontwikkelde landen. Hieronder wordt een tabel weergegeven over analfabetisme. Om de resultaten te bekomen, werd een EAS getrokken uit de wereldbevolking en werd aan de geselecteerde personen gevraagd of ze in een ontwikkelingsland dan wel een ontwikkeld land wonen, en of ze analfabeet zijn of niet.

Analfabeet	Ontwikkelingsland	Ontwikkeld land
Ja	82	16
Nee	771	585

- a) Bereken χ^2 voor deze kruistabel
- b) Is het nodig Cramer's V te berekenen? Leg uit!
- c) Maak een tabel met de gestandaardiseerde residuen en leg uit wat deze betekenen.

5

Een onderzoeker veronderstelt dat het democratisch karakter van een land bepalend is voor de sociale gelijkheid in dat land. Je vindt de bruto gegevens waar hij mee werkt in onderstaande tabel.

- a) Bereken het meeste relevante percentageverschil.
- b) Bereken de oddsratio voor landen met een hoog democratisch karakter versus landen met een laag democratisch karakter, waarbij je hoge sociale gelijkheid vergelijkt met lage sociale gelijkheid.

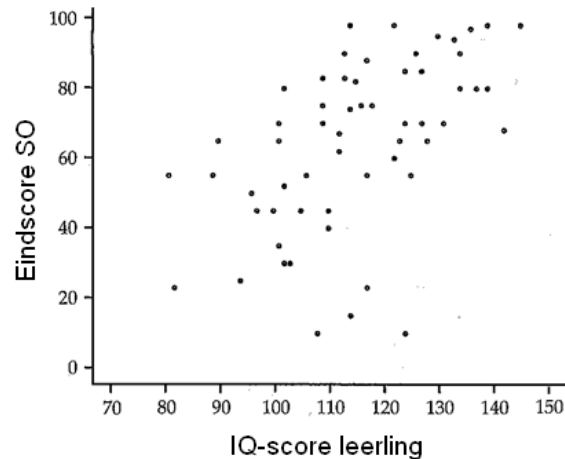
Land	Democratisch karakter van de staatsinrichting	Sociale gelijkheid
Lapland	Hoog	Hoog
Cyprus	Laag	Hoog
Ecuador	Laag	Laag
Columbia	Laag	Laag
Verenigde Emiraten	Hoog	Hoog
Guatemala	Laag	Laag
Noorwegen	Hoog	Hoog
Nederland	Hoog	Hoog
Paraguay	Laag	Hoog
Finland	Hoog	Laag
Trinidad	Laag	Laag
Japan	Laag	Laag
Estland	Hoog	Laag
Costa Rica	Hoog	Laag
Turkije	Laag	Hoog
Frankrijk	Hoog	Hoog
Portugal	Laag	Laag
Spanje	Laag	Hoog
Denemarken	Hoog	Hoog
Rusland	Hoog	Hoog

2.2. Relaties tussen metrische variabelen

1

De UGent voerde een onderzoek uit in het Vlaamse Secundair onderwijs onder 457 leerlingen uit het zesde jaar. De figuur hieronder toont ons de samenhang tussen de IQ-score van deze leerlingen en hun score op het einde van het jaar.

- Verwachten we hier een positieve of negatieve samenhang?
- Zijn er uitschieters? Zo ja, op welke manier wijken ze af van het globale patroon?
- Als we de uitschieters weglaten, is de samenhang dan bij benadering lineair?
- Wat kunnen we op zicht zeggen over de sterkte van het verband?
- Wat kunnen we op zicht zeggen over de sterkte van het effect?



2

Een studente van de faculteit PSW van Oxford en haar vriendinnen gaan op reis naar Barcelona. Daar aangekomen duiken ze het nachtleven in en de volgende morgen worden ze allemaal wakker met een nieuwe vakantie liefde naast zich. De studente besluit uit te zoeken of er een verband bestaat tussen hun eigen lichaamslengtes en die van hun nieuwe veroveringen. De gegevens lijtte ze op in de tabel hieronder (lengte in inches).

Vrouwen	66	64	66	65	70	65
Mannen	72	68	70	68	71	65

- Teken de scatterplot. Verwacht je dat de relatie positief of negatief is?
- Bepaal de covariantie en de correlatie tussen beide variabelen
- Wat gebeurt er met de correlatie indien alle mannen 6 inch kleiner geweest waren?
- Wat gebeurt er met de correlatie indien we de lengte uitdrukken in cm in plaats van inches? (1 inch is 2,54 cm)
- Mocht het nu zo zijn, dat elke vakantie liefde net 3 inches langer is dan de vrouw in kwestie, hoe ziet de correlatie er dan uit?
- Stel zowel de ongestandaardiseerde als de gestandaardiseerde regressie-vergelijking op
- Een vriendin komt wat later. Ze is 170,2 cm groot. Wat is de verwachte lengte van haar vakantie liefde als je weet dat 170,2 cm gelijk is aan 67 inches (1 inch = 2,54 cm)?

3

Het Braziliaanse vrouwenteam maakte een sterke indruk tijdens een belangrijke golfinterland. Hieronder vind je de resultaten per speler voor elke ronde.

	Spelernummer										
Ronde	1	2	3	4	5	6	7	8	9	10	11
Ronde 1	89	90	87	95	86	81	105	83	88	91	79
Ronde 2	94	85	89	89	81	76	89	87	91	88	80

- Teken de scatterplot. Er is één outlier. Welke?
- Hieronder krijg je twee regressievergelijkingen. Eén werd getekend met, de ander zonder de outlier. Teken beide lijnen in je scatterplot. Welke vergelijking werd getekend zonder de outlier? Hoe weet je dit?
- $\hat{y} = 20,49 + 0,754x$
- $\hat{y} = 50,01 + 0,410x$
- Welke regressievergelijking geeft het beste de data weer?

4

Een ruimtelijke planner gaat de associatie na tussen de rijkdom van een gemeente (uitgedrukt in hoeveel promille van het Bruto Nationaal Product dat een stad bezit) en het vruchtbaarheidscijfer van die stad (het aantal nieuwgeborenen per 100 inwoners van die stad). De planner berekent de regressievergelijking, die de associatie tussen beide variabelen weergeeft, als $\hat{y} = 15,24 - 1,42x$; de correlatie tussen beiden bedraagt $-0,52$ en de variantie van het vruchtbaarheidscijfer $6,66$. De gegevens van de 8 gemeenten vind je in onderstaande tabel.

Stad	Rijkdom	Vruchtbaarheidscijfer
Turnhout	7,2	3,7
Antwerpen	8,1	4,2
Bergen	7,4	3,6
Mechelen	5,3	4,2
Brugge	6,9	4,8
Charleroi	6,2	10,3
Gent	6	8
Luik	5,7	8,3

- Bepaal de verwachte waarden \hat{y}_i
- Bepaal de residuen e_i
- Bereken de variantie van deze residuen. Wat betekent dit cijfer?
- Bereken de verklaarde variantie

5

Professionele gamers verdienen miljoenen euro's door computerspelletjes te spelen. Een onderzoeksteam van de faculteit PSW wou graag weten hoe je spelend rijk kan worden, dus onderzochten ze of het aantal uren dat je speelt een invloed heeft op de opbrengst die je haalt uit het streamen van je spel. Een rondvraag bij de 6 beste Gentse gamers leverde de volgende gegevens per weekdag op (5 werkdagen in 1 week).

Aantal uren gamen	Opbrengst life stream	Bijkomende gegevens
11	405	$s_x^2 = 6,8$
8	327	$s_y^2 = 8604$
9	274	$\bar{x} = 9$
7	193	$\bar{y} = 288$
6	171	
13	358	

- Bereken de covariantie en de correlatie
- Stel de ongestandaardiseerde regressierechte op
- Van een student wordt verwacht dat hij/zij ongeveer 32 uur per week studeert. Als we in plaats van te studeren zouden gamen en streamen, hoeveel zouden we dan kunnen verdienen? Merk op dat 32 uur per week gelijk is aan 6,4 uur per weekdag
- Bereken en interpreteer de determinatiecoëfficiënt
- Bereken de residuele variantie aan de hand van de determinatiecoëfficiënt

2.3. De limieten van de beschrijvende statistiek

1

Teken het correcte pijltjesdiagram (interactie-effect, gemeenschappelijke oorzaak, verstrengeling of onrechtstreeks verband) bij volgende onderzoeksresultaten.

- a) Het Belgisch leger stelt vast dat er meer kalende soldaten zijn in de hoogste klassen (generaal, korporaal, etc) dan bij de laagste klassen.
- b) Het nemen van voedingssupplementen heeft een positief effect op spierontwikkeling. Dit effect is echter veel sterker voor mannen dan voor vrouwen.
- c) Het bijwonen van een rock- of popconcert is voor een deel te verklaren door inkomen. Mensen met een laag inkomen zullen minder vaak naar zo'n concert gaan. Maar opleiding heeft een minstens even sterk effect: mensen met een hogere opleiding gaan vaker naar een rock- of pop concert.
- d) Studenten uit West-Vlaanderen halen minder goede punten op de Gentse universiteit dan studenten uit Oost-Vlaanderen.
- e) Bedenk zelf een voorbeeld van een interactie-effect.

2

Politologen en sociologen slaan de handen in elkaar om een grootschalig onderzoek op te zetten naar attitudes in de bevolking omtrent de meest recente immigratiegolf in België en hoe de huidige politieke partijen daarop inspelen. Eén van de tabellen uit het lijvige onderzoeksrapport vind je hiernaast terug. Deze tabel presenteert de resultaten van 2 statistische modellen, waarbij de invloed van verschillende onafhankelijke variabelen op de afhankelijke variabele 'perceptie van bedreiging door migranten' (schaal 0-10) onderzocht wordt.

	Model 1	Model 2
	b	b
Leeftijd	0,023	0,019
Aantal jaren gestudeerd	-0,358	-0,318
Vertrouwen in politieke instellingen	-	-0,184
Vertrouwen administratie	-	-0,074
Vertrouwen gerecht	-	-,0129
Vertrouwen leger	-	0,089
R²	0,208	0,304
n	1554	1554

- Benoem en interpreteer de effecten uit model 1.
- Wat gebeurt er als de vertrouwens-variabelen aan het model worden toegevoegd?
- Welke model heeft de grootste verklarende kracht?

3

Veerle Maerlevalck, onderzoekster bij de UGent, zet een studie op poten die de kwaliteit van het schoolleven van adolescenten analyseert. Een belangrijke indicator van die kwaliteit is de mate waarin leerlingen betrokken zijn op hun school. Veerle gaat in een 1ste model na of de sociaaleconomische status (de economische situatie), het leerjaar (gaande van 1-6) en de studiepunten van leerlingen een invloed hebben op die schoolbetrokkenheid. In een 2de model bekijkt ze bijkomend of verbondenheid met leraren de gevonden relaties kan verklaren. In een 3de model voegt ze verbondenheid met ouders toe. De onderzoekster produceert voor een verslag de onderstaande tabel:

	Model 1	Model 2	Model 3
	B	B	B
SES	0,096	0,096	0,001
Leerjaar	0,540	0,000	0,000
Punten	0,890	0,320	0,220
Verbondenheid met leraren	-	0,216	0,180
Verbondenheid met ouders	-	-	0,306
R ²	0,110	0,263	0,300

- Benoem en interpreteer de effecten van het 1^{ste} model.
- Welke variabele heeft in het 1^{ste} model het grootste effect op schoolbetrokkenheid? En in het 2^{de} en 3^{de} model?
- Als je de verbondenheid met leraren toevoegt in het 2^{de} model, wat gebeurt dan met de variabelen van het 1^{ste} model? Interpreteer en teken telkens een bijhorend pijltjesdiagram.
- Als je verbondenheid met ouders toevoegt in het 3^{de} model, wat gebeurt dan met de variabelen van het 2^{de} model? Interpreteer en teken telkens een bijhorend pijltjesdiagram.
- Welke variabele voegt het meest verklarende kracht toe aan het model: verbondenheid met leraren of verbondenheid met ouders?
- Formuleer nu je belangrijkste conclusies uit deze tabel.

2. Inductieve statistiek

3.1. Studie van de systematiek van het toeval: kansrekenen

3.2. Stochastische variabelen

3.3. Studie van steekproefvariabiliteit

3.4. Basistools van de inductieve statistiek

1

Stel de nulhypothese en de alternatieve hypothese op. Moet je eenzijdig of tweezijdig testen?

- a) Gemiddeld wegen katten 3,5 kilogram. Dierenarts Elke verwacht dat het gemiddelde gewicht van de Britse korthaar verschillend is. Ze neemt een EAS van 30 Britse kortharen.
- b) Gemiddeld wegen katten 3,5 kilogram. Dierenarts Elke verwacht dat de Britse korthaar gemiddeld meer weegt. Ze neemt een EAS van 30 Britse kortharen.
- c) In Congo sterven jaarlijks ongeveer 20% van de nieuwgeborenen. Unicef hoopt dit jaar het sterftcijfer te verlagen door actieve ondersteuning van 50 Congolese kraamklinieken. Het voorbije jaar stierven per kraamkliniek gemiddeld 2900 baby's. Op het einde van dit jaar blijken gemiddeld nog 2680 baby's gestorven te zijn per kraamkliniek. Kan Unicef op basis van deze cijfers aan hun sponsors vertellen dat de toestand op het terrein significant verbeterde?
- d) De directeur van Volvo berekende dat de productie van een auto maximaal 120 minuten mag duren om winstgevend te zijn. Omdat hij in het verleden reeds merkte dat de fabriek in Gent niet efficiënt genoeg werkte, trekt hij er een EAS van 50 auto's. De directeur berekent dat de gemiddelde productiesnelheid in de Gentse fabriek 130 minuten bedraagt, met een standaardafwijking van 6. Kan de directeur hieruit besluiten dat de Gentse fabriek inefficiënt is en dus beter gesloten kan worden?

3.5. Inferentie voor verwachtingen

1

Het CLB heeft berekend dat de gemiddelde score van de Vlaamse leerling op de eindexamens van het lager onderwijs 65 bedraagt, met een standaardafwijking van 6,2. De directeur van een Gentse school verwacht dat zijn school beter doet dan het gemiddelde. Hij neemt een EAS van 23 leerlingen; het gemiddelde blijkt 69,3 te zijn.

- a) Verschilt het steekproefgemiddelde significant van de Vlaamse populatie?
- b) Stel dat het Vlaamse gemiddelde niet gekend is. Bouw een 80%-betrouwbaarheidsinterval als schatter van het Vlaamse gemiddelde.

2

Er wordt een wereldwijd onderzoek verricht naar prostitutie. De EU wil graag weten of het aantal raamprostituees in Europese steden significant verschilt van steden in andere landen. Het gemiddelde aantal over alle steden bedraagt 283 prostituees met een standaardafwijking van 11. Een steekproef van 45 Europese steden toont aan dat het gemiddelde 287 bedraagt.

- a) Verschilt dit steekproefresultaat significant van het wereldwijde gemiddelde?
- b) Stel dat het wereldwijde gemiddelde niet gekend is. Bouw een 95%-betrouwbaarheidsinterval als schatter van het wereldwijde gemiddelde.

3

Uit een steekproef met omvang 1514 bekomt men het 95%-betrouwbaarheidsinterval $[44,73 ; 46,53]$ voor μ . Bepaal nu het 99% betrouwbaarheidsinterval voor μ met behulp van dezelfde steekproef.

4

De Vlaams minister-president vraagt je om een onderzoek in te stellen naar de topsalarissen bij enkele vooraanstaande banken. Uit een vooronderzoek blijkt dat de standaardafwijking € 9001 bedraagt. Op uitdrukkelijk verzoek van het ministerie mag de foutenmarge van het 95%-betrouwbaarheidsinterval rond het steekproefgemiddelde niet meer dan € 400 bedragen. Hoe groot moet je steekproef dan minstens zijn?

5

Het is juni 2010; een hittegolf trekt over België. De weermannen vermoeden dat het toch serieus warmer is dan in andere jaren in juni en wensen dit te testen. Ze nemen een EAS van 6 dagen uit juni 2010 en bepalen daarvan de gemiddelde temperatuur. Deze bedroeg 26,9 °C, met een standaardafwijking van 4,5. Ze vergelijken dit met het gemiddelde van alle dagen in juni sinds de telling, wat 24,76 bedraagt.

- a) Welke steekproevenverdeling volgt het steekproevengemiddelde? Specificeer hierbij de parameters die de verdeling bepalen.
- b) Stel dat het gemiddelde van voorgaande jaren niet gekend is. Bouw een 70%-betrouwbaarheidsinterval.
- c) Voer de significantietoets uit. Is het inderdaad statistisch significant warmer dan in alle voorgaande jaren?

6

De politiedienst verantwoordelijk voor zedenfeiten heeft vastgesteld dat alle meisjes jonger dan 15 in België gemiddeld 8 keer per maand online worden lastiggevallen door iemand met slechte bedoelingen. Ze verwachten dat meisjes die hun flyer 'online/offline: let op met wie je chat!' hebben gelezen minder vaak geconfronteerd worden met dergelijke onaangename contacten. Ze contacteerden een EAS van 34 meisjes die de flyer hebben doorgenomen en stelden vast dat zij gemiddeld 7 keer per maand worden lastiggevallen, met een standaardafwijking van 4,3.

- a) Construeer een 80%-betrouwbaarheidsinterval.
- b) Verschillen de meisjes die de flyer hebben gelezen significant van de gehele onderzoekspopulatie wat betreft het online lastigvallen worden? Op welk significantieniveau?

7

Edward en Marijke zijn fervente zwemmers en ongelofelijke competitiebeesten. Na één van hun vele wedstrijden tegen elkaar, waarbij Edward het onderspit delfde, begint Marijke aan haar overwinningsdans. Edward werpt echter op dat individuele wedstrijden een slecht beeld geven en daagt Marijke uit voor een langlopende competitie. Op willekeurige momenten gaan Edward en Marijke elk apart gaan zwemmen en houden ze bij hoeveel minuten ze over 1 km deden. Je kunt de willekeurige momenten als een EAS beschouwen. Edward zal dan zijn statistische kennis gebruiken om voor eens en voor altijd na te gaan welke de beste is, maar eigenlijk verwacht hij wel dat Marijke het beter gedaan heeft. Na twee maanden komen de zwemmers opnieuw samen.

Edward heeft 56 keer gezwommen en haalde daarbij een gemiddelde tijd van 23 minuten, met een standaardafwijking van 6,45. Marijke heeft 24 keer gezwommen en haalde een gemiddelde tijd van 19 minuten, met een standaardafwijking van 9,65.

- a) Verschillen de steekproefgemiddeldes significant van elkaar? Op welk significantieniveau?
- b) Bepaal het 95%-betrouwbaarheidsinterval.

8

De Amerikaanse overheid is bezorgd om het gewicht van haar inwoners. Ze vermoedt dat het BMI van vrouwen in een relatie verschillend is van het BMI van single vrouwen. Van de 13 geselecteerde single vrouwen bedraagt het BMI gemiddeld 21, met een standaardafwijking van 4,3; terwijl het BMI van de 21 geselecteerde vrouwen in een relatie gemiddeld 26 bedraagt met een standaardafwijking van 8,3.

- a) Bouw een 80%-betrouwbaarheidsinterval op rond dit verschil.
- b) Is dit verschil significant?

3.6. Inferentie voor fracties

1

15% van de gehuwde Vlamingen heeft de huidige partner al bedrogen. In een peiling bij 2500 Limburgers, geven er 293 aan dat ze al ontrouw geweest zijn.

- a) Stel dat het Vlaamse percentage niet gekend is. Bouw een 90%-betrouwbaarheidsinterval.
- b) Verschillen Limburgers significant van de gehele Vlaamse bevolking?

2

De scouts krijgt vaak het verwijt dat de vereniging zich teveel toespitst op jongens en de meisjes wat in de kou laat staan. De VVKSM wenst dit te testen en bekijkt of er inderdaad een verschil is tussen de fracties meisjes in de populatie (0,5) en die in de jeugdbeweging, zonder te verwachten of er nu meer of minder meisjes zijn in de jeugdbeweging. Men trekt een EAS van leden uit de scouts. Van de 93 geselecteerde leden, zijn er 37 meisjes.

- a) Stel een 97%-betrouwbaarheidsinterval op.
- b) Voer de significantietoets uit. Verschil de steekproeffractie significant van de populatiefractie? Op welk significantieniveau?
- c) Wat is de standaardfout en de verwachting van de steekproevenverdeling die je gebruikte voor de berekening van het betrouwbaarheidsinterval? Welke gebruikte je voor het uitvoeren van de significantietoets?

3

Het Centrum voor Drugsbestrijding wil nagaan of er een geslachtsverschil bestaat met betrekking tot softdruggebruik bij ex-gedetineerden. Er werd een EAS getrokken van 324 mannen, waarbij 56 aangaven inderdaad softdrugs te gebruiken en 67 vrouwen, waarvan er 9 positief antwoordden.

- a) Stel een 95%-betrouwbaarheidsinterval op.
- b) Verschillen de steekproeffracties significant van elkaar? Op welk significantieniveau?

4

Politicalogen doen vaak onderzoek naar verschillen in stemgedrag tussen verschillende kiesarrondissementen. Ze gaan ervan uit dat, tijdens de vorige verkiezingen, Oost-Vlamingen meer op de sp.a stemden dan West-Vlamingen en wensen dit te testen. Daartoe trekken ze een EAS van 465 Oost-Vlamingen en een EAS van 198 West-Vlamingen. Uit hun resultaten blijkt dat 68 Oost-Vlamingen en 17 West-Vlamingen bij de vorige verkiezingen voor de sp.a gestemd hebben.

- a) Hoeveel bedraagt de foutenmarge van het 80%-betrouwbaarheidsinterval?
- b) Verschillen beide steekproeffracties significant van elkaar? Op welk niveau?

3.7. Inferentie voor kruistabellen

1

Het centrum voor gendergelijkheid wenst eens een onderzoek te doen naar een oude stereotype opvatting die bestaat: is het zo dat er een significant verband tussen geslacht en het al dan niet gebruiken van geweld? Ze voeren een grootschalig onderzoek uit en bekomen uiteindelijk volgende tabel. Voer de significantietoets uit.

	Man	Vrouw	Totaal
Geweld	3266	210	3476
Geen geweld	7491	1550	9041
Totaal	10757	1760	12517

2

Het centrum wenst nog een andere misvatting de wereld uit te helpen. Al jaren wordt gesteld dat mannen een grotere neiging hebben om racistisch te zijn. In het grootschalige onderzoek werd ook hiernaar gepeild (zie tabel). Is er een significant verband tussen geslacht en mate van racisme?

	Helemaal akkoord	Akkoord	Ik twijfel	Niet akkoord	Helemaal niet akkoord	Totaal
Man	6	23	101	355	538	1023
Vrouw	3	16	102	462	1130	1713
Totaal	9	39	203	817	1668	2736

DEEL C: ZELFSTUDIE

1. Univariate beschrijvende statistiek

1.1. Basisconcepten

1*

Kind en Gezin doet een massale bevraging van de Vlaamse gezinnen. Eén van de vragen die in de vragenlijst voorkwam, ging over het aantal kinderen dat gezinnen hebben. De gegevens vind je hieronder terug.

Aantal kinderen	Aantal gezinnen
0	21
1	23
2	45
3	12
4	3
5	1
6	0

- Wat zijn hier de eenheden? En de variabele?
- Hoeveel gezinnen hebben 2 kinderen? Welke proportie is dit?
- Geef het percentage gezinnen dat minimum 3 kinderen heeft

2*

Uit de resultaten van de studentenbarometer - een vragenlijst die werd verstrekt aan alle studenten aan de UGent - is gebleken dat studenten aan de faculteit Politieke en Sociale Wetenschappen gemiddeld 2,5 glazen bier per week meer drinken dan studenten aan de faculteit Psychologie.

- Wat is de populatie?
- Wat is een statistische eenheid?
- Wat zijn de variabelen?
- Op welk meetniveau zijn de variabelen gemeten?

3

Een onderzoeker van de UGent wil weten hoe regenachtig en hoe zonnig de regio's van de landen van de Europese Unie zijn. Hij neemt contact op met het KMI, dat gegevens heeft per regio. Deze gegevens bestaan uit het aantal liter regen per kubieke meter én het antwoord op de vraag 'Zou u zeggen dat uw regio zonnig, overwegend zonnig, noch zonnig/noch bewolkt, overwegend bewolkt of bewolkt is?'

- a) Wat is de populatie in dit voorbeeld?
- b) Wat is een statistische eenheid?
- c) Welke zijn de variabelen hier? En wat is hun meetniveau?

4

Een leraar uit een lagere school in Oudenaarde neemt het wiskunde examen af op 18 december 2008. Elk van zijn 17 leerlingen moet het examen invullen en krijgt dan ook net voor de kerstvakantie zijn rapport mee met de score op de test.

- a) Wat is de populatie in dit voorbeeld?
- b) Wat is een statistische eenheid?
- c) Welke zijn de variabelen hier? En wat is hun meetniveau?

5

Herman Schuermans wil weten wat de demografische kenmerken van zijn publiek op Rock Werchter zijn. Hij neemt het Gentse onderzoeksbureau Check-It onder de arm, die een grootschalig onderzoek instelt bij de editie van 2010. Het onderzoeksbureau gaat ervan uit dat de kenmerken van het publiek van dat jaar niet wezenlijk verschillen van andere jaren. Het onderzoeksbureau ontdekt dat Rock-Werchter-bezoekers gemiddeld genomen jong zijn, een hoog inkomen hebben en zich relatief gezien weinig wassen. Ook bemerkt men een verschil tussen de geslachten: jongens hebben vaker concerten meegepikt op de Main Stage, meisjes maken meer concerten mee in de Marquee.

- a) Wat is de populatie in dit voorbeeld?
- b) Wat is een statistische eenheid?
- c) Welke zijn de variabelen hier? En wat is hun meetniveau?

6

In Antwerpen vindt men meer woningen in verkrotte toestand dan in om het even welke andere Vlaamse gemeente.

- a) Wat is de populatie?
- b) Wat zijn de statistische eenheden?
- c) Wat zijn de variabelen?
- d) Op welk meetniveau zijn de variabelen gemeten?

7

Steven is een fervent zwemmer en gaat iedere dag gaan zwemmen. Hij is ervan overtuigd dat statistiek hem hierbij kan helpen en houdt dan ook iedere dag hoeveel meter hij gezwommen heeft. Hij schrijft ook steeds op of de zon schijnt of niet. Zo wil hij te weten komen of het weer zijn prestaties beïnvloedt. Dit houdt hij tien jaar vol: tussen zijn twintigste en dertigste levensjaar. Als hij dertig wordt, kan hij besluiten dat hij er echt wel op vooruit gegaan is in die tien jaar.

- a) Wat is de populatie?
- b) Wat is een statistische eenheid?
- c) Wat zijn de variabelen? En hun meetniveau?

8*

Wat is het meetniveau van de bestudeerde variabele? Geef ook aan of het om een discrete of continue variabele gaat.

- a) Wachten op de tram uitgedrukt in het aantal verlopen seconden tussen aankomst aan de halte en aankomst van de tram
- b) Het rijksregisternummer van personen die bevroegd worden in een enquête
- c) Bij een onderzoek over het personenvervoer in België wordt voor elke wagen die wordt waargenomen het merk van de wagen genoteerd
- d) Een leraar doet een spellingstest door 5 willekeurige woorden uit een woordenboek te dicteren. Hij geeft een 0 als hij één verkeerd gespeld woord tegenkomt en legt de test opzij en kent een 1 toe als de leerling alle woorden juist spelt
- e) Het aantal boeken dat iedere dag wordt uitgeleend in de bib
- f) De afstand die verschillende personen in 5 minuten kunnen lopen

9

Een groepje veeartsen wil een idee krijgen van de vermeende agressiviteit van Pitbulls. Ze noteren van iedere Pitbull die hun praktijk betreedt, het aantal aanvallen die de honden op menselijk vlees uitvoerden.

- a) Wat is de bestudeerde variabele?
- b) Hoe wordt dit kenmerk precies gemeten?
- c) Noem een statistische eenheid
- d) Wat is de bestudeerde populatie?

10

Sinds een aantal jaar neemt de politie tijdens de oudejaarsperiode in het kader van de BOB-campagne een ademtest af van een aantal Vlaamse bestuurders. Van de bestuurders die in de afgelopen eindejaarsperiode werd gecontroleerd op alcohol legde 5,5 procent een positieve ademtest af. Dat is meer dan tijdens de eindejaarsperiode vorig jaar, toen dit 4,4 procent bedroeg.

- a) Welke variabelen worden hier bekeken?
- b) Welk meetniveau hebben deze variabelen?
- c) Noem een statistische eenheid
- d) Wat is de bestudeerde populatie?
- e) Maakt men gebruik van een steekproef of wordt de hele populatie bestudeerd?

11

Een groepje criminologen bestudeert verkeersagressie. Ze tellen daartoe het totaal aantal PV's opgemaakt voor verkeersagressie voor elk Vlaams arrondissement.

- a) Wat is de bestudeerde variabele ?
- b) Hoe wordt dit kenmerk precies gemeten?
- c) Noem een statistische eenheid
- d) Wat is de bestudeerde populatie?

12

"12% van alle Belgen leeft in armoede."

- a) Wat is de bestudeerde variabele?
- b) Op welk niveau wordt dit kenmerk gemeten?
- c) Noem een statistische eenheid
- d) Wat is de bestudeerde populatie?
- e) Maakt men gebruik van een steekproef of wordt de hele populatie bestudeerd?

13

Testaankoop stelt een onderzoek op om na te gaan welk Duits automerk – Opel of Volkswagen – het meest degelijk is. De organisatie neemt daartoe een steekproef uit alle mensen die het afgelopen jaar een auto gekocht hebben. Van de 13376 respondenten die meldden dat zij een Opel gekocht hadden, hadden er 2942 in het afgelopen jaar de reparatiedienst moeten inschakelen. Van de 480 respondenten die een Volkswagen gekocht hadden, meldden er slechts 192 dat ze de reparatiedienst nodig hadden gehad.

- Welk merk is het meest betrouwbaar?
- Wat is de populatie?
- Wat zijn de variabelen? En wat is hun meetniveau?

14*

De toeristische dienst van de Kust vroeg aan 100 eendagstoeristen welke snack ze prefereren tijdens een hete namiddag op het Vlaamse strand.

Snack	Aantal toeristen
Pannenkoek	35
Suikerspin	12
Ijsje	26
Donut	4
Wafel	23

- Is de tabel een datamatrix of een frequentietabel?
- Wat is het meetniveau van de variabele?

1.2. Frequentieverdelingen

1*

De Vlaams minister van welzijn is geïnteresseerd in het datinggedrag van de Vlamingen. Het onderzoeksbureau Check-It noteerde de leeftijd (afgerond naar de dichtstbijzijnde verjaardag) van 50 internetgebruikers die via een online datingsite een levenspartner zochten. De gegevens vind je hieronder terug.

46	41	32	43	21	31	37	37	49	32	51	50	34	20	41
39	53	40	50	28	36	64	21	54	25	46	45	32	34	34
48	30	44	35	33	28	30	27	40	28	44	48	39	43	29
20	25	35	34	41										

- Wat is de variatiebreedte?
- Welke 5 klassen zou je kunnen maken, wanneer je probeert om zoveel mogelijk met gelijke klassenbreedte te werken?
- Bepaal vervolgens de klassenmiddens
- Stel een frequentietabel op aan de hand van deze klassenindeling

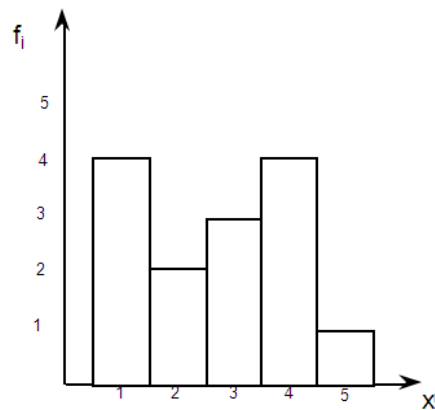
2

De minister van defensie is van plan een forse besparingsmaatregel in te voeren. Voor die van kracht wordt, wil hij eerst de verdeling van zijn troepenmacht over de verschillende officierenposities nagaan. Zijn stafmedewerkers komen met volgende tabel op de proppen. (Onderluitenant is de laagste graad, Luitenant Generaal de hoogste). Geef weer in een grafiek.

Graad	f_i
Onderluitenant	1200
Eerste luitenant	800
Kapitein	1100
Commandant	1800
Majoor	830
Luitenant Kolonel	510
Kolonel	220
Generaal Majoor	30
Luitenant Generaal	10

3*

Bepaal de proportie van waarde 3.



1.3. Maten van positie

1

De minister van administratieve vereenvoudiging geeft het onderzoeksbureau Check-It de opdracht om een algemene rondvraag te organiseren bij de huidige ambtenaren. Eén van de belangrijke kenmerken was het aantal maanden dat ambtenaars in hun huidige functie tewerkgesteld waren. De resultaten hiervan vind je, gegroepeerd in klassen, terug in onderstaande tabel. We beschouwen het aantal maanden hier als een discrete variabele.

Aantal maanden	f_i
0-11	13
12-23	40
24-35	78
36-47	146
48-59	196
60-71	169
72-83	118
84-95	59
96-107	50
108-119	14
120-131	16
132-143	0
144-155	0
156-167	0
168-179	1

- Wat is een statistische eenheid in dit voorbeeld?
- Hoe groot is n ?
- Bereken de mediaan, de kwartielen, het rekenkundig gemiddelde en het 80e percentiel.
- Interpreteer deze maten

2*

Na een nieuwe nationale IQ-test, uitgezonden door een commerciële omroep, zijn de IQ-scores van een aantal prominente BV's uitgelekt in een populair roddelblad. Het tijdschrift publiceert onderstaande tabel. Bereken de meest relevante maat van centraliteit.

Bekende Vlaming	IQ-score
Karen Damen	116
Martin Heylen	126
Samson	74
Danny Verstraeten	127
Jeff Hoeyberghe	107
Walter Grootaers	107
Erwin Vervecken	120

3*

De toeristische dienst van de Vlaamse kust wil weten hoe tevreden de toeristen zijn van hun daguitstap naar de kust. Eén van de vragen luidt: "Was u tevreden met de bediening in de horecazaken?" De verdeling van de toeristen over de antwoorden vind je terug in onderstaande frequentietabel.

	Aantal toeristen
Zeer ontevreden	33
Ontevreden	84
Noch tevreden, noch ontevreden	102
Tevreden	63
Zeer tevreden	48

- Wat is het meetniveau van de variabele?
- Bereken alle relevante maten van positie.

4*

Het Centrum voor Leerlingen Begeleiding (CLB) uit West-Vlaanderen heeft zijn jaarlijkse onderzoek bij de leerlingen van het Lager Onderwijs voltooid. In totaal kwamen 6254 leerlingen uit het derde jaar langs. Eén van de belangrijke kenmerken die in de gaten worden gehouden, is de lichaamslengte van de leerlingen. De gegevens van het onderzoek vind je terug in onderstaande frequentietabel. Bereken alle relevante maten van positie.

Lichaamslengte	Aantal leerlingen
[100-111[534
[111-116[712
[116-121[2021
[121-126[1947
[126-131[553
[131-136[487

5

Larry Boudini is een beginnend goochelaar. Van zijn mentor, de grote Fronticelli, moet hij bijhouden hoeveel trucs er elke avond lukken. Larry begint aan zijn rondreis door het Belgische land en een maand later klopt hij aan bij Fronticelli. Hij toont hem onderstaande tabel.

- Is bovenstaande tabel een datamatrix of een frequentietabel?
- Vul de frequentietabel verder aan met cumulatieve frequenties, proporties en cumulatieve proporties
- Bereken de relevante maten van positie

Voorstelling	Aantal gelukke trucs
1	6
2	2
3	8
4	6
5	4
6	2
7	5
8	3
9	6
10	9
11	5
12	4
13	5
14	6
15	2
16	1

Voorstelling	Aantal gelukke trucs
17	5
18	6
19	4
20	8
21	3
22	6
23	3
24	1
25	5
26	6
27	4
28	7
29	2
30	9
31	5

6

Bij een enkelvoudige aselechte steekproef van werknemers in een multinational wordt via een korte schriftelijke vragenlijst nagegaan hoe tevreden men is over de catering in het bedrijf. 901 personen uit de steekproef zijn "zeer tevreden", 1025 werknemers zijn "tevreden", 1143 werknemers zijn "noch tevreden, noch ontevreden", 408 personen zijn "ontevreden" over de catering en 159 personen zijn ronduit "zeer ontevreden". Welke van de volgende beweringen zijn correct? Meerdere correcte beweringen kunnen voorkomen. Beweringen die je niet kan controleren aan de hand van de informatie die je hier krijgt, moet je als verkeerd beschouwen.

- Het 30^{ste} percentiel bedraagt afgerond 1091
- De onderzoekspopulatie bestaat hier uit alle individuen die aan de enquête deelnamen
- De beste maat van centraliteit die we hier kunnen berekenen is gelijk aan de categorie "tevreden"
- We mogen in dit geval cumulatieve relatieve frequenties berekenen

7*

De directie van het pretpark Hoepla wenst een profiel op te stellen van haar bezoekers. Ze neemt daartoe contact op met de Universiteit Gent om een publieksonderzoek uit te voeren. De onderzoekers nemen een EAS uit de bezoekers die in een week in de zomervakantie naar het pretpark komen en vragen hen enkele gegevens te verstrekken. Eén belangrijk kenmerk was de leeftijd van de ondervraagden. De gegevens worden hieronder weergegeven.

16	25	14	39	85	62	4	46	23
12	16	19	18	46	75	23	34	28
39	85	62	49	78	32	16	85	10
9	41	49	73	62	51	59	20	17

- Wat is de variatiebreedte van deze gegevens?
- Hoeveel bedraagt de klassenbreedte indien je 5 klassen wil maken met dezelfde klassenbreedte?
- Verdeel de gegevens onder in 5 klassen met klassenbreedte 16 (de laatste klasse zal een andere klassenbreedte hebben). Maak een frequentietabel op van deze gegevens. Geef ook de cumulatieve frequentie, de proporties en de cumulatieve proporties erin weer
- Bepaal de kwartielen

1.4. Maten van spreiding

1*

De toeristische dienst van de Vlaamse kust wil weten hoe tevreden de toeristen zijn van hun daguitstap naar de kust. Eén van de vragen luidt: "Was u tevreden met de bediening in de horecazaken?" De verdeling van de toeristen over de antwoorden vind je terug in onderstaande frequentietabel. Bereken alle relevante maten van spreiding.

	Aantal toeristen
Zeer ontevreden	33
Ontevreden	84
Noch tevreden, noch ontevreden	102
Tevreden	63
Zeer tevreden	48

2*

Onderzoeksbureau Check-It krijgt een belangrijke opdracht: men moet van de Vlaamse Overheid nagaan hoe het met de opleiding van etnisch-culturele minderheden gesteld is. Het gerenommeerde onderzoeksbureau zet zich onmiddellijk aan de slag en binnen een mum van tijd heeft men de studie klaar. Minister van Onderwijs Pascal Smet krijgt volgende frequentietabel onder zijn neus.

Leeftijd bij afstuderen	Frequentie
16	11
18	106
19	154
20	120
21	99
22	45
23	45
25	23

- Wat is het meetniveau van de variabele?
- Bereken alle relevante maten van spreiding
- Zijn er outliers? Zo ja, welke waarden hebben ze?

3

Herneem de opgave van oefening C.1.3.5.

- Bereken alle relevante maten van spreiding

4

De Gezinsbond brengt haar jaarlijkse resultaten naar buiten. Eén van de belangrijke taken van de vereniging is het geboortecijfer in de gaten te houden. Hieronder vind je drie tabellen terug, die het aantal kinderen per gezin weergeven voor gezinnen waarvan de moeder ervoor gekozen heeft om huisvrouw te zijn (en dus niet te gaan werken), voor gezinnen waarvan de vader die keuze gemaakt heeft en voor gezinnen waar beide ouders werken.

Gezinnen met huismoeder	
Aantal kinderen	Aantal gezinnen
0	23
1	22
2	28
3	31
4	14

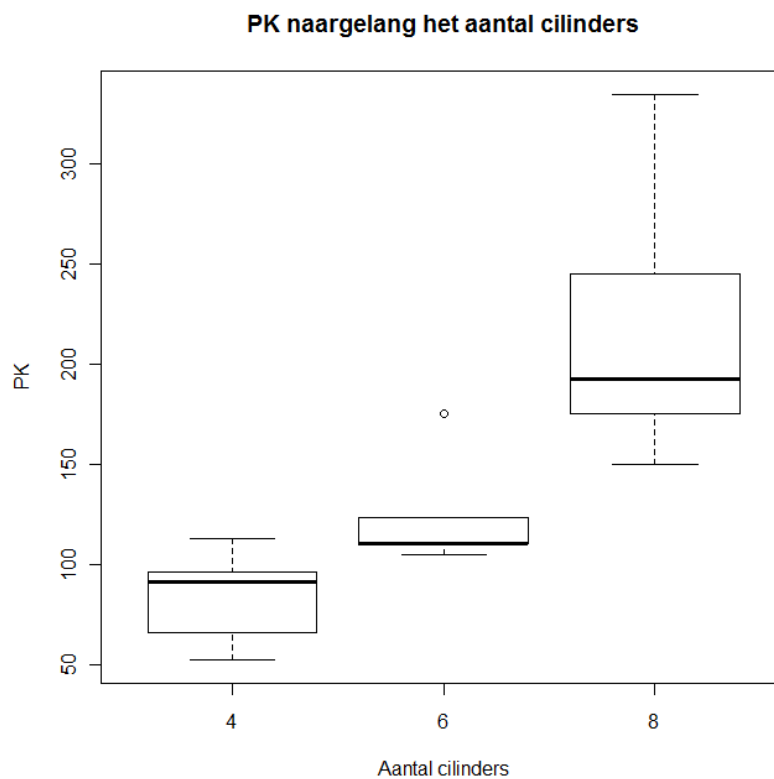
Gezinnen met huisvader	
Aantal kinderen	Aantal gezinnen
0	9
1	12
2	12
3	4
4	3

Gezinnen met werkende ouders	
Aantal kinderen	Aantal gezinnen
0	45
1	31
2	36
3	28
4	7

- Zijn deze tabellen datamatrices of frequentietabellen?
- Bereken van elke tabel het gemiddelde en de mediaan
- Bereken voor elke tabel de standaardafwijking. Wat kun je hieruit besluiten?

5

Een Amerikaans automagazine heeft een onderzoek gedaan naar de verhouding tussen het aantal cilinders van een automotor en de paardenkracht (PK) die daarmee gepaard gaat. Onderstaande grafiek geeft de resultaten weer aan de hand van drie boxplots.



- Bij hoeveel cilinders is de variatie in PK's het laagst
- Hoe denk je dat de frequentieverdeling is bij auto's met 8 cilinders? Symmetrisch, rechts asymmetrisch of links asymmetrisch? Maak een schets.
- Het gemiddeld aantal PK voor alle auto's is 146,7 terwijl de mediaan slechts 123 is. Hoe kan je dit verschil verklaren?

6

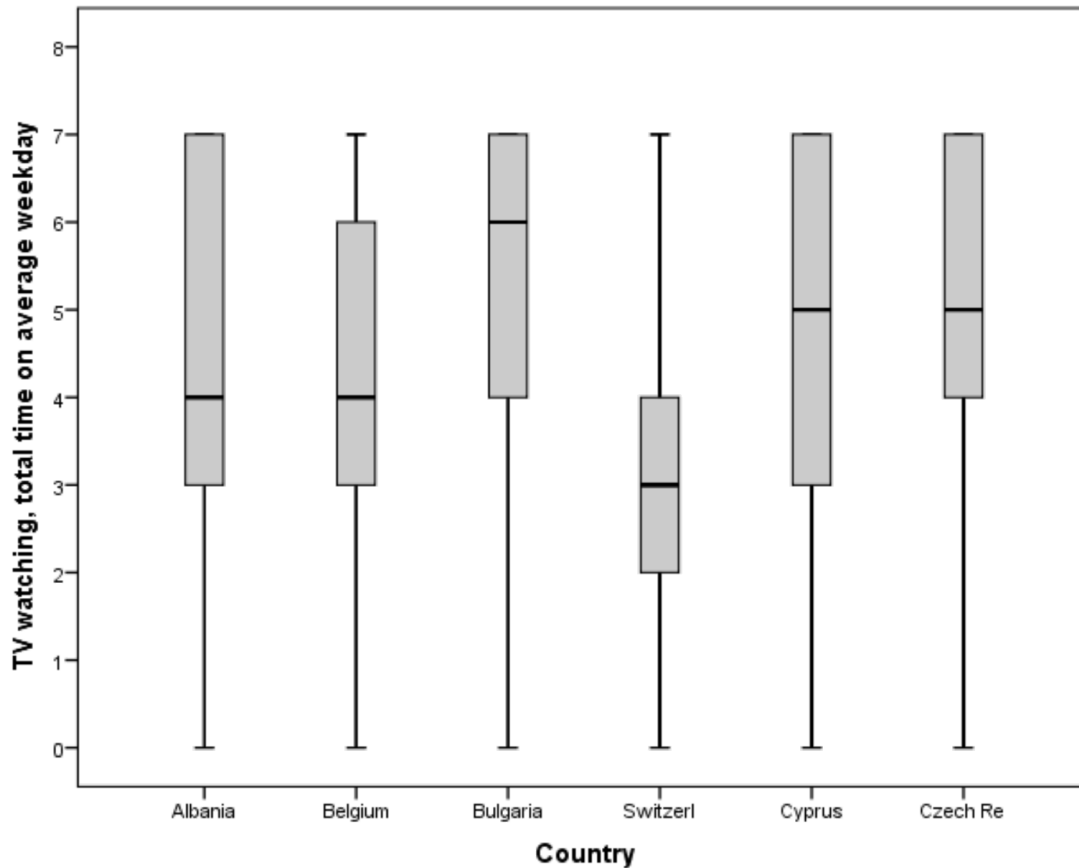
Op aandringen van zijn ouders heeft Pierre een overzicht gemaakt van al zijn studieresultaten die hij behaald heeft tijdens zijn universitaire carrière. Onderstaande tabel geeft een overzicht.

Punten op 20	Cumulatieve frequentie
0	0
1	0
2	1
3	2
4	2
5	3
6	3
7	3
8	5
9	8
10	13
11	16
12	18
13	24
14	29
15	32
16	33
17	34
18	34
19	34
20	34

- Vul deze tabel aan met absolute frequenties, proporties en cumulatieve proporties
- Wanneer Pierre de resultaten voorlegt aan zijn ouders, presenteert hij dan best het gemiddelde of de mediaan? Waarom?
- Bereken alle spreidingsmaten die van toepassing zijn

7*

De Europese Commissie wil nagaan hoeveel uren men gemiddeld per dag naar televisie kijkt. Hieronder vind je de resultaten terug voor zes Europese landen. Welke stellingen zijn correct en welke zijn foutief?



- Het percentage Albanezen dat gemiddeld tussen de 3 en 4 uur televisie kijkt per dag is even groot als het percentage Bulgaren dat tussen de 4 en 6 uur televisie kijkt
- Het percentage Zwitsers dat gemiddeld 3 uur of minder naar televisie kijkt, is kleiner dan het percentage dat 3 uur of meer kijkt per dag
- Op het eiland Cyprus wordt gemiddeld 5 uur per dag naar televisie gekeken
- De spreiding in het gemiddeld aantal uren televisie kijken per dag is groter in Albanië dan in Zwitserland

1.5. Vorm van een verdeling

1

De Gezinsbond brengt haar jaarlijkse resultaten naar buiten. Eén van de belangrijke taken van de vereniging is het geboortecijfer in de gaten te houden. Hieronder vind je drie tabellen terug, die het aantal kinderen per gezin weergeven voor gezinnen waarvan de moeder ervoor gekozen heeft om huisvrouw te zijn (en dus niet te gaan werken), voor gezinnen waarvan de vader die keuze gemaakt heeft en voor gezinnen waar beide ouders werken. Wat is de vorm van elk van deze verdelingen?

Gezinnen met huismoeder	
Aantal kinderen	Aantal gezinnen
0	23
1	22
2	28
3	31
4	14

Gezinnen met huisvader	
Aantal kinderen	Aantal gezinnen
0	9
1	11
2	13
3	4
4	3

Gezinnen met werkende ouders	
Aantal kinderen	Aantal gezinnen
0	45
1	31
2	36
3	28
4	7

2

Het Nationaal Instituut voor de Statistiek is geïnteresseerd in de evolutie van het geboortecijfer in België. In een vooronderzoek bij 104 willekeurig gekozen Belgische vrouwen stellen ze de volgende vraag: "Hoeveel kinderen heeft u?".

Aantal kinderen	Aantal vrouwen
0	30
1	35
2	21
3	13
4	3
6	2

- Bepaal de vorm van de verdeling aan de hand van het berekenen van maten van positie en/of spreiding.
- Welke maten heb je hiervoor berekend?
- En welke vorm heeft deze verdeling?

3

Een Vlaamse jeugdbeweging wil wel eens weten hoe hun leeftijdsverdeling eruit ziet. Een jonge PSW-student uit Gent, die toevallig in de leiding zit, besluit zijn kennis in de praktijk om te zetten en verzamelt de gegevens van alle leden. Hieronder vind je de leeftijdsverdeling terug.

- Groep de waarden in 4 klassen met gelijke klassenbreedte.
- Teken de boxplot voor deze verdeling (zonder dat hij in klassen is ingedeeld)
- Welke vorm heeft deze verdeling?

Leeftijd	Aantal gasten
6	6
7	7
8	9
9	9
10	13
11	11
12	12
13	13
14	11
15	23
16	21
17	27

4*

Een Belgische wandelclub trekt geregeld naar het buitenland om er lange wandelingen af te leggen. De voorzitter houdt minutieus bij hoeveel kilometer elke wandeling telde. De gegevens vind je hieronder terug.

- Stel de frequentietabel op
- Teken een boxplot voor de verdeling
- Welke vorm heeft deze verdeling?

Wandeling	km
1	10
2	14
3	12
4	10
5	13
6	9
7	14
8	15
9	9
10	10
11	19
12	12

5*

Elk jaar vindt de Marathon van Brussel plaats. George, een fervent marathonliefhebber, wil nagaan hoe goed de mensen uit de top 200 van de wereld het dit jaar doen. Hij wacht ze bij de aankomst op en vraagt hen hoeveel minuten ze over de marathon gedaan hebben (afgerond tot op de minuut). De gegevens vind je hieronder terug:

Aantal minuten	Aantal renners
128	1
129	3
130	2
131	4
132	8
133	7
134	6
135	9
136	13
137	15
138	18

Aantal minuten	Aantal renners
139	19
140	28
141	23
142	14
143	11
144	6
145	7
146	1
147	3
148	2

- Wat is het meetniveau van de variabele?
- Is deze verdeling symmetrisch of asymmetrisch?

1.6. Transformaties van een variabele

1.7. Dichtheidskrommen

1*

Sabrina Fosters, een wereldvermaarde dierenarts uit Gent heeft een artikel geschreven waarin ze bewijst dat de lengte van Baobabs, een bomensoort uit Afrika en Australië, perfect normaal verdeeld is, met een gemiddelde van 14 m en een standaardafwijking van 4,5 m. De wetenschappelijke wereld staat op zijn kop.

- Welk percentage van de Baobabs is korter dan 8 m?
- Welke proportie van alle Baobabs is langer dan 18 m?
- Hoe lang zijn de langste 7 % van alle Baobabs?
- Hoe groot is de proportie van Baobabs die hoger is dan 10 m, maar kleiner is dan 16 m?
- Hoe lang zijn de kortste 25 % van alle Baobabs?

2

Jan Van Telleghem, een maritiem bioloog, leest het blad 'Nature' maandelijks en zijn aandacht wordt dan ook getrokken door het wereldveranderde stuk van Sabrina Fosters. Hij duikt onmiddellijk de zee in en begint de lengte van haaien te meten. Niet veel later komt hij tot de ontdekking dat de lengte van deze dieren perfect normaal verdeeld is met een gemiddelde van 1,5 m en een standaardafwijking van 0,30 m.

- Welk percentage van de haaien is korter dan 1 m?
- Welke proportie van alle haaien is langer dan 2,5 m?
- Hoe lang zijn de langste 10 % van alle haaien?
- Hoe groot is de proportie haaien die langer is dan 1 m, maar kleiner is dan 2 m?
- Hoe lang zijn de kortste 5 % van alle haaien?

3*

De Lijn Oost-Vlaanderen stelt een intern onderzoek in. Men wil de effectiviteit van het tramverkeer in Gent meten. Eén van de resultaten wees uit dat de tijd dat tramlijn 21 erover doet om van de Zwijnaardebrug naar Melle Leeuw te rijden, perfect normaal verdeeld is volgens $N(60, 5)$.

- a) Welk percentage van de trams rijdt langer over het traject dan 67,5 minuten?
- b) Hoeveel minuten doen de 5% snelste trams over het traject?
- c) Welke proportie van de trams doet minder lang dan 75 minuten, maar langer dan 55 minuten over het traject?
- d) Als de Lijn op haar aankondigingen plaatst dat de tram na 57 minuten op haar eindbestemming aankomt, hoeveel procent van de trams rijdt dan op tijd?
- e) Als de Lijn in haar jaarlijkse doelstelling opgenomen heeft dat 30 % van de trams op tijd moeten aankomen, welke ritduur moet de Lijn dan op haar aankondigingen plaatsen?

4

Het WWF is geïnteresseerd in de toestand van het Braziliaanse regenwoud. Men voert gedurende enkele maanden onderzoek uit. Daarbij meet men onder andere het aantal doeroecoeli's (een nachtaapje) per willekeurig gekozen stukje van 1 are. Hierbij komt men tot de verbluffende ontdekking dat het aantal doeroecoeli's per 1 are normaal verdeeld is, volgens $N(23, 7)$.

- a) Bij hoeveel procent van de stukjes van 1 are uit het regenwoud vindt men minder dan 10 doeroecoeli's?
- b) Hoeveel procent van de stukjes van 1 are bevat meer dan 35 doeroecoeli's?
- c) Hoeveel doeroecoeli's vind je in de 10 % gebieden met de hoogste populatie van deze beestjes?
- d) Hoeveel doeroecoeli's vind je in de 15 % gebieden met de laagste populatie van deze beestjes?
- e) Een gebied is voor WWF gezond als het meer dan 20, maar wegens overpopulatie ook minder dan 40 doeroecoeli's bevat. Hoeveel procent van de gebieden van 1 are zijn gezond?

5

Mia Vellezout, de beroemde Nederlands-Belgische hoogspringster, heeft zich ingeschreven voor de Olympische Spelen. Het Olympisch comité stelde een nieuwe maatregel op waarbij hoogspringers, net als in het boksen en het worstelen, in klassen worden opgedeeld. Bij hoogspringers gebeurt dat op basis van hun lichaamslengte. De klassen zijn: - 160 cm; 160 – 170; 170 – 180; 180-190; 190 +. Mia wordt ingedeeld in de klasse van 170-180. De lichaamslengte van hoogspringers is perfect normaal verdeeld, volgens $N(180, 10)$.

- a) Hoeveel procent van alle hoogspringers wordt ingedeeld in de laagste categorie?
- b) Hoeveel procent van alle hoogspringers wordt ingedeeld in de hoogste categorie?
- c) Hoeveel procent van alle hoogspringers is kleiner dan 178 cm?
- d) Hoeveel procent van alle hoogspringers wordt ingedeeld in de klasse van Mia?

6

De Trappisten van de Sint-Sixtusabdij in Westvleteren zijn volop bezig met een controle van het productieproces van hun wereldberoemde trappistenbier: Westvleteren. Na het controleren van een random sample van 100 flesjes stellen ze vast dat hun bottelmachine een lichte afwijking vertoont: de inhoud van de flesjes is niet exact 33cl, maar volgt een normaalverdeling: $N(33, 2)$.

- a) Aangezien de monniken een eerlijk product willen verkopen, willen ze dat erin alle flesjes minstens 32cl en maximum 34cl zit. Hoeveel procent van alle flesjes voldoen aan deze vereiste?
- b) Pater Boudewijn, commercieel verantwoordelijke, vraagt zich af hoeveel procent van de flesjes meer dan 35cl bevat.

7*

Een bedrijf wenst iemand aan te werven. Tijdens de eerste selectieronde moeten de sollicitanten ($n=803$) een psycho-technische proef afleggen. De resultaten zijn perfect normaal verdeeld, $N(45, 8)$. Alleen de sollicitanten met een score die hoger is dan 59 worden voor een gesprek uitgenodigd. Hoeveel sollicitanten zullen dit zijn (afgerond cijfer)? Slechts één bewering is hier correct.

- a) 0 sollicitanten
- b) Afgerond 771 sollicitanten
- c) Afgerond 612 sollicitanten
- d) Afgerond 51 sollicitanten
- e) Afgerond 32 sollicitanten

8

De radarcontroles op de snelweg registreren de snelheid van alle voertuigen die voorbijrijden. Enkel de voertuigen die sneller rijden dan 123 km per uur worden echter effectief 'geflitst' en mogen zich aan een Proces Verbaal verwachten. Uit een analyse van de gegevens blijkt dat de gemiddelde snelheid, zowel van personenwagens als van moto's, normaal verdeeld is. Bij de personenwagens bedraagt de gemiddelde snelheid 115 km per uur, met een standaarddeviatie van 5 km per uur. De gemiddelde snelheid van de moto's bedraagt 120 km per uur, met een standaarddeviatie van 5 km per uur. Dit heeft als gevolg dat moto's 5 keer zoveel risico lopen om geflitst te worden als auto's.

Onder druk van de motorrijders die argumenteren dat het voor hen veiliger is iets sneller te rijden dan de personenwagens wordt de snelheid waarbij effectief geflitst wordt opgetrokken. Voortaan wordt er pas effectief geflitst vanaf 126 km per uur. Hoeveel hoger is het risico voor een motorrijder om geflitst te worden tegenover een automobilist na invoering van deze maatregel? (Nota: de snelheden van personenwagens en moto's zijn normaal verdeeld...)

- a) De verhouding is ongewijzigd: motorrijders lopen nog steeds 5 keer zoveel risico als automobilisten om geflitst te worden
- b) Motorrijders lopen nu nog 3,2 keer zoveel risico als automobilisten om geflitst te worden
- c) Motorrijders lopen nu 8,3 keer zoveel risico als automobilisten om geflitst te worden
- d) Motorrijders lopen nu nog 4,3 keer zoveel risico als automobilisten om geflitst te worden
- e) Motorrijders lopen nu nog 2,7 keer zoveel risico als automobilisten om geflitst te worden

9

Een gevangenisdirecteur wil een bevraging organiseren bij de oudste 10 % van de gevangenispopulatie. De leeftijden in de gevangenis zijn normaal verdeeld, met gemiddelde 37,8 jaar en standaardafwijking 5,6 jaar. Vanaf welke leeftijd zullen de gevangenen aan de bevraging kunnen deelnemen?

- a) De gevangenen vanaf de leeftijd van (afgerond) 42,5 jaar
- b) De gevangenen vanaf de leeftijd van (afgerond) 45 jaar
- c) De gevangenen vanaf de leeftijd van (afgerond) 47 jaar
- d) De gevangenen vanaf de leeftijd van (afgerond) 50 jaar
- e) We kunnen dit niet berekenen bij een gebrek aan gegevens

2. Bivariate beschrijvende statistiek

2.0. Studie van samenhangen en effecten

1*

Geef in de volgende voorbeelden aan of de bivariate relatie symmetrisch of asymmetrisch is. Indien het een asymmetrische relatie is, moet je ook aangeven wat de onafhankelijke en afhankelijke variabele is.

- a) Het aantal vissen in de Blaarmeersen en het aantal vissers
- b) Het aantal gevangenen in een gevangenis en de kans op rellen
- c) Het aantal treinen die rijden in een bepaald land en het BNP van dat land

2

Geef in de volgende voorbeelden aan of de bivariate relatie symmetrisch of asymmetrisch is. Indien het een asymmetrische relatie is, moet je ook aangeven wat de onafhankelijke en afhankelijke variabele is.

- a) Het aantal kinderen in een gezin en het percentage van het inkomen dat aan voeding gespendeerd wordt
- b) Het humeur van mensen en het weer
- c) Het inkomen van twee broers
- d) De afgelegde afstand met een auto en het dieselverbruik
- e) Het aantal pagina's in een boek en het aantal lezers

2.1. Relaties tussen categorische variabelen

1*

De organisatie die waakt over de auteursrechten op onder andere muziek, SABAM, wenst te onderzoeken of ook particulieren moeten betalen voor de muziek die ze thuis laten afspelen. Dit kadert in de strijd tegen piraterij op het internet. Om te monitoren wie, wat en hoeveel men muziek laat spelen thuis, voert de organisatie een pilootonderzoek uit bij 1000 Vlamingen. Een opmerkelijk resultaat situeert zich tussen de verschillende geslachten met betrekking tot de antwoorden op de vraag: "Hoe vaak speelt u thuis muziek af per week?"

	< 10	> 10
Man	153	296
Vrouw	221	330

- Bereken het meest relevante percentageverschil.
- Bereken de oddsratio mannen ten opzichte van vrouwen (>10 vs < 10).
- Bereken chi-kwadraat voor deze tabel.
- Bereken Cramer's V. Wat kun je hieruit besluiten?

2

De Vlaamse minister van onderwijs start een grootschalig onderzoek op naar pestgedrag in het Vlaamse secundair onderwijs. Check-It, het ondertussen wereldvermaarde onderzoeksbureau, bezorgt binnen de kortste keren het onderzoeksrapport aan de minister. Eén van de resultaten belicht het eventuele geslachtsverschil in het pesten.

	Jongens	Meisjes	Totaal
Niet pester	4613	5530	10143
Pester	1132	568	1700
Totaal	5745	6098	11843

- Bereken de oddsratio van jongens tegenover meisjes (pesters vs niet-pesters) en interpreteer.
- Bereken chi-kwadraat en interpreteer.

3

Het centrum voor Gelijke Kansen en Racismebestrijding voert een onderzoek uit. Ze wensen na te gaan of allochtonen benadeeld worden op de hedendaagse arbeidsmarkt. Daartoe trekken ze een steekproef uit ex-werkzoekenden die aangesloten waren bij de VDAB. Men vraagt de respondenten hoe vaak men uiteindelijk moest solliciteren voordat men zijn huidige werk gevonden had. Uiteindelijk komt men tot onderstaande kruistabel.

Etniciteit	Aantal keer gesolliciteerd				
	1	2	3	4	5
Autochtoon	32	15	8	4	1
Allochtoon	19	14	7	5	3

- Bereken de oddsratio van autochtoon tov allochtoon voor respondenten die tweemaal vs viermaal gesolliciteerd hebben. Wat zegt dit getal je?
- Bereken Cramer's V en interpreteer.

4*

Het Franse Centre de Sport d'Hiver Transalpine (CSHT) merkt dat skiën de laatste jaren weer populairder wordt dan snowboarden in de Franse Alpen. Om na te gaan of het wel nog zin heeft om ieder jaar snowparken en boardercrosses aan te leggen, nemen ze het onderzoeksbureau Check-It onder de arm om een grootschalig onderzoek uit te voeren bij de sneeuwsportbeoefenaars van het seizoen 2009-2010. Eén van de vragen was of er een leeftijdsverschil optreedt tussen snowboarders en skiërs. Hieronder krijg je de kruistabel die de variabelen leeftijd (minder of meer dan 60 jaar) en type sneeuwsportbeoefening met elkaar in verband brengt.

	Snowboarder	Skiër
-60	296	473
+60	128	725

- Bereken het meest relevante percentageverschil.
- Bereken de gestandaardiseerde residuen. Wat kun je hieruit besluiten?
- Bereken chi-kwadraat en phi. Wat kun je hieruit besluiten?

5

Een jonge onderzoeker in de tandheelkunde zit op een avond naar tv te kijken. Tijdens de commerciële boodschappen passeert een reclame die het merk 'Snowy White' aanprijst. De onderzoeker gelooft weinig van de reclamespot en zet een experiment op waarbij hij een groep mensen 20 jaar lang het product laat gebruiken en een andere groep mensen – een testgroep – een ander, niet gedefinieerd product laat gebruiken. 20 jaar later, wanneer de onderzoeker al professor is, komen de resultaten binnen. De professor bekijkt onmiddellijk de tabel die het aantal gaatjes dat de respondent opgelopen heeft in de voorbije 20 jaar in verband brengt met het product dat hij gebruikte:

- Stel een grafiek op van de marginale verdeling van het aantal gaatjes dat mensen opgelopen hebben in die 20 jaar.
- Bereken het gemiddeld aantal gaatjes dat alle mensen in de steekproef opgelopen hebben.
- Bepaal het percentageverschil tussen de mensen die 'Snowy white' gebruikten en zij die het zwart product gebruikten wat betreft het hebben van 6-8 gaatjes.
- Bereken de odds van de mensen die 'Snowy White' gebruikten op het hebben van 3 tot 5 gaatjes versus 0-2 gaatjes hebben
- Bereken de oddsratio van zwart product tov Snowy white voor 6-8 gaatjes vs 0-2 gaatjes

Aantal gaatjes	Merk tandenborstel	
	Snowy White	Zwart product
0-2	23	13
3-5	35	17
6-8	10	28

6

In opdracht van een Vlaams tienertijdschrift hebben Job en Saskia afgelopen zaterdag 300 toevallige voorbijgangers bevroegd in de Veldstraat in Gent. Ze vroegen aan elk van hen hoe vaak ze thuis op de weegschaal gaan staan per maand. In onderstaande tabel worden de resultaten weergegeven.

	Hoe vaak ga je op de weegschaal staan?			
	Nooit	≥ 0 en ≤ 5	> 5 en ≤ 10	> 10
Man	83	10	17	20
Vrouw	39	68	39	24

- Bereken de odds voor vrouwen om tussen de 0 en 5 keer op de weegschaal te staan tegenover niet op de weegschaal te staan
- Bereken chi-kwadraat voor deze tabel
- Bereken Cramer's V. Wat kun je hieruit besluiten?

7

Roger, een rasechte politieman houdt in zijn vrije tijd wat statistieken bij over de behandelde misdaden op zijn afdeling. In zijn laatste rapport besteedt hij uitvoerig aandacht aan het belang van ervaring voor het oplossen van een misdaad met licht geweld.

	Niet opgelost	Opgelost
Met ervaring	43	93
Zonder ervaring	65	97

- Bereken de chi-kwadraat en interpreteer het resultaat
- Wat is de odds op het niet oplossen van een misdaad voor agenten met ervaring?

8

Op het einde van het schooljaar organiseert de jeugdbeweging van Elke elk jaar een klein festival. Ze bouwen hiervoor twee verschillende podia op. In een eerste tent (=rockcity) programmeren ze hoofdzakelijk rock muziek, terwijl in de tweede tent (=techtown) enkel elektronische muziek wordt geprogrammeerd. Om een beter zicht te kunnen krijgen op de leeftijdssamenstelling van de fuifgangers in elke tent doet Elke dit jaar een rondvraag op het festival. Het aantal aanwezige fuifgangers in elke tent, per leeftijdscategorie vind je terug in onderstaande tabel.

	< 14 jaar	> 13j en < 17	> 16j en < 21	>20
Rockcity	39	117	43	183
Techtown	21	18	68	192

- Bereken de chi-kwadraat en interpreteer het resultaat
- Wat is de odds om ouder dan 20 jaar te zijn in Rockcity?
- In welke tent vind je proportioneel gezien de meeste min 14 jarigen?
- Wat is de odds om jonger dan 14 te zijn in Techtown?

9

De Gentse huisvestingsdienst gaat na hoeveel studenten op kot zitten in Gent en hoeveel telkens naar het ouderlijke huis pendelt. Volgende tabel toont, voor mannelijke en vrouwelijke studenten uit 1e bachelor aan de universiteit Gent, wie wel (ja) en niet (nee) op kot zit.

	Nee	Ja
Man	375	298
Vrouw	342	433

- Bereken het meest relevante percentageverschil
- Bereken de oddsratio van mannen tov vrouwen voor wel (tegenover niet) op kot zitten.

2.2. Relaties tussen metrische variabelen

1*

Geef voor elk geval aan wat er mis is.

- 'Er bestaat een grote correlatie tussen de leeftijd van Vlamingen en het feit of ze al dan niet kaal zijn.'
- 'We vonden een grote correlatie ($r = 1,09$) tussen de door studenten gegeven beoordeling over de kwaliteit van het lesgeven van stafleden en de gehaalde studieresultaten'
- De correlatie tussen de hoeveelheid regenval en het aantal maïsplanten in een veld bleek $r = 0,23$ l/m² te zijn

2

Het eenzame hartenbureau krijgt een klant over de vloer. Het valt de medewerkers onmiddellijk op dat de vrouw zeer korte benen heeft. Dit blijkt ook haar grootste zorg te zijn: "Vrouwen met korte benen geraken nooit aan een man!" klaagt ze. Het bureau belooft haar de zaak tot op de bodem uit te zoeken en zet begint enkele vroegere klanten op te bellen, op toevallige basis. Uiteindelijk presenteert het eenzame hartenbureau volgende resultaten aan de vrouw, als ze de volgende keer langskomt. Het gaat over een datamatrix, waarbij de benenlengte van de ondervraagde vrouwen in verband wordt gevraagd met het aantal ex-liefdespartners dat de vrouw gehad heeft.

	Beenlengte	Aantal ex-partners
1	102	4
2	98	3
3	111	3
4	105	6
5	93	1

- Bereken de covariantie
- Bereken de correlatiecoëfficiënt en interpreteer
- Stel de ongestandaardiseerde regressievergelijking op
- De vrouw met de korte benen heeft een benenlengte van 83 cm. Hoeveel ex-liefdespartners moet ze volgens het eenzame hartenbureau al gehad hebben?

3*

Jos Vercammen, een leraar wiskunde uit Ninove, wil nu eens voor eens en voor altijd de proef op de som nemen en nagaan of kinderen met een hoger IQ het inderdaad beter doen in zijn klas. Hij trekt een EAS van 4 leerlingen uit zijn klas, onderwerpt hen aan een IQ-test en vergelijkt deze gegevens aan het eind van het schooljaar met hun eindpercentages. De gegevens vind je hieronder terug:

IQ	Eindscore Secundair onderwijs
98	65
103	80
135	80
70	85

- Hoeveel bedraagt n hier?
- Teken de scatterplot
- Bereken de covariantie en de correlatiecoëfficiënt. Hoe sterk is de samenhang tussen beide variabelen?
- Bepaal de ongestandaardiseerde en gestandaardiseerde regressievergelijking.
- Teken de (ongestandaardiseerde) regressierechte in je scatterplot
- Als je ervan zou uitgaan dat het toch bewezen is, ondanks deze data, dat IQ en eindscore positief samenhangen, kun je dan verklaren waarom we hier zo'n zwakke en kleine relatie vinden, die bovendien, tegen de verwachtingen in, zelfs negatief is?

4

Een schacht uit Politeia moet als opdracht voor zijn doop een statistische opgave uitwerken. De praeses geeft hem als opdracht dat hij op de openingscantus moet optekenen hoeveel pinten iedereen gedronken heeft en wat het promille alcohol in hun bloed is. De schacht maakt onderstaande tabel op:

- Teken de scatterplot van deze bivariate verdeling.
- Hoeveel bedraagt de covariantie hier? En de correlatiecoëfficiënt? En de determinatiecoëfficiënt? Hoe interpreteer je deze laatste maat?
- Een student, die meedoet aan de volgende cantus, drinkt 15 pinten. Hoeveel promille alcohol verwacht je in zijn bloed?
- Bereken de residuen.

	Nummer van de schacht									
	1	2	3	4	5	6	7	8	9	10
Pinten	8	4	7	8	7	5	3	2	9	11
Promille	1,18	0,43	0,82	1,43	1,63	0,67	0,65	0,20	1,49	1,35

5

Een tekenfilmstudio voert een publieksonderzoek uit. Ze willen nagaan welke kenmerken van hun films aansluiten bij hun doelpubliek: de kinderen jonger dan 14 jaar. Een opmerkelijke vaststelling van het onderzoek was de relatie tussen het aantal kijkers van hun doelpubliek en de duur van de film. Hieronder krijg je gegevens van zes van hun films:

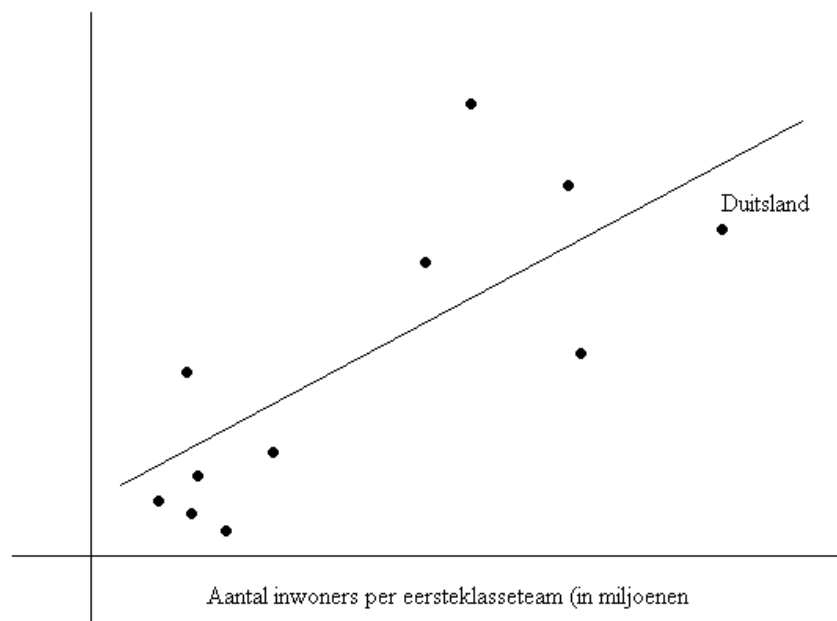
	Duur van de film (in minuten)	Aantal bezoekers < 14 jaar
Sjake en de droomfabriek	74	73
De grote zeemeerman	81	78
Knaagje en het moerasmonster	97	68
Picasso: Back to Guernica	69	81
Peter Pot	63	89
Wonderland 2: de wraak van Alice	89	61

- Hebben we hier te maken met een symmetrische of een asymmetrische relatie? In het laatste geval, wat is de onafhankelijke en wat is de afhankelijke variabele?
- Bereken de covariantie en de correlatie tussen beide variabelen. Interpreteer.
- Stel de ongestandaardiseerde en de gestandaardiseerde regressievergelijking op.
- Indien een film twee uur duurt, hoeveel bezoekers onder de 14 jaar verwachten we dan?
- Indien de duur van een film 0,6 standaardafwijkingen boven de gemiddelde duur van de gegeven 6 films ligt, hoeveel standaardafwijkingen boven of onder het gemiddeld aantal bezoekers onder de 14 jaar van de gegeven 6 films ligt het aantal bezoekers onder de 14 jaar van die film dan? Over hoeveel bezoekers gaat het dan in absolute aantallen?
- Hoeveel procent van de variantie in het aantal bezoekers onder de 14 jaar wordt verklaard door de duur van de film?

6*

Een onderzoek wil nagaan waarom in sommige landen de voetbalclubs over een veel groter budget beschikken dan in andere landen. De hypothese luidt dat hoe groter de potentiële 'thuismarkt' is, hoe groter het budget is. We bekijken 11 Europese landen. We zijn geïnteresseerd in het effect van aantal inwoners per eersteklasseteam (het aantal inwoners in een land, in miljoenen, gedeeld door het aantal voetbalteams in eerste klasse) op het gemiddelde budget van een eersteklasseteam (uitgedrukt in miljoenen EURO). We weten dat B 0,74 bedraagt, dat b 7,32 bedraagt en het intercept 3,59 bedraagt. Welke van de onderstaande beweringen zijn correct? Meerdere correcte beweringen kunnen voorkomen.

Gemiddelde budget eersteklasseteams (in milj. Euro)



- 74% van de variantie in gemiddeld budget kunnen we verklaren door (de variantie in) het aantal inwoners per eerste klasse team
- Een Europees land is hier een statistische eenheid
- Vermits we hier te maken hebben met een positief verband kunnen we besluiten (verwachten) dat hoe meer inwoners per eerste klasse team er in een bepaald land zijn, hoe groter het gemiddelde budget van een eerste klasse team in dat land zal zijn
- In een land waar er 12 teams uitkomen in eerste klasse, met 10 350 000 inwoners, verwachten we een gemiddeld budget van 9,9 miljoen Euro per eerste klasse team
- Duitsland heeft het hoogste gemiddelde budget per eerste klasse team

Guido, sinds jaar en dag de vaste leerkracht lichamelijke opvoeding van een stedelijke basisschool brengt de conditie van zijn leerlingen in kaart. Hij maakt hiervoor gebruik van hun resultaten op de gevreesde biep-test (of shuttle run test). De directeur van de school, meester Vermaert, is een trouw voorstander van de leuze: "Mens sana in corpore sano". Daarom vraagt hij zich af of er een verband is tussen de resultaten van de lichamelijke test en de resultaten van de leerlingen op het algemeen eindexamen. In onderstaande tabel vind je de resultaten terug van klas 6b voor de biep-test (in aantal minuten dat de leerling de test kan volhouden) en hun respectievelijke resultaten op het examen.

- Hoeveel bedraagt n hier?
- Teken de scatterplot. Verwacht je een positieve of negatieve relatie?
- Bereken de covariantie en de correlatiecoëfficiënt. Interpreteer.
- Bepaal de ongestandaardiseerde en de gestandaardiseerde regressievergelijking.
- Vul de tabel aan met een extra kolom die de gestandaardiseerde waarden bevat.
- Bereken nog een extra kolom met de residuen.

Biep-test (in min.)	Eindscore Secundair onderwijs
9	65
10	73
8	70
14	79
15	89
12	69
9	65
10	67
12	72
17	83
7	70
6	61
13	76
15	87
8	67
9	68
10	70
13	73
14	78

8

Frank is ervan overtuigd dat het aantal kijkers naar het weerbericht dat hij dagelijks presenteert toeneemt als de temperaturen stijgen. Hij denkt immers dat mensen dan meer interesse hebben in het weer. Zijn collega, Sabine, denkt echter dat de kijkcijfers zullen afnemen als buiten de temperaturen toenemen. Zij gelooft dat mensen geen zin om binnen te zitten kijken naar de televisie als het buiten goed weer is. Om hun gelijk te bewijzen vragen ze de kijkcijfers op van de afgelopen twee weken bij de studiedienst van de openbare omroep. De resultaten vind je terug in onderstaande tabel.

- Teken de (ongestandaardiseerde) regressierechte
- Hoeveel kijkers verwacht je als de temperatuur morgen 27 graden bedraagt?
- Bereken de determinatiecoëfficiënt en interpreteer.
- Wie heeft er gelijk, Sabine of Frank?

Temperatuur	Kijkcijfers x 1000
21	121
18	132
23	123
20	146
21	122
24	118
29	91
32	89
31	101
29	76
33	86
28	84
24	120
21	143

9*

Hangt het dagelijks eten van fruit samen met een betere gezondheid? Een bivariate regressieanalyse gaat dit na. De onafhankelijke variabele is FRUIT (hoeveel stukken fruit at men tijdens de voorbije maand gemiddeld per dag), de afhankelijke variabele is ZIEKTE (hoeveel dagen bleef men tijdens het voorbije half jaar thuis omwille van ziekte). Onderstaande tabel toont de resultaten van de bivariate regressieanalyse. Welke van de volgende beweringen zijn correct? Meerdere correcte beweringen kunnen voorkomen.

Effect van Fruit op Ziekte			
Onafhankelijke	B	b	a
Fruit	-0,58	-1,14	6,21

- Afgerond 34% van de variantie in ZIEKTE wordt verklaard door de variantie in fruit
- Indien Els gemiddeld per dag twee stukken fruit meer at dan Marie, dan verwachten we dat Els afgerond 2 dagen minder thuis bleef wegens ziekte dan Marie
- Voor iemand die 0 dagen thuis bleef wegens ziekte, verwachten we dat hij of zij 6,21 stukken fruit at
- Bij een toename met één standaardafwijking in het aantal stukken fruit dat men at, verwachten we een afname met 1,14 standaardafwijkingen in het aantal ziektedagen
- Joris at gemiddeld 2,5 stukken fruit per dag en bleef het vorige halve jaar 4 dagen thuis. De residueel bedraagt afgerond 0,64

10

Onderstaande tabel toont afwijkingen ten opzichte van het gemiddelde op variabele X en afwijkingen ten opzichte van het gemiddelde op variabele Y, voor vijf eenheden. We geven je ook de varianties mee: $s^2_x = 174,24$; $s^2_y = 252,492$.

$(x_i - \bar{x})$	$(y_i - \bar{y})$
-20,20234	-24,99958
-2,20031	-3,00004
0,80005	3,99999
5,80061	5,99991
15,80185	17,99972

- Bereken de covariantie.
- Bereken de correlatiecoëfficiënt.
- Bereken de ongestandaardiseerde richtingscoëfficiënt.
- Bereken de volledige gestandaardiseerde regressievergelijking (we gaan uit van een effect van x op y).

11

Archeoloog Ernest Voetjes heeft een nieuwe theorie gelanceerd: hij vermoedt dat het hersenvolume tijdens de menselijke evolutie continu aangroeide tijdens het Pleistoceen en dat er dus een lineaire relatie bestaat tussen het hersenvolume van primaten en het tijdsverloop. Hij veronderstelt dat het hersenvolume in het jaar – 5 Mj (miljoen jaar) 540 cc bedroeg en dat het hersenvolume toeneemt met 20 cc per Mj. Wat is volgens Voetjes de regressievergelijking tussen deze twee variabelen? (tip: blijf als eenheid voor het jaartal Mj gebruiken)

12

De twee populairste dranken, na het traditionele pilsbier, op de jaarlijkse Gentse feesten zijn Duvel en Irish koffie. Een gerenommeerd drankenfabrikant heeft op de eerste zaterdagavond bijgehouden hoeveel Duvels en Irish coffees er per uur zijn verkocht.

Uur	Duvels (x100)	Irish coffee (x100)
22u	28	0
23u	32	1
24u	36	1
1u	38	3
2u	21	5
3u	20	7
4u	18	9
5u	14	10
6u	10	11
7u	4	10

- Wat is het meetniveau van de verschillende variabelen?
- Stel een bivariate scatterplot op.
- Kunnen we hier de correlatiecoëfficiënt berekenen? Zo ja, hoeveel bedraagt deze?
- Kunnen we hier regressie toepassen? Zo ja, stel de ongestandaardiseerde regressievergelijking op.

13

Een bekend consumentenonderzoeksbureau doet een vergelijkend onderzoek naar de kwaliteit van koffiepads van verschillende merken. Ze geven per fabrikant een score op 10 voor de smaak van hun koffie en geven daarnaast ook weer hoeveel percent van de gemalen koffie in de pads van Braziliaanse afkomst is.

- a) Bereken de best passende associatiemaat
- b) Veronderstel dat de smaak van de koffie afhankelijk is van het percentage van de gemalen koffie in de pads dat van Braziliaanse afkomst is. Bereken de regressierechte.
- c) Er komt een nieuwe fabrikant op de markt wiens koffiepads 40% Braziliaanse koffie bevat. Hoeveel punten verwacht je dat die zal krijgen?

Fabrikant	Braziliaanse afkomst (in %)	Smaak (op 10)
1	36	6
2	18	3
3	11	4
4	24	5
5	45	7
6	63	8
7	4	4
8	13	4
9	74	8
10	4	3

14

Nele, een maritiem biologe, staat bijna elke dag in de file wanneer ze van haar werk naar haar huis, in Oostende, terugkeert. Onlangs heeft ze opgemerkt dat ze de lengte van de file tijdens de zomer kan voorspellen aan de hand van het weerbericht de dag voordien. De laatste twee weken heeft ze de lengte van de file en het voorspelde aantal graden bijgehouden.

Lengte in km	Voorspelde graden
3	17
9	20
10	36
2	15
4	16
12	34
6	17
8	21
1	15
2	16
0	13
12	30
10	28
9	27

- Stel de regressievergelijking op aan de hand van deze tabel.
- Ze voorspellen morgen 15 graden, hoeveel kilometer file verwacht je?
- Bereken de correlatiecoëfficiënt en interpreteer.
- Het effect van het voorspelde weer op de lengte van de file is veel groter als er ook geen regen wordt voorspeld. Teken aan de hand van een pijltjesdiagram deze relaties en bespreek.
- Hoeveel percent van de variantie in de lengte van de file wordt hier verklaard?

15

In onderstaande tabel staan de afwijkingen van het gemiddelde van twee variabelen. De variantie van x bedraagt 21,6; de variantie van y 46,4. Het gemiddelde van x bedraagt 4, het gemiddelde van y is 10.

$x - x_{\text{gem}}$	$y - y_{\text{gem}}$
1	9
2	-1
-1	3
7	4
-2	-5
-7	-10

- Bereken de covariantie en de correlatiecoëfficiënt. Interpreteer.
- Hoeveel procent van de variantie in y wordt verklaard door de variantie in x ?
- Stel de ongestandaardiseerde en de gestandaardiseerde regressievergelijking op.
- Welke waarde van y verwachten we als $x = 6$?

16*

John, de coach van het Gentse Universitaire Roei team, is sterk geïnteresseerd in statistiek en wil zijn kennis daarvan toepassen op zijn coachwerk voor het team. Een maand lang houdt hij bij hoe lang zijn team trainde per training en hoe snel ze telkens de volgende dag de Gentse watersportbaan aflegden. Hij ziet de duur van de training als onafhankelijke variabele en de tijdsduur waarbinnen zijn team de volledige lengte van de watersportbaan kan afleggen als afhankelijke. Beide zijn in minuten uitgedrukt. De tabel met resultaten vind je hieronder terug.

	Nummer van de trainingssessie							
	1	2	3	4	5	6	7	8
Duur training	86	51	73	96	56	83	69	83
Tijdsduur	25	29	28	18	29	28	29	28

- Hoeveel bedraagt de correlatiecoëfficiënt? Interpreteer.
- Hoeveel bedraagt de determinatiecoëfficiënt? Interpreteer.
- Het team doet mee aan een wedstrijd op de Watersportbaan: om ter snelst naar het andere eind. Als de coach zijn team 78 minuten laat trainen, hoe snel kunnen we dan verwachten dat het team erover zal doen?
- Bereken de residuele variantie.

2.3. De limieten van de beschrijvende statistiek

1

Geef uw kind niet zomaar een naam

wo 09/09/2009 - 11:00

U denkt beter twee keer na over welke naam u aan uw kind geeft. Hoewel een naam niet de oorzaak is van slechte schoolprestaties, is er wel degelijk een verband tussen bijvoorbeeld Kelly en een minder rapport.



De naam Vicky kreeg met de serie Little Britain een ietwat marginale bijklank

Uit een peiling van een Britse oudervereniging blijkt dat een op de drie Britse leerkrachten "verwacht" dat iemand met een bepaalde naam meer aandacht nodig zal hebben in de klas. Als er een Chelsea of Jessica in de klas zit, denkt de leraar of lerares sowieso dat het om lastposten gaat. De helft van de leerkrachten heeft al een oordeel klaar als hij voor de start van het schooljaar de namenlijst overloopt. Ongeveer 70 procent heeft het zelfs moeilijk om "originele" namen van een aantal leerlingen uit te spreken.

Of Belgische leerkrachten ook dezelfde vooroordelen hebben is nog niet echt onderzocht, maar een studie van de VUB stelde een paar jaar geleden wel vast dat er een verband is tussen de naam van een kind en zijn of haar schoolprestaties. Uiteraard is het niet zozeer de naam, maar wel de omgeving waarin het kind opgroeit cruciaal voor de ontwikkeling. Laagopgeleide ouders geven hun kinderen meestal namen zoals Kelly, Tamara en Cindy, terwijl hoogopgeleiden eerder Thomas, Ruben of Eva prefereren.

Het is bekend dat kinderen uit hoogopgeleide gezinnen meer bagage en kansen meekrijgen en bijgevolg ook beter scoren op school. Dat versterkt dan nog eens het beeld dat bestaat over die namen. Een vicieuze cirkel met andere woorden, waarin reële achterstand en hardnekkige vooroordelen hand in hand gaan.

Bron: www.deredactie.be, 09/09/2009

- Welk soort verband wordt naar voren geschoven in de eerste alinea onder de foto?
- Welk soort verband is de verklaring in de tweede alinea?

2

Onderzoek van het Centrum voor Genderstudies wees op een wel heel opmerkelijk verband tussen de kans op het succesvol liften en het tonen van de bovenbenen. Bij vrouwen is het namelijk zo dat het tonen van die bovenbil tot gevolg heeft dat die kans stijgt, terwijl die kans net daalt wanneer mannen hun bovenbil tonen. Welk soort verband vindt het centrum hier?

3

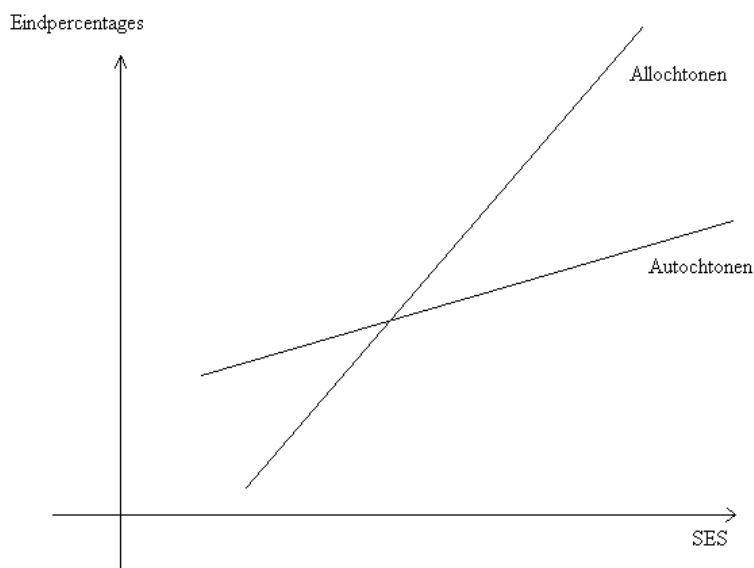
Onderzoek toont aan dat studenten die veel bier drinken, gemiddeld lagere punten behalen op het vak statistiek dan studenten die weinig bier drinken. De studie verklaarde dit door het aantal hersencellen dat gedood wordt door overmatige bierconsumptie. Welk soort verband toont deze studie aan?

4

De Koepel van studentenverenigingen vecht bovenstaande resultaten aan. 'Wij geloven niet dat alcohol de hersencellen doodt, maar dat er een effect is van het drinken van veel bier staat blijkaar vast. De universiteit zorgt er echter voor dat de lessen zeer vroeg vallen, waardoor zeker een deel van de resultaten van onze studenten verklaard wordt door hun slaapttekort!' Welk soort verband is hier volgens de Koepel aan het werk?

5*

Het Minderheden Forum publiceerde eind 2009 haar rapport over de situatie van allochtonen in het Vlaamse onderwijssysteem. Eén van de belangrijke bevindingen had te maken met de sociaal-economische status van allochtonen, opgesteld aan de hand van het opleidingsniveau en het inkomen van hun ouders. De bevinding wordt grafisch uitgebeeld in onderstaande figuur:



- Leg in je eigen woorden wat hier aan de hand is.
- Welk soort pijljesdiagram kun je hierbij tekenen?

6

De Orde der Geneesheren krijgt een klacht binnen van een patiënt, die al jaren in het ziekenhuis 'Vol goede hoop' verblijft. De patiënt hield al geruime tijd een bestandje bij van de patiënten die passeren in het ziekenhuis. Hij vroeg hen telkens hoe vaak ze in hun leven al opgenomen zijn in het ziekenhuis en bracht dit in verband met het aantal tanden die ze nog in hun mond hadden zitten. Hij kwam tot de onthutsende ontdekking dat er een sterk verband bestaat tussen het aantal ziekenhuisopnames en het aantal tanden. De Orde krijgt volgende tabel onder ogen:

Aantal ziekenhuisbezoeken	Aantal tanden	
	0-16	17-32
0-5	17	67
6-10	11	40
11-15	9	18
16-20	48	9
21-25	63	11

- a) Bereken chi-kwadraat en Cramer's V van bovenstaande tabel. Klopt het dat er zo'n sterk verband is?
- b) De Orde verwacht dat er meer aan de hand is. Ze vraagt de oorspronkelijke data op van de patiënt en maakt een leeftijdsverdeling van de respondenten. Uiteindelijk bekomt men volgende driedimensionele tabel:

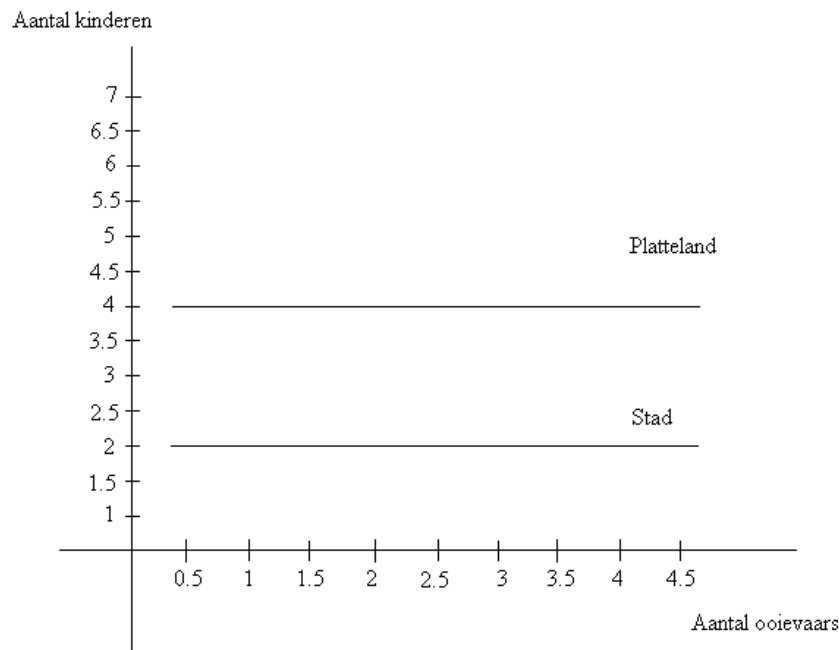
Aantal ziekenhuisbezoeken	-50		+50	
	Aantal tanden		Aantal tanden	
	0-16	17-32	0-16	17-32
0-5	11	66	6	1
6-10	7	39	4	1
11-15	3	17	6	1
16-20	1	3	47	6
21-25	1	3	62	8

- c) Bereken chi-kwadraat en Cramer's V voor beide subtabellen (dus voor de -50-jarigen en de +50-jarigen). Vind je nog steeds een sterk verband tussen aantal ziekenhuisbezoeken en aantal tanden? Wat is hier aan de hand?

7

Rick Birdie, leider van de radicale beweging 'Storks did it!' en dan ook een notoir tegenstander van de theorie dat kinderen gemaakt worden door man en vrouw, heeft eindelijk bewijs gevonden voor zijn bewering. Hij nam een grootschalige steekproef uit de Belgische bevolking en bracht twee variabelen met elkaar in verband: het aantal ooievaars dat overvliegt en het aantal kinderen dat geboren wordt. Zijn resultaten, die hij opstuurt naar het tijdschrift 'Nature', tonen aan dat de correlatie tussen de twee variabelen 0,76 bedraagt. De standaardafwijking van de variabele 'aantal ooievaars' bedraagt 0,6, op een gemiddelde van 0,8. De standaardafwijking van het aantal baby's is 0,46, op een gemiddelde van 3,3.

- Stel de enkelvoudige regressievergelijking op van de bivariate relatie tussen 'aantal ooievaars' en 'aantal baby's'. Teken de rechte in een assenstelsel.
- De vereniging 'it's all love' gelooft de bewering van Rick Birdie niet en denkt dat het verband eigenlijk aan een andere variabele ligt: namelijk of de mensen op het platteland of de stad woonden. Ze publiceren volgende figuur in het volgende nummer van 'Nature'. Wat is er hier aan de hand?



8*

Ishane wil aan de hand van onderstaande tabel meer te weten komen over de attitudes omtrent holibi's in Marokko. De attitudes omtrent holibi's wordt gemeten aan de hand van een schaal van 0 tot en met 10, waarbij een hogere score een meer negatieve attitude weergeeft. Het eerste model houdt rekening met het effect van sociaaleconomische status (een hoger cijfer betekent een hogere status) en het aantal jaren dat men studeerde. Het tweede model bevat tevens het effect van de urbanisatiegraad van de woonplaats, waarbij een hoger cijfer op meer verstedelijking wijst. In model drie wordt gecontroleerd voor leeftijd (gemeten in aantal jaren). In het laatste en vierde model wordt tot slot gecontroleerd voor het aantal broers of zussen.

	Model 1	Model 2	Model 3	Model 4
	b	b	b	b
Aantal jaren gestudeerd	-0,511	-0,466	-0,387	-0,311
Sociaaleconomische status	-0,478	-0,298	-0,241	-0,212
Urbanisatiegraad woonplaats		-0,078	-0,077	-0,048
Leeftijd			0,002	0,002
Aantal broers of zussen				0,000
R ²	0,13	0,18	0,22	0,21

- Interpreteer het effect van het aantal jaren gestudeerd op de attitude omtrent holibi's in model 1.
- Als we de urbanisatiegraad van de woonplaats in rekening brengen in model 2, wat gebeurt er dan met het effect van sociaaleconomische status dat reeds in model 1 opgenomen werd? Interpreteer en teken bijhorend pijltjesdiagram.
- Als we leeftijd in rekening brengen in model 3, wat gebeurt er dan met de variabele urbanisatiegraad woonplaats uit model 2? Interpreteer en teken bijhorend pijltjesdiagram.
- Welk model heeft de grootste verklarende kracht?
- Is het nuttig om de variabele Aantal broers of zussen aan het model toe te voegen?
- Welke variabele voegt het meest verklarende kracht toe?
- Welke variabele heeft in model 3 het grootste effect op attitudes omtrent holibi's?

3. Inductieve statistiek

3.1. De studie van de systematiek van het toeval: kansrekenen

3.2. Stochastische variabelen

3.3. Studie van steekproefvariabiliteit

3.4. Basistools van de inductieve statistiek

3.5. Inferentie voor verwachtingen

1*

Natuurpunt Vlaanderen doet een onderzoek. Ze willen nagaan hoeveel bomen er gemiddeld in een bos staan in Vlaanderen. Ze weten uit vorig onderzoek dat de standaardafwijking van het aantal bomen per Vlaams bos 2483 bedraagt. Uit een EAS van 83 bossen berekent men een gemiddeld aantal bomen per bos van 8496, met een standaardafwijking van 1800.

- Geef het 95%-betrouwbaarheidsinterval op.
- Zonder het te berekenen, zal het 86%-betrouwbaarheidsinterval smaller of breder zijn dan het 95%-betrouwbaarheidsinterval?
- Bereken nu het 86%-betrouwbaarheidsinterval.
- Zonder het te berekenen, zal een 95%-betrouwbaarheidsinterval in een steekproef met 150 bossen smaller of breder zijn dan diegene die je berekend hebt in deelvraag a?
- Indien Natuurpunt een maximale foutenmarge van 200 vooropgesteld had voor een 95%-betrouwbaarheidsinterval, hoeveel bossen hadden ze dan in de steekproef moeten opnemen?

2*

De organisatie van Rimpelrock wil nagaan of ze inderdaad de ouderen uit onze Vlaamse populatie aantrekken. Ze weten dat de gemiddelde leeftijd van de Vlaamse bevolking 46 jaar bedraagt, met een standaardafwijking van 15,76. Ze verwachten dus dat de leeftijd van de mensen die naar het festival komen, hoger ligt dan dit populatiegemiddelde. De organisatie neemt een EAS van 25 man uit de mensen die ooit het muziekfestival bijwoonden. De gemiddelde leeftijd bedroeg 57,6, met een standaardafwijking van 9,4.

- a) Stel de hypothesen op van de significantietoets. Moeten we twee- of eenzijdig toetsen?
- b) Wat is de verwachting van het steekproefgemiddelde? En de standaardafwijking?
- c) Bereken de toetsingsgrootte.
- d) Wat is de overschrijdingskans?
- e) Is er een significant verschil tussen de leeftijd van de mensen die naar Rimpelrock komen en de totale Vlaamse populatie? Op welk niveau is dit significant?
- f) Wat kunnen we besluiten uit deze significantietoets?

3*

Een leraar uit het Lager Onderwijs in Neder-Overheembeek weet niet of zijn leerlingen het beter of slechter doen dan het landelijke gemiddelde op een algemeen aanvaarde leesniveau-test. Dat landelijk gemiddelde bedraagt volgens het CLB 65, met een standaardafwijking van 8,63. De leraar trekt een EAS van 12 leerlingen uit zijn klas. Hun gemiddelde bedroeg 58, met een standaardafwijking van 6,32.

- a) Is er een significant verschil tussen het leesniveau van zijn leerlingen en de totale leerlingenpopulatie in Vlaanderen? Op welk niveau is dit significant?
- b) Wat kunnen we besluiten uit deze significantietoets?

4*

Van de hele populatie UGent-studenten weten we dat ze gemiddeld 6,5 glazen alcohol per week drinken. We vragen ons nu af of studenten aan de faculteit PSW significant verschillen van de totale populatie en voeren daarom een onderzoek uit bij een EAS van 132 studenten. Het gemiddelde aantal glazen dat zij per week dronken was 8,19 met een standaardafwijking van 11.

- a) Kunnen we op het 0,05 niveau besluiten dat er een significant verschil is tussen de gehele studentenpopulatie en de studenten aan de faculteit PSW?
- b) Stel dat het populatiegemiddelde niet gekend is. Bouw een 95%-betrouwbaarheidsinterval.

5*

Volgens een studie van de Weight Watchers bedraagt voor vrouwen tussen de leeftijd van 40 en 50 het gemiddelde van de dagelijkse calorie-inname 1781 calorieën per dag. De standaardafwijking is 31 calorieën. Je neemt zelf een EAS van 200 vrouwen uit deze leeftijdsgroep.

- a) Wat is de standaardafwijking van het steekproefgemiddelde?
- b) Wat is de foutenmarge van het betrouwbaarheidsinterval bij 95 % betrouwbaarheid?

6*

De Wereldgezondheidsorganisatie gaat na of de verspreiding van Mexicaanse griep verschilt tussen het Europese continent en Afrika. Daartoe neemt ze een EAS van Europese landen en Afrikaanse landen. In de 13 geselecteerde Europese landen worden gemiddeld 21 mensen ziek, met een standaardafwijking van 4,3, terwijl de resultaten in de 21 geselecteerde Afrikaanse landen uitwijzen dat daar gemiddeld 26 mensen per dag ziek worden, met een standaardafwijking van 8,3.

- a) Bouw een 80%-betrouwbaarheidsinterval op rond dit verschil.
- b) Is dit verschil significant?

7*

Je wil een steekproefgemiddelde bepalen, zodanig dat er slechts een kans van 5 % bestaat dat een steekproef van 22 kinderen gemiddeld genomen dit gemiddelde of lager behaalt. Je weet dat de verwachting in de populatie (met normaal verdeelde scores) 13,6 bedraagt en de standaardafwijking in de populatie gelijk is aan 3,1. Bepaal dit steekproefgemiddelde.

8

In een onderzoek naar wangedrag van kinderen op school werd aan leerkrachten gevraagd 282 leerlingen, die via een EAS werden uitgekozen, een cijfer te geven voor uiteenlopende aspecten die betrekking hadden op de mate waarin het kind wangedrag vertoont op school. Een aspect was: "spijbelt de leerling vaak?". De cijfers werden uitgedrukt op een vijf-puntsschaal van 0 tot 4. Het gemiddelde was 2,22 met standaardafwijking 1,03.

- a) Stel de steekproevenverdeling op. Is deze normaal verdeeld? Zo ja, leg uit waarom.
- b) Mogen we hier een betrouwbaarheidsinterval op berekenen?
- c) Als het mogelijk is, bepaal dan het 99%-betrouwbaarheidsinterval.

9

De federale overheidsdienst voor volksgezondheid heeft vastgesteld dat een adolescent in Vlaanderen gemiddeld 6 keer frieten eet per week. Ze verwachten dat bij adolescenten die lichamelijke opvoeding studeren dit gemiddelde lager ligt. Ze bevragen 56 studenten LO en stellen vast dat het gemiddelde 4,5 keer per week is, met een standaardafwijking van 1,2.

- a) Wat zijn de parameters van de steekproevenverdeling van het steekproefgemiddelde?
- b) Voer de significantietoets uit.
- c) Stel dat het Vlaamse gemiddelde niet gekend is. Bouw een 90%-betrouwbaarheidsinterval.

10

Twee onafhankelijke enkelvoudige aselechte steekproeven in de maanden juli en september moeten helpen om economische schattingen te verkrijgen rond het verloop van de tarweprijzen in de VS. In een steekproef van 90 metingen tijdens de maand juli bedraagt de gemiddelde tarweprijs \$ 2,95 en de standaardfout van het gemiddelde \$ 0,023. In een steekproef van 45 metingen tijdens de maand september bedraagt de gemiddelde prijs \$ 3,61 en de standaardfout van het gemiddelde \$ 0,029.

- a) Bereken op basis van de standaardfouten ook de standaardafwijkingen in beide steekproeven.
- b) Gebruik een significantietoets om te onderzoeken of de verwachte tarweprijs in juli en september verschilde.

11

Kurt Cobain, frontman van Nirvana, stond erom bekend dat hij geregeld een gitaar naar de verdommenis hielp. Zijn biograaf ontdekte dat het aantal gitaren dat Kurt per maand kapot sloeg, grofweg normaal verdeeld is volgens $N(8, 2,59)$.

- a) Wat is de kans dat Kurt tijdens een willekeurig gekozen maand minder dan 6 gitaren kapot sloeg? Welk soort verdeling gebruiken we om deze vraag op te lossen (2 antwoorden)?
- b) We nemen een steekproef van 16 willekeurige maanden uit de carrière van Kurt Cobain. Wat is de steekproevenverdeling van het gemiddeld aantal gitaren dat Kurt kapot sloeg per maand?
- c) Wat is de kans dat het gemiddeld aantal gitaren dat Kurt kapot sloeg per maand in die steekproef lager ligt dan 6?
- d) Welk percentage leunt het dichtst aan bij de realiteit? Die bij a) of die bij c)?

12

In een discussieprogramma op de radio vraagt men aan luisteraars hoeveel ze verdienen. Iedereen die dat wil mag gewoon bellen en zijn salaris doorgeven. 958 personen reageren en dit levert een gemiddeld salaris op van 14400 euro per jaar en een standaardafwijking van 990 euro. Daaruit berekent men een 95% betrouwbaarheidsinterval voor het inkomen van de luisteraars van 14337 euro tot 14463 euro. Is dit resultaat te vertrouwen? Waarom wel/niet?

13

In 1987 was het gemiddelde rendement van alle aandelen van de beurs van New York gelijk aan $\mu = -3,5\%$. (dat wil zeggen: in 1987 daalden die aandelen gemiddeld 3,5% in waarde). De standaardafwijking van alle rendementen was ongeveer gelijk aan $\sigma = 26\%$. De verdeling van het jaarlijkse rendement op aandelen is ruwweg normaal verdeeld.

- a) Welk percentage van alle aandelen is in 1987 in waarde gedaald? (dit betekent: hadden een rendement lager dan 0%)
- b) Stel dat je een portefeuille zou bezitten van vijf aandelen, aselekt gekozen uit alle aandelen op de beurs van New York. Wat zijn de verwachting en de standaardafwijking van de gemiddelde rendementen van aselekt gekozen portefeuilles van vijf aandelen? Wat is de kans dat jouw aselekt gekozen portefeuille van vijf aandelen een gemiddeld rendement lager dan 0% kende?
- c) Verklaar het verschil tussen dit antwoord en je antwoord op a)

14

Omdat bussen meer kosten dan trams en ze ook meer CO₂ uitstoten, onderzoekt De Lijn de effectiviteit van bussen in Gent. De focus ligt daarbij op buslijn 73, die net als tramlijn 21/22 het station verbindt met het Zuid. Men verwacht echter niet dat de bus sneller of trager naar het Zuid rijdt dan de tram. In de audit werd een EAS genomen van 23 ritten bij buslijn 73, die resulteerde in een gemiddelde rijtijd van 16,32 minuten (standaardafwijking=8,6). Er werd ook een EAS van 36 ritten genomen bij tramlijn 21/22. Hieruit bleek dat de gemiddelde rijtijd van de trams 22,47 minuten bedraagt (standaardafwijking= 4,2). Verschillen de lijnen significant van elkaar?

15

Landbouwonderzoekers beplanten 15 percelen met een nieuwe maïsvariëteit. De gemiddelde opbrengst bedraagt 123,8 bushels per acre. Neem aan dat deze opbrengsten in de populatie normaal verdeeld zijn, met een standaardafwijking in de populatie van 10 bushels per acre.

- a) Bepaal een 90%-betrouwbaarheidsinterval.
- b) Bepaal een 95%-betrouwbaarheidsinterval.
- c) Bepaal een 99%-betrouwbaarheidsinterval.
- d) Hoe veranderen de foutmarges terwijl het betrouwbaarheidsniveau toeneemt?

16

UNIZO voert een onderzoek uit aan de Vlaamse Kust. Het blijkt dat over de ganse kust, in het hoogseizoen, gemiddeld 267 ijsjes per dag verkocht worden. UNIZO publiceert de resultaten in een rapport dat verstuurd wordt naar alle kustburgemeesters. De burgemeester van Wenduine denkt echter dat zijn stad het beter doet, aangezien er net dat jaar veel ijssalons bijgekomen zijn. Hij neemt het onderzoeksbureau Check-It in dienst en laat hen een EAS van 46 dagen nemen uit alle dagen van het hoogseizoen. Het steekproefgemiddelde bedroeg 285 ijsjes, met een standaardafwijking van 56.

- a) Stel dat je het populatiegemiddelde niet weet. Bouw een 95%-betrouwbaarheidsinterval.
- b) Stel dat je het populatiegemiddelde niet weet. Bouw een 80%-betrouwbaarheidsinterval.
- c) Wat zullen de nulhypothese en de alternatieve hypothese zijn van de significantietoets die Check-It moet uitvoeren? Moet je een- of tweezijdig toetsen?
- d) Voer de significantietoets uit. Wat kun je hieruit besluiten?

17

Studentenclub 'De gewillige katers' is de beruchtste van de hele Overpoort. Geregeld doen de leden haar naam alle eer aan en zitten ze met barstende koppijn in de lessen. De rector roept de praeses bij zich met de boodschap dat het nu wel voldoende geweest is en dat de studentenclub haar gedrag dringend moet matigen. De praeses werpt op dat de leden van zijn studentenclub niet méér katers hebben dan andere leden van studentenclubs. De rector wil bewijzen en de praeses zet daartoe een onderzoek op poten. Hij weet dat het gemiddeld aantal keer dat alle studenten van studentenclubs met een kater in de les zitten 3,7 bedraagt en eigenlijk verwacht hij wel dat zijn leden daarboven zitten. Hij neemt een EAS van 12 leden uit zijn studentenclub en vraagt hen hoe vaak ze het vorige academiejaar met een kater in de les zaten. Het gemiddelde in de steekproef bedroeg 6 keer, met als standaardafwijking 2,3.

- a) Wat is de standaardafwijking van de steekproevenverdeling van het gemiddelde?
- b) Wat is de nulhypothese en alternatieve hypothese van de significantietoets?
- c) Voer de significantietoets uit. Wat kun je besluiten?

3.6. Inferentie voor fracties

1

Quick en Flupke lopen over straat. Quick zegt tegen Flupke dat hij een meester is in het mikken van een geldstuk in een riool. Hij houdt vol dat hij 75 % van de worpen vanop drie meter kan scoren. Flupke gelooft dat Quick overdrijft en daagt Quick uit om 37 keer te werpen. Je kunt deze worpen als een EAS beschouwen. Quick scoort 27 van de 37 worpen.

- a) Construeer een 90%-betrouwbaarheidsinterval.
- b) Spreekt Quick de waarheid? Met andere woorden, verschilt het steekproefresultaat significant van de populatiefractie?

2*

Een grote frisdrankproducent heeft onlangs het recept van hun best verkopende frisdrank lichtjes aangepast. Een marketingbureau krijgt te opdracht om na te gaan of het nieuwe recept beter aansluit bij de consumenten. Het bureau deelt een EAS van 320 blikjes met het oude recept uit en stelt vast dat 83% hiervan positief worden beoordeeld door de consument. Van de blikjes met het nieuwe recept delen ze een EAS van 673 exemplaren uit, 532 hiervan worden positief beoordeeld.

- a) Is er een significant verschil tussen het oude en het nieuwe recept in de proportie mensen dat de frisdrank positief beoordeelt?
- b) Stel een 99% betrouwbaarheidsinterval van het verschil op en interpreteer.

3

In Texas, Amerika, lopen nog steeds veel goudzoekers rond. Vooral het stadje 'Old Gold Hulch' is erg in trek, omdat er gezegd wordt dat daar nog steeds veel goud gevonden wordt. In Amerika gaat steeds meer geld naar onderzoek en de gouverneur van Texas heeft besloten om eens te bekijken of er het voorbije jaar in 'Old Gold Hulch' inderdaad meer goud gevonden werd dan in de rest van zijn staat. Van alle actieve *golddiggers* in zijn staat heeft 12% minstens een stukje van 1 cm² goud gevonden. Uit een EAS van 12 *golddiggers* die het voorbije jaar aan het werk waren in 'Old Gold Hulch', bleek dat 4 van hen het voorbije jaar minstens een stukje van 1 cm² goud gevonden had in het stadje.

- a) Stel een 90%-betrouwbaarheidsinterval op.
- b) Stel een 96%-betrouwbaarheidsinterval op.
- c) Voer de significantietoets uit. Wat kun je hieruit besluiten?

4

De Gentse schepen van cultuur wil een plan opstellen om de Gentse jeugd meer naar de stedelijke musea te krijgen. Daarom geeft hij de opdracht aan de Faculteit Politieke en Sociale wetenschappen om een profiel op te stellen van het publiek dat nu aangetrokken wordt door de musea. Medewerkers van de faculteit, die er vanuit gaan dat jongeren minder vaak de musea bezoeken, besluiten om een EAS te nemen van de jongeren uit Gent (jonger dan 30 jaar) en een EAS van volwassenen (ouder dan 30 jaar). Van de 376 geselecteerde jongeren geven er 65 aan dat ze het SMAK al eens bezocht hebben. Van de 687 geselecteerde volwassenen geven er 156 aan het SMAK al bezocht te hebben.

- a) Stel een 70%-betrouwbaarheidsinterval op.
- b) Stel een 99,5%-betrouwbaarheidsinterval op.
- c) Voer de significantietoets uit. Wat kun je hieruit besluiten?

5

Er wordt vaak gezegd van Ronaldinho dat het één van de beste voetballers van zijn tijd is. De FIFA verwacht inderdaad dat Ronaldinho beter is dan gemiddeld en wenst dit eens onder de loep te nemen en neemt een EAS uit alle trappen op het doel die Ronaldinho nam in wereldbekerwedstrijden. Men weet dat, bij alle voetballers op wereldbekerwedstrijden, 6 % van de trappen op het doel succesvol zijn. Bij de EAS van 31 gekozen doeltrappen van Ronaldinho, waren er 3 succesvol.

- a) Stel een 95%-betrouwbaarheidsinterval op.
- b) Stel een 60%-betrouwbaarheidsinterval op.
- c) Voer de significantietoets uit. Wat kun je hieruit besluiten?

6

Vinkenzetters-club 'Bij de vinken' het net zijn jaarlijkse tornooi achter de rug. De winnaar, Patrick Vogels, staat te pronken met zijn beker aan de toeg. Aan iedereen die het wil horen vertelt hij zijn verhaal: "Het is geen wonder dat ik win! Alle vogels die ik heb maken significant meer lawaai dan de totale populatie vogels!" Erick, de tweede, is het daar echter niet mee eens en de mannen besluiten de proef op de som te nemen. Patrick neemt een EAS van 12 van zijn vogels en gaat na of ze in de eerste vijf minuten nadat ze uit hun kooi genomen zijn, het gewenste 'Suskwiet' produceren. Dat was het geval voor 3 van de geselecteerde vogels. Erick weet dat in de totale populatie vinken, de fractie vogels die inderdaad het gewenste geluid maken in de eerste vijf minuten dat ze hun kooi verlaten hebben, 15% bedraagt.

- a) Stel een 95%-betrouwbaarheidsinterval op
- b) Hoeveel bedraagt de standaardfout van de steekproevenverdeling die je gebruikt voor het opstellen van dit betrouwbaarheidsinterval? En hoeveel bedraagt de standaardfout van de steekproevenverdeling die je zou gebruiken voor het uitvoeren van een significantietoets?
- c) Klopt het dat de steekproeffractie significant verschilt van de populatie van alle vogels? En indien zo, op wel niveau is dit dan significant?

7*

Trey Parker en Matt Stone, bedenkers en makers van de populaire tekenfilmreeks 'South park', vragen zich af welk doelpubliek ze eigenlijk bereiken met hun serie. Ze nemen het internationaal bekende onderzoeksbureau Check-It onder de arm. Het bureau zet zich aan het werk en een aantal maanden later ligt het uiteindelijke dossier op het bureau van Parker en Stone. Op pagina 7 kunnen ze lezen dat het bureau een EAS nam van 120 jongens en 314 meisjes. Van de geselecteerde jongens gaven er 65 aan al meer dan 3 afleveringen van de serie gezien te hebben, terwijl dat aantal bij de geselecteerde meisjes 103 bedroeg.

- a) Stel een 73%-betrouwbaarheidsinterval op.
- b) Voer de significantietoets uit. Wat kun je hieruit besluiten?

8

Yoghurtfabrikant 'Yogh-up' heeft een mannetje in een supermarkt geplaatst. De winkelspion heeft maar één opdracht: kijken of het inderdaad zo is dat blonde mensen inderdaad meer yoghurt van hun merk eten dan mensen met een andere haarkleur. Om die vraag op te lossen heeft de spion een EAS genomen van 36 blonde mensen die de winkel binnenkomen en een EAS van 48 bij de mensen met een andere haarkleur die de winkel binnenkomen. Van de 36 blonde mensen in de steekproef, kochten er 12 inderdaad minstens één potje van 'Yogh-up'. Van de 48 mensen met een andere haarkleur in de steekproef, namen er 15 minstens één potje.

- a) Stel een 80%-, 96%- en 99%-BI op.
- b) Voer de significantietoets uit. Is het inderdaad zo dat significant meer blonde mensen minstens één potje 'Yogh-up' kochten?

3.7. Inferentie voor kruistabellen

1

Een jonge onderzoeker in de tandheelkunde zit op een avond naar tv te kijken. Tijdens de commerciële boodschappen passeert een reclame die het merk 'Snowy White' aanprijst. De onderzoeker gelooft weinig van de reclamespot en zet een experiment op waarbij hij een groep mensen 20 jaar lang het product laat gebruiken en een andere groep mensen – een testgroep – een ander, niet gedefinieerd product laat gebruiken. 20 jaar later, wanneer de onderzoeker al professor is, komen de resultaten binnen. De professor bekijkt onmiddellijk de tabel die het aantal gaatjes dat de respondent opgelopen heeft in de voorbije 20 jaar in verband brengt met het product dat hij gebruikte:

Aantal gaatjes	Merk tandenborstel	
	Snowy White	Zwart product
0-2	23	13
3-5	35	17
6-8	10	28

- Hoeveel procent van de mensen die tussen de 3 en 5 gaatjes hadden, gebruikten het zwarte product? Is dit een rij- of kolompercentage?
- Bereken chi-kwadraat en cramer's V.
- Is het verband tussen beide variabelen significant? Op welk significantieniveau?

2*

De Vlaams Minister van Onderwijs en de Federale Minister van Volksgezondheid bundelen de krachten. Ze zetten, met behulp van de Gentse universiteit, een grootschalig onderzoek op naar het gezondheidsgedrag van Vlaamse jongeren en nemen daartoe een representatieve steekproef. Hieronder vind je de verdeling naar onderwijstype terug.

Roken	ASO	TSO	BSO	KSO
Nee	451	321	280	53
Ja	251	110	121	40

- Bereken chi-kwadraat voor deze kruistabel.
- Hoeveel vrijheidsgraden heeft de steekproevenverdeling voor deze kruistabel?
- Is er een significant verband tussen beide variabelen?

3.8. Overzichtsoefeningen

1

WWF voert al jaren campagne tegen het opsluiten van dieren in dierenparken. Eén van hun argumenten is dat dieren zich in gevangenschap minder voortplanten dan dieren die in de wildernis leven. De natuurliefhebbers baseren zich hiervoor onder andere op hun onderzoek, waarbij ze een EAS van krokodillen in gevangenschap vergeleken met een EAS van krokodillen in de wildernis. Het gemiddeld aantal eieren dat krokodillen in hun levensloop leggen bedroeg bij de 15 dieren in gevangenschap 57, met een standaardafwijking van 12,76; bij de 56 dieren in de wildernis bedroeg dit gemiddelde 65, met een standaardafwijking van 23,83.

- a) Bouw een 70%-betrouwbaarheidsinterval.
- b) Bouw een 98%-betrouwbaarheidsinterval.
- c) Voer de significantietoets uit. Wat kun je hieruit besluiten?

2*

Een enkelvoudige aselechte steekproef van 75 kleinhandelaars (winkels) toont een gemiddelde verkoop van 52 stuks van een bepaald apparaat, met een standaardafwijking van 13 stuks. Gedurende diezelfde maand van het vorige jaar gaf een steekproef van 53 winkels een gemiddelde verkoop van slechts 49 stuks aan, met een standaardafwijking van 11 stuks. Deze stijging van 49 naar 52 stuks maakt een marketingmanager dolblij en hij beweert prompt overal dat de verkoop met "x" procent steeg. Ga na of de manager hier niet de hard van stapel loopt. Kan hij er zeker van zijn dat er een stijging in verkoop plaats vond?

3*

Studio 100 is geïnteresseerd in de doelgroep die ze bereiken met hun pretpark 'Plopsaland' in De Panne. Ze nemen daartoe een kleine telefoonenquête af bij een aantal mensen. Eén van de meest markante bevindingen was de leeftijd van de Plopsaland bezoekers. De gegevens vind je hieronder terug. Is er een significant verband tussen beide variabelen?

Plopsaland bezocht?	Man	Vrouw
Ja	297	325
Nee	795	762

4

Het CLB voert een onderzoek uit in het Vlaamse Secundaire onderwijs. Bedoeling is om na te gaan of pesters en niet-pesters significant verschillen in het aantal beste vrienden waarover ze beschikken. Men verwacht dat pesters meer beste vrienden hebben dan niet-pesters. Uit een EAS van 21 pesters blijkt dat zij een gemiddeld aantal beste vrienden van 7 hebben, met een standaardafwijking van 4,6. Uit de EAS van de 47 niet-pesters blijkt hun gemiddeld aantal beste vrienden op 4 te liggen, met een standaardafwijking van 2,8.

- a) Bouw een 90%-betrouwbaarheidsinterval.
- b) Bouw een 95%-betrouwbaarheidsinterval.
- c) Voer de significantietoets uit. Wat kun je hieruit besluiten?

5

Een bekende telefoongigant weet dat mensen gemiddeld 6 keer per week bellen. Flair, het bekende vrouwenblad, vraagt zich af of vrouwen verschillen van dit gemiddelde, al stellen ze niet echt een duidelijke richting van het verschil voorop. Het blad neemt een EAS van 25 vrouwen en vraagt hen hoeveel keer ze per week iemand opbellen. Het gemiddelde van deze steekproef bedroeg 9, met een standaardafwijking van 4,5.

- a) Stel dat je het populatiegemiddelde niet weet. Bouw een 90%-betrouwbaarheidsinterval.
- b) Stel dat je het populatiegemiddelde niet weet. Bouw een 99 %-betrouwbaarheidsinterval.
- c) Voer de significantietoets uit. Wat kun je hieruit besluiten?

6

Op een universiteit wordt een tevredenheidsonderzoek gehouden. De gemiddelde respons van 1406 geselecteerde eerstejaarsstudenten op de vraag of ze "Zich welkom voelen op de universiteit" was 3,9 (op 5), met een standaardafwijking van 0,98. Neem aan dat deze 1406 studenten een enkelvoudige aselechte steekproef vormen uit alle eerstejaarsstudenten. Bereken een 99%-betrouwbaarheidsinterval.

7

Een touragentschap wil te weten komen welke bestemmingen ze moet uitkiezen voor de volgende wintermaanden. Omdat in deze maanden vooral gepensioneerden op reis gaan, is men vooral geïnteresseerd in deze bevolkingsgroep. Men neemt het, ondertussen gerenommeerde, onderzoeksbureau Check-It onder de arm, die prompt een onderzoek uitvoert bij de Vlaamse bevolking. Enkele van de gegevens vind je in de kruistabel hieronder. Is er een significant verband tussen beide variabelen?

Ooit op Tenerife geweest?	Niet-gepensioneerden	Gepensioneerden
Ja	498	209
Nee	804	384

8

Rudy is een rasechte doe-het-zelver, volop bezig met de afwerking van zijn woning. In een plaatselijke houthandelaar heeft hij net een lot hout gekocht dat hij gaat gebruiken voor de aanleg van de parketvloer in zijn living. De verkoper vertelt hem echter dat, van alle planken, 6.33 % onbruikbaar is wegens houtrot of te grote jaarringen. Voor zijn woning kocht Rudy een lot van 75 planken, die je als een EAS uit alle planken kunt bekijken. Na controle blijkt dat hij 5 onbruikbare planken heeft.

- Stel een 90%- en een 99%-BI betrouwbaarheidsinterval op.
- Heeft Rudy significant meer planken die onbruikbaar zijn?

9

In het honkbalseizoen van 1991 speelden de "Minnesota Twins" 81 thuiswedstrijden en 81 uitwedstrijden. Ze wonnen 51 thuiswedstrijden en 44 uitwedstrijden. Beschouw deze als enkelvoudige aselechte steekproeven uit grote populaties van uit- en thuiswedstrijden van de Twins. Hoeveel voordeel biedt een thuiswedstrijd? Bereken de foutenmarge voor een 90%-betrouwbaarheidsinterval.

10

De kringwerking van de katholieke kerk in Vlaanderen doet op regelmatige basis een onderzoek bij een representatieve steekproef voor Vlaanderen. Een van hun vragen peilt naar het belang van het katholieke geloof in het dagelijkse leven van de Vlaming. De respondenten kunnen antwoorden op een schaal van 1 (helemaal niet belangrijk) tot 5 (heel belangrijk). De EAS van 153 West-Vlamingen scoorde als volgt op deze vraag $N(3,5, 0,4)$. Voor de EAS van 207 Oost-Vlamingen waren de scores als volgt verdeeld: $N(2,8, 1,3)$.

- a) Is er een significant verschil tussen het resultaat van West- en Oost-Vlaanderen?
- b) Geef het 90% betrouwbaarheidsinterval.

11*

Bioscoopbezoek is een vorm van culturele participatie waarop studenten traditioneel heel hoog scoren. De afgelopen jaren is de prijs van een filmticketje echter serieus gestegen, terwijl het (illegaal) downloaden van films steeds toegankelijker is geworden. Een medewerker van een grote cinemaketen heeft daarom een vergelijkend onderzoek gedaan in Gent. Vorig jaar interviewde hij een EAS van 434 studenten, 321 daarvan waren de afgelopen 6 maanden minstens 1 keer naar de cinema geweest. Dit jaar interviewde hij een EAS van 376 studenten, waarvan er 216 minstens 1 keer naar een film zijn gaan kijken het afgelopen half jaar.

- a) Is er een verschil in het aantal studenten dat naar de bioscoop is geweest tussen dit jaar en vorig jaar?
- b) Bouw een 95%-betrouwbaarheidsinterval.

12*

Gegeven is een 95% betrouwbaarheidsinterval rond een steekproeffractie $[0,2 - 0,6]$. Wat is de omvang van de steekproef waaruit deze fractie werd getrokken?

13

Uit een steekproef met omvang 816 bekomt men het 90%-betrouwbaarheidsinterval $[12,5; 17,8]$ voor μ . Bepaal nu het 95% betrouwbaarheidsinterval voor μ met behulp van dezelfde steekproef.

14

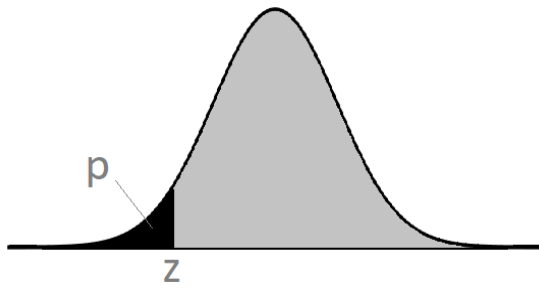
Op sociale media duiken meer en meer video's op van de hype "Mannequin Challenge"; 60% van de jongeren zou zich al hebben laten uitdagen. Professor Dewelf is benieuwd naar hoeveel studenten communicatiewetenschappen al deelgenomen hebben en trekt een EAS van 230 studenten. 150 studenten hebben al een filmpje gepost. Bereken een 83%-betrouwbaarheidsinterval om een uitspraak te kunnen doen over de volledige groep.

15*

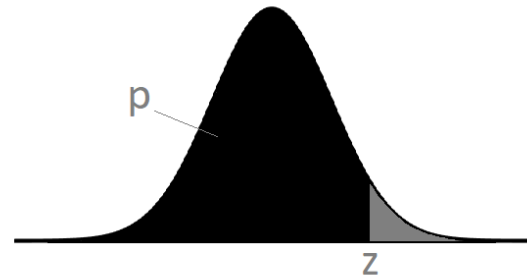
Melissa doet onderzoek naar de gemiddelde leeftijd waarop mannen hun eerste kind krijgen, aangezien er momenteel enkel populatiegegevens ter beschikking zijn omtrent de leeftijd waarop vrouwen voor het eerst moeder worden. Op basis van een steekproef van 623 mannen, weet Melissa met 80% betrouwbaarheid dat de populatieverwachting tussen 28,5 en 32,7 ligt. Bepaal het 96% betrouwbaarheidsinterval voor de populatieverwachting op basis van deze steekproef.

DEEL D: BIJLAGEN

Bijlage 1: z-tabel

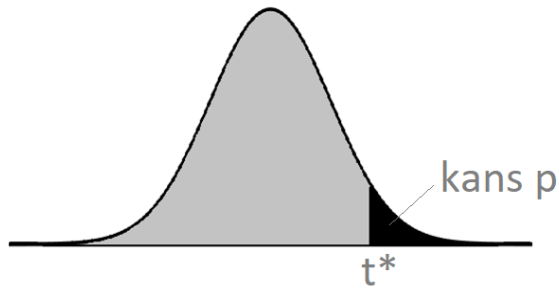


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641



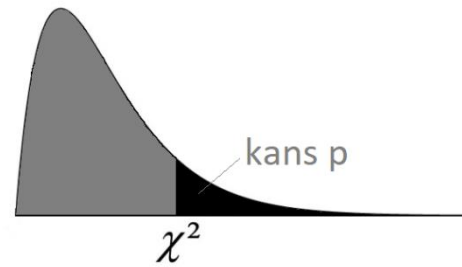
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Bijlage 2: t-tabel



Df	Staartkans p (Éenzijdig)											
	0.25	0.2	0.15	0.1	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.706	15.895	31.821	63.657	127.321	318.309	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.089	22.327	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.215	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.50%	99.80%	99.90%
	Betrouwbaarheidsniveau											

Bijlage 3: Chi-kwadraattabel



Df	Staartkans p (Tweezijdig)											
	0.25	0.2	0.15	0.1	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47	20.00
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.52	22.11
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46	24.10
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32	26.02
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12	27.87
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88	29.67
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59	31.42
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26	33.14
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91	34.82
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53	36.48
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12	38.11
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70	39.72
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25	41.31
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79	42.88
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31	44.43
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82	45.97
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31	47.50
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80	49.01
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27	50.51
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73	52.00
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18	53.48
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62	54.95
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83	54.05	56.41
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22	55.48	57.86
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59	56.89	59.30
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97	58.30	60.73
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70	62.16
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40	76.09
50	56.33	58.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66	89.56
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61	102.69
80	88.13	90.41	93.11	96.58	101.88	106.63	108.07	112.33	116.32	120.10	124.84	128.26
100	109.14	111.67	114.66	118.50	124.34	129.56	131.14	135.81	140.17	144.29	149.45	153.17

Bijlage 4: Formularium

Residuenen (χ^2) = $f(o) - f(e)$	Gestandaardiseerde residuen $(\chi^2) = \frac{f(o)-f(e)}{\sqrt{f(e)}}$
$\chi^2 = \sum \frac{(f(o)_{ij} - f(e)_{ij})^2}{f(e)_{ij}}$	$\phi = \sqrt{\frac{\chi^2}{n}}$
$V = \sqrt{\frac{\chi^2}{n(k-1)}} \quad (k = \min(r, k))$	$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n [y_i - (a + b_{x_i})]^2$
$\text{cov}(X, Y) = S_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$	$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$
$r_{xy} = \frac{1}{n-1} \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1} \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{n-1}}}$	$r_{xy}^2 = \frac{S_y^2}{S_y^2}$
$r_{xy} = \frac{S_{xy}}{S_x S_y}$	$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$
$a = \bar{y} - b\bar{x}$	$b = \frac{S_{xy}}{S_x}$
$b = r_{xy} \frac{S_y}{S_x}$	

$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$
$\mu_{\bar{x}} = \mu$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
$\mu_{\bar{x}} = \mu$	$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$
$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$	$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
$\mu_{\hat{p}} = p$	$SE_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}}$
$\mu_{\hat{p}} = p$	$SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$	$SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$
$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$	$SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

Bijlage 5: Oplossingen C-oefeningen

1. Univariate beschrijvende statistiek

1.1. Basisconcepten

1*

- a) Een Vlaams gezin is de eenheid, het aantal kinderen is de variabele
- b) 45 gezinnen hebben 2 kinderen; de proportie is 0,43
- c) 15,24%

2*

- a) Alle studenten aan de UGent
- b) Een student aan de UGent
- c) 'Faculteit' en 'aantal glazen bier per week'
- d) Faculteit: 'nominaal'; aantal glazen bier per week: ratio

3

- a) De populatie zijn hier alle regio's uit alle landen van de Europese Unie
- b) Eén regio uit een land uit de Europese Unie
- c) We hebben 2 variabelen:
 - Het aantal liter regen per kubieke meter → ratio
 - De rating van de weerkundige uit de regio → ordinaal

4

- a) De populatie zijn hier de 17 leerlingen uit de klas van de leraar
- b) Eén leerling uit de klas van de leraar
- c) We hebben 1 variabele: de score op de test → ratio

5

- a) De populatie is hier het Werchter publiek: alle mensen die ooit naar Rock Werchter gegaan zijn.
- b) Eén Rock-Werchter-bezoeker
- c) We hebben 6 variabelen:
 - a. leeftijd → ratio/interval

- b. inkomen → ratio
- c. aantal keren gewassen → ratio
- d. geslacht → nominaal
- e. aantal concerten in Marquee → ratio
- f. aantal concerten op de Main Stage → ratio
- g. OF i.p.v. 2 vorige puntjes: plaats waar men meeste concerten meepikte → nominaal

6

-
- a) Alle Vlaamse gemeenten
 - b) Eén Vlaamse gemeente
 - c) Het aantal woningen in verkrotte toestand
 - d) ratio

OF:

- a) Woningen
- b) Eén woning
- c) De Vlaamse gemeente en al dan niet verkrot
- d) Beiden nominaal

7

-
- a) Alle dagen tussen zijn twintigste en dertigste levensjaar
 - b) Eén dag tussen zijn twintigste en dertigste levensjaar
 - c) Het aantal meter dat hij gezwommen heeft op een dag → ratio
- Of: de zon schijnt of niet → nominaal

8*

-
- a) Ratio discreet
 - b) Nominaal
 - c) Nominaal
 - d) ordinaal
 - e) ratio, discreet
 - f) ratio, continu

9

- a) agressie
- b) aantal aanvallen ooit (op menselijk vlees)
- c) een pitbull
- d) alle pitbulls die alle praktijken van alle veeartsen (in dit groepje) betreden

10

- a) 'al dan niet positief testen' en 'welke eindejaarsperiode (deze of vorige)'
- b) beiden nominaal
- c) een Vlaamse bestuurder
- d) alle Vlaamse bestuurders (enkel als representatieve steekproef, anders enkel de gecontroleerde bestuurders)
- e) steekproef

OF:

- a) 'Proportie positief testende bestuurders'
- b) ratio
- c) een eindejaarsperiode
- d) beide eindejaarsperiodes
- e) steekproef

11

- a) verkeersagressie
- b) gemeten als aantal pv's voor verkeersagressie
- c) een Vlaams arrondissement
- d) alle Vlaamse arrondissementen

12

- a) Al dan niet in armoede leven
- b) Nominaal
- c) een Belg
- d) alle Belgen
- e) Populatie

13

-
- a) Merk A betrouwbaarder want
- Merk A = $2942/13\ 376 = 0,22$
 - Merk B = $192/480 = 0,40$
- b) Mensen die het afgelopen jaar een auto gekocht hebben
- Het merk van de auto → nominaal
 - Al dan niet de reparatiedienst ingeschakeld hebben → nominaal

14*

-
- a) De tabel is een frequentietabel
- b) Nominaal

1.2. Frequentieverdelingen

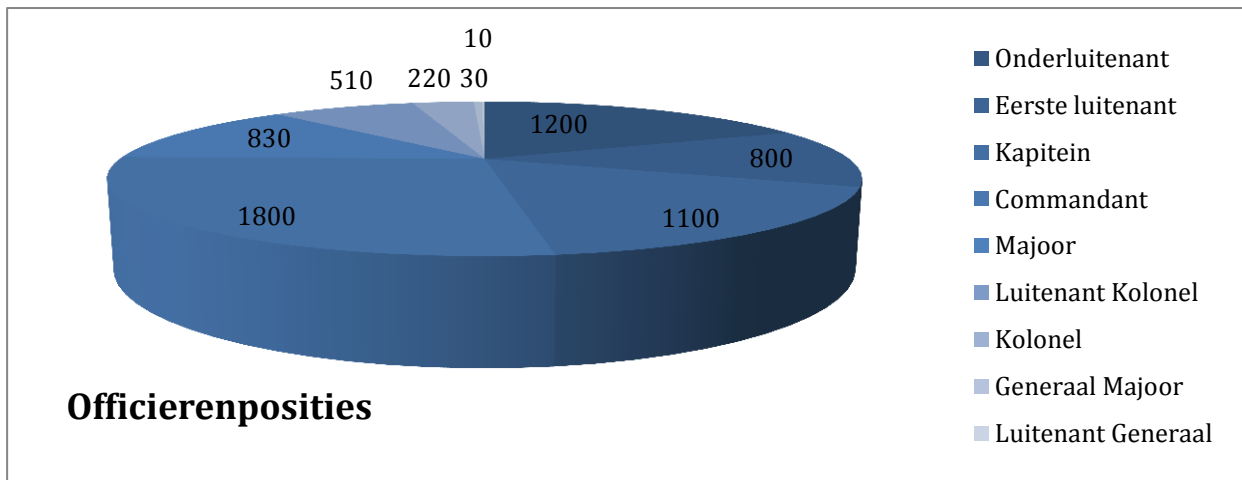
1*

-
- a) $VB = 64 - 20 = 44$
- b) $KB = (44+1)/5 = 9 \rightarrow$ klassen: $[20,29[$; $[29,38[$; $[38,47[$; $[47,56[$; $[56,65[$
- c) 24 ; 33 ; 42 ; 51 ; 60
- d)

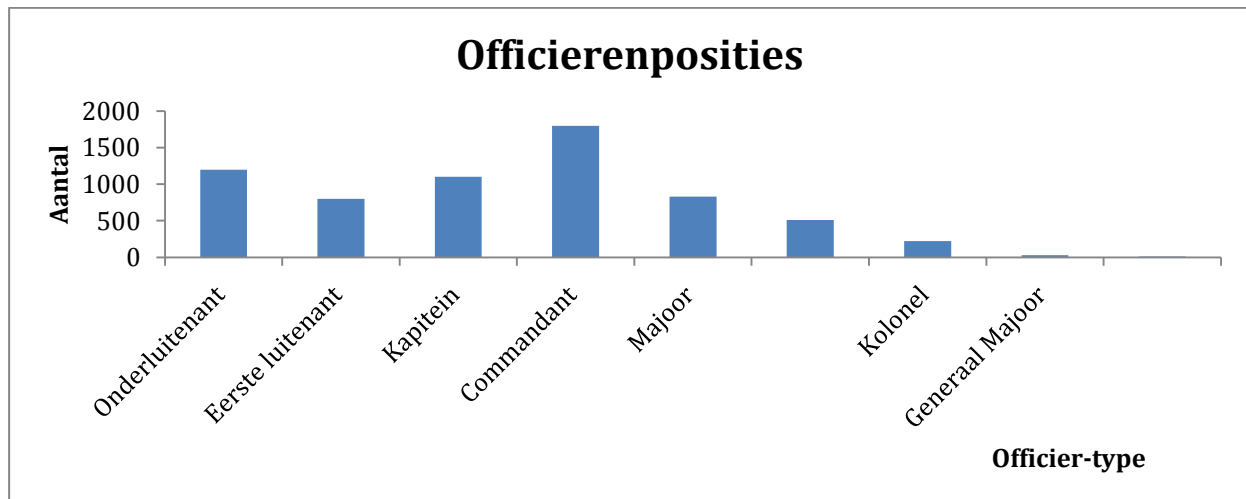
	f_i
$[20,29[$	10
$[29,38[$	17
$[38,47[$	15
$[47,56[$	7
$[56,65[$	1

2

Voorbeeld van weergave in een taartdiagram:



Voorbeeld van weergave in een staafdiagram:



3*

$$n = 14 (=4+2+3+4+1) \rightarrow f_3 = 3 \rightarrow p_3 = f_3/n = 3/14$$

1.3. Maten van positie

1

- een ambtenaar
- $n = 13+40+78+\dots+1 = 900$
- + d)

Mediaan

Me = waarde((n + 1)/2)-de waarneming = waarde 450,50-de waarneming = 48-59 maanden

Kwartielen

Eerste kwartiel: = waarde $((n + 1)/4)$ -de waarneming = waarde van de 225,25-ste waarneming

$Q_1 = P_{25} = 36$ -47 maanden

$Q_2 = Me = 48$ -59 maanden

Derde kwartiel: =waarde $(3(n + 1)/4)$ -de waarneming =waarde van de 675,75-ste waarneming

$Q_3 = P_{75} = 72$ -83 maanden

Rekenkundig gemiddelde

$X_{gem} = (13 \cdot 5,5 + 40 \cdot 17,5 + \dots + 1 \cdot 173,5) / 900 = 59,95$

80ste percentiel

Waarde van de $80 \cdot (n+1)/100$ -ste waarneming = waarde van 720,8-ste waarneming

$p_{80} = 72$ -83 maanden

2*

Frequentietabel:

IQ-score	Freq	Cum freq
74	1	1
107	2	3
116	1	4
120	1	5
126	1	6
127	1	7

Meest relevante maat is het gemiddelde, want meetniveau = ratio.

Gemiddelde = $(1 \cdot 74 + 2 \cdot 107 + \dots + 1 \cdot 127) / 7 = 111$

(Modus = 107 ; Mediaan = waarde $(7+1)/2^{\text{de}}$ waarneming = 116)

3*

a) Ordinaal

b) Modus = Noch tevreden, noch ontevreden; Me = Noch tevreden, noch ontevreden;
 Q_1 = Ontevreden; Q_3 = Tevreden

Antwoord	f_i	Cum. f_i	p_i	Cum. p_i
Zeer ontevreden	33	33	0.1	0.1
Ontevreden	84	117	0.2545	0.3545
Noch tevreden, noch ontevreden	102	219	0.3091	0.6636
Tevreden	63	282	0.1909	0.8545
Zeer tevreden	48	330	0.1455	1

4*

Mo = 116-120

Me = 116-120

Xgem = 119,93

 Q_1 = 116-120 Q_3 = 121-125

x_i	m_i	f_i	Cum. f_i	p_i	Cum. p_i
100-110	105	534	534	0.0854	0.0854
111-115	113	712	1246	0.1138	0.1992
116-120	118	2021	3267	0.3232	0.5224
121-125	123	1947	5214	0.3113	0.8337
126-130	128	553	5767	0.0884	0.9221
131-135	133	487	6254	0.0779	1

5

a) Een datamatrix

b)

x	f	cum f	p	cum p
1	2	2	0.0645	0.0645
2	4	6	0.1290	0.1935
3	3	9	0.0968	0.2903
4	4	13	0.1290	0.4194
5	6	19	0.1935	0.6129
6	7	26	0.2258	0.8387
7	1	27	0.0323	0.8710
8	2	29	0.0645	0.9355
9	2	31	0.0645	1.000

- c) Modus = 6; Mediaan = 5; $Q_1 = 3$; $Q_3 = 6$; $X_{gem} = 4,77$

6

- a) Fout. Het gaat om de WAARDE van de 1091e eenheid en deze waarde is 'noch ontevreden, noch tevreden' (rangschik de waarden van laag naar hoog).
- b) Fout. De onderzoekspopulatie = de volledige populatie werknemers aangezien hier een enkelvoudige aselechte steekproef uit de populatie 'werknemers van de multinational' werd gebruikt.
- c) Dit klopt. De mediaan bevindt zich inderdaad in deze categorie.
- d) Dit klopt, want het gaat om een ordinale variabele.

7*

- a) $VB=81$
- b) $KB = 16,4$

x_i	f_i	cum f_i	p_i	cum p_i
4-19	11	11	0.306	0.306
20-35	7	18	0.194	0.500
36-51	8	26	0.222	0.722
52-67	4	30	0.111	0.833
68-85	6	36	0.167	1.000

- c)
- d) $Q_1 = [4; 19]$; $Q_3 = [52; 67]$
- e) $X_{gem} = 37,72$

1.4. Maten van spreiding

1*

- a) Alle maten van spreiding zijn onmogelijk: ordinaal meetniveau

2*

x_i	f_i	Cum. f_i	p_i	Cum. p_i
16	11	11	0.0182	0.0182
18	106	117	0.1758	0.1940
19	154	271	0.2554	0.4494
20	120	391	0.1990	0.6484
21	99	490	0.1642	0.8126
22	45	535	0.0746	0.8872
23	45	580	0.0746	0.9619
25	23	603	0.0381	1

- a)
- b) Ratio
- c) $IKA = 2; s^2 = 3.34; s = 1.83$
- d) Ja, de eenheden met waarden 16 en 25 (34 outliers)

3

- a) Een datamatrix

x_i	f_i	Cum. F_i	p_i	Cum. P_i
1	2	2	0,0645	0,0645
2	4	6	0,1290	0,1935
3	3	9	0,0968	0,2903
4	4	13	0,1290	0,4194
5	6	19	0,1935	0,6129
6	7	26	0,2258	0,8387
7	1	27	0,0323	0,8710
8	2	29	0,0645	0,9355
9	2	31	0,0645	1

- b)
- c) Interkwartielafstand = 3; $s^2 = 4,78; s = 2,19$

4

- a) Frequentietabellen

b)

- a. 'Gezinnen met huismoeder':

$x_{gem} = 1,9237 = 1,92$; Mediaan: Waarde van de 59,5-ste waarneming $\rightarrow 2$

- b. 'Gezinnen met huisvader':

$x_{gem} = 1,50$; Mediaan: Waarde van de 20,5-ste waarneming $\rightarrow 1$

- c. 'Gezinnen met werkende ouders':

$x_{gem} = 1,4626 = 1,46$; Mediaan: Waarde van de 74-ste waarneming $\rightarrow 1$

- c) 'Gezinnen met huismoeder': $s = 1,31$
'Gezinnen met huisvader': $s = 1,18$
'Gezinnen met werkende ouders': $s = 1,24$

Besluit:

Groep 1 (huismoeder): grootste gemiddelde en spreiding

Groep 2 (huisvader): kleinste spreiding

Groep 3 (werkende ouders): laagste gemiddelde

5

- a) 6
- b) Rechts asymmetrisch
- c) In totaal zijn er meer uitschieters naar boven (in PK) bij hoger aantal cilinders (vooral bij 8) dan bij kleiner aantal cilinders (vooral 4) naar beneden. Dit trekt het totale gemiddelde meer naar boven dan de mediaan, waardoor die kleiner wordt dan het gemiddelde.

6

Punten op 20	Cumulatieve frequentie	Absolute frequenties	Proporties	Cumulatieve proporties
0	0	0	0,000	0,000
1	0	0	0,000	0,000
2	1	1	0,029	0,029
3	2	1	0,029	0,059
4	2	0	0,000	0,059
5	3	1	0,029	0,088
6	3	0	0,000	0,088
7	3	0	0,000	0,088
8	5	2	0,059	0,147
9	8	3	0,088	0,235
10	13	5	0,147	0,382
11	16	3	0,088	0,471
12	18	2	0,059	0,529
13	24	6	0,176	0,706
14	29	5	0,147	0,853
15	32	3	0,088	0,941
16	33	1	0,029	0,971
17	34	1	0,029	1,000
18	34	0	0,000	1,000
19	34	0	0,000	1,000
20	34	0	0,000	1,000
TOTAAL		34		

a)

b) Gemiddelde = 11,35; mediaan = 12. De mediaan is beter, aangezien het gemiddelde gevoeliger is voor uitschieters (2 en 3/20)

c) $sx^2 = 11,87$; $sx = 3,45$; $IKA = 4$

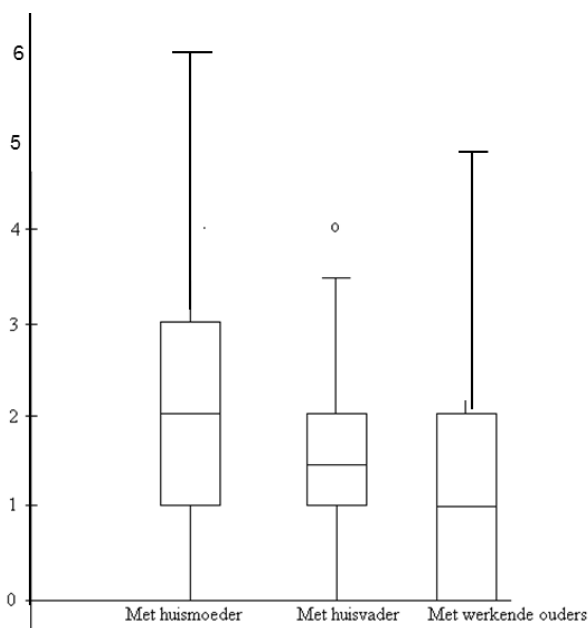
7*

- a) correct
- b) fout
- c) fout
- d) correct

1.5. Vorm van een verdeling

1

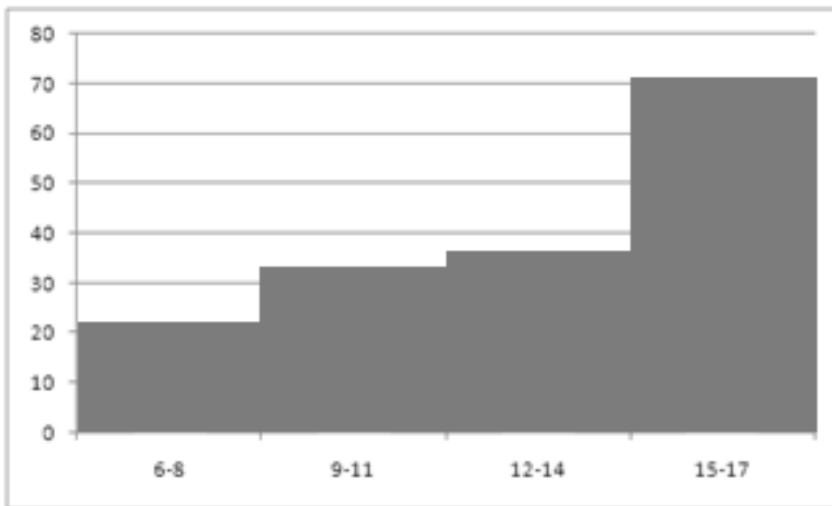
Weergegeven in boxplots:

**2**

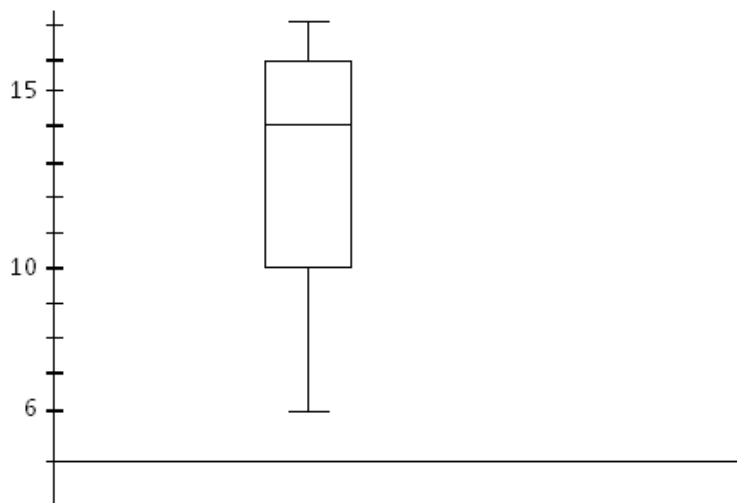
- a) Het rekenkundig gemiddelde (1,35) is groter dan de mediaan (1). Het is dus een rechts-asymmetrische verdeling.
- b) rekenkundig gemiddelde en mediaan

3

- a) Zie histogram:



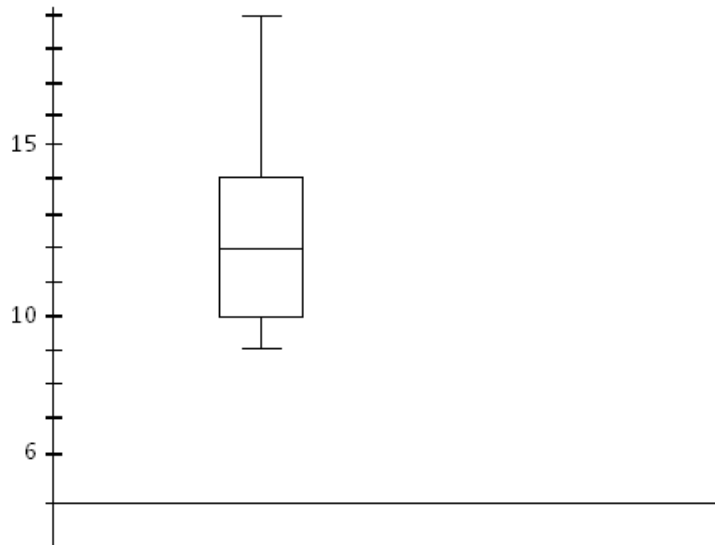
b)

c) $X_{\text{gem}} = 12,94$; $Me = 14 \rightarrow$ Links-asymmetrisch

4*

X_i	f_i
9	2
10	3
12	2
13	1
14	2
15	1
16	0
19	1

a)



b)

c) $X_{\text{gem}} = 12,25$; $Me = 12 \rightarrow$ Rechts-asymmetrisch**5***

a) Ratio

b) De verdeling is linksscheef. ($X_{\text{gem}} = 138,70$; $Me = 139$)

1.6. Transformaties van een variabele
--

1.7. Dichtheidskrommen

1*

a) $z = (8-14)/4,5 = -1,33$

$p(z < -1,33) = 0,0918$

9,18 % van alle baobabs

b) $z = (18-14)/4,5 = 0,89$

 $p(z > 0,89)$: 2 mogelijkheden:

$p(z > 0,89) = 1 - p(z < 0,89)$

$p(z < 0,89) = 0,8133$

$p(z > 0,89) = 1 - 0,8133 = 0,1867$

$p(z > 0,89) = p(z < -0,89)$

$p(z < -0,89) = 0,1867$

dus proportie van alle baobabs = 0,19 of 18,67%

c) $p = 0,07$ groter $\rightarrow p = 0,93$ kleiner $p = 0,93 \rightarrow$ niet in tabel

$p = 0,9292 \rightarrow z = 1,47$

$p = 0,9306 \rightarrow z = 1,48$

$$\rightarrow p = 0,93 \quad z = (1,47 + 1,48) / 2 = 1,475$$

$$x = (1,475 * 4,5) + 14 = 20,64$$

dus minstens 20,64 meter

$$d) \quad z = (10-14)/4,5 = -0,89$$

$$z = (16-14)/4,5 = 0,44$$

$$p(z > -0,89 \text{ en } z < 0,44) = p(z < 0,44) - p(z < -0,89)$$

$$p(z < 0,44) = 0,6700$$

$$p(z < -0,89) = 0,1867$$

$$e) \quad p(z < 0,25) - p(z < -1,63) = 0,6700 - 0,1867 = 0,4833$$

$$p = 0,25 \rightarrow z = -0,675$$

$$x = (-0,675 * 4,5) + 14 = 10,96$$

Dus: maximum 10,96 m

2

a) 4,75%

b) $p(z > 3,33) = 0,0004$ of 0,04%

c) minstens 1,89m

d) 0,9050 of 90,5%

e) maximum 1,01m

3*

a) 6,68 %

b) 51,78

c) 0,84

d) 27,43 %

e) 57,38 minuten

4

a) 3,14 %

b) 4,36 %

c) 32,00 doeroecoeli's

d) 15,76 doeroecoeli's

e) 65,89 %

5

- a) 2,28%
- b) 15,87%
- c) 42,07%
- d) 34,13%

6

- a) $z_1 = 0,5 \rightarrow p(z < 0,5) = 0,6915$
 $z_2 = -0,5 \rightarrow p(z < -0,5) = 0,3085$
 $p(-0,5 < z < 0,5) = 0,6915 - 0,3085 = 0,383$
 38,3 % van de flesjes vallen binnen het interval
- b) $z = 1 \rightarrow p(z > 1) = 0,1587$
 \rightarrow In 15,87 % van de gevallen bevat het flesje meer dan 35 cl

7*

'e' is het juiste antwoord (afgerond 32 sollicitanten)

8

'c' is het juiste antwoord:

kans auto's: $z = 2,2$; $p(z > 2,2) = 0,0139$; kans moto's: $z = 1,2$; $p(z > 1,2) = 0,1151$

$0,1151/0,0139 = 8,3$: de kans voor moto's is dus 8,3 keer groter dan die van auto's

9

'b' is het juiste antwoord (de gevangenen vanaf de leeftijd van (afgerond) 45 jaar)

2. Bivariate beschrijvende statistiek

2.o. Studie van samenhangen en effecten

1*

- a) Het verband is symmetrisch: zowel het aantal vissen kan ervoor zorgen dat er meer of minder vissers komen vissen. Net zoals het aantal vissers de grootte van het visbestand kan bepalen.
- b) Asymmetrisch: het aantal gevangenen (onafh) \rightarrow kans op rellen
- c) Het is een asymmetrisch verband en het BNP is de onafhankelijke variabele

2

- a) Asymmetrisch: Aantal kinderen (onafh) → percentage inkomen aan voedsel
- b) Asymmetrisch: Het weer (onafh) → het humeur van mensen
- c) Symmetrisch
- d) Asymmetrisch: Afgelegde afstand (onafh) → dieselverbruik
- e) Asymmetrisch: Aantal pagina's (onafh) → aantal lezers

2.1. Relaties tussen categorische variabelen**1***

- a) $65,92 - 59,89 = 6,03$ percentagepunten
- b) $1,93/1,49 = 1,30$
- c) $\text{Chi}^2 = 3,88$
- d) Cramer's $V = 0,06$. De relatie tussen beide variabelen is eerder zwak.

2

- a) oddsjongens = $0,2454$; oddsmeisjes = $0,1027$ → oddsratio = $2,39$
De verhouding tussen pesters en niet-pesters ligt $2,39$ keer hoger bij de jongens dan bij de meisjes
- b) $\text{chi}^2 = 259,16$
De grootte van Chi^2 is moeilijk te interpreteren, aangezien die ook afhangt van n en het aantal rijen en kolommen. Chi^2 zegt niets over de sterkte van het verband!

3

- a) $1,34$; De verhouding tussen mensen die tweemaal solliciteerden voor ze hun job hadden en zij die dat viermaal deden, ligt bij autochtonen $1,34$ keer hoger dan bij allochtonen.
- b) Cramer's $V = \sqrt{(2,86 / (108 * (2-1)))} = 0,16$ → Het is een zwak verband

4*

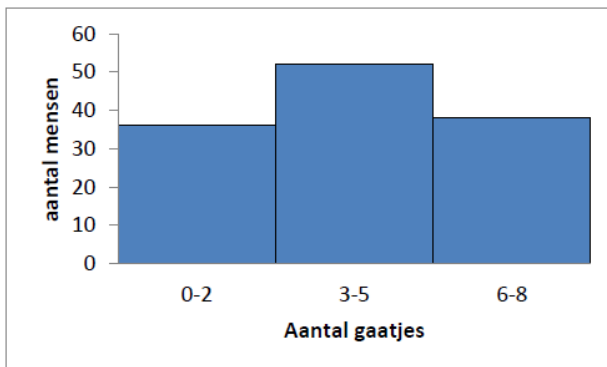
- a) $38,49\% - 15,00\% = 23,49$ percentagepunten. Dit wordt berekend op twee rijpercentages
- b) Gestandaardiseerde residuen:

	Snowboard	Skiër
-60	6.70	-3.99
+60	-6.36	3.78

We kunnen hieruit besluiten dat elke geobserveerde waarde meer dan twee standaardafwijkingen afwijkt van de verwachte waarde indien er geen verband zou zijn. We kunnen dus stellen dat er inderdaad steeds een verband is: met andere woorden mensen onder de 60 hebben een grotere kans om te snowboarden en een lagere kans om te skiën, bij mensen boven de 60 is dat net omgekeerd.

- c) $\chi^2 = 115,59$; $\phi = 0,27$. Uit chi-kwadraat kunnen we weinig besluiten z , maar uit ϕ kunnen we besluiten dat de relatie tussen beide variabelen matig is.

5



- a)
- b) Hiervoor maak je gebruik van de marginale verdeling van het aantal gaatjes. Stel eerst een frequentietabel op, en bereken via het klassenmidden het gemiddelde:

	f_i	m_i	$f_i * m_i$
0-2	36	1	36
3-5	52	4	208
6-8	38	7	266

$$x_{\text{gem}} = 4,05$$

- c) $14,71 - 48,28 = -33,57 \rightarrow 33,57$ percentagepunten. Dit is een verschil tussen twee kolompercentages
- d) $35/23 = 1,52$
- e) Oddszwart product = $28/13 = 2,15$; oddsSnowy white = $10/23 = 0,4347 \rightarrow \text{oddsratio} = 2,15/0,4347 = 4,95$

6

- a) 1,74
- b) 63,73
- c) 0,46: De associatie is medium tot sterk

7

- a) 2,11
b) 0,46

8

- a) 73,39
b) 0,92
c) Rockcity (10%)
d) 0,08

9

- a) $55,72\% - 44,13\% = 11,59$ percentagepunten
b) $\text{odds mannen} = 0,7947$; $\text{odds vrouwen} = 1,2661 \rightarrow \text{oddsratio} = 0,7947/1,2661 = 0,63$

10*

- a) $70\% - 28,57\% = 41,43$ procentpunten

		Temperatuurverschil		Totaal
		Groot	Groot	
Bewolking	Ja	3	7	10
	Rijpercentages	30%	70%	100%
	Nee	5	2	7
	Rijpercentages	71,43%	28,57%	100%
Totaal		8	9	17
Percentageverschil =		41,43	41,43	

- b) $2,3333 / 0,4 = 5,83 \rightarrow$ Interpretatie: de odds voor een klein versus een groot temperatuurverschil is 5,83 keer groter voor bewolkte dagen dan voor heldere dagen. Gebaseerd op:
- Odds voor bewolkte dagen = $7/3 = 2,3333 \rightarrow$ Interpretatie: Er zijn 2,33 keer meer bewolkte dagen met een klein temperatuurverschil dan met een groot temperatuurverschil
 - Odds voor heldere dagen = $2/5 = 0,4 \rightarrow$ Interpretatie: Er zijn 0,4 keer meer heldere dagen met een klein temperatuurverschil dan een groot temperatuurverschil = er zijn 2,5 keer minder heldere dagen met een klein temperatuurverschil dan met een groot temperatuurverschil (reciproque)
- c) 3,93

$$\chi^2 = \sum \frac{(f(o)_i - f(e)_i)^2}{f(e)_i}$$

f(o)		T° verschil		Totaal
		Groot	Klein	
Bewolking	Ja	3	7	10
	Nee	5	2	7
Totaal		8	9	17

f(e)		T° verschil		Totaal
		Groot	Klein	
Bewolking	Ja	5	5	10
	Nee	3	4	7
Totaal		8	9	17

f(o) – f(e)	kwadrateren	/ f(e)
3 – 5 = -2	(-2) ² = 4	4 / 5 = 0,8
5 – 3 = 2	2 ² = 4	4 / 3 = 1,3333
7 – 5 = 2	2 ² = 4	4 / 5 = 0,8
2 – 4 = -2	(-2) ² = 4	4 / 4 = 1
Som =		3,9333

- d) Om de sterkte van het verband te bepalen, moeten we \emptyset nog berekenen.
 $\emptyset = \sqrt{(3,9333 / 17)} = \sqrt{0,2314} = 0,4810 \rightarrow$ volgens de vuistregel is dit een eerder sterk verband.

2.2. Relaties tussen metrische variabelen

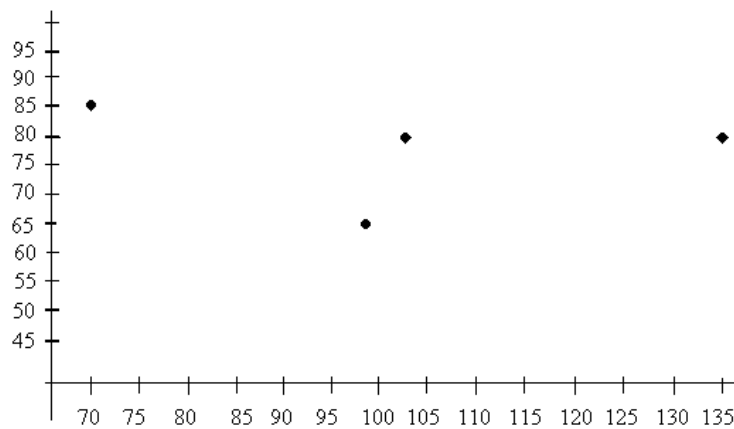
1*

- Een correlatie kun je enkel berekenen tussen twee metrische variabelen, en het al dan niet kaal zijn is nominaal
- De correlatie kan niet groter zijn dan 1
- De correlatie is een onbenoemd getal en behoeft dus geen meeteenheid

2

- $s_{xy} = 6,85$
- $r_{xy} = 0,55$. De samenhang tussen beide variabelen is relatief sterk.
- $Y = -11,87 + 0,15 * X$
- 0,58 \rightarrow Afgerond moet de vrouw al ongeveer 1 ex-liefdespartner gehad hebben.

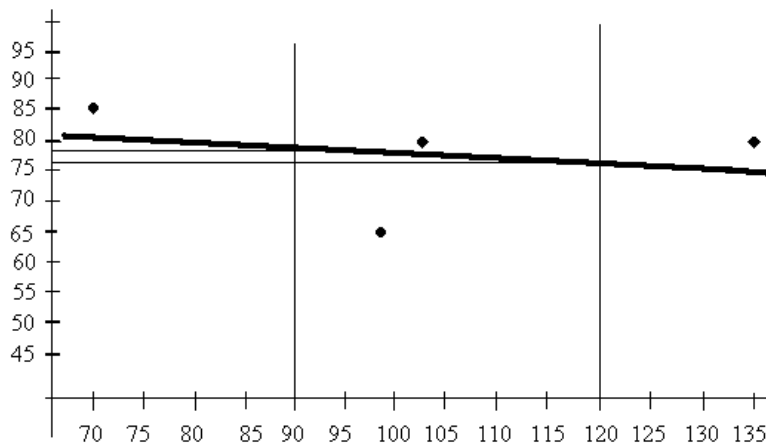
3*

a) $n = 4$ 

b)

c) $s_{xy} = -35$; $r_{xy} = -0,15$. Er is een zwakke negatieve samenhang tussen beide variabelen.

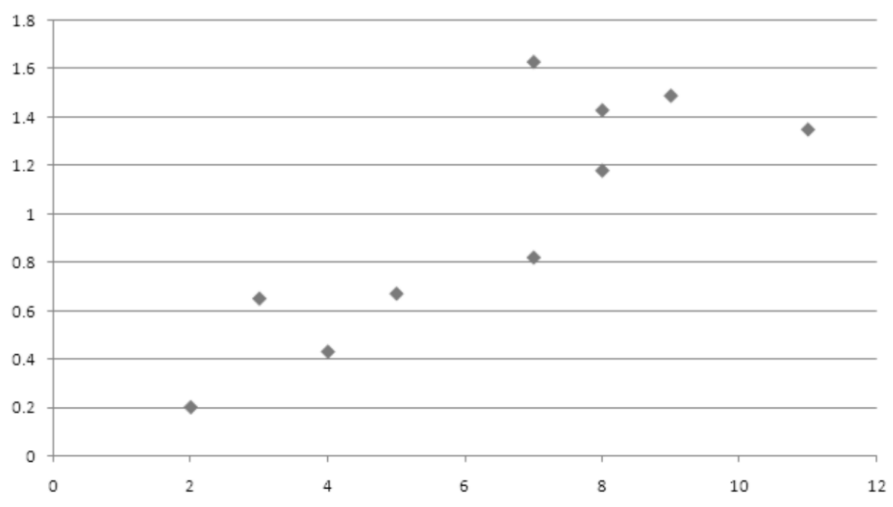
d) Eenvoudige regressievergelijking: $\hat{y} = 82,51 - 0,05 \cdot x$ gestandaardiseerde regressievergelijking: $z^{\wedge}y = -0,15 \cdot z^x$



e)

f) We vinden hier zo'n zwakke en kleine relatie, omdat de leerling met een IQ van 70 zo'n hoge score heeft gehaald. Hoogstwaarschijnlijk hebben we hier met een outlier te maken, die onze regressielijn naar zich toe trekt en zelfs laat overheulen naar de negatieve kant. Indien we informatie hadden over meer leerlingen, zou dit waarschijnlijk niet gebeuren.

4



- a)
- b) $s_{xy} = 1,1856$; $r_{xy} = 0,8448$ ($s_x = 2,8363$; $s_y = 0,4948$); $r^2 = 0,7137 \rightarrow 71,37\%$ van de variantie in het aantal promille alcohol in het bloed wordt verklaard door de variantie in het aantal pinten dat gedronken wordt.
- c) Als $x=15$, dan $\hat{y}=2,2526$ ($x_{\text{gem}} = 6,4$; $y_{\text{gem}} = 0,985$; $\hat{y} = 0,0416 + 0,1474x$) \rightarrow We verwachten 2,25 promille in zijn bloed

Pinten	Promille	\hat{y}	e_i
8	1,18	1,2208	-0,0408
4	0,43	0,6312	-0,2012
7	0,82	1,0734	-0,2534
8	1,43	1,2208	0,2092
7	1,63	1,0734	0,5566
5	0,67	0,7786	-0,1086
3	0,65	0,4838	0,1662
2	0,2	0,3364	-0,1364
9	1,49	1,3682	0,1218
11	1,35	1,6630	-0,3130

5

- a) Een asymmetrische relatie. Onafh: Duur van de film. Afh: Het aantal bezoekers onder de 14 jaar.
- b) $s_{xy} = -106,8$; $r_{xy} = -0,85$. De relatie tussen beide variabelen is heel sterk te noemen. Bovendien is ze negatief, wat wil zeggen dat, hoe langer de duur van een film, hoe minder bezoekers onder de 14 jaar we verwachten.
- c) Enkelvoudige vergelijking: $\hat{y} = 127,05 - 0,66 \cdot x$. Gestandaardiseerde vergelijking: $z^{\hat{y}} = -0,85 \cdot z_x$
- d) Verwachte y : $\hat{y} = 127,05 - 0,66 \cdot 120 = 47,82$

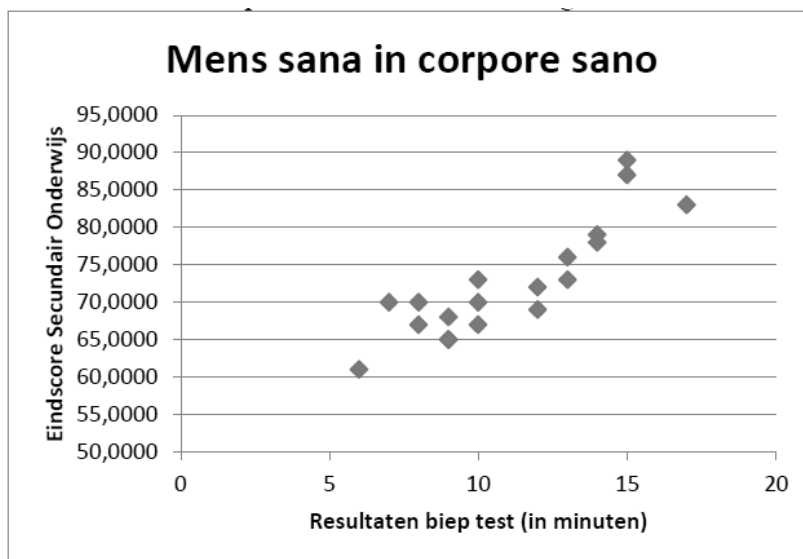
- e) Aantal standaardafwijkingen boven of onder een gemiddelde \bar{z} berekenen via gestandaardiseerde regressievergelijking: $z = -0,85 \cdot 0,6 = -0,51$. Absolute aantallen: omzetten naar y . dus: $y = s_y \cdot z_y + y_{\text{gem}} = 69,96$
- f) Determinatiecoëfficiënt berekenen: $r^2 = -0,85^2 = 0,72$. 72 % van de variantie in het aantal bezoekers onder de 14 jaar wordt verklaard door de duur van de film.

6*

- a) Fout
b) Correct
c) Correct
d) Correct
e) Fout

7

- a) 19
b) We verwachten een positieve relatie/samenhang:



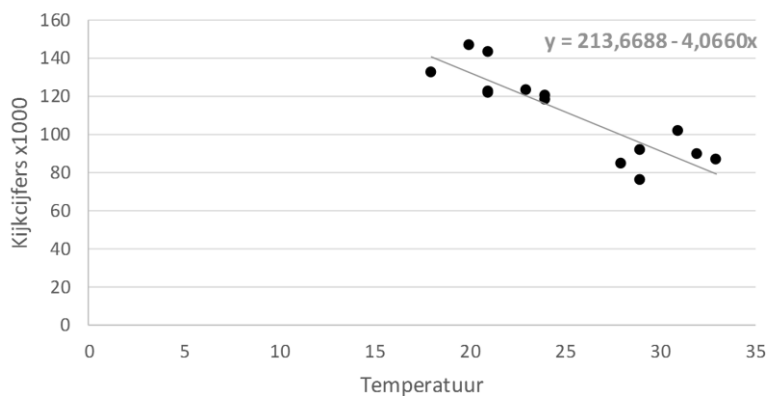
- c) covariantie = 20,25
d) Eenvoudige regressievergelijking: $y = 48,89 + 2,15x$
e) Gestandaardiseerde regressievergelijking: $z_y = 0,87z_x$

f) Gestandaardiseerde waarden:

Biep-test (in min.)	Eindscore Secundair onderwijs	Biep test (in min.) gestandaardiseerd	Eindscore Secundair onderwijs gestandaardiseerd
9	65	-0,6855	-1,0209
10	73	-0,3599	0,0347
8	70	-1,0111	-0,3611
14	79	0,9425	0,8265
15	89	1,2681	2,1461
12	69	0,2913	-0,4931
9	65	-0,6855	-1,0209
10	67	-0,3599	-0,7570
12	72	0,2913	-0,0972
17	83	1,9193	1,3543
7	70	-1,3367	-0,3611
6	61	-1,6623	-1,5488
13	76	0,6169	0,4306
15	87	1,2681	1,8822
8	67	-1,0111	-0,7570
9	68	-0,6855	-0,6251
10	70	-0,3599	-0,3611
13	73	0,6169	0,0347
14	78	0,9425	0,6945

8

Kijkcijfers weerbericht naar temperatuur



- a)
- b) We verwachten 103887 kijkers (want als $x = 27$, dan $y^{\wedge} = 103,8868 \rightarrow$ variabele kijkcijfers uitgedrukt in x1000).
- c) $r^2 = 0,7763$ dus temperatuur verklaart 77,63% van de variatie in kijkcijfers
- d) Negatieve relatie, dus Sabine heeft gelijk.

9*

- a) Correct → In het bivariate geval geldt dat $B = r$. Dus $R^2 = (-0,58)^2 = 0,3364 = 34\%$
- b) Correct → Interpretatie richtingscoëfficiënt $b =$ "als x met 1 toeneemt, hoeveel neemt y dan toe?". Hier dus: als x met 2 toeneemt, dan neemt y met $2 \cdot -1,14$ toe.
- c) Fout → Interpretatie intercept a : "als $x = 0$, dan is $y = a$ ". Maar niet: als $y = 0$, dan $x = a$
- d) Fout → Interpretatie gestandaardiseerde richtingscoëfficiënt B , maar cijfers voor b
- e) Correct → $y = 6,21 - 1,14 \cdot x$ dus verwachte waarde $= 6,21 - 1,14 \cdot 2,5 = 3,36$. Dus residueel $= 4 - 3,36 = 0,64$

10

- a) $s_{xy} = 834,0832 / (n-1) = 834,0832 / 4 = 208,5208$
- b) $r_{xy} = s_{xy} / (s_x \cdot s_y) = 208,5208 / (13,20 \cdot 15,89) = 208,5208 / 209,7480 = 0,9941$
- c) $b = s_{xy} / s_x^2$ (covariantie/variantie X) $= 208,5208 / 174,24 = 1,20$
- d) $B =$ correlatiecoëfficiënt $= 0,9941$; $A = 0$. Vergelijking: verwachte $z_y = 0,9941 \cdot z_x$

11

20 cc per Mj toename = verschil in y bij verschil 1 eenheid in $x = b$

Bij -5 Mj : 540 cc

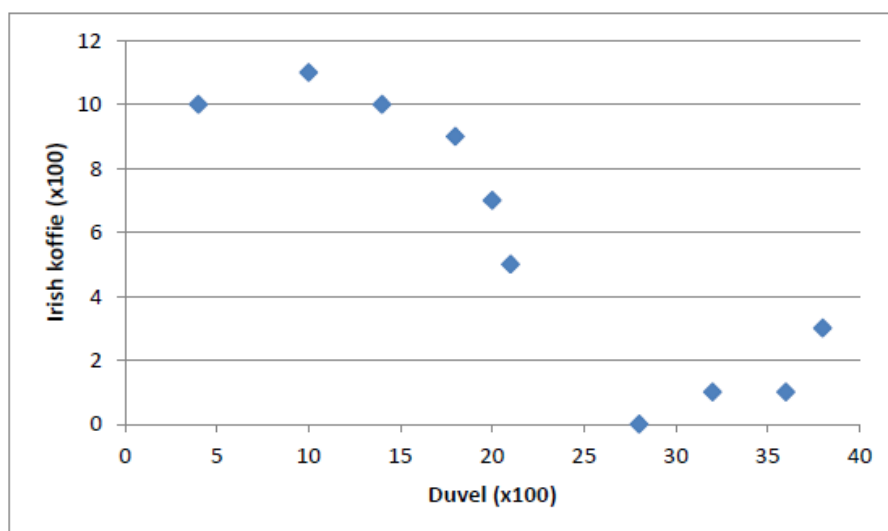
$$y = a + b \cdot x \Rightarrow a = y - (b \cdot x)$$

$$a = 540 \text{ cc} - (20 \text{ cc} \cdot (-5)) = 640 \text{ cc}$$

$$\text{Dus: } \hat{y} = 640 + 20x$$

12

- a) Aantal duvels en Irish koffie's zijn ratio variabelen. (uur is meeteenheid)



b)

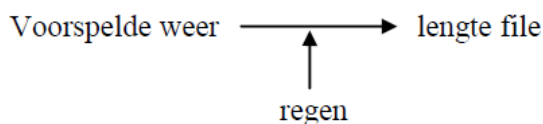
- c) Het heeft geen zin om hier de correlatie te berekenen, aangezien er geen lineair verband is (zie scatterplot).
- d) nee, want geen lineaire relatie.

13

- a) Correlatie $r = 0,95$
- b) Aangezien het hier gaat over een asymmetrisch verband (de smaak van de koffie is afhankelijk van het percentage Braziliaanse koffie in de pads), kunnen we regressie toepassen. Regressierechte: $y = 3,03 + 0,07x$
- c) 6

14

- a) $\hat{y} = -4,34 + 0,49x$ (met x = voorspelde graden en y = filelengte)
- b) 2,97 km
- c) $r = 0,89$: sterke positieve samenhang.



- d) Dit is een interactie-effect.
- e) 79,44% van de variantie

15

- a) $s_{xy} = 112 / (6-1) = 22,40$
 $r_{xy} = 22,4000 / (\sqrt{21,6} * \sqrt{46,4}) = 0,71$
- b) $R^2 = 0,7076^2 = 0,5007 \rightarrow 50,07\%$ van de variantie in y wordt verklaard door de variantie in x
- c) Eenvoudige regressievergelijking:
 $b = 0,7076 * (\sqrt{46,4} / \sqrt{21,6}) = 1,0371$
 $a = 10 - (1,0371 * 4) = 5,8516$
 $\hat{y} = 5,85 + 1,04x$

Gestandaardiseerde regressievergelijking:

$$B = r = 0,71$$

$$z_y = 0,71 * z_x$$

- d) $\hat{y} = 5,85 + (1,04 * 6) = 12,09$

16*

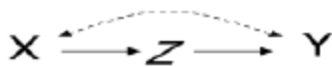
- a) $s_{xy} = -42,25$
 $r_{xy} = -0,7264 \rightarrow$ Er is een sterk negatief verband tussen x (trainingsduur) en y (duur afleggen watersportbaan)
- b) $R^2 = (-0,7264)^2 = 0,5277 \rightarrow 52,77\%$ van de variantie in y (duur afleggen watersportbaan) wordt verklaard door de variantie in x (trainingsduur)
- c) $\hat{y} = 39,9959 - (0,1775 * x)$
 $\hat{y} = 39,9959 - (0,1775 * 78) = 26,1509$
 Dus: We verwachten dat het team 26,15 minuten over de omloop van de watersportbaan zal doen.

Duur training (x)	Tijdsduur (y)	\hat{y}	e_i
86	25	24,52	0,48
51	29	30,82	-1,82
73	28	26,86	1,14
96	18	22,72	-4,72
56	29	29,92	-0,92
83	28	25,06	2,94
69	29	27,58	1,42
83	28	25,06	2,94

2.3. De limieten van de beschrijvende statistiek

1

- a) Intermediaire variabele (indirect verband) (x: naam; z: verwachtingen leerkracht; y: prestaties):

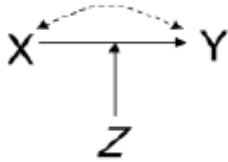


- b) Gemeenschappelijke oorzaak (z: omgeving; x: naam; y: prestaties):



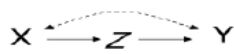
2

Het gaat hier over een interactie-effect (z: geslacht; x: het tonen van de bovenbil; y: kans op succesvol liften):



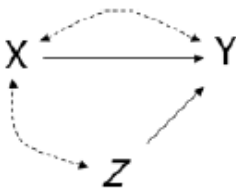
3

Intermediaire variabele (indirect verband) (x: aantal pinten bier; z: aantal hersencellen dat gedood wordt; y: punten voor statistiek):



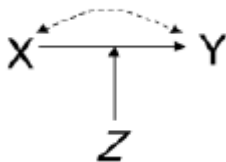
4

Verstrengeling (x: aantal pinten bier; z: slaapttekort van studenten; y: punten voor statistiek):



5*

- Allochtonen hebben een hogere richtingscoëfficiënt voor de relatie tussen SES en hun eindpercentages dan autochtonen. Een verschil van 1 punt op de SES-schaal maakt dus een groter verschil in eindpercentages dan voor autochtonen. In gewone taal wil dit zeggen dat de impact van de sociaaleconomische status op de eindpercentages groter is voor allochtonen dan voor autochtonen.
- Dit is een interactie-effect



6

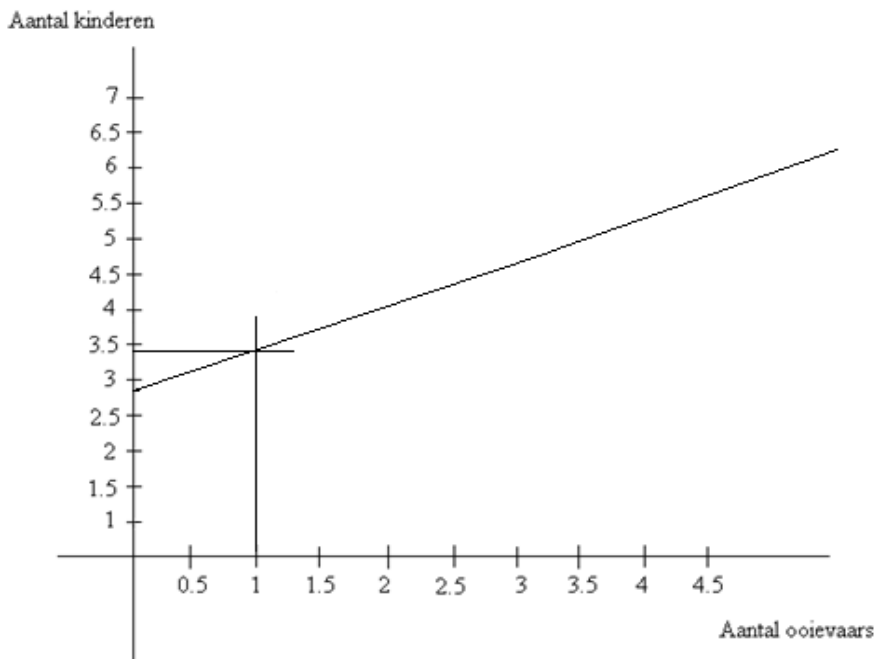
- chi-kwadraat bedraagt 113; Cramer's V is 0,62. Dit laatste wijst op een sterk verband.
- Voor de -50-jarigen: chi-kwadraat is 0,65; Cramer's V is 0,026. Dit wijst op een zwak verband.

Voor de +50-jarigen: chi-kwadraat is 0,42; Cramer's V is 0. Dit wijst op een zeer zwak verband

- c) Er is hier sprake van een gemeenschappelijke oorzaak: de leeftijd van de patiënten bepaalt zowel hun aantal ziekenhuisopnames als het aantal tanden in hun mond.

7

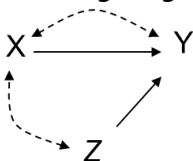
- a) $b = 0,58$; $a = 2,83$



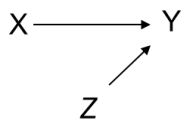
- b) Als de resultaten gesplitst worden naar platteland versus stedelijke omgeving, valt het verband tussen aantal ooievaars en aantal baby's weg. Hier is een gemeenschappelijke oorzaak aan het werk: op het platteland zijn er meer ooievaars, maar ook meer baby's

8*

- a) Voor elk jaar dat men langer studeert, scoort men 0,511 eenheden lager op de attitudeschaal.
- b) Er is een samenhang tussen urbanisatiegraad en sociaaleconomische status. Door deze samenhang verkleint het effect van sociaaleconomische status wanneer we controleren voor het effect van urbanisatiegraad in model 2. Dit is een voorbeeld van verstremgeling.



- c) Een ongewijzigd effect (of minimaal verschil) wijst erop dat er geen verband is tussen urbanisatiegraad en leeftijd.



- d) Model 3 heeft de grootste verklarende kracht, want de proportie verklaarde variantie (R^2) is het grootst.
- e) Neen; door toevoeging van deze variabele zien we dat R^2 afneemt. Model 4 is dus nodeloos complex.
- f) De variabele urbanisatiegraad woonplaats, want de toename van de proportie verklaarde variantie (R^2) is het grootst wanneer we deze toevoegen ($0,18 - 0,13 = 0,05$).
- g) Dat kunnen we niet afleiden uit deze tabel. Om de grootte van effecten te kunnen vergelijken, hebben we namelijk de gestandaardiseerde regressiecoëfficiënten B nodig. In deze opgave is enkel b gegeven, met verschillende maateenheden.

3. Inductieve statistiek

- 3.1. De studie van de systematiek van het toeval: kansrekenen
 3.2. Stochastische variabelen
 3.3. Studie van steekproefvariabiliteit
 3.4. Basistools van de inductieve statistiek

3.5 Inferentie voor verwachtingen

1*

- a) Gegevens: $\sigma = 2483$; $n = 83$; $x_{\text{gem}} = 8496$; $s = 1800$
 95%-BI $\rightarrow z^* = 1,96$
 $8496 - (1,96 * 2483 / \sqrt{83}) \leq \mu \leq 8496 + (1,96 * 2483 / \sqrt{83})$
 $8496 - 534,1896 \leq \mu \leq 8496 + 534,1896$
 $7961,81 \leq \mu \leq 9030,19$
 Dus: een 95%-BI rond het steekproefgemiddelde is $[7961,81; 9030,19]$.
 Conclusie: we kunnen met 95% betrouwbaarheid stellen dat het gemiddeld aantal bomen per bos in Vlaanderen ligt tussen 7962 en 9030.
- b) Smaller, want in ruil voor een lagere betrouwbaarheid bekomen we een exactere schatter.

- c) De kritieke z-score van een 86%-BI kunnen we niet terugvinden in de t-tabel. We moeten dus op zoek in de z-tabel.

Als 86% van de steekproefgemiddelden in het interval moet liggen, dan ligt 14% buiten het interval ($100\% - 86\% = 14\%$). Die 14% ligt verspreid over de linker- en rechterkant van de verdeling, dus blijft per staart 7% over (want symmetrische verdeling). De positieve z-score kunnen we dan op 2 manieren bekomen (cfr. oefensessie 3):

De negatieve z-score opzoeken die hoort bij een kans van 0,07 en daar de absolute waarden van nemen

De positieve z-score bereken ($0,07 + 0,86 = 0,93$, of $1 - 0,07 = 0,93$) en opzoeken

We vinden dus een z-score van 1,475.

Dus: 85%-BI $\rightarrow z^* = 1,475$

$$8496 - (1,475 * 2483 / \sqrt{83}) \leq \mu \leq 8496 + (1,475 * 2483 / \sqrt{83})$$

$$8496 - 402,0049 \leq \mu \leq 8496 + 402,0049$$

$$8093,9951 \leq \mu \leq 8898,0049$$

Dus: een 86%-BI rond het steekproefgemiddelde is [8094; 8898].

Conclusie: we kunnen met 86% betrouwbaarheid stellen dat het gemiddeld aantal bomen per bos in Vlaanderen ligt tussen 8094 en 8898.

- d) Smaller; als de steekproefomvang toeneemt zal de spreiding van de steekproevenverdeling afnemen (d.i. minder variatie tussen steekproefgrootheden).

e) $m = 200 = 1,96 * 2483 / \sqrt{n}$

Daaruit volgt: $\sqrt{n} = (1,96 * 2483) / 200$

Dus: $n = ((1,96 * 2483) / 200)^2 = 592,11$

Besluit: Ze hadden minstens 593 bossen in de steekproef moeten opnemen.

2*

a) $H_0: \mu = 46$

$H_a: \mu > 46 \rightarrow$ rechts eenzijdig testen

b) Verwachting van het steekproefgemiddelde = 46

Standaardafwijking van het steekproefgemiddelde = $15,76 / \sqrt{25}$

Dus de steekproevenverdeling volgt $N(46; 3,1520)$

c) $z = (57,6 - 46) / 3,152 = 3,68$

d) $p(z \geq 3,68) < 0,0002$

e) Vergelijken met een significantieniveau van 5%: $0,0002 \leq 0,05$

Vergelijken met een significantieniveau van 1%: $0,0002 \leq 0,01$

Vergelijken met een significantieniveau van 0,1%: $0,0002 \leq 0,001$

Het verschil is dus significant tot op het 0,1%-niveau.

- f) Mensen die Rimpelrock bijwonen hebben een significant hogere leeftijd dan de rest van de Vlaamse bevolking.

3*

- a) 1. Hypotheses

$H_0 : \mu = 65 ; H_a : \mu \neq 65 \rightarrow$ tweezijdig toetsen

2. Steekproevenverdeling

$\sigma_{\text{gem}} = 2,49 ; \mu_{\text{gem}} = 65 \rightarrow N(65; 2,49)$

3. Toetsingsgrootte

$z = (58 - 65) / 2,49 = -2,81$

4. Overschrijdingskans

$p = 0,0025$. Tweezijdig dus $p = 2 * 0,0025 = 0,005$.

5. Conclusie?

Vergelijken met een significantieniveau van 5%: $0,005 \leq 0,05$

Vergelijken met een significantieniveau van 1%: $0,005 \leq 0,01$

Vergelijken met een significantieniveau van 0,1%: $0,005 > 0,001$

Dus significant op het 10%, 5% én het 1%-niveau, maar niet 0,1% niveau.

- b) We kunnen dus besluiten dat de nulhypothese niet klopt. Er is een significant verschil tussen de leerlingen in Neder-Overheembeek en de populatie.

4*

- a) Significantietoets:

Hypotheses:

$H_0 : \mu = 6,5$

$H_a : \mu \neq 6,5$

Steekproevenverdeling:

$t(6,5; 11 / \sqrt{132})$ met $df = 131$

Toetsingsgrootte:

$t = (8,19 - 6,5) / (11 / \sqrt{132}) = 1,69 / 0,9574 = 1,77$

Overschrijdingskans:

$P(t \geq 1,77)$ bij $df = 100$: $p < 0,05$

Maar! Tweezijdig: $p < 2 * 0,05 \rightarrow p < 0,1$

Conclusie:

Vergelijken met een significantieniveau van 5%: $0,1 > 0,05$

Vergelijken met een significantieniveau van 1%: $0,1 > 0,01$

Vergelijken met een significantieniveau van 0,1%: $0,1 > 0,001$

Dus niet significant op het 0,05-niveau en dus ook niet de rest.

- b) 95%-betrouwbaarheidsinterval:

$df = 131 \rightarrow t^*$ bij $df = 100 = 1,984$

$$8,19 - (1,984 * 11 / \sqrt{132}) \leq \mu \leq 8,19 + (1,984 * 11 / \sqrt{132})$$

$$8,19 - 1,9024 \leq \mu \leq 8,19 + 1,9024$$

$$6,29 \leq \mu \leq 10,09$$

5*

a) $\sigma_{\text{gem}} = 2,19$

b) $m = 1,96 * 2,19 = 4,29$

6*

a) Gegeven:

$$n_1 = 13 \quad x_{1\text{gem}} = 21 \quad s_1 = 4,3 \quad s^2_1/n_1 = 1,4223$$

$$n_2 = 21 \quad x_{2\text{gem}} = 26 \quad s_2 = 8,3 \quad s^2_2/n_2 = 3,2805$$

$$(x_{1\text{gem}} - x_{2\text{gem}}) = -5 \text{ dus } 5$$

$$df = 12 \text{ dus } t^* = 1,356$$

$$BI: 5 - 1,356 * \sqrt{(1,4223 + 3,2805)} \leq \mu_1 - \mu_2 \leq 5 + 1,356 * \sqrt{(1,4223 + 3,2805)}$$

$$2,06 \leq \mu_1 - \mu_2 \leq 7,94$$

$$[2,06; 7,94]$$

b) 1. Hypotheses

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

2. Steekproevenverdeling?

$$n_1 = 13 \quad x_{1\text{gem}} = 21 \quad s_1 = 4,3 \quad s^2_1/n_1 = 1,4223$$

$$n_2 = 21 \quad x_{2\text{gem}} = 26 \quad s_2 = 8,3 \quad s^2_2/n_2 = 3,2805$$

$$(x_{1\text{gem}} - x_{2\text{gem}}) = -5$$

$$(\mu_1 - \mu_2) = 0$$

$$t(0; \sqrt{(1,4223 + 3,2805)}) \text{ met } df = 12$$

3. Toetsingsgrootheid

$$t = (-5 - 0) / \sqrt{(1,4223 + 3,2805)} = -2,31$$

4. Overschrijdingskans

$$\text{Als } df = 12 \text{ dan } P(t \leq -2,31) < 0,02$$

$$P < 0,04 \text{ (tweezijdig!)}$$

5. Conclusie?

$$\text{Vergelijken met een significantieniveau van 10\%: } 0,04 \leq 0,10$$

$$\text{Vergelijken met een significantieniveau van 5\%: } 0,04 \leq 0,05$$

$$\text{Vergelijken met een significantieniveau van 1\%: } 0,04 > 0,01$$

$$\text{Vergelijken met een significantieniveau van 0,1\%: } 0,04 > 0,001$$

Dus significant op het 10% niveau, het 5% niveau maar niet op het 1% niveau en op het 0,1% niveau.

7

a) $t(2,22; 1,03/\sqrt{282})$ De verdeling is daarenboven niet normaal verdeeld, want de variabele is zelf niet normaal verdeeld, ze is discreet $\{0,1,2,3,4\}$. We kunnen echter wel een steekproevenverdeling opstellen wegens centrale limietstelling (CLS).

b) Ja, we mogen dit berekenen

c) $2,06 < \mu < 2,38$

99%-betrouwbaarheidsinterval:

$df = 281 \rightarrow t^*$ bij $df = 100 = 2,626$

$2,22 - 2,626 * (1,03/\sqrt{282}) < \mu < 2,22 + 2,626 * (1,03/\sqrt{282})$

$2,06 \leq \mu \leq 2,38$

8

a) De steekproevenverdeling is $t(2,22; 1,03/\sqrt{282})$. Het gaat hier om een t-verdeling, want we kennen de standaardafwijking van de populatie niet. Let op, de variabele zelf is niet normaal verdeeld, want ze is discreet \rightarrow waarden $\{0, 1, 2, 3, 4\}$. We kunnen echter wel een steekproevenverdeling opstellen omdat CSL voldaan is.

b) Mogen we hier een betrouwbaarheidsinterval op berekenen? Betrouwbaarheidsinterval voor steekproefgemiddelde (xgem) mag

c) $df = 281 \rightarrow t^*$ bij $df = 100 = 2,626$

$2,22 - 2,626 * (1,03/\sqrt{282}) \leq \mu \leq 2,22 + 2,626 * (1,03/\sqrt{282})$

$2,059 \leq \mu \leq 2,381$

9

a) $t(6; 1,2/\sqrt{56})$ met $df = 55$

b) Hypotheses:

$H_0: \mu = 6$

$H_a: \mu < 6$

Steekproevenverdeling:

$t(6; 1,2/\sqrt{56})$ met $df = 55$

Toetsingsgrootheid

$t = (4,5 - 6)/(1,2/\sqrt{56}) = -1,5/0,1604 = -9,35$

Overschrijdingskans

$P(t \leq -9,35)$ bij $df = 50: p < 0,0005$

c) BI: $4,23 - 4,77$

t bij $df=55$ voor 90% BI= 1.676. Dus BI wordt: $4.5 \pm 1.676 * (1.2/\sqrt{55})$

10

- a) M.b.t. de meting in juli is deze 0,2182; m.b.t. de meting in september is deze 0,1945
 $s_1(\text{juli}) = 0,023 \cdot \sqrt{90} = 0,2182$
 $s_2(\text{sept}) = 0,029 \cdot \sqrt{45} = 0,1945$
- b) $t = -17,84 \rightarrow p < 0,0005$. Tweezijdig dus $p < 0,001$. Significant op het 5%-niveau, op het 1%-niveau en op het 0,1%-niveau.

11

- a) $z = -0,77 \rightarrow p = 0,2206$. We gebruiken de populatieverdeling of de steekproevenverdeling met $n = 1$
- b) $N(8; 0,65)$
- c) $z = -3,08 \rightarrow p = 0,0010$
- d) zie bij c). Zie de Centrale Limiet Stelling

12

Nee, er is geen EAS getrokken uit de luisteraars en daardoor mogen we de principes van betrouwbaarheidsintervallen niet zomaar toepassen. Hier werkt men met zelfselectie: respondenten mogen zichzelf selecteren, en dat levert nooit een toevallige steekproef op.

13*

$$p = 0,05 \rightarrow z = -1,645$$

$$x_{\text{gem}} = -1,645 \cdot \sigma / \sqrt{n} + \mu$$

$$= (-1,645 \cdot 3,1 / \sqrt{22}) + 13,6 = 12,51$$

14

Dit is inferentie voor verschil tussen 2 verwachtingen.

Stap 1: hypothesen

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

Stap 2: steekproevenverdeling?

$n_1 = 23$	$x_{1\text{gem}} = 16,32$	$s_1 = 8,6$	$s^2_1/n_1 = 3,2157$
$n_2 = 36$	$x_{2\text{gem}} = 22,47$	$s_2 = 4,2$	$s^2_2/n_2 = 0,4900$

$$(x_{1\text{gem}} - x_{2\text{gem}}) = -6,15$$

$$(\mu_1 - \mu_2) = 0$$

t-verdeling $(0; \sqrt{3,2157 + 0,4900})$ met $df = 22$

Stap 3: toetsingsgrootheid

$$t = (-6,15 - 0) / \sqrt{(3,2157 + 0,4900)} = -3,19$$

Stap 4: overschrijdingskans

Als $df = 22$ dan $P(t \leq -3,19) < 0,0025$

Hier tweezijdig, dus $P < 0,005$

Stap 5: conclusie

Dit is significant op het 10% niveau, het 5% niveau, het 1% niveau maar niet op het 0,1%.

15

- a) $123,8 - 1,645 * 2,58 \leq \mu \leq 123,8 + 1,645 * 2,58$ $119,56 < \mu < 128,04$
- b) $123,8 - 1,96 * 2,58 \leq \mu \leq 123,8 + 1,96 * 2,58$ $118,74 < \mu < 128,86$
- c) $123,8 - 2,576 * 2,58 \leq \mu \leq 123,8 + 2,576 * 2,58$
 $117,15 \leq \mu \leq 130,45$
- d) Als het betrouwbaarheidsniveau toeneemt, worden de foutenmarges groter

16

- a) $df = 45 \rightarrow$ BI: 268,31 - 301,69
- b) $df = 45 \rightarrow$ BI: 274,24 - 295,76
- c) $H_0 : \mu = 267 ; H_a : \mu > 267 \rightarrow$ eenzijdig toetsen
- d) Toetsingsgrootheid: $t = 2,18 \rightarrow$ overschrijdingskans: $p < 0,02$. Significant op het 10%- en het 5%-significantieniveau, maar niet op het 1%-significantieniveau, noch op het 0,1%-niveau. In Wenduine werden inderdaad significant meer ijsjes verkocht dan in de rest van de Vlaamse kuststeden.

17

- a) 0,66
- b) $H_0 : \mu = 3,7. H_a : \mu > 3,7$
- c) $p < 0,005$. We kunnen besluiten dat de leden van 'De gewillige katers' significant meer met een kater in de les zitten dan leden van andere studentenclubs

3.6 Inferentie voor fracties

1

a) Foutenmarge :

 z^* bij 90 % BI: 1,645 $\hat{p} = 27/37 = 0,7297$ $m = 1,645 * \sqrt{((0,7297 * 0,2703)/37)}$ $m = 1,645 * 0,0730$ $= 0,1201$

Betrouwbaarheidsinterval:

 $\hat{p} = 0,7297$ $0,7297 - 0,1201 \leq p \leq 0,7297 + 0,1201$ $0,61 \leq p \leq 0,85$

b) Hypotheses

 $H_0: p = 0,75$ $H_a: p < 0,75 \rightarrow$ eenzijdig!

Steekproevenverdeling

z-verdeling: $N(0,75; \sqrt{(0,75 * 0,25)/37})$

Toetsingsgrootheid

 $z = (0,7297 - 0,75) / 0,0712$ $z = -0,29$

Overschrijdingskans

 $p(z \leq -0,29) = 0,39$ Conclusie: $0,39 > 0,05 \rightarrow$ niet significant!

2*

a) $z = 1,49 \rightarrow p = 0,0681$. Niet significant op het 0,05-niveau

b) BI = [-0,03 - 0,11]

3

a) BI: 0,11 - 0,56

b) BI: 0,05 - 0,61

c) $z = 2,27 \rightarrow p = 0,0116$. Significant op het 10%- en het 5%-niveau, maar niet op het 1%-niveau, noch op het 0,1%-niveau. We kunnen besluiten dat er inderdaad meer golddiggers in 'Old Gold Hulch' zijn die het voorbije jaar goud gevonden hebben dan in de rest van Texas.

4

-
- a) BI: 0,03 - 0,08
- b) BI: -0,02 - 0,12
- c) $z = -2,08 \rightarrow p = 0,0188$. Significant op het 10%- en het 5%-niveau, maar niet op het 1%-niveau, noch op het 0,1%-niveau. We kunnen besluiten dat jongeren inderdaad minder aangaven het SMAK bezocht te hebben dan ouderen.

5

-
- a) BI: -0,01 - 0,20
- b) BI: 0,05 - 0,14
- c) $z = 0,86 \rightarrow p = 0,1949$. Niet significant, noch op het 10%-, het 5%-, het 1% - of het 0,1%-significantieniveau. We kunnen niet besluiten dat Ronaldinho een grotere fractie van zijn doeltrappen scoort dan andere voetballers tijdens de wereldbekerwedstrijden.

6

-
- a) BI: 0,01 - 0,50
- b) Betrouwbaarheidsinterval: $SE^p = 0,13$; Significantietoets: $SE^p = 0,10$
- c) $z = 0,97 \rightarrow p = 0,1660$. De man heeft ongelijk: zijn steekproef verschilt niet significant van de resultaten voor alle vogels

7*

-
- a) 73%-BI $\rightarrow z^*=1,105 \rightarrow BI = [0,16; 0,27]$
- b) 1. Hypothesen: $H_0 : p_1 = p_2$; $H_a : p_1 \neq p_2$
 2. Steekproevenverdeling:
 $\hat{p} = (65+103) / (120+314) = 0,3871$
 $SE(D) = \sqrt{[0,3871*(1-0,3871)*((1/120) + (1/314))]} = \sqrt{[0,3871*0,6129*0,0115]}$
 $= \sqrt{0,0126} = 0,0520$
 $\rightarrow N(0; 0,0520)$
 3. Toetsingsgrootte
 $(\hat{p}_1 - p_2) = 0,5417 - 0,3280 = 0,2137$
 $z = (0,2137 - 0) / 0,1122 = 4,11$
 4. overschrijdingskans
 $p(z \geq 4,11) < 0,0002$
 Tweezijdige test, dus $0,0002 * 2 = 0,0004$

5. Conclusie

Vergelijken met $\alpha = 5\%$: $0,0004 < 0,05 / 0,01 / 0,001$ dus geen significant verschil. De nulhypothese kan dus verworpen worden. We kunnen besluiten dat er een significant verschil bestaat tussen de fractie jongens die naar 'South park' kijken, en de fractie meisjes die dat doen (al kunnen we niets zeggen over de richting van dat verschil).

8

- a) 80%-BI: $z^* = 1,282$; BI: $-0,11 - 0,15$
 96%-BI: $z^* = 2,054$; BI: $-0,19 - 0,23$
 99%-BI: $z^* = 2,576$; BI: $-0,25 - 0,29$
- b) $\hat{p} = 0,3214$; $SED = 0,1030 \rightarrow z = 0,20 \rightarrow p = 0,4207$. Het verschil is niet statistisch significant

3.7 Inferentie voor kruistabellen

1

- a) $17/52 = 0,3269 \rightarrow 32,69\%$. Dit is een rijpercentage
 b) $\chi^2 = 18,45$; Cramer's $V = 0,39$
 c) Significant op het 10%-, 5%-, 1%- en 0,1%-niveau

2*

- a) $\chi^2 = 18,63$
 b) $df = 3$
 c) $p < 0,0005$. Significant op het 10%-, het 5%-, het 1% - en het 0,1%-significantieniveau.

3.8 Overzichtsoefeningen

1

- a) $df = 14 \rightarrow$ BI: $3,07 - 12,93$
 b) $df = 14 \rightarrow$ BI: $-4,02 - 20,02$
 c) $t = -1,75 \rightarrow p < 0,10$. Significant op het 10%-niveau, maar niet op het 5%-, het 1%- en het 0,1%-significantieniveau. We kunnen niet besluiten dat krokodillen in de wildernis meer eieren leggen dan de krokodillen in gevangenschap (niet sig op het 5%-niveau).

2*

1. Hypothesen

$$H_0 : \mu_1 = \mu_2 ; H_a : \mu_1 > \mu_2$$

2. Steekproevenverdeling

$$n_1 = 75 \quad x_{gem1} = 52 \quad s_1 = 13 \quad \rightarrow s^2/n = 13^2/75 = 2,2533$$

$$n_2 = 53 \quad x_{gem2} = 49 \quad s_2 = 11 \quad \rightarrow s^2/n = 11^2/53 = 2,2830$$

$$SE = \sqrt{(2,2533 + 2,2830)} = 2,1299$$

$$x_{gem1} - x_{gem2} = 52 - 49 = 3$$

3. Toetsingsgrootheid

$$t = (3 - 0) / 2,1299 = 1,409$$

4. Overschrijdingskans

$$p < 0,1$$

5. Conclusie

Er is geen significant verschil in gemiddelde verkoop, dus kunnen we er niet zeker van zijn dat er een stijging in verkoop plaatsvond.

3*

f(o)	man	vrouw	totaal
ja	297	325	622
nee	795	762	1557
totaal	1092	1087	2179

f(e)	man	vrouw	totaal
ja	312	310	622
nee	780	777	1557
totaal	1092	1087	2179

f(o)-f(e)	(f(o)-f(e)) ²	/f(e)
-15	225	0,7212
15	225	0,2885
15	225	0,7258
-15	225	0,2896
som=		2,0251

$$\text{Chi}^2 = 2,03 \text{ en } df = 1$$

$p < 0,20$ dus het verband is niet significant, noch op het 10%-, het 5%-, het 1% - of het 0,1%- significantieniveau

4

a) $df = 20 \rightarrow \text{Bl: } 1,13 - 4,87$

b) $df = 20 \rightarrow \text{Bl: } 0,74 - 5,26$

c) $t = 2,77 \rightarrow p < 0,01$. Significant op het 10%-, het 5%- en het 1% -niveau, maar niet op het 0,1%-significantieniveau. We kunnen besluiten dat pesters meer beste vrienden hebben dan niet-pesters.

5

a) $df = 24 \rightarrow BI: 7,46 - 10,54$

b) $df = 24 \rightarrow BI: 6,48 - 11,52$

c) $t = 3,33 \rightarrow p < 0,0025$. Tweezijdig dus: $p < 0,0025 * 2 \rightarrow p < 0,005$. Significant op het 10%-, het 5%- en het 1%-significantieniveau, maar niet op het 0,1%-niveau. We kunnen besluiten dat er een significant verschil is tussen vrouwen en de totale populatie wat betreft het aantal keren bellen per week, maar we kunnen niets zeggen over de richting van dit verschil.

6

$$t^* = 2,581 \rightarrow BI: 3,9 - 2,581 * (0,98 / \sqrt{1406}) \leq \mu \leq 3,9 + 2,581 * (0,98 / \sqrt{1406})$$

$$BI = [3,83; 3,97]$$

7

$p < 0,25$. Niet significant, noch op het 10%-, het 5%-, het 1% - of het 0,1%-significantieniveau

8

a) 90% BI: $0,02 - 0,11$

99%-BI: $-0,01 - 0,14$

b) Tweezijdige toets. $z = 0,12$; $p = 0,9044$. Geen significant verschil tussen de populatie en de steekproef fractie

9

$$z^* = 1,645 \rightarrow m = 1,645 * 0,0775 = 0,13$$

10

a) $t = 7,299 \rightarrow p < 0,0005$

Tweezijdig dus $p < 0,0010$

Dus significant verschil tussen West- en Oost-Vlamingen wat betreft het belang van geloof in hun dagelijks leven. H_0 verwerpen.

b) BI: $[0,54; 0,86]$

11*

a) 1. Hypotheses

$$H_0: p_1 = p_2$$

$$H_a: p_1 \neq p_2$$

2. Steekproevenverdeling

Standaardfout:

$$\hat{p} = (321+216)/(434+376) = 0,6630$$

$$SED = \sqrt{((0,6630*0,3370)*(1/434 + 1/376))}$$

z-verdeling: $N(0; 0,0334)$

3.Toetsingsgrootheid

$$z = (0,7396-0,5745)/0,0334 = 4,94$$

4.Overschrijdingskans

$$p(z \geq 4,96) < 0,0002 \rightarrow p < 0,0004$$

5.Conclusie: significant tot op het 0,1%-niveau

b) $BI = [0,10 ; 0,23]$

12*

95%-BI van fractie $\rightarrow z^* = 1,96$

Midden van $[0,2; 0,6] = 0,4 \rightarrow BI$ berekend adhv $0,4 \pm m$

$$m = 1,96 * \sqrt{\frac{0,4*(1-0,4)}{n}}$$

Vergelijking met 1 onbekende: $0,2 = 0,4 - 1,96 * \sqrt{\frac{0,4*(1-0,4)}{n}}$

OF: $0,6 = 0,4 + 1,96 * \sqrt{\frac{0,4*(1-0,4)}{n}}$

$$\rightarrow n = 23,05$$

Dus: steekproefomvang van 24 personen (conservatief zijn!).

13

Gemiddelde van de steekproef = 15,15 (= midden interval = $(17,8 + 12,5) / 2$)

Foutenmarge m van 90%-BI = $17,8 - 15,15 = 2,65$

t^* voor 90%-BI = 1,66

$$\rightarrow m = t * s/n \rightarrow s/n = 1,5964$$

t^* voor 95%-BI = 1,984

Foutenmarge m van 95%-BI = $1,984 * 1,5964 = 3,1673$

Nieuwe BI = $15,15 \pm 3,1673 = [11,98; 18,32]$

14

$$\hat{p} = 150 / 230 = 0,6522$$

$$83\% \text{-BI} \rightarrow z^* = 1,375$$

$$m = 0,0432$$

$$\text{BI} = 0,6522 \pm 0,0432 = [0,61; 0,70]$$

15*

$$\text{Gemiddelde van de steekproef} = 30,6 \text{ (= midden interval} = (28,5 + 32,7) / 2)$$

$$\text{Foutenmarge } m \text{ van } 80\% \text{-BI} = 32,7 - 30,6 = 2,1$$

$$t^* \text{ voor } 80\% \text{-BI} = 1,29$$

$$\rightarrow m = t^* s/n \rightarrow s/n = 1,6279$$

$$t^* \text{ voor } 96\% \text{-BI} = 2,081$$

$$\text{Foutenmarge } m \text{ van } 96\% \text{-BI} = 2,081 * 1,6279 = 3,3877$$

$$\text{Nieuwe BI} = 30,6 \pm 3,3877 = [27,21; 33,99]$$