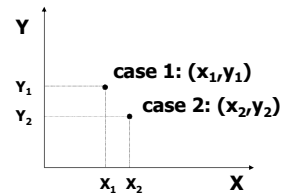


Statistiek
Deel 2. Bivariate beschrijvende statistiek

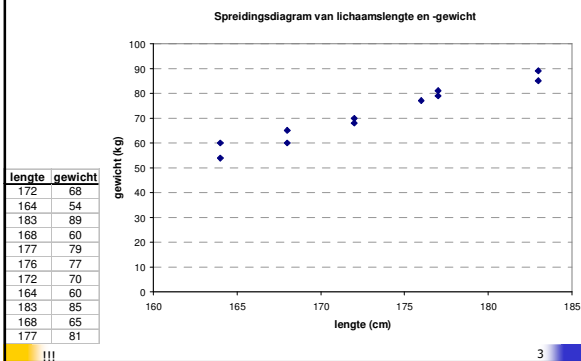
2.3. Correlatie- en regressie-analyse
2.4. Limieten van de bivariate beschrijvende statistiek: Opstap naar multivariate analyse

2.3. Correlatie- en regressie-analyse
Spreidingsdiagram (scatterplot)

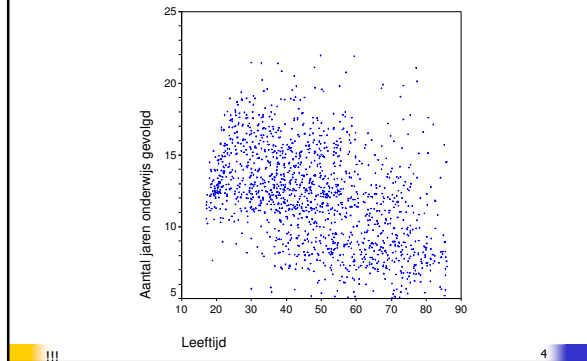
- grafische weergave van de gezamenlijke frequentieverdeling van metrische variabelen:
 - assenstelsel met waarden van X op X-as en van Y op Y-as
 - elke waarneming aanduiden als punt met coördinaten (x_i, y_i)
 - puntenwolk: vorm geeft samenhang weer



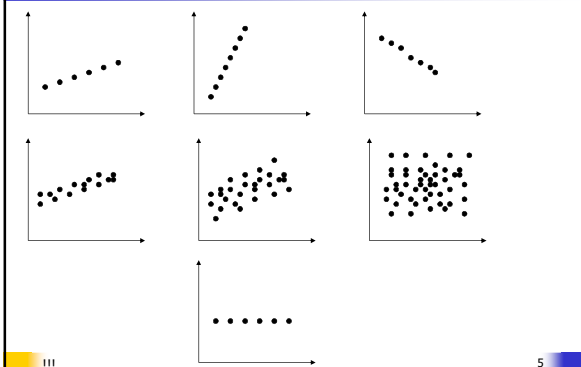
voorbeeld (fictief)



vb. Leeftijd - Aantal jaren onderwijs gevolgd



voorbeelden van puntenwolken



Regressie- en correlatieanalyse

- samenvatten puntenwolk:
 - eenvoudigste samenvatting: lijn positief - negatief ; steil - plat
→ schatten via regressie-analyse (sterkte/grootte van effect)
 - hoe goed is samenvatting ?
hoe dicht liggen punten bij lijn
→ correlatie (sterkte/grootte van samenhang)

Enkelvoudige lineaire regressie analyse

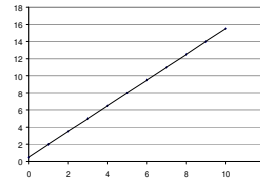
- 'enkelvoudig'
 - bivariaat: 1 onafhankelijke variabele
- 'lineair'
 - eenvoudigste samenvatting van puntenwolk: rechte
- 'regressie'
 - techniek om positie lijn te bepalen
- voor a-symmetrische relaties (effectrelaties)

2,3

7

Regressierechte

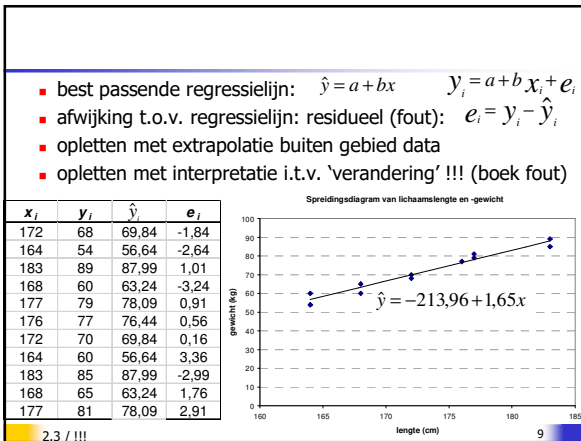
- rechte: $y = a + bx$
 - a : intercept
 - waarde Y als X=0 (snijpunt regressierechte en Y-as)
 - b : richtingscoëfficiënt (regressiecoëfficiënt)
 - helling regressierechte
 - hoeveelheid verschil in Y bij verschil 1 eenheid in X
 - richting en grootte van effect van X op Y



$$y = 0,5 + 1,5x$$

2,3

8



2,3 / III

9

Kleinste kwadratenregressie (Ordinary Least Squares)

- regressielijn best passend als goede samenvatting puntenwolk = som residuelen zo klein mogelijk
 - maar: positieve en negatieve residuelen heffen elkaar op => kwadraat nemen van residuelen
 - OLS: a en b zo bepalen dat de som van de gekwadrateerde residuelen zo klein mogelijk is
- $$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n [y_i - (a + b x_i)]^2$$
- functie minimaliseren via partiële afgeleide naar a en b en gelijk te stellen aan 0

2,3 / III

10

	(1)	(2)	(3)	(4)	(5)	(6)
	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	(3)x(4)	(3) ²
	172	68	-1,09	-3,64	3,97	1,19
	164	54	-9,09	-17,64	160,33	82,64
	183	89	9,91	17,36	172,06	98,19
	168	60	-5,09	-11,64	59,24	25,92
	177	79	3,91	7,36	28,79	15,28
	176	77	2,91	5,36	15,60	8,46
	172	70	-1,09	-1,64	1,79	1,19
	164	60	-9,09	-11,64	105,79	82,64
	183	85	9,91	13,36	132,42	98,19
	168	65	-5,09	-6,64	33,79	25,92
	177	81	3,91	9,36	36,60	15,28
rek.gem.:	173,09	71,636			som: 750,36	454,91

oplossing:

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{750,36}{454,91} = 1,65$$

$$a = \bar{y} - b\bar{x} = 71,64 - (1,65 \times 173,09) = -213,96$$

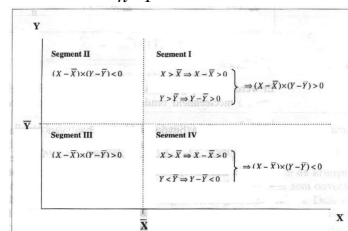
III

11

Covariantie

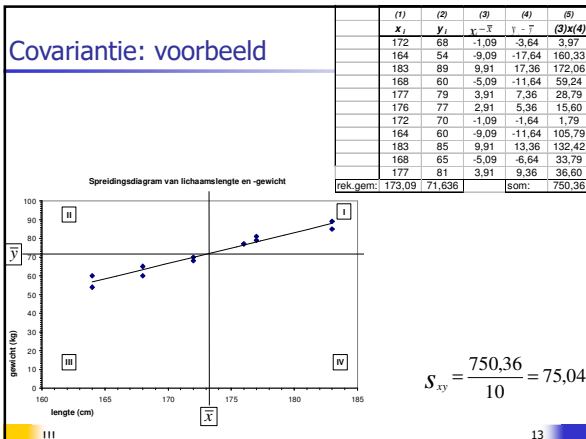
- drukt uit in welke mate 2 metrische variabelen samen variëren (co-variëren)

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



III

12



- interpretatie:
 - symmetrische maat
 - teken: positieve vs. negatieve relatie
 - nul: geen samenhang
 - absolute waarde: geen vaste interpretatie van sterkte samenhang <= afhankelijk van meeteenheid variabelen
 - oplossing: correlatiecoëfficiënt
 - = covariantie berekend op gestandaardiseerde variabelen (omgerekend naar z-scores)
- $$Z_{x_i} = \frac{x_i - \bar{x}}{s}$$

Correlatiecoëfficiënt

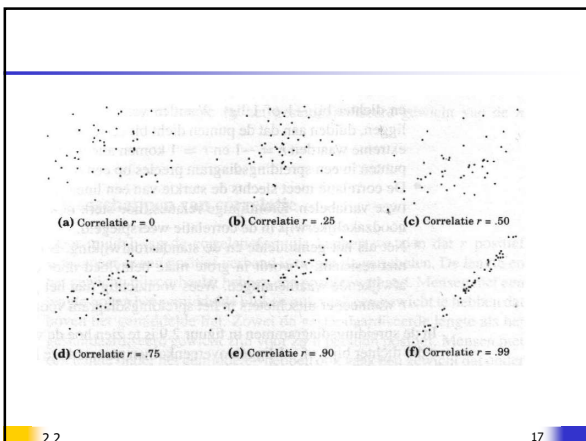
- covariantie berekend op z-getransformeerde variabelen (vaste schaal X en Y + gelijke spreiding)

$$\frac{\sum_{i=1}^n Z_x Z_y}{n-1} = \frac{\sum_{i=1}^n \frac{(x_i - \bar{x})}{s_x} \frac{(y_i - \bar{y})}{s_y}}{n-1} = \frac{1}{n-1} \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1} \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{n-1}}}$$

$$= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = r_{xy}$$

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^n \frac{(x_i - \bar{x})}{s_x} \frac{(y_i - \bar{y})}{s_y} = \frac{s_{xy}}{s_x s_y}$$

- interpretatie:
 - symmetrische maat van lineaire samenhang
 - teken: positieve vs. negatieve relatie
 - nul: geen samenhang
 - [-1,+1]
 - hoe verder van 0, hoe dichter punten bij best passende regressierechte
 - in vb.: $r_{xy} = \frac{75,04}{6,75 \times 1,37} = 0,98$
- alternatieve berekeningen b: $b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2} = r_{xy} \frac{s_y}{s_x}$

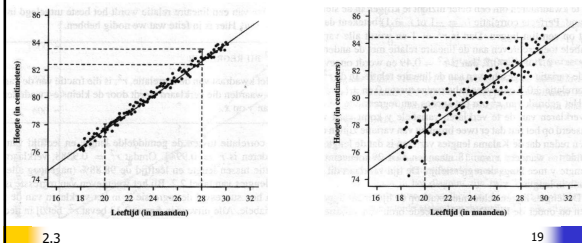


- ### B: Gestandaardiseerde b
- b schaalgevoelig: bij wijziging van meeteenheid, andere waarde
 - vb.: $\hat{y} = -213,96 + 1,65x$
 - als $x' = x/100$: $\hat{y} = -213,96 + 164,95x'$ (lengte in m i.p.v. cm)
 - oplossing: B: richtingscoëfficiënt berekend na z-transformatie X,Y
 - schalonaafhankelijk
 - interpretatie: bij verschil van 1 standaardafwijking in X, hoeveel standaardafwijkingen verschil verwacht in Y
 - in bivaariaat geval = r
- $$B = r_{xy} \frac{s_y}{s_x} = r_{xy}$$

Determinatiecoëfficiënt: r^2

- r^2 = proportie van de variantie in Y die verklaard wordt door X
- in vb.: $(0,98)^2 = 0,9604$
 → lengte verklaart 96,04% van variatie in gewicht

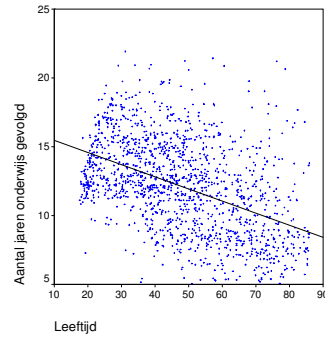
$$r^2_{xy} = \frac{S_{xy}^2}{S_x^2 S_y^2}$$



2,3

19

leeftijd -> aantal jaren onderwijs (APS2002)



$$\hat{Y} = 16,29 - 0,087X$$

$$a = 16,29$$

$$b = -0,087$$

$$r = -0,46$$

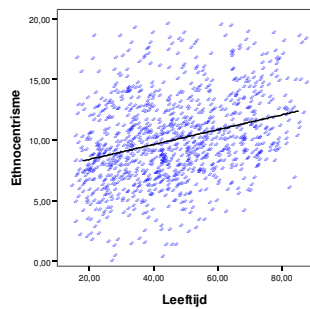
$$r^2 = 0,21$$

$$B = -0,46$$

III

20

leeftijd -> ethnocentrisme (APS 2002)



$$\hat{Y} = 7,17 + 0,061X$$

$$a = 7,17$$

$$b = 0,061$$

$$r = 0,30$$

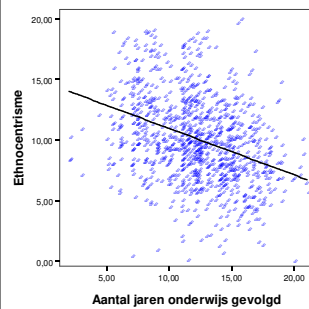
$$r^2 = 0,09$$

$$B = 0,30$$

III

21

jaren onderwijs -> ethnocentrisme (APS2002)



$$\hat{Y} = 14,7 - 0,382X$$

$$a = 14,7$$

$$b = -0,382$$

$$r = -0,37$$

$$r^2 = 0,13$$

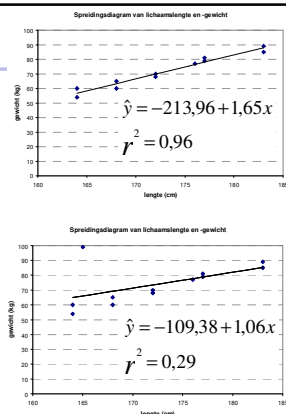
$$B = -0,37$$

III

22

Outliers

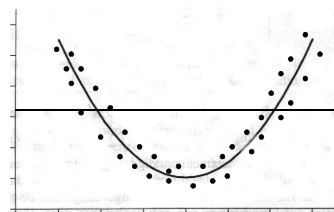
- uitschieters:
sterke invloed op regressie-oplossing (trekken regressielijn naar zich toe)



2,4 / III

23

Non-lineariteit



$$b = 0$$

$$r = 0$$

2,4 / III

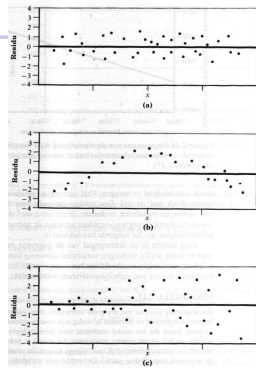
24

Residuen-analyse

- indien regressielijn goede samenvatting data: geen systematisch patroon in residuen (a)

(b) non-lineariteit

(c) heteroscedasticiteit



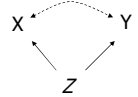
2.4

25

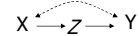
2.4. Opstap naar multivariate analyse

- bivariate effectrelatie \neq causale relatie

- beschrijvende vs. inductieve statistiek
- cross-sectionele vs. longitudinale data
- verborgen variabelen (Z) / multivariate analyse types van relaties met meer dan 2 variabelen:
 - gemeenschappelijke afhankelijkheid / wederzijdse oorzaak



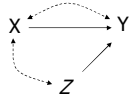
- intermediaire variabele = mediator variabele (mediatie)



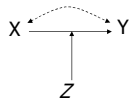
2.5 / III

26

- verstrengeling



- interactie-effect = moderatie-effect



→ multivariate statistiek: netto-effecten

2.5 / III

27

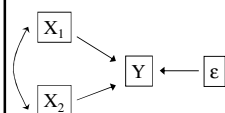
- maar zelfs met correcte toepassing MVA-technieken, blijft:

- cross-sectionele vs. longitudinale data
- mogelijkheid verborgen variabelen => vaststellen effectrelatie \neq causale relatie
- ev. wel sterke indicatie voor causale relatie, indien:
 - sterke samenhang
 - statistische controle voor diverse andere kenmerken (MVA)
 - samenhang wordt consistent gevonden in vele onderzoeken
 - vermeende oorzaak gaat vooraf aan het effect

2.5 / III

28

Het multivariate model met 2 onafhankelijke variabelen



- Y: afhankelijke variabele
- X_1 en X_2 : onafhankelijke variabelen
- a: intercept (verwachte Y als X_1 en $X_2 = \text{nul}$)
- b_1 : netto-effect van X_1 op Y (met X_2 constant gehouden) (statistisch gecontroleerd voor X_2)
- b_2 : netto-effect van X_2 op Y (met X_1 constant gehouden) (statistisch gecontroleerd voor X_1)
- e: cumulatief effect alle andere oorzaken Y

$$\hat{Y} = a + b_1 X_1 + b_2 X_2$$

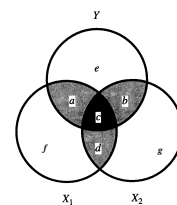
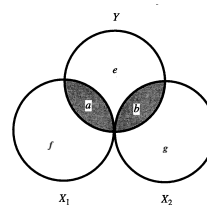
$$Y = a + b_1 X_1 + b_2 X_2 + e$$

$$Z_i = B_1 Z_{X_1} + B_2 Z_{X_2} + e$$

III

29

Concept van gedeelde variantie



III

30

een illustratie

- 2 afzonderlijke bivariate regressie-analyses:

$$ETNOCENTRISME = 14,7 - 0,382 OPLEIDING + e$$

$$(r^2 = 0,13 = a + c)$$

$$ETNOCENTRISME = 7,17 + 0,061 LEEFTIJD + e$$

$$(r^2 = 0,09 = b + c)$$

- maar: $r_{LEEFTIJD-OPLEIDING} = -0,46$

$$r^2_{LEEFTIJD-OPLEIDING} = 0,22 (= c + d)$$

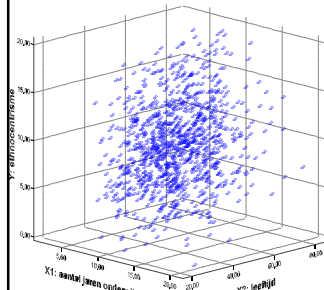
- multivariate analyse:

$$ETNOCENTRISME = 14,7 - 0,305 OPLEIDING + 0,031 LEEFTIJD + e$$

$$(R^2 = 0,15 = a + b + c)$$

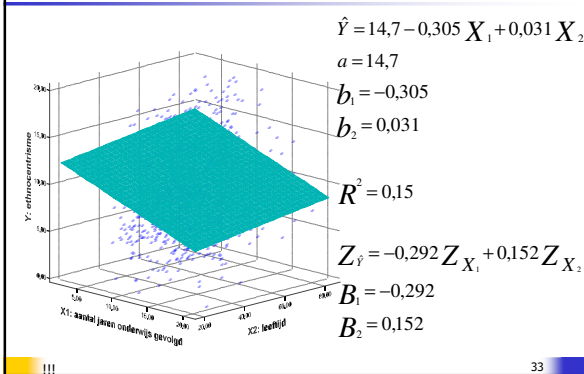
31

jaren onderwijs (X₁) + leeftijd (X₂)
-> etnocentrisme (Y) (APS2002)



32

jaren onderwijs (X₁) + leeftijd (X₂)
-> etnocentrisme (Y) (APS2002)



$$\hat{Y} = 14,7 - 0,305 X_1 + 0,031 X_2$$

$$a = 14,7$$

$$b_1 = -0,305$$

$$b_2 = 0,031$$

$$R^2 = 0,15$$

$$Z_{\hat{Y}} = -0,292 Z_{X_1} + 0,152 Z_{X_2}$$

$$B_1 = -0,292$$

$$B_2 = 0,152$$

33

Uitbreiding naar meer onafhankelijke variabelen

- eenvoudige uitbreiding naar drie en meer onafhankelijke variabelen:

$$\hat{Y} = a + b_1 X_1 + b_2 X_2 + b_3 X_3$$

$$\hat{Y} = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + \dots$$

34

een illustratie:
perceptie van bedreiging door migranten (Y)

- Regressie-analyse van perceptie van bedreiging door migranten - gestandaardiseerde regressie-coëfficiënten

	B
Leeftijd	0,034
Opleiding	-0,318
Lidmaatschap verenigingen	-0,067
Vertrouwen politieke instellingen	-0,184
Vertrouwen administratie	-0,074
Vertrouwen gerecht	-0,129
Vertrouwen leger	0,089
R²	0,304

(+ ook controle voor geslacht en levensbeschouwing)
(bron: Meuleman B. & J. Billiet (2005) De evolutie van etnische dreiging tussen 1991 en 2004 en de relatie met institutioneel vertrouwen. in: Vlaanderen Gepeild 2005, pp.37-60).

35

een (fictieve) illustratie van een stapsgewijs opgebouwd model

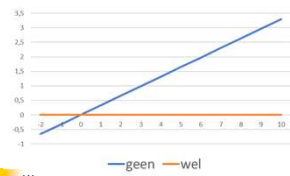
- Y: vooroordelen bij leerkrachten t.a.v. leerlingen met een migratieachtergrond (hoe hoger, hoe meer vooroordeel)

	Model 1	Model 2	Model 3	Model 4	Model 5
	B	B	B	B	B
Leeftijd	0,150	0,150	0,150	0,000	0,000
Urbanisatiegraad woonplaats		0,188	0,009	0,001	0,001
Aantal vrienden met migratieachtergrond			-0,085	-0,085	-0,085
Aantal jaren ervaring				0,240	0,240
Zelf migratieachtergrond					-0,9
R²	0,07	0,15	0,18	0,25	0,35

36

(fictieve) illustratie van interactie-effect

	Model 6
(...idem als vorige...)	B
Aantal jaren ervaring * zelf <u>geen</u> migratieachtergrond	0,330
Aantal jaren ervaring * zelf <u>wel</u> migratieachtergrond	0
R ²	0,39



III

37

STATISTIEK

OEFENBOEK

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

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.

VOORWOORD

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 Ufora, 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. Deze oefeningen kan je ook terugvinden op Ufora, voorzien van uitgebreide elektronische feedback. De C*-oefeningen dien je telkens verplicht zelfstandig na de oefenles te maken.

Veel succes!

INHOUDSOPGAVE

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
2.1	Studie van samenhangen en effecten	9
2.2	Relaties tussen categorische variabelen	10
2.3	Relaties tussen metrische variabelen	12
2.4	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: VOORBEREIDENDE 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
2.1	Studie van samenhangen en effecten	33
2.2	Relaties tussen categorische variabelen	33
2.3	Relaties tussen metrische variabelen	36
2.4	De limieten van de beschrijvende statistiek	40
3	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.....	54
1.3	Maten van positie	56
1.4	Maten van spreiding	60
1.5	Vorm van een verdeling	65
1.6	Transformaties van een variabele	68
1.7	Dichtheidskrommen	68
2	Bivariate beschrijvende statistiek.....	72
2.1	Studie van samenhangen en effecten	72
2.2	Relaties tussen categorische variabelen	72
2.3	Relaties tussen metrische variabelen	77
2.4	De limieten van de beschrijvende statistiek	88

3	Inductieve statistiek	92
3.1	De studie van de systematiek van het toeval: kansrekenen	92
3.2	Stochastische variabelen	92
3.3	Studie van steekproefvariabiliteit	92
3.4	Basistools van de inductieve statistiek	92
3.5	Inferentie voor verwachtingen	92
3.6	Inferentie voor fracties	98
3.7	Inferentie voor kruistabellen	101
3.8	Overzichtsoefeningen	102

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

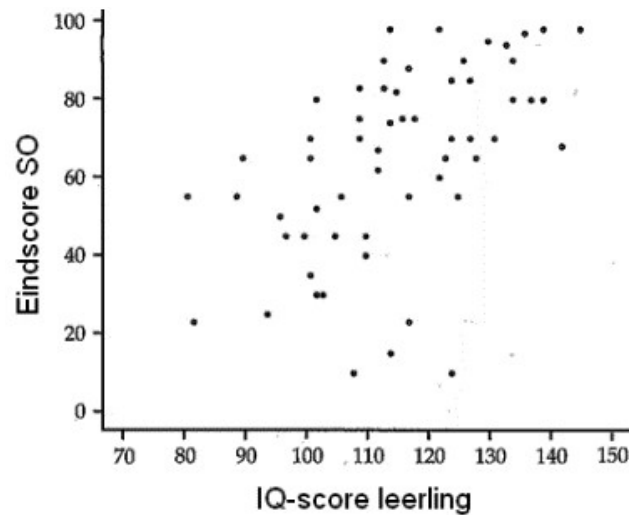
Land	Democratisch karakter van de staatsinrichting	Sociale gelijkheid
Lapland	Hoog	Hoog
Cyprus	Laag	Hoog
Ecuador	Laag	Laag
Columbia	Laag	Laag
Verenigde Emiraten	Hoog	Hoog
Guatemala	Laag	Laag
Noorwegen	Hoog	Hoog
Nederland	Hoog	Hoog
Paraguay	Laag	Hoog
Finland	Hoog	Laag
Trinidad	Laag	Laag
Japan	Laag	Laag
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

2.3 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 vakantieliefde 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 lijstte 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 vakantieliefde net 3 inches langer is dan de vrouw in kwestie, hoe ziet de correlatie er dan uit?
- Stel zowel de ongestandaardiseerde als de gestandaardiseerde regressievergelijking op
- Een vriendin komt wat later. Ze is 170,2 cm groot. Wat is de verwachte lengte van haar vakantieliefde 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

- a) Teken de scatterplot. Er is één outlier. Welke?
- b) 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?
- a. $\hat{y} = 20,49 + 0,754x$
- b. $\hat{y} = 50,01 + 0,410x$
- c) 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

2.4 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	0,000
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 1ste model.
- Welke variabele heeft in het 1ste model het grootste effect op schoolbetrokkenheid? En in het 2de en 3de model?
- Als je de verbondenheid met leraren toevoegt in het 2de model, wat gebeurt dan met de variabelen van het 1ste model? Interpreteer en teken telkens een bijhorend pijltjesdiagram.
- Als je verbondenheid met ouders toevoegt in het 3de model, wat gebeurt dan met de variabelen van het 2de 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.

4

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?

5

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?

6

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?