

6.4 Totaalbeeld

Samenvatten

Je hebt nu alle theorie van het onderwerp ‘Verbanden en verschillen’ doorgewerkt. Het is nu tijd om een overzicht over het geheel te krijgen.

Begrippenlijst

- verband tussen variabelen — correlatiecoëfficiënt — trendlijn/regressielijn
- verschil van kwalitatieve variabelen — cumulatief verschilpercentage — kruistabel en phi
- verschil van kwantitatieve variabelen — boxplots vergelijken — effectgrootte — verschiltoets van het gemiddelde

Activiteitenlijst

- statistisch verband tussen twee variabelen herkennen — verschil statistisch verband en oorzakelijk verband herkennen
- twee methoden om het verschil tussen kwalitatieve variabelen in een getal uit te drukken toepassen — vuistregels gebruiken
- boxplots vergelijken met behulp van vuistregels — een tekentoets uitvoeren — een verschiltoets van het gemiddelde uitvoeren

Achtergronden

Je hebt diverse manieren gezien om statistische variabelen te vergelijken of juist naar een statistisch verband tussen twee variabelen te zoeken. Wie hebben al die methoden toch bedacht?

- De *correlatiecoëfficiënt* r voor het bepalen of er tussen twee kwantitatieve variabelen een (statistisch) verband bestaat, is ook bedacht door **Karl Pearson (1857–1936)**, hoewel de trendlijn zelf al ouder was en door **Carl Friedrich Gauss (1777–1855)** met de kleinste-kwadraten-methode is afgeleid.
- De *phi* voor het bepalen of in een twee bij twee kruistabel de twee kwalitatieve variabelen weinig, een beetje of veel van elkaar verschillen, is bedacht door **Karl Pearson**, één van de grondleggers van veel statistische methoden. In feite hangt dit getal nauw samen met de correlatiecoëfficiënt r , in dit geval voor twee punten in een assenstelsel.
- De *boxplot* wordt gebruikt om twee kwantitatieve variabelen te vergelijken. Hij is voor het eerst geïntroduceerd door de Amerikaanse wiskundige **John Tukey (1915–2000)**. Tukey was één van de grondleggers van de data-analyse.
- Het *hypothese toetsen* wordt ook wel gebruikt om twee kwantitatieve variabelen te vergelijken (o.a. tekentoets en verschiltoets van het gemiddelde). De begrippen nulhypothese en significantieniveau kregen vooral betekenis door de Britse bioloog en statisticus **Ronald Aylmer Fisher (1890–1962)**. Hij schreef het invloedrijke boek ‘Statistical methods for research workers’ in 1925.



Figuur 1 Karl Pearson

Testen

Opgave 1

Van een groep bovenbouwleerlingen met wiskunde A/C in hun profiel is vastgelegd in hoeveel vreemde talen zij eindexamenen doen. Dit is ook gebeurd voor een groep bovenbouwleerlingen met wiskunde B in hun profiel. Bekijk de tabel met de gegevens.

aantal vreemde talen	wiskunde A/C	wiskunde B
één	7	16
twee	26	58
drie	10	30
vier	0	7
	43	111

Tabel 1

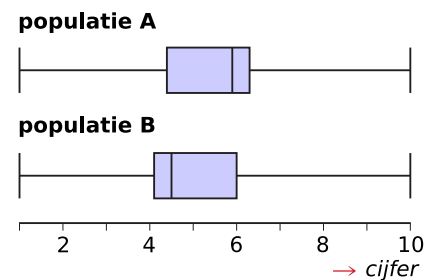
- Leg uit waarom je het verschil in aantal gekozen vreemde talen tussen de groepen met wiskunde A/C en wiskunde B het best kunt onderzoeken met behulp van $\max V_{CP}$.
- Laat zien dat voor de gegevens in deze tabel geldt: $\max V_{CP} = 10\%$.
- Van welke mate van verschil is er sprake volgens de vuistregels voor $\max V_{CP}$?

(bron: cTwo)

Opgave 2

Bekijk de boxplots van de gegevens van twee populaties.

- Om welk type gegevens gaat het hier?
- Geef van elk van de volgende beweringen aan of de bewering, de conclusie of beide wel of niet juist zijn en beargumenteer waarom dat volgens jou zo is.
 - De boxplots zijn even breed dus de beide populaties zijn gelijk.
 - De boxen van beide boxplots zijn even breed dus beide populaties zijn gelijk.
 - De boxen van beide boxplots verschillen van opbouw: beide populaties verschillen dus veel van elkaar.



Figuur 2

Opgave 3

In 1947 hielden de wiskundigen Freudenthal en Sittig een statistisch onderzoek ten behoeve van een nieuw maatsysteem voor vrouwenkleding in opdracht van het warenhuis De Bijenkorf. Zij lieten daarbij een grote verscheidenheid aan lichaamsmaten opmeten van 5001 vrouwen. Zij vonden onder andere een sterke correlatie tussen de taille (de omtrek van het lichaam gemeten ter hoogte van de navel) en de bovenwijdte (de omtrek van het lichaam gemeten ter hoogte van de borst). De gevonden correlatiecoëfficiënt bedroeg ongeveer 0,9058.

In hun rapport 'De juiste maat' geven zij voor de bovenwijdte b een gemiddelde van 97,99 cm met een standaardafwijking van 10,12 cm en voor de taille t een gemiddelde van 80,45 cm met een standaardafwijking van 10,80 cm.

- Welke betekenis heeft deze hoge correlatie tussen b en t voor een spreidingsdiagram van deze twee variabelen?
- Stel een vergelijking op van de regressielijn voor b .
- Geef een statistisch verantwoorde schatting van de bovenwijdte van een vrouw met een taille van 90 cm.

Opgave 4

Bekijk de tabel met de (geschatte) woordenschat die eentalige Nederlandse kinderen hebben.

Wat mag je, op basis van de gegevens, uit deze tabel concluderen omtrent het verband tussen de leeftijd en de woordenschat van een eentalig Nederlands kind?

(bron: www.cbsdeakker.nl)

leeftijd	woordenschat
4	3000
5	3800
6	4500
7	5200
8	6000
9	8500
10	11000
11	14000
12	17000

Tabel 2

Opgave 5

Cito meet kennis en vaardigheden in door middel van van toetsen. Cito heeft onderzocht of er een verschil in vaardigheden is tussen vwo- en gymnasiumleerlingen. Cito heeft dit gedaan voor de onderwerpen Nederlands leesvaardigheid, Nederlands woordenschat, Engels leesvaardigheid, rekenen, wiskunde en taalverzorging. Bekijk de tabel.

leerling	vak	aantal leerlingen	aantal scholen	gemiddelde	standaard-afwijking	effect-grootte
gym	nlv	1654	87	284,24	15,77	0,55
gym	nws	1598	84	296,71	22,15	0,58
gym	elv	1605	84	297,53	29,65	0,48
gym	rek	1873	87	293,12	24,45	0,6
gym	wis	1872	87	234,93	13,77	0,62
gym	tv	1566	84	302,10	29,53	1,16
vwo	nlv	7400	170	275,68	15,62	
vwo	nws	6961	165	284,53	20,61	
vwo	elv	6955	160	284,32	26,80	
vwo	rek	7148	163	279,31	22,74	
vwo	wis	7145	163	227,15	12,08	
vwo	tv	6978	165	275,32	21,49	

Figuur 3

- Schrijf de statistische variabelen uit dit onderzoek op en geef aan of ze kwalitatief of kwantitatief zijn.
- In de laatste kolom staat de effectgrootte. Bereken zelf de effectgrootte voor Nederlands leesvaardigheid.
- Schrijf de conclusie op voor dit verschil.
- Controleer de effectgrootte voor taalverzorging. Wat valt je op? Verandert de conclusie?

Opgave 6

Patiënten die voor een behandeling enige tijd in een ziekenhuis worden opgenomen, lopen tijdens dit verblijf het risico een infectie te krijgen. Zo'n infectie wordt een zorginfectie genoemd. Een deel van de zorginfecties ontstaat na een operatie.

In de periode 2007 tot en met 2012 is een steekproef gehouden onder een deel van de Nederlandse ziekenhuizen. Bekijk de tabel met enkele resultaten hiervan.

	aantal
patiënten	95299
patiënten die een zorginfectie hebben opgelopen	4694
geopereerde patiënten	32664
geopereerde patiënten die een zorginfectie hebben opgelopen	1286

Tabel 3

Je mag aannemen dat de patiënten in deze ziekenhuizen representatief zijn voor alle patiënten die in een Nederlands ziekenhuis worden opgenomen.

Onderzoek in hoeverre er verschil is in het krijgen van een zorginfectie tussen geopereerde en niet-geopereerde patiënten.

(bron: pilotexamen 2016 - II havo A)

Opgave 7

Bekijk de resultaten van 19 leerlingen voor het SE en het CE.

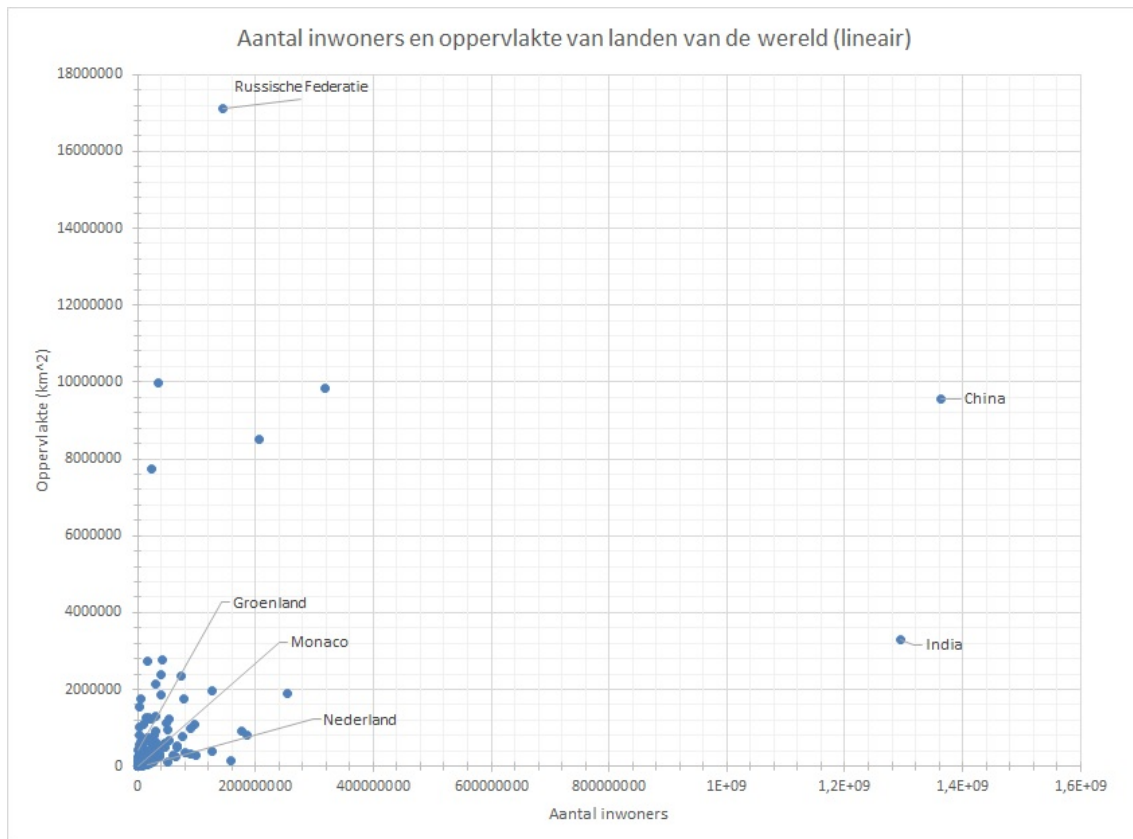
leerling	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
SE-cijfer	6,0	6,7	5,8	7,1	5,4	6,5	8,8	6,9	7,9	5,1	6,1	6,1	6,4	7,4	5,9	6,2	7,1	6,8	6,3
CE-cijfer	6,4	6,3	5,2	6,5	5,4	6,1	9,0	6,8	7,5	5,6	6,0	6,5	6,0	6,5	6,0	6,6	7,0	6,6	6,4

Tabel 4

Neem aan dat de gemiddelden van de steekproeven normaal verdeeld zijn. Ga na of uit een verschiltoets met een significantieniveau van 5% volgt dat het SE beter is gemaakt dan het CE.

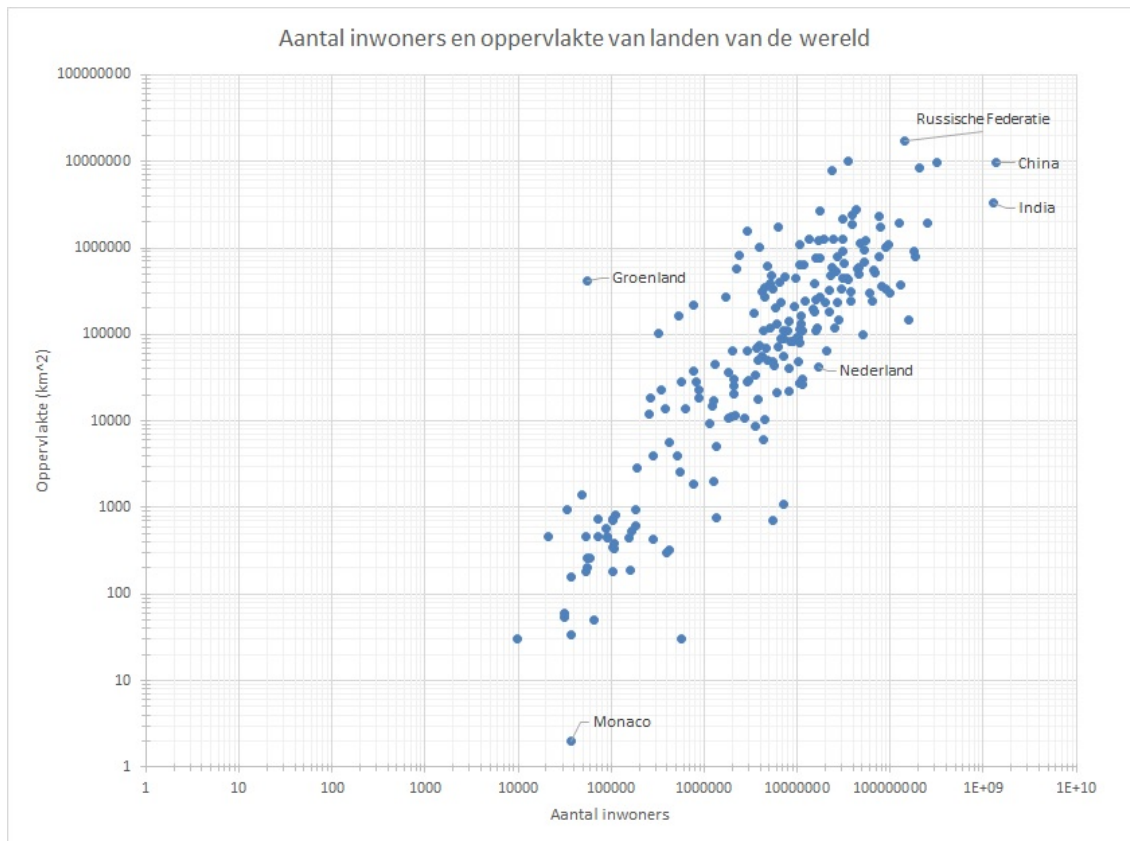
Toepassen

Opgave 8: Trendlijn op logaritmisch papier



Figuur 4

In de figuur hierboven is voor een groot aantal landen het aantal inwoners en de oppervlakte weer-gegeven. Het is duidelijk dat aflezen hier erg moeilijk is. Een manier om dit te verbeteren is het gebruiken van logaritmische schalen. Dat zie je in de volgende figuur.



Figuur 5

- a** Het land dat de meeste inwoners telt, is China.
Lees af hoeveel inwoners China heeft. Geef je antwoord in miljarden, afgerond op één decimaal.
- b** Lees af hoe groot het oppervlak van China is. Geef je antwoord in miljoenen, afgerond op één decimaal.
- c** Het land met de grootste oppervlakte is Rusland.
Bereken het aantal inwoners per km^2 voor China en Rusland.
In de figuur is ook de trendlijn getekend.
De vergelijking van de trendlijn is: $A = \frac{1}{10000} \cdot P^{\frac{4}{3}}$,
waarin A de oppervlakte in km^2 en P de bevolkingsomvang is.
- d** Bereken met deze formule het aantal inwoners van een land van 200000 km^2 .
Geef je antwoord in miljoenen, afgerond op één decimaal.



© 2024

Deze paragraaf is een onderdeel van het Math4All wiskundemateriaal.

Math4All stelt het op prijs als onvolkomenheden in het materiaal worden gemeld en ideeën voor verbeteringen in de content of dienstverlening kenbaar worden gemaakt. Klik op  in de marge bij de betreffende opgave. Uw mailprogramma wordt dan geopend waarbij het emailadres en onderwerp al zijn ingevuld. U hoeft alleen uw opmerkingen nog maar in te voeren.

Email: f.spijkers@math4all.nl

Met de Math4All Foliostaat kunnen complete readers worden samengesteld en toetsen worden gegenereerd. Docenten kunnen bij a.f.otten@math4all.nl een gratis inlog voor de maatwerkdienst aanvragen.
