

Basisvaardigheden

Toegepaste statistiek



Noordhoff

Gert-Jan Reus & Hans van Buuren

3^e druk

Basisvaardigheden Toegepaste Statistiek HO



Basis- vaardigheden Toegepaste Statistiek HO

Gert-Jan Reus

Hans van Buuren

Derde druk

Noordhoff Groningen

Ontwerp omslag: Shootmedia

Omslagillustratie: Shootmedia

Eventuele op- en aanmerkingen over deze of andere uitgaven kunt u richten aan: Noordhoff Uitgevers bv, Afdeling Hoger Onderwijs, Antwoordnummer 13, 9700 VB Groningen of via het contactformulier op www.mijnnoordhoff.nl.

De informatie in deze uitgave is uitsluitend bedoeld als algemene informatie. Aan deze informatie kunt u geen rechten of aansprakelijkheid van de auteur(s), redactie of uitgever ontleen.

0 / 20



© 2020 Noordhoff Uitgevers bv, Groningen/Utrecht, Nederland.

Deze uitgave is beschermd op grond van het auteursrecht. Wanneer u (her)gebruik wilt maken van de informatie in deze uitgave, dient u vooraf schriftelijke toestemming te verkrijgen van Noordhoff Uitgevers bv. Meer informatie over collectieve regelingen voor het onderwijs is te vinden op www.onderwijsauteursrecht.nl.

This publication is protected by copyright. Prior written permission of Noordhoff Uitgevers bv is required to (re)use the information in this publication.

ISBN (ebook) 978-90-01-75356-6

ISBN 978-90-01-57528-1

NUR 916

Voorwoord

Regelmatig wordt een beroep gedaan op statistiek. Afhankelijk van je vooropleiding en eigen interesse maakt dit de uitdaging om zelf inzicht te krijgen in statistische wetmatigheden groter.

Met dit boek kun je zelfstandig je statistische vaardigheden op het gewenste peil brengen.

Bij de keuze van de onderwerpen is uitgegaan van de behoeften van de doelgroep: eerstejaarsstudenten aan opleidingen in het hoger onderwijs. De structuur is steeds hetzelfde: op de linkerpagina de theoretische achtergrond met een voorbeeld en op de rechterpagina opgaven om de kennis zelfstandig te toetsen. Achter in het boek staan de antwoorden op de opgaven.

Behalve het boek zorgt ook de website www.basisvaardigheidentoegepastestatistiek.noordhoff.nl

voor de nodige ondersteuning. Op de website staan de uitgebreide uitwerkingen van de opgaven. Je kunt zo controleren of de door jou gekozen aanpak de juiste is geweest. Maak hier gebruik van, de antwoorden achter in het boek zijn ‘slechts’ indicatief.



Op de website vind je een instaptoets. Met deze instaptoets krijg je vooraf een beeld van je eigen kwaliteiten. De toets heeft dezelfde hoofdstukopbouw als het boek. Aan het eind van de toets krijg je per hoofdstuk een overzicht van je prestaties. Zo kun je snel zien welke onderdelen je in ieder geval goed moet bestuderen.

Je kunt verder oefenen met de extra opgaven op de website. Deze hebben hetzelfde niveau als de opgaven in het boek. Je komt direct te weten of je de opgave goed hebt gemaakt. Soms krijg je informatie over de manier waarop je de opgave had kunnen aanpakken.

Een woord van dank gaat nog uit naar drs. Eva Lorbach, docent en onderzoeker aan de faculteit der Maatschappij- en gedragswetenschappen, afdeling Psychologie Universiteit van Amsterdam, vanwege haar

rol als sparringpartner aan het begin van dit traject. Dr. Esther Bakker en dr. Tilly Houtmans, universitair docenten aan de Open Universiteit Nederland, hebben nauwgezet onze eerste teksten bekeken en nagerekend.

Ondanks de zorg en aandacht kunnen er foutjes in de tekst geslopen zijn, statistisch gezien kan het ook bijna niet anders.

Bij deze derde druk is een aantal opgaven aangepast. Zodoende zijn er meer opgaven buiten de economische en sociale wetenschapscontext opgenomen. Daarnaast is op verzoek de paragraaf over centrummaten bij gegroepeerde gegevens geheel aangepast.

Gert-Jan Reus

Coauteur van *Basisvaardigheden Toegepast Rekenen*

Gert-Jan Reus is hogeschooldocent bij Fontys Hogescholen.

Hans van Buuren

Coauteur van onder andere *Statistiek voor de sociale wetenschappen, Onderzoek, de Basis en Onderzoek in de Gezondheidszorg*.

Hans van Buuren is voormalig universitair hoofddocent onderzoekscompetenties bij de faculteit Psychologie van de Open Universiteit Nederland.

Inhoud

1 Inleiding statistiek

- 1.1 Statistiek 10
- 1.2 Variabelen en meetniveaus 12
- 1.3 Discrete en continue variabelen 14
- 1.4 Onafhankelijke en afhankelijke variabelen 16
- 1.5 Ordenen van gegevens 18

2 Gegevens presenteren

- 2.1 Tabellen 20
- 2.2 Steel-en-bladdiagram 22
- 2.3 Lijndiagram 24
- 2.4 Staafdiagram, stapeldiagram en cirkeldiagram 26
- 2.5 Histogram 28
- 2.6 Frequentiepolygoon 30
- 2.7 Rapporteren van onderzoek 32

3 Univariaten - Centrummaten

- 3.1 Modus bij ongegroepeerde gegevens 34
- 3.2 Mediaan bij ongegroepeerde gegevens 36
- 3.3 Gemiddelde bij ongegroepeerde gegevens 38
- 3.4 Centrummaten bij gegroepeerde gegevens 40

4 Univariaten - Spreidingsmaten

- 4.1 Variantie bij ongegroepeerde gegevens 42
- 4.2 Standaarddeviatie en range bij ongegroepeerde gegevens 44
- 4.3 Interkwartielafstand bij ongegroepeerde gegevens 46
- 4.4 Spreidingsmaten bij gegroepeerde gegevens 48
- 4.5 Boxplot 50

5 Kansen

- 5.1 Volgordeproblemen 52
- 5.2 Kansdefinities 54
- 5.3 Verwachtingswaarde 56
- 5.4 Rekenregels voor kansen 58
- 5.5 Experiment met en zonder teruglegging 60
- 5.6 Vrijheidsgraden 62

6 Kansverdelingen

- 6.1 Normale verdeling 64
- 6.2 Standaardnormale verdeling 66
- 6.3 Overschrijdingskans 68
- 6.4 Binomiale verdeling 70
- 6.5 Poisson-verdeling 72

7 Schatten

- 7.1 Populatie en steekproef 74
- 7.2 Statistisch schatten 76
- 7.3 Punt-schatten van het gemiddelde 78
- 7.4 Punt-schatten van de variantie en proportie 80
- 7.5 Punt-schatten bij twee populaties 82
- 7.6 Intervalschatten 84
- 7.7 Omvang van de steekproef 86

8 Toetsen

- 8.1 Toetsen 88
- 8.2 Toetsprocedure – algemeen 90
- 8.3 Toetsingsprocedure (1) – Overschrijdingskans 92
- 8.4 Toetsingsprocedure (2) – Kritieke grenswaarde 94
- 8.5 Toetsingsprocedure (3) – Betrouwbaarheidsinterval 96

9 Univariaten toetsen

- 9.1 Hypergeometrische toets van een aantal 98
- 9.2 Binomiaaltoets 100
- 9.3 Binomiaaltoets benaderd door normale verdeling 102
- 9.4 Chikwadraat: goodness-of-fittoets voor een verdeling 104
- 9.5 Z-toets voor een gemiddelde 106
- 9.6 T-toets voor een gemiddelde 108

10 Associatie

- 10.1 Samenhang en causaal verband 110
- 10.2 Percenteren en vergelijken 112
- 10.3 Statistische onafhankelijkheid 114
- 10.4 Chikwadraat (χ^2) en statistische onafhankelijkheid 116

11 Samenhang toetsen

- 11.1 Associatiemaat Cramér's V 118
- 11.2 Chikwadraat (χ^2) - associatietoets 120
- 11.3 Samenhang tussen twee dichotome variabelen 122
- 11.4 Toets 2×2-tabel: Chikwadraat (χ^2)-toets 124
- 11.5 Toets 2×2-tabel: Fisher-Exact-toets 126
- 11.6 Cohen's kappa (κ) 128

12 Correlatie

- 12.1 Pearson's correlatiecoëfficiënt ρ of r 130
- 12.2 Spearman's rangcorrelatiecoëfficiënt ρ_s (ρ_{rs}) of (r_s) 132
- 12.3 Kendall's rangcorrelatiecoëfficiënt τ (tau) 134
- 12.4 Lineaire regressie van Y op X : $\hat{Y} = a + bX$ 136

13 Gepaarde waarnemingen

- 13.1 Odds ratio 138
- 13.2 Relatief risico 140
- 13.3 Tekentoets 142
- 13.4 De t-toets voor gepaarde waarnemingen 144
- 13.5 Verschiltoets voor fracties 146

14 Tijdreeksen

- 14.1 Indexcijfer 148
- 14.2 Tijdreeks, trendwaarde en voortschrijdende gemiddelden 150
- 14.3 Trend- en seizoencomponent 152
- 14.4 Multiplicatief model 154

Antwoorden 156

Bijlagen

- Tabel 1 Standaardnormale verdeling 172
- Tabel 2 Cumulatieve binomiale verdeling 174
- Tabel 3 Poisson-verdeling 177
- Tabel 4 Cumulatieve Poisson-verdeling 179
- Tabel 5 t -verdeling 181
- Tabel 6 Enkele kritieke grenzen voor de χ^2 -verdeling 182

Trefwoordenregister 183



1.1 Statistiek

Data

Voorbeeld 1

Uit nieuwsgierigheid heb je aan je groepsgenoten een aantal vragen gesteld. Vervolgens heb je deze gegevens overzichtelijk willen weergeven:

Naam	Leeftijd	Geslacht	Op kamers	Beoefende sport
Bart	18	M	Ja	Voetbal
Lisa	19	V	Ja	Tennis
Erik	17	M	Nee	Voetbal
Anne	18	V	Nee	Fitness
Rick	18	M	Nee	Niet

Een ander woord voor gegevens is **data**. De datamatrix is de tabel waarin alle verzamelde (ruwe) gegevens worden weergegeven. De **onderzoekseenheid** (observatie-eenheid of analyse-eenheid) is datgene waarover je gegevens hebt verzameld. De onderzoekseenheid kan een individu of een groep individuen zijn. Datgene wat gemeten of geobserveerd wordt, is de **variabele**. Een variabele kan verschillende waarden aannemen. De **waarde** die een onderzoekseenheid op een variabele krijgt, is de score (meetwaarde of waarnemingsuitkomst). In de datamatrix staan de onderzoekseenheden in de (horizontale) rijen van de tabel en de variabelen in de (verticale) kolommen.

Statistiek

Statistiek is het verzamelen, ordenen, samenvatten en analyseren van data. Bij **beschrijvende statistiek** orden je gegevens en presenteert deze op aansprekende wijze. Vaak gebruik je hierbij tabellen en grafieken.

Voorbeeld 2

Het Centraal Bureau voor de Statistiek rapporteert dat in het jaar 2013 een Nederlander gemiddeld 45 duizend liter water gebruikte.

Vaak beschik je om praktische redenen (geld, tijd) niet over gegevens van de gehele groep (**populatie**) waarin je geïnteresseerd bent. Je hebt vaak wel gegevens van een **steekproef** uit de populatie. Je gebruikt **verklarende** (of inductieve) **statistiek** om van de steekproef te generaliseren naar de populatie.

Voorbeeld 3

Uit een steekproef van 500 stemgerechtigde Nederlanders bleek dat het CDA waarschijnlijk tussen de 24 en de 27 kamerzetels zou behalen als de verkiezingen nu gehouden zouden worden.

Opgaven

- 1** Bepaal van voorbeeld 1 de volgende aspecten:
 - a** de onderzoekseenheid;
 - b** de variabelen;
 - c** de waarden van de variabelen.

- 2** In een bestemmingsplan van gemeente A is voorzien in de aanleg van een geluidswal bij een nieuwe wijk om de overlast van een in de nabijheid liggende spoorlijn te beperken.

Nu zijn er bij de bewoners van de huizen vlak bij het spoor stemmen opgegaan om de gemeente te verzoeken de geluidswal niet aan te leggen, omdat voor hen het voordeel van een wijds uitzicht ruimschoots opweegt tegen het nadeel van de geluidsoverlast. Anderzijds zijn er bewoners van de kern van de wijk die wél een geluidswal willen, omdat het niet aanbrengen ervan voor hen alleen maar nadelen heeft (geen voordeel van een wijds uitzicht).

Om in deze kwestie een beslissing te nemen, besluit de gemeente via een enquête alle gezinshoofden in de wijk de vraag voor te leggen of hun gezin voor of tegen het aanbrengen van een geluidswal is. Na verwerking van de gegevens blijkt dat van de groep randbewoners 20% voor de aanleg van een geluidswal is, terwijl van de groep kernbewoners 75% voor de aanleg is.

Wat zijn in deze situatie de onderzoekseenheden?

- 3** Geef bij de volgende situaties aan of het gaat om beschrijvende of verklarende statistiek.
 - a** Eke wedt graag op voetbalwedstrijden. Voor de komende speelronde heeft hij de wedstrijd FC Utrecht – FC Twente uitgekozen. Eke maakt gebruik van de voorafgaande competitiewedstrijden om de kans te bepalen of FC Utrecht van FC Twente zal winnen.
 - b** Eke heeft berekend dat hij dit jaar €50 verloren heeft met het ieder weekend wedden op voetbalwedstrijden.
 - c** Leonie heeft, als hoofd van de EHBO-afdeling, een half jaar bijgehouden hoeveel spoedgevallen iedere dag het ziekenhuis bezoeken. Ze wil op deze manier aantonen dat er verschil in werkdruk is in het week-einde in vergelijking met een doordeweekse dag.



1.2 Variabelen en meetniveaus

Je bent geïnteresseerd in een bepaald kenmerk van wat je onderzoekt. Een *variabele* is een kenmerk dat verschillende waarden kan aannemen. Voorbeelden van variabelen zijn geslacht (man of vrouw), leeftijd, belangstelling voor ruimtevaart op een schaal van 1 tot 5, buitentemperatuur in februari. Deze kenmerken hebben verschillende **meetniveaus**. Het meetniveau van een variabele bepaalt welke statistische technieken je kunt gebruiken. Je moet dus weten wat het meetniveau van een variabele is.

Nominaal meetniveau

Een variabele op **nominaal** meetniveau heeft categorieën die een naam krijgen (vandaar ‘nominaal’) en gelijkwaardig zijn. Geslacht is een nominale variabele met als categorieën man en vrouw.

Ordinaal meetniveau

Een variabele op **ordinaal** meetniveau heeft ook categorieën, maar je kunt deze ordenen op grootte (vandaar ‘ordinaal’). Belangstelling voor ruimtevaart op een schaal van 1 tot 5 met 1 (totaal geen belangstelling) tot 5 (volledig ruimtevaartgek) is een ordinale variabele. Hoewel je de categorieën 1 tot en met 5 kunt ordenen, betekent een 5 niet vijf keer zoveel belangstelling als een 1.

Variabelen op nominaal en ordinaal meetniveau noem je **kwalitatieve variabelen**.

Interval/ratiomeetniveau

Met een **kwantitatieve variabele** kun je rekenen. Je onderscheidt het interval- en ratiomeetniveau. Bij een variabele op **intervalmeetniveau** zijn de afstanden tussen elkaar opvolgende getallen even groot (vandaar ‘interval’). Temperatuur is een variabele op intervalmeetniveau. Het interval tussen 1 en 2 graden Celsius is even groot als het interval tussen 10 en 11. Je kunt echter niet zeggen dat 2 graden Celsius twee keer zo warm is als 1 graad Celsius.

Een variabele op **ratiomeetniveau** heeft dezelfde kenmerken als een variabele op intervalmeetniveau, plus een absoluut nulpunt. Leeftijd is een variabele die je meet op rationiveau. Twee jaar is twee keer zo oud als één jaar. In de praktijk is het verschil tussen het interval- en ratiomeetniveau voor vrijwel alle statistische technieken niet relevant.

Opgaven

- 1 Vul het onderstaande schema over meetniveaus en eigenschappen in.

Eigenschap	Meetniveau			
	Nominaal	Ordinaal	Interval	Ratio
Onderscheid				
Ordering				
Gelijke verschillen				
Verhoudingen kloppen				

- 2 Petra wordt op straat aangesproken om mee te doen aan een smaaktest. Uit nieuwsgierigheid doet ze mee. Vervolgens krijgt ze de volgende vraag voorgelegd:

‘U moet bepalen in hoeverre de twee producten, die u zo meteen te zien krijgt, een verfrissende smaak hebben. De smaak van product I is vastgesteld op een score van 10. U moet hieraan de smaak van product II relateren. Als product II volgens u drie keer zo verfrissend is, dan geeft u 30 punten. Als product II volgens u maar de helft zo verfrissend is, dan geeft u 5 punten.’

Petra houdt zich, na de tekst twee keer goed gelezen te hebben, netjes aan de instructie.

Welk meetniveau wordt hier door de onderzoekers gebruikt?

- 3 De directie van een onderwijsinstelling wil weten wie de beste docent is. De studenten wordt gevraagd een rapportcijfer te geven aan iedere docent. Welk meetniveau hanteert de directie?

- 4 Leonie heeft op de EHBO-afdeling een aantal gegevens bijgehouden van binnengekomen patiënten. Geef steeds aan van welke soort meetniveau sprake is bij de volgende variabelen:

- a de nationaliteit van de patiënt.
- b totaal aantal behandelde patiënten aan het einde van de dag.
- c hoogst genoten opleiding van de patiënt.
- d haarkleur van de patiënt.
- e lengte van de patiënt.
- f leeftijd van de patiënt.



1.3 Discrete en continue variabelen

Variabelen verschillen niet alleen in meetniveau maar ook in het aantal mogelijke *waarden* die ze kunnen aannemen.

Een **discrete variabele** neemt alleen maar bepaalde waarden aan, zonder de tussenliggende waarden.

Voorbeeld 1

Het aantal inwoners van een stad is altijd een geheel getal; halve inwoners bestaan niet. Op 1 oktober 2008 had Amsterdam 755 269 inwoners.

Een **continue variabele** kan wel alle tussenliggende waarden aannemen.

Voorbeeld 2

Het gewicht van een brood kan 803,456 gram zijn.

In de praktijk zijn er natuurlijk beperkingen aan de nauwkeurigheid van de meting. Discrete en continue variabelen zijn beide kwantitatief, dus gemeten op interval/rationiveau.

Continuïteitscorrectie

Soms wil je vanwege het rekengemak de discrete variabelen behandelen als continue variabelen.

Voorbeeld 3

Je wilt weten hoe groot de kans is dat iemand vier koppen koffie per dag drinkt. Als je het aantal koppen koffie zou zien als een continue variabele, dan kom je met alle resultaten tussen 3,5 en 4,5 uit op vier koppen koffie.

Het omzetten van een bepaalde discrete grenswaarde in een geschikte grenswaarde voor de continue verdeling noem je de **continuïteitscorrectie**. Het gaat steeds om een correctie van 0,5. De waarde '4' uit een discrete verdeling komt overeen met het gebied 3,5 tot 4,5 van een continue verdeling. De waarden 4 en groter van de discrete verdeling komen overeen met $k \geq 3,5$ van de continue verdeling. De waarden kleiner dan 4 van de discrete verdeling komen overeen met de waarde $k < 3,5$ van de continue verdeling. De *continuïteitscorrectie* mag altijd worden toegepast en moet worden toegepast als de steekproefomvang $n < 50$. Als de steekproefomvang erg groot is ($n \geq 200$) is het effect van de *continuïteitscorrectie* verwaarloosbaar.

Opgaven

- 1** Leonie heeft op de EHBO-afdeling een aantal gegevens bijgehouden van binnengekomen patiënten. Geef steeds aan of er sprake is van een continue of discrete variabele:
 - a** de wachttijd voordat de patiënt geholpen wordt.
 - b** totaal aantal behandelde patiënten aan het einde van de dag.
 - c** het aantal niet-verzekerde patiënten.
 - d** het gewicht van de patiënt.
 - e** lengte van de patiënt.
 - f** geboortjaar van de patiënt.

- 2** Puck heeft onderzoek gedaan naar het snoepgedrag bij mede-studenten van haar opleiding. Ze heeft hiervoor aan 100 studenten gevraagd of ze minstens iedere dag snoep of chips eten. Vervolgens heeft ze aangegeven welk gedeelte van de steekproef minstens iedere dag snoept. Is hier sprake van een continue of discrete variabele?

- 3** Geef aan of 'het aantal huwelijken per jaar in een land' een continue of een discrete variabele is.

- 4** Een tentamen bestaat uit 10 vragen. Geef steeds aan of er sprake is van een continue of discrete variabele:
 - a** aantal goed beantwoorde vragen.
 - b** behaalde score voor de juist beantwoorde vragen (dus voor vraag 1 kan de student maximaal 10 punten halen, voor vraag 2 maximaal 6 enzovoort).
 - c** proportie (het aantal goed beantwoorde vragen gedeeld door het totaal).
 - d** behaald eindcijfer voor het tentamen.
 - e** hoeveelheid tranen, in milliliters, bij degenen die een onvoldoende hebben gehaald voor het tentamen.
 - f** aantal tranen bij degenen die een onvoldoende hebben gehaald voor het tentamen.

- 5** Beoordeel de volgende stelling (juist of onjuist):
'Bij een experiment heb je altijd te maken met discrete variabelen.'



1.4 Onafhankelijke en afhankelijke variabelen

Bij je onderzoek is het van belang dat je weet welke *relatie* er bestaat tussen de verschillende variabelen.

De **onafhankelijke variabele** (*oorzaak*) is het gegeven waarop je de onderzoekseenheden indeelt. De onafhankelijke variabele kan door de onderzoeker gemanipuleerd worden om de relatie tussen de verschillende variabelen te achterhalen en te verklaren.

De **afhankelijke variabele** (*gevolg*) is datgene wat je wilt onderzoeken. Dit is het gegeven dat varieert ten opzichte van de onafhankelijke variabele.

Voorbeeld 1

Een docent zegt aan het begin van het studiejaar dat het bijwonen van de lessen zorgt voor een hoger resultaat bij het tentamen.

Het aantal bijgewoonde lessen zijn in dit geval de onafhankelijke variabele en het tentamenresultaat de afhankelijke variabele.

Voorbeeld 2

Bij het gezegde ‘hoe ouder, hoe wijzer’ is leeftijd de onafhankelijke variabele. Als je op zoek bent naar de leeftijd waarop een vrouw met een bepaald opleidingsniveau haar eerste kind krijgt, is leeftijd de afhankelijke variabele.

Regressie- en variantieanalyse

Je gebruikt **regressieanalyse** om te onderzoeken of en hoe een *continue variabele* door een andere *continue variabele* wordt verklaard of voorspeld. Je gebruikt **variantieanalyse** als het gaat om een *onafhankelijke variabele* op *kwalitatief meetniveau*.

Voorbeeld 3

Je wilt weten wat de invloed van alcohol op het reactievermogen is.

Als je het alcoholpromillage steeds netjes meet, dan is de onafhankelijke variabele continue en kun je aan de slag met de regressieanalyse.

Weet je alleen maar of er ‘veel’, ‘middelmatig’ of ‘weinig’ gedronken is door de proefpersonen, dan is de onafhankelijke variabele gemeten op ordinaal meetniveau (kwalitatief) en moet je aan de slag met variantieanalyse.

Opgaven

- 1** De directeur van het Oorlogsmuseum Overloon gaat ervan uit dat er een verband bestaat tussen het aantal bezoekers en de voorspelde kans op neerslag.

Wat is de onafhankelijke variabele: aantal bezoekers of voorspelde kans op neerslag?

- 2** Enkele journalisten hebben onderzocht of internationaal georiënteerde banken eerder willen meewerken aan het ‘witten’ van zwart geld dan nationaal georiënteerde banken. Zij bezochten beide soorten banken met een koffertje met zogenaamd zwart geld en vroegen of de banken een mogelijkheid zagen het geld te witten. Welke variabele is in dit onderzoek de onafhankelijke variabele?

- 3** De wielrenners die deelnemen aan de Tour de France, worden geregeld blootgesteld aan onderzoeken. Voor de start van de Tour de France is de hematocrietwaarde ((het percentage rode bloedcellen in het bloed) van de renners gemeten. Vervolgens wordt gekeken of de hematocrietwaarde verandert door de geleverde inspanning.

Wat is de onafhankelijke variabele: hematocrietwaarde of het leveren van inspanning?

- 4** De sportkoepel NOC*NSF financierde een onderzoek naar het nut van hoogtestages. De onderzochte atleten vertoonden tijdens het wedstrijdseizoen sterk dalende hematocrietwaarden, maar dat was niet terug te zien in hun wedstrijdresultaten. Daarmee zou het nut van bijvoorbeeld epo-gebruik en bloed doping op losse schroeven komen te staan.

Je weet van de onderzochte atleten alleen of ze ‘veel’, ‘weinig’ of ‘niet’ op hoogtestage zijn geweest.

Welk type analyse moet je dan gebruiken: regressieanalyse of variantie-analyse?

1.5 Ordenen van gegevens

Bij het ordenen en presenteren van de verzamelde gegevens moet je rekening houden met degenen die naar het onderzoek kijken.

Klassenindeling

Als je de gevonden uitkomsten in groepen wilt presenteren, moet je van tevoren nadenken over een **klassenindeling**. Voor elke uitkomst moet er één plaats zijn, dus de klassen mogen elkaar niet overlappen.

Absolute, relatieve en cumulatieve frequentie

De (*absolute*) **frequentie** is het aantal keer dat de gebeurtenis is geconstateerd.

Bij de **relatieve frequentie** (*fractie of proportie*) geef je de frequentie als fractie (of in procenten) weer.

Door te kijken naar het aantal waarnemingen beneden een bepaalde grenswaarde bepaal je de **cumulatieve frequentie**. Als je van de opeenvolgende klassen steeds de bijbehorende fracties (of percentages) optelt, heb je te maken met de **relatieve cumulatieve frequentie**.

Voorbeeld 1

Een docent heeft de volgende tentamengegevens verzameld van 50 studenten:

1 student behaalde een 1, 2 studenten een 2, 3 studenten een 3, 6 studenten een 4, 8 studenten een 5, 9 studenten een 6, 11 studenten een 7, 7 studenten een 8, 2 studenten een 9 en 1 student een 10.

Resultaat	Frequentie	Cumulatieve frequentie	Relatieve frequentie	Relatieve cumulatieve frequentie
1	1	1	0,02	0,02
2	2	3	0,04	0,06
3	3	6	0,06	0,12
4	6	12	0,12	0,24
5	8	20	0,16	0,40
6	9	29	0,18	0,58
7	11	40	0,22	0,80
8	7	47	0,14	0,94
9	2	49	0,04	0,98
10	1	50	0,02	1

Opgaven

- 1** Geef de controleberekening van voorbeeld 1 voor:
- cumulatieve frequentie.
 - relatieve frequentie.
 - cumulatieve relatieve frequentie.
- 2** Een museumdirecteur heeft voor de maand juni bijgehouden hoeveel bezoekers er iedere dag gekomen zijn. Hieronder staat zijn lijstje:

312	354	301	312	348
330	296	283	327	319
286	315	336	293	292
352	363	308	299	309
289	364	333	348	317
296	354	324	304	316

- Presenteer de gegevens in een frequentieverdeling met een klassebreedte van 10.
 - Stel de cumulatieve frequentie voor de museumdirecteur op.
 - Wat ‘vertelt’ de cumulatieve frequentie de museumdirecteur?
- 3** Christiaan moet zijn verkoopresultaten van de afgelopen maand morgen inleveren bij zijn baas. Ilex, de speelse beagle pup van het gezin, heeft echter een paar flinke happen uit het overzicht genomen. Het onderstaande is nog wel leesbaar:

Orderomvang	Frequentie	Relatieve frequentie	Cumulatieve frequentie
0 - < 500	?	?	?
500 - < 1 000	?	0,06	12
1 000 - < 2 000	?	?	18
2 000 - < 5 000	9	?	?
5 000 - < 10 000	?	0,2	?
10 000 - < 25 000	?	?	?
Totaal	50	1	

Vul de ontbrekende getallen in.