

Inhoud

Inleiding	15
-----------	----

Deel I	Beschrijvende statistiek	17
--------	--------------------------	----

1	Tabellen, grafieken en kengetallen	19
---	------------------------------------	----

1.1	Case Game 16	20
1.2	Populatie en steekproef	22
1.3	Meetniveaus	23
1.4	De frequentieverdeling	25
1.5	Grafieken	26
1.6	Centrummaten	31
1.7	Spreidingsmaten	38
1.8	Trefwoorden	43
1.9	Samenvatting	43
1.10	Kennisopgaven	44
1.11	Toegepaste opgaven	44
1.12	Competentieprikkels	47

2	Lineaire regressie en correlatie	49
---	----------------------------------	----

2.1	Case Mediamarkt	50
2.2	Het spreidingsdiagram	51
2.3	Het lineaire verband	51
2.4	Het lineaire model	53
2.5	Kleinste-kwadratenschatters	54
2.6	De voorspelling	55
2.7	De correlatiecoëfficiënt	56
2.8	Trefwoorden	57
2.9	Samenvatting	57
2.10	Kennisopgaven	57
2.11	Toegepaste opgaven	58
2.12	Competentieprikkels	60

3	Tijdreeksen	63
3.1	Case Wegenwacht	64
3.2	Tijdreeksmodellen	65
3.3	Het voortschrijdend gemiddelde	68
3.4	De berekening van seizoensindices	72
3.5	Het bepalen van de trendlijn	74
3.6	De toepassing van het model	76
3.7	Trefwoorden	77
3.8	Samenvatting	77
3.9	Kennisopgaven	78
3.10	Toegepaste opgaven	78
3.11	Competentieprikkel	81

Deel II Kansrekening en kansverdelingen 85

4	Kansrekening	87
4.1	Case H&M	88
4.2	Kansdefinities	89
4.3	Kansregels	92
4.4	Permutaties en combinaties	97
4.5	Trefwoorden	101
4.6	Samenvatting	101
4.7	Kennisopgaven	101
4.8	Toegepaste opgaven	102
4.9	Competentieprikkel	106
5	De binomiale verdeling	107
5.1	Case Samsung	108
5.2	Kansvariabelen	109
5.3	De binomiale verdeling	112
5.4	Verwachte waarde en standaarddeviatie van de binomiale verdeling	116
5.5	Trefwoorden	117
5.6	Samenvatting	117
5.7	Kennisopgaven	118
5.8	Toegepaste opgaven	118
5.9	Competentieprikkel	121

6	De normale verdeling	123
6.1	Case Pinlock	124
6.2	De standaardnormale verdeling	125
6.3	De normale verdeling	128
6.4	De sigmagebieden	130
6.5	Het gemiddelde en de normale verdeling	131
6.6	De benadering van de binomiale verdeling met de normale verdeling	134
6.7	Trefwoorden	138
6.8	Samenvatting	138
6.9	Kennisopgaven	139
6.10	Toegepaste opgaven	139
6.11	Competentieprikkels	142
Deel III	Schatten en toetsen	145
7	Werken met een gemiddelde	147
7.1	Case Mercedes	148
7.2	Betrouwbaarheidsinterval één gemiddelde	149
7.3	Minimale steekproefgrootte één gemiddelde	154
7.4	Betrouwbaarheidsinterval één gemiddelde beperkt eindige populatie	156
7.5	Minimale steekproefgrootte één gemiddelde beperkt eindige populatie	158
7.6	Trefwoorden	158
7.7	Samenvatting	159
7.8	Kennisopgaven	159
7.9	Toegepaste opgaven	160
7.10	Competentieprikkels	165
8	Toetsen op één gemiddelde	165
8.1	Case Philips	166
8.2	Het formuleren van hypothesen	166
8.3	De toetsingsgrootte	167
8.4	Het kritieke gebied	168
8.5	De conclusie	169
8.6	Een andere benadering	170
8.7	Een- en tweezijdig toetsen	172
8.8	Fout van de eerste en tweede soort	173

- 8.9 De p-waarde 175
- 8.10 Trefwoorden 178
- 8.11 Samenvatting 178
- 8.12 Kennisopgaven 178
- 8.13 Toegepaste opgaven 179
- 8.14 Competentieprikkel 181

- 9 Toetsen op twee gemiddelden 185**
 - 9.1 Case kantoorkosten Amsterdam en Brussel 185
 - 9.2 Toetsen op twee gemiddelden: onafhankelijke steekproeven 186
 - 9.3 Gepaarde toets op twee gemiddelden 188
 - 9.4 Trefwoorden 192
 - 9.5 Samenvatting 192
 - 9.6 Kennisopgaven 192
 - 9.7 Toegepaste opgaven 193
 - 9.8 Competentieprikkel 195

- 10 Kleine aantallen 199**
 - 10.1 Case KLM 200
 - 10.2 Betrouwbaarheidsinterval gemiddelde bij kleine aantallen 201
 - 10.3 Toets op één gemiddelde en kleine aantallen 203
 - 10.4 Toets op twee gemiddelden en kleine aantallen 204
 - 10.5 Gepaarde toets gemiddelde bij kleine aantallen 207
 - 10.6 Trefwoorden 210
 - 10.7 Samenvatting 210
 - 10.8 Kennisopgaven 210
 - 10.9 Toegepaste opgaven 211
 - 10.10 Competentieprikkel 214

- 11 Toetsen op standaarddeviaties 217**
 - 11.1 Case Nike 217
 - 11.2 Toets op één standaarddeviatie 218
 - 11.3 Toets op twee standaarddeviaties 221
 - 11.4 Trefwoorden 224
 - 11.5 Samenvatting 224
 - 11.6 Kennisopgaven 224
 - 11.7 Toegepaste opgaven 224
 - 11.8 Competentieprikkel 227

12	Werken met een percentage	229
12.1	Case prijzenoorlog	230
12.2	Betrouwbaarheidsinterval één percentage z-benadering	231
12.3	Betrouwbaarheidsinterval één percentage Poisson-benadering	233
12.4	Minimale steekproefgrootte één percentage	235
12.5	Betrouwbaarheidsinterval één percentage beperkt eindige populatie	236
12.6	Minimale steekproefgrootte één percentage beperkt eindige populatie	237
12.7	Trefwoorden	239
12.8	Samenvatting	239
12.9	Kennisopgaven	240
12.10	Toegepaste opgaven	240
12.11	Competentieprikkels	243
13	Toetsen op percentages	247
13.1	Case accountants falen	248
13.2	Toets op één percentage z-benadering	249
13.3	Toets op één percentage Poisson-benadering	251
13.4	Toets op gelijkheid van twee percentages	254
13.5	Toets op verondersteld verschil van twee percentages	257
13.6	Toets op meer dan twee percentages	259
13.7	Trefwoorden	262
13.8	Samenvatting	262
13.9	Kennisopgaven	263
13.10	Toegepaste opgaven	263
13.11	Competentieprikkels	268
14	Analyse van een kruistabel	271
14.1	Case Esprit	271
14.2	Kruistabel en percenteren	272
14.3	De chikwadraattoets	274
14.4	Trefwoorden	277
14.5	Samenvatting	277
14.6	Kennisopgaven	277
14.7	Toegepaste opgaven	278
14.8	Competentieprikkels	281

15	Variantieanalyse	283
15.1	Case de beste verpakking	283
15.2	Variantieanalyse	285
15.3	De Tukey-toets	291
15.4	Trefwoorden	292
15.5	Samenvatting	292
15.6	Kennisopgaven	293
15.7	Toegepaste opgaven	293
15.8	Competentieprikkel	295
16	Toetsen en schatten lineaire regressie en correlatie	299
16.1	Case AB Seafood	299
16.2	Het spreidingsdiagram	300
16.3	Regressievergelijking	302
16.4	Voorspellingsinterval	304
16.5	Toets op correlatie	306
16.6	Trefwoorden	308
16.7	Samenvatting	309
16.8	Kennisopgaven	309
16.9	Toegepaste opgaven	309
16.10	Competentieprikkel	312
17	Niet-parametrische toetsen	315
17.1	Case B&C Products	315
17.2	De tekentoets	317
17.3	De Wilcoxon-toets	321
17.4	De Mann-Whitney-toets	325
17.5	De Kruskal-Wallis-toets	331
17.6	De rangcorrelatiecoëfficiënt van Spearman	334
17.7	Trefwoorden	337
17.8	Samenvatting	337
17.9	Kennisopgaven	338
17.10	Toegepaste opgaven	338
17.11	Competentieprikkel	342

18	Het keuzeproces	335
18.1	Case iedereen tv-manager	345
18.2	Het keuzeproces	347
18.3	Trefwoorden	353
18.4	Samenvatting	353
18.5	Toegepaste opgaven	353
	Bijlagen	359
	Tabel standaardnormale verdeling	359
	Tabel t-verdeling	361
	Tabel chikwadraatverdeling	363
	Tabel F-verdeling	365
	Tabel binomiale verdeling	366
	Tabel Poisson-verdeling	370
	Tabel betrouwbaarheidsinterval voor μ met de Poisson-verdeling	372
	Tabel Wilcoxon-toets	373
	Tabel Mann-Whitney-toets	374
	De Poisson-verdeling	373
	Formuleblad	377
	Register	391
	Over de auteurs	395



Tabellen, grafieken en kengetallen

Het bedrijfsleven besteedt veel aandacht aan marktonderzoek. Welke mening hebben consumenten over een product of dienst? Wat moet er eventueel worden verbeterd? Wat vinden consumenten van een advertentie of van de prijs-kwaliteitverhouding van een product? Bij de analyse van een marktonderzoek wordt veelal gebruikgemaakt van tabellen, grafieken en getallen die de verzamelde gegevens samenvatten.

In dit eerste hoofdstuk maak je kennis met de beschrijvende statistiek. Door middel van een case laten we zien op welke manieren je gegevens kunt weergeven (oftewel beschrijven). Na een uitleg over zogenoemde meetniveaus kijken we naar de constructie van tabellen en grafieken. Vervolgens komen de berekening van modus, mediaan, gemiddelde en standaarddeviatie aan bod.

Leerdoelen

Kennis:

- Je kent de vier meetniveaus nominaal, ordinaal, interval en ratio.
- Je kent de begrippen populatie en steekproef.
- Je weet aan welke eisen een tabel dient te voldoen.
- Je kent diverse grafieken.
- Je kent de meest voorkomende centrummaten en spreidingsmaten.

Vaardigheid:

- Je kunt een tabel maken rekening houdende met de eisen die hieraan gesteld worden.
- Je kunt diverse grafieken maken.
- Je kunt de meest voorkomende centrummaten en spreidingsmaten berekenen.

1.1 Case Game 16

De IBL-student James Blackmoore loopt stage bij een groot ICT-bedrijf in Chicago, Verenigde Staten. Hier wordt koortsachtig gewerkt aan nieuwe computerspellen voor de nieuwe Playstation, die over een halfjaar wereldwijd gelanceerd zal worden. Een aantal spellen zit nog in de testfase. In deze fase worden aan twintig testpersonen vragenlijsten voorgelegd. In de vragenlijst worden alle onderdelen van een spel doorlopen op onder andere gebruikersvriendelijkheid, plezier en uitdaging.



Playstation

De antwoorden op de vragenlijsten worden in een computerbestand gezet. Aan James wordt gevraagd om de gegevens van het spel Game 16 te analyseren. Op basis van de analyse dient hij voor de directie een presentatie te houden zodat helder wordt wat de eerste reacties uit de markt zijn. Eventuele gebreken kunnen nog worden aangepast.

Als James de gegevens ontvangt en opent op de computer, realiseert hij zich dat er verschillende soorten vragen met bijbehorende data zijn in de vragenlijst. Bijvoorbeeld:

- 1 Wat is je geslacht? ☐ man
☐ vrouw
- 2 Wat vind je van de vormgeving van Game 16? ☐ zeer goed
☐ goed
☐ noch goed, noch slecht
☐ slecht
☐ zeer slecht
- 3 Hoeveel geld geef je per kwartaal uit aan games?dollar

De ingevulde antwoorden zien er als volgt uit:

Tabel 1.1 Dataonderzoek Game 16

Geslacht	Vormgeving Game 16	Uitgave games
man	zeer goed	30
man	zeer goed	80
man	goed	150
man	noch goed, noch slecht	135
man	goed	175
man	zeer goed	35
man	goed	75
man	noch goed, noch slecht	70
man	goed	80
man	goed	185
vrouw	goed	40
vrouw	goed	75
vrouw	goed	80
vrouw	goed	40
vrouw	slecht	120
vrouw	goed	75
vrouw	zeer goed	25
vrouw	goed	75
vrouw	noch goed, noch slecht	30
vrouw	goed	60

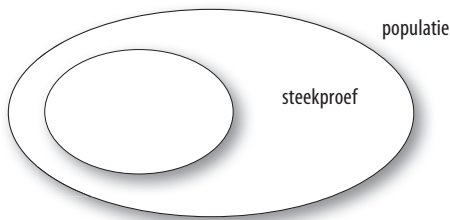
Zal hij tabellen maken of grafieken, of allebei? Moet hij iets berekenen en kengetallen laten zien? Hoe moet James de verkregen gegevens analyseren zodat de weergave van de resultaten zo helder mogelijk is?

1.2 Populatie en steekproef

De situatie in de case komt veel voor: er wordt onderzoek gedaan en de verzamelde gegevens dienen te worden geanalyseerd. Het liefst zouden we alle personen die het spel gaan kopen als testpersoon willen hebben. In dat geval onderzoeken we de *populatie*, in marketingtermen de totale doelgroep. In de praktijk zijn dat veel te veel mensen om mee te laten doen aan het onderzoek. We ondervragen meestal een gedeelte van de populatie. We spreken van het nemen van een *steekproef*.

Definitie Populatie: alle objecten of personen waar het onderzoek op van toepassing is.

Definitie Steekproef: een gedeelte van de populatie.



Figuur 1.1 Steekproef en populatie schematisch weergegeven

Het aantal elementen van een populatie wordt genoteerd met een N . In de case zijn er bijvoorbeeld 5 miljoen Amerikanen die het spel zullen gaan kopen, dus $N = 5$ miljoen.

Het aantal elementen in een steekproef wordt daarentegen genoteerd met n . In de case gaat het om een onderzoek onder 20 personen, dus $n = 20$.

We willen de gegevens van de steekproef uit de case generaliseren naar de populatie. Maar omdat de steekproef klein is, geven de resultaten van dit onderzoek slechts een indicatief beeld. Naarmate de steekproef groter is, ontstaat er een meer nauwkeurige indruk.

1.3 Meetniveaus

Om de gegevens van de vragenlijst uit de case op de juiste wijze te kunnen analyseren, moet je eerst goed kijken naar de gebruikte antwoordschaal.

Schaal: een logische weergave van de antwoordmogelijkheden.

Definitie

Er zijn verschillende soorten schalen. Bij elke schaal hoort een *meetniveau*.

Meetniveau: het type schaal.

Definitie

Er zijn vier meetniveaus:

- nominaal
- ordinaal
- interval
- ratio.

We laten deze vier meetniveaus de revue passeren.

Nominaal

Bij een *nominaal meetniveau* is er alleen sprake van verschillende antwoordmogelijkheden. Er zit geen systeem in deze antwoorden; er is geen logische volgorde. Een voorbeeld van dit meetniveau is de vraag naar geslacht. De antwoorden man en vrouw zijn verschillend, maar de volgorde in de vragenlijst is willekeurig. Een ander voorbeeld van een nominaal meetniveau is de vraag welke supermarkt iemand het meest heeft bezocht in het afgelopen jaar. Er zijn tal van antwoorden mogelijk, bijvoorbeeld:

- Super de Boer
- Albert Heijn
- Plus
- Jumbo
- Edah
- Overig.

Ook hier kunnen de antwoordmogelijkheden in elke willekeurige volgorde worden gezet.

Nominaal meetniveau: verschillende antwoordmogelijkheden zonder logische volgorde.

Definitie

Ordinaal

Bij een *ordinaal meetniveau* is er sprake van zowel verschillende antwoorden als een logische volgorde. Een voorbeeld van deze schaal is de vraag naar de mening over de vormgeving van Game 16 met de antwoorden:

- zeer goed
- goed
- noch goed, noch slecht
- slecht
- zeer slecht.

Je kunt zien dat er verschillende antwoorden zijn en dat er een logische volgorde is. Eventueel kun je beginnen met zeer slecht, maar je gaat deze antwoorden niet door elkaar gooien. Aan de *volgorde* ontleent dit meetniveau zijn naam.

Definitie *Ordinaal meetniveau*: nominaal + volgorde.

Interval

Bij een *intervalmeetniveau* is er ook sprake van een logische volgorde. Daarnaast hebben de verschillen tussen de antwoorden een eenduidige betekenis. Een voorbeeld is de temperatuur in graden Celsius. Als iemand zegt dat 70 graden Celsius 10 graden meer is dan 60 graden Celsius, dan is dat een even groot verschil als tussen 70 en 80 graden Celsius. Opvallend aan de intervalschaal is het ontbreken van een natuurlijk nulpunt. Zo betekent 0 graden Celsius iets anders dan 0 graden Fahrenheit.

Kenmerkend voor dit meetniveau is dus dat het *interval* tussen twee antwoorden een eenduidige interpretatie kent, naast de logische volgorde.

Definitie *Intervalmeetniveau*: ordinaal + eenduidige betekenis interval.

Ratio

Het *ratiomeetniveau* heeft naast de kenmerken van het intervalmeetniveau een eenduidige betekenis voor de verhouding (ratio) tussen twee getallen. Merk op dat de verhouding bij een intervalmeetniveau geen eenduidige betekenis heeft. Als we bijvoorbeeld kijken naar graden Celsius, dan is het niet zo dat als het 10 graden Celsius is, het twee keer zo warm is als 5 graden Celsius. Dit is wel zo bij het ratiomeetniveau. Als iemand bijvoorbeeld \$ 100 per kwartaal uitgeeft aan games, geeft hij tweemaal zo veel uit als iemand die \$ 50 per kwartaal aan games besteedt.

Bij een ratiomeetniveau is er sprake van een natuurlijk nulpunt. Als je € 0 hebt, heb je geen geld, maar als het 0 graden Celsius is, kun je niet zeggen dat er geen temperatuur is. Door dit natuurlijke nulpunt krijgt de verhouding tussen twee getallen de eenduidige betekenis.

Ratiomeetniveau: interval + natuurlijk nulpunt.

Definitie

Opdracht 1

Bepaal van de drie vragen uit de case het meetniveau.

Als je de antwoorden op een vragenlijst in een computerbestand hebt gezet, noem je elke vraag een variabele. Bij een variabele meet je een kenmerk van een persoon of een object. Variabelen kunnen kwantitatief of kwalitatief zijn. Bij een *kwantitatieve variabele* hebben we te maken met getallen, bijvoorbeeld de uitgave aan games per persoon per kwartaal. Bij een *kwalitatieve variabele* kijk je naar kenmerken die niet weergegeven worden door getallen, zoals het geslacht en de mening over Game 16.

1.4 De frequentieverdeling

Nu we weten welk meetniveau elke vraag heeft, kunnen we tabellen maken. Bij een nominaal en ordinaal meetniveau maken we een tabel waar bij elk antwoord het aantal en het percentage vermeld worden. Bij een interval- of ratiomeetniveau maken we eerst klassen; daarna vermelden we bij elke klasse het aantal en het percentage.

Dit betekent dat voor de vraag naar geslacht de volgende tabel gemaakt kan worden:

Tabel 1.2 Geslacht testpersonen Game 16

Geslacht	Aantal	Percentage
man	10	50
vrouw	10	50
totaal	20	100

Bron: Onderzoeksverslag Game 16, Fenthworth Company, 2013

Het is gebruikelijk om naast de aantallen en percentages van de antwoorden ook het totaal in de tabel te zetten. Als je gegevens overneemt uit een ander onderzoek, vergeet dan niet om een bronvermelding onder de tabel te zetten. Verder moet je bij het gebruik in een verslag of presentatie altijd een titel vermelden.

Omdat in deze tabel frequenties (aantallen) gebruikt worden, is de naam van deze tabel een *frequentieverdeling*.

Bij de vraag naar de mening over de vormgeving kunnen we ook een frequentieverdeling maken:

Tabel 1.3 Mening vormgeving Game 16

Mening vormgeving Game 16	Aantal	Percentage
zeer goed	4	20
goed	12	60
noch goed, noch slecht	3	15
slecht	1	5
zeer slecht	0	0
totaal	20	100

Tot slot kijken we naar de vraag naar de besteding aan games per kwartaal. We maken eerst de volgende klassenindeling:

- 0 tot 50 dollar
- 50 tot 100 dollar
- 100 tot 200 dollar.

Opdracht 2

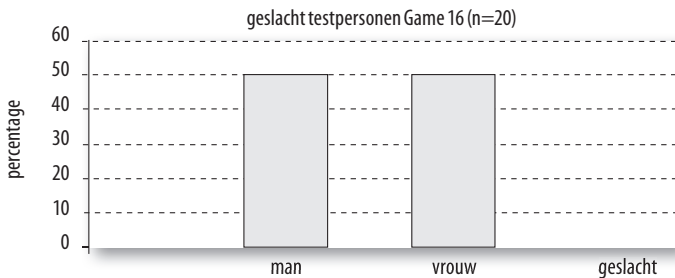
Maak de frequentieverdeling van de besteding aan games per kwartaal per persoon.

1.5 Grafieken

Behalve frequentieverdelingen kun je in de case ook grafieken maken. Er zijn heel veel grafieken die je met de pc heel snel kunt maken. We behandelen de meest voorkomende grafieken.

Het staafdiagram

Bij een nominaal meetniveau kunnen we een staafdiagram maken. In de case heeft geslacht een nominaal meetniveau. We gebruiken de frequentieverdeling van de vorige paragraaf om de grafiek te maken.



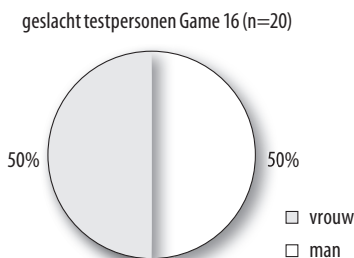
Figuur 1.2 Voorbeeld frequentieverdeling nominaal meetniveau

We zien een eenvoudige figuur. Toch valt er nog wel iets te vertellen over het staafdiagram:

- In de titel vermelden we het aantal personen dat meegedaan heeft aan het onderzoek ($n =$).
- Op de y-as zetten we niet de aantallen, maar de procenten.

Het cirkeldiagram

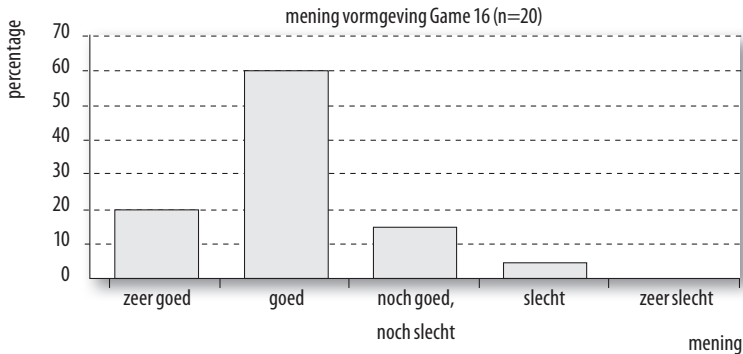
Bij een nominaal meetniveau kunnen we ook een cirkeldiagram maken. We gebruiken dezelfde gegevens als in de vorige paragraaf.



Figuur 1.3 Voorbeeld cirkeldiagram nominaal meetniveau

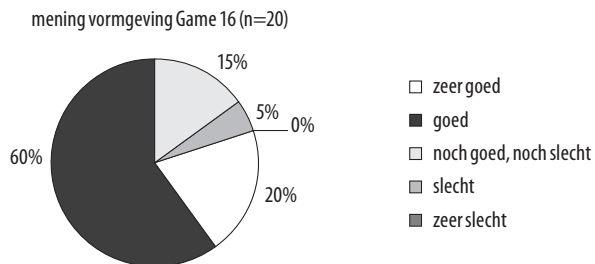
De vraag is dus of we nu een cirkeldiagram of een staafdiagram moeten gebruiken in het verslag. Het antwoord hierop is dat je zelf mag kiezen, het hangt van je persoonlijke voorkeur af.

Ook bij een ordinaal meetniveau kunnen we een staafdiagram en een cirkeldiagram maken. Zo wordt het staafdiagram van de vraag naar de mening over de vormgeving van Game 16:



Figuur 1.4 Voorbeeld staafdiagram ordinaal meetniveau

Het cirkeldiagram bij de vraag naar de mening over de vormgeving van Game 16 wordt bijvoorbeeld:



Figuur 1.5 Voorbeeld cirkeldiagram ordinaal meetniveau

Het histogram

Bij een interval- of ratiomeetniveau maken we een histogram. We kijken naar vraag 3 uit de case; deze ging over de besteding aan games per kwartaal.

Tabel 1.4 Besteding games per kwartaal

Klasse (in dollars)	Aantal	Percentage
0 tot 50	6	30
50 tot 100	9	45
100 tot 200	5	25

Een opvallende klasse in de tabel is de laatste klasse. Deze is veel breder dan de andere twee klassen, die even breed zijn. Dit doet zich vaak voor bij een variabele op interval- of ratiomeetniveau. Vanaf een bepaalde waarde wordt er dan een 'restklasse' gebruikt, omdat er anders heel veel klassen nodig zijn, waarin bovendien vaak maar weinig waarden vallen. In het voorbeeld is het aantal respondenten dat \$ 100 tot \$ 200 aan games besteedt lager dan de aantallen in de andere klassen, ondanks het feit dat deze klasse breder is dan de andere twee.

Het gevolg is echter dat vergelijking van de klassen niet helemaal eerlijk is. Dit probleem wordt opgelost met de zogenoemde *frequentiedichtheid*. Deze kan als volgt worden berekend:

- Kies een geschikte eenheid van klassenbreedte. Vaak is dit de kleinste klassenbreedte.
- Bepaal van alle klassenbreedtes de verhouding tot de eenheid van klassenbreedte.
- Deel alle frequenties door deze verhouding.

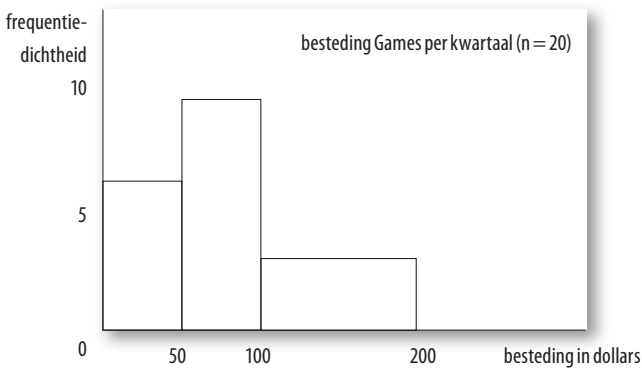
Aan de hand van de vraag over de uitgave aan games per kwartaal bekijken we wat er gebeurt als we dit doen:

Tabel 1.5 Besteding games per kwartaal met frequentiedichtheid

Klasse (in dollars)	Klassenbreedte	Eenheid	Aantal	Frequentiedichtheid
0 tot 50	50	1	6	6
50 tot 100	50	1	9	9
100 tot 200	100	2	5	2,5

- De geschikte eenheid van klassenbreedte is \$ 50 dollar.
- De laatste klasse is tweemaal zo breed.
- De frequentiedichtheid in de eerste twee klassen is hetzelfde als de frequentie. In de laatste klasse interpreteren we de frequentiedichtheid als volgt: per eenheid van \$ 50 vinden we 2,5 personen gemiddeld. Dat wil zeggen dat er van \$ 100 tot \$ 150 gemiddeld 2,5 personen zijn en van \$ 150 tot \$ 200 ook. Samen zijn dit 5 personen.

Het histogram wordt getekend door op de x-as de klassen in dollars te vermelden en op de y-as de frequentiedichtheid.



Figuur 1.6 Voorbeeld histogram

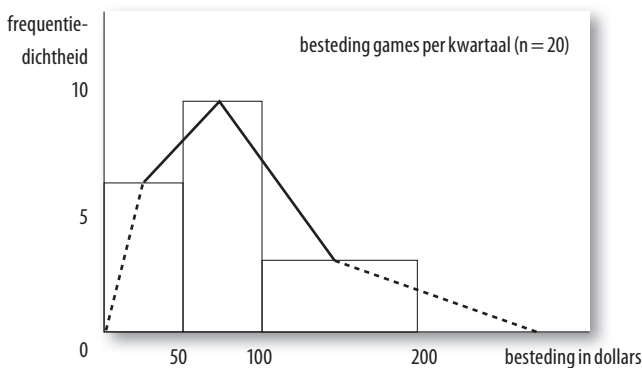
Uit de frequentieverdeling en uit het histogram halen we de *modale klasse*.

Definitie *Modale klasse*: de klasse met de grootste frequentiedichtheid.

In het histogram en de frequentieverdeling kunnen we zien dat dit de klasse is van \$ 50 tot \$ 100.

Het frequentiepolygoon

Uit het histogram kun je het *frequentiepolygoon* afleiden. Je neemt daarvoor het midden van elk blok uit het histogram en deze verbind je met elkaar. Aan de uiteinden neem je een fictief klassenmidden en trek je een gestippelde lijn. Als de grafiek zoals hier bij de oorsprong begint, trek je een stippellijn vanuit de oorsprong naar het eerste klassenmidden. Zo ontstaat een optisch verzorgde grafiek.



Figuur 1.7 Voorbeeld frequentiepolygoon

Het relatief cumulatief frequentiepolygoon

Standaard worden bij een aantal statistieksoftwarepakketten de percentages in een frequentieverdeling gecumuleerd (opgeteld). Zo kun je in één oogopslag zien hoeveel procent van de consumenten tot bijvoorbeeld \$ 100 besteedt, namelijk 75%.

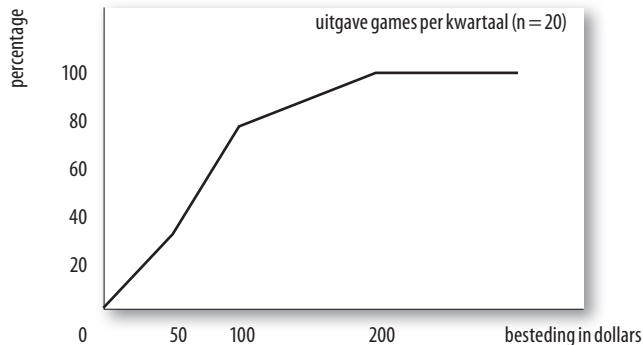
Tabel 1.6 Besteding games per kwartaal met gecumuleerde percentages

Klasse (in dollars)	Aantal	Percentage	Gecumuleerde percentages
0 tot 50	6	30	30
50 tot 100	9	45	75
100 tot 200	5	25	100

Op basis van de laatste kolom kunnen we het *relatief cumulatief frequentiepolygoon* tekenen. Deze figuur kan als volgt worden getekend:

Je begint bij de oorsprong 0.

- Daarna verbind je 0 ter hoogte van het gecumuleerde percentage van de eerste klasse bij de *rechtergrens* van de klasse in het histogram.
- Zo ga je door met verbinden van rechtergrens tot rechtergrens.
- Als je bij 100% bent, trek je een horizontale lijn.



Figuur 1.8 Voorbeeld relatief cumulatief frequentiepolygoon

1.6 Centrummaten

Nadat we tabellen en grafieken gemaakt hebben, dringt zich de vraag op of we kengetallen moeten berekenen waarmee we de gegevens kunnen samenvatten. We behandelen in deze paragraaf een aantal centrummaten.

Definitie Centrummaat: een getal dat iets zegt over het centrum van verzamelde getallen.

Er zijn drie belangrijke centrummaten:

- modus;
- mediaan;
- gemiddelde.

De modus

Bij de bespreking van het histogram heb je gezien dat de *modale klasse* de meest voorkomende klasse (de hoogste balk) in een histogram is. Bij elke rij getallen kun je de *modus* vaststellen.

Definitie Modus: de waarde die het meeste voorkomt in een rij getallen.

We laten dit zien aan de hand van een voorbeeld:

0 0 0 1 1 2

In de bovenstaande rij komt de waarde 0 het meeste voor; de modus is dus 0. Een ander voorbeeld is de volgende rij:

0 0 0 1 1 1 2 2

Nu zijn er twee waarden die het meeste voorkomen, namelijk 0 en 1. We noemen deze rij getallen *bimodaal*, er zijn twee getallen die de modus zijn.

Ook in de case kunnen we de modus bepalen voor de genoemde drie vragen. Allereerst bekijken we de frequentietabel van geslacht:

Tabel 1.7 Modus bepalen van geslacht

Geslacht	Aantal	Percentage
man	10	50
vrouw	10	50
totaal	20	100

We zien dat mannen en vrouwen even vaak genoemd worden. Geslacht is dus bimodaal.

Dan kijken we naar de mening over de vormgeving van Game 16:

Tabel 1.8 Modus bepalen mening vormgeving Game 16

Mening vormgeving Game 16	Aantal	Percentage
zeer goed	4	20
goed	12	60
noch goed, noch slecht	3	15
slecht	1	5
zeer slecht	0	0
totaal	20	100

De modus is hier gelijk aan het antwoord goed. Tot slot kijken we naar de besteding aan games per kwartaal per persoon:

Tabel 1.9 Modus bepalen besteding games per kwartaal

Besteding games	
30	40
80	75
150	80
135	40
175	120
35	75
75	25
70	75
80	30
185	60

Opdracht 3

Bepaal de modus bij de bestedingen aan games per kwartaal per persoon.

De mediaan

De tweede centrummaat die we bespreken, is de *mediaan*.

Mediaan: de middelste waarde na rangschikking van klein naar groot.

Definitie

We laten die zien aan de hand van een voorbeeld:

60 50 70

Wat is de mediaan van deze rij getallen? Eerst dienen we de getallen van klein naar groot te rangschikken:

50 **60** 70

De middelste waarde 60 is nu de mediaan. Vervolgens kijken we naar een ander voorbeeld:

60 50 70 80

We vragen ons wederom af wat de mediaan is. We rangschikken van klein naar groot:

50 **60** **70** 80

Er zitten twee waarden in het midden, namelijk 60 en 70. In dit soort gevallen spreken we af dat de mediaan het gemiddelde is van 60 en 70. De mediaan is dus 65.

Het laatste fenomeen doet zich altijd voor bij een even aantal getallen. Je zoekt na rangschikking de twee middelste waarden op en neemt hierover het gemiddelde. Dan heb je de mediaan gevonden!

In de case van Game 16 is de mediaan niet te bepalen bij geslacht vanwege het nominale meetniveau, maar wel bij de vragen over de besteding aan games per kwartaal en de mening over de vormgeving van Game 16.

Tabel 1.10 Mediaan bepalen mening vormgeving Game 16

Mening vormgeving Game 16	Aantal	Percentage
zeer goed	4	20%
goed	12	60%
noch goed, noch slecht	3	15%
slecht	1	5%
zeer slecht	0	0%
totaal	20	100%

De antwoorden in de tabel zijn al gerangschikt. We zoeken in volgorde waarneming 10 en 11. Bij beide waarnemingen is het antwoord goed. De mediaan is dus het antwoord goed.

Dan kijken we naar de bepaling van de mediaan bij de vraag naar besteding aan games per kwartaal. Allereerst rangschikken we alle bedragen van klein naar groot:

Tabel 1.11 Mediaan bepalen besteding games per kwartaal

Besteding games gerangschikt	
25	75
30	75
30	80
35	80
40	80
40	120
60	135
70	150
75	175
75	185

Opdracht 4

Wat is de mediaan van de bestedingen aan games per kwartaal per persoon?

Het gemiddelde

We onderscheiden twee soorten gemiddelden:

- het steekproefgemiddelde
- het populatiegemiddelde.

Steekproefgemiddelde: wordt berekend door alle gevonden waarden in de steekproef op te tellen en te delen door n = het aantal elementen in de steekproef. *Definitie*

Het steekproefgemiddelde van de getallen

50 60 70

is gelijk aan $(50 + 60 + 70)/3 = 60$.

De formule voor het steekproefgemiddelde is:

$$\bar{x} = \frac{\sum x_i}{n}$$

Waarbij:

- $\bar{x}$ = de notatie voor het steekproefgemiddelde.
- n = het aantal elementen van de steekproef.
- het symbool Σ wordt uitgesproken als ‘de som van’, dit is het sommatieteken uit de wiskunde.
- de getallen worden aangeduid met x_i , bijvoorbeeld $x_1 = 50$.

Definitie **Populatiegemiddelde:** wordt berekend door alle gevonden waarden in de populatie op te tellen en te delen door N = het aantal elementen in de populatie.

De formule voor het populatiegemiddelde is:

$$\mu = \frac{\sum x_i}{N}$$

Waarbij:

- μ = de notatie voor het populatiegemiddelde;
- N = het aantal elementen in de populatie.

Rekentechnisch worden steekproefgemiddelde en populatiegemiddelde op dezelfde wijze uitgerekend als je een aantal getallen hebt. Het verschil tussen het populatiegemiddelde en het steekproefgemiddelde is dat het populatiegemiddelde het gemiddelde is dat je graag zou willen weten. Omdat het meestal te veel tijd en geld kost om iedereen te ondervragen, volsta je in je onderzoek met het steekproefgemiddelde. Je schat dus het populatiegemiddelde met het steekproefgemiddelde.

Opdracht 5

Bepaal het steekproefgemiddelde van de bestedingen aan games per kwartaal per persoon.

Gemiddelde bij een frequentieverdeling

Als je deskresearch uitvoert, dan kan het je overkomen dat je wel een frequentieverdeling vindt, maar niet het bijbehorende gemiddelde. Zo zou je de frequentieverdeling van uitgave aan games per persoon per kwartaal bij een ander onderzoek kunnen gebruiken als secundaire informatie:

Tabel 1.12 Frequentieverdeling besteding games per kwartaal

Klasse (in dollars)	Aantal	Percentage
0 tot 50	6	30
50 tot 100	9	45
100 tot 200	5	25
totaal	20	100

Hoe groot is nu het steekproefgemiddelde op basis van deze tabel? In deze frequentieverdeling zijn alle verzamelde getallen gecompriemd in de gegeven klassen. Het is niet meer mogelijk om het exacte gemiddelde uit te rekenen als we alleen maar deze tabel hebben. Wel kunnen we een goede *benadering* vinden van het gemiddelde met behulp van de klassenmiddens:

- Bepaal van elke klasse het klassenmidden.
- Vermenigvuldig het klassenmidden met het aantal van deze klasse.
- Tel de uitkomsten hiervan op: we hebben een benadering van de som van de uitkomsten!
- Deel door het aantal waarnemingen.

We illustreren dit aan de hand van het voorbeeld:

Tabel 1.13 Hulptabel bepaling gemiddelde

Klasse (in dollars)	Aantal	Midden	Aantal * midden
	f_i	m_i	$f_i m_i$
0 tot 50	6	25	$6 \cdot 25 = 150$
50 tot 100	9	75	675
100 tot 200	5	150	750
totaal	20		1.575

- De klassenmiddens zijn 25, 75 en 150. Ga dit na!
- Deze vermenigvuldigen we met de aantallen, bijvoorbeeld $6 \cdot 25 = 150$ is het totaal van de eerste klasse.
- We tellen de uitkomsten hiervan op en vinden een totaal van alle waarnemingen van \$ 1.575. Er zijn 20 mensen die samen bij benadering \$ 1.575 aan games uitgeven per kwartaal.
- Het gemiddelde is $\$ 1.575/20 = \$ 78,75$ per persoon per kwartaal aan uitgave aan games. Vergelijk dit met het steekproefgemiddelde op basis van losse gegevens in de vorige paragraaf van \$ 81,75. Je ziet dat er een verschil is: het correcte steekproefgemiddelde is \$ 81,75, de benadering op basis van de frequentieverdeling is \$ 78,75.

Het gemiddelde dat we uitgerekend hebben, noemen we het *gewogen gemiddelde*. We wegen immers elk midden met de bijbehorende frequentie.

De formule voor het gewogen steekproefgemiddelde is:

$$\bar{x} = \frac{\sum f_i m_i}{n}$$

In de tabel van het voorbeeld zie je de benodigde notatie staan:

- f_i = de frequentie van klasse i , bijvoorbeeld $f_1 = 6$;
- m_i = het midden van klasse i , bijvoorbeeld $m_1 = 25$.

De formule voor het gewogen populatiegemiddelde is:

$$\mu = \frac{\sum f_i m_i}{N}$$

Het verschil met de vorige formule is dat we nu delen door N = het aantal elementen in de populatie. Voor de berekening maakt het niets uit of we te maken hebben met een steekproefgemiddelde of populatiegemiddelde.

1.7 Spreidingsmaten

In deze paragraaf kijken we naar een aantal *spreidingsmaten*.

Definitie **Spreidingsmaat**: een getal dat iets zegt over de spreiding van een aantal getallen.

Om het begrip spreiding duidelijk te maken geven we de volgende twee rijen getallen:

Rij 1: 50 60 70

Rij 2: 0 60 120

Als je naar de twee rijen kijkt, dan blijkt dat het gemiddelde van beide 60 is. Toch zijn beide rijen niet gelijk. De afstand tussen de getallen van de tweede rij is groter dan de afstand tussen de getallen van de eerste rij. Klaarblijkelijk is het gemiddelde niet voldoende om de rij getallen samen te vatten. We gebruiken een spreidingsmaat om een rij getallen verder te karakteriseren.

We onderscheiden twee spreidingsmaten:

- de spreidingsbreedte;
- de standaarddeviatie.

De spreidingsbreedte

Spreidingsbreedte: de hoogste waarde minus de laagste waarde (dit wordt ook wel *range* genoemd).

Kijken we naar de getallen:

50 60 70

Dan is de spreidingsbreedte $70 - 50 = 20$.

De spreidingsbreedte wordt gebruikt om snel een eerste indruk te krijgen van de spreiding van een rij getallen. Helaas is de spreidingsbreedte niet optimaal omdat niet alle gegevens gebruikt worden.

Opdracht 6

Bepaal de spreidingsbreedte bij de bestedingen aan games per kwartaal per persoon.

De standaarddeviatie

Er zijn twee standaarddeviaties:

- de steekproefstandaarddeviatie;
- de populatiestandaarddeviatie.

De *steekproefstandaarddeviatie* wordt als volgt berekend:

- Bereken het steekproefgemiddelde.
- Bepaal het verschil tussen alle getallen en het steekproefgemiddelde.
- Kwadrateer deze verschillen.
- Tel al deze kwadraten op.
- Deel door het aantal elementen minus 1.
- Neem tot slot de wortel.

Dit alles is een hele mond vol. We gaan dit illustreren met de volgende getallen, die we beschouwen als onze steekproef:

50 60 70

Deze getallen plaatsen we in een tabel om een en ander overzichtelijk weer te geven:

1.14 Hulptabel bepaling standaarddeviatie

Getal	Verschil	Kwadraat
50	$50 - 60 = -10$	100
60	$60 - 60 = 0$	0
70	$70 - 60 = 10$	100
totaal	0	200

- Bereken het steekproefgemiddelde. Het gemiddelde van 50, 60 en 70 is $(50 + 60 + 70)/3 = 60$.

Het verschil van het gemiddelde met 50, 60 en 70 is:

$$50 - 60 = -10$$

$$60 - 60 = 0$$

$$70 - 60 = 10$$

- Het kwadraat van deze verschillen is: $(-10)^2 = 100$; $0^2 = 0$; $10^2 = 100$.
- De som van de kwadraten is $100 + 0 + 100 = 200$.
- Delen door het aantal elementen in deze steekproef minus 1 is delen door 2, dus $200/2 = 100$.
- Tot slot nemen we de wortel, dus $\sqrt{100} = 10$.

De formule van de *steekproefstandaarddeviatie* is:

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

Berekening van deze spreidingsmaat geschiedt in de praktijk met behulp van de grafische rekenmachine of een softwareprogramma als Excel. De berekening gaat dan, ook bij veel gegevens, vliegensvlug.

De formule voor de *populatiestandaarddeviatie* luidt:

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

Belangrijkste verschillen met de formule van de steekproefstandaarddeviatie zijn:

- We gebruiken in de formule het populatiegemiddelde.
- We delen niet door $n - 1$ = het aantal steekproefgetallen minus 1, maar door N = aantal populatiegetallen.

Als de getallen 50, 60 en 70 samen alle waarnemingen in de hele populatie zouden vormen (de populatie bestaat dus uit drie eenheden), dan is de berekening praktisch hetzelfde als bij de steekproefstandaarddeviatie. Het verschil is dan dat we 200 delen door 3 in plaats van door 2. We vinden $\sigma = \sqrt{(200/3)} = \sqrt{66,67} = 8,17$.

In de case berekent James de steekproefstandaarddeviatie op basis van de 20 verzamelde gegevens met de pc. Hij vindt $s = \$ 47,77$.

Het verschil tussen de populatiestandaarddeviatie en de steekproefstandaarddeviatie is hetzelfde als het verschil tussen het populatiegemiddelde en het steekproefgemiddelde. De populatiestandaarddeviatie bereken je als je alle gegevens van een kenmerk in de populatie hebt. Meestal heb je een steekproef en schat je de populatiestandaarddeviatie met de steekproefstandaarddeviatie.

De volgende vuistregels zijn in het algemeen van toepassing bij grote aantallen gegevens:

- 68% van de gegevens zit in tussen het gemiddelde en plus of min de standaarddeviatie.
- 95% van de gegevens zit in tussen het gemiddelde en plus of min tweemaal de standaarddeviatie.

Opdracht 7

In de case hebben we slechts 20 gegevens voor uitgave aan games per kwartaal. Controleer desalniettemin of de vuistregels hier ongeveer van toepassing zijn of helemaal niet.

Standaarddeviatie bij een frequentieverdeling

Zowel de populatiestandaarddeviatie als de steekproefstandaarddeviatie is te bepalen op basis van alleen een frequentieverdeling.

De formule voor de populatiestandaarddeviatie luidt in dit geval:

$$\sigma = \sqrt{\frac{\sum f_i (m_i - \mu)^2}{N}}$$

Voor de steekproefstandaarddeviatie is de formule:

$$s = \sqrt{\frac{\sum f_i (m_i - \bar{x})^2}{n - 1}}$$

We zien wederom het kenmerkende verschil dat bij de populatiestandaarddeviatie gedeeld wordt door het aantal getallen N en bij de steekproefstandaarddeviatie door het aantal getallen n minus 1.

We nemen voor de berekening van de steekproefstandaarddeviatie de frequentieverdeling van de besteding aan games per persoon per kwartaal. Hier was het steekproefgemiddelde op basis van de tabel \$ 78,75, zoals eerder uitgerekend.

Tabel 1.15 Hulptabel bepaling standaarddeviatie

Klasse (in euro's)	f_i	m_i	$m_i - \bar{x}$	$(m_i - \bar{x})^2$	$f_i (m_i - \bar{x})^2$
0 tot 50	6	25	$25 - 78,75 = -53,75$	2.889,0625	$6 * 2.889,0625 = 17.334,3750$
50 tot 100	9	75	-3,75	14,0625	126,5625
100 tot 50	5	150	71,25	5.076,5625	25.382,8125
totaal	20				42.843,7500

- We halen van elk klassenmidden het gemiddelde \$ 78,75 af.
- Dan bepalen we het kwadraat van deze verschillen.
- De kwadraten vermenigvuldigen we met de frequentie van elke klasse en vervolgens tellen we op.
- We vinden dan in totaliteit 42.843,75.
- Dan delen we door $20 - 1 = 19$ en nemen de wortel.
- Het resultaat is $s = \sqrt{(42.843,75 / 19)} = \$ 47,49$.

Vergelijk dit met de steekproefstandaarddeviatie die James gevonden heeft in de vorige paragraaf op basis van de losse gegevens van $s = \$ 47,77$. Wederom zie je dat $s = \$ 47,77$ op basis van de losse gegevens de correcte steekproefstandaarddeviatie is en dat $s = \$ 47,49$ een benadering hiervan is.

De berekening van de populatiestandaarddeviatie volgt op soortgelijke wijze. In de formule van de populatiestandaarddeviatie wordt het populatiegemiddelde gebruikt en wordt, zoals gezien, gedeeld door N .

1.8 Trefwoorden

centrummaten	nominaal
cirkeldiagram	ordinaal
frequentiedichtheid	populatie
frequentieverdeling	populatiegemiddelde
gemiddelde	populatiestandaarddeviatie
gewogen gemiddelde	ratio
histogram	relatief cumulatief frequentiepolygoon
interval	spreidingsbreedte
kwalitatief	spreidingsmaten
kwantitatief	staafdiagram
mediaan	standaarddeviatie
meetniveau	steekproef
modale klasse	steekproefgemiddelde
modus	steekproefstandaarddeviatie

1.9 Samenvatting

We hebben gezien dat er vier meetniveaus zijn: nominaal, ordinaal, interval en ratio. Bij nominaal en ordinaal kunnen we direct een frequentieverdeling maken, bij interval en ratio maken we eerst klassen, daarna een frequentieverdeling. Bij een nominaal en ordinaal meetniveau zijn basisgrafieken die veel gebruikt worden het staafdiagram en het cirkeldiagram. Bij het interval- en ratiomeetniveau worden het histogram, frequentiepolygoon en het relatief cumulatief frequentiepolygoon veel gebruikt. Verder hebben we gezien dat de belangrijkste centrummaten modus, mediaan en gemiddelde zijn; de belangrijkste spreidingsmaten zijn de spreidingsbreedte en de standaarddeviatie. Ook is aandacht besteed aan de opsplitsing in populatie en steekproef voor de berekening van gemiddelde en standaarddeviatie. Tot slot is gekeken hoe we gemiddelde en standaarddeviatie op basis van een frequentieverdeling kunnen benaderen.

1.10 Kennisopgaven

- 1 a Wat is het verschil tussen een nominaal meetniveau en een ordinaal meetniveau?
- b Wat is het onderscheid tussen een ordinaal meetniveau en een intervalmeetniveau?
- c Wat is het verschil tussen een intervalmeetniveau en een ratiomeetniveau?
- 2 Wat is het verschil tussen een steekproef en een populatie?
- 3 Welke informatie dient er in een frequentieverdeling te staan?
- 4 a Geef twee typen grafieken die je kunt maken op nominaal en ordinaal meetniveau.
- b Geef drie typen grafieken die je kunt maken op interval- en ratiomeetniveau.
- 5 a Welke drie centrummaten kun je noemen?
- b Noem twee spreidingsmaten.
- 6 a Wat is het verschil tussen een steekproefgemiddelde en een populatiegemiddelde?
- b Wat is rekentechnisch het verschil tussen de steekproefstandaarddeviatie en de populatiestandaarddeviatie?

1.11 Toegepaste opgaven

Opgave 1

Geef van de volgende variabelen aan welk meetniveau ze hebben:

- Winst in euro's per maand.
- Mening nieuwe dvd Marco Borsato met als antwoordmogelijkheden:
zeer goed;
goed;
noch goed, noch slecht;
slecht;
zeer slecht.
- IQ.
- Bezit van een iPod met als antwoordmogelijkheden ja en nee.
- Uitgave aan bioscoop per kwartaal in euro's.

Gamification leidt tot meer site engagement

Gamification zorgt er voor dat engagement op websites groeit. Door de motiverende principes van gaming in een website te verwerken, nemen reacties gemiddeld met 13% toe. Social sharing op Facebook en Twitter groeit met 22% en nieuwe content wordt veel vaker

bekeken. Dit stelt onderzoeksbureau Gigga, dat miljarden interacties van gebruikers bij bedrijven als Pepsi, Nike en Dell analyseerde.

Bron: CLOU, tijdschrift voor marketing, informatie en research, april 2013

Kader 1.1

Opgave 2

In kader 1.1 lees je over een onderzoek naar het toepassen van gamification op websites.

In het artikel is onder andere gebruikgemaakt van onderstaande frequentieverdeling:

Tabel 1.16 Frequentieverdeling toename reacties website

Toename reacties website	Aantal interacties (in miljoenen)	Percentages
ja		13
nee		87
Totaal	2.000	100

- Welke aantallen moeten er in de lege cellen staan?
- Teken het bijbehorende cirkeldiagram.

Opgave 3

Bij een non-profitorganisatie wordt onderzoek gedaan naar de interne communicatie. Een van de aspecten die aan de orde komen, is de vraag welke mening men heeft over het personeelsblad. Van de resultaten heeft men de volgende tabel gemaakt:

Tabel 1.17 Frequentieverdeling mening personeelsblad

Mening personeelsblad	Aantal
zeer interessant	120
interessant	340
neutraal	475
oninteressant	50
zeer oninteressant	15
totaal	1.000

- a Bepaal de bijbehorende percentages.
- b Maak een staafdiagram over de mening van het personeelsblad.
- c Bepaal de modus en de mediaan.

Opgave 4

TYC Europe is het Europese distributiecentrum van TYC Brother Industrial Co. Ltd, een Taiwanese onderneming. Vanuit het hoofdkantoor in Almere worden originele en vervangende automaterialen (lampen, spiegels en condensoren) geleverd aan grote zakelijke klanten in Europa.

In een groot onderzoek in Nederland wordt onder andere gevraagd hoeveel geld automobilisten bereid zijn neer te leggen voor exclusieve voor- en achterlichten voor een auto. Van de eerste 50 respondenten staan hieronder de bedragen in euro's:

Tabel 1.18 Gegevens bereidheid uitgave lichten auto

600	1.400	4.000	2.250	1.300
750	800	1.500	3.500	1.500
1.200	1.000	1.600	1.800	1.000
500	1.500	1.200	2.000	900
2.000	2.500	800	800	1.750
1.600	1.200	1.100	1.000	1.600
1.350	4.000	1.150	1.400	750
1.800	900	500	1.900	3.000
900	750	750	500	2.750
1.000	1.000	1.300	600	2.500

- a Maak een frequentietabel met zes klassen.
- b Teken het bijbehorende histogram.
- c Teken in het histogram tevens het frequentiepolygoon.
- d Teken het relatief cumulatief frequentiepolygoon.

Opgave 5

Bereken van de gegevens van opgave 4:

- a Het steekproefgemiddelde.
- b De modus.
- c De mediaan.
- d De spreidingsbreedte.
- e De steekproefstandaarddeviatie.

Opgave 6

In een concurrentenanalyse wordt gekeken naar vijf regionale concurrenten van transporteur Quickman. Van deze concurrenten is achterhaald hoeveel geld afgelopen jaar geïnvesteerd is in innovatieve ontwikkelingen (in euro's):

50.000 25.000 70.000 125.000 50.000

Van Quickman zelf is de investering afgelopen jaar € 40.000.

- Bereken van de zes ondernemingen het populatiegemiddelde, de modus en de mediaan.
- Bereken van de zes ondernemingen de populatiestandaarddeviatie en de spreidingsbreedte.
- Leg uit waarom je hier σ uitrekent en niet s .

Opgave 7

In een steekproef is aan financieel adviseurs gevraagd wat hun brutojaarinkomen in euro's bedraagt:

Tabel 1.19 Brutojaarinkomen financieel adviseurs

Bruto-inkomen in euro's per jaar	Aantal
25.000 tot 40.000	12
40.000 tot 55.000	35
55.000 tot 70.000	24
70.000 tot 85.000	8
85.000 tot 120.000	6
totaal	85

- Bereken het steekproefgemiddelde.
- Bereken de steekproefstandaarddeviatie.

1.12 Competentieprikkels

In kader 1.2 staat een artikel over het gebruik van QR-codes.

Outdoor actief

Outdoor-advertenties worden steeds vaker interactief door middels van QR-codes, truckscreens en interactie door middel van beweging. 11% van de Europese consumenten heeft wel eens een QR-code gescand in een outdoor-advertentie, blijkt uit onderzoek van CBS Outdoor, terwijl 54% weet wat een QR-code is.

Bron: CLOU, tijdschrift voor marketing, informatie en research, april 2013



Kader 1.2

- a Teken een cirkeldiagram waarin de bekendheid met een QR-code wordt weergegeven.

Veronderstel dat tevens de volgende tabel bij dit onderzoek gemaakt kan worden:

Tabel 1.20 Bekendheid QR-code naar geslacht

Bekendheid QR-code/geslacht	Man	Vrouw
ja	502	309
nee	249	442
Totaal	751	751

- b Teken in één staafdiagram het verschil tussen mannen en vrouwen ten aanzien van de bekendheid met een QR-code.
- c Welk meetniveau heeft de bekendheid met een QR-code?