



J. J. Groot

BEDRIJFSSTATISTIEK MBA



Noordhoff Uitgevers

1e druk



·MBA·

Bedrijfsstatistiek MBA



Bedrijfsstatistiek

MBA

drs. J.J. Groot

Eerste druk

Noordhoff Uitgevers Groningen | Houten

Ontwerp omslag: Dorèl en anderen, Groningen
Omslagillustratie: ©Rex / Hollandse Hoogte

Eventuele op- en aanmerkingen over deze of andere uitgaven kunt u richten aan:
Noordhoff Uitgevers bv, Afdeling Hoger Onderwijs, Antwoordnummer 13, 9700 VB
Groningen, e-mail: info@noordhoff.nl

2 3 4 5 / 16 15 14 13 12

© 2008 Noordhoff Uitgevers bv Groningen/Houten, The Netherlands.

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever. Voor zover het maken van reprografische verveelvoudigingen uit deze uitgave is toegestaan op grond van artikel 16h Auteurswet 1912 dient men de daarvoor verschuldigde vergoedingen te voldoen aan Stichting Reprorecht (postbus 3060, 2130 KB Hoofddorp, www.cedar.nl/reprorecht). Voor het overnemen van gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (artikel 16 Auteurswet 1912) kan men zich wenden tot Stichting PRO (Stichting Publicatie- en Reproductierechten Organisatie, postbus 3060, 2130 KB Hoofddorp, www.cedar.nl/pro).

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.

ISBN (ebook) 978-90-01-85683-0

ISBN 978-90-01-71356-0

NUR 916

Woord vooraf

Bedrijfsstatistiek is bedoeld als een basisboek bij de studie in de bedrijfsstatistiek. Leidraad bij de keuze van de onderwerpen en het niveau waarop deze zijn besproken, is het Examenprogramma voor Moderne Bedrijfsadministratie, opgesteld door de Stichting Nederlandse Associatie voor Praktijkexamens (november 2007).

Kenmerkend voor deze methode is dat de diverse onderwerpen grondig worden besproken. De theoretische uiteenzettingen worden zoveel mogelijk toegelicht met één of meer uitgewerkte voorbeelden. Ook aan het taalgebruik is grote zorg besteed. De grondige behandeling van de leerstof en de vele voorbeelden zorgen ervoor dat een meer zelfstandige bestudering door de gebruikers mogelijk wordt.

De hoofdpunten uit de leerstof worden systematisch aan het eind van elk hoofdstuk nog eens samengevat. Een beperkte zelfcontrole is mogelijk, doordat alle hoofdstukken met een reeks kennisvragen worden afgesloten. De margewoorden geven de gebruiker houvast.

Bij dit basisboek is een Vraagstukkenbundel samengesteld. De opgenomen vraagstukken volgen nauwgezet de hoofdstukindeling van dit boek. Per hoofdstuk loopt de moeilijkheidsgraad van de vraagstukken geleidelijk op van zeer eenvoudig tot het examenniveau.

Hier en daar is kort ingegaan op het gebruik van een rekenmachine met statistische functie en op het gebruik van een spreadsheet. Over het algemeen geldt dat het gebruik van statistische programma's alleen zinvol is als men op de hoogte is van de begrippen en technieken, en de uitkomsten kan interpreteren.

De voorgaande editie van dit boek is tot stand gekomen in nauwe samenwerking met oud-collega P. van Kampen (†) die daarbij zijn grote didactische vaardigheden en ervaring heeft ingezet.

Opmerkingen over de inhoud van dit boek die kunnen leiden tot verbetering in de volgende druk zijn welkom.

U kunt uw op- en aanmerkingen doorgeven aan:

Noordhoff Uitgevers bv, Afdeling Hoger onderwijs, Antwoordnummer 13, 9700 VB, Groningen, e-mail: info@noordhoff.nl

Februari 2008

J.J. Groot



Inhoud

- 1 Statistisch onderzoek 11**
 - 1.1 Het begrip statistiek 12
 - 1.2 Het begrip massa 12
 - 1.3 De fasen van het statistisch onderzoek 12
 - Samenvatting 19
 - Vragen 20

- 2 Tabellen 21**
 - 2.1 De terminologie 22
 - 2.2 Het samenstellen van een tabel 23
 - 2.3 Tabellen met één of meer ingangen 24
 - Samenvatting 26
 - Vragen 27

- 3 Grafieken 29**
 - 3.1 Het coördinatenstelsel 30
 - 3.2 Het maken van een grafiek 33
 - 3.3 Keuze van de schaalverdeling 34
 - 3.4 Soorten grafieken 35
 - 3.5 Grafische computerprogramma's 46
 - Samenvatting 47
 - Vragen 49

- 4 Verhoudingsgetallen 51**
 - 4.1 Percentages 52
 - 4.2 Enkelvoudige indexcijfers 54
 - 4.3 Enkelvoudige hoeveelheds-, prijs- en waarde-indexcijfers 56
 - 4.4 De factorverwisselingstoets 58
 - 4.5 Samengestelde indexcijfers 59
 - 4.6 Samengesteld gewogen indexcijfers 60
 - 4.7 Samengesteld gewogen indexcijfers: de directe methode 60
 - 4.8 Samengesteld gewogen waarde-indices 62
 - 4.9 De samengesteld gewogen prijsindex 62
 - 4.10 De samengesteld gewogen hoeveelhedsindex 63
 - 4.11 Samengesteld gewogen indices: de indirecte methode 63
 - 4.12 Samengesteld gewogen waarde-indices 64
 - 4.13 Samengesteld gewogen prijsindices 66
 - 4.14 Samengesteld gewogen hoeveelhedsindices 67
 - 4.15 Samengesteld gewogen indexcijfers volgens het systeem Laspeyres 67
 - 4.16 Economische interpretatie 68
 - 4.17 Keuze basisperiode 69
 - 4.18 Basisverleggen 69
 - 4.19 Het koppelen van indexcijfers 71
 - 4.20 Defleren 72

- 4.21 Nominaal en reëel loon 73
- 4.22 De arbeidsproductiviteit 74
- 4.23 Productie-indexcijfers 75
- Samenvatting 77
- Vragen 80

5 Frequentieonderzoek 83

- 5.1 Frequentieverdeling 84
- 5.2 Soorten frequentieverdelingen 85
- 5.3 Het maken van een frequentietabel 87
- 5.4 Het aanduiden van klassen 93
- 5.5 Ongelijke klassebreedten 95
- 5.6 Het histogram 96
- 5.7 Relatieve frequenties 99
- 5.8 Polygonen 101
- 5.9 Gecumuleerde frequenties 104
- 5.10 Rekenkundig interpoleren 107
- 5.11 De ogief (= gecumuleerde frequentiecurve) 112
- 5.12 Spreadsheets 115
- Samenvatting 115
- Vragen 117

6 Karakteristieke grootheden 119

- 6.1 Modus, mediaan en rekenkundig gemiddelde uit individuele waarnemingsuitkomsten 120
- 6.2 Keuze uit drie middenwaarden 122
- 6.3 Modus, mediaan en rekenkundig gemiddelde vanuit een frequentietabel 124
- 6.4 Wel frequenties, geen klassen 128
- 6.5 Kwartielen 130
- 6.6 Wat is spreiding? 130
- 6.7 De variatiebreedte 131
- 6.8 Rekenmachine en spreadsheet 131
- Samenvatting 132
- Vragen 134

7 Correlatie 137

- 7.1 Doel van correlatieonderzoek 138
- 7.2 Het spreidingsdiagram 138
- 7.3 Soorten correlatie 140
- 7.4 De lineaire regressievergelijking 142
- 7.5 De regressiecoëfficiënt a en de regressieconstante b nader bekeken 147
- 7.6 Inter- en extrapoleren in een spreidingsdiagram 149
- Samenvatting 150
- Vragen 153

- 8** **Tijdreeksanalyse** 155
- 8.1 De componenten van een tijdreeks 156
 - 8.2 De bepaling van de trend (voortschrijdende gemiddelden) 157
 - 8.3 De bepaling van het additieve seizoenpatroon 161
 - 8.4 De rest- of toevallige component 163
 - 8.5 Seizoengedrag 164
 - 8.6 Het onderkennen van seizoen gedrag 165
 - 8.7 Een reeks met multiplicatief seizoen gedrag 166
 - 8.8 Residuen bij een reeks met multiplicatief seizoen gedrag 169
 - 8.9 Corrigeren voor het seizoen 169
 - 8.10 Prognose 171
- Samenvatting 173
- Vragen 176

- 9** **Bedrijfseconomische statistiek** 179
- 9.1 Begrippen in de economische statistiek 180
 - 9.2 Bedrijfseconomische kengetallen 180
 - 9.3 De Z-grafiek 183
 - 9.4 Enkele kanttekeningen bij de Z-grafiek 188
 - 9.5 De break-evenanalyse 189
 - 9.6 Enkele kanttekeningen bij de break-evenanalyse 194
 - 9.7 De 20-80-regel, concentratiecurve 195
- Samenvatting 198
- Vragen 200

Register 201



Statistisch onderzoek



1

- 1.1 Het begrip statistiek
 - 1.2 Het begrip massa
 - 1.3 De fasen van het statistisch onderzoek
- Samenvatting
Vragen

In de economische werkelijkheid van alledag worden alle handelingen van degenen die deelnemen aan het productieproces geregistreerd en verwerkt in administraties of overzichten.

Deze overzichten worden vaak gepubliceerd in de vorm van productieverslagen, jaarverslagen, personeelsstatistieken, inkomensstatistieken, import- en exportstatistieken enzovoort.

Die rapporten worden dan weer gebruikt om het beschreven verschijnsel nader te onderzoeken. In dit proces speelt de statistiek (paragraaf 1.1) een vitale rol, waarbij het er op aankomt de massa (paragraaf 1.2) nauwkeurig te omschrijven.

In dit hoofdstuk zullen wij ingaan op de wijze waarop gegevens systematisch worden verzameld. Zo'n statistisch onderzoek is aan strikte regels onderhevig (paragraaf 1.3). Wij besteden vooral aandacht aan de regels die gelden en de technieken die worden toegepast.

1.1 Het begrip statistiek

Het woord statistiek komt voor in twee betekenissen:

- a* een statistiek (meervoud: de statistieken): een publicatie waarin cijfermatige gegevens worden verstrekt over een bepaald verschijnsel. Bijvoorbeeld: een 'bevolkingsstatistiek', een 'ongevallenstatistiek', 'de statistieken wijzen uit dat ...';
- b* de statistiek (hiervan bestaat geen meervoud): een wetenschap (net zoals de natuurkunde, de economie enzovoort).

Definitie statistiek

Statistiek is de wetenschap van het verzamelen, verwerken en bestuderen van gegevens over massale verschijnselen met het doel een cijfermatig inzicht te geven in die verschijnselen.

1.2 Het begrip massa

Massa/populatie

De massa (ook wel populatie genoemd) is de verzameling te onderzoeken elementen.

De massa kan bijvoorbeeld bestaan uit:

- alle kiesgerechtigde Nederlanders;
- alle op één dag in een bepaalde fabriek geproduceerde gloeilampen;
- de werknemers van een vestiging van een supermarktketen.

Dit laatste voorbeeld laat zien dat het aantal elementen van een massa niet zeer groot hoeft te zijn.

De omschrijving (definitie) van een element van de populatie wordt 'eenheid van telling' genoemd, zoals bijvoorbeeld een kiesgerechtigde Nederlander, een geproduceerde gloeilamp, een werknemer.

1.3 De fasen van het statistisch onderzoek

We onderscheiden bij een statistisch onderzoek de volgende fasen:

- 1 de voorbereiding;
- 2 het verzamelen van het grondmateriaal;
- 3 het verwerken van het grondmateriaal;
- 4 het analyseren van de verkregen resultaten;
- 5 de rapportage.

De voorbereiding

De voorbereiding is erg belangrijk. Een slechte voorbereiding kan tot gevolg hebben dat de verkregen gegevens niet juist of onvolledig zijn of zelfs helemaal niet binnenkomen.

Een statistisch onderzoeker moet het onderzoek voorbereiden door:

- duidelijk het doel van het onderzoek vast te stellen;
- nauwkeurig de te onderzoeken populatie te definiëren;
- precies aan te geven over welke kenmerken van die populatie hij gegevens wil verzamelen;
- vast te stellen hoe, wanneer en door wie het grondmateriaal verzameld zal worden.

Bij een onderzoek naar bijvoorbeeld het eigen woningbezit zal men moeten aangeven wat men onder een eigen woning verstaat. Is een flat een woning? Is een permanent bewoonde stacaravan een woning en hoe zit het met een woonboot? En wanneer spreken we van bezit? Is eigendom hetzelfde als bezit? Een ander voorbeeld.

■ Voorbeeld

We willen de mening weten van de bewoners uit onze buurt over het plan van een ondernemer om een discotheek in de buurt te vestigen. Vanzelfsprekend zullen we op een plattegrond precies de grenzen van de buurt kunnen aangeven. Maar nu verder. Aan wie gaan we de vragen stellen? Als element kunnen we bezwaarlijk het gezin nemen. Want wat moeten we precies onder een gezin verstaan?

(Vormen twee werkende, vrijgezelle broers, die samen in een huisje wonen, één of twee gezinnen?) Dan krijgen we ook nog het probleem dat een van de gezinsleden namens het gezin de vragen moet beantwoorden. En hoe moet het dan, als de verschillende gezinsleden verschillende meningen hebben? Het is misschien verstandiger, dat we aan alle volwassen buurtbewoners de vragen gaan stellen. Het element wordt in dat geval dus de volwassen buurtbewoner. Maar wat verstaan we onder volwassen? We zullen in dat geval een bepaalde geboortedatum moeten noemen. Mensen die op of voor die datum geboren zijn, rekenen we tot de volwassenen.

Proefonderzoek

De statistisch onderzoeker doet er vaak verstandig aan eerst een proefonderzoek te houden. Dat is een onderzoek met als doel op kleine schaal de inhoud van het onderzoek, de gekozen statistische definities en de onderzoeksmethode in de praktijk te toetsen en bij gebleken ondeugdelijkheid te herzien.

Het verzamelen van het grondmateriaal

Het waarnemingsmateriaal in onbewerkte vorm wordt ook wel het grondmateriaal genoemd. Het grondmateriaal kan bestaan uit de resultaten van tellen of meten en uit antwoorden op aan personen gestelde vragen.

Primaire statistiek

Het meest gebruikelijke is *primaire statistiek*. We spreken van primaire statistiek als de bij het onderzoek gebruikte gegevens door de statisticus worden verzameld door feitelijke waarneming van een aantal individuele elementen.

Secundaire statistiek

Het is niet altijd nodig om het grondmateriaal door eigen waarneming te verzamelen. Soms is het mogelijk grondmateriaal of een deel ervan bij een andere instantie op te vragen. In zo'n geval spreken we van *secundaire statistiek*. Secundaire statistiek is een onderzoek waarbij gegevens worden gebruikt die al *eerder voor een ander doel* werden verzameld door anderen of door de statisticus zelf, zoals onderstaand voorbeeld illustreert.

■ Voorbeeld

Iemand die een onderzoek instelt naar het opnemen van ouderschapsverlof door het personeel van een bedrijf en daarbij geïnteresseerd is in de leeftijden en huwelijkse staat, zou dat aan alle personeelsleden zelf kunnen vragen. Maar, hij zal de gegevens beter kunnen opvragen bij de personeelsafdeling van het bedrijf. Andere voorbeelden van statistieken waarvoor het grondmateriaal op secundaire wijze wordt verkregen, zijn de statistieken van de buitenlandse handel en de statistieken van inkomens- en vermogensverdeling.

CBS

Deze statistieken worden door het Centraal Bureau voor de Statistiek (CBS; de voornaamste overheidsinstelling op het gebied van de statistiek in Nederland) samengesteld op grond van gegevens die door transportbedrijven en de Douanediens, respectievelijk de Belastingdienst ter beschikking worden gesteld.

Voor de statisticus is primair onderzoek met het daaruit voortkomende primaire materiaal in vele opzichten duidelijk: hij kent de definities van de massa, de kenmerken, de eenheden van meting, de nauwkeurigheid van meting en de betrouwbaarheid van de waarnemingsuitkomsten uit eigen ervaring. Bij secundair onderzoek en materiaal is die kennis er in veel mindere mate en moet hij vertrouwen op de door een andere onderzoeker gehanteerde definities en gevolgd aanpak.

Kortom, het voordeel van primaire statistiek: het grondmateriaal is voor de onderzoeker duidelijker en betrouwbaarder dan bij secundair onderzoek; het nadeel van primair onderzoek: de kosten zijn hoger en het vraagt meer tijd dan secundair onderzoek.

Enquête

Een vorm van primaire statistiek is het verzamelen van gegevens door het stellen van vragen aan personen. We noemen dat een *enquête*. Een enquête kan mondeling of schriftelijk worden gehouden.

Het voordeel van een mondelinge enquête is dat de enquêteur de bedoeling van de vragen kent en zo nodig aan de ondervraagde enige toelichting kan geven. Ook is een voordeel dat de enquêteur net zo lang doorgaat totdat alle vragen zijn beantwoord. Zo wordt voorkomen dat de geënquêteerde midden in de vragenlijst stopt en vele vragen onbeantwoord blijven. Een ander voordeel is dat de statisticus het van tevoren vastgestelde aantal personen kan enquêteren, omdat de enquêteurs net zo lang doorgaan met vragen tot het gewenste aantal is bereikt. De nadelen van een mondelinge enquête zijn de hoge kosten aan arbeidsloon voor de instructie en het inzetten van enquêteurs(trices).

Steeds meer wordt door enquêteurs en enquêtrices gebruikgemaakt van een laptop. Het computerprogramma loodst de enquêteur door de vragenlijst. Hij typt de antwoorden, die meteen worden ingelezen. Na een aantal enquêtes te hebben afgenomen, kan de enquêteur thuis met een modem via de telefoonlijn de resultaten doorgeven aan de computer van het onderzoeksbureau.

Telefonische enquête

Een speciale vorm van een mondelinge enquête is de telefonische enquête. Ten opzichte van de 'aan huis'-enquête kan daarmee aanzienlijke tijdswinst worden geboekt. Ook de telefonische enquête wordt computergestuurd. Met tientallen tegelijk worden de enquêtes afgenomen. Zo kan in één of twee dagen (avonden) een enquête onder 1 500 Nederlanders worden gehouden.

Bij de schriftelijke enquête wordt de vragenlijst aan de te enquêteren personen toegezonden. De ondervraagde vult de vragenlijst zelfstandig in en geeft of stuurt de ingevulde vragenlijst terug aan de onderzoeker. Het voordeel van schriftelijk enquêteren is dat de kosten betrekkelijk laag zijn. Bovendien kan de geënquêteerde op een moment dat het hem schikt rustig de tijd nemen voor het invullen van de vragenlijst. Het nadeel van een schriftelijke enquête is dat de onderzoeker maar moet afwachten hoeveel procent van de vragenformulieren ingevuld terugkomt, en dat kan wel eens vrij laag zijn.

Non-respons

De statisticus kan proberen de *non-respons* (het percentage ondervraagden dat niet meewerkt) zo klein mogelijk te maken door een presentje te beloven als er wordt meegewerkt.

Als de statistisch onderzoeker voor primaire statistiek heeft gekozen, moet hij nóg een belangrijke keuze maken: gaat hij alle elementen uit de populatie onderzoeken of alleen maar een gedeelte daarvan?

Volledige telling

We spreken van een *volledige telling* als alle elementen worden onderzocht. Een onderzoek onder een gedeelte van de populatie heet een *steekproef*.

Steekproef

Van een steekproef wordt gebruikgemaakt om verschillende redenen. Vaak ontbreekt de tijd en/of het geld om een volledige telling te houden. Het voordeel ten opzichte van een volledige telling is dan:

- een kostenbesparing en
- een tijdsbesparing (de uitslag van het onderzoek is sneller bekend).

Soms is een volledig onderzoek onmogelijk, omdat anders alle elementen bij het onderzoek verloren zouden gaan (bijvoorbeeld bij het testen van de belastbaarheid van zekeringen, het meten van de brandduur van lampen enzovoort).

Een onderzoek naar de *structuur* (samenstelling) van een populatie kan meestal zonder bezwaar worden beperkt tot een steekproef. De ervaringswet die dat rechtvaardigt heet de *Wet van de grote aantallen*.

Wet van de grote aantallen

Deze wet luidt als volgt:

naarmate het aantal waarnemingen toeneemt, zal het percentage van de waargenomen elementen dat een bepaalde eigenschap bevat tot een grenswaarde naderen en daarvan bij verdere opvoering van het aantal waarnemingen niet meer afwijken.

Dit wordt in het volgende voorbeeld geïllustreerd:

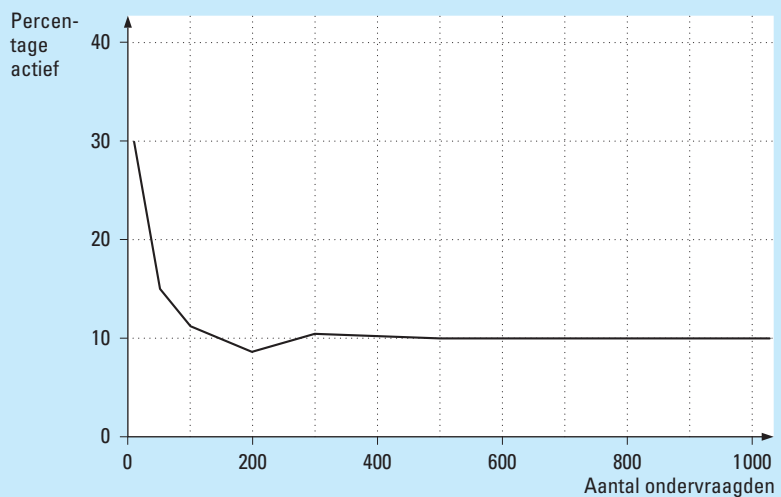
■ Voorbeeld

Bij de ingang van een voetbalstadion wordt aan de mannelijke bezoekers gevraagd of zijzelf in clubverband actief voetballen of hebben gevoetbald. Tussentijdse tellingen leiden tot het volgende overzicht.

| Aantal ondervraagden | Aantal actief | Percentage actief |
|----------------------|---------------|-------------------|
| 10 | 3 | 30,0 |
| 50 | 8 | 16,0 |
| 100 | 12 | 12,0 |
| 200 | 16 | 8,0 |
| 300 | 31 | 10,3 |
| 400 | 40 | 10,0 |
| 500 | 49 | 9,8 |
| 750 | 74 | 9,9 |
| 1000 | 98 | 9,8 |

We zien dat met het toenemen van het aantal ondervraagden het percentage actieve voetballers aanvankelijk nogal schommelt, maar geleidelijk aan worden de schommelingen kleiner en uiteindelijk stabiliseert in dit voorbeeld de uitkomst zich bij 9,8% (zie figuur 1.1). Het heeft geen zin om het aantal ondervraagden te vergroten, er zal geen andere uitkomst meer worden verkregen. Klaarblijkelijk is voor dit onderzoek de steekproef zo wél groot genoeg.

Figuur 1.1 Steekproefpercentage bij toenemend aantal ondervraagden



Representatief

Er dient voor gezorgd te worden dat een steekproef voldoende representatief is. Een steekproef is representatief als zij dezelfde eigenschappen (structuur) heeft als de populatie waaruit zij is getrokken. Een onderzoeker kan de representativiteit van de steekproef bevorderen door:

Aselect trekken

- de elementen aselect te trekken: de elementen uit de massa hebben dan alle dezelfde kans om tot de steekproef te behoren (denk aan de balletjes in een lottospel).
- een voldoende groot aantal elementen in de steekproef op te nemen.

Controle van het grondmateriaal

Het verzamelde grondmateriaal moet – voordat het kan worden verwerkt – worden gecontroleerd op volledigheid en juistheid. Die controle bestaat uit drie fasen:

1 Bepaal of het benodigde aantal formulieren is terugontvangen

Is dit niet het geval, probeer dan de ontbrekende formulieren alsnog te verzamelen. Vaak worden schriftelijke enquêtes anoniem gehouden. De ondervraagde hoeft geen naam op het antwoordformulier te zetten. Dat heeft tot gevolg dat de onderzoeker wél weet aan wie hij enquêteformulieren heeft toegezonden, maar niet van wie hij het formulier heeft terugontvangen. Aan iedereen moet dus weer worden gevraagd om, als ze dat nog niet gedaan hebben, alsjeblijft mee te werken.

2 Bepaal of op de formulieren alle vragen zijn beantwoord

Is dat niet het geval, dan moet zorgvuldig worden beoordeeld of het steeds om een bepaalde vraag gaat. Dat zou erop kunnen duiden dat bij een deel van de ondervraagden weerstand bestaat tegen het beantwoorden van die vraag. De wél op die vraag verkregen antwoorden zijn dan niet representatief voor de gehele populatie.

3 Bepaal waar mogelijk of de vragen juist zijn beantwoord

Soms zijn er in de beantwoording van verschillende vragen bepaalde tegenstrijdigheden aan te wijzen. Voor dat doel worden soms in een enquête speciale controlevragen ingebouwd. Als er in de beantwoording een berekening (bijvoorbeeld een optelling) voorkomt, dan moet de uitkomst worden gecontroleerd.

De onderzoeker moet proberen het onvolledige aangevuld te krijgen of vermoedelijke fouten verbeterd te krijgen, maar mag uiteraard nooit uit zichzelf in de enquêteformulieren wijzigingen aanbrengen.

Door tal van oorzaken kunnen fouten in het grondmateriaal sluipen. We onderscheiden:

- a toevallige fouten en
- b systematische fouten.

Toevallige fouten

Toevallige fouten

Toevallige fouten kunnen in verschillende richtingen invloed uitoefenen. Voorbeelden: fouten door vergissingen, afleesfouten, afrondingsfouten en dergelijke. De nadelige invloed op de uitkomst van het totale onderzoek is meestal gering, omdat toevallige fouten (van verschillende elementen) elkaar geheel of nagenoeg geheel opheffen bij optelling.

Systematische fouten

Systematische fouten

Systematische fouten zijn fouten waarvan de invloed steeds in dezelfde richting wordt uitgeoefend. Voorbeelden: fouten door een afwijking in de meetapparatuur of door het belang dat ondervraagden menen te hebben bij een bepaald antwoord enzovoort. Systematische fouten zullen het gedeelte van het onderzoek waarop zij betrekking hebben vertekenen.

Het verwerken van het grondmateriaal

Een groot deel van het verwerken van het grondmateriaal bestaat uit tellen. Eerst worden totaalstellingen gemaakt. Daarna worden de gegevens gesorteerd.

Het tellen gebeurt bij grotere onderzoeken nog maar zelden met de hand. Meestal worden de gegevens van de formulieren ingebracht in een computer en vastgelegd in een databestand. Met de computer en een statistisch programma gaat de verwerking van het grondmateriaal zeer snel.

Door ordening en telling van de basisgegevens ontstaat een beter overzicht van het te onderzoeken verschijnsel. Daarna kunnen de typische statistische berekeningen worden toegepast. Het beschrijven van zulke technieken zal de hoofdmoot van dit boek vormen. Daarop komen we dus later uitgebreid terug.

Om de verwerking van het grondmateriaal af te ronden dienen de resultaten van het tellen, ordenen en berekenen te worden gepresenteerd in de vorm van tabellen en grafieken.

Het analyseren van de verkregen resultaten

Ten slotte zal de onderzoeker het verwerkte materiaal analyseren, conclusies trekken en tot bevindingen komen. De uitkomsten van het onderzoek kunnen betrekking hebben op:

- de *omvang* van de populatie;
- de *structuur* van de populatie;
- de *ontwikkeling* in de tijd van het onderzochte verschijnsel;
- *overeenkomsten* of *verschillen* met andere populaties;
- de *samenhang* tussen verschijnselen.

De statistisch onderzoeker moet de conclusies kort en duidelijk beschrijven.

De rapportage

Rapport

Bij het opstellen van zijn rapport dient de statisticus rekening te houden met de lezers voor wie het is bestemd. Is de groep lezers een groot, niet-statistisch geschoold publiek, dan moet hij typische vakuitdrukkingen en formules achterwege laten. In dat geval gaat het meer om de strekking van de resultaten. Grafieken zullen voor deze groep lezers het rapport begrijpelijker maken.

Het verdient vaak aanbeveling om al vóór in het rapport de uitkomsten te vermelden. Veel lezers zijn namelijk allereerst geïnteresseerd in de conclusies van het onderzoek.

Als het onderzoek periodiek herhaald wordt, moet het rapport steeds op dezelfde wijze worden ingedeeld, zodat onderlinge vergelijking tussen de opeenvolgende rapporten eenvoudig mogelijk is.

Samenvatting

- 1 De fasen van statistisch onderzoek zijn:
 - a* de voorbereiding;
 - b* het verzamelen van grondmateriaal;
 - c* het verwerken van grondmateriaal;
 - d* het analyseren van de verkregen resultaten;
 - e* de rapportage.
- 2 De wet van de grote aantallen luidt: met het toenemen van het aantal waarnemingen nadert de relatieve frequentie van een verschijnsel tot een grenswaarde en zal daarvan bij verdere toename van het aantal waarnemingen niet meer afwijken. Deze wet rechtvaardigt het nemen van steekproeven om structuren te bepalen.
- 3 Verzameld grondmateriaal moet voor het verder wordt verwerkt, worden gecontroleerd op:
 - volledigheid en
 - juistheid.
- 4 Opeenvolgende toevallige fouten hebben een willekeurige richting en een willekeurige grootte. Doordat ze elkaar geheel of gedeeltelijk compenseren is hun invloed op het totale onderzoek meestal gering.
- 5 Systematische fouten werken in dezelfde richting en zullen daardoor de uitkomsten van een onderzoek vertekenen.
- 6 Een steekproef dient representatief te zijn voor de gehele populatie. De steekproef moet daarom voldoende groot zijn en de elementen moeten aselekt worden getrokken.

Vragen

- 1.1 Waarmee houdt de statistiek zich bezig?
- 1.2 Er wordt een statistisch onderzoek gehouden bij de zes filialen van een winkelbedrijf. Vormen deze zes filialen in statistisch opzicht een massa? Licht het antwoord toe.
- 1.3 De leiding van een bedrijf is geïnteresseerd in het verschijnsel wanbetaling in een bepaalde periode. Iemand doet het voorstel om voor die periode het aantal niet-betaalde facturen uit te drukken in een percentage van het aantal uitgeschreven facturen. Zou dit een goede indruk geven van het verschijnsel wanbetaling?
- 1.4 Waarom wordt er vaak een (beperkt) proefonderzoek gehouden?
- 1.5
 - a Wat is het verschil tussen primaire en secundaire statistiek?
 - b Wat is het voordeel van primaire statistiek boven secundaire statistiek?
 - c Welke nadelen kleven aan primair statistisch onderzoek?
- 1.6 Het CBS stelt onder andere statistieken van particuliere inkomens en vermogens samen. Het grondmateriaal komt van de Rijksbelastingdienst. Het is denkbaar dat deze gegevens afwijken van de werkelijkheid. Geef aan hoe en waarom.
- 1.7 Bij een studie naar actieve en passieve sportbeoefening in ons land is men geïnteresseerd in het gedeelte van de toeschouwers bij voetbalwedstrijden dat zelf voetbalt of heeft gevoetbald. Voor dit doel houdt men voor en tijdens een wedstrijd van een club uit de eredivisie een enquête onder het publiek.
 - a Is het resultaat geldig voor de toeschouwers in ons land?
 - b Is het resultaat geldig voor de toeschouwers van eredivisieclubs in ons land?
 - c Is het resultaat geldig voor de toeschouwers van de bewuste voetbalclub?
- 1.8 Lang niet altijd wordt een volledig onderzoek gehouden, maar volstaat men in plaats daarvan met een steekproefonderzoek. Noem daarvoor twee redenen.
- 1.9 Hoe kan een statisticus bevorderen dat de uitkomst van een steekproefonderzoek representatief is voor de gehele populatie?
- 1.10 Wat betekent het begrip eenheid van telling? Geef voorbeelden. (zie paragraaf 1.2)