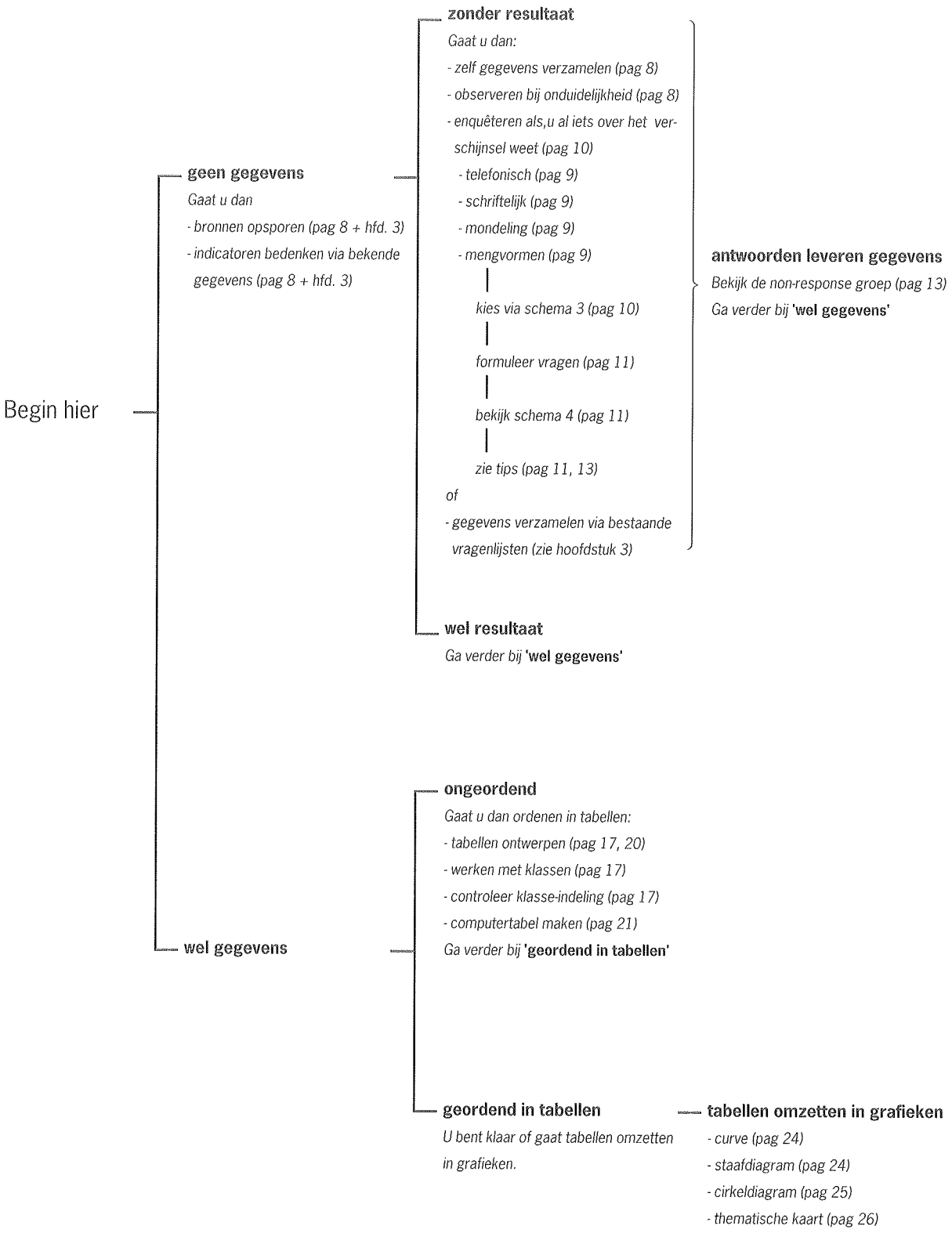
Inhoudsopgave hoofdstuk 8: Werken met cijfers

Dit hoofdstuk is een bewerking van het Basisboek Statistiek (LBVM 1986) aangevuld met ervaringen opgedaan op cursussen en themadagen over het onderwerp statistiek. De bewerking is uitgevoerd door K.B. Blits.

Zoekschema

Hier treft u twee schema's aan.
Met **Zoeken en ordenen** ziet u snel waar u in het hoofdstuk informatie vindt over het zoeken naar en het ordenen van gegevens.
Het andere schema - **Analyseren van cijfers** - helpt u op dezelfde wijze waar het gaat om het maken van analyses.

Zoeken en ordenen



Analyseren van cijfers

Via centrale tendentie: deze analyses geven aan waar het zwaartepunt van een aantal gegevens ligt

		pag.
de verdeling van de gegevens is regelmatig, er zijn geen grote uitschieters	→ Gebruik het rekenkundig gemiddelde	27
de verdeling van de gegevens is niet regelmatig, er zijn grote uitschieters	→ Gebruik de modus	28
de gegevenselementen wegen voor u niet alle even zwaar	→ Gebruik het gewogen rekenkundig gemiddelde	28
U wilt schommelingen in een reeks cijfers afvlakken	→ Gebruik het voortschrijdend gemiddelde	29

Via spreiding: deze analyses geven informatie over onderlinge verschillen van waarnemingsuitkomsten

U wilt weten of bepaalde uitkomsten uitzonderlijk hoog of laag zijn	→ Gebruik dan de standaardafwijking	30

Via reeksen cijfers: deze analyses hebben tot doel reeksen eenvoudiger/begrijpelijker weer te geven

U wilt de ontwikkeling van een bepaald verschijnsel weergeven over een aantal periodes	→ Gebruik een indexcijfer	33
U wilt het aandeel van een bepaald verschijnsel in het totaal weergeven over een periode	→ Gebruik percentuele verdelingen	35
U wilt de toe- of afname van een bepaald verschijnsel van het ene naar het andere jaar weergeven	→ Gebruik percentuele toe- of afname	36
U wilt een beter/ander perspectief aan uw gegevens geven	→ Gebruik kengetallen	37

Via rangordecorrelatie: deze analyses geven weer in hoeverre twee rijen met rangordes met elkaar overeenstemmen

U kunt twee rijen gegevens in een rangorde zetten en wilt zien of de neergezette rangordes met elkaar overeenstemmen	→ Gebruik rangordecorrelatie	47

Via correlatie: deze analyses geven weer in hoeverre twee variabelen met elkaar samenhangen

U heeft gegevens over twee variabelen en wilt zien of die samenhangen	→ Kijk eerst of de samenhang logisch is → Maak een correlatiepuntenwolk	45-47 44

Inleiding en leeswijzer

In de vorige hoofdstukken bent u diverse keren het woord analyseren tegengekomen. In hoofdstuk 3 'Prioriteitenstelling en criminaliteitsanalyse' wordt in de inleiding gesteld: 'In dit hoofdstuk gaan wij er van uit dat er bij het stellen van prioriteiten op het gebied van criminaliteitspreventie gebruik gemaakt wordt van (cijfermatige) gegevens over criminaliteit, overlast en onveiligheidsgevoelens. Het verzamelen en analyseren van deze gegevens wordt ook wel aangeduid met de term criminaliteitsanalyse'. Hoofdstuk 3 gaat vooral in op de tactische kant van het gebruiken van gegevens: wat wil ik met de gegevens bereiken en hoe doe ik dat? Het geeft ook inzicht in diverse gegevensbronnen die er zijn. Hierbij wordt aangegeven hoe die wel en hoe die niet gebruikt moeten worden.
In hoofdstuk 3 wordt niet ingegaan op het maken van berekeningen. Er wordt bijvoorbeeld wel over indexcijfers en procentuele verdelingen gesproken, maar niet aangegeven hoe die zijn te berekenen.
Over die berekeningen en een aantal andere vindt u in dit hoofdstuk meer. U zult het een en ander voorgeschoteld krijgen uit het brede terrein van de statistiek. Bij veel mensen bestaat een schrikbeeld over de statistiek. Men denkt dan aan allerlei ingewikkelde formules.
Gelukkig kunt u bij de analysewerkzaamheden zoals die in het Basisboek worden voorgestaan al een heel eind komen met een beperkt aantal bewerkingen, die niet zo ingewikkeld zijn.
Door als uitgangspunt te nemen dat de bewerkingen in ieder geval uitgevoerd moeten kunnen worden met de meest eenvoudige calculator op creditcardformaat (mits voorzien van een toets met het √-teken) vallen ingewikkelde berekeningen vanzelf af.

Het werken met cijfers heeft twee kanten. Enerzijds verwacht men van u dat u met cijfermateriaal aan de slag gaat, anderzijds vraagt men u aangeleverd cijfermateriaal van andere te beoordelen. Als u weet hoe een bewerking gedaan moet worden en wat de betekenis ervan is, dan bent u ook in staat om te zien of anderen via dezelfde bewerkingen wel een juiste voorstelling van zaken geven. Zelfs als u niet zelf aan de slag gaat, dan nog kan het daarom nuttig zijn gedeelten van dit hoofdstuk door te lezen. Statistieken zijn leugens, zo luidt immers een bekend gezegde!

Leeswijzer

In paragraaf 2 wordt uitgelegd wat statistiek is en worden een aantal termen toegelicht. Het meest krijgt u te maken met gegevens die al verzameld zijn. Soms moet u echter zelf aan de slag. In het eenvoudigste geval zijn de verzamelinstrumenten zoals vragenlijsten al ontwikkeld.

Als u een lokale slachtofferenquête wilt uitvoeren, kunt u de Politiemonitor ter hand nemen. Het wordt moeilijker als u zelf onderzoek moet gaan opzetten. Over wat daarbij komt kijken en hoe u een vragenlijst moet maken gaat paragraaf 3.
In paragraaf 4 vindt u aanwijzingen in verband met het gebruik van een eenvoudige calculator op creditcardformaat. Het is namelijk een praktijkervaring dat men niet optimaal omgaat met de mogelijkheden, die zelfs de meest eenvoudige creditcardcalculator te bieden heeft. Als men de mogelijkheden kent, kan men zich tijd besparen. Ook blijkt steeds weer dat de functies en het gebruik van sommige toetsen niet duidelijk zijn. Uitleg hierover is in deze paragraaf te vinden. Paragraaf 5 gaat in op hoe men van ongeordend cijfermateriaal tot geordende overzichten (tabellen) kan komen. Ook wordt kort aangegeven op welke wijze dergelijke overzichten grafisch weergegeven kunnen worden.
In paragraaf 6 worden een aantal berekeningswijzen uitgelegd die u bij diverse onderdelen van uw analyses van pas kunnen komen. Het zijn eenvoudig uit te voeren berekeningen uit het domein van de beschrijvende statistiek, zoals indexcijfers, gemiddelden, relatieve verdelingen en standaardafwijking. De beschrijvende statistiek houdt zich bezig met het zo goed mogelijk beschrijven van hetgeen men in de werkelijkheid vindt. Daarnaast bestaat ook nog de verklarende statistiek.
U krijgt dus maar een klein gedeelte van het terrein van de statistiek voorgeschoteld. Het zou echter niet goed zijn van het overblijvende gedeelte niets te behandelen. In de vakliteratuur, met name die waarin onderzoeksresultaten worden gepresenteerd, komt men termen als correlatie, chi-kwadraat en a-select tegen. Het is handig als u zich een voorstelling kan maken waarover men het heeft. In paragraaf 7 wordt een aantal termen nader belicht.

Omschrijving

Statistiek is: het verzamelen, ordenen en analyseren van gegevens over massaverschijnselen.

- **het verzamelen:** het gaat hier om het bij elkaar harken van diverse gegevens die worden gemeten.
- **het ordenen:** de gegevens worden geordend via bepaalde bewerkingen.
- **het analyseren:** de geordende gegevens worden nader bekeken via bepaalde bewerkingen.
- **massaverschijnselen:** statistiek gaat over een veelheid van gegevens, niet over individuele gegevens.

Soorten statistiek

Er zijn twee soorten statistiek:

- **de beschrijvende statistiek:** geeft weer wat er in de werkelijkheid is gevonden in termen van hoe veel, hoe groot, hoe lang enz. Conclusies uit dit gedeelte van de statistiek kunnen niet anders zijn dan: het een is meer dan het ander, het een is groter dan het ander, het een is minder lang dan het ander enz. Er kunnen geen conclusies worden getrokken in termen van oorzaak en gevolg, hoewel dit vaak wordt gedaan.
  Voorbeeld: Per 1 februari 1975 werd de valhelmplicht voor bromfietsers van kracht. Na dat tijdstip werden er aanmerkelijk minder bromfietsdiefstallen gepleegd. Het is verleidelijk te stellen dat dit komt omdat het stelen voor de potentiële daders door die valhelmplicht moeilijker wordt. Op basis van de gegevens kan men dat echter niet stellen. Men kan zuiver gezien alleen stellen: 'Het aantal bromfietsers met een helm op is op ± 100% komen te liggen (meer dan voorheen) en tegelijkertijd daalt het aantal bromfietsdiefstallen (minder dan voorheen). Wij vermoeden dat dit komt omdat .....'.

- **de verklarende statistiek:** hier wordt getracht juiste conclusies over oorzaak en gevolg en onderlinge verbanden te trekken. Vaak wordt er met steekproeven gewerkt en vertaalt men de conclusies over een klein gedeelte van het geheel naar het geheel al dan niet in de vorm van voorspellingen.
  Voorbeeld: Uit diverse analyses komt naar voren dat delinquentie op jeugdige leeftijd een belangrijke voorspeller is van criminaliteit op latere leeftijd, vooral als er frequent contact is geweest met politie en justitie.

Meetniveaus

Meten is het volgens bepaalde afspraken toekennen van waarden aan eigenschappen van diverse zaken. In de meeste landen meet men volgens het metrieke systeem (bijvoorbeeld meter, centimeter), in Angelsaksische landen volgens andere systemen (bijvoorbeeld inch, yard). Welk van de twee systemen men ook gebruikt, er kan nauwkeurig iets worden aangegeven. Meten doet men echter ook als men zegt: 'Dit produkt is beter dan dat'. Hier wordt iets veel minder nauwkeurig aangegeven.
Er zijn vier soorten niveaus waarop gemeten kan worden. Het niveau heeft te maken met wat voor soort bewerkingen er op de verzamelde gegevens losgelaten kunnen worden. En dat houdt weer verband met hoe en wat men meet. Te onderscheiden zijn het nominale, het ordinale, het interval- en het ratiomeetniveau. In schema 1 is weergegeven wat de kenmerkende verschillen tussen deze meetniveaus zijn.

Meetmethoden

Wees u er van bewust dat er niet alleen figuurlijk met diverse maten wordt gemeten, maar bij het vastleggen van gegevens óók en vooral letterlijk. Zo kan vandalisme aan openbare voorzieningen weergegeven worden in de totale schade, in de schade vanaf een bepaald bedrag of in alleen de schade van het verwijderen van graffiti. Soms wordt de schade weergegeven als alle materiaal- en loonkosten tezamen, soms ook niet. Het is dus niet alleen van belang te zien op welk niveau er gemeten wordt maar ook te weten welke meetafspraken er gemaakt zijn. Zijn de afspraken niet gelijk, dan loopt men de kans appels met peren te vergelijken.

Meetnauwkeurigheid

Hoe nauwkeurig men ook meet, ergens komt er na de komma een punt, waarbij men zegt dat het wel nauwkeurig genoeg is geweest. Bij 1:3 = 0,3333333 enz., wil men genoegen nemen met twee cijfers achter de komma, evenals bij 2:3 = 0,6666666 enz. Dan moet er iets gebeuren met wat er achter die twee cijfers achter de komma staat.
Hierover zijn de volgende afspraken gemaakt:

- de cijfers 1, 2, 3 en 4 worden naar beneden afgerond, dus worden gelezen als een 0 vanaf het punt waar men stopt, het laatste cijfer achter de komma verandert niet.
- de cijfers 5, 6, 7, 8 en 9 worden naar boven afgerond, dus worden gelezen als + 1 vanaf het punt waar men stopt en wordt bijgeteld bij het laatste cijfer achter de komma.

Dus in onze voorbeelden
1:3 = 0,3333333 is te lezen als
1:3 = 0,33 = 0,33
2:3 = 0,6666666 is te lezen als
2:3 = 0,66 = 0,67

Schema 1: Meetniveaus

SCHAAL	KENMERK	TOEGESTANE BEWERKINGEN
Nominale schaal		
– de waarden zijn alleen te onderscheiden in naam, b.v.: Nederlands, Engels, Duits etc. man, vrouw	– naam	– geen bewerkingen
Ordinale (of hiërarchische schaal)		
– de waarden hebben een naam – de waarden zijn ook nog in volgorde te plaatsen, bijv.: uitvoerend, middenkader-, leidinggevend personeel	– naam – volgorde	– alleen aangeven groter of kleiner,
Interval schaal		
– de waarden hebben een naam – er is een volgorde aan te geven – en de onderlinge afstand in de volgorde is bekend, bijv.: 01.00 uur, 02.00 uur etc.	– naam – volgorde – onderlinge afstand in de volgorde bekend	– optellen – aftrekken
Ratio schaal		
– de waarden hebben een naam – er is een volgorde aan te geven – de onderlinge afstand in de volgorde is bekend – en er is een nulpunt, bijv.: een meetlat	– naam – volgorde – onderlinge afstand in de volgorde bekend – nulpunt bekend	– alle bewerkingen mogelijk

3 Het verzamelen van gegevens

Inleiding
In hoofdstuk 3 (Prioriteitenstelling en criminaliteitsanalyse) is o.a. beschreven dat u voor het verzamelen van gegevens gebruik kunt maken van reeds bestaande bronnen (bv. de politiestatistiek van het CBS en de slachtofferenquêtes).
De vraag die u zich altijd moet stellen voordat u zelf met een vragenlijst aan de slag gaat is: 'Zijn de gegevens die ik wil hebben, al eens verzameld?'
Gegevens verzamelen kost zowel tijd als geld, dus als deze al voorhanden zijn kunt u op beide aspecten besparen.

Ga er niet te snel vanuit dat er geen gegevens over een bepaald verschijnsel zijn. Er zijn een groot aantal instanties, die registreren. Veel van wat die instanties registreren wordt verwerkt in de CBS-statistieken. Via het 'Statistisch zakboek' (een jaarlijkse CBS-uitgave) kunt u al een indruk krijgen van wat er zoal geregistreerd wordt.
Het CBS verwerkt niet alle aan haar toegezonden gegevens (het heeft ze dus wel); niet iedere instantie zendt al haar gegevens naar het CBS (er is dus meer). Navraag doen bij het CBS over het een en ander kan geen kwaad. U kunt zo ook op het spoor van gegevensleveranciers komen.
Ook door uw eigen organisatie worden in het kader van de werkzaamheden vaak statistieken vervaardigd, die u wellicht kunt gebruiken.

Het kan voorkomen dat u gegevens en cijfers nodig heeft die die bestaande bronnen niet kunnen leveren. U zult er dan zelf naar op zoek moeten gaan en u besluit een onderzoek uit te (laten) voeren. Er staan twee wegen open:
a) voor dat onderzoek maakt u gebruik van een of meer modulen van de Politiemonitor (zie hoofdstuk 3) met als belangrijkste voordeel dat u gegevens en cijfers krijgt die eenvoudig met andere onderzoeken die ook met de Politiemonitor zijn uitgevoerd, zijn te vergelijken. Een tweede voordeel is dat de vragenlijst in principe al vast is gelegd.
b) U besluit zelf een vragenlijst te gaan maken.

Over dit laatste gaan de volgende paragrafen.
Als u geen gegevensbronnen kunt vinden, heeft u nog een mogelijkheid, voordat u zelf aan het verzamelen slaat. U kunt kijken of u bij de wel bekende gegevens indicatoren vindt die uw verschijnsel min of meer (redelijk) benaderen. Zo zal bijvoorbeeld 'binding aan de maatschappij van de jeugd' gemeten kunnen worden via de indicator 'spijbelgedrag'. Daarover hebben scholen wel gegevens. Door daar gebruik van te maken bespaart u zich de moeite van het maken van vragenlijsten.
Het is een hele kunst om goede indicatoren te vinden. U kunt er van op aan, dat als u ze gebruikt, u er altijd (dit geldt ook voor het hier gegeven voorbeeld) commentaar op zal krijgen in de zin van 'Het is geen goede indicator, want .....'. Uw antwoord moet dan zijn 'Geeft u mij maar een betere!'

3.1 Waarnemingstechnieken: observeren en vragen

Als u zelf gegevens gaat verzamelen, kunt u kiezen tussen twee waarnemingstechnieken: observeren en vragen.
Bij **observatie** registreert de onderzoeker de gedragingen van de te onderzoeken groep.
Bij **vragen** registreert de onderzoeker de antwoorden van mensen (we spreken over: respondenten) op door hem gestelde vragen.

Observeren
In het schema 2 zien we dat bij observeren een onderscheid gemaakt kan worden tussen **par-**

Schema 2: Waarnemingstechnieken

Observeren	Vragen
1. Participerende observatie – Zelf meedoen – Anderen laten meedoen 2. Externe observatie – Zelf observeren – Anderen laten observeren	1. Mondeling 2. Schriftelijk 3. Collectief (groeps-interview) 4. Telefonisch 5. Mengvormen

ticiperende observatie en externe observatie. Bij participerende observatie is de onderzoeker opgenomen in de groep die hij bestudeert. Naast de rol van onderzoeker vervult hij de rol van groepslid: hij doet echt mee met de groep. In het kader van een onderzoek sluit iemand zich bijvoorbeeld aan bij een supportersclub van een voetbalclub waarvan de supporters nogal eens bij voetbalvandalisme zijn betrokken. Bij **externe observatie** vervult de onderzoeker alleen de onderzoeksrol. Bij de beslissing om wel of niet een verkeerslicht te plaatsen wordt een kruispunt op één dag het aantal passerende verkeersdeelnemers geteld.

Observatie is een waarnemingstechniek die, behalve voor verschijnselen die goed zijn waar te nemen (zoals vandalisme), vaak gebruikt wordt om een eerste indruk van een probleem te krijgen. Hoe zou het in elkaar zitten, wat is zinvol om nader te onderzoeken? Is het eigenlijk wel zinvol om het te onderzoeken? Observatie is natuurlijk alleen geschikt voor zaken, die via waarneming te achterhalen of af te leiden zijn. Als het u echter te doen is om 'meningen van buurtbewoners', dan zult u weinig aan observatie hebben. Meningen zitten in iemands hoofd en om die te achterhalen zult u dus meestal die mensen moeten gaan ondervragen.

Vijf manieren om mensen met vragen te benaderen

Bij de vraag: 'Hoe benader ik de mensen?' kunnen we een onderscheid maken tussen:

1. **De mondelinge benadering:** de onderzoeker (of enquêteur) ondervraagt iemand aan de hand van een vragenlijst. Bijvoorbeeld: de CBS-slachtofferenquête, de Politiemonitor.
2. **De schriftelijke benadering:** een lijst met vragen wordt aan iemand opgestuurd (of overhandigd) met het verzoek de vragen in te vullen. De ingevulde lijst wordt dan teruggestuurd of opgehaald.
3. **De collectieve benadering:** een groepje mensen wordt in een ruimte gezet. Met legt hen bepaalde vragen voor. Discussie in de groep behoort tot de mogelijkheden. Het geheel wordt genotuleerd of op band opgenomen en later door de onderzoeker uitgewerkt.
4. **De telefonische benadering:** de onderzoeker vuurt via de telefoon een aantal vragen af. Deze methode wordt veelal door de grote enquêtebureaus in Nederland gebruikt.
5. **Talloze mengvormen:** Bijvoorbeeld: U gaat bij de mensen langs en stelt ze een aantal vragen (mondelinge benadering), u laat een vragenlijst achter met het verzoek die in te vullen en terug te sturen (schriftelijke benadering).

Hoe kiest u de beste manier van benadering?

De voorgaande benaderingen gaan we op een aantal punten met elkaar vergelijken. Op grond daarvan kunt u elke keer weer kiezen, welke benadering voor uw doel het beste is.

De mate van respons

Hoe meer respons (= mensen die de vragen beantwoorden) u krijgt, des te beter. De ervaring is dat bij een mondelinge benadering de respons rond de 60% ligt, dat wil zeggen dat 60% van de personen die benaderd worden de vragen beantwoorden. Bij de telefonische benadering ligt dat wat lager, ± 50%. Bij een schriftelijke enquête is 40% vaak al veel. Een collectieve benadering kan 100% respons opleveren als de te ondervragen groepen makkelijk als groep te benaderen zijn (bijvoorbeeld een schoolklas).

Volledigheid en kwaliteit van de antwoorden

Mensen geven vaak onvolledige antwoorden als ze de vragen niet goed begrijpen, of als zij vragen niet willen of durven te beantwoorden. Over het algemeen zullen onvolledige antwoorden het minst voorkomen bij de mondelinge en de collectieve benadering. In die gevallen kan de interviewer immers het makkelijkst ingrijpen door een toelichting te geven of door de vraag op een andere manier te stellen. Kwaliteit van de antwoorden overlapt sterk met volledigheid. Soms kan men de kwaliteit verhogen door te werken met visuele hulpmiddelen (foto's, films) of met kaartjes ('welke van deze vier vindt u het mooiste').

Diepgang van de vragen

Bij mondelinge enquêtes kan men sterk doorvragen, bij schriftelijke veel minder. Bij een schriftelijke enquête liggen de vragen vast.

Zekerheid dat de bedoelde respondent zelf antwoordt

Zijn de antwoorden wel echt van het gezinshoofd, de autobezitter, de scholier? Als u dat zeker weet is dat natuurlijk het beste. De minste zekerheid hebt u in dit opzicht bij de schriftelijke benadering.

De aard van het onderwerp

Als het onderwerpen betreft waarbij persoonlijke en/of vertrouwelijke informatie wordt gevraagd ligt een persoonlijke en mondelinge benadering meer voor de hand dan een schriftelijke.

In schema 3 staan de voor- en nadelen van de verschillende wijzen van bevragen samengevat.

Schema 3: Voor- en nadelen van vier ondervragingsmethoden

Benadering	Respons	Kosten	Volledigheid kwaliteit	Diepgang	Respondent zelf	Vertrouwelijke c.q. persoonlijke informatie
Mondeling	60%	hoog	goed	groot	100% zeker	+
Schriftelijk	40%	laag	slecht	gering	onzeker	-
Collectief	100%*	laag*	goed*	groot*[1]	zeker	-
Telefonisch	50%	+/-	+/-	+/-	minder zeker, maar navragen	-

3.2 De enquête

Een enquête is een lijst met vragen waarmee u over een bepaald verschijnsel meer te weten kan komen.
Een enquêtevraag is een meetinstrument, dat tot doel heeft om de waarde te bepalen van iets of iemand op een bepaalde variabele.
Een variabele kan een kenmerk van iemand zijn (lengte, geslacht, woonplaats), maar het kan ook een oordeel of mening van iemand zijn.

Voorbeeld:
- Wij willen het lichaamsgewicht van meneer X weten.
  De variabele is: lichaamsgewicht;
  Het meetinstrument is: de weegschaal;
  De waarde van meneer X op de variabele 'gewicht' is: dat wat de weegschaal aangeeft (bijvoorbeeld 85 kilo of 170 pond).
- Wij willen de mening van meneer X weten over belastingontduiking.
  De variabele is: mening over belasting ontduiking.
  Het meetinstrument is een vraag over de mening.

Meestal zullen we niet alleen geïnteresseerd zijn in het gewicht van die ene meneer: we willen van een hele groep mensen (uit de onderzoekspopulatie) weten wat hun waarde op die variabele is. Deze mensen duiden we aan met 'elementen uit de populatie'.

Kortom: door middel van het onderzoek wordt van elementen uit de populatie hun waarde op een variabele bepaald.

Vaak geven we deze waarde weer door middel van een code.

Bijvoorbeeld:
- Populatie: bezoekers van de introductieavond
- Variabele: inkomen
- Waarde: 10.000/10.500/11.000/etc.
- Code: 100/105/110/etc.

of:
- Populatie: bezoekers van de introductieavond
- Variabele: mening over de avond
- Waarde: 'geen mening'/'beter dan de vorige'/'slechter dan de vorige'
- Code' 0/1/2

Van belang is dat variabelen en waarden meer of minder 'hard' kunnen zijn. Inkomen uitgedrukt in guldens is bijvoorbeeld een harde waarde, terwijl het oordeel over een introductieavond uitgedrukt in 'beter dan de vorige/slechter dan de vorige' een zachte is. In statistische termen uitgedrukt is het eerste voorbeeld een meting in de ratioschaal, het tweede in de ordinale schaal.

Soorten vragen

Men kan een vragenlijst of een vraaggesprek in verschillende vormen 'verpakken'. Belangrijke onderscheidingen zijn:

- **Gestructureerd versus ongestructureerd**

Vraaggesprekken kunnen liggen tussen twee

[1] Bij deze waarderingen geldt dat alle groepsleden meedoen. Indien in een groep slechts een paar mensen steeds het woord voeren komen de waarderingen veel lager uit. Van dit risico moet men zich goed bewust zijn.

uitersten: volledig gestructureerd en totaal ongestructureerd. Bij een **totaal ongestructureerd vraaggesprek** legt u van tevoren niet vast welke vragen u gaat stellen en welke mogelijke antwoorden u verwacht. U stelt wel vragen om het gesprek op gang te houden, maar het blijft een gesprek dat alle kanten op kan gaan. Een ongestructureerd vraaggesprek is handig als u nog weinig weet over het onderwerp of over de mensen die u onderzoekt. Door het verloop van het gesprek heel open te houden krijgt u maximale informatie. Een **volledig gestructureerd vraaggesprek** houdt in, dat u werkt met een vragenlijst waarin de vragen en de volgorde waarin u ze stelt al zijn vastgelegd. Vaak zijn de mogelijke antwoorden van tevoren ook al vastgelegd. De respondent moet dan kiezen uit die mogelijke antwoorden. Het grote voordeel van een gestructureerde vragenlijst is dat de informatie die van verschillende respondenten verkregen wordt, onderling goed te vergelijken is en gemakkelijk te verwerken is. U stelt aan iedereen precies dezelfde vragen. Om een goed gestructureerde vragenlijst te kunnen maken zult u meestal eerst moeten gaan praten (ongestructureerde vraaggesprekken houden) met een aantal mensen uit de groep die u wilt onderzoeken, om te weten te komen welke vragen u in welke volgorde moet stellen.

- **Open versus gesloten**

De mate van gestructureerdheid van een vragenlijst heeft alles te maken met het aantal open en gesloten vragen dat men in de vragenlijst heeft opgenomen. Open vragen zijn vragen waarop de respondent kan antwoorden wat hij of zij wil. Met open vragen krijgt men veel informatie, maar het verwerken van de antwoorden vergt meestal veel tijd. Gesloten vragen zijn vragen waar de respondent moet kiezen uit van tevoren vastgestelde antwoorden, bijvoorbeeld: ja/nee/geen mening. Met gesloten vragen krijgt men minder informatie, maar de antwoorden zijn wel snel te verwerken: men telt eenvoudigweg het aantal respondenten dat een bepaald antwoord gaf (bijvoorbeeld: 35x ja, 20x nee, 6x geen mening).
Gesloten vragen kunt u alleen maken:

- als u vrij simpele informatie wilt vergaren (als u bijvoorbeeld alleen leeftijd, geslacht en woonbuurt van de ondervraagden wilt weten); en/of
- als u heel goed op de hoogte bent van de mogelijke antwoorden die de ondervraagden kunnen gaan geven.

In dat laatste geval geldt weer, dat u beter eerst eens (ongestructureerd) kunt gaan praten met mensen uit de te onderzoeken groep, zodat u een goed beeld krijgt van de mogelijke antwoorden die u kunt verwachten.
In schema 4 worden de voor- en nadelen van open en gesloten vragen weergegeven.

Schema 4: Voor- en nadelen van open gesloten vragen

	Open	Gesloten
Men krijgt *zoveel* mogelijk informatie:	ja	nee
Men krijgt zoveel mogelijk *dezelfde* informatie:	nee	ja
Snel te verwerken door de onderzoeker:	nee	ja
Makkelijk voor de vragensteller:	nee	ja
Makkelijk voor de respondent:	nee	ja

Tips bij het maken van vragen voor een vragenlijst 3.3

Het maken van een vragenlijst is minder eenvoudig dan het lijkt. U zult zelf wel eens vragenformulieren ingevuld hebben, waarbij u dacht: 'Wat bedoelen ze nu?' Denkt u maar eens aan het aangifteformulier voor de inkomstenbelasting ..... .

Hierna volgen enkele tips die u kunt gebruiken bij het maken van een vragenlijst.

Tip 1: Beperk uw vragenlijst

Vragen staat vrij, maar of u antwoord krijgt is een tweede. Hoe langer de vragenlijsten zijn, des te minder kans u heeft dat alles nauwgezet en volledig wordt ingevuld.
Men komt vaak in de verleiding veel vragen te stellen. U kunt zelf het aantal vragen beperken door u steeds met enig cynisme af te vragen: 'Als ik straks het antwoord weet, wat doe ik er

dan mee'? U kunt bijvoorbeeld in het kader van een bedrijfsonderzoek vragen 'Wilt u meer verdienen'?. Het ligt voor de hand dat iedereen daarop volmondig 'ja' zegt. Aan dat 'ja' heeft u alleen iets, als u werkelijk in staat bent de inkomens te beïnvloeden.

Tip 2: Ken uw doelgroep
U moet vragen stellen waarvan de inhoud en de gebruikte begrippen voor de ondervraagden ook duidelijk zijn. Als u veel bij inbraakpreventieprojecten bent betrokken kunt u zich wat bij het begrip technopreventie voorstellen. Dit zal bij de gemiddelde Nederlander niet het geval zijn. Een vraag als 'Heeft u uw huis met technopreventieve middelen beveiligd?' zal niet voor iedereen duidelijk zijn. En als dat wel het geval mocht zijn is er weer een verschillend niveau van kennis over technopreventieve maatregelen. U moet hierover nadenken. Anders loopt u de kans antwoorden te krijgen die niet met elkaar te vergelijken zijn noch over de zaak zelf gaan.

Tip 3: Houdt de vragen kort
Korte vragen zijn meestal eenvoudiger te beantwoorden, dan lange ingewikkelde vragen. Hoe langer een vraag hoe meer interpretatiemogelijkheden er zijn.

Tip 4: Stel geen vragen waarin algemene termen voorkomen
De vraag 'Kijkt u vaak naar de t.v.' legt de bepaling van 'vaak' bij degene, die antwoordt. Als u 'ja' of 'nee' krijgt weet u eigenlijk nog niets. Beter is een vraag als 'Heeft u deze week meer dan vier dagen naar de t.v. gekeken'? Uw bepaling van 'vaak' is dan, 'meer dan vier dagen per week t.v. gekeken'.

Tip 5: Stel enkelvoudige vragen
Hoewel het gemakkelijk is om te controleren of een vraag wel of niet enkelvoudig is ziet men toch vaak meervoudige vragen waarbij meer dan een onderwerp tegelijkertijd ter sprake komt. Evenals bij de controle of u algemene termen in uw vragen gebruikt moet u zich bij de geformuleerde vragen afvragen wat het antwoord 'ja' of 'nee' betekent. Wat bijvoorbeeld zegt het antwoord 'nee' op de volgende vraag 'stelt u er prijs op om via themadagen of een cursus verder over het onderwerp gebouwde omgeving en criminaliteit geïnformeerd te worden'?[2]

Wil degene die 'nee' antwoord nu:
- helemaal geen informatie hebben?
- geen informatie via themadagen of een cursus hebben?
- of wel informatie maar niet via een cursus?
- of wel informatie maar niet via themadagen?

Een correcte vraagstelling had kunnen zijn:
- Stelt u er prijs op over het onderwerp gebouwde omgeving en criminaliteit geïnformeerd te worden? ja/nee
- Zo ja, op welke wijze wilt u dat
  - via themadagen? ja/nee
  - via cursussen? ja/nee
  - anders nl. ..? ja/nee

Tip 6: Formuleer positief
Wees u er van bewust dat bij negatief geformuleerde vragen het antwoord 'nee' gegeven kan worden in de zin van 'inderdaad nee' of 'nee, juist wel'. De vraag 'Vindt u dat de politie niet had moeten ingrijpen?' kan op die twee neemanieren worden beantwoord. De vraag kan dan beter zijn 'De politie heeft ingegrepen. Vindt u dat een goede of een slechte handelwijze?'

Tip 7: Geef geen aanleiding tot sociaal wenselijke antwoorden
Sociaal wenselijke antwoorden zijn antwoorden die niet zozeer op basis van de eigen mening worden gegeven, maar meer op basis van wat men denkt dat men 'behoort' te antwoorden. Dat 'behoort' kan afgeleid worden van wat men denkt, dat de vraagsteller graag wil horen of wat men in het maatschappelijk verkeer als 'behoorlijk' veronderstelt. Als een interviewer een vraag stelt als' 'U bent toch tegen racisme?' dan verwacht hij via 'toch' eigenlijk een positief antwoord. Op een algemene vraag als 'Bent u voor of tegen de doodstraf' zal men in Nederland als maatschappelijk 'behoorlijk' een 'ik ben tegen' antwoord veronderstellen. In de V.S. zal dit anders liggen. Het oproepen van sociaal wenselijke antwoorden ligt vaak heel subtiel.

Tip 8: Gebruik een oneven aantal antwoordcategorieën
Bij een oneven aantal antwoordmogelijkheden (vanaf drie) is er altijd een midden dat als neutraal punt kan dienen. Bij de antwoordmogelijkheden heel goed – goed – slecht – heel slecht – mist u het neutrale punt, terwijl een neutraal antwoord meestal tot de mogelijkheden behoort. Er zijn echter ook mensen die het gebruik van het neutrale punt niet voorstaan, omdat de ondervraagden te snel op dat 'veilige' neutrale punt zouden gaan zitten.

Tip 9: Gebruik even zware antwoordcategorieën
Waar van toepassing moet men zorgen, dat de

[2] Overgenomen uit een vragenlijst naar aanleiding van een themadag voor de politie in 1987.

antwoordcategorieën even zwaar zijn. Een slechte reeks is bijvoorbeeld heel goed – goed – noch goed/noch slecht – minder goed – slecht.
De categorie 'slecht' is in gevoelswaarde niet even zwaar als 'heel goed'. De categorie 'minder goed' past niet erg in zwaarte bij de categorie 'goed'. Beter zou zijn: heel goed – goed – noch goed/noch slecht – slecht – heel slecht.

Tip 10: Doe niet te veel een beroep op iemands geheugen

Vragen als 'Bent u de afgelopen drie jaar op het politiebureau geweest?' zijn niet zo eenvoudig te beantwoorden als het lijkt. Die ene keer, dat u aangifte deed van de diefstal van uw fiets was dat nu vijf of vier jaar geleden of maar twee jaar terug?

Tip 11: Test uw vragenlijst

Het maken van (ongewilde) fouten ligt vaak heel subtiel. U doet er goed aan uw conceptvragenlijst eerst uit te testen. U kunt hiervoor uw collega's gebruiken. Het liefst moet u de test doen bij een aantal mensen uit de groep die u wilt gaan ondervragen. Geef daarbij een kopietje van de voorgaande tips en laat ze aan de hand daarvan kritisch kijken.

Het probleem van de non-respons

U heeft tenslotte een goede vragenlijst ontwikkeld. Met de vragenlijst gaat u aan de slag. Meestal gaat u niet de hele groep die u wilt onderzoeken benaderen, dat is te kostbaar. U laat een steekproef samenstellen. Als iedereen uit de steekproef de vragenlijst beantwoordt zijn er geen problemen. De ervaring leert echter dat zoiets vrijwel nooit gebeurt. Er zijn altijd mensen, die geen zin hebben om een vragenlijst in te vullen of op welke wijze dan ook geen medewerking verlenen aan onderzoeken. Bij een telefonische enquête overdag zult u een aantal mensen niet thuis treffen. Denk maar aan tweeverdieners.

U krijgt dus vrijwel altijd minder binnen dan u uitgezet heeft. Laten wij aannemen dat u een respons hebt van 60%. U bent dik tevreden. De keerzijde is echter dat 40% niet heeft geantwoord. U bent er dan niet, als u dat alleen maar aangeeft. U kunt weliswaar het geluk hebben dat de 40% niet-antwoorders in feite weer een steekproef uit de steekproef is, maar u kunt ook de pech hebben dat het helemaal niet zo is en dat die 40% niet-antwoorders juist een speciale categorie vormen.
Bij een enquête over onrustgevoelens, die thuis wordt afgenomen door interviewers zou het wel eens zo kunnen zijn dat die 40% niet-antwoorders juist de angstige mensen betreffen. Die laten niet graag wildvreemden binnen, al kunnen zij aantonen voor een enquête te komen. Als u er zich niet van bewust bent dat dit speelt krijgt u dus een te zonnig beeld. Immers vooral de minder angstigen hebben aan de enquête meegedaan.

Er wordt met de enquête over de onrustgevoelens een voorbeeld gegeven, waarbij het non-respons probleem vrij eenvoudig te achterhalen valt. Maar meestal kost het een hoop denken zoekwerk voor dat u kunt vaststellen of de non-respons groep een uitzonderlijke groep vormt, waardoor de resultaten beïnvloed kunnen worden. Wees u er in ieder geval bewust van, dat het non-respons probleem kan spelen en dat het om de vraag gaat of de respondenten een goede afspiegeling vormen van de totale groep.

Inleiding

Bij de statistische bewerkingen is het gebruik van een calculator onontkoombaar.
Calculatoren zijn in allerlei soorten en maten te koop. De nieuwste trend is de databank, die naast de mogelijkheid van het opslaan van allerlei gegevens (in letters en/of cijfers) ook een calculatorgedeelte heeft.
Bij het samenstellen van het hoofdstuk 'Werken met cijfers' was het uitgangspunt dat de bewerkingen uitgevoerd moesten kunnen worden met een eenvoudige calculator op creditcardformaat (verder te noemen creditcardcalculator). **De enige eis was dat daar wel een toets met het √ teken op voorkomt.** (Een aantal vooral oudere creditcard-calculators heeft dat teken niet. Op sommige recente creditcard-databanken komt het teken ook niet voor).
Dit soort calculators kunt u zich voor enkele guldens aanschaffen.
Er is een snelle methode om te zien of u met een zeer simpele calculator te maken heeft of met één van een hoger niveau. Als u de volgende bewerking uitvoert:

123:789 = zal het antwoord zijn
0,1558935 (bij 8-cijferige display).

Als u dit antwoord weer x789 doet zou er 123 uit moeten komen. Dit gebeurt alleen bij calculators van een hoger niveau. Onze calculator (met 8-cijferige display) geeft als antwoord 122,99997. Een zeer geringe afwijking maar toch!
Een calculator is opgebouwd uit een display (beeldscherm), uit een aantal cijfertoetsen en aantal functietoetsen. In de volgende paragrafen gaan wij daarop in.

Het gebruik van de display

Via het beeldscherm kan men het verloop van de berekeningen volgen. Het is nuttig om elke keer direct na het intoetsen van een getal te kijken of u dat goed gedaan heeft. Op dat moment is het nog mogelijk een fout te herstellen.

Elke calculator heeft een maximaal bereik. Wordt dit bereik overschreden, dan verschijnt op het beeld een E (van **Error**). Bij een 8-cijferige display krijgt men bij het intoetsen van

789456 x 789456 = E 62324077.

Ook verschijnt er een E als u een bewerking uitvoert die niet mogelijk is, bijvoorbeeld als men drukt 6–8 = –2 en daarna drukt √ (er kan geen wortel getrokken worden uit een negatief getal).
Alle uitkomsten met een E kunt u niet gebruiken. U zult overigens wel merken dat uw calculator geblokkeerd is. U kunt niet verder. U moet eerst de E opheffen. Dit kunt u doen door op de (A)C toets te drukken.

Op het beeldscherm kan ook een M (van **Memory**) verschijnen. Daarmee wordt aangegeven, dat u een bewerking uitvoert waarbij u gebruik maakt van het geheugen (zie voor verdere uitleg 'gebruik geheugengroep').

Als u de bewerking 123:789 = uitvoert is het antwoord 0,1558935 (bij 8-cijferige display) en als u toetst 123456:123= krijgt u 1003,7073 (bij 8-cijferige display). De komma in het antwoord verschuift vanzelf. In het eerste geval was de uitkomst een getal met zeven cijfers achter de komma, in het tweede met vier. De calculator heeft het systeem van de 'drijvende' komma.

Bij het gebruiksklaar maken is de display van belang. De calculator is gebruiksklaar als er op display geheel rechts te lezen is 0.. Als er een getal verschijnt anders dan 0.. druk dan op de (A)C-toets.
Indien er nog een M of een E te lezen valt moet ook die eerst opgeheven worden. Voor de M drukt men op de MC of (A)C-toets, voor de E op de (A)C-toets.

Bij de calculators worden twee energiebronnen gebruikt, nl. batterijen of lichtcellen. Soms is er een combinatie van beiden.
Bij beide soorten kan men een on(aan) en off(uit)-toets aantreffen.
Bij calculators die werken op lichtcellen moet men soms als er geen on/off-toets is op de (A)C-toets drukken voordat men aan de slag kan gaan.

Het gebruik van de cijfer- en functietoetsen

De cijfertoetsen

De cijfertoetsen zullen weinig problemen opleveren. Wil men 789 op de display hebben, dan drukt men achtereenvolgens op de 7, de 8 en de 9. Wil men 7,89 op het scherm hebben, dan drukt men op de 7, de punt (=komma), de 8 en de 9.

De functietoetsen

Groep functietoetsen voor basisberekeningen

: de toets voor het delen 6 : 3 = 2
x de toets voor het vermenigvuldigen 6 x 3 = 18
– de toets voor het aftrekken 6 – 3 = 3
+ de toets voor het optellen 6 + 3 = 9
= hiermee wordt een bewerking afgesloten.

De bewerking is ook in die zin afgesloten dat direct daarna getallen voor een volgende bewerking kunnen worden ingetoetst. Er hoeft niet eerst op de (A)C-toets gedrukt te worden dus:
6 : 3 = 2 6 x 3 = 18 en dus niet 6 : 3 = 2 (A)C 6 x 3 = 18.

: gecombineerd met =
De : toets kan in combinatie met de =toets gebruikt worden om de in de calculator aanwezige mogelijkheid van een constante deling, dus een deling door steeds hetzelfde getal, te benutten. Bijvoorbeeld:

3:2= 1,5	door in te toetsen	3 : 2 =	(4 handelingen)[3]
18:2= 9	door daarna te doen	18 : 2 =	(4 handelingen)
37:2=18,5	door daarna te doen	37 : 2 =	(4 handelingen)
42:2=21	door daarna te doen	42 : 2 =	(4 handelingen)
50:2=25	door daarna te doen	50 : 2 =	(4 handelingen)
Totaal			20 handelingen

Maar het kan ook met de combinatie : =

3:2= 1,5	door in te toetsen	3 : 2 =	(4 handelingen)
18:2= 9	door daarna te doen	18 =	(2 handelingen)
37:2=18,5	door daarna te doen	37 =	(2 handelingen)
42:2=21	door daarna te doen	42 =	(2 handelingen)
50:2=25	door daarna te doen	50 =	(2 handelingen)
Totaal			12 handelingen

Bij de eerste methode waren in het totaal 20 handelingen nodig, bij de tweede 12. U spaart dus tijd. De delingen in het voorbeeld zijn simpel. Het kost echter tijd om steeds het getal 321,123 in te voeren als u daardoor wil delen. Bij de tweede methode sluit u verder intoetsfouten uit omdat u maar één keer het constante getal hoeft in te voeren en niet meerdere keren.

x gecombineerd =
Op ongeveer dezelfde wijze kan men de combinatie X met = bij een vermenigvuldiging met een constant getal benutten. U moet hier alleen uitvinden hoe die in uw calculator zit. Er zijn twee mogelijkheden. Ofwel het eerste getal is constant, ofwel het tweede. Bij de deling is dat altijd het tweede. Proberen dus.
U toetst in:

5 x 3 = antwoord 15 hierna doet u
4 =
2 =

Met een antwoordenrijtje van
4 = 20
2 = 10
is het getal 5 constant (het eerste)

Met een antwoordenrijtje van
4 = 12
2 = 6
is het getal 3 constant (het tweede)

Als het de bedoeling was 5 als constant getal te gebruiken is het bij het eerste rijtje in orde en kan overgegaan worden de x gecombineerd met = methode te gebruiken. Als het tweede rijtje eruit komt is echter de drie constant en moet dit hersteld worden door niet 5 x 3 = in te toetsen maar 3 x 5 =, daarna kan de x gecombineerd met = methode worden toegepast. Er zijn gelukkig maar weinig calculatoren die de uitkomst van het tweede rijtje opleveren maar u kunt altijd die uitzondering zijn.

x gecombineerd met =.
De combinatie x met = kan ook nog voor kwadrateren gebruikt worden. Wil men het kwadraat van 91 berekenen dan zal men normaal gesproken intoetsen 91 x 91 = antwoord 8281. Het kan net iets korter door 91 x = antwoord 8281. Een handeling minder en vermijding van intoetsfouten.

Het gebruik van de geheugengroep
In de calculator zit de mogelijkheid het resultaat van een bewerking 'even opzij te zetten'. Dit kan van belang zijn om bewerkingen met minder handelingen uit te voeren. Maar eerst de diverse toetsen.

M+ Door het indrukken van deze toets voegt men een uitkomst van een bewerking toe aan het geheugen. Er verschijnt een M (Memory) in de display.

[3] Het intoetsen van een getal uit hoeveel cijfers dat ook bestaat wordt beschouwd als één handeling.

M– Door het indrukken van deze toets trekt men een uitkomst van een bewerking af van het geheugen. Er verschijnt een M in de display.

MR soms RM R van Read M van Memory Door het indrukken van deze toets leest men het eindresultaat af van het optellen c.q. aftrekken van de bewerkingsuitkomsten via de M+ c.q. M– toets.

MC soms CM C van Clear Door het indrukken van deze toets wordt het geheugen schoon gemaakt. Soms ziet met een toets MR/c. Een keer drukken laat het eindresultaat zien, nogmaals drukken wist het geheugen schoon. Als er geen MC/CM toets is, kan er gewist worden via de A(C) toets.

Het handigst is het gebruik van het geheugen bij zgn. ketting-berekeningen zoals (6x9) + (8x10) + (9+30) + (8x16) = 289. Zonder gebruikmaking van het geheugen zou men de volgende handelingen moeten doen:

intoetsen 6 x 9 = antwoord 54 noteren
(5 handelingen)
intoetsen 8 x 10 = antwoord 80 noteren
(5 handelingen)
intoetsen 9 x 3 = antwoord 27 noteren
(5 handelingen)
intoetsen 8 x 16 = antwoord 128 noteren
(5 handelingen)
daarna intoetsen 54 + 80 + 27 + 128 = antwoord 289 (8 handelingen). In het totaal zijn er 28 handelingen geweest. Met het gebruik van het geheugen worden dat er veel minder. Het gaat als volgt:

intoetsen 6 x 9 M+ antwoord M 54
(4 handelingen)
intoetsen 8 x 10 M+ antwoord M 80
(4 handelingen)
intoetsen 9 x 3 M+ antwoord M 27
(4 handelingen)
intoetsen 8 x 16 M+ antwoord M 128
(4 handelingen)
intoetsen MR/RM of Mr/c antwoord M 289
(1 handeling)
In het totaal dus 13 handelingen in plaats van 28!

Let op 1: het drukken op M+/M– toets heeft twee functies : enerzijds wordt het resultaat in het geheugen ingevoerd anderzijds sluit het de berekening af als ware er op de = toets gedrukt. Dat hoeft u dus niet nog eens te doen.
Let op 2: na het uitlezen van het eindresultaat moet u eerst het geheugen schonen voordat u er verder gebruik van gaat maken voor andere bewerkingen.

Het gebruik van de (A)C en C(E) groep

(A)C
Met deze toets worden alle voorgaande resultaten van bewerkingen gewist. Blijft er een M op de display staan, dan zit er nog wat in het geheugen dat nog gewist moet worden via de MC/CM of MR/C toets.
Nu eens vindt men een AC toets dan weer een C toets

C (E)
(E van Entry).
De C(E) toets kan men gebruiken om het laatst ingevoerde getal te corrigeren. Dit moet gedaan worden voordat andere toetsen worden ingedrukt of van het geheugen gebruik wordt gemaakt. Men wil bijvoorbeeld intoetsen 6 x 9 =. Per abuis toetst men 6 x 8 in. Op dit moment kan met de C(E) toets de vergissing gecorrigeerd worden dus 6 x C/E (op de display komt nu 0 het foutieve getal is gereduceerd tot 0). Nu kan de 9 weer ingevoerd worden. De C/E toets is vooral gemakkelijk als er een fout wordt gemaakt bij het eind van een lange reeks bewerkingen. U hoeft dan niet alles overnieuw te doen.

Het gebruik van de groep bijzondere toetsen
De enige bijzondere toets van belang voor ons op de zeer eenvoudige creditcard-calculator is de √ toets. Wij hebben die straks voor een aantal berekeningen nodig. Men toetst eerst een getal in en gebruikt dan de √-toets. Een druk op = toets is niet noodzakelijk. Het getal moet wel positief zijn. Probeert men het bij negatieve getallen dan krijgt men E.

Waarschuwing
Indien u niet een calculator van het zeer eenvoudige type gebruikt wees er dan op attent, dat het voorgaande niet zonder meer van toepassing is op calculators van een hoger niveau. U zult echt de gebruiksaanwijzing moeten consulteren als u merkt dat iets niet lukt, zoals hiervoor is beschreven.

Het ordenen van cijfers in tabellen en grafieken

Inleiding
Nadat u aan het verzamelen bent geweest heeft u de bekende 'brij' aan gegevens voor u liggen. De 'brij' is vaak ongeordend. Het komt maar zelden voor dat u gegevens hebt die u zonder meer voor uw doeleinden kunt gebruiken.

U moet ze gaan ordenen. Omdat u met de geordende gegevens iets wilt bereiken, zorgt u ervoor dat u de geordende gegevens overzichtelijk weergeeft. Daarvoor gebruikt u tabellen en grafieken.

Het ordenen in tabellen

Het ordenen van cijfers geschiedt via zgn. frequentietabellen. Daarin wordt het aantal malen dat een bepaalde waarneming voorkomt (frequentie) weergegeven.

De basis van een tabel is een rechthoek die getallen bevat en die verdeeld is in kolommen en regels (schema 5).

Schema 5: Schematische opbouw van een tabel

	K	K	
	O	O	REGEL
	L	L	REGEL
	O	O	REGEL
	M	M	

Er zijn frequentietabellen zonder klassen en met klassen.

Voorbeeld 1: frequentietabel zonder klassen
In een door winkeldiefstallen geplaagd warenhuis wordt een tijdje bijgehouden hoeveel winkeldiefstallen er per dag worden vastgesteld. Het resultaat is de volgende reeks.
7 6 5 8 6 5 5 6 7 6 9 4 5 6 6 3 7 8 8 7 7 6 7 5 6

Deze reeks is niet erg overzichtelijk, er moet geordend worden. Wij maken eerst een turftabel.

Aantal winkeldiefstallen per dag	Geturfd	Frequentie
10		0
9	I	1
8	III	3
7	I𝍸	6
6	II𝍸	8
5	𝍸	5
4	I	1
3	I	1
2		0
Totaal aantal frequenties		25

De frequentietabel is dezelfde maar zonder de kolom 'geturfd'.

Voorbeeld 2: frequentietabel met klassen
Bij een snelheidsmeting van motorvoertuigen binnen de bebouwde kom werden de volgende 40 waarnemingen van de snelheid in km/h genoteerd.
47 63 65 63 78 53 60 62 59 75 55 87 63 57 58 35 90 58 80 60 57 58 63 54 53 58 57 62 59 65 50 68 65 30 70 77 59 60 73 33
Als u hiervan een gewone frequentietabel maakt, dan wordt het er niet duidelijker op, omdat het aantal waarnemingen te groot is om een overzichtelijke tabel te krijgen. Immers men zou een cijferreeks van 30 t/m 90 krijgen waarbij sommige waarden 'leeg' blijven. In zulke gevallen is het verstandiger de resultaten te groeperen in **klassen**.

Algemene tips bij het maken van klassen zijn:
- gebruik niet te veel klassen
- deel de klassen in met hele getallen
- zorg voor een ondubbelzinnige indeling
- zorg voor logische indeling.

Met deze aanwijzingen gaan wij nu de rij waarnemingen uit voorbeeld 2 bekijken.

Tip 1: Gebruik niet te veel klassen
Wij zien in de rij cijfers, dat de laagst gemeten snelheid 30 km/h was en de hoogste 90 km/h. Die waarnemingen bepalen het gebied waarover wij klassen willen maken. Het verschil is 60 km/h. Als wij klassen van 5 km/h maken zijn er minimaal 12 klassen nodig, bij 10 km/h komen wij uit op minimaal 6. Waar wij tenslotte voor kiezen, hangt af van de vraag hoeveel informatie verloren kan gaan en dat heeft weer te maken met de logische indeling waarover zo direct meer.

Tip 2: Deel de klassen in met hele getallen
Het heeft weinig zin om klassen van 5 km/h aan te geven met bijvoorbeeld 40,5–44,5 in plaats van 40–44. Het leest niet eenvoudig. De informatie is niet exacter geklasseerd, in beide gevallen staat er dat er een bepaalde klasse van 5 km/h breed is.

Tip 3: Zorg voor een ondubbelzinnige klasse-indeling
Hoe men de klassen ook indeelt, het mag niet voorkomen dat een bepaalde waarneming in twee klassen geplaatst kan worden. Een indeling met de klassen 0–10, 10–20 enz. is dus fout omdat de waarneming 10 zowel in de ene als in de andere klasse geplaatst kan worden. Wel kan 0–9, 10–19, of 0–< 10, 10–< 20 gebruikt worden (< = minder dan).

Tip 4: Zorg voor een logische indeling
De waarnemingen willen wij straks ergens voor gebruiken. Dat gebruik bepaalt voor een groot gedeelte de logica van de klasse-indeling nu. In ons voorbeeld zijn snelheden gemeten. Het kan zijn dat de politie dit doet om een snelheidscontrole-actie binnen de bebouwde kom voor te bereiden. Het beleid bij die actie is dat iedereen die zich aan de 50 km/h houdt een attentie krijgt. Dat met degenen die harder dan 50 km/h maar niet harder dan 60 km/h rijden wordt 'een praatje' gehouden. Tegen degenen die harder dan 60 km/h rijden wordt onverbiddelijk een proces-verbaal opgemaakt. Bij die ideeën zou een logische klasse-indeling zijn:

Gemeten snelheden op de Beestweg
20 januari 1994

Snelheid in km/h	Frequentie
20–30	1
31–40	2
41–50	2
51–60	17
61–70	11
71–80	6
81–90	1
Totaal aantal frequenties	40

Is het echter de bedoeling om te zien of een bepaalde weg binnen de bebouwde kom in aanmerking komt voor een bord maximum 70 km/h en de gemeten snelheden bij de beoordeling daarvan een rol spelen, dan zou de volgende klasse-indeling logisch zijn:

Gemeten snelheden op de Beestweg
20 januari 1994

Snelheid in km/h	Frequentie
minder dan 50 km/h	5
meer dan 50 km/h	28
meer dan 70 km/h	7
Totaal aantal frequenties	40

Voorbeeld 3: Frequentietabel met cumulatieve verdelingen
In sommige gevallen is een frequentietabel handig die cumulatieve verdelingen aangeeft. Hierbij worden de frequenties van de opeenvolgende klassen opgeteld bij die van de vorige klasse. De frequenties worden als het ware gestapeld (gecumuleerd). In het voorbeeld vlak hiervoor van de tabel met de klasse-indeling die gebruikt kon worden voor de voorbereiding van de politieactie krijgen wij dan het volgende:

Gemeten snelheden op de Beestweg
20 januari 1994

Snelheden in km/h	Frequentie	Cumulatieve frequentie
20–30	1	1
31–40	2	3*
41–50	2	5
51–60	17	22
61–70	11	33
71–80	6	39
81–90	1	40
Totaal aantal frequenties		40

*Dus frequentie klasse 20–30 + frequentie klasse 31–40 + enz.

In de laatste klasse (81–90) moet bij de kolom cumulatieve frequentie dan altijd aan het eind het totaal aantal frequenties komen te staan (40). De cumulatieve frequentie wordt vooral in percentages gegeven.
Een tabel met de cumulatieve frequenties in % is handig als men **snel** wil zien hoe groot het aantal frequenties is dat onder of, boven een bepaalde waarde ligt.

Dus:

Gemeten snelheden op de Beestweg
20 januari 1994

Snelheden in km/h	Frequentie	Frequentie in %	Cumulatieve frequentie in %
20–30	1	2,5	2,5
31–40	2	5	7,5
41–50	2	5	12,5
51–60	17	42,5	55,0
61–70	11	27,5	82,5
71–80	6	15	97,5
81–90	1	2,5	100
Totaal aantal frequentie	40		100%

Vanuit deze tabel is het gemakkelijk te zien dat slechts 12,5% van de gemeten voertuigen 50 km/h of minder reed, 55% niet harder dan 60 km/h en 17,5% harder dan 71 km/h.

In het voorgaande is ingegaan op hoe u via turftabellen naar frequentietabellen gaat. In deze paragraaf wordt nader ingegaan op het maken van tabellen. Zoals reeds gezegd is een tabel een rechthoek waarin getallen staan en die verdeeld is in kolommen en regels.

Dat is de meest simpele tabel. Een wat ingewikkelder tabel ziet er anders uit. Hierna wordt een voorbeeld van zo'n tabel gegeven. U moet dan allerlei kolommen en regels goed gaan benoemen.

Schema 6: Model van een tabel

Titel

	Verklaring van b			Verklaring van c			
Verklaring van a	onderverdeling van b1	onderverdeling van b2	onderverdeling van b3	onderverdeling van c1	onderverdeling van c2	onderverdeling van c3	Totaal
a1							totaal a1
a2							totaal a2
a3							totaal a3
a4							totaal a4
a5							totaal a5
totaal	totaal b1	totaal b2	totaal b3	totaal c1	totaal c2	totaal c3	generaal

REGELS

KOLOMMEN

Bron:

Toelichting

Titel: Hier komt uw creativiteit van pas. De verklaring moet kort maar toch volledig zijn. In ieder geval moet blijken het wat, het waar en het wanneer. Een kop in romanvorm is niet gewenst.

a, a1, a2, a3, a4, a5: De kolom waar deze vakjes in voorkomen wordt de voorkolom genoemd. In deze kolom worden in a1, a2, a3, a4 en a5 de categorieën weergegeven waarin de getallen per regel zijn verdeeld, bijvoorbeeld de klasse-indeling van de rijsnelheden (0-50, etc.). In het bovenste vakje wordt een korte verklaring van de categorieën gegeven, in het voorbeeld 'Rijsnelheden 1993 in km/h'.

b en c. In deze hokjes wordt aangegeven waar de verticale getallenreeksen op slaan. Zo zou de b 'Amsterdam' en c 'Den Haag'. Via de hokjes b1, b2 en b3 kan Amsterdam weer onderverdeeld worden in Amsterdam-centrum, Amsterdam-noord, Amsterdam-zuid, via c1, c2 en c3 Den Haag in Den Haag-centrum, Den Haag-noord, Den Haag-zuid.

lege hokjes. Dit zijn de cellen.

bron: Meestal zijn de cijfers niet zelf verzameld. Dan moet de gebruikte bron zo mogelijk met vermelding van een jaartal worden vermeld.

Een geheel ingevulde tabel volgens het beschreven model zou kunnen zijn:

Schema 6a: Gemeten rijsnelheden van personenauto's in een aantal Amsterdamse en Haagse wijken in 1993

Rijsnelheden 1993 in km/h	Amsterdam			Den Haag			Totaal
	centrum	noord	zuid	centrum	noord	zuid	
0–50	50	25	30	40	30	25	200
51–60	70	80	50	60	70	55	385
61–70	10	33	30	8	10	25	116
71–80	5	15	5	2	8	4	39
81–90	–	10	10	–	4	10	34
Totaal	135	163	125	110	122	119	774

Bron: gemeentelijke statistieken Den Haag en Amsterdam 1994

U ziet in de tabel twee cellen met een –. Dat betekent dat er geen waarnemingen in die klasse rijsnelheden zijn geweest. Er zijn een aantal afspraken voor bijzondere gevallen in de cellen.

- een • wordt neergezet, als het gegeven ontbreekt, maar wel kan voorkomen (In onze tabel zou een • bijvoorbeeld kunnen betekenen dat de snelheidsmeetapparatuur op een bepaald moment het niet deed).
- niets wordt neergezet als het gegeven logischerwijs niet kan voorkomen.
- een * wordt neergezet als het gegeven voorlopig is
- een - als het gegeven exact nul is
- een 0 als het gegeven afgerond is op nul
- een x als het gegeven wel voorhanden is, maar geheim.

Een veel gehoorde term is het aantal **ingangen** van een tabel. Hiermee wil men alleen maar aangeven via hoeveel aspecten de getallen via de kolommen en regels zijn bekeken. In onze ingevulde tabel hebben wij drie aspecten: de gemeten rijsnelheden 1993 in km/h, (twee)steden en (drie)wijken. In het totaal zijn dus drie ingangen gebruikt. U kunt een bepaalde cel alleen maar 'bereiken' via de drie ingangen. Een cel kan zijn rijsnelheid (1) 81–90 km/h, in de stad (2) Den Haag en in de wijk (3) Zuid. De frequentie in die cel is 10. Men kan tabellen met zoveel ingangen maken als men wil. Te veel ingangen betekent echter een te onoverzichtelijke tabel. Als u veel ingangen heeft kunt u beter een aantal aparte tabellen maken waarin de informatie gedoseerd wordt gegeven.

Computertabellen 5.3

In de praktijk zult u steeds meer te maken krijgen met tabellen uit de computer.
De opzet is iets anders dan een normale tabel. Om de opzet te begrijpen gaan wij eerst naar een 'gewone' tabel met willekeurig gekozen cijfers kijken.

	kolom 1	kolom 2	totaal
regel 1	10	15	25
regel 2	20	35	55
totaal	30	50	80

De cel met het getal 10 heeft op drie manieren invloed in de tabel. Naar kolom gezien is er een bijdrage voor het totaalgetal 30, naar regel gezien een bedrage voor het totaalgetal 25 en tenslotte draagt het getal 10 bij aan het totaal generaal 80. Die diverse bijdragen kan men in procenten uitdrukken. Voor de ingang naar kolom is 10, 33% van 30, voor de ingang naar regel is 10, 40% van 25, voor de ingang naar totaal is 10, 13% van 80. Op dezelfde wijze kunnen wij naar de andere cellen kijken. Wij bekijken nu met deze wetenschap een voorbeeld van een computertabel.

		AUTO	Page 1 or 1	
	Count Row Pct Col Pct Tot Pct	ja	nee	Row
Functie		1.00	2.00	Total
medisch	1	5 71.4 3.6 2.6	2 28.6 1.0	7 3.6 3.6
paramedisch	2	26 76.5 18.6 13.3	8 23.5 14.3 4.1	34 17.3
verpleegkundig	3	60 75.9 42.9 30.6	19 24.1 33.9 9.7	79 40.3
civiel/technisch	4	14 58.3 10.0 7.1	10 41.7 17.9 5.1	24 12.2
administratief	5	26 66.7 18.6 13.3	13 33.3 23.2 6.6	39 19.9
stafdienst	6	9 69.2 6.4 4.6	4 30.8 7.1 2.0	13 6.6
	Column Total	140 71.4	56 28.6	196 100.0

Number of Missing Observations: 12

In de kop zien wij staan:

Count is het aantal waarnemingen.
Row PCT is hetzelfde als de percentuele bijdrage van een waarneming in het regeltotaal.
Col PCT is hetzelfde als de percentuele bijdrage van een waarneming in het kolomtotaal.
Tol PCT is hetzelfde als de percentuele bijdrage van een waarneming in het totaalgeneraal.

Als wij nu de cel 'medisch ja' bekijken zien wij

		ja
		1.00
	1	5
medisch		71.4
		3.6
		2.6

In de ene cel is alle informatie die in de kop staat (count, row pct, col pct en tot pct) tegelijkertijd vermeld

	ja	
	1	5 (count)
medisch		71.4 (row.pct)
		3.6 (col. pct)
		2.6 (tot. pct)

Dus 5 is 71.4% van 7 (7 = row total = totaal van de regel zie voorbeeld tabel)
5 is 3.6% van 140 (140 = colum total = totaal van de kolom zie voorbeeld tabel)
5 is 2.6% van 196 (196 = totaal-generaal zie voorbeeld tabel)
Op dezelfde wijze moeten de andere cellen worden gelezen. Het cijfer 1 bij 'medisch' is het regelnummer. In de voorbeeld tabel zijn er 6 regels. Het cijfer 1.00 bij 'ja' geeft het kolomnummer weer. In de voorbeeldtabel zijn er twee kolommen.

U ziet in de voorbeeldtabel dat bij de getallen in de Row Total en Column Total maar één percentage is vermeld. Dat is logisch: de diverse totalen leveren alleen een bijdrage voor het totaal-generaal. In de tabel worden die weergegeven als percentage van het totaal-generaal.

Onderaan de tabel ziet u 'number of missing observations' staan. Hiermee wordt het aantal weergegeven, waarvan wel een waarneming zou kunnen zijn, maar waarvan dat om de een of andere reden niet gebeurt is. Het kan bijvoorbeeld zijn dat iemand per ongeluk de vraag waarop deze tabel sloeg heeft overgeslagen.

Tabellen omzetten in 'beelden' 5.4

Overzicht mogelijkheden

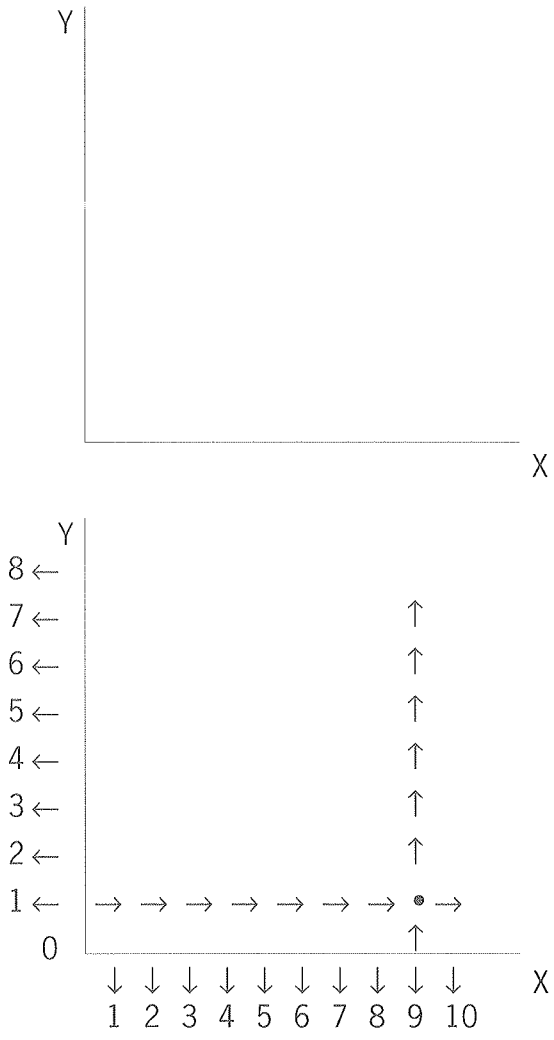
Frequentietabellen geven op zich wel voldoende informatie. De ervaring leert dat voor de meeste lezers het gemakkelijker is als men frequentietabellen in beelden omzet.

De mogelijke beelden zijn:
- staven (histogrammen)
- lijnen (curven)
- vlakke meetkundige figuren: de cirkel
- (thematische) kaarten

Lijnen en Staven

Het omzetten van frequentietabellen in staven en lijnen is in beide gevallen gebaseerd op het grondprincipe van een assenstelsel. Het eenvoudigste bevat twee lijnen die elkaar raken onder een rechte hoek, waar zij elkaar raken is het nulpunt. Er zijn afspraken over hoe wij de lijnen van het stelsel benoemen. De verticale lijn is de Y as, de horizontale de X as. De grondgedachte nu van het stelsel is dat als wij beide assen van een indeling voorzien, wij een plaats binnen rechthoek exact kunnen bepalen door een soort kruismeting (schema 7):

Schema 7: Assenstelsel



Het punt X=9, Y=1 kan nu als volgt bepaald worden. Een punt X=9 ligt in ieder geval op de lijn die van de 9 op X as recht omhoog gaat. Het tweede gegeven is Y=1 en dat ligt in ieder geval op de lijn die horizontaal in een rechte hoek van Y=1 naar rechts vertrekt. Op het kruispunt van die lijnen ligt het punt X=9, Y=1.

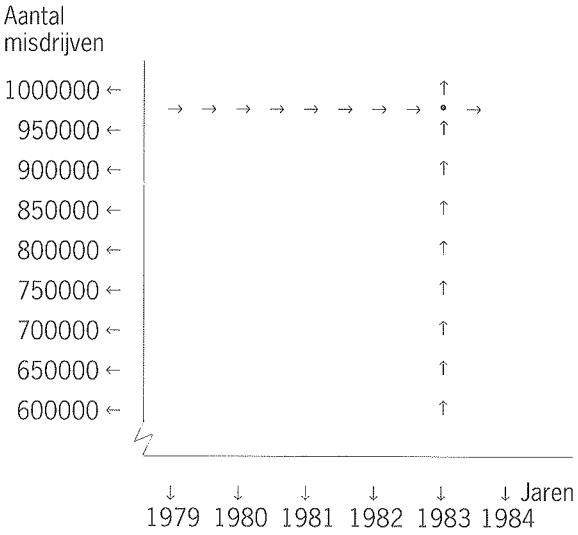
In de volgende tabel worden eigenlijk ook zulke kruispunten gegeven.

MISDRIJVEN IN NEDERLAND

Jaar	Aantal misdrijven
1979	613094
1980	695993
1981	800235
1982	909975
1983	973216

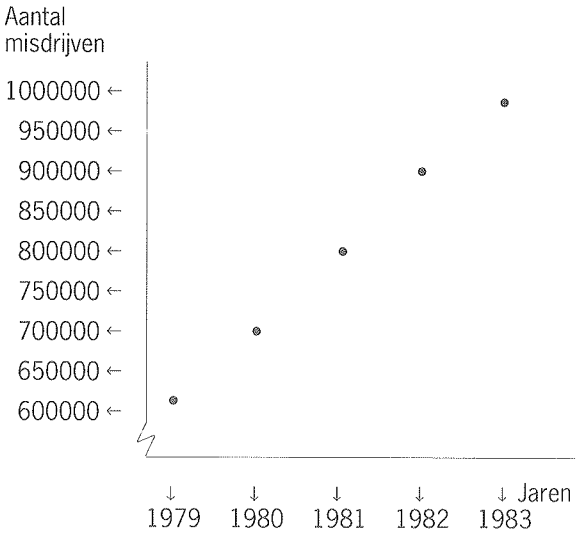
Bron: CBS Statistisch Zakboek, 1985.

Wij gaan die kruispunten dan ook in een assenstelsel zetten. Het **zig-zag**teken in de hierna volgende figuren staat voor 0 tot 600.000. Als wij dit niet deden zouden wij voor informatie die nu niet van belang is toch een indeling moeten maken.



Bron: CBS Statistisch Zakboek, 1989

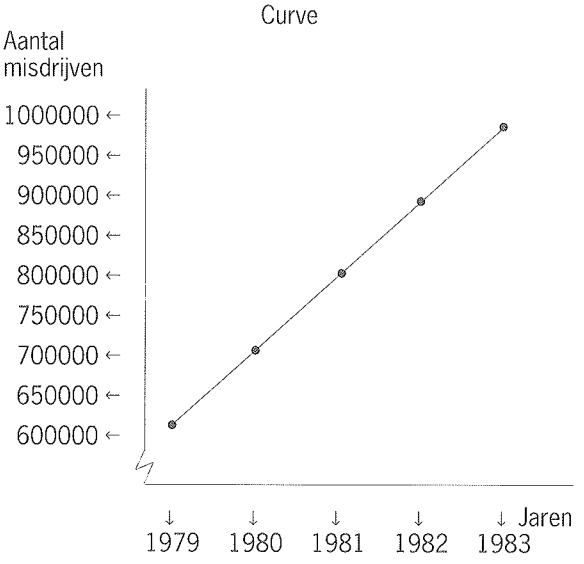
Het kruispunt '1983 – aantal misdrijven' bepalen wij zoals hiervoor met het X en Y voorbeeld is gedaan. Dus 1983 bevindt zich in ieder geval op de lijn die van 1983 op de x-as loodrecht omhoog gaat, 973216 in ieder geval op de lijn die vanaf dat punt op de y-as loodrecht horizontaal naar rechts gaat. Op het kruispunt van de twee lijnen ligt het gezochte punt. Als wij dit voor alle punten doen krijgen wij het volgende



Bron: CBS Statistisch Zakboek, 1985

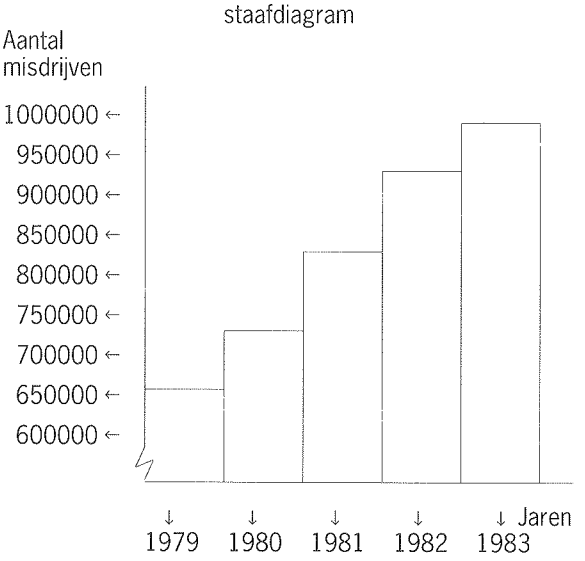
Vanuit deze basisfiguur kan men ofwel met een curve werken ofwel met een staafdiagram.

Curve



Bron: CBS Statistisch Zakboek, 1985

staafdiagram



Bron: CBS Statistisch Zakboek, 1985

In het kader van dit hoofdstuk voert het te ver om verder in te gaan op het hoe en wat van de diverse lijnen en staven. Als u het basisprincipe onder de knie heeft bent u al een heel eind. Wel is het van belang het volgende te weten:

- Een curve wordt gebruikt om een ontwikkeling aan te geven.
- Met een staafdiagram wordt ook een ontwikkeling weergegeven. Bovendien kunnen de staven weer gebruikt worden voor nadere onderverdelingen. In ons voorbeeld zou men per staaf kunnen onderverdelen in lichte en zware criminaliteit.
- Naast voor het aangeven van een ontwikkeling wordt een staafdiagram ook gebruikt om een onderverdeling van een verschijnsel grafisch weer te geven, bijvoorbeeld hoe de verdeling naar haarkleur is bij duizend personen.
- Een histogram is een staafdiagram waarvan de staven altijd tegen elkaar aanliggen. Bij andere staafdiagrammen hoeft dat niet. Een histogram wordt gebruikt om een onderverdeling van een bepaald verschijnsel grafisch weer te geven.

Cirkeldiagram

De cirkel wordt vaak gebruikt om procentuele verdelingen weer te geven. Wij hebben bijvoorbeeld de volgende frequentietabel.

Totaal aantal misdrijven in gemeente Z in 1992

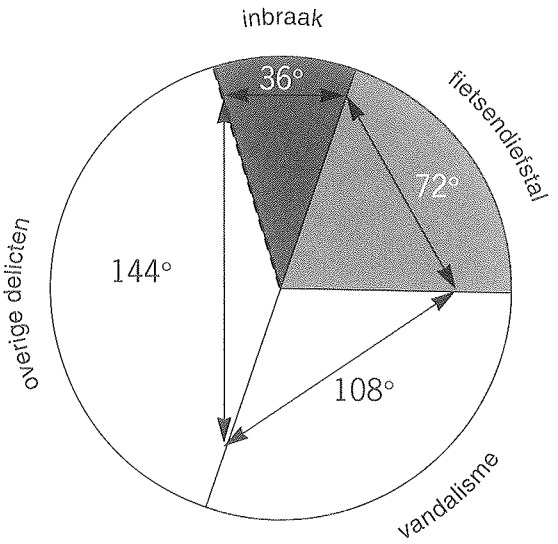
Delict	Frequentie	Frequentie in percentage totaal
Inbraak	50	10
Fietsendiefstal	100	20
Vandalisme	150	30
Overige delicten	200	40
Totaal	500	100%

Nu wil men de laatste kolom in beeld brengen. Dit doet men via het cirkeldiagram. Dat is op een simpele manier te maken. Een cirkel heeft 360% graden. Het delict inbraak zal in de cirkel een segment van 36 (10% van 360) graden zijn, het delict fietsendiefstal 72 (20% van 360) graden het delict vandalisme 108 (30% van 360) graden en de overige delicten 144 (40% van 360) graden
(36 + 72 + 108 + 144 = 360 graden)

Nu moeten wij die segmenten nog intekenen. Wellicht kunt u dit via een computer doen. Als u het helemaal zelf wilt doen, gaat dit met behulp van een gradenboog tamelijk gemakkelijk.

- maak een cirkel
- trek vanuit het middelpunt een lijn naar de buitenkant van de cirkel (hier gestippeld weergegeven)
- deze lijn dient als nulpunt. Teken daarna met behulp van een gradenboog vanaf die lijn het eerste segment in
- neem de tweede lijn weer als nulpunt en teken het tweede segment in met de klokrichting mee.
- neem de derde lijn als nulpunt enz. tot u klaar bent.

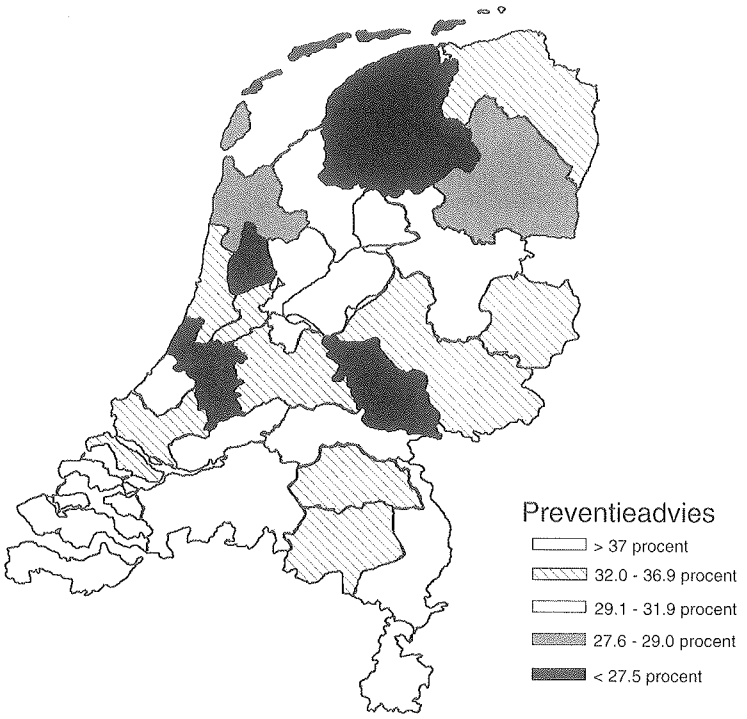
Schema 8: Het opbouwen van een cirkeldiagram

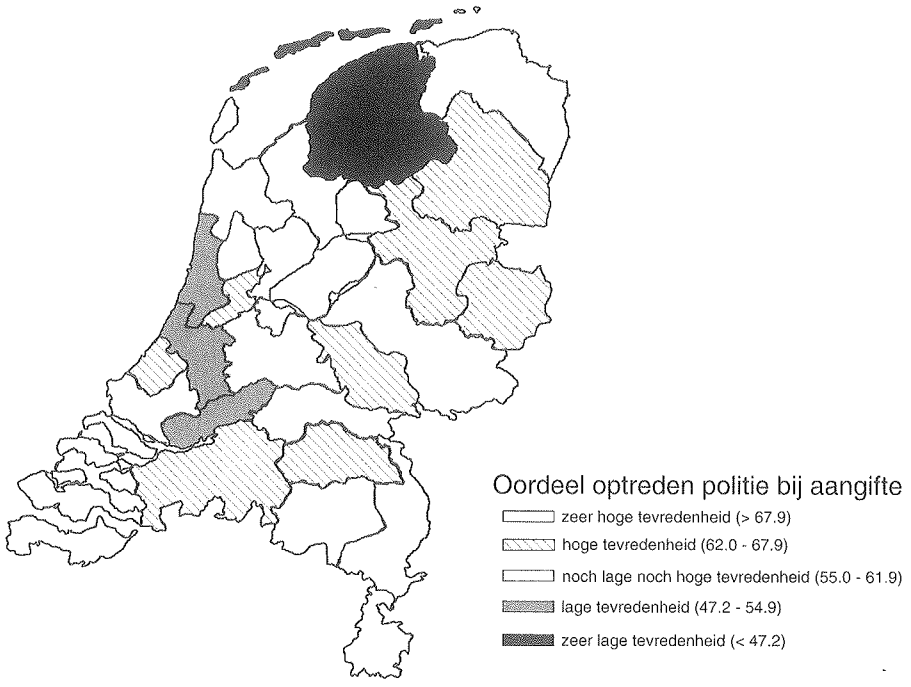


Ook hier wordt volstaan u vertrouwd te maken met het basisprincipe. Door gewoon aan de slag te gaan en bij uw collega's om raad te vragen zult u het maken van cirkeldiagrammen beter onder de knie krijgen, dan dat hier een uitvoerige uitleg volgt over het hoe en waarom en de haken en ogen.

(Thematische) kaarten

Het maken van (thematische)kaarten wordt minder vaak gebruikt, hoewel men daarmee vaak geconstateerde verschijnselen in een bepaald gebied goed in beeld kan brengen, met name als het gaat om de weergave van meer of minder. Het is een goede manier van 'hot spots' weer te geven. Hierna volgen twee voorbeelden van zulke kaarten.





Bron: landelijke rapportage Politiemonitor Bevolking 1993. Ministerie van Binnenlandse Zaken, Ministerie van Justitie, Den Haag 1993.

Bewerkingen[4] in verband met analyses

Inleiding

Na het ordenen komt het analyseren. In de voorgaande paragraaf hebben we de frequentietabel leren kennen als een hulpmiddel om waarnemingsmateriaal te ordenen en in een gecomprimeerde vorm te presenteren wel of niet in 'beelden'.
Een frequentietabel verschaft een snel inzicht in de structuur van de gegevens. Men kan onmiddellijk aflezen tussen welke uitersten de waarnemingsuitkomsten variëren en – wat belangrijker is – men kan gemakkelijk overzien op welke wijze de elementen van het totaal aantal gegevens (de massa) tussen deze uitersten verdeeld liggen. Voor een eerste indruk is dit voldoende, maar er kan meer mee gebeuren. U kunt de gegevens via bepaalde berekeningen meer zeggingskracht geven en kan aangeven wat de betekenis er nu eigenlijk van is.

Het rekenkundig gemiddelde

Vaak wil men een beslissing baseren op het belangrijkste of meest voorkomende aspect (centrale tendentie) van geordende reeksgegevens. Dit doet men dan via het rekenkundig gemiddelde: RG.
Het RG is één getal, dat een indruk geeft van wat in een verdeling het belangrijkst is.
Het berekenen is erg eenvoudig: het gaat om het delen van alle individuele waarnemingsuitkomsten gedeeld door het totaal aantal waarnemingen.
In een formule $RG = \frac{\Sigma x}{N}$

Onthoudt dat het Σ teken in alle formules betekent 'de som van' en N altijd voor het totaal aantal waarnemingen staat.

Omdat wij een reeks getallen door één getal weergeven gaat er informatie verloren. Wij hebben bijvoorbeeld een klas onder onze hoede. Bij het bepalen van de rapportcijfers zien wij het volgende:

Piet 3,9,3,9,3,9,3,9 $RG = \frac{3+9+3+9+3+9+3+9}{8} = \frac{48}{8} = 6$

Kees 6,6,6,6,6,6,6,6 $RG = \frac{6+6+6+6+6+6+6+6}{8} = \frac{48}{8} = 6$

(totaal aantal waarnemingen van ieder 8)

Rekenmachine: 3+9+3+9+3+9+3+9 = 48 : 8 = 6

Beiden krijgen een 6 op het rapport. De informatie dat Piet briljant maar lui is en Kees ijverig, maar het net kan bolwerken gaat verloren. Dit aspect moet bij het werken met gemiddelden niet uit het oog verloren worden.
Soms is het berekenen van een rekenkundig gemiddelde niet zo eenvoudig als in voorgaand voorbeeld.
Bezien wij bijvoorbeeld de volgende tabel.

Opkomst voorlichtingsavond voor mensen tussen de 20 en 25 jaar in A.

Leeftijd (in jaren)	Aantal bezoekers
20	8
21	10
22	16
23	28
24	30
25	40
Totaal	132

Er zijn nu twee gegevens: de leeftijd in jaren en het aantal mensen van die leeftijd die gekomen zijn. Die beide gegevens moeten vermenigvuldigd worden om de bijdrage daarvan voor Σx te krijgen dus:

8 x 20 = 160
10 x 21 = 210
16 x 22 = 352
28 x 23 = 644
30 x 24 = 720
40 x 25 = 1000

RG = 3086 : 132 = 23,4 (afgerond)

Rekenmachine
8 x 20 = M+
10 x 21 = M+
16 x 22 = M+
28 x 23 = M+
30 x 24 = M+
40 x 25 = M+ MR : $M^R/_C$: 132 = 23,4 (afgerond)

[4] Bij alle rekenvoorbeelden in deze paragraaf is gebrtuik gemaakt van een creditcard-calculator met een 8-cijferig display en waar het getal voor de constante vermenigvuldiging het eerst ingetoetst wordt.

Gebruik van het rekenkundig gemiddelde
Het rekenkundig gemiddelde is in de statistische praktijk een veel gehanteerde waarde. Toch geeft het rekenkundig gemiddelde in een aantal gevallen geen goede indicatie over de ligging van het zwaartepunt in de massa.
Dit kan het geval zijn als het gemiddelde wordt vertekend door een aantal extreme waarden. Als men bij huiszoekingen bij drugshandelaren gemiddeld 100 gram heroïne in beslag neemt, dan zullen de individuele inbeslagnames daarvan bijvoorbeeld tussen de 90 en 110 gram liggen. Als men nu een 'recordvangst' van 20 kg heroïne in één keer doet, dan zal het duidelijk zijn dat die vangst het rekenkundig gemiddelde omhoog zal trekken terwijl in de dagelijkse praktijk inbeslagnames blijven schommelen tussen de 90 en 110 gram. In zo'n geval met extreme waarden moet een andere maat voor de weergave van het zwaartepunt worden gevonden. Men gebruikt dan de **modus**: de waarde van de waarnemingsuitkomst die het meest is voorgekomen. Ter illustratie:

In beslaggenomen heroïne in oktober 1992
Situatie 1: Geen extreme waarde.

gram	aantal keren
91	2
95	3
97	4
103	5
105	6

RG = 100 gr. Dit is goed bruikbaar.

Situatie 2: Met extreme waarde

gram	aantal keren
91	2
95	3
97	4
103	5
105	6
1000	1

RG zou zijn 143 gr. Dit geeft geen goed beeld. Daarom nemen we hier de modus: 105 gr.

Bij verdelingen die niet symmetrisch zijn opgebouwd, heeft men met een soortgelijk probleem te maken. De inkomensverdeling in Nederland is daar een voorbeeld van. Uit het Statistisch Zakboek 1991 ontlenen wij volgende tabel. In de tabel staat de verdeling van het aantal huishoudens naar klassen van het besteedbaar inkomen.

Aantal huishoudens naar klassen van het besteedbaar inkomen 1987

Besteedbaar inkomen x ƒ 1000	aantal huishoudens x 1 mln.
≤ 0	0,03
> 0 – < 10	0,2
10 – < 20	0,78
20 – < 30	1,31
30 – < 40	1,28
40 – < 50	0,88
50 – < 60	0,55
60 – < 70	0,25
70 – < 80	0,15
80 – < 90	0,13
90 – < 100	0,10
≥ 100	0,12

Er is een relatief kleine groep met zeer weinig inkomen, een grote midden groep en daarna een lange staart met de hogere inkomens. Die hogere inkomens zorgen er voor, dat in een andere tabel van het statistische zakboek is te lezen dat het rekenkundig gemiddeld besteedbaar inkomen voor 1987 ligt op ƒ 37.200,–. Het modale inkomen – de naam zegt het al – wordt via de modus gemeten.
Wij moeten dan in deze tabel kijken naar de waarnemingsuitkomst, die het vaakst voorkomt. Het grootste aantal huishoudens ligt in de klasse ƒ 20.000,– < ƒ 30.000,–. De waarden in die klasse liggen in ieder geval onder het gemiddeld besteedbaar inkomen van ƒ 37.200,–.
Bij belastingmaatregelen is de modus een goede maat om aan te geven wat een bepaalde maatregel voor de grootste groep mensen betekent. Als men het rekenkundig gemiddelde zou nemen zou men in bovenstaande tabel niet bij de grootste groep mensen terecht komen.

Het gewogen rekenkundig gemiddelde
Tot nu toe was er sprake van een **ongewogen** rekenkundig gemiddelde: alle waarnemingsuitkomsten worden even zwaar meegeteld (gewogen) om het rekenkundig gemiddelde te bepalen. Daarnaast kennen wij het **gewogen** rekenkundig gemiddelde: elk van de waarnemingsuitkomsten krijgt een bepaald gewicht toegekend. Dit wordt gedaan om verschillen in de bijdrage van bepaalde gegevens bij de berekening van het gemiddelde tot uiting te brengen.

Berekening
Op een school heeft men bij het bepalen van rapportcijfers het volgende systeem. Het eindcijfer van het eerste schoolkwartaal telt één keer mee, dat van het tweede kwartaal twee en dat van het derde kwartaal drie keer. Via

deze gegevens wordt het eindcijfer van de gehele schoolperiode berekend. De gedachte hierachter is dat de stof in het begin van het jaar eenvoudiger is dan aan het eind van het jaar. Bovendien moet men aan het eind van het jaar meer stof beheersen dan aan het begin. Er zijn volgens deze redeneringen verschillen in de bijdrage bij het berekenen van het gemiddelde. Afgerond wordt verder vanaf 0,5 naar boven. Een 3,5 wordt dus een 4; 3,4 geeft een 3 als rapportcijfer. Vóór het systeem werd ingevoerd hanteerde men voor het bepalen van het eindcijfer het rekenkundige gemiddelde van de drie kwartaalcijfers.

Piet behaalde de volgende cijfers voor Economie
1e kwartaal 5
2e kwartaal 5
3e kwartaal 6
Het RG is Σ 5 + 5 + 6 : 3 = 5,33. Op basis van het oude systeem hiervan zou hij een 5 krijgen als eindcijfer. De school kent echter gewichten toe aan de cijfers. Dit moeten wij in de berekening meenemen. Het wordt nu

1e kwartaal telt één keer dus	5 x 1 = 5
2e kwartaal telt twéé keer dus	5 x 2 = 10
3e kwartaal telt drie keer dus	6 x 3 = 18
	Σ 33

Wij kunnen nu niet door drie delen, immers door anders te tellen hebben wij als het ware voor meer waarnemingen gezorgd 1+2+3=6 waarnemingen. Het eindcijfer wordt via het gewogen gemiddelde 33 : 6 = 5,5.
Piet krijgt nu een 6 en gaat daarom nog net over. De algemene formule voor het gewogen rekenkundige gemiddelde is

$$GRG = \frac{\Sigma W^1x + \Sigma W^2y}{W^1 + W^2}$$

Σ betekent 'de som van' of 'sommeer'
W wil zeggen de wegingsfactor (W^1 voor X, W^2 voor Y)
X betekent de waarnemingsuitkomst van 'type 1'
Y betekent de waarnemingsuitkomst van 'type 2' enz.

Rekenmachine: Eerst optellen $W^1 + W^2$ enz.
In ons voorbeeld
1 + 2 + 3 = 6. Deze uitkomst noteren, vervolgens
5 x 1 M+
5 x 2 M+
6 x 3 M+
MR of MR/C : 6 = 5,5

Het voortschrijdend gemiddelde

Het voortschrijdend gemiddelde is een rekenkundig gemiddelde, maar dan op een bepaalde manier berekend. Het voortschrijdend gemiddelde wordt gebruikt als men in een bepaalde reeks waarnemingen schommelingen wil afvlakken.

Berekening

Het principe van het voortschrijdend gemiddelde is eenvoudig. In de hierna staande reeks van 8 waarnemingen kan men de afspraak maken het 1e en 2e kwartaal samen te nemen en daarna door 2 te delen. Het eerste voorschrijdend gemiddelde is dan 250. Dit getal telt men bij het derde kwartaal op en deelt weer door twee. Dat getal telt men op bij het vierde kwartaal en deelt dat door twee. Op deze manier werkt men de rij verder af. Het gaat er dus om dat via een bepaald patroon de rij gegevens wordt aangepakt.

Men zou ook de afspraak kunnen maken: kwartaal 1 + kwartaal 2 + kwartaal 3 delen door drie, dat getal optellen met kwartaal 4, delen door drie, dat getal weer optellen met kwartaal 1 van 1991 enz.
Het gaat dus om de afspraak.

Zie voorbeeld hieronder.

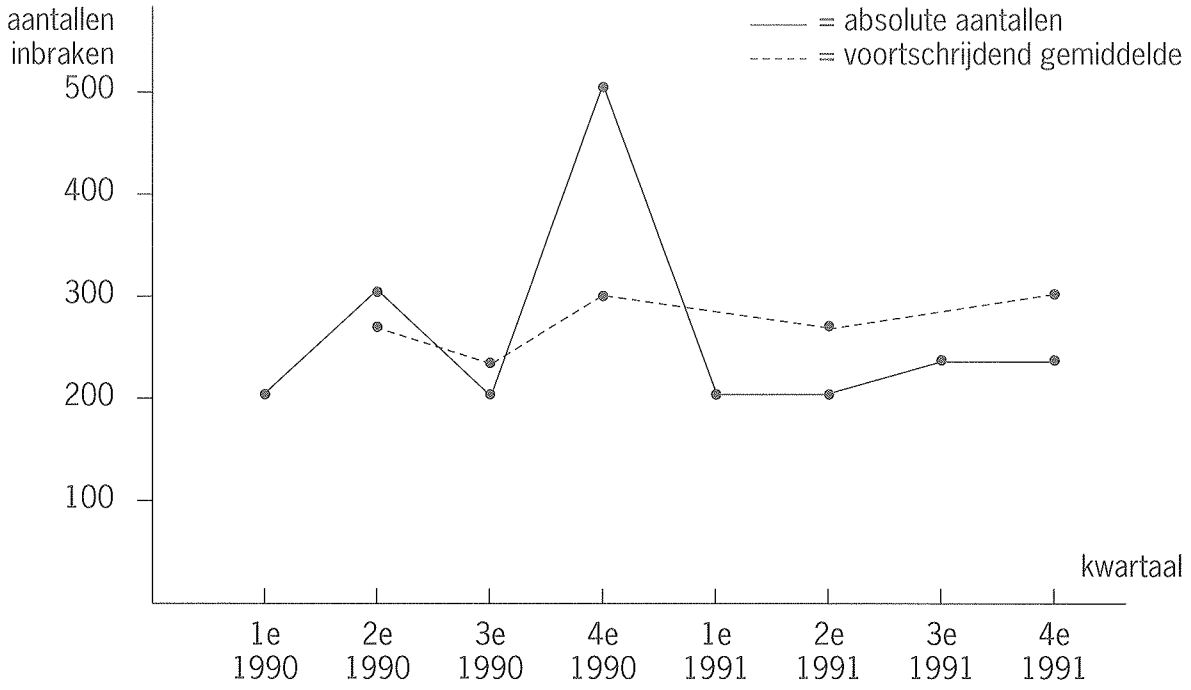
Aantal inbraken over de kwartalen 1990 van 1991 in de stad A

		Berekening	Voortschrijdend gemiddelde
1990	1e kwartaal 200		
	2e kwartaal 300	200+300:2=	250
	3e kwartaal 210	250+310:2=	230
	4e kwartaal 500	230+500:2=	365
1991	1e kwartaal 215	enz.	290
	2e kwartaal 200	enz.	245
	3e kwartaal 325	enz.	285
	4e kwartaal 335	enz.	310

De getallenreeks van het voortschrijdend gemiddelde is minder grillig dan die van de geconstateerde inbraken per kwartaal. Men kan hiermee voorkomen dat er na het vierde kwartaal 1990 een overdreven aandacht voor het delict inbraken na de uitschieter in dat kwartaal komt of dat de aandacht gezien de resultaten van het tweede en derde kwartaal 1991 verslapt. Met het voortschrijdend gemiddelde geeft men meer de algemene ontwikkeling van een verschijnsel over een langere periode weer en legt minder de nadruk op de schommelingen bij de waarnemingen over dat verschijnsel in die periode.

Het werken met het voortschrijdend gemiddelde heeft wel als consequentie dat er voor de eerste waarnemingsperiode geen waarde voor een voortschrijdend gemiddelde kan worden ingevuld. Men moet dus een stuk vòòr de grote schommelingen beginnen met het berekenen van het eerste voortschrijdend gemiddelde.

Wij laten nog op grafische wijze het verschil tussen het verloop van de absolute aantallen en het verloop via de voortschrijdende gemiddelden van het hier gebruikte voorbeeld zien.



6.2 Het analyseren via de spreiding

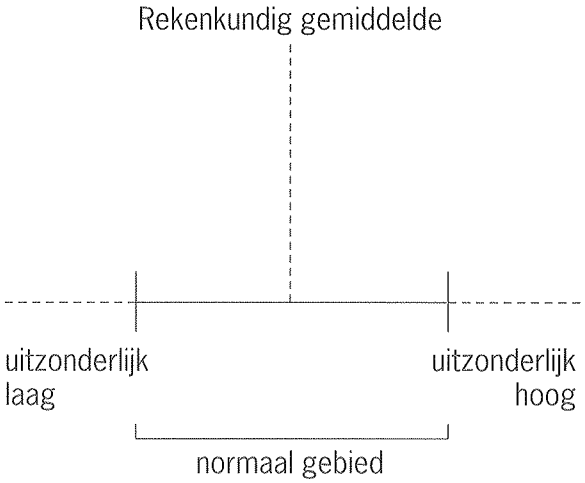
Algemeen

Maatstaven van de centrale tendentie geven aan waar het zwaartepunt van een verdeling ligt. Wij hebben gezien, dat maatstaven voor de centrale tendentie het best gebruikt kunnen worden als er een regelmatige opbouw van de waarnemingen is en als er geen grote extreme uitschieters zijn. Dit ligt anders bij de maatstaven van spreiding. Maatstaven van spreiding geven juist informatie over de onderlinge verschillen tussen de waarnemingsuitkomsten. Er zijn een aantal spreidingsmaten. Hier zal alleen de standaardafwijking worden behandeld.

De standaardafwijking, wat is dat?

Het gaat te ver aan te geven op welke uitgangspunten de berekeningen van de standaarddeviatie gebaseerd zijn. De standaardafwijking (ook standaarddeviatie genoemd) geeft aan in hoeverre waarnemingsuitkomsten van elkaar verschillen en afwijken van het rekenkundig gemiddelde. Voor het dagelijks gebruik is vooral de standaardafwijking van belang om te bezien of bepaalde waarnemingen die wij gedaan hebben uitzonderlijk hoog of uitzonderlijk laag zijn. Beslissingen worden vaak op uitschieters gebaseerd, terwijl die uitschieters eigenlijk (in ieder geval statistisch) niet als uitzonderlijk beschouwd kunnen worden. Met behulp van de standaardafwijking bepalen wij het gebied, dat rond het rekenkundig gemiddelde van een verdeling als 'normaal' is te beschouwen. Normaal wordt hier in statistische zin gebruikt. In beeld gebracht ziet het er zo uit:

Schema 9: De standaardafwijking in beeld



Via de standaardafwijking geeft men het gebied aan **rond** het rekenkundig gemiddelde van een verdeling dat als normaal is te beschouwen. Vandaar dat men ook spreekt van plus en min de standaardafwijking. Dat plus en min is dan gerekend vanaf het rekenkundig gemiddelde. Als men de standaardafwijking uitrekent en daarna met twee vermenigvuldigt en daarna het gebied plus en min vanaf het rekenkundig gemiddelde bepaalt valt ongeveer 95% van de waarnemingen in het als normaal te beschouwen gebied. In de resterende 5% vallen waarnemingen die statistisch gezien uitzonderlijk hoog of uitzonderlijk laag zijn. Bij één keer de standaardafwijking valt 70% van de waarnemingen binnen het normale gebied en 30% is dan uitzonderlijk hoog of uitzonderlijk laag. Om nu in 30% van de gevallen te spreken van uitzonderlijk is minder gewenst, vandaar dat twee keer de standaardafwijking gebruikelijk is.

Een voorbeeld voor het gebruik van een standaarddeviatie

Bij de planning van werkzaamheden kan het van belang zijn om te weten of bepaalde waarnemingen die gedaan zijn er statistisch gezien er uit springen of niet. De nadruk ligt op **statistisch gezien**. De statistiek bemoeit zich alleen om het cijfermateriaal en houdt zich niet bezig met overwegingen van andere (bijvoorbeeld politieke) aard.
In een bepaalde stad vond men de volgende cijfers.

Fietsdiefstallen bij het station in de gemeente A per dag in 1991

Dagen	Gestolen fietsen
maandag	120
dinsdag	150
woensdag	130
donderdag	140
vrijdag	160
zaterdag	180
zondag	170

In het kader van een project vond men de politie bereid één dag in de week te gaan posten bij het station met het oogmerk enkele heterdaadbetrappingen te realiseren. De vraag is nu welke dag dan het beste daarvoor in aanmerking zou komen. Als men de tabel op zich bekijkt zou men vermoedelijk met het gezonde verstand voor de zaterdag kiezen en in ieder geval niet voor de maandag. Dat zijn immers de uitschieters naar boven en naar beneden in de tabel. De korpschef is met de zaterdag echter niet zo gelukkig. In de gemeente vinden een aantal structurele activiteiten (zoals de markt) op zaterdag plaats. Deze vragen al een verhoogde inzet van de surveillancecapaciteit. Het enige, dat hem dan overblijft is aan te tonen dat de waarneming 180 eigenlijk niet als extreem is te beschouwen en dat een inzet op een andere dag dus ook zou kunnen, zelfs op maandag als de 120 ook niet als extreem is te beschouwen.

Berekenen

De korpschef laat van de gegevens de standaarddeviatie berekenen om te zien of hij dat in ieder geval statistisch kan aantonen.
De berekening wordt nu van stap tot stap weergegeven. De standaarddeviatie wordt weergegeven door S of σ. De algemene formule is:

$$S \text{ (of } \sigma) = \sqrt{\frac{\Sigma D^2}{N}}$$

Het teken Σ kwamen wij al tegen het betekent 'de som van'.
Het teken D betekent hier: het verschil van een gevonden waarde met het rekenkundig gemiddelde.
Het teken N kwamen wij ook al tegen het betekent: het aantal waarnemingen.
Wij gaan nu stap voor stap de formule toepassen op ons voorbeeld. Die gaf de volgende waarnemingenreeks:

maandag	120
dinsdag	150
woensdag	130
donderdag	140
vrijdag	160
zaterdag	180
zondag	170

De sleutel tot het berekenen van de standaarddeviatie is het rekenkundig gemiddelde

$$RG = \frac{\Sigma x}{N} = \frac{120+150+130+140+160+180+170}{7} = \frac{1050}{7} = 150$$

rekenmachine:
120+150+130+140+160+180+170=1050:7=150.
Antwoord noteren.

Wij moeten nu D berekenen en daarna D^2.
Bij het berekenen van D krijgt u negatieve of positieve getallen. U kunt echter gewoon het verschil noteren omdat het verschil gekwadrateerd wordt. Daarbij is plus of min niet van belang, de uitkomst is altijd positief. Wij maken nu een overzicht als volgt. De bewerking met de rekenmachine staat er bij.

		RG	D verschil waarneming met RG	D^2	Rekenmachine
maandag	120	150	30	900	120–150=30 x 30=900
dinsdag	150	150	0	0	150–150= 0 x 0= 0
woensdag	130	150	20	400	130–150=20 x 20=400
donderdag	140	150	10	100	140–150=10 x 10=100
vrijdag	160	150	10	100	160–150=10 x 10=100
zaterdag	180	150	30	900	180–150=30 x 30=900
zondag	170	150	20	400	170–150=20 x 20=400
				Σ 2800	

Wij kunnen nu de formule

$S \text{ of } \sigma = \sqrt{\frac{\Sigma D^2}{N}}$ invullen. $S \text{ of } \sigma = \sqrt{\frac{2800}{7}} = \sqrt{400} = 20$

Rekenmachine: 2800 : 7 = 400 $\sqrt{}$ 20
Het als normaal te beschouwen gebied is het rekenkundig gemiddelde plus of min twee keer de standaarddeviatie. Dus:
(RG + 2 S of σ) : 150 + (2 x 20) = 190
(RG–2 S of σ) : 150–(2 x 20) = 110
Het als (statistisch) normaal te beschouwen gebied ligt tussen de waarden 110 en 190.

De betekenis hiervan is dat (**statistisch** gezien) geen enkele waarde in de uitzonderlijk hoog of uitzonderlijk laag categorie valt. Volgens deze uitkomst kan voor het posten elke dag van de week gekozen worden. Noch de waarde 120 noch die van 180 vallen buiten het statistisch als normaal te beschouwen gebied. Voor het gevoel is dat wel zo, voor de statistiek dus niet. De korpschef kan dit inbrengen en trachten met deze berekening andere participanten te overtuigen dat er op elke andere dag dan zaterdag gepost kan worden, omdat de zaterdag hem niet goed uitkomt.

6.3 Het analyseren via indexcijfers, procentuele verdelingen en kengetallen

Algemeen

Het is niet ongebruikelijk dat men een aantal reeksen cijfers onderling met elkaar wil vergelijken. Er worden u bijvoorbeeld vragen voorgelegd als: geef een indruk van de ontwikkeling van het delict fietsendiefstal in vergelijking met dat van autodiefstal; hoe is de ontwikkeling van het aandeel van de delicten fietsendiefstal en vandalisme in het totaal aantal misdrijven, en dalen of stijgen het aantal fiets- en bromfietsendiefstallen nu werkelijk over de afgelopen jaren of niet.

- Voor een indruk van de ontwikkeling van een bepaald verschijnsel kunnen wij gebruik van **indexcijfers** maken.
- Voor het weergeven van de ontwikkeling van het aandeel van een bepaald verschijnsel kunnen wij **percentuele verdelingen** maken.
- Als u de toe- c.q. afname van een bepaald jaar in vergelijking met het voorafgaande jaar wilt aangeven doet u dat met de **percentuele toe- of afname**.

– Bij vragen van daalt of stijgt iets nu werkelijk of niet maken wij gebruik van **kengetallen.**

Indexcijfers

Het idee dat achter de indexcijfers schuilt is simpel. Absoluut cijfermateriaal in een reeks achter elkaar gezet is niet altijd direct op een eenvoudige wijze te overzien, zeker niet als het om de ontwikkeling gaat. Bij bewerkingen met het indexcijfer stellen wij één waarneming op 100 en gaan van daaruit de overige waarnemingen uit de reeks via die 100 uitdrukken. Wij laten eerst zien dat indexcijfers vanuit het oogpunt van overzichtelijkheid ons voordeel bieden. Daarna volgt de uitleg van de berekening.

Fietsendiefstal in Nederland

Jaar	Absoluut	Indexcijfer
1986	167.559	100
1987	163.148	97
1988	172.036	103
1989	193.942	116

Bron: CBS Statistisch Zakboek, 1991

De rij indexcijfers geeft ons inderdaad een makkelijk te overzien inzicht in de ontwikkeling. Wij stelden hier 1986 op 100 en laten de ontwikkeling daarna zien.

De berekening van indexcijfers

Het uitgangspunt is dat wij een getal uit de reeks op 100 stellen en van daaruit de overige getallen uit de reeks via die 100 uitdrukken.

De berekening gaat het eenvoudigst als volgt: wij gebruiken de cijfers uit het voorbeeld van hiervoor.
1986 aantal gestolen fietsen 167.559 = 100
1/100 gedeelte van dat aantal = 1.675,59
als wij nu elk volgend en/of voorgaand getal door dat 1/100 gedeelte delen krijgen wij een antwoord dat uitgedrukt is in die 100 van drie cijfers voor de komma en een aantal achter de komma. Het eerste cijfer achter de komma is hierbij van belang voor het afronden.
Indexcijfers worden met hele getallen aangegeven.
Dus indexcijfer 1987 aantal gestolen fietsen is
163.148 : 1.675,59 = 97,367494 = 97
indexcijfer 1988 aantal gestolen fietsen is
172.036 : 1.675,59 = 102,67189 = 103
indexcijfer 1989 aantal gesloten fietsen is
193.942 : 1.675,59 = 115,74549 = 116

Rekenmachine: (gebruik makend van de constante deling, zie par.4)
163.148 : 1.675,59 = 97,367494 = 97
172.036 : = 102,67189 = 103
193.942 : = 115,74549 = 116

In de praktijk kunt u indexcijfers heel goed gebruiken zoals uit het volgende voorbeeld zal blijken.

U krijgt de volgende twee reeksen onder ogen.

Aantallen inbraken en adviesaanvragen in de gemeente A 1986–1991

	1986	1987	1988	1989	1990	1991
1. inbraakmeldingen	2763	2471	2931	1997	3024	3036
2. adviesaanvragen	712	791	832	833	916	956

Bron: RCID, 1992

De vraag is nu of u snel kunt aangeven of de ontwikkeling van het aantal inbraakmeldingen gelijke tred met het aantal adviesaanvragen houdt of niet. U gaat dat via indexcijfers doen. De cijfers van 1986 stellen wij op 100. Daarna gaan wij per reeks de op 1986 volgende indexcijfers berekenen. U maakt een lijstje.

Inbraak-meldingen	Absoluut	Indexcijfer	Advies-aanvragen	Absoluut	Indexcijfer
1986	2763	100	1986	712	100
1987	2471	89	1987	791	111
1988	2931	106	1988	832	117
1989	2997	109	1989	833	124
1990	3024	109	1990	916	129
1991	3036	110	1991	956	134

Rekenmachine berekening indexcijfers (constante deling):

reeks Inbraakmeldingen 2471 : 27,63 = (antwoord), 2931 = (antwoord), 2997 = (antwoord) 3024 = (antwoord), 3036 = antwoord (A/C)

reeks Adviesaanvragen 791 : 7,12 = (antwoord), 832 = (antwoord), 833 = (antwoord) 916 = (antwoord), 956 = (antwoord)

De vergelijking van de twee rijen indexcijfers maakt ons duidelijk dat de ontwikkeling van inbraakmeldingen minder stijgt dan die van de adviesaanvragen. Dit zou voor de werkzaamheden bijvoorbeeld kunnen betekenen dat als de inbraakmeldingen blijven stijgen er voor een aantal jaren meer beroep wordt gedaan op de adviescapaciteit dan gegeven kan worden, tenzij er personeelsuitbreiding komt!

Waarschuwing 1: De keuze van een bepaald basisjaar kan bepalend zijn voor de waarde, die men gaat geven aan een ontwikkeling. Nemen wij bijvoorbeeld de volgende gegevens uit het Criminaliteitsbeeld van Nederland (DCP 1990).

Diefstal motorfietsen in Nederland

Jaar	Diefstal motorfietsen
1980	1476
1985	1861
1986	2225
1987	1739
1988	1532
1989	1974
1990	1801

De jaren 1980 en 1986 zijn uitspringers. In 1980 word het minste aantal motorfietsen gestolen, in 1986 het meeste in vergelijking met de andere jaren. Kiest men 1980 als basisjaar, dan zijn alle andere aantallen hoger en kan men via de indexcijfers de suggestie wekken van een opgaande lijn. Kiest men daarentegen 1986 als basisjaar dan kan men zeker voor de jaren daarna een dalende tendens suggereren. In tabel gezet:

Jaar	Diefstal motorfietsen	Index 1980=100	Index 1986=100
1980	1476	100	
1985	1861	126	
1986	2225	151	100
1987	1739	118	78
1988	1532	104	69
1989	1974	134	89
1990	1801	122	81

Als u dus alleen maar indexcijfers voorgeschoteld krijgt op basis waarvan men u bepaalde beslissingen wil laten nemen, moet u op uw hoede zijn. Het kan immers zijn, dat er met het basisjaar gegoocheld is om een bepaalde suggestie te wekken.

Waarschuwing 2: Met het berekenen van indexcijfers laat u de informatie over de absolute aantallen verdwijnen. U geeft alleen ontwikkelingen aan. Voor beslissingen zijn echter niet alleen de ontwikkelingen van belang maar ook de absolute getallen. Wij kunnen dit in de volgende tabel, die ontleend is aan Criminaliteitsbeeld van Nederland (DCP 1990), zien.

Ter kennis gekomen vernielingen, absoluut en geïndexeerd, 1986–1989

	1986	1987	1988	1989
beschadiging auto	44.894	48.951	48.595	50.755
vernieling aan openbaar vervoer	1.489	1.119	1.155	1.299
vernieling aan openbaar scholen	12.237	13.643	14.496	16.373
overige vernielingen	45.058	47.106	51.140	55.206
totaal vernielingen	103.678	110.819	115.386	123.633
beschadiging auto	100	109	108	133
vernieling aan openbaar vervoer	100	75	78	87
vernieling aan openbaar scholen	100	111	118	134
overige vernielingen	100	105	113	123
totaal vernielingen	100	107	111	119

Bij de indexcijfers zien wij een behoorlijke groei van de vernielingen aan openbare scholen in vergelijking met de andere categorieën. Als wij de indexcijfers uitsluitend als basis voor een beslissing zouden nemen, komen wij in de verleiding ons op de vernieling aan openbare scholen te gaan fixeren. De absolute cijfers geven echter aan dat het absoluut aantal beschadiging auto in de loop der jaren tussen de 3 à 3,5 keer zo groot is als dat van vernieling aan openbare scholen. Dat gegeven kan ook meespelen bij onze beslissingen. Daarom nu iets over percentuele verdelingen

Percentuele verdelingen

Indexcijfers geven een indruk over het verloop van een bepaald verschijnsel. Dit is niet (altijd) voldoende. De ontwikkeling van een verschijnsel kan op zich verontrustend zijn. Van belang is echter ook of het verschijnsel een grote omvang heeft of niet. Verschijnselen van geringe omvang die snel toenemen, zijn zorgwekkend. Nog erger is het met verschijnselen van grote omvang die snel toenemen. Daarom is er ook behoefte aan informatie over het aandeel van een bepaald verschijnsel in het totaal. Dit kunnen wij doen via percentuele verdelingen.

Berekenen

De berekeningen van een percentuele verdeling lijken op die van een indexcijfer. We laten het even zien.

Fietsendiefstallen bij het station in de gemeente A per dag in 1991

	Gestolen fietsen	Percentage van het totaal
maandag	120	(11) 11
dinsdag	150	(14) 14
woensdag	130	(12) 12
donderdag	140	(13) 13
vrijdag	160	(15) 15
zaterdag	180	(17) 18
zondag	170	(16) 17
totaal	1050	100%

Wij stellen het totaal van 1050 op 100%. Als wij nu 1% deel van dat getal steeds delen op de getallen van de dagen van de week krijgen wij een antwoord van twee cijfers voor komma en een aantal achter de komma. Alleen het eerste cijfer achter de komma is belangrijk voor het afronden.

Rekenmachine (gebruik makend van de constante deling (zie par. 4)

120 : 10,5 = (antwoord) 150 = (antwoord) 130 = (antwoord) 140 = (antwoord) 160 = (antwoord) 180 = (antwoord) 170 = (antwoord).

Wij hebben nu voor elke dag het percentuele aandeel in het totaal van alle dagen tezamen berekend. De gevonden afgeronde uitkomsten staan tussen haakjes vermeld in de tabel onder 'percentage van het totaal'. Let op: als men alle tussen de haakjes staande percentage (11+14+12+13+15+17+16) optelt komt men uit op 98% in plaats van 100%. Dat komt door het afronden. Hoewel het niet strikt noodzakelijk is ervoor te zorgen dat men op 100% uitkomt, oogt het soms beter. Wil men toch op

100% uitkomen, dan moet men als volgt te werk gaan.

Men zou tot op één cijfer na de komma kunnen afronden. Men krijgt dan de reeks 11,4; 14,3; 12,4; 13,3; 15,2; 17,1; 16,2 opgeteld 99,9%. Men is er dan bijna, er mist nog 0.1% om het geheel mooi op 100% te krijgen. Gebruikelijk is om het geheel 'kloppend' te maken door zelf wat te schuiven. Daarvoor is een simpele vuistregel, altijd schuiven in de hoogste aantallen, dan is de fout die men aanbrengt het minst groot. Het veranderen van een percentage van 4 naar 3 (1%) is een fout van 25% van het begingetal. Het veranderen van een percentage van 80 naar 79 (1%) is een fout van 1,25% van het begingetal. In onze voorbeeldtabel hebben wij dan ook een verandering aangebracht in de twee hoogste percentages en zo een en ander kloppend gemaakt.
In het voorbeeld met de afronding met één cijfer achter de komma veranderen wij 17,1% in 17,2% om op 100% uit te komen.

Wij passen het voorgaande toe op de aan het Criminaliteitsbeeld van Nederland (DCP, 1990) ontleende tabel die wij ook bij de indexcijfers tegenkwamen. De uitkomst is als volgt:

Ter kennis gekomen vernielingen, absoluut en in aandeel van het totaal, 1986–1989

	1986	1987	1988	1989
beschadiging auto	44.894	48.951	48.595	50.755
vernieling aan openbaar vervoer	1.489	1.119	1.155	1.299
vernieling aan openbare scholen	12.237	13.643	14.496	16.373
overige vernielingen	45.058	47.106	51.140	55.206
totaal vernieling	103.678	110.819	115.386	123.633
beschadiging auto	43	44	42	41
vernieling aan openbaar vervoer	1	1	1	1
vernieling aan openbare scholen	12	12	13	13
overige vernielingen	44	43	44	45
totaal vernielingen	100%	100%	100%	100%

Rekenmachine: 1986 44.894 : 1036,78 = (antwoord), 1489 = (antwoord), 12.237 = (antwoord), 45.058 = (antwoord), A/C
1987 48.951 : 1108,19 = (antwoord), 1119 = (antwoord), 13.643 = (antwoord), 47.106 = (antwoord), A/C
1988 enz.
1989 enz.

Let op: indexcijfers gaan over de ontwikkeling van **éénzelfde verschijnsel over verschillende tijdsperioden**, procentuele verdelingen gaan over het aandeel van **verschillende verschijnselen** over **eenzelfde tijdsperiode!**

Als wij ons alleen laten leiden door procentuele verdelingen in de hiervoor gegeven tabel, dan zou de beslissing om iets te gaan doen vallen op de categorie 'beschadiging van auto'. De categorie 'overige vernielingen' scoort ook hoog. Het is echter een restcategorie waarin diverse soorten vernielingen zijn ondergebracht. Daarom kiezen wij deze categorie niet. Via de indexcijfers kwamen wij tot een andere conclusie, namelijk ons richten op 'vernieling aan openbare scholen'.

U moet nu niet moedeloos worden en denken wat heb ik nu aan al dat gereken. Eerder gaat het erom dat u diverse zijden van een probleem laat zien, en aangeeft hoe u tot een besluit komt. Het kan zijn dat u ondanks de percentuele verdelingen goede redenen hebt om de ontwikkelingen van de verschijnselen te laten prevaleren bij uw keuze. U moet dan met argumenten komen. Als u door anderen alleen percentuele verdelingen krijgt voorgeschoteld, waarop beslissingen worden gebaseerd, weet u nu dat het de moeite loont om over dezelfde gegevens indexcijfers te vragen dan wel te berekenen!

Het berekenen van percentuele toe- of afname

Soms wil men over een bepaalde periode het verschil in ontwikkeling van jaar tot jaar weten. Dit kan het beste met het berekenen van een percentuele toe- of afname. Wij laten de berekening zien aan de hand van een verzonnen voorbeeld.

Misdrijven 1988–1991 in de stad A

1988	4000
1989	3600
1990	3960
1991	4070

Bron: Stadskoerier, 3 oktober 1992

De vragen zijn nu: 'Hoeveel nam het aantal misdrijven toe of af in 1989 ten opzichte van 1988; in 1990 ten opzichte van 1989 en 1991 ten opzichte van 1990?'.
Wij beantwoorden de vraag voor 1988 ten opzichte van 1989. In 1989 waren er 3600 misdrijven, in 1988 4000. Er is dus een bemoedigende ontwikkeling nl. een vermindering met 400. Dat getal 400 moeten wij nu als percentage uitdrukken van 4000 **(toe- of afname ten opzichte van 1988)**.
Eén procent van 4000 = 40,
dus 400 = 400 : 40 = 10%.
Er is dus een achteruitgang van 10% aan te geven door –10%.
Nu nog als voorbeeld: 1991 ten opzichte van 1990.
In 1990 waren er 3960 misdrijven en in 1991 4070, een minder goede ontwikkeling. Een verhoging met 110. Dat getal moeten wij als percentage uitdrukken van 3960. Een procent van 3960 = 39,6 dus 110 = 110/39,6 = 3% (afgerond). Er is dus een verhoging van 3% aan te geven door + 3%.

Als wij voor het voorbeeld alles hebben uitgerekend, krijgen wij misdrijven 1988–1991 in de stad A de ontwikkeling in percentuele toe- of afname met het voorstaande jaar als volgt te zien:

Absoluut	Percentuele toe- of afname	
1988	4000	
1989	3600	– 10%
1990	3960	+ 10%
1991	4070	+ 3%

Kengetallen

Kengetallen gebruiken is eigenlijk een bijzondere manier van het maken van vergelijkingen. In die zin dat men zelf materiaal bij elkaar zoekt om een reeks cijfers in een beter perspectief te krijgen. Uit de reeds eerder genoemde DCP-publikatie 'Criminaliteitsbeeld van Nederland' (1990) ontlenen wij volgende cijfers.

Aantal gestolen (brom)fietsen in Nederland over een aantal jaren absoluut en geïndexeerd:

Jaar	Gestolen fietsen		Gestolen bromfietsen	
	absoluut	index	absoluut	index
1986	167.559	100	19.084	100
1987	163.148	97	16.924	87
1988	172.036	103	18.243	96
1989	193.942	116	17.102	90

In eerste instantie spreken de cijfers voor zich. Fietsdiefstallen bevinden zich in stijgende lijn, bromfietsdiefstallen laten een dalende lijn zien. In de krant heeft u echter gelezen, dat er steeds meer nieuwe fietsen worden verkocht en dat de verkoop van nieuwe bromfietsen achter blijft. De gedachte vat nu bij u post dat de aan u geleverde gegevens wel **een beeld** geven, maar vermoedelijk **een te beperkt**. Het is denkbaar dat een groter park aan fietsen op zich al meer fietsdiefstallen oplevert en een kleiner wordend park bromfietsen op zich minder bromfietsdiefstallen mogelijk maakt. U gaat op zoek naar gegevens over het voertuigenpark. In hoofdstuk 3 van het Basisboek worden dit de **achtergrondgegevens** genoemd.

Het CBS heeft gegevens over het voertuigenpark in Nederland. U vindt het volgende in het 'Statistisch Zakboek'.

Jaar	Fietsenpark	Bromfietsenpark
1986	11.517.000	564.000
1987	11.441.000	516.000
1988	11.695.000	516.000
1989	11.865.000	474.000

U moet nu zoeken naar een mogelijkheid om de ontwikkeling van het aantal (brom)fietsdiefstallen aan die van het park te koppelen. Wij gaan een **diefstalrisico** berekenen. Dat kan door het aantal (brom)fietsdiefstallen uit te drukken in een percentage van het park. Wij geven één voorbeeld van hoe de berekening gaat. De vraag is bijvoorbeeld voor 1986 hoeveel is 167.559 (het aantal gestolen fietsen) uitgedrukt in een percentage van 11.517.000 (het park). In wezen zien wij dezelfde berekening als bij de indexcijfers en percentuele verdelingen. Het getal 11.517.000 is 100%, 1% daarvan is 115.170. Het getal 167.559 levert dan 167.559 : 115.170 = 1,5% op (afgerond).

Als we alle risicopercentages hebben uitgerekend, komen wij op het volgende:

Diefstalrisico van (brom)fietsen uitgedrukt in het het aantal diefstallen als percentage van het (brom)fietsenpark

	Diefstalrisico fietsen	Diefstalrisico bromfietsen
1986	1,5	3,4
1987	1,4	3,3
1988	1,5	3,5
1989	1,6	3,6

Als wij de oorspronkelijke cijfers zo weergeven krijgen wij een ander beeld. Bij de fietsen is het diefstalrisico min of meer stabiel, bij de bromfietsen stijgt het. Op basis van de **risicocijfers** zouden wij meer aandacht moeten besteden aan bromfietsdiefstallen, terwijl de absolute aantallen een daling van het aantal diefstallen van bromfietsen te zien gaf. Ook zien wij dat op deze manier weergegeven het diefstalrisico voor bromfietsen ongeveer twee keer zo groot is als voor fietsen. Wederom kan onze keuze waarop wij ons gaan richten anders worden, dan aanvankelijk verondersteld.

Wij geven nog een voorbeeld van een ander veel gebruikt kengetal nl. het **inbraakrisico**. Dat is het aantal inbraken uitgedrukt in het aantal woningen. Voor 1988 vinden wij voor de gemeente Diemen totaal 302 inbraken en voor de gemeente Oosterhout 304. Er gebeuren absoluut gezien nagenoeg evenveel inbraken. Omdat wij in beide gemeenten bekend zijn weten wij dat Diemen minder groot is dan Oosterhout. Via het CBS krijgen wij gegevens over het woningenbestand. Dat was in 1988 voor Diemen 7138 en voor Oosterhout 17288.

Het inbraakrisico (aantal inbraken uitgedrukt in een percentage van het aantal woningen) is voor Diemen 4,23 en voor Oosterhout 1,76. Dit geeft weer een ander beeld dan de absolute cijfers.

Bij **inbraak** is het zinvol de cijfers in verband te brengen met **woningen**. In eerste instantie wordt in woningen ingebroken. Bij andere delicten zoals **straatroof** is het niet zo zinvol naar het aantal woningen te relateren, maar bijvoorbeeld naar **inwoners**. Bij (brom)fietsdiefstallen is het zinvol naar het park te relateren. Er zijn dus diverse mogelijkheden. Kies welke mogelijkheid u het beste vindt passen bij een delict. Uw keuze wordt uiteraard beperkt door wat er beschikbaar is aan gegevens.

U zult, als naar inwoners wordt gerelateerd, verschillende aantallen tegenkomen. Iets wordt gerelateerd naar 100 inwoners, naar 1000 inwoners, naar 10.000 inwoners enz.
In de praktijk zal de keuze worden gemaakt door te kijken naar:

- **Hoe vaak komt een delict voor?** Als een delict weinig voorkomt zet men dat eerder tegen een groter aantal inwoners af. Er komen anders weinig zeggende getallen uit, bijvoorbeeld:
  10 moorden in een stad van 500.000 inwoners is
  2 moorden op 100.000 inwoners
  0,2 moord op 10.000 inwoners
  0,002 moord op 1000 inwoners
- **De wijze waarop andere cijfers zijn berekend.** Als er een verslag komt over bankovervallen in steden gerelateerd aan het aantal banken aldaar, dan zult u dat moeten hanteren als u het ook voor uw stad wilt berekenen. Dat kunt u goed vergelijken.

6.4 Schattingen maken van de totale omvang van een delict

Vaak wordt de vraag gesteld, **hoe het nu staat met de totale omvang van een bepaald delict.** Hiervan kan op zijn hoogst een schatting worden gemaakt. Het werkelijk aantal voorgekomen delicten is niet bekend en zal ook niet bekend worden.
Een schatting kan echter wel gemaakt worden.

Het is echter een grove globale schatting omdat gewerkt wordt met cijfers uit verschillende bronnen. Bij deze bronnen worden verschillende meetinstrumenten gebruikt. Wij zagen bij paragraaf 3 onder punt 4 'meetmethoden' al dat het dan oppassen geblazen is.

Het gaat hier om het combineren van gegevens van de CBS-politiestatistieken en die van de slachtofferenquêtes. Bij het samenstellen van de CBS-politiegegevens wordt gewerkt met delictomschrijvingen uit het Wetboek van Strafrecht; bij de slachtofferenquêtes wordt meer met termen uit het dagelijks taalgebruik gewerkt. De CBS-politiegegevens gaan over alle aangiften, ook die van jeugdigen, bedrijven, overheden, toeristen en niet-Nederlands sprekende burgers. De slachtofferenquêtes worden alleen gehouden bij Nederlands sprekende burgers van 15 jaar en ouder. Bij de slachtofferenquêtes wordt tenslotte naar een

beperkt aantal delicten gevraagd. U kunt dus niet voor alle delicten schattingen van het totaal aantal geven.

Het berekenen van die schatting is relatief eenvoudig. U moet er wel voor zorgen dat de beide gegevenssoorten op hetzelfde jaar betrekking hebben. Als dat niet kan – de slachtofferenquêtes worden meestal niet elk jaar opnieuw gehouden – dan moet u wel vermelden hoe u te werk bent gegaan.
U heeft bijvoorbeeld twee gegevens.
CBS 'Aantal vernielingen 1991': 624.
'Slachtofferenquête aangiftepercentage vernielingen 1991': 24%
U stelt dan:
624 = 24% (het aantal aangegeven vernielingen). Van hieruit moeten wij op 100% komen. Dat gaat als volgt.
624 : 24% = 1% = 26; 100% = 100 x 26 = 2600.
Dus schatting totaal aantal vernielingen in 1991 is 2600.
In een formule gegoten:

$$\frac{\text{CBS-aantal} \times 100}{\text{aangiftepercentage uit slachtofferenquête}}$$

Inleiding

In dit gedeelte zal achtergrondinformatie over een aantal statistische begrippen en technieken gegeven worden. Het is de bedoeling van de informatie dat u zich een voorstelling kunt maken bij die technieken en begrippen. U weet wat er mee bedoeld wordt of hoe het globaal in zijn werk gaat. U hoeft de technieken niet zelf te kunnen toepassen. Bij wijze van uitstapje zal een eenvoudige rangorde-correlatieberekening worden uitgelegd. Bij de behandeling van X^2 moet de berekening wel worden gevolgd anders is het moeilijk het begrip X^2 beter tot leven te brengen.

De steekproef

Wij trekken steekproeven omdat het vaak te kostbaar is een hele groep te onderzoeken. Een steekproef trekken is een werkwijze om te komen tot een verantwoorde bepaling van het gedeelte van de hele groep waarvan de onderzoeksresultaten zo veel mogelijk overeenkomen met hetgeen je zou vinden als je het geheel onderzoekt. Het is wel mogelijk om aan elke Nederlander ouder dan 18 jaar te vragen hoe zijn ervaringen als automobilist zijn, maar ontzettend duur. Men probeert dan een groep samen te stellen die een afspiegeling van de Nederlandse automobilist vormt. De uitkomsten van die groep worden dan als geldend voor alle automobilisten in Nederland beschouwd.

Een goede steekproef trekken is niet eenvoudig. Dat is werk voor specialisten. Het is zelfs niet mogelijk om vuistregels te geven. Toch hebben wij er op een of andere manier ook mee van doen. Stel, u wilt een steekproefgewijze regionale slachtofferenquête houden, wat dan? De specialisten, die verstand hebben van steekproeftrekken zijn net computersoftwareverkopers; zij zullen altijd naar **gebruikseisen** vragen. Zij moeten dan in ieder geval de volgende zaken weten:

a. Welke **grenzen van zekerheid** van uitspraken wil je uit de steekproef hebben, 90%, 95%, 99%? Hoe hoger de zekerheid, hoe waarschijnlijker de uitspraken zijn, maar des te groter de steekproef moet zijn.
b. Welke **eigenschappen van het geheel** wil je terugzien in het deel (de steekproef)? Het maakt voor de grootte van een steekproef uit of je bijvoorbeeld alleen maar wil onderscheiden naar man/vrouw, of dat je man/vrouw gecombineerd met geloof, inkomen enz. wil terugzien. In het laatste geval zal de steekproef groter moeten zijn.
c. Welke **variabelen** wil je onderzoeken? Wil je alleen weten of er bij iemand ingebroken is, of moet dat gecombineerd naar stadswijk, naar woningsoort, naar dag, enz? Het zal duidelijk zijn dat naar mate het aantal variabelen toeneemt de steekproef ook groter zal worden.

Over de gebruikerseisen kan van te voren nagedacht worden. Op zich is dat alleen denkwerk, geen cijferwerk. U kunt dus aan de slag.

Het begrip a-select

Vaak duikt het begrip a-select op. 'De steekproef was niet a-select en geeft een vertekend beeld'. Wat houdt dat nu in? Heel eenvoudig gezegd: iedereen moet een even grote kans hebben in de steekproef te komen, dan is er sprake van a-select. Als de één een grotere kans heeft dan een ander is er dus iets niet goed. Dit lijkt allemaal heel makkelijk, maar dat is het niet. Ook hier is het oppassen geblazen. Stel: wij willen een a-selecte steekproef via telefoonnummers uit het telefoonboek trekken. Dan kunt u niet lukraak aan de slag gaan. Degenen met een geheim nummer, of zonder telefoon, vallen bijvoorbeeld niet in de prijzen; die hebben dus **geen** kans om in de steekproef te komen.
Degenen die meerdere nummers hebben, bijvoorbeeld ministeries, artsen hebben, meer kans in de steekproef te komen.
U kunt dan ongewild een zgn. onder- c.q. overrepresentatie van bepaalde groepen krijgen. Als een steekproef om de een of andere reden niet a-select is kunnen er vertekeningen in de resultaten optreden.

De betrouwbaarheid van een meetinstrument

Betrouwbaarheid slaat op het feit of wij met een bepaald meetinstrument steeds hetzelfde meten of niet.
Wij zagen al dat schade door vandalisme op verschillende manieren gemeten kan worden door gemeenten. Zelfs het gebruik van een goed meetinstrument - de gulden - kan onbetrouwbare metingen opleveren als wij bij de ene meting de schade inclusief BTW en bij de andere exclusief BTW weergeven en dat niet vermelden. Bij de meting met de BTW is de gulden als het ware groter dan bij de meting exclusief BTW.

De validiteit van een meetinstrument

Simpel gesteld is een meetinstrument valide als wij er mee meten wat wij bedoelen te meten, met andere woorden dat de werkelijk-

heid goed wordt weergegeven. Het gebruik van geregistreerde schade door vandalisme in gemeenten is soms naast niet betrouwbaar ook niet valide om het 'vandalismeprobleem' er mee weer te geven. Immers door de diverse soorten registraties wordt vaak een ander stuk van het probleem weergegeven.

De chi-kwadraat-toets als techniek

Uit het in 1981 verschenen WODC-rapport over de proef 'Goed Gemerkt' in Deventer het volgende citaat.

'Van de ondervraagden behoorde 54% tot de rijkspolitie en 46% tot de gemeentepolitie. De kansen op gunstige effecten van Goed Gemerkt worden bij de rijkspolitie systematisch hoger geschat dan bij de gemeentepolitie. Binnen de rijkspolitie denkt bijvoorbeeld 95% dat hun collega's positief tegenover Goed Gemerkt staan, terwijl dit bij de gemeentepolitie slechts 32% is ($X^2 = 26,3$; $p = 0,001$ en $\varphi = 1$)'.
'.

Tot nu toe was de laatste zin van het citaat duister, meer in het bijzonder als het ging over '$X^2 = 26,3$, $p = 0,001$ en $\varphi = 1$'.
Daarin gaan wij nu wat licht brengen.

De X^2 is een vrij grove maat die bij **steekproeven** aangeeft in hoeverre de gevonden waarden gezet in een tabel afwijken van hetgeen te verwachten valt.
Wij kijken dus of er een wezenlijk verschil is tussen hetgeen wij gevonden hebben en datgene dat er naar verwachting had moeten zijn.

U kunt bij de X^2 toets de volgende symbolen tegenkomen

- χ Grieks hoofdletter voor de klank chi
- X i.p.v. Grieks hoofdletter onze X, in de Engelse literatuur ook wel aangeduid als 'chi-square'.
- φ kleine letter voor de klank fi geeft het aantal vrijheidsgraden aan. In de Engelse literatuur ook wel aangeduid met: df (degrees of freedom)
- p dit geeft aan voor welke graad van zekerheid gekozen is bijvoorbeeld $p = 0,05$ (95% zekerheid); $p = 0,001$ (99,9% zekerheid).

Berekenen

Om een beter idee van een X^2 te krijgen is dit helaas niet voldoende. Wij volgen aan de hand van een eenvoudig voorbeeld de berekening stap voor stap.
Het is hier niet de bedoeling dat u zelf X^2 gaat uitrekenen van diverse tabellen. Het gaat om het inzicht. De formule voor X^2 kan men in verschillende vormen tegenkomen.

We geven er twee:

$$X^2 = \Sigma \frac{(W \text{ (werkelijk aantal)} - V \text{ (verwacht aantal)})^2}{V \text{ (verwacht aantal)}}$$

$$X^2 = \Sigma \frac{(fo \text{ (frequency observed)} - fe \text{ (frequency expected)})^2}{fe \text{ (frequency expected)}}$$

In de formule staan dus de werkelijk gevonden aantallen en de te verwachten aantallen.

De formule ziet er niet erg ingewikkeld uit, maar de vraag is wel hoe wij komen tot de te verwachten aantallen. Dit wordt bij deze toets gedaan door gebruik te maken van de zgn. **randtotalen**. Randtotalen zijn de totalen die aan de rand van een tabel komen. In de onderstaande tabel zijn de randtotalen omkaderd.

Werkelijk gevonden waarden

32	468	500	
54	575	629	
86	1043	1129	

Als wij nu op basis van deze tabel een tabel met 'te verwachten waarden' moeten maken, dan gaan wij als volgt te werk. Wij maken een lege tabel met **dezelfde** randtotalen als die van de tabel met de werkelijk gevonden waarden, dus zo:

Verwachte waarden

		500
		629
86	1043	1129

Nu maken wij het wat ingewikkelder, nl. zo:

a	b	500
c	d	629
86	1043	1129

Elke letter neemt een plaats in. Die plaats noemen wij een cel. De verwachting voor cel a berekenen wij als volgt. Het randtotaal van de rij waarin a staat (horizontaal gezien) wordt vermenigvuldigd met het randtotaal van de rij waarin a staat (verticaal gezien) en gedeeld door het totaal van de randtotalen. U ziet, dat

hier iets gebeurt, dat wij al eerder tegen kwamen in een andere vorm bij de computertabel: men maakt gebruik van de totalen die door cel a worden beïnvloed (pag. ..). In de tabel hieronder wordt het gellustreerd.

$$\frac{500 \times 86}{1129} = 38 \text{ (afgerond)}$$

De verwachting voor alle cellen kunnen wij nu gaan uitrekenen.

$$\text{cel a} = \frac{500 \times 86}{1129} = 38 \quad \text{(afgerond)}$$

$$\text{cel b} = \frac{500 \times 1043}{1129} = 462 \quad \text{(afgerond)}$$

$$\text{cel c} = \frac{629 \times 86}{1129} = 48 \quad \text{(afgerond)}$$

$$\text{cel d} = \frac{629 \times 1043}{1129} = 581 \quad \text{(afgerond)}$$

De tabel met de te verwachten aantallen wordt dus:

38	462	500
48	581	629
86	1043	1129

Nu hebben wij een tabel met werkelijk geworden waarden en een tabel met te verwachten waarden. Die zetten wij naast elkaar.

Werkelijk gevonden			**Verwacht aantal**		
32	468	500	38	462	500
54	575	629	48	581	629
86	1043	1129	86	1043	1129

$$X^2 = \sum \frac{(W - V)^2}{V}$$

De Σ betekent dat wij de formule een aantal keren op de getallen moeten loslaten en de resultaten daarvan moeten optellen. Wij maken de tabellen nu 'formuleklaar' als volgt.

Werkelijk gevonden (w)			**Verwacht aantal (v)**		
32 (aw)	468 (bw)	500	38 (av)	462 (bv)	500
54 (cw)	575 (dw)	629	48 (cv)	581 (dv)	629
86	1043	1129	86	1043	1129

Op elk cellenpaar waarin de werkelijk gevonden waarde (bijvoorbeeld: aw) en de te verwachten waarde staan (bijvoorbeeld: av) laten wij nu de formule los. Elk cellenpaar levert een bijdrage voor de X^2. Bij deze berekening uitkomsten noteren tot twee cijfers na de komma

$$\text{voor cel a:} \quad \frac{(32-38)^2}{38} = 0{,}95$$

$$\text{voor cel b:} \quad \frac{(54-48)^2}{48} = 0{,}75$$

$$\text{voor cel c:} \quad \frac{(468-462)^2}{462} = 0{,}08$$

$$\text{voor cel d:} \quad \frac{(575-581)^2}{581} = 0{,}06$$

$$1{,}84$$

We vinden dus $X^2 = 1{,}84$. Dat is dan dat. Maar hoe nu verder?

Het begrip vrijheidsgraden

Wij kunnen alleen maar verder als wij onderstaande tabel kunnen hanteren.

Tabel X^2 verdelingen
Waarden van X^2 (afhankelijk van het aantal vrijheidsgraden) behorende bij een aantal vaak toegepaste eenzijdige overschrijdingskansen

	eenzijdige overschrijdingskans (%)			
φ	5	2½	1	½
1	3,84	5,02	6,63	7,88
2	5,99	7,38	9,21	10,6
3	7,81	9,35	11,3	12,8
4	9,49	11,1	13,3	14,9
5	11,1	12,8	15,1	16,7
6	12,6	14,4	16,8	18,5
7	14,1	16,0	18,5	20,3
8	15,5	17,5	20,1	22,0
9	16,9	19,0	21,7	23,6
10	18,3	20,5	23,2	25,2
11	19,7	21,9	24,7	26,8
12	21,0	23,3	26,2	28,3
13	22,4	24,7	27,7	29,8
14	23,7	26,1	29,1	31,3
15	25,0	27,5	30,6	32,8
16	26,3	28,8	32,0	34,3
17	27,6	30,2	33,4	35,7
18	28,9	31,5	34,8	37,2
19	30,1	32,9	36,2	38,6
20	31,4	34,2	37,6	40,0
22	33,9	36,8	40,3	42,8
24	36,4	39,4	43,0	45,6
26	38,9	41,9	45,6	48,3
28	41,3	44,5	48,3	51,0
30	43,8	47,0	50,9	53,7

De getallen 5, 2½, 1 en ½ slaan op de graad van zekerheid van onze uitspraak, hier respectievelijk 95%, 97,5%, 99% en 99,5%.
Men schrijft dat op als p = 0,05 (1–0,95); p = 0,025 (1–0,975); p = 0,01 (1–0,99) en p = 0,005 (1–0,995). Hierbij staat het getal 1 voor 'de volmaakte waarheid'. Van die waarheid halen wij dat gedeelte af van wat wij voldoende als 'volmaakte waarheid' vinden. Het overblijvende geven wij aan met de p. Blijft de φ als onbekende over. De φ slaat op het begrip vrijheidsgraden. Dit wil zeggen: het minimaal aantal cellen waarvan ik de waarde moet weten om de andere cellen te kunnen invullen (dan moeten wèl de randtotalen bekend zijn).
Voor tabellen geldt dat:
φ = (het aantal kolommen –1) x (het aantal regels –1)
In onze gebruikte tabel (pag. 41) van de werkelijk gevonden waarden:

	kol. <1>	kol. <2>	
regel 1:	32		500
regel 2:			629
	86	1043	1120

Dus = (2–1) x (2–1) = 1 x 1 = 1. Er is dus één vrijheidsgraad. Dat wil zeggen als men in onze tabel de randtotalen weet, men maar van één cel het getal hoeft te weten om de tabel verder in te kunnen vullen. Dus bijvoorbeeld 32.
Wij kijken nog even of de φ-formule ook klopt voor een andere tabel bijvoorbeeld:

6	7	8		31
				12
10	12	10	11	43

φ = (4–1) x (2–1) = 3 x 1 = 3. Als wij de randtotalen weten dan is de serie drie bekende getallen 6, 7 en 8 voldoende om de rest in te vullen. De φ-formule gaat er van uit, dat men slim gekozen getallen neemt. Als wij bijvoorbeeld de getallen 6, 7 en 4 nemen, dan kunnen wij de tabel toch niet verder invullen.

Als wij nu de tabel over de X^2 bezien dan staat 'φ' in de meest linkse kolom en daaronder de vrijheidsgraden die men kan vinden.
Wij hadden in ons voorbeeld gevonden φ = 1 en bij 95% zekerheid is de X^2 dan 3,84. Wij vonden een X^2 = 1,84. Aanmerkelijk minder dus dan het getal 3,84.
Het getal 3,84 is op te vatten als een keerpunt. Vindt men X^2- waarden die beneden dit keerpunt liggen, dan wijkt het gevondene van wat er verwacht kan worden niet af. Vindt **men waarden boven dit keerpunt, dan wijkt het gevondene wel af van hetgeen men verwachten kan.** Wij kunnen dat vermelden maar over het waarom daarvan weten wij dan nog niets. Kijkt u nog maar eens terug naar het citaat in het begin uit het 'Goed Gemerkt' rapport. Er staat niet in waarom de kansen systematisch hoger worden ingeschat, **alleen maar dat het zo is.**

De correlatie als techniek
Het begrip 'correlatie' betekent dat er een verband tussen twee of meer verschijnselen bestaat.
Als men tegelijkertijd naar de onderlinge samenhang van meer dan twee verschijnselen kijkt, spreekt men van een multivariantie-analyse. Dat is een zeer ingewikkelde materie. Wij

gaan hier niet verder dan de verbanden tussen **twee** verschijnselen.

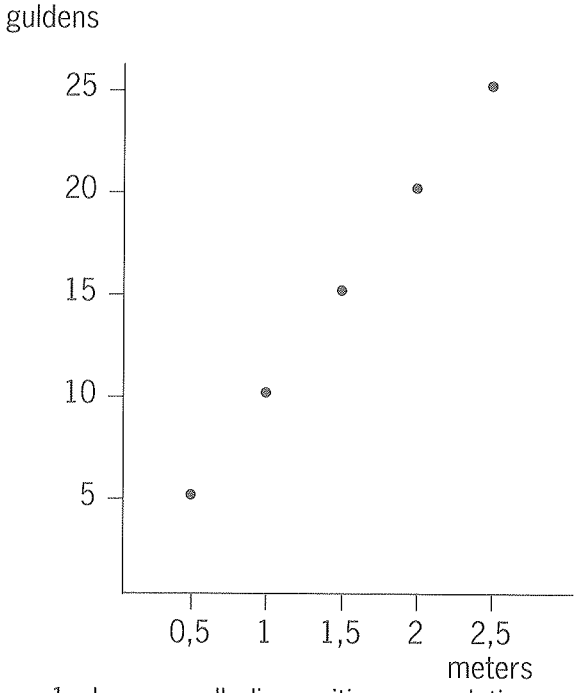
Symbolen en formule
Correlatie wordt verschillend aangeduid:
ρ (Griekse letter voor de r, de klank is rho)
r (in plaats van de Griekse letter onze letter)

De meest algemene formule is:

$$r = \frac{n.xy - \Sigma x.y}{(n.\Sigma x^2 - (\Sigma x)^2\ (n.\Sigma y^2 - (\Sigma y)^2)}$$

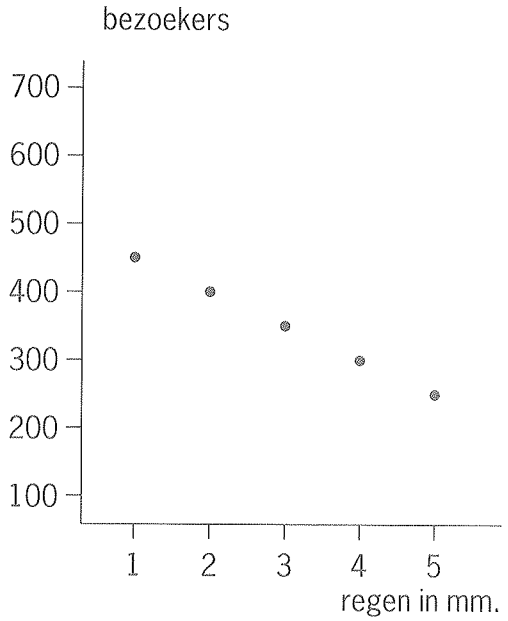
In die formule geeft x de gevonden waarde van het ene verschijnsel weer en y die van het andere verschijnsel.

De uitkomst van een correlatie wordt in een getal weergegeven dat ligt tussen + 1 en –1. Bij getallen die tussen de 0 en + 1 liggen spreken wij van een **positieve** correlatie en bij die tussen de 0 en –1 van een **negatieve** correlatie.
Bij een volledige positieve correlatie, + 1, wordt het ene verschijnsel met dezelfde factor groter als het andere. Als hout per strekkende meter voor ƒ 10,– wordt verkocht en er geen kwantumkortingen worden gegeven en er een bestelling komt voor een stuk van 50 cm., 1m., 1,5 m. en 2,5 m., dan moet er respectievelijk ƒ 5,–, ƒ 10,–, ƒ 15,–, en ƒ 25,– betaald worden. Als wij dit in een assenstelsel zetten krijgen wij het volgende.



r=+1, dus een volledig positieve correlatie.

Een volledig negatieve correlatie wordt weergegeven door –1. Hierbij geldt als het ene verschijnsel groter wordt, wordt het andere even veel kleiner. Bij een gelegenheid voor openlucht recreatie zou kunnen gelden dat hoe meer regen er valt, des te minder bezoekers er komen. Een voorbeeld zou dan in een assenstelsel het volgende opleveren.



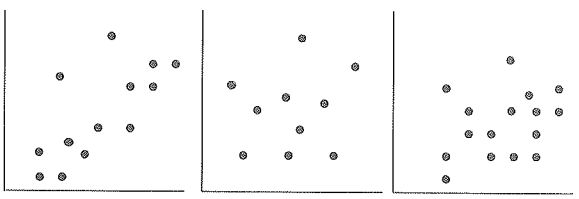
r= –1

Hieronder is de betekenis van de waarde van de correlatie coëfficiënt en de aard van het verband tussen twee variabelen schematisch weergegeven:

Schema 10: Betekenis gevonden waarden van een correlatie-efficiënt

negatief						positief
–1	–0,6	–0,3	0	0,3	0,6	+1
sterk verband	matig verband	geen verband		matig verband	sterk verband	

De in het voorgaande gegeven voorbeelden van een volledig positieve dan wel negatieve correlatie komen in de praktijk nauwelijks voor. De gevonden punten liggen vrijwel nooit zo, dat er een rechte lijn door is te trekken.
Als wij punten gaan intekenen kunnen wij uit de gemaakte puntenwolk al redelijk aflezen hoe met de verbanden staat.



a. Sterk verband
b. Geen verband
c. Matig verband

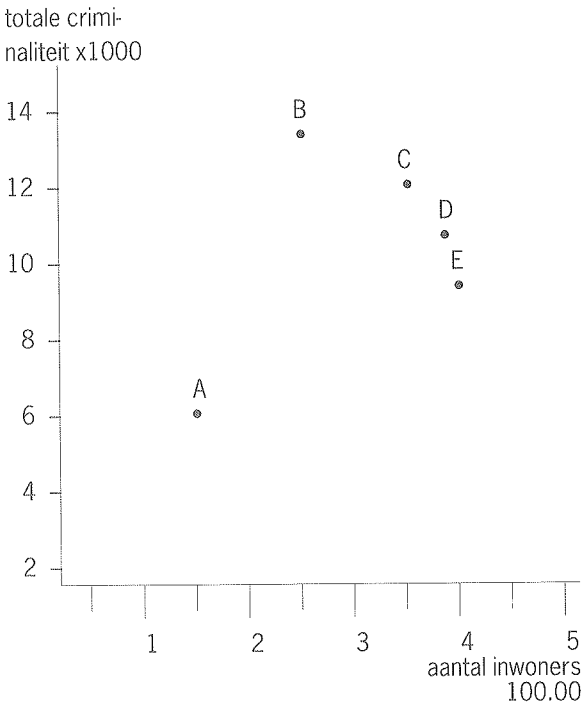
In de middelste figuur is op het oog geen verband te ontdekken. Wanneer we het verband zouden uitrekenen door middel van de correlatiecoëfficiënt zou het best kunnen dat er wel een zwak verband bestaat. Zo'n verband is dan meestal statistisch niet interessant.

Wij illustreren de puntenwolk nog aan de hand van twee verzonnen voorbeelden. Een algemeen verondersteld verband is: hoe groter de steden hoe meer criminaliteit. Wij gaan meten en vinden in regio A de volgende gegevens

Regio A

	Inwoners	Totale criminaliteit
stad A	150.200	7.320
stad B	246.500	13.220
stad C	362.900	12.040
stad D	385.100	10.330
stad E	401.600	9.400

Getekend als puntenwolk



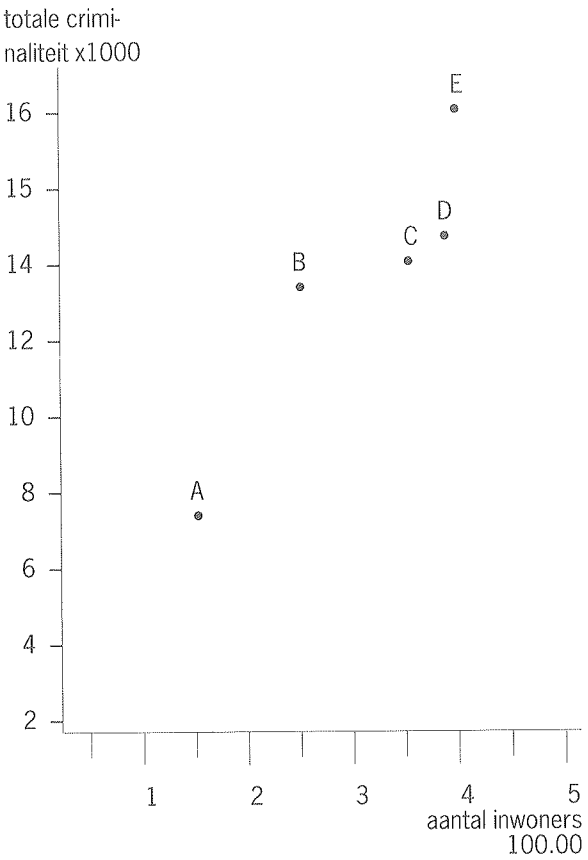
De punten liggen niet zo, dat er van geen enkel verband gesproken kan worden. Een matig verband zit er in, maar wel dicht tegen een zwak verband aan. Dit laatste is juist, de r is voor deze tabel 0,31.

In regio B vinden wij het volgende:

Regio B

	Inwoners	Totale criminaliteit
stad A	150.200	7.320
stad B	246.500	13.220
stad C	362.900	14.000
stad D	385.100	14.500
stad E	401.600	16.000

In een puntenwolk overgebracht



De puntenwolk of -configuratie benadert het plaatje van een sterk verband. Dit is ook juist, de r is hier 0,93.
Door middel van het intekenen van puntenwolken is al een redelijk inzicht te krijgen in hoe sterk c.q. zwak een verband zal zijn.

Valkuilen bij correlatie

Met de formule in de hand kunnen wij van diverse reeksen cijfers de onderlinge correlatie berekenen. Er zijn een groot aantal verschijnselen die in de tijd gezien eenzelfde of een omgekeerd verloop hebben. Op basis daarvan en **alleen** op basis daarvan zal er tussen twee verschijnselen een hoge correlatie kunnen bestaan. Een echte correlatie bestaat er pas als er inderdaad een oorzakelijk verband bestaat. Alle andere gevonden correlaties noemt men valse of schijnverbanden.

Door na te denken kunnen wij (soms) er achter komen of een ons voorgeschotelde correlatie inderdaad oorzaak en gevolg weergeeft of dat het eigenlijk maar een schijncorrelatie betreft. Wij kunnen dit gaan doen via het weergeven van het ons voorgeschotelde verband met een pijl.
In het hiervoor gebruikte 'hout' voorbeeld is het duidelijk.

verschijnsel A ⟶ verschijnsel B
(lengte hout) (hoogte prijs)

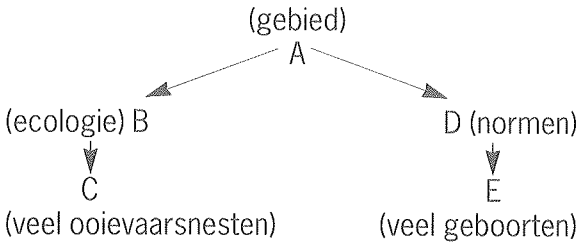
Logisch redenerend is er een oorzakelijk verband.

Een bekend voorbeeld van een schijncorrelatie is de zeer sterke correlatie die werd gevonden tussen het jaarlijks aantal geboorten en het aantal ooievaarsnesten in de overeenkomstige jaren in een bepaald gebied. Zou het dan toch waar zijn dat ooievaars iets met geboorten te maken hebben?

Men probeert ons in dit geval wijs te maken dat:

A ⟶ B
(ooievaarsnesten) (geboorten)

Logisch redenerend weten wij, dat dit niet waar is. Wij laten het er niet bij zitten want er zal vermoedelijk een onderliggende factor zijn die zowel het aantal ooievaarsnesten als het aantal geboorten beïnvloedt. Wij hebben dan niet alleen gesteld dat er een schijncorrelatie is (dat is gemakkelijk), maar ook aangegeven hoe een en ander wel in elkaar steekt (dat is heel wat moeilijker).
Na onderzoek vinden wij dat er in het aangegeven gebied veel ooievaarsnesten zijn omdat de omgeving zich daarvoor leent en dat in hetzelfde gebied ook veel geboorten plaatsvinden omdat daar bepaalde normen over het gewenste kindertal heersen.
Dan krijg je schematisch het volgende beeld.



In het ooievaarsvoorbeeld zal iedereen er snel van te overtuigen zijn dat het oorspronkelijk gesuggereerde verband niet juist is. In de praktijk beginnen hier echter al de problemen. Het is daarna nog minder eenvoudig om uit te vinden hoe de vork wel in de steel steekt.

Wij geven daarom nog een voorbeeld, waar wij al eerder mee aan de slag waren: de valhelmplicht voor bromfietsers en de daling van het aantal bromfietsdiefstallen na de inwerkingtreding daarvan. Een citaat: 'In Nederland is het aantal bromfietsdiefstallen blijvend met de helft gedaald sinds de invoering van de verplichte bromfietshelm. Het stelen van bromfietsen is hierdoor kennelijk te lastig en riskant geworden voor veel van de potentiële daders van dit soort delicten[5]. Het woord kennelijk geeft nog een slag om de arm, maar er wordt in ieder geval iets gesuggereerd:

A ⟶ B ⟶ C
(bromfiets helmplicht) (stelen te lastig en riskant voor potentiële daders) (minder bromfiets diefstallen)

In tegenstelling tot het ooievaarsvoorbeeld is het ons voorgeschotelde niet onlogisch. Integendeel het lijkt heel logisch. U moet dus een goede reden hebben (uitgezocht hebben hoe de vork dan wel in de steel zit) voordat u kan stellen, dat het verband wel logisch lijkt, maar (vermoedelijk) niet helemaal zo loopt als wordt gesuggereerd.

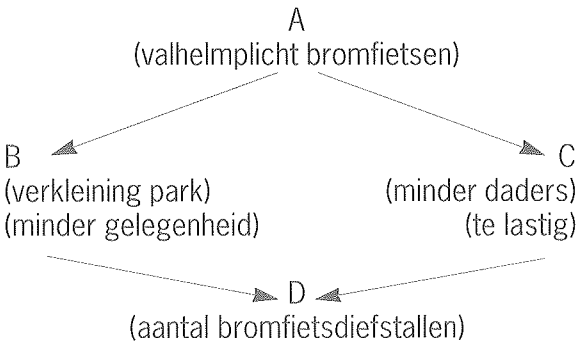
Het is meestal een kwestie van toeval dat iemand die de daarvoor benodigde kennis bezit, het gesuggereerde verband onder ogen krijgt en zegt 'Dat klopt niet'. Dat was hier ook het geval. Iemand die vroeger bij het ministerie van Verkeer en Waterstaat werkte en nauw betrokken was geweest bij de invoering van de helmplicht herinnerde zich, dat door die helmplicht de belangstelling voor de bromfiets als vervoermiddel sterk terug liep. De verkoop van nieuwe bromfietsen stagneerde sterk. Er kwamen gewoon minder bromfietsen op de weg. Minder bromfietsen: dus minder gelegenheid voor diefstallen. De valhelmplicht werd op 1 februari 1975 van kracht.

Het bromfietspark had in de jaren vlak daarna het volgende verloop

Bromfietspark Nederland	
1975	1.650.000
1976	1.400.000
1977	1.200.000
1978	1.100.000
1979	900.000
1980	800.000

[5] dr. J.J.M. van Dijk, Criminaliteit als keerzijde, pag. 33.

Een sterke daling van het voertuigenpark, een reducering tot de helft van 1975 tot 1980. Het ons eerst gesuggereerde verband ligt dus anders. Hoe precies zou nader onderzoek moeten uitwijzen.
In schema krijgen wij nu:



De oorspronkelijke suggestie kan niet blijven staan, in ieder geval moet de achteruitgang van het bromfietspark ook genoemd worden. U ziet hoezeer u op uw hoede moet zijn. Dit klemt des te meer, omdat er soms beslissingen worden genomen op de basis 'voor de hand' liggende oorzakelijke verbanden, die achteraf niet of minder waar waren.

Een eenvoudige rangordecorrelatie

Wij maken nog een uitstapje naar een correlatieberekening. Gekozen is voor een rangordecorrelatie, die u in de praktijk kunt toepassen en makkelijk te berekenen is. Rangordes worden overal gebruikt: de top tien delicten, de zwarte zes en ga zo maar door.
In sommige gevallen kunnen wij van twee reeksen gegevens een rangorde bepalen en zien of er tussen die rangordes een verband bestaat. Dit verband wordt met de bekende waarde r weergegeven. U moet zich dat als volgt voorstellen.

rangorde			
A	B	A	B
1	1	1	5
2	2	2	4
3	3	3	3
4	4	4	2
5	5	5	1
volledig positief (r = +1)		volledig negatief (r = –1)	

De rangordecorrelatie is een grove maat. Als wij naar pag..... terugbladeren zien wij dat wij een rangorde als een ratioschaal behandelen, hetgeen eigenlijk niet kan. Niettemin kunnen wij ons voordeel er mee doen, gezien het volgende voorbeeld.
In een werkgroep kwam een rapport van becommentariëring ter tafel. In dat rapport stond deze tabel.
De begeleidende tekst was:

Verdeling van het aantal jachthavenligplaatsen (bron: CBS 1982) en het gemiddeld percentuele aandeel geregistreerde delicten per regio (1981 tot en met 1987)

Regio	% delicten	% ligplaatsen
Utrecht	22,5	9,3
Rijnmond	9,3	11,0
Zuid Holland Midden	8,0	8,2
Friesland	7,8	14,6
Noord Holland Oost	7,1	2,5
Noord Brabant West	7,0	3,5
Noord Holland West	5,6	3,9
Limburg Zuid	4,8	0,4
Noord Holland Noord	3,7	8,3
Gelderland Noord	2,9	6,0

'Uit de tabel blijkt dat er een samenhang bestaat tussen de verdeling van het aantal ligplaatsen en de verdeling van het aantal gepleegde delicten: in de regio's met veel ligplaatsen vinden ook veel diefstallen plaats'.

Een lid van de werkgroep merkte op, dat de uitspraak 'in de regio's met veel ligplaatsen vinden ook veel diefstallen plaats' voor discussie vatbaar was. Immers Utrecht heeft 9,3% (hoog) ligplaatsen en 22,5% (hoog) delicten, Noord-Holland Noord heeft 8,3% (hoog) ligplaatsen en 3,7% (laag) delicten.
Hoe nu dit gevoelen via een statistische bewerking waarmaken. Het zou -grof- met een rangordecorrelatie kunnen. Immers volgens de uitspraak 'veel ligplaatsen' geven veel diefstallen, kunnen de gegevens omgezet worden in een rangorde. Wij kunnen dan zien in hoeverre de rangordes met elkaar overeenstemmen uitgedrukt in r.
De formule voor de gebruikte rangordecorrelatie is:

$$r = 1 - \frac{6 \Sigma D^2}{N (N^2 1)}$$

waarbij:
N = het aantal geobserveerde paren (in dit geval 10)
D = het verschil tussen twee rangordes binnen een bepaald paar.

Wij moeten nu de getallen van het % delicten in een rangorde zetten. Dat is eenvoudig, die lopen al van hoog tot laag. De ligplaatsen moeten nu nog in een rangorde worden geplaatst aangezien straks de paren via het verschil in rangorde worden vergeleken kunnen de cijfers van het % ligplaatsen niet zomaar van hoog tot laag gezet worden. Zij moeten blijven staan het kost even wat moeite om de rangorde goed aan te brengen. Daarna moet D worden berekend. Zoals bij andere berekeningen al bleek is het niet nodig + of – aan te geven, als daarna gekwadrateerd wordt. Het teken Σ kwamen wij al meer tegen. Het betekent 'tel op'. Wij maken nu een werktabel om een en ander in te vullen.

(werktabel

Regio	% delicten	% ligplaats	Rangorde % delicten	Rangorde % ligplaats	Verschil in rangorde = D	D^2
Utrecht	22,5	9,3				
Rijnmond	9,3	11,0				
Zuid-Holland N	8,0	8,2				
Friesland	7,8	14,6				
Noord-Holland O	7,1	2,5				
Noord Brabant W	7,0	3,5				
Noord-Holland W	5,6	3,9				
Limburg Z	4,8	0,4				
Noord Holland N	3,7	8,3				
Gelderland N	2,9	6,0				
						Σ

(ingevulde tabel)

Regio	% delicten	% ligplaats	Rangorde % delicten	Rangorde % ligplaats	Verschil in rangorde = D	D^2
Utrecht	22,5	9,3	1	3	2	4
Rijnmond	9,3	11,0	2	2	0	–
Zuid-Holland N	8,0	8,2	3	5	2	4
Friesland	7,8	14,6	4	1	3	9
Noord-Holland O	7,1	2,5	5	9	4	16
Noord Brabant W	7,0	3,5	6	8	2	4
Noord-Holland W	5,6	3,9	7	7	0	–
Limburg Z	4,8	0,4	8	10	2	4
Noord Holland N	3,7	8,3	9	4	5	25
Gelderland N	2,9	6,0	10	6	4	16
						Σ 82

De formule $r = 1 - \frac{6 \sum D^2}{N (N^2 - 1)}$ is nu eenvoudige in te vullen

$$r = 1 - \frac{6 \times 82}{10 (100 - 1)} = 1 - \frac{492}{990} = 0{,}50$$

Wij zien dus dat er maar een matig verband is. Ons gevoelens kunnen wij nu staven.
De tabel is in de uiteindelijke publikatie vervallen. De rangordecorrelatie had zijn werk gedaan!

Slotopmerkingen

Er is een grote hoeveelheid literatuur over statistiek en de daarbij gangbare methoden en technieken. Wellicht dat u na het lezen van dit hoofdstuk u daar wat verder in wilt verdiepen. Een goed begin daarbij zou zijn het boek 'Statistiek in Woorden' van A. Slotboom (Wolters, 1987) te raadplegen.
De belangrijkste opmerking blijft dat u er vooral zelf mee aan de slag moet gaan en blijven. Dit is een zich versterkend proces. Als u er eenmaal mee begonnen bent en resultaten hebt geboekt, dan gaat u er meer gebruik van maken en .....