

VAN
TOFU
KRIJG
JE
GEHEUGEN-
VERLIES



VAN
TOFU
KRIJG
JE
GEHEUGEN-
VERLIES*

Coen de Bruijn

* Gekonkel en gestuntel met statistiek
in media, politiek en reclame



2010 Uitgeverij Het Tweede Gezicht, Den Haag

HE
L
T
W
E
E
D
E
G
E
Z
I
C
H
T
U
I
T
G
E
V
E
R
I
J

© 2010 Coen de Bruijn / Het Tweede Gezicht, Den Haag
www.hettweedegezicht.com

Een uitgave via www.schrijversmarkt.nl
Eerste druk 2010

Ontwerp omslag en binnenwerk Valetti, www.valetti.nl
Drukkerij Wilco



Mixed Sources

Productgroep uit goed beheerde
bossen, gecontroleerde bronnen
en gerecydeerd materiaal.
www.fsc.org Cert no. CU-COC-803902
© 1996 Forest Stewardship Council

ISBN 978-94-6141-002-3 / NUR 740

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke wijze en/of door welk ander medium ook, zonder voorafgaande schriftelijke toestemming van de uitgever.

INHOUDSOPGAVE*

*Voorwoord	6
*Hoofdstuk 1 De inleiding – De maker, gebruiker en consument van statistieken	11
*Hoofdstuk 2 De oorzaak – Over de bakermat van slechte statistiek	23
*Hoofdstuk 3 De psychologie – Het zit allemaal tussen onze oren	49
*Hoofdstuk 4 De steekproef – Iene, miene, mutte is de baas	63
*Hoofdstuk 5 Het percentage – De Frans Bauer van de statistiek	81
*Hoofdstuk 6 Het gemiddelde – De gulden middenweg	97
*Hoofdstuk 7 De voorspelling – Glazen bollen en risico's	111
*Hoofdstuk 8 De grafiek – Als duizend woorden meer zeggen dan één plaatje	133
*Hoofdstuk 9 De lijsten – Algemeen Beschaafd Statistiek	165
*Dankwoord	174
*Bronnen & Nice to visit	176
*Lijst met supporters	183

VOORWOORD*

Om maar direct met de deur in huis te vallen: dit boek gaat over statistiek. Uit ervaring weet ik dat meer dan driekwart van de lezers nu de neiging voelt om dit boek opzij te schuiven. Behoor jij tot die doelgroep? Lees dan even door. Hier volgt waarom.

Ten eerste is ons leven een aaneenschakeling van statistiek, of we nou willen of niet. We worden ermee geboren en we gaan ermee dood. Nog geen vijf minuten op deze wereld en we krijgen al een rapportcijfer in de vorm van de APGAR-score. In rap tempo volgen groeischema's bij het consultatiebureau, Cito-toetsen, rapportcijfers, trouw- en scheidingsstatistieken, consumentenpatronen, inflatiecijfers of economische groei. Tenslotte eindigen we allemaal in een sterftestatistiek.

Ten tweede gebruikt bijna iedereen wel eens statistiek om een boodschap mee te onderbouwen of andermans boodschap mee te ontkrachten. De bergen gegevens die ons dagelijks non-stop via de media en reclame bereiken, staan bol van feitjes, onderzoeksresultaten en opinies. Vervolgens gebruiken politici, beleidsmakers, reclamemakers, verkopers, collega's of vrienden die gegevens om er hun beleid of mening op te baseren.

Ten derde vertrouwen we blindelings op cijfers en dat is gevaarlijk. Cijfers zien we als feiten. Als iets in cijfers is gevat dan *is het zo*. Veel van de statistieken in dit boek zijn afkomstig uit

de media, politiek en reclame. Ze lijken misschien onschuldig en zijn dat soms ook, maar beseft dat statistiek de basis vormt voor veel belangrijke beslissingen. Bedrijfsstrategieën, reorganisaties, massaontslagen, overheidsbeleid, -uitgaven of veroordelingen in een rechtbank. Al die keuzes zijn gebaseerd op cijfers. We proberen continu de complexe samenleving om ons heen te begrijpen, onder andere met behulp van cijfers. Statistiek heeft op die manier een enorme impact op ons dagelijks leven. Wij gaan er automatisch vanuit dat de aan ons gepresenteerde gegevens en informatie kloppen, want ze zijn gebaseerd op *Statistisch Onderzoek of Significante Resultaten*. Om de interpretatie van de resultaten wat makkelijker te maken, worden de cijfers vaak pakkend weergegeven in plaatjes en grafieken. Wie zijn wij om daar vraagtekens bij te zetten? Eén plaatje zegt meer dan duizend woorden en de meesten van ons zijn toch geen wiskundigen of statistici. Juist *daarom* moeten we er vraagtekens bij zetten.

Ten vierde laten statistische gegevens zich erg goed lenen voor kwalijk gekonkel en tenenkrommend gestuntel in media, reclame en politiek. Ik wil het hebben over die valkuilen. We hebben er namelijk allemaal mee te maken.

Statistiek en statistici staan bij mij over het algemeen hoog aangeschreven. Dit boek is dan ook niet geschreven om het vakgebied af te kraken. Zoals gezegd wil ik de lezer laten kennismaken met een aantal statistische valkuilen. Die behandel ik per hoofdstuk. De hoofdstukken laten zich los van elkaar lezen. Begin met het eerste hoofdstuk, maar voel je daarna niet verplicht om de volgorde aan te houden. Ik verwacht

geen kennis van statistiek bij de lezers. Waar nodig leg ik begrippen uit. Het laatste hoofdstuk bevat een lijst met de meest voorkomende statistische termen. Iedereen die weet wat een gemiddelde en een percentage is, is in staat om het hele boek - hopelijk met veel plezier - te lezen.

De aanleiding tot het schrijven van dit boek was *How to lie with statistics* van Darell Huff. Dit standaardwerk uit 1954 las ik eind jaren negentig toen ik werkte bij SPSS, een bedrijf dat statistische software ontwikkelt. Door het boek van Huff raakte ik gefascineerd in het gebruik van cijfers, overal om me heen. Nog steeds geldt *How to lie with statistics* als de belangrijkste bron voor alles wat daarna over misbruik van statistiek is geschreven. Huff schreef zijn boek in een tijd dat kranten en radio de voornaamste nieuwsbronnen waren. Politiek speelde zich ergens ver buiten de huiskamer af en statistiek was niet besteed aan de gewone man. Dat is anno 2010 wel anders. Het gebruik van statistiek is drastisch toegenomen maar daar zijn nog altijd weinig mensen zich van bewust. Statistische gegevens worden te pas en te onpas gebruikt en misbruikt. Het zou goed zijn als iedereen zich daarvan bewust is, te beginnen bij middelbare scholieren. Als mensen al vroeg betrouwbare en onzinnige statistiek van elkaar leren onderscheiden, zullen ze sneller een kritisch oog daarvoor ontwikkelen.

Toen ik Huffs boek las, was internet nog traag en onoverzichtelijk. Je moest je door bergen sites en ellenlange blogs heen worstelen voor wat informatie. Nieuws kreeg je via de krant, het *NOS Journaal* of *RTL Nieuws*. Met een mobiele telefoon

kon je alleen maar bellen. Eigenlijk wilde je dat niet eens want dat was iets voor patsers. Maurice de Hond kenden we toen voornamelijk van zijn doemscenario's rondom de *millenium-bug*, waarin hij ons aanspoorde om te gaan hamsteren.

Tegenwoordig worden we gebombardeerd met informatie. Gelukkig sorteert *Google* de hoogtepunten voor ons en helpt *Wikipedia* ons bij het vinden van de nieuwe waarheid. Alles komt sneller dan het licht en in korte *tweets* op ons af. Nieuws halen we via *nu.nl* of de gratis bladen. Kranten zijn er nog wel, maar worden compacter. De hoeveelheid achtergrondinformatie krimpt in. Bijna dagelijks verschijnt 's lands bekendste opiniepeiler, Maurice de Hond, in de media. Zelfs als er geen verkiezingen zijn.

Dit boek gaat dan ook in op de wisselwerking tussen statistiek en onze digitale maatschappij. Voordeel van deze digitale maatschappij is dat we steeds beter in staat zijn om informatie te filteren. We scannen gemakkelijk door gegevens en kiezen datgene waar we naar op zoek zijn. Nadeel is dat we minder kritisch zijn over de informatie die ons wordt aangeboden. Er komt gewoonweg te veel op ons af. We nemen er de tijd niet meer voor om het aanbod nauwkeurig onder de loep te nemen. Dat geldt zeker voor cijfers en statistieken. Hopelijk helpt dit boek daarbij en ben je na het lezen in staat om snel, selectief *en* kritisch naar cijfers en statistieken te kijken. Dan is mijn doel bereikt.

I. DE INLEIDING*

* De maker, gebruiker en consument van statistieken

*Cijfers kunnen niet liegen, maar
leugenaars kunnen wel cijferen.*

Toon Verhoeven

Statistiek is een middel, geen doel. Statistici verzamelen gegevens en zetten ze om in cijfers en data. Die verzameling van data kan over van alles gaan. Leeftijden, salarissen, voorkeuren, meningen, lengtes en ga zo maar door. Om een conclusie te kunnen trekken op basis van data, maken statistici gebruik van verschillende wiskundige technieken. Die lopen uiteen van zeer eenvoudig (het *gemiddelde* of het *percentage*) tot zeer complex (de *multivariate variantie-analyse*). Omdat de gebruikte technieken afkomstig zijn uit de wiskunde, wekt statistiek soms de indruk dat het een exacte wetenschap is. Het lijkt al snel alsof de resultaten uit statistische berekeningen vaststaande feiten zijn. Het draait in de statistiek echter niet om de vraag of iets waar is, maar om de vraag of we redelijkerwijs kunnen aannemen dat iets waar is.

Uit de gekozen methode rollen cijfers. Die zeggen vrij weinig tot niets. Pas wanneer aan deze cijfers conclusies worden verbonden, krijgen ze betekenis. Een percentage is gewoon een getal totdat je het omschrijft als “*veel*” of een “*enorme stijging*”. Vaak zijn meerdere conclusies mogelijk. Het is maar net wat je vraagstelling is of vanuit welke invalshoek je kijkt naar de uitkomsten. Het interpreteren van de uitkomsten van statistische berekeningen is dan ook mensenwerk. Dit boek gaat daarover. Verschillende wiskundige technieken en modellen komen aan bod, maar ze vormen niet de kern van het boek. Het gaat om de interpretatie van die technieken en modellen en om de valkuilen die je daarbij tegenkomt.

Wie krijgen te maken met statistiek? Natuurlijk de statistici, oftewel, de *makers* van statistiek. De resultaten die statistici

op basis van statistiek verkrijgen, worden weer gebruikt door andere groepen mensen. Ik noem ze de *gebruikers*. Dan zijn er nog mensen die statistische gegevens tot zich nemen. Dat zijn de *consumenten*. Men kan dus op drie verschillende manieren betrokken raken bij statistiek: als maker, gebruiker of als consument.

De makers

Deze groep ontwikkelt de cijferbrij en levert de resultaten. Makers kunnen onderzoekers zijn, statistici, academici, peilingbureaus of adviseurs. Kortom, een zeer diverse groep die statistieken en grafieken produceert. De één pas na jarenlang onderzoek, de ander al na een korte enquête onder voorbijgangers op straat.

De meeste makers creëren hun statistieken naar aanleiding van een vraag. Een maker van statistieken kan zelf met een vraag zitten of een opdracht krijgen om het antwoord te vinden. Tot op zekere hoogte is iedereen een maker, we berekenen immers allemaal wel eens een gemiddelde of percentage. De makers die ik in dit boek bedoel, produceren statistieken beroepshalve. Het zijn niet allemaal statistici of wiskundigen. Dat lijkt voor simpele technieken als het percentage en het gemiddelde ook niet nodig. Toch zullen we in de loop van dit boek zien, dat een gedegen achtergrond in de statistiek nuttig is voor wie zich vanuit z'n werk bezighoudt met de wondere wereld van statistieken.

De maker heeft drie belangrijke activiteiten te voltooien. Ten eerste moet hij de gestelde vraag op een correcte wijze verta-

len naar een geschikte methode. Met die methode verzamelt de maker vervolgens gegevens waarmee de vraag kan worden beantwoord. Het kan gaan om totaal uiteenlopende vragen zoals: “Wat vindt de Nederlander van het ingevoerde rookverbod?”, “Wat is de kans dat een bepaald product aanslaat bij een specifieke doelgroep?”, of “Hoeveel geld wordt er in Nederland jaarlijks witgewassen?”. Met behulp van statistiek wil de vragensteller daarop antwoorden vinden. Dat lijkt misschien makkelijker dan het is. Aan de maker de schone taak om een methode te kiezen waarmee niet zomaar een antwoord kan worden gevonden, maar een antwoord dat zo dicht mogelijk in de buurt van de werkelijkheid komt. Zo’n methode kan een vragenlijst zijn, een interview, een groep observeren of een andere vorm van dataverzameling.

Is de keuze voor een geschikte methode eenmaal gemaakt dan moet de geselecteerde methode worden toegepast. Eerst de vraag vertalen en dan correct uitvoeren. Als het goed is, vinden de activiteiten in deze volgorde plaats. Soms gebeurt dat niet. Het komt bijvoorbeeld voor, dat een vragenlijst wordt opgesteld en afgenomen zonder dat men vooraf bedenkt of de vraagstelling wel leidt tot de gewenste gegevens. We zullen in dit boek voorbeelden zien van gekozen technieken met een verkeerde toepassing.

Na de keuze voor techniek en toepassing volgt de derde stap. De maker moet een conclusie verbinden aan de resultaten. Het kan voorkomen dat informatie naar boven borrelt die op voorhand onbekend was. Dat maakt statistiek natuurlijk erg interessant, maar logischerwijs verbindt de maker een con-

clusie naar aanleiding van de gestelde vraag.

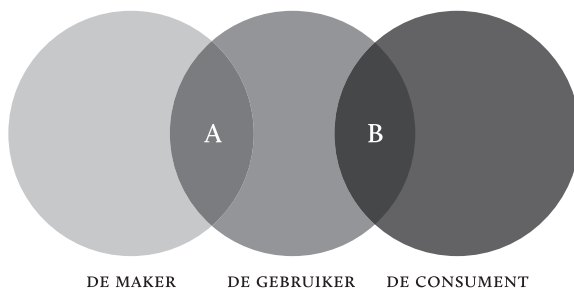
Voor de makers van statistische gegevens bestaat een belangrijke valkuil. Een maker gaat soms net zo lang op zoek naar gegevens tot het gewenste antwoord is gevonden. Blijkt nu dat de gevonden data niet het gewenste resultaat opleveren, dan vraagt de maker gewoon een nieuwe groep mensen naar hun mening. Zelfs als de data beschikbaar zijn, kan het voorkomen dat er net zo lang met statistische formules wordt gesleuteld totdat er een gewenst antwoord uitkomt. Dat is een typisch geval van *“Torture your data until they confess”*. Je redeneert naar je uitkomst toe en dat levert natuurlijk geen betrouwbare onderzoeksresultaten op.

De gebruikers

Nadat de makers hun werk hebben gedaan, is het aan de gebruikers om er iets mee te doen. Gebruikers zijn de boodschappers die de resultaten van statistiek onderzoek uitdragen. Ze zijn zelf niet of nauwelijks betrokken bij de totstandkoming van de cijfers, maar gebruiken ze voor hun eigen verhaal.

De groep van gebruikers is zeer divers en bestaat uit zowel opdrachtgevers als willekeurige gebruikers. Opdrachtgevers geven de makers opdracht om een antwoord te vinden op een vraag die bij ze leeft. Dit is weergegeven door het cijfer A in figuur 1. Zo kan een producent van geneesmiddelen een commercial willen maken over een bepaald medicijn en daarom zoeken naar bewijzen voor de genezende werking ervan. Een krant wil misschien een artikel schrijven over de mening van Partij voor de Vrijheid (PVV) - stemmers over onze samenle-

ving en een onderzoeksbureau bijbehorend cijfermateriaal laten aandragen. Of een minister besluit onderzoek te laten uitvoeren naar de mening van mensen over filevorming, zodat kan worden aangetoond dat het gevoerde beleid succesvol is.



Figuur 1: De drie partijen die hebben te maken met statistiek

De groep gebruikers kan ook bestaan uit mensen die cijfers en resultaten willekeurig gebruiken, namelijk zonder dat ze een link hebben met de maker. Zo kan iedereen op de website van het Centraal Bureau van de Statistiek (CBS) legio statistische gegevens vinden over de meest uiteenlopende onderwerpen. Veel peilingbureaus hebben ook een website waar je onderzoeksresultaten kunt vinden. Een krant kan een persbericht overnemen van een peilingbureau, dat vervolgens wordt gebruikt op de website van een belangenvereniging. Daar kan iemand die informatie vinden en weer in een verhitte discussie gebruiken op een verjaardagsborrel. Statistieken worden op deze manier vaak (her)gebruikt, zonder dat de makers daarvan in de meeste gevallen op de hoogte zijn gebracht.

Bij dit (her)gebruik van statistische gegevens loert het gevaar van manipulatie en misinterpretatie. Het is namelijk de gebruiker die achteraf de meest geschikte cijfers bij zijn vraag zoekt. De statistische gegevens zijn in negen van de tien gevallen echter niet gemaakt om de specifieke vraag van de (her)gebruiker te beantwoorden. Sterker nog, de gebruiker interpreteert de resultaten en geeft er een uitleg aan. Dat is de taak van de maker, maar de link met de maker ontbreekt. Nog een grote valkuil voor dit type gebruiker is, dat statistische resultaten de status van vaststaande feiten krijgen en ze ook als zodanig worden gepresenteerd door de gebruiker. Dit komt niet zelden door een gebrek aan kennis over statistiek: “Als de maker het zegt, dan zal het wel kloppen”. De gebruiker bepaalt dus in belangrijke mate hoe de cijfers worden overgedragen. Dat is in figuur 1 weergegeven met de letter B. Veel voorbeelden in dit boek gaan over (bewuste) missers in het gebruik van de resultaten. Opvallend is dat gebruikers zich zelden verantwoorden voor hun keuzes voor methoden of dataverzameling. Als een gebruiker zonder verdere toelichting maar op een duidelijke manier verwijst naar de bron van de betreffende statistische gegevens, dan is dat in de praktijk al heel wat.

De consumenten

In zekere zin zijn we allemaal consumenten van statistieken. Ook makers en gebruikers lezen kranten, surfen op het internet en kijken televisie. Met een statistiekconsument bedoel ik echter de groep mensen die niet is betrokken bij de totstandkoming van statistieken en die geen invloed heeft op de wijze van presentatie van statistische gegevens.

Zoals eerder aangegeven, zijn statistieken overal om ons heen. We zijn ons er niet altijd van bewust, maar we bevinden ons overdag gewild of ongewild in een *statistiekconsumerende* toestand. We maken ook aan de lopende band keuzes op basis van deze statistieken. Dat varieert van statistieken in de krant over de samenleving op basis waarvan we onze politieke voorkeur bepalen, tot de verminderde percentages verzadigd vet op een verpakking waar ons oog in een winkel terloops op valt.

Als consument zijn we gemakzuchtig en naïef wanneer het gaat om statistiek. We gaan er van uit dat de gebruiker zijn huiswerk heeft gedaan en heeft gecontroleerd of de maker correct tot de resultaten is gekomen. De meeste consumenten nemen niet de moeite om het gat tussen hen en de maker te dichten. Hoe vaak vraagt iemand een onderzoeksrapport op naar aanleiding van een krantenartikel of televisie-uitzending? We gaan er ten onrechte van uit dat de gepresenteerde statistieken altijd een correcte weergave zijn van dat wat de maker heeft geproduceerd.

Het komt ook regelmatig voor dat bepaalde statistieken de consument helemaal niet bereiken. De maker handelt in opdracht van de gebruiker maar die is om uiteenlopende redenen niet altijd van plan om de onderzoeksresultaten openbaar te maken. Bedrijven kunnen er wat van, maar ook de overheid is er goed in. Neem het programma *Overheid voor de Toekomst* dat eind 2009 van start ging. Een omvangrijk programma dat bestond uit veertien modules. Daarin werden thema's behandeld als ICT, samenwerking, ambtenaarschap,

staatsrecht en politiek-ambtelijke verhoudingen. Ook werden congressen georganiseerd, onderzoek gedaan en rapporten geschreven. Allemaal met als hoofdthema: “Hoe maken we de overheid klaar voor de toekomst?”. Door de val van het kabinet halverwege 2010 was het einde van dat programma snel beklonken. Grote kans dat we als statistiekconsument van al deze rapporten nooit iets gaan horen. Of dat erg is, is een andere vraag.

De consumenten van dit boek

Maker, gebruiker of consument: dit boek is geschreven voor iedereen. De makers zien dat hun resultaten vaak op een hele andere manier bij de consument terecht komen dan zij ooit voor ogen hadden. Voor de gebruiker staan er legio adviezen in waarmee ze kritisch kunnen kijken naar wat de maker tot stand heeft gebracht en waarmee ze dit op een eerlijke en transparante manier aan de consument over kunnen brengen. Tenslotte is het voor de statistiekconsument handig om zich de adviezen eigen te maken. De kans is namelijk groot dat niet alle gebruikers dit boekje lezen of er wat mee willen doen.

Ik heb de drie groepen bewust als gescheiden groepen gepresenteerd. Natuurlijk is die scheiding relatief. Makers kunnen gebruikers zijn, gebruikers kunnen consumenten zijn, etc. Toch is in praktijk meestal sprake van het volgende scenario: een onderzoeksbureau verricht onderzoek, al dan niet in opdracht van een derde partij of persoon. In het gepubliceerde rapport staan de belangrijkste resultaten en conclusies. Afhankelijk van de opdrachtgever, bereiken de resultaten vervolgens via de media de consument. Eén beroepsgroep kan

dus in verschillende rollen zitten, bijvoorbeeld *consultants*. Als voormalig adviseur weet ik dat consultancy bureau's zelf ook statistieken maken door onderzoeken uit te voeren en de resultaten te verwerken in hun adviezen.

MAKERS*	GEBRUIKERS*	CONSUMENTEN*
* Wetenschappers	* Journalisten	* Krantlezers
* Onderzoekers	* Reclamemakers	* TV-kijkers
* Opiniepeilers	* Marketeers	* Radioluisteraars
* Consultants	* Politici	* Internetters
* Statistici	* Advocaten	* Bedrijven
	* Consultants	* Rechters

Figuur 2: Makers, gebruikers en consumenten - een niet uitputtende greep uit de verschillende groepen.

De komende twee hoofdstukken richten zich op het gestuntel van achtereenvolgens gebruikers en consumenten. In de daaropvolgende vijf hoofdstukken gaat het vooral over het gekonkel van makers en gebruikers. Elk hoofdstuk wordt afgesloten met een aantal *tips & tricks*, bedoeld voor zowel gebruikers als consumenten. Daarmee hoop ik deze twee groepen wat middelen in handen te geven om door de cijfers heen te kijken. Het boek sluit af met een samenvattend hoofdstuk. Deze samenvatting bevat een korte woorden- en vragenlijst. Het is bedoeld als checklist waarmee je kritisch kunt kijken naar statistische gegevens, zodat je ze op waarde kunt