

De signaal-ruisverhouding van tabellen

Onderzoekers vragen zich ongetwijfeld wel eens af waarom ze onderzoek doen. Het antwoord op die vraag zal verschillen, maar om uiteenlopende redenen zal een onderzoeker behoefte hebben de resultaten van zijn onderzoek onder andermans ogen te brengen. Sterker nog, de onderzoeker zal er naar streven dat met de resultaten van zijn onderzoek iets wordt gedaan. Dan zal hij moeten rapporteren, liefst zo helder mogelijk. Uiteraard maakt het verschil of voor vakgenoten wordt geschreven of (ook) voor niet-vakgenoten. Aan een aantal eisen zal elke rapportage moeten voldoen 1). Is aan die eisen niet voldaan, dan verslechtert de signaal-ruisverhouding zodanig dat de boodschap niet meer overkomt. Het is dus zaak in een rapportage het signaal (de boodschap) zo krachtig mogelijk te maken en de ruis (niet-relevante of verstorende elementen) zo gering mogelijk.

Een belangrijk onderdeel van onderzoeksrapportages zijn *tabellen*. Omdat vooral de compositie van tabellen vaak stiefmoederlijk wordt behandeld, wordt hierna ingegaan op (het minimaliseren van) de ruis in tabellen. Daarbij is het onderscheid tussen *gegevens* en *informatie* van belang. Globaal gezegd zijn gegevens los van elkaar staande elementen, terwijl informatie is te beschouwen als zodanig bewerkte en gepresenteerde gegevens, dat hieraan inzichten kunnen worden ontleend 2).

Veel lezers hebben problemen met het lezen/interpreteren van tabellen, hetgeen nog versterkt wordt door het gebrek aan compositie dat veel tabellen kenmerkt. Ehrenberg gaat in op de presentatie van (getalsmatige) gegevens, met name in tabelvorm 3). Hij geeft enkele saillante voorbeelden hoe door omzetting van tabellen de signaal-ruisverhouding van die tabellen aanzienlijk kan worden opgevoerd. Door een aantal simpele regels te volgen, kunnen veel tabellen aanzienlijk worden verbeterd. Deze regels worden hier in de vorm van tien geboden gegeven, onderverdeeld in drie groepen: getallenrepresentatie, verhouding tabel-tekst, en lay-out en overige onderwerpen. De tien regels worden vervolgens toegepast op een voorbeeldtabel.

Getallenrepresentatie

1. Afronden

Rond af op twee effectieve cijfers,

dat wil zeggen op twee cijfers afgezien van nullen. Voor afronding zijn verschillende argumenten aan te voeren:

- soms zijn opgenomen decimalen niet significant. Dan is opname van die decimalen onzin;
- vaak gaat het er om een globale indruk van een verschijnsel te krijgen en dan is een grove afronding veelal voldoende;
- tot twee effectieve cijfers afgeronde getallen zijn beter te onthouden;
- het is soms verleidelijk een tabel in een artikel voor de *archieffunctie* te gebruiken, maar archiveren moet elders geschieden.

2. Ordenen

Orden tabelregels en tabelkolommen op grootte. Dan is (veel duidelijker) te zien waar het om gaat, dat wil zeggen door zo te handelen komt informatie beschikbaar, die eerst ontbrak.

3. Randwaarden

Geef randtotalen of randgemiddelden om een ijkpunt te hebben voor de overige getallen in de tabel. Getallen in regels en kolommen tellen soms, door afrondingsverschillen, niet op tot 100%. Desondanks moet in de totaalregel 100% staan. In een voetnoot kan vermeld worden dat getallen door afronding niet steeds tot 100 optellen.

4. Getallenstructuur

Laat de onderliggende structuur van getallen zien: zet kolommen in een logische volgorde, met op elkaar afgestemde eenheden. Zet getallen die moeten worden vergeleken in kolommen in plaats van in regels: getallen in een kolom laten zich sneller onderling vergelijken dan getallen in een regel. Het onderliggende verband tussen kolommen onderling en tussen rijen onderling kan zonodig expliciet worden aangegeven.

Verhouding tabel-tekst

5. Autarkie

Zorg dat tabellen op zichzelf duidelijk zijn. Dat houdt in dat ze onafhankelijk van de bijbehorende toelichtende tekst te begrijpen zijn.

6. Tabel-tekststructuur

De structuur van een tabel en van de bijbehorende tekst moeten elkaar ondersteunen en dus de boodschap versterken. Bespreek een tabel in de tekst. Geef daarbij de belangrijkste/op-

vallendste patronen en uitzonderingen op die patronen aan. Geef zo mogelijk steeds een voorbeeld van hoe een tabel moet worden gelezen. Laat de volgorde waarin een tabel wordt besproken, overeenstemmen met de structuur van de tabel. Bespreek de regels in de tabel van boven naar beneden. Begin met een bespreking van de totaalkolom. Ga daarna van links naar rechts de kolommen langs. Zorg dat bij verwijzing naar getallen uit een tabel en in de tekst en in de tabel exact dezelfde getallen staan. Als de behoefte bestaat cijfers uit een tabel in de tekst af te ronden duidt dit er op dat het in de tabel ook kan.

Lay-out en overige onderwerpen

7. Tabelkop

Maak een korte zinvolle tabelkop. Dat geldt ook voor subkopjes in de tabel.

8. Cijferconventie

Hanteer een duidelijke en consistente cijferconventie. De cijferconventie van het Centraal Bureau voor de Statistiek (CBS) is een voor de hand liggende kandidaat 4). Het CBS is uiteindelijk de grootste cijferfabrikant van Nederland.

9. Afkortingen

Verklaar afkortingen - voor zover ze niet duidelijk zijn - in een voetnoot bij de tabel.

10. Bron

Vermeld de bron van alle gegevens.

1) Hierbij valt te denken aan bij voorbeeld een systematische structuur, een consistent begrippenkader, het verklaren van nieuwe begrippen.

2) Op het onderscheid wordt nader ingegaan in C.A. de Kam en C.J. Wiebrens, Inleiding en probleemstelling, in: C.A. de Kam, C.J. Wiebrens en F.G. van Herwaarden (red.), *Bouwstenen voor inkomensbeleid en sociale zekerheid*, blz. 20-23.

3) A.S.C. Ehrenberg, The problem of numeracy, in: *The American Statistician*, mei 1981, jg. 35, nr. 2, blz. 67-71.

4) Deze staat bij voorbeeld in het *Statistisch Zakboek* van het CBS onder 'verklaring der tekens'.

. = gegevens ontbreken

* = voorlopig cijfer

X = geheim of te weinig waarnemingen

- = nihil

Tabel 1. Geschatte kosten in guldens per persoon per uur voor verschillende vrijetijdsactiviteiten

	Bedrag in gld. per huishouden per jaar (1)	Percentage personen die de activiteit in een week verricht (2)	Gemiddeld aantal minuten door (2) besteed (3)	Kosten in gld. per persoon per uur (4)
Bezoek aan concert, opera, toneel, ballet, cabaret	22	4,7	163	1,13
Filmbezoek	20	5,2	135	1,13
Bezoek aan museum, galerie, openluchtmuseum, tentoonstelling	2	4,9	126	0,13
Lezen van kranten, tijdschriften en boeken	471	a)	a)	0,59
Radio luisteren	128	43,7	143	0,81
Televisie kijken	303	96,3	635	0,20
Luisteren naar platenspeler of bandrecorder	151	34,9	110	1,56
Doen aan sport	252	32,4	191	1,61
Bezoek aan sportwedstrijden	47	17,1	151	0,72
Bezoek café, snackbar, sociaal-cultureel centrum, wijkcentrum, clubhuis, buurthuis, dans- en discoavond	303	34,1	183	1,92
Afgelopen jaar met vakantie geweest	1.441	61,0	562	1,67

a) Deze gegevens zijn niet direct beschikbaar. De kosten in guldens per persoon per uur is op een andere manier berekend, namelijk door het gewogen gemiddelde van de leesactiviteiten te nemen. In totaal wordt er per huishouden 795 uur per jaar gelezen. Bron: zie voetnoot 2.

Tabel 2. Vrijtijdsbesteding: tijd en kosten per persoon, 1980

Activiteit	Alle personen	Personen die de activiteit verrichten		
	Bedrag per persoon a) in gld/jaar (1)	in procenten (2)	Gemiddelde tijd in uur/jaar (3)	Kosten per persoon b) in gld/uur (4)
- Bezoek café, snackbar, sociaal-cultureel centrum, wijkcentrum, clubhuis, buurthuis, dans- en discoavond	100	34	160	1,90
- Afgelopen jaar met vakantie geweest	500	61	490	1,70
- Doen aan sport	87	32	170	1,60
- Luisteren naar platenspeler of bandrecorder	52	35	100	1,60
- Bezoek aan concert, opera, toneel, ballet, cabaret	8	5	140	1,10
- Filmbezoek	7	5	120	1,10
- Radio luisteren	44	44	120	0,81
- Bezoek aan sportwedstrijden	16	17	130	0,72
- Lezen van kranten, tijdschriften en boeken	160	c)	c)	0,59
- Televisie kijken	100	96	550	0,20
- Bezoek aan museum, galerie, openluchtmuseum, tentoonstelling	1	5	110	0,13
- Alle activiteiten

a) Gemiddeld over alle personen, ongeacht of ze de activiteit wel of niet verrichten.
b) Kol 4 = kol 1/(0,01 x kol 2 x kol 3).

c) Deze gegevens zijn niet direct beschikbaar. De kosten in guldens per persoon per uur zijn op een andere manier berekend, namelijk door het gewogen gemiddelde van de leesactiviteiten te nemen. In totaal wordt er per huishouden 795 uur per jaar gelezen. Bron: Centraal Bureau voor de Statistiek (CBS), *Budgetonderzoek 1980, 1985*.

Voorbeeld

Het is niet moeilijk in willekeurig welk blad gepubliceerde tabellen te vinden die op één of meer van de hiervoor genoemde punten zijn te verbeteren. Aan de *ESB* van 8 april 1987 is een voorbeeld ontleend (zie tabel 1) waarin interessante gegevens staan, die op een overzichtelijker manier kunnen worden gepresenteerd 5).

De oorspronkelijke tabel is weergegeven in tabel 1.

Ik heb tabel 1 op een aantal punten verbeterd, waarna tabel 2 resulteert. De punten waarop verbetering is aangebracht worden in dezelfde volgorde besproken als de hiervoor opgesomde tien geboden.

Afronden

De kolommen zijn op twee effectieve cijfers afgerond. In kolom 1 zijn enkele cijfers niet zo afgerond, omdat ik niet beschikte over de onderliggende gegevens. De vraag is of de getallen in kolom 4 tot op de cent significant zijn. Los daarvan heeft het verschil in aantal gepresenteerde effectieve cijfers (2 of 3) in bij voorbeeld kolom 4 van tabel 1 verschillende resoluties tot gevolg: één op de 13 cent is 8%, één op de 192 cent is 0,5%. Voor het betoog in het bewuste artikel is het overigens niet no-

dig de getallen heel nauwkeurig te geven.

Ordenen

Er is op de laatste kolom geordend. Ook op andere kolommen zou kunnen worden geordend. De tabelkop suggereert evenwel deze keuze, evenals het feit dat in de tekst alleen kolom 4 wordt besproken en nog wel in geordende vorm. In de tweede in het artikel van Bakker voorkomende tabel kan dezelfde regelvolgorde worden aanhouden als in de hier besproken tabel.

Randwaarden

Er zijn randgemiddelden als ijkpunt toegevoegd, waardoor (sneller) is te zien of een activiteit boven of onder het gemiddelde ligt. Bij gebrek aan de onderliggende gegevens zijn puntjes opgenomen.

Getallenstructuur

De eenheden van de getallen in kolom 1 en 3 en daarmee ook de getallen zelf zijn gewijzigd. Kolom 1 is gewijzigd van 'gld/jaar per huishouden' in 'gld/jaar per persoon'. Een huishouden bestaat gemiddeld uit 2,91 personen. Het bedrag per persoon is dus gevonden door te delen door 2,91. Kolom 3 is gewijzigd van 'minuten/week' in 'uur/jaar'. De eenheden in deze kolommen zijn aangepast, omdat zowel in de

tabel als in de bijbehorende tekst de kosten per persoon per uur centraal zijn gesteld. Voorts wordt ook in de voetnoot als eenheid 'uur/jaar' gebruikt, terwijl in kolom 3 oorspronkelijk 'minuten/week' is gebruikt. Deze wijziging maakt het mogelijk het onderlinge verband tussen de kolommen direct te doorzien. Ten overvloede is dit verband in een voetnoot geëxpliciteerd.

Autarkie

Er is een directe bronvermelding toegevoegd in plaats van de oorspronkelijke indirecte. Voorts is het (zie vorige punt) niet meer nodig de tekst uitputtend te bestuderen alvorens de ge-

0 (0,0) = het getal is minder dan de helft van de gekozen eenheid
niets (blank) = een cijfer kan op logische gronden niet voorkomen

< = minder dan, resp. kleiner dan

> = meer dan, resp. groter dan

≦ = minder dan, resp. kleiner dan of gelijk aan

≧ = meer dan, resp. groter dan of gelijk aan 1986 - 1987 = 1986 tot en met 1987

1986/1987 = het gemiddelde over de jaren 1986 en 1987

1986/87 = oogstjaar, boekjaar, schooljaar, enz., beginnend in 1986 en eindigend in 1987.

5) B.F.M. Bakker, *Draagkracht en vrijetijdsbesteding*, *ESB*, 8 april 1987, blz. 337-340.

tallenstructuur voorzien kan worden.

Tabel-tekststructuur

Wat kolom 1 tot en met 3 betreft zijn er twee mogelijkheden. Ofwel ze zijn niet essentieel/informatief (het gaat uiteindelijk om kolom 4) en dan worden ze terecht niet in de tekst besproken. Ofwel ze zijn wel essentieel/informatief en dan moeten ze wel in de tekst worden besproken. Duid begrippen steeds op dezelfde manier aan. Bij voorbeeld niet in de tabel 'kosten' en in de tekst 'uitgaven'.

Tabelkop

De tabelkop is verkort door minder essentiële elementen weg te laten. De subkopjes zijn zoveel mogelijk van ruis ontdaan. Toevoeging van 'activiteiten in een week verricht' in kolom 2 aan 'percentage personen' is overbodig. Het kopje van kolom 3 is door verwijzing naar kolom 2 verwarrend. Onderaan de subkopjes zijn (duidelijker) eenheden afzonderlijk toegevoegd.

Bron

Aan de tabelkop is het jaar toegevoegd, waarop de gegevens betrekking hebben.

Tot slot zijn enkele taalfoutjes verbeterd. Na deze transformatie hebben we tabel 2. De informatiewaarde van tabel 2 is duidelijk toegenomen 6). Informatie die eerst niet in de tabel aanwezig was, komt nu boven drijven. Er is veel beter te zien hoe de getallen in kolom 4 tot stand zijn gekomen. Het is nu ook mogelijk de eventuele correlatie tussen de kolommen te zien: het zou in de lijn der verwachtingen liggen, dat bij dalende kosten per uur (kolom 4) de deelname stijgt (kolom 2) en dat ook de aan een activiteit bestede tijd stijgt (kolom 3). Dit blijkt echter niet het geval te zijn. Het ontbreken van deze verbanden is niet of heel moeilijk uit de oorspronkelijke tabel 1 te halen.

Nu zou de lezer kunnen denken dat dit voorbeeld gezocht is, maar een snelle blik in de tot nu toe in de jaargang 1987 verschenen afleveringen leert dat vele tabellen voor verbetering vatbaar zijn. Ook de onderwerpen van de artikelen waarin die tabellen te vinden zijn lopen zeer uiteen. Zie bij voorbeeld de volgende bladzijden van *ESB*, jg. 1987: blz. 100 (spaarmotieven), blz. 129 (handelsbalanssaldi), blz. 188 (inkomensverbetering), blz. 279 (scheepbouwbeleid), blz. 422 (joint ventures).

Conclusie

Het is mogelijk dat de hier gegeven tabeltransformatie op onderdelen niet helemaal klopt omdat de onderliggende gegevens niet beschikbaar waren. Desondanks zijn de 'tien geboden' op hun merites te waarderen. Uit de hier gegeven tabeltransformatie en uit andere voorbeelden die voor het oprapen liggen, blijkt dat de compositie van tabellen meer aandacht vergt. Daardoor kan het ruisniveau dalen, waardoor de informatiewaarde van tabellen aanzienlijk kan stijgen en volstaan kan worden met minder toelichtende tekst. Met andere woorden: 'Had I more time, I would write you a shorter letter'.

Floor Gerard van Herwaarden

Werkzaam op het Sociaal en Cultureel Planbureau te Rijswijk. De auteur dankt ir. A.W. van Herwaarden en mr. C.A. de Kam voor hun commentaar op een eerdere versie van dit artikel.

6) Het is overigens opmerkelijk dat de kosten voor filmbezoek f. 1,10 per uur zijn, terwijl de toegangsprijs voor de bioscoop in de orde van f. 12 voor twee uur ligt.
