DRUK: TEN HAGEN N.V. - DEN HAAG

(4)

EERSTE JAARGANG - NUMMER 3 - SEPTEMBER 1968

tijdschrift voor marktonderzoek

NEDERLANDSE VERENIGING VAN MARKTONDERZOEKERS

N.V. MARKON MARKETING-KONSULENTEN

In het kader van de introductie van een Marketingplan voor de komende jaren wordt het bestaande assortiment en de mogelijke uitbreiding daarvan bij een onzer relaties — een grote industriële onderneming in het westen van het land — aan een nauwkeurige analyse onderworpen.
Tegen deze achtergrond is de behoefte gerezen de nu nog kleine marktonderzoek-afdeling uit te breiden met een ervaren

MARKTONDERZOEKER

Zijn taak zal o.m. bestaan uit:

- Het analyseren van de E.E.G.-markt voor de bestaande en nieuwe produkten van de onderneming
- Het opzetten en uitvoeren resp. doen uitvoeren door externe bureaus van onderzoeken voor de produkten waarbij op industrieel marktonderzoek de nadruk ligt
- Het interpreteren van interne en externe gegevens en het verwerken ervan in een rapport ten dienste van de beleidsbeslissingen.

Geboden wordt:

- Een verantwoordelijke functie met ontplooiingsmogelijkheden binnen de zich uitbreidende marktonderzoek-afdeling waarbij eigen initiatieven en t.z.t. ook leidinggevende kwaliteiten tot hun recht moeten komen
- Goede arbeidsvoorwaarden, waaronder directe opname in het pensioenfonds. Zonodig hulp bij de huisvesting.

Gevraagd wordt:

- Leeftijd tussen 28 en 35 jaar
- Ervaring op het terrein van marktonderzoek, bij voorkeur op industrieel gebied
- Redelijke kennis van de moderne talen
- Bij voorkeur kennis en vorming op academisch niveau.

Degenen, die menen in aanmerking te komen voor deze functie, wordt verzocht een schriftelijke sollicitatie in te dienen, met pasfoto, t.a.v. de heer M. J. J. Veraart, N.V. Markon Marketing Konsulenten, J. J. Viottastraat 52 te Amsterdam.

Degenen, die meer bijzonderheden willen weten alvorens te solliciteren kunnen desgewenst een afspraak maken voor een persoonlijk onderhoud. Telefoon (020) - 762323.

tijdschrift voor marktonderzoek

uitgave: nederlandse vereniging van marktonderzoekers

Jaargang 1, september 1968, nummer 3

Redaktiecommissie: drs. H. Emanuel, voorzitter; F. A. A. Claasen secretaris; drs. C. C. J. Koning; drs. P. H. van Westendorp.

Redaktieadres: Kwartellaan 20, 's-Gravenhage.

Administratie: p/a Mr. C. C. Th. van Andel, Keizersgracht 586, Amsterdam-C.

Abonnementsprijs: *f* 20,—; studenten en cursisten marktonderzoek L.O.I. *f* 10,—, los nummer *f* 6,—.

kwaliteit beoordelen? geef het maar een cijfer

door J. Stapel

Algemeen wordt aangenomen dat attitude-schalen op z'n best niet meer dan ordinale schalen zijn. Zij zetten proefpersonen in het gelid, in een rangorde, meer niet. Resumerende beschrijvingen van schaal-resultaten in termen van medianen, rekenkundige gemiddelden e.d. zijn dan ook ongewenst en steekproeftoetsen op zulke reken-uitkomsten worden niet toegepast. Dat is een gangbare opvatting.

Maar nu is er in de Verenigde Staten ene Stanley Smith Stevens en die houdt zich bezig met onderzoek (in de voetsporen van o.m. Fechner) over de relaties tussen stimuli en sentaties. Zijn methode lijkt bijzonder veel op het gebruik van numerieke schalen, op „cijfers geven". Een van de eenvoudigste proefnemingen bestaat uit het laten schatten door proefpersonen van lengten of lengte-verschillen. *Gemiddeld* blijkt (let wel: rekenkundige gemiddelden) dat gemiddelde geschatte lengte $= C \times (\text{opgemerkte lengte})^n$. (Waarin in dit geval de constante (C) en de exponent (n) iets kleiner zijn dan één. M.a.w. lengte-maten worden gemiddeld op het oog iets onderschat). Stevens heeft minder eenvoudige dingen laten schatten[1]) — steeds in cijfers —: helderheden van lichtbronnen b.v. en elektrische spanningen. (Hoeveel hoger schat u het voltage van schok 2 in vergelijking met schok 1). Hoewel proefpersonen onderling aanzienlijk verschillen in de cijfers die zij toekennen en zich vaak onzeker voelen over hun beoordelingen, blijken de *gemiddelde* cijfers stabiel en repro-

[1]) S. S. Stevens: „The quantification of sensation", 1959.

duceerbaar. M.a.w. 100 volt levert, bij verschillende steekproeven van proefpersonen, steeds weer hetzelfde gemiddelde cijfer. (En, dáár gaat het Stevens om, steeds komt er een patroon uit, beschrijfbaar als „Gemiddeld proefpersonen-cijfer" = C x $(\text{stimulus})^n$). C en exponent n zijn zeer verschillend voor verschillende onderzochte objekten. De stimulus wordt steeds technisch/objektief gemeten: centimeters, volt, enz.). De toegediende prikkels liggen natuurlijk op een *ratio-schaal* (0 tot x cm, 0 tot y volt, enz.). Van de reakties wordt hetzelfde verondersteld, *uiteraard omdat zij beschouwd worden als schattingen van op een ratio-schaal afzetbare kwantiteiten.* Vandaar ook dat Stevens zo onbekommerd met gemiddelden werkt.

Zodra men in marktonderzoek objekten laat „raten", zelfs indien dat b.v. op een numerieke schaal van 0 tot 10 gebeurt, bekruipt menige onderzoeker gerede twijfel of hij ooit b.v. een rekenkundig gemiddelde der uitkomsten zou mogen vermelden, laat staan b.v. de standaard-deviatie van zo'n reken-uitkomst. Maar als wij nu eens, net als voor spanningen of lichtsterkte, zouden veronderstellen dat menig kwantitatief objekt extern meetbaar zou zijn in ratio-schaal-termen? Zodat wij een cijferwaardering door ondervraagden zouden mogen opvatten als schattingen van kwantificeerbare grootheden? (Dus net zoals docenten cijfers geven (zelfs voor een opstel), terwijl proefwerkcijfers — zelfs van verschillende docenten — weer vrolijk gemiddeld worden voor een rapportcijfer . . . tot het bittere of vreugdevolle einde van zakken of slagen toe). Zou dat nu zo gek zijn voor een schaalwaardering door consumenten van b.v. de voordeligheid van een bepaald tandpasta-merk? Dat komt tenslotte neer op „aantal keren tanden poetsen per tube ad prijs X". Of voor een cijfer geven voor de veronderstelde brandstofzuinigheid van een automerk? Dat komt neer op een index-cijfer (b.v. tussen 0 en 10) toekennen voor „aantal kilometers per liter benzine".

Het is eigenlijk verbluffend met welk gemak consumentensteekproeven waarderingscijfers e.d. toekennen aan multidimensionale items als vormgeving van produkten, smaken, geuren, handigheid van verpakking, „interessantheid" van advertenties, enz. enz. En van steekproef tot steekproef en geruime tijd lang blijken frequentie-verdelingen, mediaan, modus, gemiddelde enz. veelal hoogst stabiel en reproduceerbaar. Sterker nog: een verschil in de tijd op zulke schaalwaarderingen blijkt praktisch altijd te volgen op wezenlijke, objektief vaststelbare veranderingen in het objekt. Zou men, ten aanzien van de vrees om kwalitatieve objekten met cijfers te lijf te gaan, niet nog een stapje verder kunnen gaan en als volgt redeneren: Indien consumenten moeiteloos een cijfer afgeven, dan „zien" zij het objekt ook als meetbaar op een interval- of ratioschaal en ergo: dat is het dan ook, want wij meten dat „zien", dat beoordelen door subjekten en niet de kwaliteit van het objekt zelf. Dan zijn wij trouwens weer dicht bij het economische gebeuren, waarin alles in prijzen gemeten wordt, dus op de ratio-schaal van guldens en centen. (Het is trouwens tòch goed om in *markt*-onderzoek het rinkelen der kasregisters nooit geheel te vergeten).

Mochten deze vluchtige bespiegelingen voer voor de statistici en psychologen onder onze lezers zijn, des te beter. Is men het er mee oneens: Fijn!

kwaliteit beoordelen? geef het maar een cijfer . . . . als je durft?

(enkele kanttekeningen bij Stapel's overpeinzingen)

door drs. P. H. van Westendorp

Wordt eigenlijk wel in het algemeen aangenomen dat „attitudeschalen op z'n best niet meer dan ordinale schalen zijn", zoals Stapel stelt? Dit is een opmerking die misschien te vinden is in meer gedegen leerboeken, doch de praktijk leert, dat vrijwel alle marktonderzoekers aan hun attitude- en ratingschalen intervaleigenschappen toekennen zonder veel gewetensbezwaren. Er worden scores opgeteld en afgetrokken, gemiddelden berekend en standaarddeviaties, alle operaties die strikt genomen ontoelaatbaar zijn bij slechts ordinale schalen. Er is in de praktijk van een uitgebreid „als-of-gedrag".

De gangbare opvatting is zeker niet, dat - aldus Stapel - „beschrijvingen van schaalresultaten in termen van medianen, rekenkundige gemiddelden e.d. ongewenst zijn" en dat „steekproeftoetsen op zulke rekenuitkomsten niet worden toegepast". Het berekenen van een mediaan (gebaseerd op de frequentieverdeling der scores) is een toelaatbare operatie bij ordinale scaling. Er zijn talrijke non-parametrische (verdelingsvrije) toetsen (zoals Wilcoxon en Mann Whitney U[1]) om zulke steekproefuitkomsten te toetsen. In het algemeen kan worden opgemerkt, dat het systeem der getallen een aantal nuttige eigenschappen heeft voor het beschrijven van delen van de werkelijkheid. Men kan er zeker niet de hele werkelijkheid adequaat mee beschrijven, noch kan men die beschrijving altijd op hetzelfde niveau uitvoeren. De eigenschappen waar het om gaat zijn op verschillende plaatsen nauwkeurig omschreven[2]), het betreft hier klassifikatie, rangordening en meting ofwel echte kwantificering. Als we proefpersonen een deel van de werkelijkheid laten beschrijven met behulp van een getallensysteem maken deze onbewust gebruik van de logische eigenschappen daarvan waarbij een mogelijkheid is, dat hetzelfde systeem der natuurijlke getallen wordt toegepast bij operaties op verschillend niveau. Maakt men gebruik van meetinstrumenten, d.w.z. hulpmiddelen waarmee de waarneming nauwkeuriger gekwantificeerd kan worden (lineaal, spanningsmeter, psychogalvanometer en dergelijke) dan maakt men gebruik van de interval of ratio-eigenschappen van het getallensysteem. Dit vereist echter operationalisering waarbij wij bedoelen,

1) Zie: Siegel S., Non Parametric Statistics Mac Graw Hill 4-1956.
2) Torgerson W. S., Theory and problems of scaling. Wiley New York 1958.

dat aanneembaar gemaakt moet worden, dat de „werkelijkheid" (waargenomen) aansluit bij de eigenschappen van het getallensysteem of, anders gezegd in de termen van de logische positivisten, daarmee verbindbaar is.

Een voorbeeld: Wij beschouwen drie ziekten: griep, roodvonk en typhus. Dit zijn drie identificeerbare ziekten (ordinale scaling!) die vrijwel iedereen die ze kent in de volgorde van „ernst" zal plaatsen (griep = minst ernstig). Voor de leek houdt daarmee de meetmogelijkheid definitief op: een uitspraak als „typhus is vier maal zo ernstig als roodvonk" of als „griep is half maal zo ernstig als roodvonk" heeft buiten de rangschikking die bewaard blijft geen enkele betekenis omdat de „leek" geen waarneembare eigenschappen zal „verbinden" met de interval of zelfs ratio-eigenschappen van het getallensysteem. Deze operationalisering is niet onmogelijk, maar van belang is vooral dat er verschillende mogelijkheden zijn die leiden tot totaal verschillende scaling:

— het aantal patiënten dat aan de ziekte overlijdt
— het aantal ziektedagen der patiënt
— het temperatuurverloop
— het aantal artsen-visites per patiënt

enz. waarbij dan een uitspraak in kwantitatieve (interval) zin een preciese betekenis krijgt die te allen tijde blijft samenhangen met de gekozen operationalisering.

Er zijn duideijlk tal van waarnemingen of gewaarwordingen die *niet* te operationaliseren zijn op de intervalniveau, maar wel op nominaal of zelfs ordinaalniveau. Ik vind ossehaas lekkerder dan rookworst en rookworst weer lekkerder dan gehakt, doch heeft het zin de afstanden tussen deze vleesprodukten te bepalen zonder operationalisering? Wat is de betekenis van een uitspraak als „ik vind zwezerik drie maal zo lekker als rookworst en rookworst vier maal zo lekker als gehakt"? Het is moeilijk hierbij een operationalisatie te bedenken, anders dan de rangordening.

De moeilijkheid met proefpersonen in het algemeen is, dat wij bij rating gebruik maken van het getallensysteem dat onder omstandigheden intervaleigenschappen heeft terwijl zij in feite slechts gebruik maken van de rangschikkingseigenschappen die steeds mede in het systeem verborgen zijn omdat operationalisering niet mogelijk is of niet konsekwent en unidimensioneel door te voeren. Aangezien dezelfde mogelijkheden van het getallensysteem zich ook onbewust opdringen aan de onderzoeker is deze maar al te snel geneigd een „zin" te ontdekken in het materiaal waar deze niet aanwezig kan zijn.

De constanties die S. S. Stevens[3]) ontdekt op het terrein der psychofysische scaling bewijzen geenszins, dat operationalisering op intervalniveau steeds heeft plaats gevonden:

[3]) Stevens S. S., Stapel cit.

1. In sommige gevallen kán er sprake zijn van deze operationalisering, bijvoorbeeld bij het schatten van afstanden. Iedere sergeant-instrukteur weet, dat men dit aan rekruten kan leren door hen begrip bij te brengen voor standaardafstanden.
2. Stabiliteit van gemiddelden of van frequentieverdelingen is een noodzakelijke, doch geenszins voldoende voorwaarde bij interval-scaling. Immers: als ik personen laat kiezen uit een willekeurige reeks getallen (bijv. 1—10) zonder verdere eisen dan is de kans goort dat de frequentieverdeling der keuzen uitstekend reproduceerbaar is omdat deze tot stand komt via het normale a-selektieproces of via bepaalde cultureel bepaalde voorkeuren voor getallen als 3 en 7.
   Sterker nog, bij klassifikatie van objekten (nominale scaling) of bij rangschikking van objekten waarbij intervalscaling volstrekt onmogelijk is treedt dezelfde constantie op in de frekwentieverdelingen bij meting - hermeting, een verschijnsel dat dus kennelijk niets te maken heeft met de veronderstelde „externe meetbaarheid van een kwalitatief objekt in ratio-schaal termen" (Stapel).

Stapel stelt: „Zou het nu zo gek zijn voor een schaalwaardering door konsumenten van b.v. de voordeligheid van een bepaald tandpasta-merk? Dat komt tenslotte neer op „aantal keren tanden poetsen per tube ad prijs X". Dat is nu precies waar het om gaat! Als de konsument inderdaad bij de „rating" voordelig deze operationalisering in het hoofd heeft, dan is intervalscaling aanneembaar. Dit mogen wij zeker niet zonder bewijsvoering aannemen want:

— uit enquêtes blijkt meestal dat de ondervraagde huisvrouw er geen flauw benul van heeft hoe vaak gepoetst kan worden met een tube (net zo min als ik weet hoeveel pijpen er uit een pakje tabak gaan),
— bij doorvragen (introspectie) blijkt dat de meeste huisvrouwen „voordelig of niet" op geheel verschillende wijze operationaliseren:
  — met de eenheidsprijs per tube
  — met de grootte van de tube
  — in relatie tot merkpreferentie (Halo-effect)
  etc.

Stapel's interval-aanname (of zelfs ratio-aanname) zou misschien gewettigd zijn als in de vraagstelling duidelijk de operationalisering was opgenomen én als deze operationalisering voor de proefpersoon mogelijk is via de waarneming. Konsumenten geven inderdaad met verbluffend gemak cijfers voor de vreemdste dingen.

Daarbij is het niet verbluffend dat er zekere konstanties waarneembaar zijn — zoals wij betoogden kunnen die meer samenhangen met de toekennings-akten dan met de relatie persoon-objekt —, maar wél, dat marktonderzoekers zo vlotjes geneigd zijn aan die cijfers eigenschappen toe te kennen die ze kennelijk voor de poerfpersonen niet hebben, of zelfs maar kúnnen hebben. De opmerking van Stapel dat „een verschil in de tijd op zulke schaalwaarden

praktisch altijd blijkt te volgen op wezenlijk objektief vaststelbare veranderingen in het objekt" lijkt mij even irrelevant ter ondersteuning van zijn stellingen omtrent interval-scaling als zijn opmerking omtrent het konstante gemiddelde. Vind ik Jetje aardiger dan Marietje en deze weer aardiger dan Fientje en verandert mijn verhouding tot Marietje dan is de kans groot dat mijn rangorde gaat variëren, ook zonder dat ik in staat ben „aardig vinden" te operationaliseren op interval-niveau. Ernstiger is, dat volgens de ervaring van veel marktonderzoekers scaling zo merkwaardig constant is, dat een verandering in objekten relatief zelden tot uitdrukking komt. Het is op het ogenblik ten enen male onmogelijk te *voorspellen* hoe scalings gaan veranderen bij een bepaalde verandering in b.v. de advertising-policy. Iedereen wordt op zijn tijd gekonfronteerd met de opmerkelijke constantie van „images" over de jaren heen waarbij het maar de vraag is, of deze constantie werkelijk een psychologische is (de traagheid van het stereotype) dan wel een artefact veroorzaakt door de relatie persoon-getal. Tenslotte: na het lezen van het lucide betoog van De Groot[4]) over de achtergronden bij het geven van schoolcijfers komt men eerder op de gedachte dat deze cijfers vrijwel niets betekenen (in de zin van fatsoenlijke meting) dan anderszins. De wijze waarop schoolcijfers worden uitgedeeld vormt eerder een krachtige ondersteuning van mijn opwekking tot voorzichtigheid met het niveau van metingen, dan een bewijsstuk voor de relevantie van het toepassen van getallensystemen bij de beoordeling van een stuk werkelijkheid. Pikante bijzonderheid: juist in het onderzoek van De Groot blijkt overduidelijk dat de toekenning van cijfers gehoorzaamt aan een aantal wetmatigheden (ook daar opmerkelijk constanties!) die niets maar dan ook niets met de prestaties van de ongelukkige slachtoffers van de beoordeling te maken hebben. Een ongelukkiger illustratie had Stapel nauwelijks kunnen kiezen, als hij tenminste niet bedoelt aan te tonen, dat het geven van cijfers in belangrijke mate een sociaal geaccepteerd stuk nonsens is binnen bepaalde groepen.

In antwoord op Stapels betoog zou ik voor de marktonderzoeker de volgende stellingen willen poneren:

1. Aanname van interval-niveau kan vrijwel bij alle gebruikelijke rating-procedures logisch niet aannemelijk worden gemaakt en is meestal irrelevant omdat deze aanname géén bijdrage vormt tot een betere kennis omtrent de werkelijkheid.
2. Men kan hoogstens stellen dat in een aantal gevallen de aanname van intervalniveau niet zal leiden tot totaal verkeerde konklusies omtrent de werkelijkheid omdat de rangschikkings- en klassifikatie-eigenschappen van het getallensysteem ook bij deze operaties behouden blijven.
3. Bij de huidige ontwikkeling der statistische methoden heeft het dikwijls geen zin onderzoekmateriaal geweld aan te doen door interval- of ratio-niveau te veronderstellen omdat bepaalde rekentechnieken dan beter toepasbaar zouden zijn. Er zijn ten slotte non-parametrische methoden te over terwijl ook op het gebied der multi-dimensionele low-level scaling vorderingen worden gemaakt[5]).
4. Verlangt men interval-scaling van een proefpersoon, of wil men als onderzoeker intervalstatus aan bepaalde gegevens toekennen dan dient de operationalisering expliciet gesteld te worden, of in de vraagstelling, of in de methodologische verantwoording.
5. Bepaalde gegevens kunnen door een proefpersoon nooit zodanig worden geoperationaliseerd dat interval-niveau aannemelijk wordt. In dergelijke gevallen suggereert het gebruik der natuurlijke getallen een pseudo ratio- of intervalkarakter.
6. Constantie-verschijnselen bij scaling op welk niveau dan ook, bewijzen niet dat er sprake is van „verborgen interval-scaling", of van intervaleigenschappen van een stuk waarneembare werkelijkheid.
7. Het is gevaarlijk uitkomsten uit de psychofysische scaling waarbij operationele verbindbaarheid met intervaleigenschappen een zekere evidentie heeft te generaliseren over attitude-meting e.d. Hier is duidelijk sprake van een heel andere klasse van metingen.

[4]) De Groot A. D., Vijven en zessen. Wolters Groningen 1966.

[5]) Coombs, C. H., A theory of data. Wiley New York 1964.

kwaliteit beoordelen? noem desnoods een prijs!

door J. Stapel

Heerlijk! Discussie, en bovendien (o.m.) in fijne huis-, tuin en keuken-termen als zwezerik, rookworst en gehakt. Ook dáárvoor dank aan Van Westendorp. Wij nemen aan dat het hem en andere lezers niet is ontgaan dat wij van een extreem „conservatief" standpunt vol „gewetensbezwaren" (niet cijferen, geen toetsen) via een tussenpositie met veel vraagtekens zijn doorgemarcheerd naar het ietwat provocerende tegendeel: „Indien consumenten moeiteloos een cijfer afgeven, dan „zien" zij het objekt ook als meetbaar op een interval- of ratioschaal en ergo: dat is het dan ook, want wij meten dat „zien", dat beoordelen door subjekten en niet de kwaliteit van het objekt zelf". Dat zou b.v. het geval zijn indien Van Westendorp, nu eens in de rol van proefpersoon, gehakt een één zou geven, rookworst een vier (4 × zo lekker als gehakt/v. W.) en zwezerik een twaalf (3 × zo lekker als rookworst/v. W.). Dat zegt dan natuurlijk niets over deze vleessoorten an und für sich. Maar het zegt wel iets (daarover zijn wij het beiden eens) over zijn rangordening, en misschien meer dan alleen de rangorden in zijn vlees-consumptie-gedrag. Indien de redaktie het toestaat, zouden wij graag nog twee onderzoek-uitkomsten te berde willen brengen. Eén komt uit een vrij recente enquête over ontwikkelingshulp en de andere uit de dissertatie van Prof. Van de Woestijne (1953). De desbetreffende enquête-vraag luidde: „Ontwikkelingshulp kost geld. Zoudt u bereid zijn er een financiële bijdrage voor te geven; eens per jaar b.v.? Hoeveel gulden per jaar bent u bereid te geven?" Van de Woestijne analyseerde de 9.454 giften (ad totaal *f* 82.862,82) die de V.P.R.O. in november 1952 voor Albert Schweitzer en zijn Lambarene ontving na een desbetreffende radio-uitzending:

V.P.R.O.-giften Schweitzer

0,98%	*f* 100,—	en meer
2,75	*f* 25,—	en meer
8,90	*f* 20,—	en meer
10,60	*f* 15,—	en meer
11,20	*f* 12,50	en meer
34,60	*f* 10,—	en meer
35,80	*f* 7,50	en meer
59,0	*f* 5,—	en meer
60,0	*f* 4,—	en meer
62,50	*f* 3,—	en meer
91,0	*f* 2,—	en meer
99,0	*f* 1,—	en meer
100,0%	*f* 0,50	

NIPO (G-177) - Bereid per jaar te geven voor ontwikkelingshulp

11,5%	*f* 100,—	en meer
34	*f* 25,—	en meer
36	*f* 25,—	en meer
38	*f* 15,—	en meer
70,5	*f* 10,—	en meer
85	*f* 5,—	en meer
85	*f* 4,—	en meer
91,5	*f* 3,—	en meer
95	*f* 2,—	en meer
100%	*f* 1,—	en meer

(100% = 418 ondervraagden.
543 wilden niets geven /n = 961)

Hoe fraai beide reeksen ook parallel lopen, noch de *f* 82.862,82 die de V.P.R.O. ontving, noch de verbale betalingsbeloften die de interviewers registreerden, noch de schijn van „hardheid" (geld) van deze gegevens, zeggen iets over intrinsieke eigenschappen van Albert Schweitzer of „ontwikkelingshulp". Zij zeggen alleen iets over de schenkers en de ondervraagden zelf, zij het in hun relatie tot resp. Schweitzer en ontwikkelingshulp.

Natuurlijk kan en mag de onderzoeker zich nu allerlei voorstellingen gaan maken over het objekt ontwikkelingshulp. Wij wierpen de vraag op of een objekt-voorstelling van b.v. een intervalkarakter nu zo gek zou zijn. Anders gezegd: indien respondenten in hun relatie tot een objekt zich lang (b.v.) een guldenslineaal opstellen, zit er wellicht een „lineaal-aspekt" aan het objekt zelf. (Hoewel als de lengte langs een lineaal werd gemeten, de dikte nog onbekend is en er tevens nog menige andere ongemeten of misschien onmeetbare dimensie aan het objekt kan zitten). Voordeligheid van een tandpastamerk kàn bewust zijn als „aantal keren poetsen per tube ad prijs X". Alleen dan, stelt Van Westendorp, zou men aan intervallen mogen durven denken. (Dus vermoedelijk eerder bij ons andere voorbeeld: „aantal kilometers per liter benzine" dat zoveel automobilisten drommels goed weten). Er kan echter ook geredeneerd worden dat, zelfs als prijs en aantal keren poetsen onbekend blijken, er toch van een voordeligheidsvoorstelling bij consumenten sprake moet zijn (desnoods alleen maar voordeligheids-suggestie of „halo"), zodra zij een schaalpositie menen te kunnen aanwijzen. Misschien zou zelfs bij anders gericht doorvragen blijken dat zij tot schattingen van de prijs per poetsing in staat zijn. Waarschijnlijk feitelijk onjuist, maar wel van betekenis voor hun oordelen en handelen.

Daarna gingen wij door naar een extremer standpunt dat in wezen hierop neerkomt: Laat de consument desnoods niet eens kùnnen weten hoe het zit met benzine-verbruik of keren tanden poetsen . . ., zodra hij subjektieve en wellicht onjuiste indrukken op een cijferachtige schaal kan samenvatten, dan rangschikt men zich in een interval en dan mag er met deze plaatsbepaling kwantitatief geëxcerseerd worden. Indien de overgrote meerderheid der docenten op bizarre gronden cijfers toekent is dat pech voor de scholieren. Maar voor overgaan of zitten blijven is het cijfergemiddelde beslissend, wat ook de „werkelijke"(?) kwaliteiten van de leerling en zijn prestaties geweest moge zijn. (Zo gek is dit voorbeeld nu ook weer niet).

„Operationeel" in de zin van „Wat gaat er gebeuren?" zijn die rapportcijfers, zijn menigmaal de waarderingscijfers van consumenten, zijn de giften die men geeft of wil geven, zijn de prijzen die men bereid is te betalen. Graag zij toegegeven dat het geven van cijfers ook een sociaal geaccepteerd stuk nonsens kan zijn en zelfs in marktonderzoek vaak te onkritisch, met te weinig gewetensbezwaren wordt toegepast. Ons dunkt echter dat de nonsens vooral daar zit, waar men in de war is over het objekt van onderzoek. Als men over de zwezerik, de rookworst en gehakt zelf uitspraken doet op grond van waarderingscijfers, is de onderzoeker wellicht vergeten wat hij aan het doen was. Als men

op grond van dezelfde cijfers uitspraken doet over de kansen dat zwezerik, rookworst of gehakt op het menu der werkelijke onderzoek-objekten: *de ondervraagde consumenten,* zullen verschijnen, is men dichter bij huis. (Het zijn dezelfde soort kansen als overgaan of zitten blijven op grond van de rekenkundige gemiddelden van zelfs onrechtvaardige docenten-cijfers). Laat ons straks, na één kanttekening bij een ander punt van Van Westendorp's boeiende betoog, de probleemstelling van onze eerste bespiegeling in andere woorden herhalen.

Maar eerst dan dit: Veel „images" zijn uitermate constant in de tijd. Dat geldt bijvoorbeeld voor de (o.m.) op een resumerende schaal geregistreerde corporate images van veel grote Nederlandse bedrijven. Maar één is er (in periodiek onderzoek) enorm op en neer gegaan: de K.L.M. Maar dáár is ook nogal iets gebeurd. Reklame en „advertising policy" blijken vaak invloeden van te geringe kracht om op scalingresultaten veel invloed uit te oefenen, tenzij men b.v. ten behoeve van pre-test vergelijkingen de verschillen welbewust stimuleert. De probleemstelling dan, waarop wij aan het eind van onze eerste bespiegeling wellicht onvoldoende duidelijk doelden, is op de eerste plaats deze: Welke „werkelijkheid" zijn wij aan het onderzoeken door de registratie van consumenten-attitudes? Leggen wij die „werkelijkheid" *in het objekt* waarover ondervraagden hun instelling weergeven, dan bestaat er kans op „sociaal geaccepteerde nonsens". Dan is men *altijd* met „alsof-gedrag" bezig. Dat mag natuurlijk best, maar het dient wel expliciet gesteld te worden.

Beperken wij die „werkelijkheid" tot onze ondervraagden (en hun uitspraken), dan blijven wij op vastere grond en kunnen ons binnen die beperking veel meer permitteren. „Zien" ondervraagden (net als de docenten) iets als uitdrukbaar op een cijferschaal, dan mogen wij hun „rapportcijfers" ook op face value nemen, zolang wij ons maar blijven realiseren dat wij met de waarderingen van deze subjekten bezig zijn, maar niet bezig met de „werkelijkheid" van het beoordeelde objekt zelf. Zolang de subjekten (consumenten en docenten b.v.) beslissen over wat er met het objekt zal gebeuren (veel of weinig kopen, wel of geen diploma toekennen) is dat ook een zinnige bezigheid. De 7 stellingen van Van Westendorp doen stuk voor stuk ter zake voor het middenstuk van onze eerste bespiegeling. (De vraagtekens werden daar niet zonder reden geplaatst). Ze zijn o.i. echter overbodig voor de eindstelling, thans, dankzij Van Westendorp's interventie, uitvoeriger en hopelijk duidelijker uitgewerkt.

factoriële opzet bij produkttest

door drs. E. N. Koster en W. H. Maltha[1])

Van het arsenaal van methoden en technieken welke in het consumenten-marktonderzoek worden gehanteerd, is de methode van de produkttest er een die bij uitstek geschikt is voor het vaststellen van de relatieve appreciatie van experimentele of reeds in de markt bestaande produkten. Het is goed om t.a.v. het begrip produkt even vast te stellen dat in marktonderzoektechnische zin daarbij niet alleen gedacht wordt aan een samenstel van fysische of chemische bestanddelen. Ook attributen als verpakking, merknaam, prijs e.d. geven aan een produkt een gevoelsmatige of reële inhoud mee, die mede bepalend zijn voor de appreciatie van het produkt. Zo worden 2 pakken Skip waarvan er één is uitgevoerd met een speciaal openingsdevice als variabele evenzeer als 2 produkten beschouwd als bijv. Omo en Sunil. In beide gevallen kan het van belang zijn middels een produkttest de relatieve appreciatie te doen vaststellen. Het verschil in appreciatie is a.h.w. het eindprodukt van een produkttest. Afhankelijk van de testopzet en de in deze test ingebrachte variabelen zal het mogelijk zijn de invloed van deze variabelen op de appreciatie te bepalen.

Doelstellingen bij een produkttest

De produkttest is, zoals reeds werd opgemerkt, een middel om de relatieve appreciatie vast te stellen van 2 of meer produkten. De objektieven welke men ermee beoogt te bereiken kunnen echter zeer uiteenlopen:

a. produktontwikkeling: naast andere methoden kan de produkttest dienen als instrument bij het zoeken naar verbetering van het produkt of naar de richting waarin het produkt verder moet worden ontwikkeld. Het accent zal dan vooral komen te liggen bij de open vraagstelling zoals die betreffende redenen van preferentie, suggesties tot verbetering, oordelen op zich e.d.

b. plaatsbepaling: het doel is hier om vast te stellen hoe het eigen merk zich qua appreciatie verhoudt tot die van zijn belangrijkste concurrenten. Hierbij zijn 2 testsituaties denkbaar, één waarbij de merknaam meespeelt als variabele (as marketed testsituatie) en één waarbij de plaatsbepaling geschied zuiver op grond van produkthoedanigheden, onbeïnvloed door factoren als merk en reclame-inspanning (blinde testsituatie). Het doel kan ook zijn om middels „ranking" een keuze te doen uit een reeks proefsamenstellingen alvorens het aldus geselecteerde produkt tegen de standaardsamenstelling of tegen de concurrent te testen.

c. produktwijziging: principieel niet anders dan de doelstelling beschreven

[1]) Beide auteurs zijn verbonden aan het Instituut voor Huishoudelijk Onderzoek te Rotterdam.

onder a). Daar er reeds een gevestigd produkt is met een eigen plaats op de markt zal het onderzoek gericht zijn op het vaststellen van de invloed op de appreciatie bij wijziging van bij voorkeur één variabele bijv. de verpakkingsvorm, de kleur, het vetgehalte, de parfumconcentratie.

d. vaststelling drempelwaarde: hier is het doel vast te stellen op welk niveau een variabele in een produkt nog net wordt onderkend. Een variat hierop is het bepalen van het niveau waarboven een procentuele toename van de variabele (bijv. parfumconcentratie) niet meer wordt ervaren.

e. factorieel onderzoek: terwijl onder c) in het algemeen slechts beoogd wordt de invloed vast te stellen op de appreciatie van een wijziging van één variabele (bijv. 2 tandpasta's waarvan uitsluitend het krijtgehalte verschilt), is de factoriële opzet er op ingericht om van meerdere variabelen, meestal factoren genoemd, tegelijkertijd te onderzoeken in hoeverre deze elk op zich alsook in onderlinge wisselwerking de „overall" preferentie beïnvloeden.

Terwijl de doelstellingen van een produkttest dus uiteen kunnen lopen, is de uitvoering ervan steeds dezelfde. Het zal vrijwel altijd neerkomen op het laten keuren van 2 produkten, meestal in blanco verpakking, door een aantal proefpersonen. Zijn er meer dan 2 produkten waarvan de relatieve appreciatie moet worden bepaald dan zal afhankelijk van de doelstelling gekozen moeten worden uit een testopzet waarbij elk der produkten paarsgewijs tegenover de standaard wordt geplaatst (bijv. A vs. C en B vs. C) en één waarbij elk der produkten paarsgewijs tegen alle andere wordt getest in een zgn. complete tournament volgens het schema A vs. B, A vs. C, B vs. C. (Wij later hier gemakshalve buiten beschouwing de testopzetten, waarbij meer dan twee produkten tegelijk door dezelfde huisvrouw worden getest). De respondente zal antwoord moeten geven op de vraag aan welk van de 2 ter keuring toegezonden of voorgelegde produkten zij de voorkeur geeft (event. verfijnd op te geven in gradaties van voorkeur). De vragenlijst zal al naar gelang de doelstelling vragen omvatten naar de algemene preferentie, de reden van deze voorkeur, beoordeling op zich van elk der produkten, suggesties tot verbetering en vragen naar de voorkeur t.a.v. speciale karakteristieken van het produkt. Vragen ter verkrijging van voor het onderzoek relevante achtergrondgegevens completeren de vragenlijst.

De factoriële opzet

Zo gemakkelijk als het is om huisvrouwen twee produkten toe te sturen en van haar een preferentie, omkleed met redenen, te verkrijgen, zo moeilijk is het vaak om de resultaten te interpreteren. Verschillen de produkten nl. op meer dan één produktcomponent, dan zijn deze verschillen alle verdiskonteerd in het eindoordeel van de ondervraagde. Aangezien het verbaliseringsvermogen van de huisvrouw beperkt is, vlucht zij vaak in zeer algemene formuleringen en maakt zij geen onderscheid tussen de verschillende eigenschappen van het produkt. Trachten we dit te ondervangen door de huisvrouw te vragen om uitsluitend een oordeel te geven op een enkel aspekt van het produkt dan levert

ook dit niet het gewenste resultaat, daar in de praktijk blijkt dat een eenmaal opgegeven voorkeur t.a.v. een bijzondere eigenschap een zodanige bewustwording van deze eigenschap teweegbrengt dat de algemene voorkeur er door wordt vertrokken, een fenomeen dat bekend is als het halo-effekt. Bovendien kan van een gewijzigde samenstelling - een niveauverandering dus van een variabele - de suggestie uitgaan als zou het produkt op een geheel ander aspekt een wijziging hebben ondergaan. Zo kan verandering van kleur van een voedingsmiddel leiden tot een gewijzigde smaakappreciatie.

Het is duidelijk dat het voor de researcher vaak ondoenlijk is om aan de hand van de onontwarbare preferentiebeschrijvingen aan te geven op welk punt het produkt nu verbeterd moet worden, in welke richting de produktontwikkeling moet worden voortgezet. In principe is het denkbaar om een testprogramma op te zetten, waarbij steeds twee produkten tegen elkaar wordt getest, die slechts op één punt verschillen. Als we echter met een groot aantal componenten te maken hebben, zou dit tot een zeer groot aantal tests leiden, terwijl bovendien de invloeden op de preferentie, uitgaande van de *wisselwerking* van verschillende componenten - m.a.w. de interakties tussen de te onderzoeken factoren - geheel niet aan de orde zouden komen. De factoriële opzet - een testprogramma van een serie proefprodukten systematisch opgebouwd volgens een bepaald schema - maakt het mogelijk de invloed op de preferentie vast te stellen van de verschillende componenten (faktoren) in hun onderlinge samenhang zonder dat de ondervraagde een voorkeur behoeft te omschrijven. Aan de hand van een (fictief) cijfervoorbeeld kan het best geïllustreerd worden op welke wijze een factoriële opzet kan bijdragen tot een snelle beantwoording van vragen welke tijdens de ontwikkeling van een produkt kunnen opkomen.

In ons voorbeeld waar de faktoren aroma, zoutgehalte en kleur van een margarinesamenstelling moeten worden onderzocht, zouden de vragen als volgt kunnen luiden: wat is de invloed op de appreciatie wanneer het aroma wordt versterkt of het zoutgehalte wordt verhoogd of de kleur „geler" wordt gemaakt en zal de verhoging van het zoutgehalte bij een versterkt aroma hetzelfde effekt op de appreciatie hebben als op het lagere niveau van het aroma, m.a.w. bestaat er een interaktie tussen aroma en zout. Zo kan de vraag ook worden gesteld of er interakties bestaan tussen aroma en kleur of zout en kleur. Waar het hier dan gaat om het bepalen van de mate waarin de faktoren individueel en in hun samenhang de appreciatie beïnvloeden, is het duidelijk dat het dus wenselijk zou zijn om de te onderzoeken faktoren op meerdere niveaus te doen testen.

Een rem op dit streven wordt gevormd door het feit dat het aantal te produceren en te testen produktcomposities snel stijgt met het aantal niveaus, hetgeen praktische bezwaren met zich brengt zowel produkttechnisch als marktonderzoektechnisch. Laten we dus ons onderzoek eenvoudig houden en de faktoren aroma, zoutgehalte en kleur op 2 niveaus onderzoeken. Het factoriële schema van de te produceren en te testen proefsamenstellingen ziet er dan als volgt uit:

Aroma	a1				a2			
Zoutgehalte	z1		z2		z1		z2	
Kleur	k1	k2	k1	k2	k1	k2	k1	k2
Code	(1)	k	z	zk	a	ak	az	azk

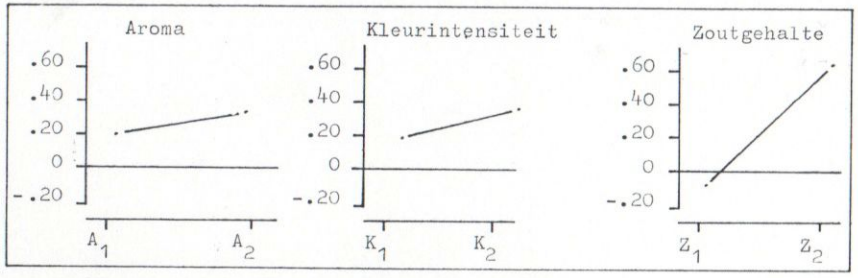
Bij een 2^3 schema (3 faktoren op 2 niveaus) is dus het aantal te testen proefsamenstellingen, composities, = 8. Om reden van praktische aard wordt aan elke compositie een code toegekend waarin tot uitdrukking wordt gebracht of een faktor op laag dan wel op hoog niveau voorkomt, door alleen die faktor te benoemen welke op hoog niveau voorkomt (een moeilijkheid vormt de compositie, waarin alle 3 factoren op laag niveau aanwezig zijn, zodat ter onderscheiding de verkorte code „(1") wordt gebezigd). We kunnen nu alle produktcomposities tegen één standaardsamenstelling testen, één welke wordt gekozen buiten de serie van de testcomposities, in welk geval we dus 8 tests krijgen; wordt één der 8 testcomposities als standaard genomen dan kunnen we volstaan met 7 tests. Een derde mogelijkheid is het testen in volledige competitie, waarbij alle composities tegen elkaar worden getest. We krijgen dan $\frac{n(n-1)}{2} = 28$ testcombinaties. Tenslotte is ook de monadische testopzet mogelijk waarbij elke testcompositie in isolatie wordt getest. Terwille van de eenvoud kiezen we het testschema waarbij elke proefsamenstelling tegen eenzelfde standaard (S) wordt getest door telkens een (ongeveer) gelijk aantal personen. Beperken we ons verder tot de uitkomsten van een „overall" preferentie gegeven in antwoord op de vraag „welke van de 2 produkten prefereert u of heeft u geen voorkeur" en de vraag naar de voorkeursintensiteit. De uitkomsten per test (1) vs. S, k vs. S, z vs. S etc. worden uitgedrukt in een „balance of preference".

Bijvoorbeeld (n per testpaar = 80):

Test-code	Beoordeling: Proefsamenstelling is				Beoordeling: Standaard is			Gewogen 3-2-1 schaal		Score		Balance of preference
	Zeer veel beter	Beslist beter	Alleen iets beter	Geen voorkeur	Alleen iets beter	Beslist beter	Zeer veel beter	Proefsamenstelling	Standaard	Proefsamenstelling	Standaard	Proefsamenstelling
(1)	1	9	16	24	20	9	1	37	41	0.46	0.51	—0.05
k	3	7	26	20	16	7	1	49	33	0.61	0.41	+0.20
z	3	8	25	22	15	6	1	50	30	0.62	0.37	+0,25
zk	1	14	28	17	13	7	0	59	27	0.74	0.34	+0.40
a	1	6	17	24	17	14	1	32	48	0.40	0.60	—0.20
ak	1	6	15	24	20	12	2	30	50	0.38	0.63	—0.25
az	5	13	31	20	10	1	0	72	12	0.90	0.15	+0.75
azk	5	22	28	22	3	0	0	87	3	1.09	0.04	+1.05

Hoofdeffekten

Het effekt van een faktor kan nu worden berekend door het verschil te nemen tussen de gemiddelde appreciaties (de balance of preference) van alle proefsamenstellingen met de betreffende faktor op niveau 1 en de gemiddelde appreciatie van alle proefsamenstellingen met deze faktor op niveau 2. We krijgen dus 3 „hoofd"-effekten welke hieronder grafisch zijn afgebeeld:



1 = laag, 2 = hoog niveau van de faktor.

De 3 hoofdeffekten worden berekend uit de balances of preferences van de proefsamenstellingen (1) t/m azk.

Voorbeeld berekening van effekt aroma:
¼(a+ak+az+azk)—¼(„1"+k+z+zk) = 0.3375—0.20 = +0.1375.

Op analoge wijze wordt berekend het effekt kleurintensiteit = 0.35—0.1875 = +0.1625 resp. het effekt zoutgehalte = 0.6125—(—0.0175) = +0.6875 (sign.).

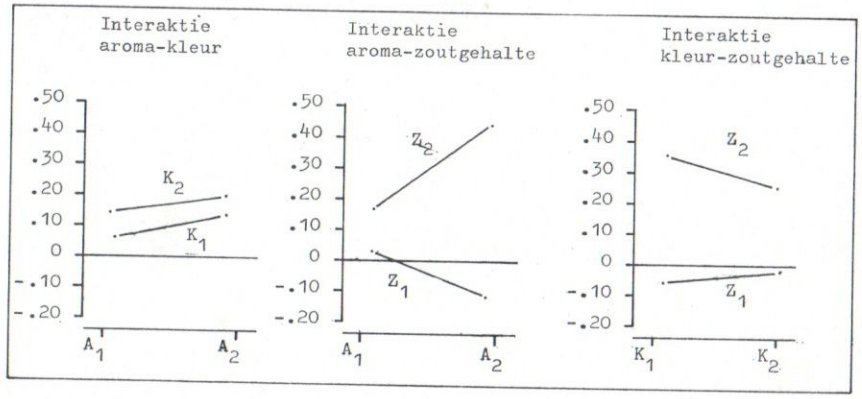
Voor de afleiding van de formule en de toetsing op significantie wordt verwezen naar de literatuur.

De conclusies welke hieruit na toetsing van de gevonden effekten getrokken kunnen worden zijn de volgende:

1. een effekt van verhoging van het aroma-gehalte is uit dit onderzoek niet aan te tonen.
2. er is een lichte indicatie dat de kleurintensiteit een gunstige invloed zou kunnen hebben op de appreciatie.
3. verhoging van het zoutgehale heeft een gunstige invloed op de appreciatie.

Eerste orde interakties

Voor het bepalen van een eventuele interactie is het nodig om vier punten vast te stellen zoals uit de volgende grafieken duidelijk wordt:



[1]) berekening interaktie aroma-kleur

$$\frac{(azk + ak)}{4} - \frac{(zk + k)}{4} - \frac{(az + a)}{4} - \frac{(z + 1)}{4} = 0.0375 \text{ (niet sign.)}$$

Analoge berekening van de interakties aroma-zoutgehalte resp. kleur-zoutgehalte levert op: +0.4375 (sign.) resp. +0.0625 (niet sign.). (Voor afleiding formule en toetsing op significantie zie literatuur).

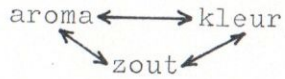
Conclusies:

1. Het aroma-effekt (A2—A1) is bij lage kleurintensiteit (K1) vrijwel gelijk aan dat bij hoge kleurintensiteit (K2), of anders gezegd het kleureffekt (K2—K1) bij laag aroma (A1) is ongeveer gelijk aan het kleureffekt bij hoog aroma (A2).
   Er is dus geen sprake van een interaktie tussen kleur en aroma of vice versa.
2. Verhoging van het aromagehalte heeft een gunstige invloed op het effekt van het zoutgehalte. Terwijl dus versterking van het aroma bij een laag zoutgehalte niets doet of mogelijk zelfs negatief werkt, blijkt versterking van het aroma bij een hoog zoutgehalte sterk positieve invloed te hebben op de appreciatie (dit verklaart waarom het hoofdeffekt van het aroma (zie blz. 7), wat berekend is voor een gemiddeld zoutgehalte per saldo positief is, zij het niet statistisch significant).
3. Invloed van de kleurintensiteit op het zouteffect is niet aantoonbaar en vice versa, m.a.w. geen interaktie tussen kleur en zoutgehalte.

Interakties van hogere orde

Tot dusverre hebben we ons alleen bezig gehouden met de hoofdeffekten en

de (1e orde) interakties tussen 2 faktoren. Daar wij 3 faktoren in ons voorbeeld hebben betrokken is ook nog de mogelijkheid aanwezig van een interaktie van hogere (2e) orde nl. die van aroma, kleur en zout in onderlinge samenhang:



Dit komt neer op het bepalen van het verschil in effekt van de interaktie tussen 2 faktoren bij hoog niveau van de 3e faktor en het effekt van de interaktie tussen dezelfde faktoren met de 3e faktor op laag niveau.

(azk — zk) — (ak — k)
— (zie hiervoor de opgegeven literatuur).
(az — z) — (a — 1)

Hier is slechts één interaktie te berekenen. In ons voorbeeld blijkt een tweede orde interaktie niet aantoonbaar.

Voordelen factoriële opzet

De merites van de factoriële opzet kunnen als volgt worden samengevat:

1. de factoriële opzet geeft maximum efficiency t.a.v. de bepaling der effekten.
2. alle resultaten worden gebruikt bij de bepaling van alle effekten.
3. een factoriële opzet is noodzakelijk als er interakties zijn (of verwacht mogen worden) om verkeerde conclusies te vermijden.
4. in de factoriële opzet wordt het effekt van elke faktor in één onderzoek gemeten bij verschillende niveaus der andere faktoren, waardoor de conclusies meer algemene geldigheid hebben.
5. voor de bepaling van de effekten en eventuele interakties is het voldoende om de algemene voorkeur te laten vaststellen.
6. de steekproefgrootte van een conventioneel onderzoek dat dezelfde resultaten zou moeten opleveren zou veel groter moeten zijn.

Literatuur:

O. L. Davies M.sc., Ph. D. (ed.) (Oliver & Boyd, Londen, 1956). The design and analysis of Industrial Experiments.
John Wiley & Sons Inc. N. York, 2nd ed. 1957. Experimental designs. William Cochran and Gertrude M. Cox.
Ned. Inst. v. Praeventieve Geneeskunde, Leiden, 1964. Inleiding tot de medische statistiek (deel 2). H. de Jonge.
London: Her Majesty's Stationery Office, repr. 1956, Industrial Experimentation. R. A. Brownlee M. A.
Penguin Books Inc. Baltimore. Repr. 1957, Facts from Figures, M. G. Moroney.

Het lezen waard

Selectief repertorium van de belangrijkste vaktijdschriften

JMR = Journal of Marketing Research
JAR = Journal of Advertising Research
COMM = Commentary
POQ = Public Opinion Quarterly
DMF = Der Marktforscher
MET = Metra
SG = Sociologische Gids
MM = Mens en Maatschappij
TPG = Tijdschrift voor Psychologie en haar grensgebieden
AD = Admap

MEDIA RESEARCH ISSUE. JAR: Vol. 8 no. 2, juni 1968.

Het juninummer van het Journal of Advertising Research is vooral gewijd aan media-onderzoek. Voor geïnteresseerden geven wij een kort overzicht van enkele van de artikelen die in dit nummer voorkomen en hun inhoud.

Schreiber R. J.: Probability Assignments for the Simulation of Media Reach and Frequency

Uit onderzoek trekt de schrijver de konklusie dat, wanneer het aantal uitgaven van een tijdschrift waarover wordt geobserveerd gering is, de proportie van de ingeziene bladen niet kan dienen als basis voor de waarschijnlijkheid dat de lezer enige uitgave heeft gezien. Als bovendien het aantal niveaus van waarschijnlijkheid gelijk is aan het aantal uitgaven (plus 1 voor degenen die geen uitgave zagen) ontstaan in sommige gevallen speciale moeilijkheden bij het toepassen van simulatietechnieken. De schrijver bespreekt mogelijke oplossingen van dit probleem, doch is niet optimistisch over de oplossing ervan.

Marc, M.: Combining Simulation and Panel Data to Obtain Reach and Frequency

Evenals het artikel van *Schreiber* heeft dit artikel betrekking op de simulatietechnieken bij het media-onderzoek (speciaal bij mediaplanning). Schrijver betoogt, dat het gebruik van leespatronen afkomstig uit de lezerspanels in belangrijke mate kunnen bijdragen tot betrouwbare simulatietechnieken. Het betoog is geïllustreerd met voorbeelden.

Ehrenberg A. S. C.: The Time and the Place for Readership Panels

In dit artikel berijdt Ehrenberg nogmaals zijn stokpaard: het gebruik van marktonderzoekgegevens bij het opbouwen van theorieën en van een betrouwbare methodologie. Lezerspanels zouden in beide gevallen een belangrijke rol kunnen spelen, belangrijker op het moment dan bij direkte toepassing op het

uit lezers-panels zijn overigens volgens de auteur nog geenszins overwonnen. terrein van media-planning. De moeilijkheden bij het gebruik van gegevens

Harvey B.: Nonresponse in T.V. Meter Panels.

Een exploratieve studie toonde aan dat non-response (weigering om deel te nemen) de resultaten van T.V.-meter onderzoek kan beïnvloeden. Aangezien wij in Nederland te maken hebben met dergelijk onderzoek (Intometrie) is dit artikel ook voor Nederlandse onderzoekers van belang. De schrijver wijst er zelf terecht op, dat zijn studie een indicatief karakter draagt: onderzoekmethode en steekproef betekenen beperkingen voor de validiteit van de uitkomsten. Aangezien echter die uitkomsten voldoende ruimte voor twijfel over laten, wordt verder - en onpartijdig - onderzoek krachtig aanbevolen.
Waarbij we ons gaarne aansluiten.

Claycamp H. J. en McClelland C. W.: Estimating Reach and the Magic of K.

In 1961 publiceerde Agostini zijn inmiddels beroemd geworden formule omtrent het netto bereik van honderden media combinaties in Frankrijk:

$$C/A = \frac{1}{1 + K.D/A}$$

waarin C = het netto bereik van een set van m media
A = de som van het bereik van ieder medium binnen de set
D = de som van het gedupliceerd bereik voor alle paren media in de set
K = 1.125.

Dit artikel belicht nader de aard van de „konstante" K. (een analytische interpretatie). Speciaal aanbevolen aan de mathematische fijnproevers onder de media-onderzoekers.

De konklusies van de schrijvers zijn:

a. K varieert slechts in geringe mate (limited range).
b. K is relatief ongevoelig voor de aard van de media, speciaal als het gaat om veel uitgaven en/of media in de kombinatie.
c. De variantie van K vermindert sterk als D/A en het aantal media in de kombinatie toenemen.
d. Schattingen van het bereik zijn in belangrijke mate ongevoelig voor bepaalde fouten in K.

De konklusie luidt, dat indien K op de juiste wijze geschat wordt voor een bepaalde set gegevens, fouten veroorzaakt door de behandeling als konstante meestal verwaarloosd kunnen worden. In het algemeen mag K niet gezien worden als universele konstante, zo merken de schrijvers op, doch wel heeft de Agostini-formule een stevige empirische basis die deze geschikt maakt voor gebruik in de praktijk. De methode der kleinste kwadraten wordt aanbevolen om K te schatten voor een bepaald empirisch materiaal. P. H. v. W.

Brody Robert P. en Scott M. C.: Personality Variables and the Consumer Decision Progress. JMR februari 1968.

De beginperiode van het psychologisch marktonderzoek (de vijftiger jaren) kende een zonnig optimisme omtrent de mogelijkheden om verbanden te ontdekken tussen de keuzeprocessen bij de konsument en de variabelen der persoonlijkheid. Dit optimisme was vooral gebaseerd op de „glittering generalities" die met zoveel succes op de onderzoekmarkt gebracht werden door de esoterische Dichterianen. Bij nadere beschouwing bleken de door deze school gerapporteerde resultaten nimmer afkomstig uit wat men in meer wetenschappelijke kringen zou beschouwen als deugdelijk onderzoek. Empirisch onderzoek met in de praktijk der psychologie beproefde methoden zette weldra een domper op het aanvankelijk enthousiasme: zoals de schrijvers van het hier besproken artikel dan ook zeggen „However the attempt to document these relationship have not been notably succesful and there is a growing feeling that the study of personality is likely to provide little insight beyond that by standard demographic data".

Dezelfde ervaringen hebben wij in Nederland opgedaan: tussen het konsumentengedrag aan de ene kant en de echte persoonlijkheidsvariabelen aan de andere kant bestaan meestal korrelaties waar geen enkele marketingman van in vervoering zou raken.

Schrijvers van het artikel waarschuwen echter tegen de neiging het kind met het badwater weg te gooien: zij menen dat het gebrek aan geconstateerde verbanden wel eens samen zou kunnen hangen met een onvolledig of onjuist theoretisch uitgangspunt op dit terrein en pogen deze konklusie te documenteren met een eigen onderzoek. Zij stellen o.m.:

a. er zijn nog betrekkelijk weinig studies waarin personality en koopgedraggegevens worden gebruikt.
b. de gebruikte persoonlijkheidsmaatstaven zouden wel eens ongeschikt kunnen zijn omdat zij ontleend zijn aan het algemene sociale onderzoek en niet samenhangen met een theoretisch model van het consumptief gedrag.

De schrijvers geven in het kort de resultaten weer van verschillende onderzoeken waarbij gewerkt werd met persoonsvariabelen. In het algemeen bleken deze variabelen slechts een onbetekend percentage der variantie te verklaren. Speciale aandacht wordt besteed aan *Evans* bekende Chevrolet-Ford studie die in brede kring wordt beschouwd als de doodsteek voor het personality onderzoek.

Theoretisch uitgangspunt bij de auteurs

De auteurs gaan er van uit, dat het keuzeproces bij de konsument een bijzondere mate van complexiteit vertoont. Een eerste eis, zo stellen zij, is groepering der variabelen.

Zij menen dat er sprake is van vier verschillende groepen:
a. variabelen van het persoons-systeem (bewuste en onbewuste behoeften b.v.)
b. variabelen van het sociale systeem (lidmaatschap van groepen b.v.)
c. exogene variabelen (zoals b.v. relatieve prijzen en aankoopgemak)
d. risico-reducerende variabelen (zoals b.v. het vertrouwen in winkels en merken).

Voorts menen de auteurs dat het keuzeproces bij de konsument vereenvoudigd bestudeerd kan worden als de invloed van elk systeem van variabelen beschouwd wordt als een funktie van de perceptie van de persoon van de keuze situatie. Drie percepties schijnen hierbij van bijzonder gewicht te zijn:
a. het waargenomen „performance risk" bij de keuze - in hoeverre menen de konsumenten dat verschillende merken op *relevante* wijze verschillen in prestatie?
b. specifiek zelfvertrouwen: hoe zeker is de persoon dat een bepaald merk aan zijn verwachtingen zal voldoen?
c. het waargenomen sociale risiko - in hoeverre meent de persoon dat anderen hem/haar zullen beoordelen op grond van zijn/haar keuze?

De auteurs stellen met nadruk, dat het hier gaat om een gesimplificeerd model van „percept al filters". De theorie leidt tot een aardig model, dat op pag. 52 van het artikel is weergegeven. Uit dit model kunnen de relevante variabelen in een bepaalde situatie worden afgeleid (relevant = de moeite waard om te meten). Het model geeft aanleiding tot een aantal predicties zoals: waar weinig performance-risk of social-risk wordt waargenomen door de konsument zal de groep der exogene variabelen van meer belang zijn. Of bijvoorbeeld: waar de vrouw de kleding ziet als een hoog social risiko doch het „performance risk" als gering (kwaliteit is niet van belang) zullen vooral de social-system variabelen (het oordeel van anderen) een belangrijke rol spelen. Of: voor mensen die de keuze van een hoofdpijnmiddel als sociaal onbelangrijk zien en de prestatie van merken als voorspelbaar en belangrijk zullen persoonsvariabelen een belangrijke rol spelen.

Toepassing

De auteurs hebben het model toegepast bij de analyse van bestaande gegevens omtrent het aankoopgedrag op het terrein van koffie (gegevens afkomstig uit het JWT panel).
Er werd in de eerste plaats gepoogd persoonlijkheids- en demografische variabelen te relateren aan merktrouwgegevens. (45 variabelen). De selektie van de kombinatie van de 10 beste variabelen voor predictie bleek slechts 3% van de variatie in de onafhankelijke variabele te verklaren, waarbij bovendien meer dan de helft van de predictie voor rekening kwam van de demografische variabelen. Dit is het gebruikelijke magere resultaat bij dergelijke analyses. Daarna hebben de auteurs eerst hun model toegepast op de gegevens. Dit model suggereert, dat personality variabelen slechts van belang zijn onder kondities van hoog waargenomen „performance-risk" en hoog specifiek zelf-

vertrouwen. Als onder deze kondities bovendien het waargenomen sociale risiko gering is, zouden persoonlijkheidsvariabelen moeten discrimineren tussen mensen die verschillende merken kiezen van hetzelfde produkt. Het zou ons te ver voeren de resultaten van de uitvoerige analyse der gegevens op basis van het model hier weer te geven. De deel-analyses die gemaakt werden op basis van de verschillende veronderstellingen die in het model zijn neergelegd leverden in het algemeen een veel groter percentage aan „verklaring" op dan eerder werd gevonden bij de „personality studies". Zo bleek dat bij analyse van uitsluitend de groep der 100% merktrouwen de predictieve waarde van een set variabelen op te lopen tot 48%. Aacht persoonlijkheidsvariabelen (op basis van testscores) leverden gezamenlijk een predictie van 32% der variantie bij deze groep.
De konklusie van de auteurs luidt dan ook, dat voorafgaande teleurstellende resultaten bij deze analyses zeer goed het gevolg kunnen zijn van de manier waarop de gegevens werden geanalyseerd - i.c. het ontbreken van een adequaat theoretisch model. In de woorden van de auteurs: „For explaining relative loyality to a family's favorite brand personality variables were found to have negliglible practical value. However when trying to discriminate the brand choice of people likely to have perceived-high performance risk and to have high specific self confidence, personality variables were very useful.
De weg die de auteurs hier wijzen komt neer op het toepassen van een klassifikatie- en filtersysteem op de variabelen voordat men begint te analyseren. Een zwak punt bij de studie is, dat de auteurs geen onafhankelijke maatstaven ter beschikking hadden omtrent „perceived performance risk", „percieved social risk" en „specific self-confidence", de filters die werden toegepast. Hiervoor werd gebruik gemaakt van veronderstellingen op het waargenomen „merkgedrag". Hoewel er ongetwijfeld veel meer research nodig is op diverse punten, menen wij dat de auteurs een belangrijke weg hebben gewezen tot het verkrijgen van betere resultaten bij het psychologisch marktonderzoek. Bestudering van hun resultaten leidt tot de konklusie dat het bij de aanvang van deze recencie gesignaleerde pessimisme tot op zekere hoogte ongefundeerd is. Personality gegevens kúnnen van belang zijn als . . . P. H. v. W.

Parfitt J. H. en Collins B. J. K.: Use of Consumer Panels for Brand Share Prediction JMR mei 1968.

De belangstelling voor consumenten-panels is in de laatste jaren aanmerkelijk toegenomen. De Attwood-groep, waarvan Parfitt en Collins deel uit maken speelt op dit terrein een belangrijke rol. Wij herinneren aan de ESOMAR-voordracht 1967 van Parfitt, waarbij deze met cijfers illustreerde, dat een Consumenten-panel in de regel een betrouwbaarder hulpmiddel vormt voor de schatting van marktaandelen, omzetten en dergelijke dan een ad hoc enquête. (Zie dit tijdschrift, entreenummer dec. 1967).

In dit artikel wordt nu het gebruik beschreven van panelgegevens bij het maken van schattingen van toekomstige marktaandelen na lancering van nieuwe produkten. In het artikel wordt de gevolgde methode uitvoerig beschreven en geïllustreerd met voorbeelden. Bij berekening wordt gebruik gemaakt van verzamelde gegevens omtrent „nieuwe koopsters" in een aantal perioden en van gegevens omtrent herhalingsaankopen in deze perioden. De methode is qua principe afkomstig van Baum en Dennis (Esomar 1961).
Bij de berekeningen (die in principe niet bijzonder gekompliceerd zijn), en de prediktie moet rekening gehouden worden met een tweetal aannamen: een uniforme hoge distributiegraad in het te bestuderen gebied en gelijkblijvende kondities op de markt. Vooral deze laatste konditie zal in de praktijk nog wel moeilijkheden opleveren tijdens een introduktieperiode van nieuwe produkten of merken. Vaak immers kan in een dergelijke periode op verhoogde aktiviteit van de konkurrentie worden gerekend.

De schrijvers besteden aandacht aan de validiteit van de methode:
a. Zijn de voorspellingen korrekt binnen zekere grenzen?
b. Zijn voorspellingen mogelijk *voordat* ook uit andere gegevens af te leiden valt, wat er op een gegeven tijdstip T op de markt aan de hand zal zijn?
De schrijvers laten zien, dat in 24 gevallen een suksesvolle voorspelling werd geleverd van het aantal dat een merk ging behalen 12 tot 18 maanden na introduktie. Daarbij dient men te bedenken, dat het hierbij slechts gaat om suksesvolle merken: onsuksesvolle merken worden dikwijls eerder weer uit de markt genomen zodat kontrole der voorspelling niet mogelijk is. Verzuimd is te vermelden, welk deel de 24 voorspellingen uitmaakten van het totaal der kontroleerbare voorspellingen zodat niet kan worden nagegaan, of er sprake is geweest van een selektie van „overtuigende gevallen".
In het algemeen is hier sprake van een belangrijk en voortreffelijk gedokumenteerd artikel waarin de gevolgde rekenmethode nog eens in extenso in een naschrift wordt beschreven.

Enkele uitwijdingen dragen belangrijk bij aan de waarde van het artikel: beschreven worden o.m. de prediktie in een zeer vroeg stadium, prediktie bij introduktie van een nieuw produkt, het bepalen van aandeels-voorspellingen in een veranderende markt en het gebruik van de gegevens bij het opbouwen van een theorie van algemene aard (vaststellen van algemene relaties). Ook ten aanzien van de werking van marketing-factoren (de invloed van speciale aanbiedingen etc.) kan met behulp van de methode het één en ander vastgesteld worden. Tenslotte zo stellen de schrijvers, kunnen de resultaten van de analyse nog gekombineerd worden met de „psychologische" gegevens die omtrent de Attwood-panels verzameld worden. Het zal duidelijk zijn, dat het voor de ondernemer van groot belang is reeds vroegtijdig de beschikking te hebben omtrent gegevens die een betrouwbare voorspelling vormen van het sukses van nieuwe merken of produkten na een zekere periode. Het werk van Parfitt en Collins geeft ons voldoende vertrouwen, dat een dergelijke voorspelling in vele gevallen tot de mogelijkheden behoort. Bovendien brengt dit werk ons dichter bij een algemene theorie op dit terrein waaraan dringend behoefte bestaat. Bestudering van dit artikel zij dan ook aanbevolen zowel aan marketingmensen als aan onderzoekers. P. H. v. W.

inhoud van dit nummer:

Artikelen

J. Stapel:
„Kwaliteit beoordelen? Geef het maar een cijfer" . . . 69

P. H. van Westendorp:
„Kwaliteit beoordelen? Geef het maar een cijfer"
. . . als je durft? 71

J. Stapel:
„Kwaliteit beoordelen. Noem desnoods een prijs" . . . 76

Drs. E. N. Koster en W. H. Maltha:
„Factoriële opzet bij produkttest" 79

Het lezen waard 86