

Het selecteren van adressen voor direct mail, maximaliseren van de verwachte winst

Dr. J.R. Bult

1 Inleiding

De belangstelling voor direct marketing is de afgelopen jaren sterk toegenomen. Dit wordt mede veroorzaakt door het feit dat mensen steeds meer hun eigen wereld creëren, waarbij zij zich minder laten leiden door wat anderen vinden of denken. Dit betekent dat afnemers op hun eigen specifieke manier en met op hun eigen smaak afgestemde producten moeten worden benaderd (zie Bult en Hoekstra 1991, en Hoekstra en Raaijmakers 1991). Dit heeft een aantal belangrijke consequenties voor het marketing management in termen van het definiëren van doelgroepen en segmenten en in combinatie hiermee het vinden van specifieke stimuli-responsrelaties. Eén van de manieren om hierop in te springen is een individuele benadering (direct marketing).

Een tweede reden voor de toegenomen belangstelling is dat de mogelijkheden van informatieverwerking de laatste jaren enorm zijn toegenomen. Bovendien zijn de kosten van het verwerken van informatie sterk gedaald. Dit betekent dat het opslaan, bewaren, analyseren en 'updaten' van individuele informatie kosten effectief kan zijn.

Ook vanuit de wetenschap is de belangstelling voor direct marketing toegenomen. Dit uit zich onder andere in een toename van het aantal aangeboden dm-cursussen aan HBO-instellingen en universiteiten en in een toename van het aantal wetenschappelijke onderzoeken op het gebied van direct marketing (zie onder andere Bult 1993a en Wedel et al. 1993). Direct marketing leent zich bij uitstek voor kwantificering van problemen.

Direct mail is één van de belangrijkste mediatypen van direct marketing. Het succes van een direct mail campagne hangt in het algemeen af

van: 1) het aanbod, 2) de elementen van communicatie (art en copy), 3) het tijdstip van communicatie en 4) het adressenbestand van potentiële kopers waarop de communicatie gericht is. In tabel 1 wordt een subjectieve verzameling van relatieve belangen van de vier factoren gegeven die men in de direct marketing literatuur tegenkomt (Roberts en Berger 1989). Deze zijn niet gebaseerd op empirisch onderzoek. Toch lijkt er binnen de direct marketing literatuur geen twijfel te bestaan over de volgorde van de factoren in het behalen van succes. De belangrijkste succesbepalende factor van direct mail is dus het selecteren van kansrijke adressen.

Tabel 1: Succesbepalende factoren voor direct mail

<i>Succesbepalende factor</i>	<i>Relatieve belang</i>
het adressenbestand van prospects	50%
het aanbod	20%
het tijdstip van communicatie	20%
de elementen van communicatie	10%

Het selecteren van doelgroepen vindt plaats in diverse geledingen van de bedrijfskolom. De aanbieders die van direct marketing gebruik maken variëren van fabrikanten (bijv. Yves

Dr. J.R. Bult (1966) studeerde econometrie aan de Rijksuniversiteit Groningen. Hij promoveerde in 1993 op het proefschrift *Target selection for direct marketing*. Hij publiceerde o.a. in *Journal of Marketing Research*, *Marketing Science*, *Journal of Applied Econometrics* en *International Journal of Research in Marketing*. Momenteel is hij als post-doc onderzoeker werkzaam bij de vakgroep Gezondheidswetenschappen (RUG) en houdt zich bezig met kosten-versus-effectiviteit studies binnen de gezondheidszorg.

Rocher) tot detaillisten. Zo kan een detaillist bijvoorbeeld gebruik maken van directe communicatiemiddelen (geadresseerde post, telefoon), en kan hij kiezen voor directe levering aan huis. Een voorbeeld van dit laatste is het Ahold-project James Telesuper. Een belangrijk verschil tussen doelgroepselectie in consumenten marketing en industriële marketing is dat er op industriële markten sprake is van een beperkt aantal potentiële kopers, terwijl het op consumentenmarkten vaak gaat om grote adressenbestanden. Statistische selectiemethoden spelen daarom in de consumenten marketing een veel belangrijker rol.

Het doel van dit artikel is onder andere inzicht te verschaffen in het gebruik van bedrijfseconomische begrippen en statistische technieken bij het selecteren van kansrijke adressen voor een direct mail campagne. Een voorbeeld hiervan is het selecteren van adressen op basis van de *profit maximization* benadering, waarbij marginale kosten worden vergeleken met verwachte marginale opbrengsten. In paragraaf 2 wordt het proces van het selecteren van adressen voor een direct mail campagne beschreven. De volgende paragraaf gaat nader in op de meest gebruikte selectietechnieken. Paragraaf 4 begint met kritiek op bestaande methoden en bevat een nieuwe selectiemethode, ontwikkeld op basis van maximaliseren van de verwachte winst (profit maximization benadering). Enkele toepassingen van deze benadering staan in paragraaf 5. Het artikel wordt afgesloten met enkele conclusies.

2 Selectieproces

Het selectieproces van een organisatie die gebruik maakt van direct mail bestaat uit vier stappen. In de eerste stap moet men kiezen voor één of meerdere beschikbare databestanden waaruit prospects geselecteerd kunnen worden. We maken onderscheid tussen interne en externe bestanden. Het interne bestand van een organisatie bevat gegevens van individuen (zowel klanten als potentiële klanten) met wie in het verleden contact is geweest. Externe bestanden worden door derden beheerd en kunnen worden gekocht of gehuurd. Voorbeelden van externe bestanden zijn de postcode informatiesystemen van Geo-Marktprofiel en Mosaïc en de individuele segmentatiesystemen zoals Omnidata en Responsplus.

Stap 2 bestaat uit het bepalen van kenmerken van individuen (selectievariabelen) uit de in stap 1 gekozen databestanden die van invloed zijn op het gedrag van potentiële kopers. Selectievariabelen worden onderverdeeld in vier categorieën, namelijk:

- 1 geografische variabelen (postcode, regio, urbanisatiegraad, etc.),
- 2 demografische variabelen (geslacht, leeftijd, inkomen, etc.),
- 3 levensstijlvariabelen (hobby's, leesgedrag, bezitskenmerken, etc.) en
- 4 gedragsvariabelen uit het verleden (tijd sinds laatste aankoop, kredietwaardigheid, etc.).

De belangrijkste selectievariabelen zijn de gedragsvariabelen uit het verleden. Deze worden ook wel RFM-variabelen genoemd (Recency, Frequency en Monetary value). Recency betreft variabelen die meten hoe lang het is geleden dat een individu iets bij de organisatie kocht. Frequency geeft aan hoe vaak een huishouden in een bepaalde tijdsbestek iets heeft gekocht. Monetary value geeft de waarde van de verrichte aankopen weer.

Nadat de direct marketeer in stap 1 de adressenbestanden en in stap 2 de selectievariabelen heeft bepaald moet hij in de derde stap van het selectieproces een statistische selectiemethode kiezen waarmee hij adressen voor de komende direct mail campagne selecteert. De laatste stap in het selectieproces bestaat uit het selecteren zelf.

3 Selectietechnieken

Selectietechnieken kunnen worden onderverdeeld in cluster-scoremodellen en individuele scoremodellen. Cluster-scoremodellen zijn modellen die aan een cluster van individuen een kans op respons toekennen. Individuele scoremodellen bepalen voor iedere individu afzonderlijk een kans op respons. De belangrijkste cluster-scoremodellen zijn 1) Kruistabelanalyse, 2) AID (Automatic Interaction Detection) en 3) CH-AID (CHi-kwadraat AID).

Kruistabelanalyse is de methode waarbij de populatie op basis van selectievariabelen in een aantal niet-overlappende clusters wordt ingedeeld. Per cluster wordt de verwachte kans op respons berekend. Bovendien wordt getoetst of deze clusterindeling significant bijdraagt in het beschrijven van de respons. Indien er sprake is van

meer dan drie selectievariabelen wordt de uitkomst onoverzichtelijk.

AID en CH-AID zijn gebaseerd op een stapsgewijze toepassing van variantie-analyse. Het doel is om de populatie in een aantal niet-overlapende clusters te verdelen zodanig dat de variantie in de te verklaren variabele (de respons) maximaal gereduceerd wordt. Deze technieken zijn in staat om interacties tussen de selectievariabelen te ontdekken. AID kan beschreven worden door een aantal stappen:

- 1 Voor iedere selectievariabele wordt de binaire splitsing bepaald die de variantie in de responsvariabele zoveel mogelijk reduceert.
- 2 Bepaal uit stap 1 de selectievariabele die voor de grootste reductie zorgt in de variantie van de responsvariabele en splits de populatie in twee delen op basis van deze selectievariabele.
- 3 De deelbestanden uit stap 2 worden vervolgens weer geanalyseerd zoals in stap 1 en stap 2 is weergegeven. Dit proces wordt gecontinueerd totdat er aan een stopcriterium is voldaan.

CH-AID doorloopt in principe dezelfde stappen maar komt tegemoet aan een aantal nadelen van AID. Zo beperkt CH-AID zich niet tot binaire splitsingen. De analyse bepaalt zelf de optimale splitsing van de selectievariabelen. Daarnaast splits CH-AID op de meest significante selectievariabele door gebruik te maken van een chi-kwadraat toets.

De belangrijkste individuele scoremodellen zijn (Judge et al. 1985): 1) het lineaire kansmodel, 2) discriminant analyse en 3) probit en logit modellen. Indien de te verklaren variabele slechts twee waarden aan kan nemen komt het lineaire kansmodel sterk overeen met discriminantanalyse (Maddala 1983). Het lineaire kansmodel beschrijft de kans dat individu i reageert op een mailing als lineaire functie van een k -tal selectievariabelen:

$$P_i = \alpha + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + u_i,$$

waarbij P_i is de kans op respons van individu i , X_j is de j -de selectievariabele van individu i , u_i is de storingsterm (deze vertegenwoordigt alle factoren die van invloed zijn op de kans op respons maar die niet worden gemeten), en α, β_j zijn de parameters die op basis van een testmailing geschat worden. P_i is een latente variabele. In de praktijk

wordt niet de kans dat individu i reageert op een mailing geobserveerd maar er wordt slechts geconstateerd of hij reageert of niet. Er wordt dus een dummy variabele Y_i^* geobserveerd die de waarde 1 aanneemt indien i reageert en de waarde 0 indien i niet reageert. Voor het schatten van de onbekende parameters wordt P_i vervangen door Y_i^* . Indien naast het gegeven of een individu al dan niet gereageerd heeft op een mailing ook het bedrag bekend is dat individu i heeft besteed, ontstaat er een andere regressievergelijking, het zogenaamde tobit model. De afhankelijke variabele is in dat geval het geldbedrag dat individu i heeft besteed. Er wordt in dit artikel niet verder ingegaan op deze vorm van regressie.

Het lineaire kansmodel heeft een aantal statistische nadelen (zie Judge et al. 1985), zoals het feit dat de variabele Y_i^* slechts twee waarden aanneemt. Hierdoor neemt de storingsterm ook slechts twee waarden aan. Dit betekent dat de gebruikelijke vooronderstellingen van normaliteit en homoskedasticiteit van de storingstermen niet opgaan, hetgeen leidt tot inefficiënte schatters van de parameters. Om hieraan tegemoet te komen zijn een aantal modellen ontwikkeld die rekening houden met het binaire karakter. De twee meest gebruikte modellen zijn het logit en het probit model. Het logit model beschrijft de relatie tussen de kans op respons en de selectievariabelen als

$$P_i = \frac{1}{e^{-\alpha - \beta_1 X_{1i} - \dots - \beta_k X_{ki}} + 1}$$

Het probit model is gespecificeerd als

$$P_i = \Phi (\alpha + \beta_1 X_{1i} + \dots + \beta_k X_{ki}).$$

Hierbij is $\Phi (\cdot)$ de cumulatieve standaard normale verdeling. Door de manier waarop het logit en het probit model zijn gespecificeerd, verdwijnt de storingsterm in de relatie tussen de kans op respons, P_i , en de verklarende variabelen, X_{1i}, \dots, X_{ki} .

Nadat één van bovenstaande modellen is geschat kan een gains chart analyse worden toegepast. De scores voor de kans op respons worden bepaald door de waarden van de selectievariabelen in het geschatte model in te vullen. Vervolgens worden de scores gerangordend van groot naar klein en gecategoriseerd in klassen van

gelijke omvang (meestal decielen). Per klasse wordt de gemiddelde respons berekend. Indien de gemiddelde respons wordt afgezet tegen de score van de individuen ontstaat de zogenaamde gains chart. Met behulp van een break-even analyse worden de potentiële respondenten ingedeeld in een kansrijk en een niet-kansrijk segment. Slechts het eerste segment wordt gemaïld. Gains chart analyse is een standaard module in direct marketing software, zoals Geo-DCS en DM-STATS van Geo-Marktprofiel en SMART van Omnidata. Toch kleven er aan gains chart analyse ook een aantal nadelen. Bult en Wansbeek (1995) hebben een selectieprocedure ontwikkeld die tegemoet komt aan deze nadelen.

4 Selecteren op basis van winstmaximalisatie

In de nieuwe methode, de *profit maximization benadering (PM-benadering)*, staat het principe van *winstmaximalisatie* centraal. Dit betekent dat deze methode gemiddeld hogere opbrengsten genereert omdat de verwachte winst van een mailing campagne wordt geoptimaliseerd. Een tweede verschil met gains chart analyse is dat deze laatste de gemiddelde kans op respons berekent voor slechts tien groepen (per deciel). Het deciel dat zich op het breekpunt bevindt vormt een dilemma voor de organisatie (moet dit deciel al dan niet gemaïld worden). Aangezien het selecteren van kansrijke adressen voor direct mail soms om meer dan één miljoen adressen gaat, betekent dit dat de beslissing ten aanzien van het breekpunt-deciel een belangrijke beslissing is, waarbij het gaat om grote investeringen. Een derde verschil met gains chart analyse betreft het feit dat gains chart analyse niet garandeert dat huishoudens met een hogere score voor de kans op respons ook daadwerkelijk een hogere voorspelde kans op respons hebben. Deze vreemde eigenschap wordt veroorzaakt omdat zowel het systematisch als het niet-systematisch gedrag van huishoudens wordt beschreven (en voorspeld). De profit maximization benadering houdt slechts rekening met dat deel van het gedrag van huishoudens dat verklaard kan worden, namelijk het systematisch gedrag.

De profit maximization benadering werkt als volgt. Stel het adressenbestand van een organisatie bevat N huishoudens en k selectievariabelen $X_1, \dots,$

X_k . Hierbij kan men denken aan wat men in het verleden heeft gekocht, hoe lang het is geleden dat men iets heeft gekocht en het betalingsgedrag uit het verleden. Andere voorbeelden van selectievariabelen zijn het aantal kinderen binnen een huishouden en het inkomen. Er wordt verondersteld dat individuen een intrinsieke neiging hebben om op een mailing te reageren. Een direct marketing organisatie kan echter deze neiging niet waarnemen. Hoe groter de neiging is om te reageren des te groter de kans dat ook daadwerkelijk gerespondeerd wordt. De intrinsieke neiging om te reageren wordt aangeduid met de variabele Y_i . Deze neiging kan gedeeltelijk verklaard worden door de selectievariabelen. De relatie tussen de neiging om te reageren, Y_i , en de selectievariabelen wordt als volgt weergegeven:

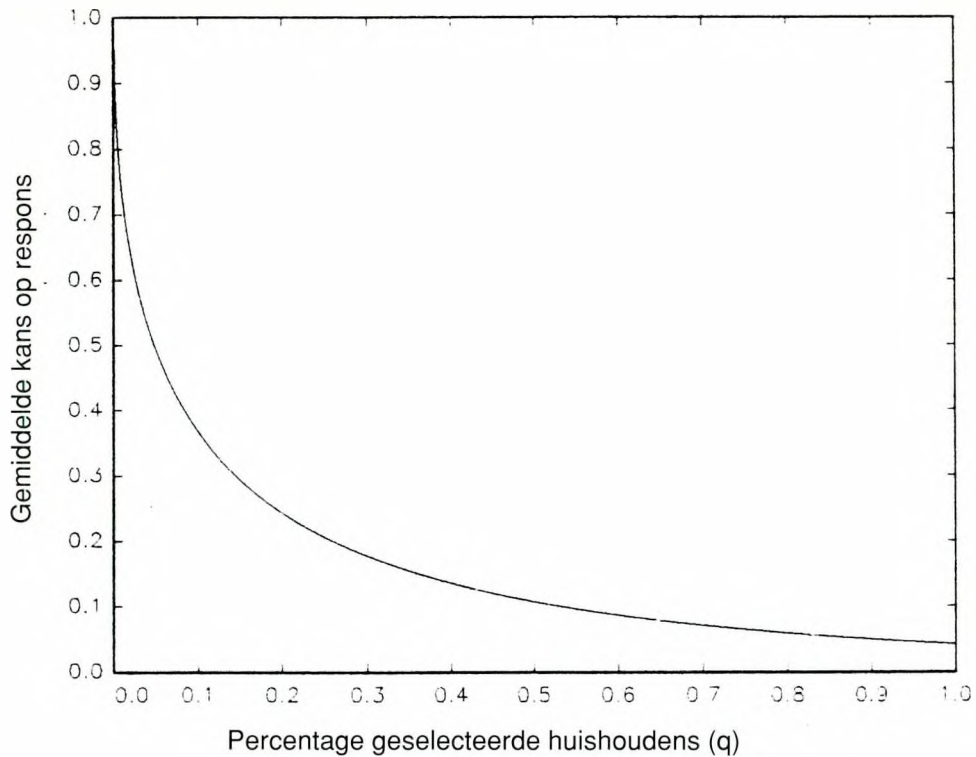
$$Y_i = \alpha + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + u_i.$$

De constante, α is een indicatie voor de gemiddelde neiging om te reageren. Daarnaast wordt de neiging verklaard door de k selectievariabelen. De invloed van de selectievariabelen wordt weergegeven door de parameters β_j ($j=1, \dots, k$). De laatste term in vergelijking (1) is de storingsterm u_i . Deze term representeert alle factoren die wel van invloed zijn op de neiging om te reageren maar die niet in het adressenbestand aanwezig zijn. De invloed van de storingsterm is vaak zeer groot. Bij het voorspellen van toekomstig responsgedrag kan slechts het systematisch deel worden gebruikt.

De selectiestrategie bestaat hieruit dat de groep van huishoudens gerangschikt wordt aan de hand van de geschatte neiging (het systematisch deel) om te reageren. Er wordt verondersteld dat individu i op een mailing reageert indien zijn of haar neiging om te reageren boven een bepaalde drempel ligt. Zonder verlies aan algemeenheid wordt deze drempel gelijk aan 0 gesteld. Stel dat Y_i^* de variabele is die weergeeft of een individu gerespondeerd heeft ($Y_i^* = 1$) of niet ($Y_i^* = 0$), dan is $Y_i^* = 1$ indien $Y_i > 0$. Nadat de waarde van de parameters op basis van een testmailing zijn geschat ($\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_k$) kan het systematisch deel van het gedrag worden bepaald. Dit wordt weergegeven met de indexvariabele n_i , dus:

$$n_i = \hat{\alpha} + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_k X_{ki}.$$

Figuur 1: Responscurve



Voor alle huishoudens wordt de waarde van n_i bepaald, door de waarden van de verklarende variabelen te vermenigvuldigen met de geschatte parameters. Vervolgens worden de huishoudens gerangordend op basis van de waarde van het systematisch deel van de neiging om te reageren.

Gegeven deze ordening is de vraag van het wel of niet mailen teruggebracht tot het bepalen van het breekpunt c . Alle huishoudens met een waarde van n_i groter dan c ontvangen een mailing; de overige huishoudens niet. Indien de waarde van c daalt zal een groter percentage van het adressenbestand gemaïld worden. Gegeven de relatie tussen c en het percentage adressen dat gemaïld wordt (weergegeven met q) kan de relatie bepaald worden tussen de gemiddelde kans op respons, gegeven dat een percentage q van 'beste prospects' gemaïld wordt. Deze relatie wordt de *responscurve* genoemd. In figuur 1 wordt een fictieve responscurve gegeven. De gemiddelde kans op respons neemt af indien een groter percentage huishoudens wordt gemaïld. Indien alle huishoudens een mailing ontvangen is de gemiddelde kans op respons gelijk aan 4,2 procent. Indien de beste 10 procent van de huishoudens een

mailing ontvangen, wordt een responspercentage van bijna 37 procent verkregen.

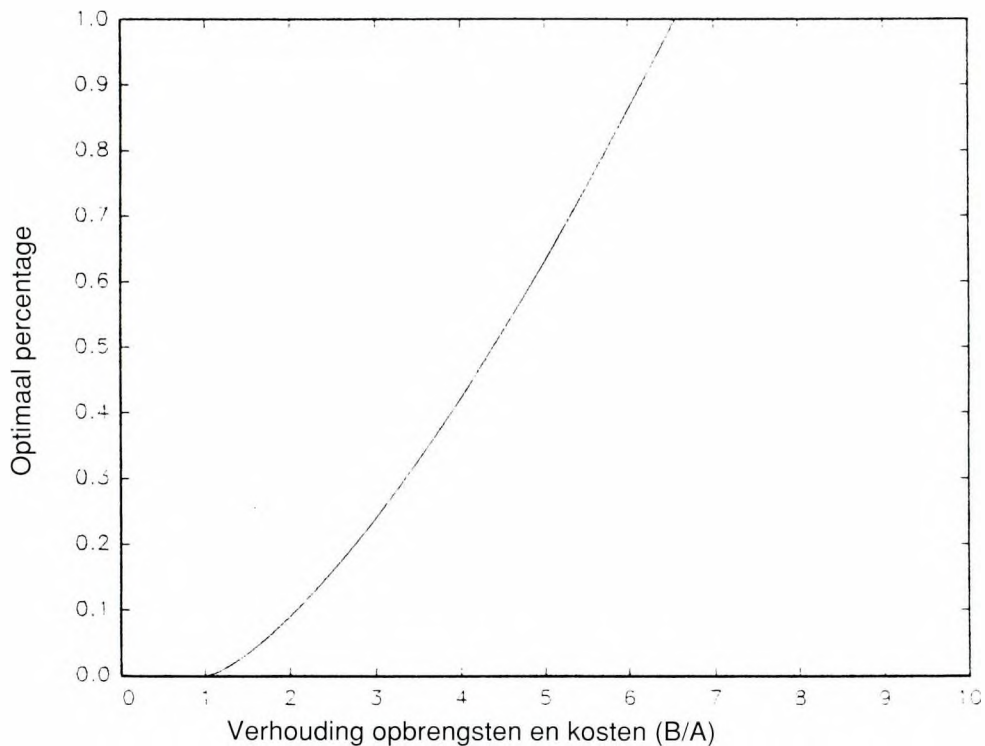
Gegeven de responscurve wordt het breekpunt zodanig bepaald dat de totale verwachte opbrengst van de direct mail campagne wordt gemaximaliseerd. Het breekpunt hangt af van twee factoren, namelijk: 1) van de kosten van de mailing en 2) van de opbrengst van een succesvolle mailing.

Stel dat de kosten van de mailing gelijk zijn aan A en de opbrengst van een succesvolle mailing gelijk is aan B . Het breekpunt wordt nu bepaald door middel van de volgende beslissingsregel: *huishouden i ontvangt een mailing indien de verwachte opbrengst van de mailing van huishouden i (zijnde de kans op respons, P_i , vermenigvuldigd met de opbrengst van een succesvolle mailing, B) groter is dan de kosten van de mailing, A , dus $BP_i > A$. Hieruit volgt dat het optimale afbreekpunt gelijk is aan:*

$$c = \frac{A}{B}.$$

Bij deze waarde van het breekpunt wordt de totale verwachte winst van de mailing gemaxima-

Figuur 2: Cutoff curve



liseerd. Gegeven de relatie tussen q en de gemiddelde kans op respons kan nu ook de relatie worden getekend tussen de verhouding van opbrengst en kosten van een mailing ($\frac{B}{A}$) en het optimale percentage huishoudens dat gemaïld dient te worden. Deze relatie wordt de *cutoff curve* genoemd. Een fictieve cutoff curve is weergegeven in figuur 2. De cutoff curve begint te stijgen vanaf $B/A = 1$. Beneden dit punt is de opbrengst van de mailing lager dan de kosten. Niemand ontvangt logischerwijs een mailing. Hoe groter de relatieve opbrengst des te groter het percentage huishoudens dat gemaïld wordt. In het geval dat B/A gelijk is aan 4 dient ongeveer 42 procent van de huishoudens een mailing te ontvangen.

5 Toepassingen van de profit maximization benadering

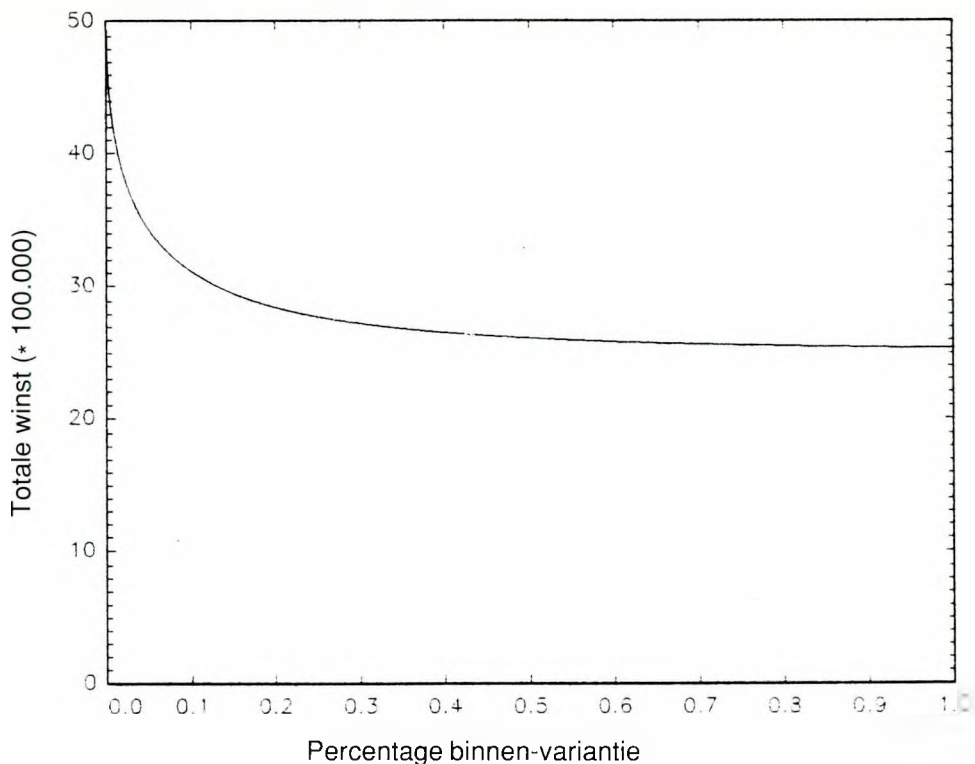
5.1 De waarde van postcode informatie

Direct marketing streeft naar het verzamelen van gegevens van individuen op een zo laag

mogelijk aggregatieniveau. Een categorie specifiek voor direct marketing ontwikkelde systemen zijn de postcode informatiesystemen (Mosaïc en Geo-Marktprofiel). Deze bestanden bevatten gegevens (zoals welstand en levensfase) van alle Nederlandse huishoudens op het aggregatieniveau van de postcode. Huishoudens in hetzelfde postcodegebied krijgen dezelfde waarden van selectiekenmerken toegekend. Nederland telt zo'n 6,4 miljoen huishoudens verdeeld over bijna 400.000 postcodes. Door een koppeling te leggen tussen postcodes en selectievariabelen is het mogelijk om een profielschets te maken van de meest-kansrijke postcodegebieden.

De gegevens uit de postcode informatiesystemen zijn geaggregeerd. Dit betekent dat op individueel niveau (huishoudens) meetfouten worden gemaakt ten aanzien van de selectievariabelen. De waarde van postcode informatiesystemen is afhankelijk van de grootte van de meetfouten. Ten einde een indicatie te geven van de variabiliteit van selectievariabelen binnen en tussen postcodegebieden is een steekproef getrokken van zo'n 40.000 huishoudens met informatie over één van de meest gebruikte (niet-RFM) selectievariabele: de autocataloguswaarde. Deze

Figuur 3: De waarde van postcode informatie



variabele wordt gebruikt als 'proxy' voor het inkomen en het bestedingspatroon van huishoudens. Ten aanzien van deze selectievariabele geldt dat de binnen-postcode variantie gelijk is aan 55 procent, de waarde van de tussen-postcode variantie gelijk is aan 45 procent. Deze getallen geven een indicatie van de mate waarin huishoudens binnen postcodegebieden (ten aanzien van de selectievariabele) op elkaar lijken.

De profit maximization benadering kan nu worden gebruikt om een indruk te krijgen van het te verwachten rendement van het gebruik van postcode-informatiesystemen. We starten op dezelfde wijze als in een profit maximization benadering zonder meetfouten. De selectievariabelen bestaan, althans voor een deel, uit postcode informatie. Gegeven de selectievariabelen wordt het optimale breekpunt bepaald zoals beschreven in de vorige paragraaf. De relatie tussen de verwachte winst, Π , en het percentage huishoudens dat gemaïld wordt, q , is

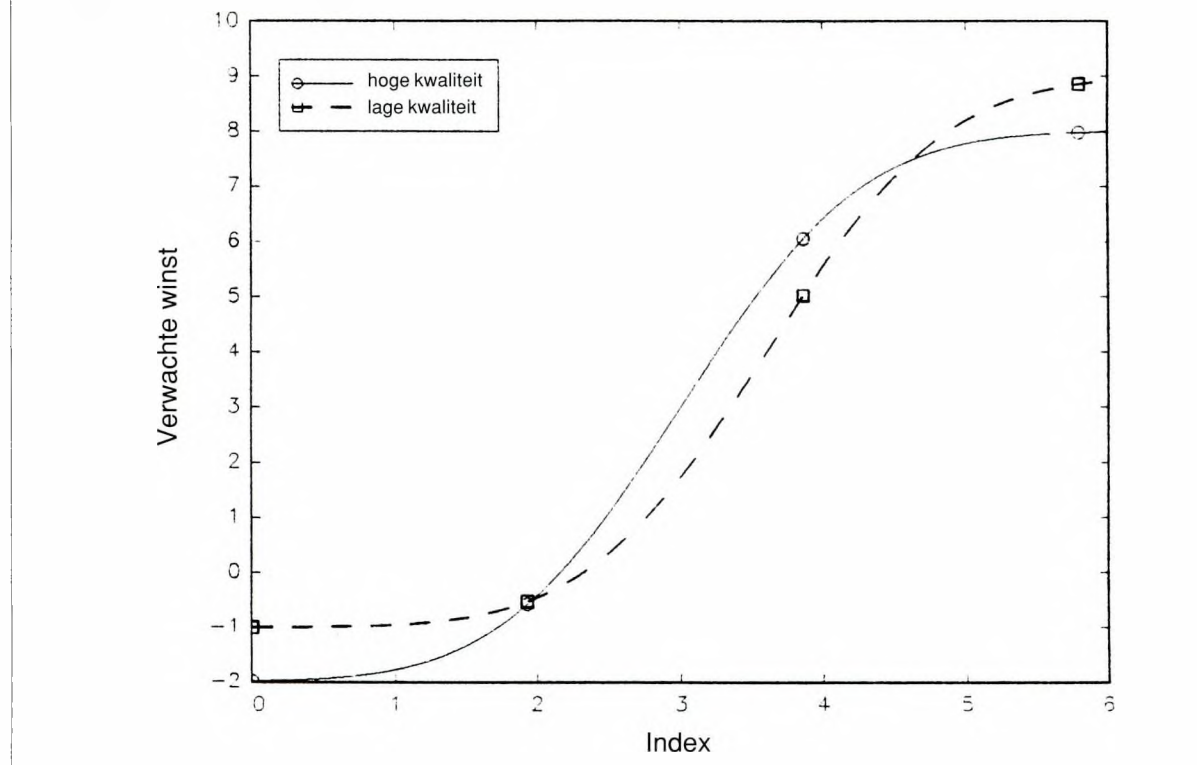
$$\Pi(q) = r(q) - q * c * N,$$

waarbij $r(q)$ is de opbrengst van de mailing

campagne gegeven q , c zijn de kosten van de mailing en N is het aantal huishoudens in het bestand. De totale kosten indien een percentage q geselecteerd wordt is gelijk aan $q * c * N$. De waarde van q is gebaseerd op selectievariabelen uit een postcode-informatiesysteem. Indien een organisatie in staat zou zijn om de gegevens te verzamelen zonder meetfouten dan zou op basis van de profit maximization benadering het optimaal percentage huishoudens berekend kunnen worden, q^{opt} . Deze waarde zal afwijken van het gekozen percentage q en dus wijkt de verwachte waarde van de winst af van de maximale winst die behaald zou kunnen worden indien de organisatie zou beschikken over perfecte informatie. De afwijking ten opzichte van de maximale winst is afhankelijk van de grootte van de binnen-variancie. De wiskundige afleiding van deze relatie wordt hier achterwege gelaten; voor een formele afleiding zie Bult (1993b).

In figuur 3 wordt een mogelijke relatie weergegeven tussen de maximaal haalbare winst gegeven de hoeveelheid binnen-variancie. De figuur laat zien dat de winst die behaald kan worden maximaal is indien er geen sprake is van meetfouten (binnen-variancie is gelijk aan 0). Indien de

Figuur 4: Multilayer mailing strategie



binnen-variantie toeneemt zal de verwachte winst afnemen tot een punt waarop de winst gelijk is aan de situatie waarbij geen gebruik wordt gemaakt van postcodegegevens (niveau 25). Met behulp van deze grafiek kan het verwacht rendement van het huren van postcode-informatiesystemen worden vergeleken met de kosten van verwerving van dergelijke systemen.

5.2 De multilayer mailing strategie

Een tweede toepassing van de profit maximization benadering, de multilayer mailing strategie, is uitvoerig beschreven in Bult, van der Scheer en Wansbeek (1994). Hieronder volgt een beknopte weergave. Het succes van direct mail hangt af van vier factoren (zie de inleiding). In de literatuur zijn verschillende studies bekend die elk van deze factoren afzonderlijk proberen te optimaliseren. Men houdt geen rekening met mogelijke interactie-effecten tussen de factoren. Het is echter niet onwaarschijnlijk dat er interactie bestaat tussen de karakteristieken van de huishoudens en de communicatieve elementen van de mailing. Charitatieve instellingen die fondsen werven met behulp van direct mail, maken bijvoorbeeld onderscheid

tussen ledenmailings (gericht op huishoudens die reeds in het bestand zitten) en wervingmailings (gericht op nieuwe donateurs). De communicatieve elementen die gebruikt worden hangen dus af van hoe lang iemand reeds in het bestand zit. Aan de andere kant zal het optimaal percentage huishoudens dat gemaïld dient te worden afhangen van de aantrekkelijkheid van een mailing. Gegeven de interactie dient een organisatie de twee factoren gelijktijdig te optimaliseren. Een methode om dit te doen is de multilayer mailing strategie.

Stel dat een organisatie twee typen mailings kan versturen, een mailing van hoge kwaliteit en een mailing van lage kwaliteit. De hoge kwaliteit mailing zal vermoedelijk meer respons genereren maar is ook duurder. De organisatie moet dus een afweging maken tussen een hogere kans op respons en de hogere kosten per contact. Op basis van de profit maximization benadering wordt voor elk huishouden de verwachte netto opbrengst berekend gegeven de waarde van de index (systematische variantie). Een huishouden ontvangt vervolgens, afhankelijk van deze berekening, de hoge kwaliteit mailing, de lage kwaliteit mailing of geen van beide. De verwachte netto opbrengst

voor iedere waarde van de index van de huishoudens kan worden afgezet in een grafiek (zie figuur 4). De hoge kwaliteit mailing wordt gestuurd naar alle huishoudens waarvan de waarde van index ligt tussen 2,10 en 4,70. Huishoudens met een hogere index (dit zijn mensen die met grote waarschijnlijkheid reageren) ontvangen de goedkopere (lage kwaliteit) mailing. Huishoudens met een index lager dan 2,10 ontvangen geen van beide.

6 Conclusie

De belangstelling voor direct marketing is de laatste jaren sterk toegenomen. Dit heeft geleid tot een toenemend aantal wetenschappelijke artikelen over direct marketing. Het vakgebied leent zich bij uitstek voor kwantificering van problemen omdat de kosten en het resultaat (wel of niet respons) van communicatie op individueel niveau kunnen worden gemeten.

Ondanks deze informatie bestaat er in de praktijk geen statistische methode die huishoudens selecteert op basis van winstmaximalisatie. De profit maximization benadering, beschreven in dit artikel, is de eerste selectiemethode gebaseerd op winstmaximalisatie.

De methode kan eenvoudig worden aangepast zodanig dat het in staat is om, bijvoorbeeld, de waarde van postcode-informatiesystemen te bepalen. Daarnaast is het eenvoudig om op basis van de profit maximization benadering simultaan verschillende succesbepalende factoren van direct mail te optimaliseren. De profit maximization selectiemethode is ontwikkeld voor het selecteren van adressen op de consumentenmarkt voor een direct mail campagne. Dezelfde technieken kunnen uiteraard ook worden toegepast voor telemarketing. Hierbij worden telefoonnummers geselecteerd voor telefonische communicatie op basis van kenmerken. De selectiemethode kan niet zonder meer worden toegepast voor het selecteren van prospects op industriële markten. Om de responscurve en de cutoff curve betrouwbaar te

kunnen schatten, uitgaande van de situatie dat de respons van een direct marketing actie laag is, dienen voldoende gegevens beschikbaar te zijn. Wat onder voldoende gegevens moet worden verstaan is moeilijk te zeggen. Dit is dus onder andere afhankelijk van de respons. Men is het er echter wel over eens dat in situaties waar sprake is van minder dan duizend potentiële adressen de in dit artikel geschetste technieken niet moeten worden toegepast.

L I T E R A T U U R

- Bult, J.R., (1993a), Semiparametric versus parametric classification models: an application to direct marketing, *Journal of Marketing Research*, 30, 380-390.
- Bult, J.R., (1993b), *Target selection for direct marketing*, Proefschrift.
- Bult, J.R. en J.C. Hoekstra, (1991), Marktonderzoek en direct marketing, in: *Jaarboek van de Nederlandse Vereniging van Marktonderzoekers 1991-1992*, Uitgeverij de Vrieseborch, Haarlem.
- Bult, J.R., H.R. van der Scheer en T.J. Wansbeek, (1994), *Multilayer mailing strategy*, Manuscript Rijksuniversiteit Groningen.
- Bult, J.R. en T.J. Wansbeek, (1995), *Optimal selection for direct mail*, *Marketing Science*, verschijnt binnenkort.
- Hoekstra, J.C. en M.C. Raaijmakers, (1991), Betekenis en groei van direct marketing, *Maandblad voor Accountancy en Bedrijfseconomie*, 65, 363-372.
- Judge, G.C., W.E. Griffiths, R.C. Hill, H. Lütkepohl en T.C. Lee, (1985), *The theory and practice of econometrics*, Wiley, New York, NY.
- Maddala, G.S., (1983), *Limited dependent and qualitative variables in econometrics*, Cambridge University Press, Cambridge.
- Roberts, M.L. en P.D. Berger, (1989), *Direct marketing management*, Prentice-Hall Inc., Englewood Cliffs, NY.
- Wedel, M., W.S. DeSarbo, J.R. Bult en V. Ramaswamy, A latent class poisson regression model for heterogeneous count data, *Journal of Applied Econometrics*, 8, 397-411.