

MULTI-MODE EN MULTI-DEVICE SURVEYS

Oratie bij leerstoel Methoden en Statistiek, in het bijzonder Mixed-Mode Survey Designs

Barry Schouten

13 februari 2018

Motivatie

Mijnheer, de Rector Magnificus, beste collega's, familie en vrienden, het is een eer om een oratie te mogen geven over de bijzondere leerstoel Methoden en Statistiek, in het bijzonder Mixed-Mode Survey Designs. De titel Multi-mode en multi-device surveys van deze oratie heeft ongeveer dezelfde strekking als die van de leerstoel maar benadrukt het gebruik van devices; wat een centraal thema is in huidige vernieuwing van surveys. Onze kinderen waren enkele weken terug bezig met een grote stad van blokken en auto's. Ik zei dat ik dit vroeger ook deed en een auto had bij de ingang van de stad als bewaker. Mijn dochter zei daarop: "Het is nu een modernere tijd, de poort herkent de auto's en dat gaat automatisch."





Het vakgebied statistiek zoekt naar patronen waarmee gebeurtenissen begrepen, gecontroleerd en eventueel gemanipuleerd kunnen worden. Het probeert daarmee relevante signalen te scheiden van niet-relevante signalen die dan als ruis worden opgevat. Dit schilderij van Jackson Pollock, genaamd Convergence (convergentie) laat een op het eerste gezicht chaotisch beeld zien. Het is echter zeer gestructureerd en volgens een vaste werkwijze gemaakt. Pollock heeft veel van dit soort schilderijen gemaakt in allerlei kleurstellingen; dit schilderij mengt juist de kleuren van CBS en universiteit Utrecht. Ik vind het één van de mooiste schilderijen die ik ken. Surveys zijn een instrument om structuur te vinden voor beleid en onderzoek maar zijn zelf ook onderhevig aan allerlei verstoringen, ook wel survey errors of survey fouten genoemd. Om die reden zijn surveys als instrument al decennia onderwerp van statistisch onderzoek.

Maar wat is eigenlijk een survey, of ook wel enquête, en wat zijn modes en devices? Een survey is de verzameling van informatie onder een (steekproef van een) gespecificeerde doelpopulatie met behulp van een vragenlijst, meestal over een specifiek thema. Een survey gaat dus over communicatie van informatie en heeft daarmee communicatiemediën nodig; dat zijn modes. Traditionele modes zijn schriftelijke surveys, telefonische surveys en persoonlijke surveys. Ongeveer 15 jaar geleden is de online mode sterk opgekomen; vooral vanwege zijn lage kosten en hoge snelheid. Tegenwoordig zijn er nog betrekkelijk weinig surveys die geen internet gebruiken. Benaderde je vroeger het internet

bijna uitsluitend via een desktop of laptop, tegenwoordig is er een veelheid aan apparaten waarmee dat kan; van de kleine smartphone tot aan de grote smart TV. Deze worden devices genoemd.

Surveys, modes en devices

- Een survey: Een vragenlijst onder een steekproef uit een doelpopulatie, meestal over een specifiek thema;
- Een survey mode: Medium waarmee respondenten voor de survey hun informatie aanleveren;
- Een survey device: Bij online aanlevering, het device waarmee contact met internet wordt gemaakt;



 Multi-mode en multi-device surveys, 13 februari 2018 

Het onderwerp van de leerstoel en van deze oratie lijken (mogelijk) wat beperkt of specifiek. Dat is zeker niet zo. Het is al verscheidene jaren bekend dat modes, en in het verlengde daarvan devices, invloed hebben op alle foutenbronnen in surveys en daarmee op de uitkomsten van surveys. Deze methode-effecten, ook wel mode effecten genoemd, komen voort uit verschillen in toegang tot een mode (heb je een smartphone, heb je een geregistreerd telefoonnummer), bereikbaarheid met een mode (neem je de telefoon op, maak je een brief open, doe je deur open), responsbereidheid met een mode (heb je tijd en ben je bereid mee te doen) en antwoordgedrag (begrijp je de vragen, ben je gemotiveerd deze te beantwoorden). Alle mogelijke foutenbronnen in surveys komen aan de orde. Als je deze eigenschappen van modes afzet tegen de grote verschillen in kosten, dan ontstaat een complex optimalisatieprobleem. Dit is eigenlijk een alledaags probleem voor survey organisaties.

Modes en devices beïnvloeden waarneming

- Selectie: Personen verschillen in toegang tot, bereikbaarheid met en geneigdheid mee te doen via modes/devices;
- Meting: Antwoorden die respondenten geven worden beïnvloed door een mode/device;

Selectie- en meetverschillen heten samen mode/device effecten;

 Multi-mode en multi-device surveys, 13 februari 2018 

Dit klinkt voor velen mogelijk nog wat abstract. Ik wil dit daarom aan de hand van een voorbeeld duidelijker maken. Ik vraag alle toehoorders (iedereen hier in de zaal) om mee te doen. Ik ga u een aantal vragen voorleggen. Wilt u bij ja gaan staan? Stelt u voor dat u wordt uitgenodigd met een brief van het CBS voor een onderzoek naar gezondheid. Deelname is niet verplicht. U kunt dat online doen via een login en wachtwoord; een internetadres is gegeven. Zou u deze vragenlijst op uw smartphone

of tablet gaan invullen? Of zou deze vragenlijst op een traditionele PC of laptop invullen? De vragenlijst duurt ongeveer een half uur. Gaat u dat afmaken? Als u nog niet hebt meegedaan of bent afgehaakt, dan benadert een interviewer u aan huis om mee te doen. Zou u dat doen? Percentages zijn in werkelijkheid 10% op smartphone/tablet, 25% op desk/laptop en 20% F2F. Ik wil u twee vragen uit de vragenlijst voorleggen. Denkt u daarbij aan de mode of het device. De eerste vraag is uw BMI, ofwel lengte en gewicht. Weet u dat en geeft u ook het correcte antwoord? De tweede vraag gaat over bewegen. Weet u in uren en minuten hoeveel u op een gemiddelde dag zit? Dit voorbeeld was een poging om u allemaal te laten opstaan en weer te laten zitten.

MODE EFFECTEN?

Onderzoek naar gezondheid met een combinatie van online en F2F


Vraag 1: Meedoen online via smartphone – tablet – lap/desktop?

Vraag 2: Blijft u 30 minuten meedoen?


Vraag 3: Alsnog meedoen met een F2F interviewer?

Vraag 4: Antwoord op vraag naar BMI, ofwel gewicht en lengte?

Vraag 5: Antwoord op aantal uren/minuten zitten op gemiddelde dag?




Multi-mode en multi-device surveys, 13 februari 2018




Deze tabel toont schattingen voor vier gezondheidsstatistieken onder de Nederlandse bevolking: % zeer goede/goede gezondheid, % roken, %obesitas (gebaseerd op BMI) en % tandartsbezoek. De eerste kolom bevat de schattingen met enkel online waarnemingen. De tweede kolom bevat schattingen voor personen die niet online meededen maar later wel F2F. Er zitten relatief grote verschillen tussen. Wat veroorzaakt dit? Krijgen we met F2F een ander soort personen of geven ze andere antwoorden of beide?

Gezondheidsenquête

Statistiek	Online	F2F
Goede gezondheid	78.0%	75.6%
Roken	19.9%	29.8%
Obesitas	12.1%	13.9%
Tandartsbezoek	82.3%	74.5%



Multi-mode en multi-device surveys, 13 februari 2018



Terug naar het onderwerp van deze oratie: Het algemene doel van de leerstoel is bij te dragen aan statistieken die nauwkeurig zijn en die vergelijkbaar zijn over de tijd en tussen personen binnen budgettaire kaders. Ik zie drie grote onderzoeksthema's die daarvoor moeten zorgen: De eerste is een veel grotere focus op antwoordgedrag en meeteffecten. De tweede is differentiatie in de benadering van personen en huishoudens; ook wel doelgroepenbenadering genoemd. De derde is vernieuwing met

nieuwe modes en devices. Deze drie thema's hangen sterk met elkaar samen. Differentiatie is een middel om om te gaan met meeteffecten. Modes en devices zijn keuzes in een differentiatie maar hebben dus ook hun invloed op statistieken. Ik ga deze drie thema's nu kort langs. Je kunt er makkelijk per thema een uur over praten. Ik beperk me tot de stand van zaken en de speerpunten die ik zie.



Hoofddoel van de leerstoel

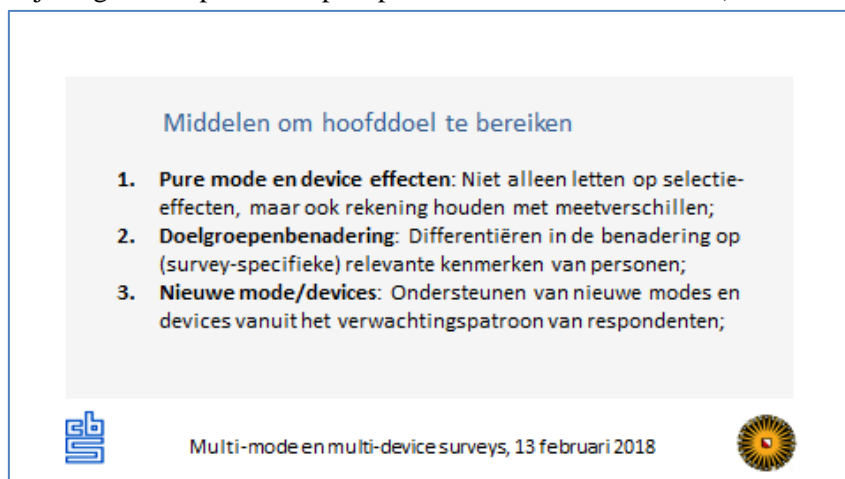
Optimaliseren van nauwkeurigheid en vergelijkbaarheid van statistieken, rekening houdend met beschikbare budgetten;

Modes/devices spelen in toenemende mate een bepalende rol

 Multi-mode en multi-device surveys, 13 februari 2018 



Pure mode effecten

Eerst de mode- en device-specifieke meeteffecten, ook wel pure mode effecten genoemd: Laten we eerst eens kijken naar de oorsprong van de pure mode effecten. Literatuur wijst op vier eigenschappen waarop modes/devices ongelijksoortig zijn en verschillen kunnen veroorzaken. Dat zijn de aanwezigheid van een interviewer, het tempo van het invullen, de presentatie van de vragenlijst en de timing. Interviewers beïnvloeden antwoordgedrag; ze kunnen motiveren en uitleggen, maar ook sociaal-wenselijkheid introduceren. Het tempo van invullen varieert per mode; telefoongesprekken gaan sneller dan persoonlijk gesprekken. De presentatie, ofwel zie je de vragenlijst en hoe, of luister je naar de vragen, maakt verschil. Tot slot, is het moment van invullen per mode/device anders (wanneer wordt de vragenlijst ingevuld, spontaan, op afspraak of wanneer het uitkomt).



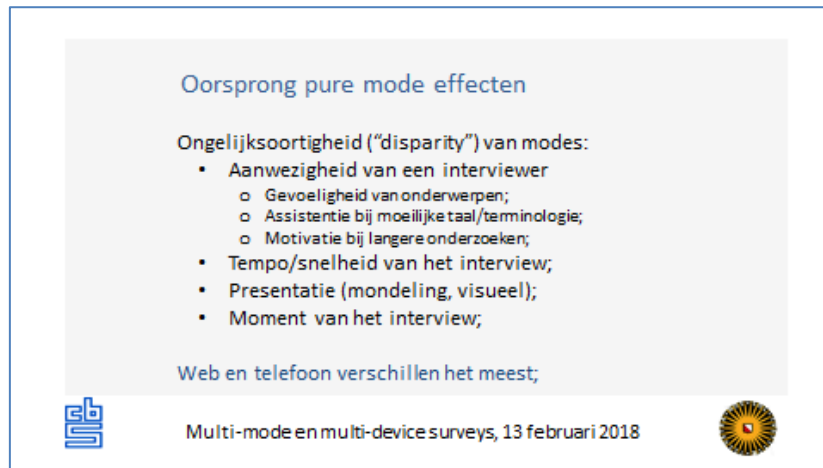
Middelen om hoofddoel te bereiken

1. **Pure mode en device effecten:** Niet alleen letten op selectie-effecten, maar ook rekening houden met meetverschillen;
2. **Doelgroepenbenadering:** Differentiëren in de benadering op (survey-specifieke) relevante kenmerken van personen;
3. **Nieuwe mode/devices:** Ondersteunen van nieuwe modes en devices vanuit het verwachtingspatroon van respondenten;

 Multi-mode en multi-device surveys, 13 februari 2018 

Pure mode effecten worden meestal als ongewenst gezien. Methodologie om deze te reduceren omvat vier opties; hier in volgorde gezet naar fase in de uitvoering van een survey. Deze volgorde is, naar mijn mening, ook de prioritering. Je kunt de effecten voorkomen door vragenlijstontwerp aandacht te geven door middel van test over meerdere modes en devices. Je kunt de effecten vermijden door sommige combinaties van modes niet te kiezen in een survey omdat ze relatief ongelijksoortig zijn, eventueel enkel voor bepaalde deelpopulaties. Je kunt de effecten corrigeren, ofwel proberen te verwijderen uit de statistieken. En je kunt de effecten stabiliseren door met vaste mode verhoudingen

te werken. Literatuur heeft (terecht) veel aandacht gehad en nog steeds voor het voorkomen van effecten. De overige opties hebben minder aandacht gehad, hoewel er wel in de laatste vijf jaar een serie artikelen is verschenen over correctie. Doelgroepenbenadering ligt dicht aan tegen vermijden van effecten als je modes en devices differentieert. Ik zal me daarom concentreren op corrigeren.





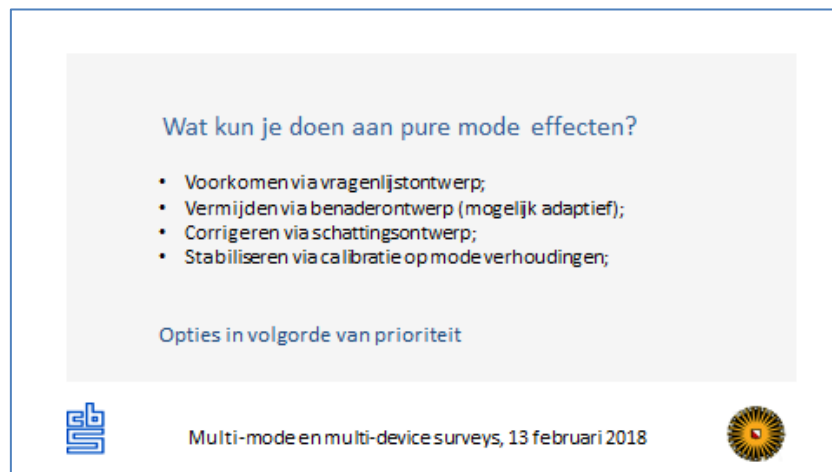
Oorsprong pure mode effecten

Ongelijksoortigheid ("disparity") van modes:

- Aanwezigheid van een interviewer
 - Gevoeligheid van onderwerpen;
 - Assistentie bij moeilijke taal/terminologie;
 - Motivatie bij langere onderzoeken;
- Tempo/snelheid van het interview;
- Presentatie (mondeling, visueel);
- Moment van het interview;

Web en telefoon verschillen het meest;



 Multi-mode en multi-device surveys, 13 februari 2018 



Wat kun je doen aan pure mode effecten?

- Voorkomen via vragenlijstontwerp;
- Vermijden via benaderontwerp (mogelijk adaptief);
- Corrigeren via schattingsontwerp;
- Stabiliseren via calibratie op mode verhoudingen;

Opties in volgorde van prioriteit



 Multi-mode en multi-device surveys, 13 februari 2018 

Correctie van pure mode effecten kan op drie manieren: Alle manieren brengen op één of andere wijze extra informatie in. De eerste werkwijze is het inbouwen van een onderliggende structuur of samenhang in de vragenlijst waarvan bekend is of verondersteld wordt dat die bestaat; zogenaamde constructen of factoren. Dit is een oude tactiek die veranderingen op de veronderstelde, latente variabelen verbindt met pure mode effecten. De vragenlijst wordt er langer van en het is noodzakelijk te veronderstellen dat respondenten onafhankelijk op elke vraag reageren. De tweede werkwijze is het koppelen van externe bronnen met relevante variabelen waarmee selectie-effecten geschat en verwijderd kunnen worden. Deze variabelen moeten wel bestaan natuurlijk. Het CBS is wat dat betreft een goede omgeving om te werken omdat het tijdelijk allerlei bronnen kan koppelen. De derde werkwijze is het doen van herhaalde metingen op dezelfde respondenten. Dit is een investering, want je neemt twee keer een deel van de vragen af. Bovendien moet je veronderstellen dat de respondent sinds de eerste meting geen ander mens is geworden. Ik ben van mening dat deze derde werkwijze, mits experimentele invloed kan worden geminimaliseerd, de meest krachtige is om pure mode effecten te schatten. Het is een tactiek die is getest op het CBS en die uitgebreid is geanalyseerd door Thomas Klausch. Thomas is hier aan de universiteit gepromoveerd enkele jaren geleden. De volgende tabel geeft een voorbeeld van de schattingen die met een dergelijk re-interview zijn gemaakt voor de Enquête Beroepsbevolking (EBB), en wel een belangrijke statistiek, de arbeidspositie van Nederlanders. De linker kolommen geven de totale mode effecten van online en telefonische waarneming afgezet tegen aan-huis waarneming; u ziet de relatieve verschillen afgezet tegen een

waarneming aan huis. De middelste en rechter kolommen geven schattingen voor de verschillen in selectie en in de antwoorden zelf. De blauwe cellen toetsen significant op 5%. Voor arbeidspositie zijn geschatte verschillen dus voornamelijk toegeschreven aan verschillen in de selectie van personen.

Hoe kun je pure mode effecten corrigeren?



- Psycho/sociometrie: Introduceer latente variabelen en schrijf verschillen in betrouwbaarheid en validiteit toe aan modes/devices;
- Calibratie: Weeg nonrespons en schrijf resterende verschillen toe aan pure mode effecten;
- Re-interview: Nodig respondenten uit voor een tweede interview in een andere mode, meestal F2F;


Multi-mode en multi-device surveys, 13 februari 2018


Methodologische opties - corrigeren

Re-interview schattingen Enquête Beroepsbevolking:

	Mode effect		Selectie-effect		Meet-effect	
	Tel	Web	Tel	Web	Tel	Web
Werkzaam	0.4%	8.9%	-0.3%	9.7%	0.7%	-0.7%
Werkloos	-3.0%	-2.6%	-1.1%	-0.9%	-1.9%	-1.7%
Overig	2.6%	-6.3%	1.4%	-9.8%	1.2%	2.4%


Multi-mode en multi-device surveys, 13 februari 2018


Dan de speerpunten: Ik zie twee vraagstukken waar in de literatuur nog weinig aandacht aan is besteed maar die voor statistische bureaus wel zeer relevant zijn. Het eerste vraagstuk is de ontwikkeling van profielen van vragenlijsten waarmee relatief snel en goedkoop besloten kan worden wat het risico is op pure mode effecten. Met die bepaling van risico kan dan gekozen worden hoeveel energie en tijd moet worden gestoken in het vermijden van de effecten via cognitieve tests en/of benaderstrategieën. Dit is het onderwerp van het promotie-onderzoek van Frank Bais aan de UU. Het gaat om de relatie tussen item kenmerken, denk aan complexiteit en gevoeligheid, respondentkenmerken, zoals opleiding en leeftijd, antwoordgedrag, zoals antwoorden weigeren of snel antwoorden, en natuurlijk de eerder genoemde eigenschappen van modes en devices. Voor 11 surveys hebben we inmiddels dergelijke profielen opgesteld en we werken momenteel aan twee nieuwe surveys, waaronder weer de gezondheidsenquête. Het tweede vraagstuk is de kosten-baten, zeg maar de business case, van re-interview designs. Deze designs vragen om een investering omdat respondenten, zoals gezegd, twee keer een antwoord geven. Deze investering kan echter qua nauwkeurigheid, meer specifiek de verwachte kwadratische fout, positief zijn. Ofwel het verlies aan precisie doordat budget (en dus steekproefomvang) geofferd moet worden voor re-interviews wordt gecompenseerd door een winst in vertekening. Dit is mogelijk als statistieken langzaam veranderen, als het risico op meetverschillen groot is door de aard van de vragen, en als de meetfouten zelf relatief constant zijn. Voor de Enquête Beroepsbevolking (arbeidspositie) pakt dit niet goed uit door de grote dynamiek in werkloosheid, maar


voor gezondheidsstatistieken is er wel een business case. Dit leidt overigens wel tot een cirkelredenering; als meeteffecten groot zijn dan kan MSE dalen, maar de re-interview is nu juist bedoeld om die effecten te schatten. Dit tweede vraagstuk heeft dus ook weer baat bij de het eerste vraagstuk, de vragenlijstprofielen.

Speerpunt vragenlijstprofielen


Interactie tussen:

- Persoonskenmerken;
- Item kenmerken;
- Antwoordgedrag;
- Eigenschappen modes/devices;

Inschattingen van risico op mode/device effecten aan de hand van vragenlijstprofielen.



Multi-mode en multi-device surveys, 13 februari 2018




Speerpunt mode/device effect correctie


Kosten-baten re-interview investering zijn positief als:

- Kernstatistieken relatief stabiel zijn in de tijd;
- Pure mode effecten relatief groot kunnen zijn;
- Meetfouten niet sterk tijdsafhankelijk zijn;

Verkennde analyse:
EBB: Te dynamisch en pure mode effecten relatief klein;
GEZO: Stabiel en pure mode effecten soms mogelijk groot;



Multi-mode en multi-device surveys, 13 februari 2018



Doelgroepenbenadering

Het tweede middel is doelgroepenbenadering, in Engels adaptive survey design genoemd. Doelgroepenbenadering gaat uit van de voor de hand liggende gedachte dat personen verschillen in hun voorkeuren, antwoordgedrag en kosten voor verschillende benaderstrategieën; hier dus verschillende modes en devices. Hoewel deze gedachte voor de hand ligt, wordt doelgroepenbenadering nog maar beperkt toegepast. Het betekent namelijk een meer complexe logistiek, vereist flexibele systemen en monitoring en het strookt niet altijd met de eisen die opdrachtgevers stellen, namelijk zoveel mogelijk precisie (ofwel respons). Desondanks is de interesse in deze werkwijze sterk gestegen vanwege de sterke druk op responscijfers, de grote variatie aan mogelijk benaderstrategieën en de automatisering van deelprocessen van waarneming.

Doelgroepenbenadering (adaptive survey design)

Adaptive survey designs wijzen personen toe aan verschillende benaderingen voor een optimale kwaliteit-kosten balans

Redenen voor toegenomen interesse in ASD:

- Meer urgentie;
- Meer meters;
- Meer opties




Multi-mode en multi-device surveys, 13 februari 2018


De vier hoofdingrediënten van doelgroepenbenadering zijn hulpinformatie, kwaliteit- en kostencriteria, benaderstrategieën en een optimalisatietactiek. Met de hulpinformatie kun je deelpopulaties onderscheiden die van belang zijn maar ook verschillend reageren, denk aan hoogte van het inkomen of iemands leeftijd. De expliciete kwaliteit- en kostencriteria, bijvoorbeeld de hoogte en spreiding van responskansen en meeteffecten, zijn nodig om keuzes te kunnen maken. Deze zijn niet altijd zo expliciet zodat doelgroepenbenadering kan leiden tot fundamentele discussies. De benaderstrategieën zijn, in dit geval de modes. De tactiek om een optimale balans te vinden is het meest statistisch en artikelen kunnen, analoog aan medische statistiek waar differentiatie al veel langer wordt onderzocht, zeer technisch zijn.

Elementen in doelgroepenbenadering

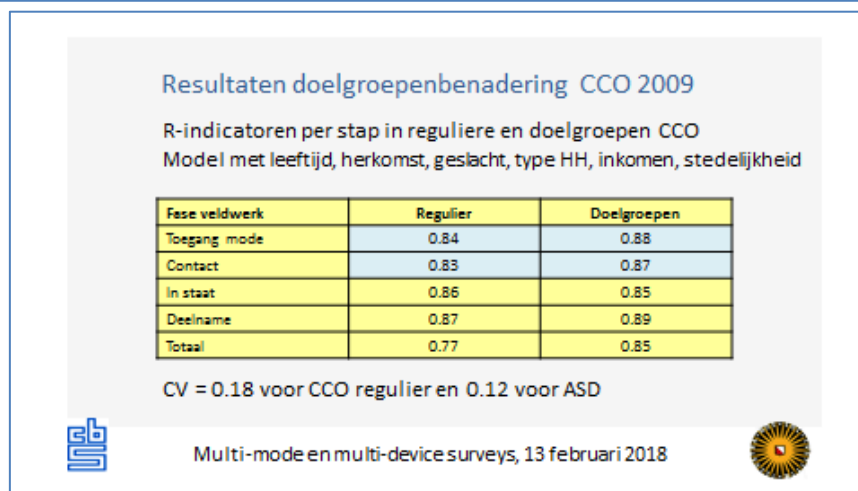
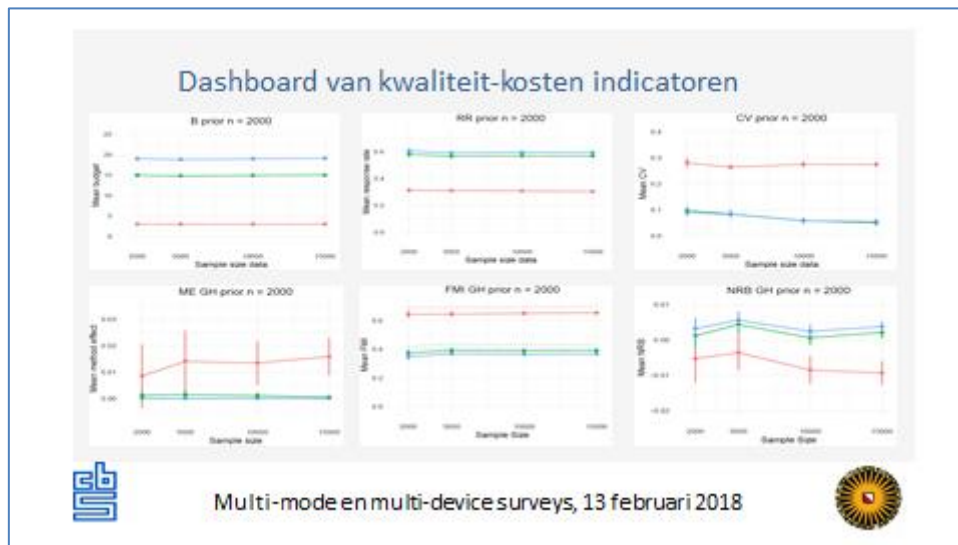
1. Hulpinformatie (kader, registers, paradata, eerdere metingen);
2. Design kenmerken/interventies;
3. Kwaliteit-kosten functies;
4. Optimalisatie strategie;

Zowel zeer statistische als praktische onderzoeksvragen.


Multi-mode en multi-device surveys, 13 februari 2018


Doelgroepenbenadering relateert sterk aan de zogenaamde plan-do-check-act cyclus. De vier ingrediënten komen daar terug en analyse is constant nodig om te kunnen inschatten hoe deelpopulaties op verschillende benaderstrategieën reageren. Het survey klimaat is namelijk altijd in beweging, al vele jaren is er een langzame daling in responscijfers, en dat geldt zeker voor de wijze van communicatie. Wat vijf jaar geleden een optimaal design was, kan nu dus niet meer optimaal zijn. Doelgroepenbenadering gaat daarom uit van zogenaamde dashboards met kwaliteit en kosten indicatoren. Dit is een voorbeeld hiervan geproduceerd door een stagiaire, Eva de Jong. Het dashboard toont zes panelen waarbij de horizontale as aangeeft hoe de indicatoren veranderen naarmate meer waarnemingen worden gedaan. In elke figuur zijn drie lijnen opgenomen, één voor enkel online waarneming (rood), één voor een korte F2F herbenadering (groen) en één voor een uitgebreide F2F herbenadering (blauw). Linksboven de kosten per eenheid, midden-boven het responscijfer, rechtsboven de variatiecoëfficiënt in responskansen, linksonder het methode-effect in kernvariabele algemene gezondheid, midden-onder de fractie missende informatie in algemene gezondheid, en rechtsonder de geschatte vertekening in algemene gezondheid. Zegt u het maar, welk design zullen we

kiezen? Dat hoeft geen uniforme keuze te zijn natuurlijk. Voor doelgroepenbenadering moeten deze indicatoren verder uitgesplitst worden naar doelgroepen. In de literatuur waar het CBS sterk aan heeft bijgedragen zijn daarvoor allerlei oplossingen bedacht.



Een simpel voorbeeld van doelgroepenbenadering is bijna 10 jaar terug toegepast op een doorlopend, maandelijks onderzoek, het Consumenten Conjunctuur Onderzoek. Op basis van de bekende leeftijd, herkomst, geslacht, type huishouden, inkomen, en stedelijkheid van de woonomgeving van steekproefpersonen, werd een doelgroepenbenadering gekozen. Doel was een zo gebalanceerd mogelijke respons op deze zes kenmerken met behoud van het responscijfer en budget. Om de mate van balans van de respons te meten zijn zogenaamde R-indicatoren gebruikt waarvan de waarde zo dicht mogelijk bij 1 moet liggen. De tabel laat de uitkomst van een doelgroepenbenadering zien tegen een uniforme aanpak (de reguliere waarneming). De onderste kolom laat zien dat de uiteindelijke respons een betere balans heeft. Deze betere balans is vooral verkregen door betere toegang tot de modes en door verbetering van het contactleggen. In benadering is het niet gelukt om ook deelname kansen mee gebalanceerd te maken.

Waar liggen de speerpunten? Opnieuw noem ik twee vraagstukken. De eerste is theoretisch. Binnen de survey methodologie bestaat kritiek op de veronderstelling dat vertekening van statistieken gereduceerd wordt door tijdens een survey te balanceren op hulpinformatie. De kritiek komt voort uit de gedachte dat deze informatie ook achteraf in het schattingsproces kan worden ingezet en daar dezelfde winst in vertekening geeft, en dat dus ook nog eens zonder een complexe logistiek. De afgelopen jaren is enige literatuur geschreven, onder andere door mezelf, over de condities waaronder

balanceren op bekende kenmerken effectief is, zelfs na correctie achteraf. Ik licht hier twee stellingen uit. De eerste stelt dat als design 1 meer variatie laat zien in responskansen dan design 2 (gemeten met de variatiecoëfficiënt) voor een willekeurige set variabelen, dat dit dan naar verwachting waar is voor alle variabelen samen. Ofwel naar verwachting is design 1 beter, zelf na correctie met die hulpvariabelen. Dat is een mooi resultaat maar zegt nog niks over balanceren. De tweede stelling zegt dat dit nog steeds waar is als je had gebalanceerd op die hulpvariabelen. De stellingen worden echter aanzienlijk krachtiger als je de relaties tussen variabelen op de populatie expliciet modelleert, omdat dan ook de winst gekwantificeerd kan worden. Vraag is of en hoe dat te doen? Dit is fundamenteel onderzoek dat feitelijk de standaard aannames over missende data vervangt door aannames over hoe variabelen tot stand zijn gekomen.


Speerpunt – Effectiviteit doelgroepenbenadering

Kan doelgroepenbenadering effectief zijn ook na inzet van dezelfde hulpinformatie in schattingsmethoden? JA


Stelling 1: Als $CV(\text{design 1}) < CV(\text{design 2})$ voor een willekeurige set hulpvariabelen, dan $ECV(\text{design 1}) < ECV(\text{design 2})$ voor alle hulpvariabelen;

Stelling 2: Stelling 1 geldt ook als data verzameling gestuurd is op een gebalanceerde respons voor een willekeurige set hulpvariabelen;

Stellingen worden nog krachtiger als covarianties tussen variabelen gemodelleerd worden. Hoe?



Multi-mode en multi-device surveys, 13 februari 2018




Speerpunt - Inbrengen expertkennis/historie


Bayesian Adaptive survey DEsign Network (BADEN):

- Expertkennis en historische survey data inbrengen;
- Expliciet rekening houden met onzekerheden via priors;
- Aanpassen kennis met data via posteriors;
- Optimalisatie doelgroepenbenadering vanuit onzekerheid;

Methodologische en statistische onderzoeksvragen



Multi-mode en multi-device surveys, 13 februari 2018



Het tweede vraagstuk gaat over het inbrengen van kennis. Zoals eerder gezegd is het survey klimaat voortdurend in beweging en is bijleren noodzakelijk. Maar hoe dat te doen? Zo'n drie jaar geleden is een zevental instituten, waaronder het CBS, begonnen met een Bayesiaanse analyse van surveys. Hoewel Bayesiaanse analyse in andere statistische toepassingen al enige tijd wordt gebruikt, is de toepassing voor monitoring van surveys geheel nieuw. Dit is eigenlijk vreemd, want dataverzamelingsafdelingen hebben meestal zeer veel kennis en ervaring. Die historische data en expertkennis over responskansen en methode-effecten voor verschillende benaderstrategieën worden in de analyse vertaald naar kansverdelingen die tijdens het veldwerk worden aangepast. Veel kan daarbij worden geleerd uit andere toepassingen, maar er zijn ook enkele specifieke elementen in surveys, namelijk de aandacht voor meetfouten en de expliciete focus op budgetten. De eerste resultaten zijn veelbelovend mits de niet-stationariteit, ofwel dynamiek in de parameters voor

doelgroepenbenadering beperkt is. Redelijk recent is aan de universiteit Utrecht een PhD begonnen, Shiya Wu, op dit onderwerp.

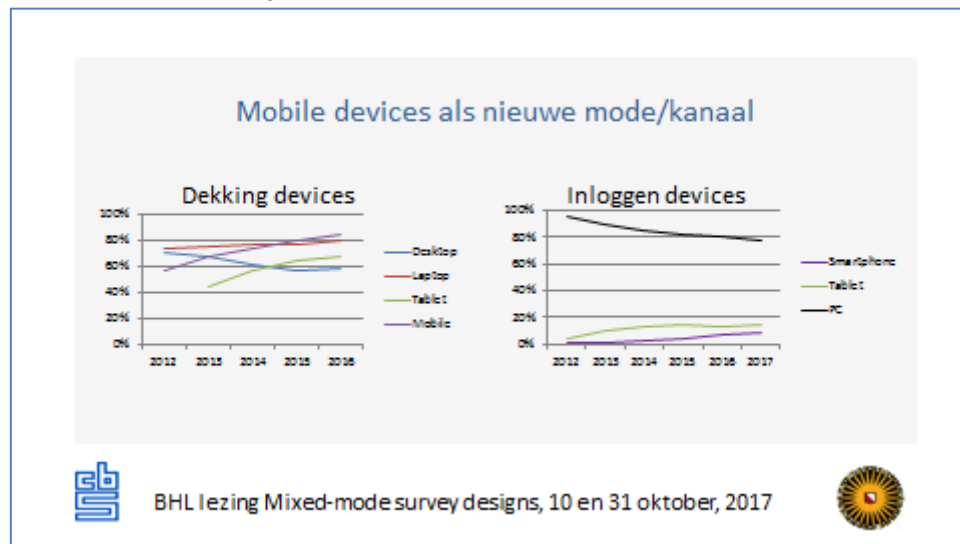
Nieuwe modes en devices

Het derde middel is vernieuwing van waarneming met allerlei devices. Nieuwe technologie verspreidt zich in een hoog tempo. Zo is er een veelheid aan mobile devices in allerlei groottes die verrassend veel kunnen opslaan, sensoren bevatten en ook nog kunnen verwerken. Een soort mini-computers met extra functionaliteit. Er zijn ook meer gespecialiseerde en wat simpelere devices die je kunt dragen zoals smartwatches en fitbits. Of waarmee je visuele effecten kunt toevoegen of oogbewegingen meten zoals smart glasses. Er zijn ook zeer gespecialiseerde devices die enkel bestaan uit maatwerksensoren, als geheel vaak Internet of Things genoemd. En dan is er allerlei kleine technologie waarmee je een stimulus kunt geven zoals bluetooth beacons en NFC tags. Kenmerkend voor al deze devices is dat ze Engelse namen hebben. Niet alles hiervan is bruikbaar voor surveys, maar de mogelijkheden groeien snel.



Introductie van mobile devices kan op twee terreinen: het invullen van vragenlijsten en het gebruik van de sensoren die ze bevatten. Het eerste is overigens niet een echte keuze; respondenten kiezen zelf voor een device. De aanvulling of vervanging van vragen met sensormetingen is vanzelfsprekend wel een keuze. Online surveys werken vanuit browsers en browsers kunnen ook sommige sensoren aansturen en uitlezen via zogenaamde API's. Dat heeft als voordeel dat een respondent niet apart software (apps) hoeft te installeren, maar leent zich niet voor langdurige metingen en voor sommige sensoren of combinaties van sensoren.

De populatiedekking van devices is sterk in beweging zoals u in deze linkerfiguur kunt zien (met dank aan Anouk Roberts en Jeldrik Bakker). Het bezit van desktops daalt geleidelijk terwijl smartphones nu van alle devices de hoogste dekking hebben in Nederland. Ook tablets zijn langzaam in dekking gestegen maar de dekking stabiliseert nu. Dat uit zich direct in de keuzes die respondenten maken. De rechterfiguur toont de aandelen online respondenten per device. Smartphones en tablets stijgen gestaag en nemen nu samen meer dan 20% van de respons voor hun rekening. Je hoeft geen ICT visionair te zijn om te vermoeden dat deze lijn zich zal doorzetten.



Hoe zit het met de devices als het gaat om pure mode effecten, ook wel device effecten genoemd? Mobile devices vertonen natuurlijk gelijkenis met desktops en laptops maar met twee belangrijke verschillen: De presentatie is anders door de schermgrootte en door het gebruik van het scherm voor navigatie. Daarnaast worden mobile devices ook zeer veel buitenshuis gebruikt waardoor de timing van de survey kan veranderen. Met name de andere presentatie noopt survey organisaties tot aanpassingen van hun surveys. Een recent artikel meldt dat in smartphone vragenlijstontwerp vijf elementen een rol spelen: grootte van tekens, grootte van knoppen, zichtbaarheid van alle opties, eenvoud in navigatie, en voorspelbaarheid over devices heen. Desondanks zijn er nog geen aanwijzingen voor grote device-effecten; mogelijk doordat smartphones nog betrekkelijk weinig worden gestimuleerd en geëvalueerd in de grotere officiële onderzoeken. De tot nu toe uitgevoerde studies zijn voornamelijk specifieke populaties en geoptimaliseerde vragenlijsten.

Dan de tweede functie, de sensoren. Een smartphone of tablet bevat al verrassend veel metingen die sensoren ondersteunen. Ik noem er een aantal: beweging/versnelling/trilling, camera, microfoon, luidspreker, GPS, Wi-Fi, GSM, Bluetooth, Near Field Communication (NFC), magnetisch veld,

vochtigheid, luchtdruk, nabijheid, licht. Dergelijke sensoren kunnen vragen in surveys aanvullen of zelfs vervangen. De belangrijkste kandidaten zijn vragen waarvan de literatuur aangeeft dat deze een groot risico op meetfouten hebben. Dat zijn vragen die een grote geestelijke inspanning vragen dan wel herinnering vereisen, zoals bijvoorbeeld het aantal uren en minuten zitten. Dat zijn vragen die kennis of informatie vragen die de respondent niet heeft, zoals de ingrediënten van alle voeding die is genuttigd. Of het zijn onderwerpen die lastig zijn te bevragen zoals iemands geestelijke toestand; daar heb je dan een hele batterij vragen voor nodig. Dit zijn ook onderwerpen die recent aan de orde zijn geweest in een Sensor data challenge die de Haagse Hogeschool (met het lectoraat Smart sensor systems van John Bolte), RIVM, Universiteit Utrecht en het CBS hebben georganiseerd. Het winnende team bestond overigens uit voormalige studenten van de UU.

Mobile device sensor data

Mobile device sensoren omvatten beweging/versnelling/trilling, camera, microfoon, luidspreker, GPS, Wi-Fi, GSM, Bluetooth, NFC, magnetisch veld, vochtigheid, luchtdruk, nabijheid, licht.

Sensor data zijn vooral interessant voor onderwerpen die:



- Een grote cognitieve inspanning of herinnering vereisen;
- Een gedetailleerde en/of esoterische kennis nodig hebben;
- Niet beantwoord kunnen worden met vragen;





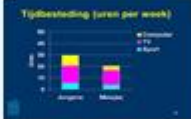

Multi-mode en multi-device surveys, 13 februari 2018




Sensor data thema's







Multi-mode en multi-device surveys, 13 februari 2018



Thema's waar we momenteel aan werken of beginnen te werken zijn koopgedrag, gezondheid, tijdbesteding, werkomstandigheden, verplaatsing/mobiliteit en woonomstandigheden. Voor al deze thema's geldt dat ze onderwerpen bevatten die voldoen aan één van de drie eerder genoemde criteria (cognitief belastend, esoterische kennis of moeilijk bevaagbaar). Per thema zou je een aparte lezing kunnen geven. Ik verwijs graag naar de intreeredes Gabriëlle Tuijthof aan Hogeschool Zuyd en van John Bolte aan de Haagse Hogeschool.

Ik zal een korte demo laten zien van sensoren om een idee te geven. Deze website is gemaakt door Ole Mussmann van het CBS. Vanuit de browser worden een aantal metingen gedaan. Op een laptop is dat niet zo spannend want die ondersteunt maar beperkt de sensoren. Mijn smartphone is al ingeschakeld op de website en laat meer mogelijkheden zien.

Mobile devices als sensoren

[Demo van enkele sensoren](#)




Multi-mode en multi-device surveys, 13 februari 2018



Speerpunten mobile devices

- Groeit de verscheidenheid aan devices?
- Apps of browser-gestuurd?
- Hoe krijgen we respondenten zover dat ze meedoen?
- Verbreden van functionaliteit van apps zodat deze meerwaarde hebben voor respondent op langere termijn?
- Kwantitatieve versus kwalitatieve sensor metingen?
- Hoe data te analyseren en verwerken (data mining)?



Multi-mode en multi-device surveys, 13 februari 2018



Wat zijn de speerpunten in de introductie van nieuwe devices? Als het gaat om ontwerp dan is een belangrijke vraag vooral de grote verscheidenheid aan modellen en schermgroottes. Hoe ontwikkelt dit zich in de nabije toekomst en hoe moeten we daar in design mee omgaan. De neiging bestaat een soort zwakste schakel aanpak te volgen waarbij de kleinere schermen/modellen maatgevend zijn. Dan is er de vraag of we in de browser blijven werken of overstappen op maatwerk apps. Zoals gezegd zal de duur en aard van de metingen bepalend zijn, maar we weten ook uit de literatuur dat meer stappen en meer keuzes voor een respondent altijd nadelig zijn. Willen Nederlanders wel een app installeren voor een onderzoek van de overheid? En ook belangrijk hoe krijgen we ze zover dat ze meedoen en vertrouwen hebben in de privacy bescherming en ethiek van de survey organisatie. In het verlengde daarvan ligt de vraag of we respondenten wat moeten terug geven, ofwel of we de app een meerwaarde moeten maken ook voor de respondent. Bijvoorbeeld doordat hij/zij over zichzelf leert. Als een respondent bereid is om mee te doen dan doemen nog twee speerpunten op: De eerste is het budget. Hoe duurder de sensoren hoe beter de nauwkeurigheid, maar onderzoeken hebben ook een zekere breedte en omvang nodig in de populatie. Waar ligt de optimale afweging tussen informatie en kosten. Ten tweede, is er de analyse van de data. Dat is vakgebied op zichzelf aan het worden, maar wordt snel onderschat. Een goed plan en gerichte waarneming zijn onvermijdelijk. Op dit terrein ligt er een sterke relatie met big data en data mining.

Afsluiting

Samenvattend, omvat de leerstoel een verscheidenheid aan zowel methodologische als statistische vraagstukken. Waarneming van personen is sterk in beweging en zal dat naar verwachting de komende jaren blijven. Vergelijkbaarheid is daarmee de belangrijkste uitdaging. Dat zou je inert kunnen maken, maar geen aanpassing is hetzelfde als geleidelijke onvergleichbaarheid. Maatwerk en innovatie zijn noodzakelijk. CBS en UU hebben dat ook herkend en werken al enige tijd aan innovatie binnen een gezamenlijk programma WIN (Waarneem-Innovatienetwerk). U ziet hier de verschillende onderzoekers en management betrokken bij WIN. De leerstoel is feitelijk niet mogelijk zonder de bijdragen van deze personen en het is belangrijk de leerstoel te zien als één van de gezichten die dit onderzoek heeft. Ik noem in het bijzonder de onderzoekers van universiteit Utrecht, Bella Struminskaya, Marieke Haan, Peter Lugtig en Vera Toepoel. Binnen WIN loopt een verscheidenheid aan complexe maar ook uitdagende projecten. Blijf ons daarom volgen of, beter nog, neem contact op.

Leerstoel Mixed-mode survey designs

- Personenwaarneming is sterk in beweging met veranderingen in communicatiekanalen → brugfunctie is zeer zinvol;
- Kansen en uitdagingen in:
 - Vergelijkbaarheid van statistieken (data kwaliteit);
 - Doelgroepenbenadering;
 - Devices (design, keuze van sensoren, nut van apps);
- CBS-UU Waarneem-Innovatie Netwerk (WIN)

Multi-mode en multi-device surveys, 13 februari 2018

Onderzoekers Utrecht

Onderzoekers CBS

Stuurgroep WIN

Multi-mode en multi-device surveys, 13 februari 2018

Ik kom dan tot de afsluiting. Ik wil graag Peter van der Heijden, Werner Raub, Kees Zeelenberg, Bart Bakker, Winfried Ypma, Wim van Nunspeet bedanken voor hun steun aan de leerstoel. En het WIN team voor de inspirerende samenwerking.

Tot slot: Ik wil deze oratie graag opdragen aan mijn moeder die altijd mijn grootste fan was.

Ik heb gezegd.