Onderzoekspracticum BCO
ANALYSEPLAN

Harry Ganzeboom

14 april 2005

Waar waren we?

- Probleemstelling, deelvragen, theorie
- Definities, conceptueel model
- Hypothesen
- Onderzoekzoeksopzet, operationalisatie
- Dataverzameling
- **Data-analyse**
- Rapportage

Soorten gegevens

- Documentatie, jaarverslagen, beleids-plannen, voorlichtingsmateriaal, reeds bewerkte (secundaire gegevens)
- Kwalitatieve primaire gegevens, bv.
  - Verslagen topical interviews
  - Observatieverslagen
- Kwantitatieve primaire gegevens, bv.
  - (Numerieke) gegevens voorgestructureerde interviews
  - Gecodeerde inhouden van documenten
  - Gestructureerde observaties

Documentatie

- Maak een zeer nauwkeurige lijst van verwijzingen. Dit is (nog) veel moeilijker dan een lijst van artikelen / boeken.
- Maak uitgetypte uittreksels van elke bron (kort). Neem daarin over saillante passages (maak informatie elektronisch beschikbaar).
- Houdt deze documenten bij de hand als je je onderzoeksverslag maakt.

Kwalitatieve gegevens

- Bij uitwerking van open interviews twee opties:
  - Bewerk tot documentair materiaal. Hierin laat je de geinterviewde niet zelf aan het woord, je doet verslag en analyseert tegelijkertijd.
  - Schrijf het gezegde volledig uit. Je neemt letterlijke passages op. Dit maakt het mogelijk om iets te vertellen door de mond van de onderzochten.
- In beide gevallen is het cruciaal dat je je materiaal elektronisch beschikbaar hebt. Dit is ook zo, als je je materiaal met kleurstiften op papier bewerkt.

Coderen kwalitatieve gegevens

- Er bestaan hiervoor gespecialiseerde computerprogramma's: Kwalitan, Nudist, N5/N6, Atlas/TI etc.
- Je komt ook een eind met pen en papier, en met een word processor.
- Gebruik cut/paste en Word-tabellen om materiaal te herordenen en van keywords (coderingen, rubriceringen) te voorzien.

Spoorboekje analyse kwantitatieve gegevens

- Datadocumentatie en -cleaning
- Van conceptueel model (hypothesen) naar causaal model (relatie tussen variabelen).
- Valide en betrouwbare meting van variabelen:
  - Selectie van valide en betrouwbare indicatoren
  - Constructie het meetinstrument
- Enkelvoudige (bivariate) relaties:
  - Conditionele gemiddelden
  - (Pearson) correlatie.
- Meervoudige (multipele) relaties:
  - Multiple regressie

Kwantitatieve gegevens: invoer en documentatie

- Kwantitatieve gegevens invoeren in data-sheet. Dat kan in Excel, maar het is handiger het gelijk in SPSS te doen.
- Keuze variabelennamen: correspondentie met vragenlijst.
- Label variabelen en values volledig (mag in het Engels).
- Voer alle 'open informatie' ook in (strings) en maak er kwantitatieve codes van.
- Documentatie: Vragenlijst & frequentieverdelingen van de uiteindelijke data & verslag van het veldwerk.

Cleaning

- Bekijk alle frequentieverdelingen en herstel onmogelijke scores.
- Bestudeer missing values: waar komen die vandaan? Zijn ze te repareren?
- Missing values zijn een groot probleem voor analyses – binnen de kortste keren ben je je data kwijt. Hou altijd in de gaten over hoeveel eenheden je analyse doet.

Causale modellen

- In een causaal model worden variabelen met elkaar verband gebracht via veronderstelde oorzaak- gevolg relaties. De relatie wordt weergegeven met een causale pijl: X→Y ('X leidt tot Y', 'verschillen in Y komen voort uit verschillen in X').
- Causale relatie zijn veronderstellingen die doorgaans te maken hebben met tijdsvolgorde. Ze berusten op kennis van de onderzochte situatie.
- Causale analyse gaat om het kwantificeren van de sterkte van de veronderstelde relaties. Een uitkomst kan zijn dat deze relatie 0 is, niet bestaat.

Statistische significantie (terminologie)

- SPSS vermeldt bij correlatie- en regressie-coëfficiënten de geschatte **overschrijdingskans**.
- De overschrijdingskans is de kans dat het resultaat (of extremer) in een enkelvoudig aselecte steekproef naar boven zou komen, terwijl in de populatie de nul-hypothese geldt.
- We spreken van een "statistisch significant" resultaat als deze kans **kleiner** is dan het gekozen **significantieniveau** (.05 of .10).
- Significantie is een kwestie van 'al dan niet', het is niet de sterkte van het verband.

Bivariate analyse

- De meeste eenvoudige causale relatie is die met één X en één Y:
  - X → Y
- In dit geval veronderstellen we dat variaties in X variaties in Y veroorzaken en dat die relatie niet verstoord wordt door andere variabelen.
- De meest voor de hand liggende meting van een X→Y relatie zijn (verschillen tussen) conditionele gemiddelden.
- We kunnen de sterkte van de relatie ook bepalen uit elke bivariate associatiecoëfficiënt. Het meest bruikbaar daarvoor is de pearson correlatie (r). Dit is equivalent aan lineair gemodelleerde conditionele gemiddelden.

Correlatie (en regressie)

- Kan strikt genomen alleen worden toegepast op gegevens van intervalniveau.
- Wordt oogluikend ook wel toegepast op ordinale (meer .. minder) gegevens.
- Ook goed interpreteerbaar voor dichotome gegevens (0/1 variabelen). Dit kan met alle variabelen.
- Voorbeeld van een correlatieberekening.

Multipele regressie

- Vaak bestuderen we de situatie met een Y (de afhankelijke of gevolg-variabele) en meerdere X-variabelen (de onafhankelijke of oorzaak-variabelen).
- Dit is met name noodzakelijk in het (zeer veel voorkomende geval) dat de X-variabelen onderling gecorreleerd zijn.
- De standaard statistische techniek is multipele regressie: in dit model bepaal je de sterkte van de relatie tussen elke X en de Y, terwijl de invloed van de andere X-variabelen constant wordt gehouden ('controlled').

Verwarrende statistische terminologie

LET OP: In de statistiek wordt hetzelfde woord onafhankelijkheid ('independence') voor twee verschillende dingen gebruikt:

- Onafhankelijke == oorzaak-variabelen.
- Onafhankelijke == ongecorreleerde variabelen.

Multipele regressievergelijking

- Let achtereenvolgens op de volgende dingen:
  - Multipele correlatie / verklaarde variantie
  - Statistische significantie van de vergelijking
  - Ongestandaardiseerde B-coëfficiënten
  - Gestandaardiseerde B-coëfficiënten.
  - Statistisch significantie van de afzonderlijke coëfficiënten.

Stapsgewijze regressie

- Een populaire wijze van analyse is stapsgewijze regressie: kijken hoe een regressievergelijking verandert als je geleidelijk aan meer X-variabelen ('controle-variabelen') toevoegt.
- Door modellen naast elkaar te zetten kun je zien hoe de onderlinge correlaties tussen X-variabelen doorwerkt.
- Stapsgewijze procedures zijn vaak dubieus, omdat er geen expliciet model aan ten grondslag ligt.

Complexe causale modellen (1)

- We kunnen spreken van complexe causale modellen wanneer we meerdere Y-variabelen hebben. Twee belangrijke situaties:
- ***Confounding***: een derde variabele (Z) is van invloed op zowel X als Y. De relatie tussen X en Y is niet causaal, maar een 'schijnverband' (spurious correlation).
- ***Intervening***: een derde variabele (Z) intervenieert tussen X en Y: X is van invloed op Y doordat (voorzover) X op Z van invloed is, en Z weer Y beïnvloedt.

Complexe causale modellen (2)

- Merk op dat het confounding en intervening tot radicaal verschillende interpretaties van dezelfde uitkomst leiden!
- Je kunt over de verschillende interpretaties geen beslissing nemen aan de hand van de statistische analyses, ze moeten berusten op je kennis van de onderzochte situatie en het onderzoeksdesign.

Betrouwbaarheid

- Betrouwbaarheid == stabiliteit van resultaten. Als je iets (op dezelfde manier) opnieuw onderzoekt, moet er hetzelfde uitkomen.
- Onbetrouwbaarheid ontstaat door toevallige (random) invloeden op je onderzoeksresultaten. Bronnen van zulke invloeden zijn bv.:
  - Steekproeffluctuaties (vandaar statistische significantie).
  - Fouten gemaakt bij beantwoording van vragen door vergissing van interviewers, respondenten, data-invoerders (meetbetrouwbaarheid).

Meetbetrouwbaarheid

- Analyse van meetbetrouwbaarheid kun he beschouwen als een causaal model, waarbij een (latente, ongemeten) variabele F van invloed is op meerdere (geobserveerde, gemeten) indicatoren f.
- Elke f is een (lineaire) functie van F en een random storingsterm e: f = b.F +e.
- We spreken van een betrouwbare meting als de relatie tussen F en f sterk is en de omvang (variantie) van de storingsterm e klein.

Multiple indicatoren

- Als $f_1 = b_2.F + e_1$ en $f_2 = b_2.F + e_2$ (etc.), dan:
  - Dat f_1 en f_2 moeten samenhangen
  - De samenhang is sterker naarmate de omvang van e1 en e_2 geringer is (en daardoor b_1 en b_2 dichter bij 1).
  - De som (gemiddelde) van f_1 en f_2 zal een betrouwbaarder meting van F zijn dan elk van f_1 en f_2 afzonderlijk.
- Het doel van een betrouwbaarheidsanalyse is een zodanig deelverzameling van je indicatoren te vinden dat hun gemiddelde de meest betrouwbare meting van F is.

Hoeveel indicatoren?

- Hoeveel indicatoren hebben we nodig om iemands geslacht nauwkeurig te bepalen? Drie!
- Hoeveel indicatoren hebben we nodig om de meetbetrouwbaarheid van afzonderlijke indicatoren te bepalen? Drie!

Cronbach's alfa (1)

- De formule van Cronbach's alfa geeft een schatting van de betrouwbaarheid van een somscore F en de onderlinge correlatie tussen de F_i-indicatoren:
  - Alfa = N.R/[1+R(N-1)]

  Waarin N het aantal indicatoren en r de gemiddelde correlatie tussen die indicatoren is.
- Alfa wordt groter naarmate:
  - R groter is
  - N groter is.
- Meetbetrouwbaarheid kun je dus vergroten door vaker te meten en door betere metingen te gebruiken. Meestal hebben deze twee een omgekeerde relatie met elkaar!

Cronbach's alfa (2)

- Een waarde van 0.70 wordt doorgaans als voldoende beschouwd.
- Enige karakteristieke waarden van alfa:
  - 2 items, R = .60, alfa = 0.75
  - 2 items, R = .40, alfa = 0.57
  - 4 items, R = .30, alfa = 0.65
  - 4 items, R = .40, alfa = 0.73
  - 6 items, R = .20, alfa = 0.60
  - 6 items, R = .30, alfa = 0.70

Kwalitatieve gegevens: betrouwbaarheid

- Onbetrouwbaarheid van uitkomsten is niet iets dat zich beperkt tot kwantitatieve metingen.
- Ook bij kwalitatieve analyse doet zich de vraag van betrouwbaarheid voor (misschien nog wel sterker):
  - Krijg je dezelfde conclusies als je na een tijdje de analyses opnieuw doet?
  - Krijg je dezelfde conclusies als twee onderzoekers afzonderlijk de analyses doen?

Beoordelaars-betrouwbaarheid (1)

- Een vorm van betrouwbaarheidsonderzoek die toegepast kan worden in kwalitatief onderzoek, is onderzoek naar consistentie tussen beoordelaars: zijn de uitkomsten voor iedereen (ongeveer) hetzelfde?
- Bijv. rangorden ondervraagde sleutelpersonen naar een vooraf afgesproken kenmerk. Of: laat meerdere beoordelaars passages uit vraaggesprekken volgens een conceptueel schema classificeren

Beoordelaars-betrouwbaarheid (2)

- Doe beoordelingen (A) geheel onafhankelijk van elkaar; (B) nadat je eerst hebt voorgesproken.
- Je kunt op deze manier een bepaald aspect van je kwalitatieve gegevens kwantitatief maken en er correlatie op berekenen.
- Denk aan jureringsystemen in sport e.d. (kunstrijden, turnen, Idols).