

DATA-ANALYSEPLAN

(Versie: 19-11-2006)

Inleiding

De manier waarop data georganiseerd, gecodeerd en gescoord (getallen toekennen aan observaties) worden en welke technieken daarvoor nodig zijn, dient in het ideale geval bepaald te worden voordat de gegevens worden verzameld. Als de data analyse bij het begin van het project wordt ontworpen, kan de onderzoeker namelijk precies bepalen welke gegevens nodig zijn en of ze in de geschikte vorm verzameld worden om de onderzoeksvragen te beantwoorden. Veel onderzoekers hebben er baat bij om in deze fase de hulp in te roepen van een statisticus om samen de data-analyses door te nemen, zodat de dataverzameling eventueel aangepast kan worden aan de eisen van de analyses. Dan kan bijvoorbeeld overdacht worden met welk meetniveau de variabelen verzameld moeten worden om bepaalde technieken mogelijk te maken en hoeveel onderzoekseenheden moeten worden betrokken bij het onderzoek om zinvolle analyses te kunnen doen.

Ter ondersteuning van een goede voorbereiding van de dataverzameling en analyse, kan een *data-analyseplan* een nuttige functie vervullen. De bedoeling van een data-analyseplan is een overzicht te geven van hoe de inhoudelijke gegevens geanalyseerd gaan worden, gegeven de verzamelde of te verzamelen data. Voorwaarden voor het opstellen van een dergelijk plan zijn in termen van de gemeten variabelen of kenmerken geformuleerd inhoudelijke vragen, duidelijkheid over de eenheden of personen waarop de analyses betrekking hebben en duidelijkheid over de meetniveaus van de variabelen.

Het is niet de bedoeling dat een data-analyseplan in steen is gehouwen. Het plan is bedoeld om richting te geven aan de dataverzameling en de daarbij behorende data-analyse en om aan te geven wat nodig is voor een succesvolle analyse. Lacunes in kennis en in de data kunnen hierdoor naar voren komen, zodat zij tijdig rechtgezet kunnen worden. Zo kan bijvoorbeeld voorkomen worden dat er een slechte aansluiting is tussen de vraagstellingen en de data, tussen de data en hun analyse en tussen de benodigde en de aanwezige data-analytische kennis bij de student en eventueel ook de begeleider. Voor scripties, werkstukken e.d. dienen de voorgestelde analyses namelijk te passen binnen de competentie van de student. De analyses moeten dus de reeds aanwezige kennis bij de student aanspreken of zodanig dicht liggen bij die kennis dat verwacht kan worden dat de student zich de kennis eigen kan maken die nodig is om de analyse te begrijpen, uit te voeren en tijdens een examen te verdedigen.

In dit document wordt uiteengezet welke onderdelen het data-analyseplan bevat, gevolgd door een voorbeeld van een formulier om de benodigde gegevens in te vullen.

Inhoud data-analyseplan

Een data-analyse plan dient de volgende elementen te bevatten:

1. Een weergave van persoonlijke gegevens, de naam van de begeleider, de titel van de scriptie en de onderzoeksvragen.
2. Een lijst van variabelen waarop de analyses betrekkingen hebben, alsmede hun meetniveaus. Daarbij moet aandacht besteed worden aan meetniveaus van de daadwerkelijk gemeten variabelen en die van eventuele onderliggende variabelen (theoretische constructen).
3. Een compacte beschrijving van de onderzoekseenheden waarop de analyses uitgevoerd gaan worden. Daarbij moeten zowel de aantallen voor overkoepelende analyses vermeld worden als die van eventuele deelgroepen (zeg jongens en meisjes).
4. Een beknopte beschrijving van de onderzoeksopzet.
5. Een overzicht van de datastructuur met name indien deze afwijkt van de standaard structuur van de data matrix van personen bij variabelen. Standaardvoorbeelden van datastructuren zijn: een serie vijfpunts-beoordelingsschalen aan personen voorgelegd en scores van personen op een serie variabelen. Ongebruikelijker zijn rangordeningen van voorkeuren (voor speelgoed bijvoorbeeld). Bijzonder zijn ook longitudinale gegevens waarbij personen verschillende malen gemeten worden op dezelfde variabelen.
6. Een controlelijst voor het opschonen van de gegevens. Het gaat om eventuele ontbrekende gegevens en uitbijters. Tevens dient te worden aangegeven wat hiermee gedaan zal worden. Ook dient vermeld te worden hoe de gegevens worden gecontroleerd op correctheid (bijv.. hoe de zwangere mannen in de data set zullen worden opgespoord; hoe gecontroleerd gaat worden op incorrect ingevoerde gegevens).
7. Een algemeen overzicht van de analyses die uitgevoerd gaan worden. Hierbij horen ook de procedures die gebruikt worden om de betrouwbaarheid en validiteit van de metingen vast te stellen (soms hebben anderen dit al gedaan, met name bij standaardmeetinstrumenten). De volgorde van analyses ziet er over het algemeen als volgt uit:
 - a. Vaststellen van datakwaliteit: correctheid gegevens, ontbrekende gegevens, betrouwbaarheid en validiteit van de meetinstrumenten (zie punt 6)
 - b. Beschrijvende analyses van de steekproef en de gegevens
 - c. Correlationele procedures om samenhangen te onderzoeken
 - d. Inferentiële methoden om de theoretisch gepostuleerde verbanden en effecten te onderzoeken
 - e. Aanvullende analyses, geïnspireerd door de proefopzet of de gegevens
8. Een gedetailleerde beschrijving van de analytische methoden die gebruikt gaan worden. Hierbij valt te denken aan beschrijvende technieken, toetsen en grafische presentaties. Hierbij dient tevens te worden aangegeven aan welke eisen de data dienen te voldoen om door deze technieken geanalyseerd te kunnen worden en in hoeverre hieraan is voldaan.

Ook dient aangegeven te worden welke statistische routines (uit SPSS of een ander statistisch pakket) hiervoor gebruikt gaan worden.

9. Een kort literatuuroverzicht waarbij aangegeven wordt waar de technieken die gebruikt gaan worden beschreven staan.

VOORBEELD DATA-ANALYSE PLAN

1.	Gegevens		
	Naam auteur/student		
	Naam begeleider		
	Titel document/scriptie		
	Centrale Onderzoeksvraag		
	Secundaire onderzoeksvraag		
	Secundaire onderzoeksvraag		
2	Criterium/Afhankelijke variabelen	Naam:	Meetniveau
		1.	
	<i>(vermeld ook onderliggend meetniveau, indien relevant)</i>	2.	
		3.	
		4.	
		5.	
	Criterium/Onafhankelijke variabelen	Naam:	Meetniveau
		1.	
		2.	
		3.	
		4.	
		5.	

3	Onderzoekseenheden + aantallen	
	Eenheid	Aantal
4	Onderzoeksopzet (geef details)	<i>Voorbeelden:</i> (Quasi-)Experiment Inventarisatie-onderzoek (surveyonderzoek) Correlatie-onderzoek Observatie-onderzoek Interviews
5	Datastructuren	<i>Voorbeelden:</i> Datamatrix van personen bij variabelen Personen bij variabelen bij condities Personen bij variabelen bij tijdstippen (herhaalde metingen opzet) Kruistabel(len) Correlatie/Covariantiematrix (-matrices)

6	Controlelijst voor opschonen van de gegevens.	Aanpak	Behandelinggegevens
	Codeerfouten: Alle variabelen gecontroleerd op codeerfouten		
	Ontbrekende gegevens: Alle variabelen gecontroleerd op ontbrekende gegevens (indien aanwezig, vermeld hoe behandeld en hoe te behandelen bij de analyses)		
	Normaliteit: Univariate verdelingen van metrische (continue) variabelen zijn gecontroleerd op normaliteit via histogrammen met normale verdelingen, scheefheid (scheefheid/standaardmeetfout ≤ 3 of zo; idem voor gepiekttheid (=kurtosis); heel erg precies hoeft dit niet).		
	Vorm verdelingen categorische variabelen Univariate categorische variabelen zijn gecontroleerd op de afwezigheid van categorieën met te weinig personen (zeg, < 5% van de scores). Een oplossing is hiervoor voorgesteld.		
	Univariate uitbijters: Alle variabelen gecontroleerd op uitbijters (indien gevonden, vermeld actie ten aanzien van die uitbijters)		
	Bivariate uitbijters: De centrale variabelen twee aan twee gecontroleerd op bivariate uitbijters. De overige data van de uitbijtende personen zijn gecontroleerd of er iets bijzonders aan de personen is (bijv. de enige miljonair in een groep stratenmakers).		

<p><i>Bivariate verbanden:</i> De centrale variabelen zijn gecontroleerd op de lineariteit van de verbanden. Bij gebrek aan lineariteit zijn beredeneerde beslissingen genomen over eventuele <i>schaaltransformaties</i> van de scores.</p>		
<p><i>Homogeniteit van varianties in spreidingsdiagrammen:</i> De spreidingsdiagrammen zijn gecontroleerd op overdreven niet-homogeniteit van varianties (heteroscedasticiteit) (Met name bij verder onderzoek met regressie is het noodzakelijk hiernaar te kijken.)</p>		
<p><i>Schaalconstructie:</i></p> <ul style="list-style-type: none"> - Bij het maken of gebruiken van schalen die uit diverse items bestaan is de betrouwbaarheid gemeten Cronbach's alpha bij metrische variabelen [$>.70$ doorgaans; $>.80$ is beter]; vaak kappa en percentage overeenstemming bij categorische data en overeenstemming van observatoren bij weinig categorieën; rangcorrelaties bij intersubjectiviteit met ordinale categorieën 		
<ul style="list-style-type: none"> - Met name bij de strenge noodzaak van een één-dimensionale schaal is gekeken of verwijdering van items de betrouwbaarheid (meest Cronbach's alpha) verbetert 		
<ul style="list-style-type: none"> - Er is gecontroleerd dat de correlaties tussen alle items van de schaal positief correleren. De schendende items zijn omgekeerd gecodeerd om te zien of daardoor alle items wel positief correleren, zo niet dan zijn deze items verwijderd (bij een eigen gemaakte schaal). 		

7 & 8.	Overzicht van voornaamste analysetechnieken behorende bij de hypothesen	Eisen aan de data	Statistische routine, b.v. in SPSS
	<i>Descriptieve analyses om overzicht te krijgen van de data</i>		
	- <i>Univariate</i>		
	-		
	- <i>Bivariaat</i>		
	-		
	- <i>Multivariaat</i>		
	<i>Analyseprocedures voor centrale hypothesen</i>		
	- 1.		
	- 2.		
	<i>Analyseprocedures voor subhypothesen</i>		
	- 1.		
	- 2.		
	- 3.		
	- 4.		

9	Literatuurlijst voor beschrijving technieken
	- ...
	- ...

Samenvattende tabel van de belangrijkste data-analytische activiteiten

Onderzoeksvraag	Onderzoekopzet gebruikt voor dataverzameling (Basisvorm, e.g. experiment, vragenlijst, observatie)	Variabelen	Status van de variabelen (criteriumvariabele, predictorvariabele, etc.)	Meetniveaus	Analyseprocedure en methode van presentatie van resultaten	Statistische toets of analysemethode (voor zover relevant)
1.						
2.						
3.						
Etc.						

