

Hermann Dülmer

76.1 Einleitung

Beim faktoriellen Survey (Vignettenanalyse) handelt es sich um ein experimentelles Design (Eifler/Leitgöb, Kapitel 13 in diesem Band), bei dem der Forscher variierende Situations- oder Personenbeschreibungen, die sogenannten Vignetten, zu einem Thema erstellt und unter einem bestimmten Gesichtspunkt beurteilen lässt. Die vom Forscher festgelegten Situations- oder Personenbeschreibungen, wie etwa das Geschlecht, die Hautfarbe, die Deutschkenntnisse, die Konfession und die Erwerbstätigkeit einer fiktiven Person bilden die unabhängigen Variablen (X-Variablen, experimentelle Einwirkungen) für das zu erklärende Urteilsverhalten (Y-Variable, Messung), also etwa dafür, wie sehr ein Befragter der jeweils beschriebenen Vignettenperson vertrauen würde. Beurteilt jeder Befragte mehr Vignetten als es X-Variablen für die Vignetten gibt, dann lässt sich für jeden einzelnen Befragten die Wichtigkeit, die er oder sie den entsprechenden Vignettenmerkmalen bei der Beurteilung der Vignetten beimisst, mit Hilfe der Regressionsanalyse schätzen. Werden die Vignetten von vielen Befragten beurteilt, dann lässt sich über alle Befragten hinweg auch der Einfluss schätzen, den bestimmte Befragtenmerkmale als weitere unabhängige Variablen auf das Urteilsverhalten (Y-Variable) ausüben. Damit lassen sich dann beispielsweise Fragen danach beantworten, ob es Unterschiede zwischen weiblichen und männlichen Befragten hinsichtlich des Vertrauens in die auf den Vignetten beschriebenen Personen gibt.

Ein Vorteil der Beurteilung konkreter Vignettenbeschreibungen besteht darin, dass dies unserem Urteilsverhalten im alltäglichen Leben erheblich näher kommt als die Beantwortung allgemeiner, häufig abstrakter oder aus dem Kontext herausgelöster Fragen. Dies dürfte vor allem der Gültigkeit (Krebs/Menold, Kapitel 35 in diesem Band) der Ergebnisse zugutekommen (Alexander/Becker 1978: 93). Da die Si-

tuations- oder Personenbeschreibungen als experimentelle Einwirkungen (Stimuli) gezielt vom Forscher variiert werden, jede Vignette quasi als „Kontrollgruppe“ für die anderen Vignetten dient und die Befragten den Fragebögen per Zufall zugewiesen werden (bzw. alle dieselben vom Forscher ausgewählten Vignetten beurteilen), so dass die Vignettenmerkmale unabhängig von den Merkmalen der Befragten sind, erlauben faktorielle Surveys eine *kausale* Erklärung (Kühnel/Dingelstedt, Kapitel 46 in diesem Band) des beobachteten Urteilsverhaltens. Die Variation der Vignettenmerkmale ist somit ursächlich für die beobachtete Änderung des Urteilsverhaltens der Befragten verantwortlich (Taylor 2006: 1196).

76.2 Der Aufbau der Vignetten: Faktoren, Merkmalsausprägungen und Antwortskala

Der faktorielle Survey lässt sich für die Untersuchung ganz unterschiedlicher Fragestellungen verwenden: In der Literatur finden sich etwa faktorielle Surveys zur sozialen Mindestsicherung (Liebig/Mau 2002), zur Gültigkeit von Normen (Beck/Opp 2001) oder zur Einwanderung (Jasso 1988). Zur Veranschaulichung des faktoriellen Designs ist in Abbildung 76.1 neben dem Einleitungstext eine Beispielvignette zum Vertrauen in andere Menschen (Bartolomé et al. 2021) wiedergegeben. Die Vignetten können, wie im verwendeten Beispiel, eher in tabellarischer Form oder als Kurzgeschichte präsentiert werden. Bei Kurzgeschichten werden die variierenden Größen häufig durch Unterstreichungen hervorgehoben (Jasso 2006: 412). Beim Vergleich beider Präsentationsformen zeigt sich, dass insbesondere weniger gebildete Personen von der größeren Übersichtlichkeit von Tabellenformaten profitieren, wodurch sich der Anteil fehlender bzw. ungültiger Antworten im Vergleich zu Textvignetten verringern lässt (Shamon et al. 2022).

76.2.1 Faktoren und Merkmalsausprägungen

Eine Hauptaufgabe bei der Erstellung eines faktoriellen Surveys besteht darin festzulegen, welche *Merkmale* (beim faktoriellen Survey als *Faktoren* oder *Dimensionen* bezeichnet) einen bedeutsamen Einfluss auf das Urteilsverhalten ausüben. Bei dieser Aufgabe kann auf existierende Theorien, Forschungsergebnisse und konventionelles Alltagswissen zurückgegriffen werden (Jasso 2006: 342). Anschließend ist eine Entscheidung über die einzubeziehenden *Merkmalsausprägungen* (*Level*) notwendig. Während die Anzahl der Merkmalsausprägungen bei vielen qualitativen Variablen wie dem Geschlecht gegeben ist, stellt sich bei quantitativen Variablen wie dem Alter die Frage, wie viele und welche Merkmalsausprägungen für den faktoriellen Survey verwendet werden sollen. Für die Schätzung einer linearen Beziehung reicht es aus statistischer Sicht aus, zwei Merkmalsausprägungen auszuwählen, die die Spannweite

Auf den folgenden Seiten werden wir Ihnen 20 ähnliche Situationen vorstellen, in denen eine Person ihr offenes Handgepäck im Zugabteil zurückgelassen hat, um zur Toilette zu gehen. Im Handgepäck befindet sich jeweils eine Brieftasche mit **etwa 100 Euro** und das **Handy**.

Bevor die Person das Zugabteil verließ, befand sich dort eine einzige weitere Person. Diese andere Person ist jeweils **35 Jahre alt**, **ledig** und besitzt **keine deutsche Staatsbürgerschaft**, unterscheidet sich aber in ihren übrigen Merkmalen von Situation zu Situation. Für jede der 20 Situationen möchten wir Sie bitten anzukreuzen, wie wahrscheinlich es Ihrer Meinung nach ist, dass das **Eigentum** der Person während ihres Toilettenganges von der anderen Person **gestohlen wurde**. Benutzen Sie dabei jeweils bitte die folgende Antwortskala:

Es wurde ...

vollkommen sicher nicht gestohlen	<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr> <td style="width: 20px; text-align: center;">0</td> <td style="width: 20px; text-align: center;">1</td> <td style="width: 20px; text-align: center;">2</td> <td style="width: 20px; text-align: center;">3</td> <td style="width: 20px; text-align: center;">4</td> <td style="width: 20px; text-align: center;">5</td> <td style="width: 20px; text-align: center;">6</td> <td style="width: 20px; text-align: center;">7</td> <td style="width: 20px; text-align: center;">8</td> <td style="width: 20px; text-align: center;">9</td> <td style="width: 20px; text-align: center;">10</td> </tr> </table>	0	1	2	3	4	5	6	7	8	9	10	vollkommen sicher gestohlen
0	1	2	3	4	5	6	7	8	9	10			

1 Frau J. befindet sich in einem Zug, der gerade in den nächsten Bahnhof einfährt. Sie kommt von der Toilette als ihr einfällt, dass sie ihr unverschlossenes Handgepäck im Zugabteil zurückgelassen hat. In dem Handgepäck befinden sich die Brieftasche mit **etwa 100 Euro** und das **Handy**.
Bevor sie das geschlossene Zugabteil verließ, befand sich dort nur eine weitere, **35jährige, ledige Person ohne deutsche Staatsbürgerschaft** mit folgenden weiteren Merkmalen:

Geschlecht:	weiblich
Hautfarbe:	weiß
Deutschkenntnisse:	spricht gebrochenes Deutsch
Konfession:	muslimisch
Erwerbstätigkeit:	Geschäftsführerin eines mittelständischen Unternehmens

Wie wahrscheinlich ist es Ihrer Meinung nach, dass das Eigentum von Frau J. während ihrer Abwesenheit von der beschriebenen Person gestohlen wurde? Es wurde ...

vollkommen sicher nicht gestohlen	<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr> <td style="width: 20px; text-align: center;">0</td> <td style="width: 20px; text-align: center;">1</td> <td style="width: 20px; text-align: center;">2</td> <td style="width: 20px; text-align: center;">3</td> <td style="width: 20px; text-align: center;">4</td> <td style="width: 20px; text-align: center;">5</td> <td style="width: 20px; text-align: center;">6</td> <td style="width: 20px; text-align: center;">7</td> <td style="width: 20px; text-align: center;">8</td> <td style="width: 20px; text-align: center;">9</td> <td style="width: 20px; text-align: center;">10</td> </tr> </table>	0	1	2	3	4	5	6	7	8	9	10	vollkommen sicher gestohlen
0	1	2	3	4	5	6	7	8	9	10			

Abb. 76.1 Vignettenbeispiel zum Vertrauen in andere Menschen (Bartolomé et al. 2021: 93, deutsche Fassung)

der Variable möglichst realitätsgetreu abdecken. Wird eine u-förmige Beziehung erwartet, so sind mindestens drei Merkmalsausprägungen erforderlich.

Die Festlegung der Anzahl der Faktoren und deren Merkmalsausprägungen entscheiden über die Größe des Vignettenuniversums, das sich aus der Anzahl aller möglichen Vignetten zusammensetzt. Für das oben angeführte Beispiel zum Vertrauen wurden die folgenden Merkmalsausprägungen in die Untersuchung einbezogen (Tab. 76.1).

Das vorliegende Beispiel setzt sich aus fünf Faktoren zusammen. Weitere Informationen über die Wertgegenstände im Handgepäck sowie das Alter, der Familienstand und die Staatsbürgerschaft der Vignettenperson wurden im vorliegenden Beispiel konstant gehalten, da deren Einfluss für die Forschungsfragen von nachrangiger Bedeutung war. Von den fünf Faktoren weisen vier jeweils zwei Merkmalsausprägungen auf. Bei der Konfession waren fünf Merkmalsausprägungen von Interesse. Das vollständige Vignettenuniversum besteht somit aus insgesamt 80 Vignetten ($2 \times 2 \times 2 \times 2 \times 5 = 80$ Merkmalskombinationen).

Bei der Auswahl der Faktoren und der Merkmalsausprägungen sollte darauf geachtet werden, dass mit zunehmender Anzahl von Faktoren die Komplexität der Aufgabenstellung für die Befragten stark ansteigt (Beck/Opp 2001: 287). Folgt man einer Empfehlung von Auspurg und Hinz (2015: 22), dann ist es angeraten, nicht mehr als sieben (plus/minus zwei) Faktoren mit jeweils zwei oder drei Merkmalsausprägungen in den faktoriellen Survey einzuschließen. Werden weniger Faktoren einbezogen, dann können mehr Merkmalsausprägungen berücksichtigt werden und umgekehrt. Anzumerken ist, dass Faktoren mit mehr Merkmalsausprägungen im Vergleich zu

Tab. 76.1 Merkmale und Merkmalsausprägungen des faktoriellen Survey zum Vertrauen

Merkmale (Faktoren/Dimensionen):	Merkmalsausprägungen (Level):
Geschlecht:	männlich weiblich
Hautfarbe:	schwarz weiß
Deutschkenntnisse:	spricht gebrochenes Deutsch spricht fließend Deutsch
Konfession:	keine evangelisch katholisch orthodox muslimisch
Erwerbstätigkeit:	arbeitslose/r Verkäufer/in Geschäftsführer/in eines mittelständischen Unternehmens

Faktoren mit weniger Merkmalsausprägungen häufig auch die größere Aufmerksamkeit auf sich ziehen, was ihren Einfluss auf das Antwortverhalten vergrößern mag (Auspurg/Hinz 2015: 20).

76.2.2 Die Antwortskala

Das Antwortverhalten der Untersuchungsteilnehmer wird beim faktoriellen Survey zumeist auf einer Rating-Skala (Franzen, Kapitel 74 in diesem Band) erfasst, wobei empirisch 11er-Skalen vorherrschen (Wallander 2009: 511). Werden jedem Befragten mehr Vignetten zur Beurteilung vorgelegt als es vorgegebene Antwortkategorien gibt, bei einer 11er-Skala also 12 oder mehr, dann müssen zwangsläufig mindestens zwei Vignetten gleich beurteilt werden. Dies würde bei differenzierterem Urteilsvermögen zwangsläufig mit einem Informationsverlust einhergehen, da die entsprechenden Unterschiede auf der Antwortskala nicht mehr erfasst werden könnten (Lodge 1981: 5). Die Vorgabe von einer zu hohen Anzahl von Antwortkategorien mag Befragte mit weniger differenziertem Urteilsvermögen hingegen überfordern. Als Alternative zur einfachen Rating-Skala kann die Magnitude-Messung verwendet werden, bei der die Befragten ihr Urteil im Vergleich zu einem zuvor festgelegten Referenzwert ausdrücken (dazu weiterführend Lodge 1981: 7, Jasso 2006: 345–346). Auspurg und Hinz (2015: 69) raten jedoch von der Verwendung der Magnitude-Messung ab, da sie zu einem recht hohen Anteil fehlender Antworten führe. In manchen Untersuchungen sind die Befragten schließlich aufgefordert, einen absoluten Betrag zu nennen (etwa den gerechten Verdienst einer beschriebenen Vignettenperson) oder die Wahrscheinlichkeit eines Ereignisses anzugeben (etwa die angenommene Wahrscheinlichkeit einer Scheidung, Jasso 2006: 344).

76.3 Die Auswahl des experimentellen Designs: Quoten- und Zufallsdesigns

Mit der Anzahl der Faktoren und der Merkmalsausprägungen steigt die Größe des Vignettenuniversums, so dass es für den einzelnen Befragten sehr schnell unmöglich wird, alle Vignetten zu beurteilen. Beim faktoriellen Survey hat jeder Untersuchungsteilnehmer daher in der Regel nur eine Stichprobe des Vignettenuniversums (Vignettenset) zu bearbeiten. In repräsentativen Umfragen sollten keinem Befragten mehr als 20 Vignetten vorgelegt werden (Sauer et al. 2011: 98). Um systematischen Ordnungseffekten (serielle Korrelationen, die dadurch entstehen können, dass zuvor beurteilte Vignetten einen Einfluss auf die Beurteilung der folgenden Vignetten ausüben) bei der Schätzung der Regressionskoeffizienten vorzubeugen, ist es angeraten, die Vignetten jedes Vignettensets in zufälliger Anordnung beurteilen zu lassen (Jasso 2006: 343). Logisch unmögliche Kombinationen, wie sie bei der Einbeziehung

von Bildung und Beruf auftreten können (etwa Physiker ohne Hauptschulabschluss), sind auszuschließen.

Zur Erzeugung der Vignettenstichproben existieren zwei grundsätzlich unterschiedliche Ansätze: dies sind zum einen Zufallsdesigns und zum anderen Quotendesigns (Dülmer 2007: 383). Die Grundidee für die Verwendung eines einfachen *Zufallsdesigns* besteht darin, das Vignettenuniversum möglichst genau durch verschiedene Vignettenstichproben gleicher Größe zu repräsentieren. Hierzu wird aus dem Vignettenuniversum für jeden einzelnen Befragten eine eigene Zufallsstichprobe (Häder/Häder, Kapitel 27 in diesem Band) gleichen Umfangs gezogen (Rossi/Anderson 1982: 40–41, Jasso 2006: 343). Die Ziehung jedes einzelnen Vignettensets erfolgt mit oder ohne Ersetzen bereits gezogener Vignetten, wobei das Ziehen ohne Ersetzen den Vorteil hat, dass keinem Befragten eine Vignette wiederholt vorgelegt wird (Dülmer 2007: 384). Vor der Erhebung ist für jede Vignettenstichprobe zu kontrollieren, ob es eine Vignettenvariable gibt, die zufallsbedingt nur eine Merkmalsausprägung aufweist (Konstante), und ob es eine Vignettenvariable gibt, die sich innerhalb eines Vignettensets perfekt durch die Vignettenvariablen anderer Faktoren vorhersagen lässt (Linearkombination). Ohne die Ersetzung solcher Vignettenstichproben lässt sich für die betreffenden Befragten nicht schätzen, welchen Einfluss die Variable auf das Urteilsverhalten dieser Person besitzt.

Alternativ wird beim *Quotendesign* (Akremi, Häder/Häder, Kapitel 26 und 27 in diesem Band) versucht, die Merkmale des Vignettenuniversums durch die Konstruktion eines oder vergleichsweise weniger Vignettensets abzubilden. Das Vignettenuniversum zeichnet sich durch die beiden Merkmale aus, dass die Faktoren untereinander unkorreliert sind (*orthogonal*) und dass die Merkmalsausprägungen innerhalb eines Faktors gleich häufig auftreten (*balanciert*). Bei den Quotendesigns lassen sich die *klassischen faktoriellen Designs* (fraktionelle faktorielle Designs und konfundierte faktorielle Designs, Gunst/Mason 1991, Kirk 1995, Steiner/Atzmüller 2006: 132–133) von den *D-effizienten Designs* (D-effiziente Designs und konfundierte D-effiziente Designs, Kuhfeld et al. 1994, Dülmer 2007: 387, 2016: 316–318) unterscheiden.

Bei den klassischen faktoriellen Designs sind die Faktoren zwar immer orthogonal, die Merkmalsausprägungen innerhalb der Faktoren aber nicht immer balanciert. Da die Faktoren orthogonal sind, lassen sich die Effekte aller Vignettenmerkmale unabhängig voneinander schätzen. Der Preis für die Verwendung eines *fraktionellen faktoriellen Designs* statt des vollständigen Vignettenuniversums besteht darin, dass bestimmte Vignettenvariablen innerhalb des verwendeten Vignettensets perfekt mit Wechselwirkungen (Interaktionen) zwischen verschiedenen Vignettenvariablen konfundiert sind (Aliasing, Gunst/Mason 1991: 40–41, Steiner/Atzmüller 2006: 126). Perfektes Aliasing zwischen der Vignettenvariable X_1 und der Wechselwirkung zwischen den Vignettenvariablen X_2 , X_3 und X_4 bedeutet beispielsweise, dass die Korrelation zwischen der Vignettenvariable X_1 und dem Interaktionsterm $X_2 \times X_3 \times X_4$ innerhalb des verwendeten Vignettensets 1 (oder aber -1) beträgt, so dass sich deren Effekte statistisch nicht mehr trennen lassen. Um eine systematische Verzerrung (Bias) des

geschätzten Regressionskoeffizienten der Vignettenvariable ausschließen zu können, ist Aliasing daher nur zulässig, wenn der Einfluss der Wechselwirkung im Vergleich zum Einfluss der Vignettenvariable (Haupteffekt) vernachlässigbar ist (Gunst/Mason 1991: 41–42). Beim *konfundierten faktoriellen Design* werden verschiedene fraktionale faktorielle Designs für die Erhebung verwendet. Hierdurch wird eine Konfundierung über verschiedene Vignettensets hinweg möglich: Liegt etwa zwischen einer Vignettenvariable und einer Wechselwirkung ein perfektes Aliasing vor und ist die hierdurch verursachte perfekte Korrelation im ersten fraktionellen faktoriellen Design positiv und im zweiten negativ, dann bietet das konfundierte faktorielle Design die Möglichkeit, zumindest die Größe der Wechselwirkung mit Hilfe einer sogenannte Set-Variable über die beiden fraktionellen faktoriellen Designs hinweg abzuschätzen.

Für manche Fragestellungen mag innerhalb einer vertretbaren Setgröße jedoch kein faktorielles Design existieren. Da die Erzeugung *fraktioneller faktorieller Designs* auf Teilbarkeitsregeln beruht, ist das Finden eines faktoriellen Designs innerhalb einer zumutbaren Setgröße immer dann besonders schwierig bzw. manchmal unmöglich, wenn die Anzahl der Merkmalsausprägungen der Faktoren auf unterschiedlichen Primzahlen basieren (also etwa 2, 3, 5, Gunst/Mason 1991: 51–52). D-effiziente Designs erlauben zumindest leichte Korrelationen zwischen den Faktoren, wobei dann quasi im Gegenzug angestrebt wird, dass die Vignettenmerkmale innerhalb eines Faktors gleich häufig auftreten (Kuhfeld et al. 1994: 545). Im Rahmen dieser Möglichkeiten lässt sich für jede gewünschte Setgröße ein D-effizientes Design finden.

Die D-Effizienz ist ein Maß für die relative Güte eines ausgewählten Designs im Vergleich zu einem balancierten, orthogonalen Design. Die Berechnung der D-Effizienz bezieht sich immer nur auf die Vignettenvariablen (und gegebenenfalls Wechselwirkungen), deren Effekte später zu schätzen sind. Werden ausschließlich qualitative Variablen verwendet, dann ist die D-Effizienz auf Werte im Bereich von 0 bis 100 beschränkt (Kuhfeld et al. 1994: 547, 549). Der Maximalwert von 100 bedeutet, dass das ausgewählte Design balanciert und orthogonal ist. Da eine größere Anzahl von Vignetten pro Vignettenset nicht unbedingt zu einer höheren oder zumindest gleichhohen D-Effizienz führt (Dülmer 2007: 394), ist ein Vergleich über verschiedene Setgrößen hinweg angeraten. Werden quantitative Variablen einbezogen, dann kann die D-Effizienz auch Werte größer als 100 annehmen. Ein Designvergleich bleibt aber auch in diesem Fall möglich.

Im Unterschied zum einfachen Zufallsdesign führt das Zusammenfügen einfacher D-effizienter Designs, die Untersuchungsteilnehmern zur Beurteilung vorgelegt wurden, weder zu einer Erhöhung der Balance noch zu einer Reduktion der Korrelationen zwischen den Faktoren. Beim *konfundierten D-effizienten Design* werden Unbalanciertheit und Korrelationen über verschiedene einfache D-effiziente Designs hinweg ausgeglichen bzw. beseitigt (Dülmer 2016: 318). Die D-Effizienz des konfundierten D-effizienten Designs ist somit höher als die des einfachen D-effizien-

ten Designs. Werden alle verwendeten D-effizienten Designs gleich häufig bearbeitet, dann ist das zusammengefügte konfundierte D-effiziente Design balanciert und orthogonal.

Die Erzeugung D-effizienter Designs erfolgt über spezialisierte Computerprogramme (SAS), wobei die D-Effizienz automatisch ausgegeben wird. Durch das Permutieren (Vertauschen) der Zuweisung von Merkmalsausprägungen der Vignetten zu den entsprechenden Variablenausprägungen des gewählten D-effizienten Designs lässt sich ein konfundiertes D-effizientes Design erzeugen (Dülmer 2016). Zur Erzeugung relativ einfacher fraktioneller faktorieller Designs werden zumeist Computerprogramme (SPSS, SAS) oder vorgefertigte Pläne aus der Literatur verwendet (etwa Gunst/Mason 1991, Kirk 1995: 549–654). Die Erzeugung komplexer konfundierter faktorieller Designs kann häufig nur über die Veränderung vorhandener, aber nicht zufriedenstellender Designs erfolgen (Steiner/Atzmüller 2006: 137). Diese Arbeit ist jedoch im Allgemeinen sehr anspruchsvoll und zeitaufwendig.

76.4 Die Auswertung faktorieller Surveys

Bei den allermeisten faktoriellen Surveys wird der Einfluss von Vignetten- und Befragtenmerkmalen auf die Beurteilung der Vignetten analysiert. Wird mehr als eine Vignette von jedem Befragten beurteilt, ist das Antwortverhalten in den Kontext eines jeden Untersuchungsteilnehmers eingebettet. Für solche hierarchisch strukturierten Daten empfiehlt sich die Auswertung über die Mehrebenenanalyse (Pötschke, Kapitel 53 in diesem Band). Aus Gründen der Übersichtlichkeit ist im Folgenden die Mehrebenengleichung für eine Vignettenvariable X und eine Variable Z für ein Befragtenmerkmal (keine Wechselwirkung zwischen X -Variable und Z -Variable) wiedergegeben. Da der Einfluss der Set-Variablen eines konfundierten faktoriellen Designs nur über die verschiedenen fraktionellen faktorielle Designs hinweg schätzbar ist, sind die Set-Variablen wie Z -Variablen auf der Befragtenebene in die Mehrebenenanalyse einzubeziehen.

Vignettenebene (Level 1): $Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + r_{ij}$

Befragtenebene (Level 2): $\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j}$
 $\beta_{1j} = \gamma_{10} + u_{1j}$

In der Zweiebenen-Regressionsgleichung steht „ i “ für die jeweilige Vignette, „ j “ für einen jeweiligen Befragten, „ Y “ für die Beurteilung einer jeweiligen Vignette, „ β “ für die befragtenspezifischen b-Koeffizienten, „ γ “ für den durchschnittlichen b-Koeffizienten über alle Befragten hinweg (Grand Mean), „ u “ für die Residuen (Zufallskomponenten) auf der Befragtenebene und „ r “ für das Residuum auf der Vignettenebene. Der Einschluss einer Zufallskomponente bedeutet, dass der b-Koeffizient eines jeden

Befragten durch einen befragtenspezifischen u -Term vom jeweiligen Grand Mean abweicht. Die statistische Mindestanforderung für die simultane Einbeziehung einer eigenen Zufallskomponente für jede Vignettenvariable (einschließlich des Intercepts und eingeschlossener Wechselwirkungen) besteht folglich darin, dass die gewählte Setgröße mehr Vignetten umfasst als b -Koeffizienten auf der Vignettenebene zu schätzen sind. Ob die Varianz eines u -Terms (Varianzkomponente) signifikant wird oder nicht, ist eine empirische Frage (Snijders/Bosker 2012: 43–44). Je weniger Varianzkomponenten signifikant sind, desto homogener ist das Antwortverhalten der Befragten. Erweisen sich alle Varianzkomponenten als nicht signifikant und können damit alle u -Terme aus der Gleichung entfernt werden, wird aus der Mehrebenenregression eine einfache OLS-Regression. In Abhängigkeit von der Heterogenität des Antwortverhaltens beruhen die geschätzten b -Koeffizienten also stärker auf den Schätzern der einzelnen Vignettensets oder auf den Schätzern der kombinierten Vignettenstichprobe.

76.5 Abschließende Bemerkungen

Die Frage nach einem möglichst optimalen experimentellen Design ist beim faktoriellen Survey eng mit der *Heterogenität* des Antwortverhaltens der Befragten verknüpft. Ein optimales Design erlaubt eine effiziente (also möglichst präzise und damit reliable sowie möglichst unverzerrte und damit valide) Schätzung aller Regressionskoeffizienten. Der Unterschied zwischen einem einfachen Quotendesign (unbalanciertes fraktionelles faktorielles oder D -effizientes Design) und einem Zufallsdesign besteht darin, dass durch das Zusammenfügen der verschiedenen Vignettensets eines Zufallsdesigns sich die D -Effizienz erhöht: Die Gesamtstichprobe der von allen Befragten bearbeiteten Vignetten erreicht somit eine größere Balance und Orthogonalität (D -Effizienz) als die einzelnen Vignettensets. Beim einfachen Quotendesign hat jedes Vignettenset hingegen die gleiche D -Effizienz wie die Gesamtstichprobe aller bearbeiteten Vignetten. Bei homogenem Antwortverhalten der Befragten werden die b -Koeffizienten mit Hilfe der OLS-Regression auf der Grundlage der Gesamtstichprobe aller bearbeiteten Vignetten geschätzt. Unter dieser Bedingung weist das Zufallsdesign mit zunehmender Anzahl der Befragten sehr schnell eine höhere D -Effizienz als ein unbalanciertes fraktionelles faktorielles Design und ein D -effizientes Design auf. Dies gilt insbesondere für kleine Setgrößen (Dülmer 2007). Bei sehr heterogenem Antwortverhalten verhält es sich umgekehrt: Das durchschnittliche Vignettenset einer Zufallsstichprobe weist eine geringere D -Effizienz auf als das Vignettenset eines D -effizienten Designs.

Bei konfundierten Quotendesigns werden verschiedene Vignettensets mit gleicher D -Effizienz eingesetzt. Wie beim Zufallsdesign erhöht sich mit dem Zusammenfügen der verschiedenen Vignettensets somit auch die D -Effizienz für die Gesamtstichprobe der von allen Befragten bearbeiteten Vignetten. Damit gewinnen

konfundierte Quotendesigns eine gewisse Unabhängigkeit von der Heterogenität des Antwortverhaltens der Befragten: Bei sehr homogenen Antworten weisen sie eine höhere D-Effizienz als das D-effiziente und das unbalancierte fraktionelle faktorielle Design auf. Bei sehr heterogenem Antwortverhalten ändert sich hingegen nichts: Das konfundierte Quotendesign besitzt in diesen Fällen eine ebenso hohe D-Effizienz wie das entsprechende unbalancierte fraktionelle faktorielle bzw. D-effiziente Design und eine höhere D-Effizienz als das Zufallsdesign. Die einzigen Designs, die die maximale D-Effizienz von 100 vollkommen unabhängig von der Heterogenität des Antwortverhaltens erreichen, sind das balancierte fraktionelle faktorielle und das balancierte konfundierte faktorielle Design. Insofern handelt es sich hierbei um ideale Designs.

Für komplexe Fragestellungen und/oder für faktorielle Surveys mit unterschiedlicher Anzahl von Merkmalsausprägungen existiert innerhalb einer zumutbaren Setgröße jedoch nur selten ein konfundiertes faktorielles Design (Steiner/Atzmüller 2006: 137, 144). Unter diesen Umständen bietet das konfundierte D-effiziente Design eine attraktive Alternative. Da die Erzeugung konfundierter D-effizienter Designs bei einer hohen Anzahl von Vignettenvariablen und/oder Merkmalsausprägungen jedoch zeitaufwendig ist, mag man unter diesen Umständen weiterhin einfache Zufallsdesigns verwenden. Sind Vignetten aufgrund logisch unmöglicher Merkmalskombinationen auszuschließen, dann lassen sich konfundierte Quotendesigns nicht anwenden. Fraktionelle faktorielle Designs und einfache D-effiziente Designs lassen sich schließlich verwenden, wenn, wie häufig in repräsentativen Umfragen, kleine Vignettensets benötigt werden und das Antwortverhalten sehr heterogen ist (Dülmer 2007: 406). D-effiziente Designs lassen sich auch noch beim Ausschluss von logisch unmöglichen Merkmalskombinationen einsetzen (Kuhfeld et al. 1994: 551). Im Unterschied zu den anderen Designs lässt sich bei einfachen Quotendesigns im Nachhinein allerdings in aller Regel nicht mehr prüfen, ob es nicht vernachlässigbar große Wechselwirkungen (Interaktionen) zwischen Vignettenvariablen gibt, die bei der Konstruktion des verwendeten Vignettensets nicht berücksichtigt wurden (Dülmer 2016: 323).

Literatur

- Alexander, Cheryl S./Becker, Henry Jay (1978): The Use of Vignettes in Survey Research. In: *Public Opinion Quarterly* 41: 93–104
- Auspurg, Katrin/Hinz, Thomas (2015): *Factorial Survey Experiments. Series: Quantitative Applications in the Social Sciences*. Thousand Oaks, CA: Sage
- Bartolomé, Edurne/Dülmer, Hermann/Coromina, Lluís (2021): Social Sustainability and Social (Dis)trust in Outgroups: Evidence from Germany and Spain Using the Factorial Survey. In: *Cuadernos Europeos de Deusto* 64: 81–109
- Beck, Michael/Opp, Karl-Dieter (2001): Der faktorielle Survey und die Messung von Normen. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 53: 283–306
- Dülmer, Hermann (2016): The Factorial Survey: Design Selection and its Impact on Reliability and Internal Validity. In: *Sociological Methods & Research* 45: 304–347
- Dülmer, Hermann (2007): Experimental Plans in Factorial Surveys: Random or Quota Design? In: *Sociological Methods & Research* 35: 382–409
- Gunst, Richard/Mason, Robert L. (1991): How to Construct Fractional Factorial Experiments. Volume 14. *The Basic References in Quality Control: Statistical Techniques*. Milwaukee, WI: ASQC Quality Press
- Jasso, Guillermina (2006): Factorial Survey Methods for Studying Beliefs and Judgments. In: *Sociological Methods & Research* 34: 334–423
- Jasso, Guillermina (1988): Whom Shall We Welcome? Elite Judgments of the Criteria for the Selection of Immigrants. In: *American Sociological Review* 53: 919–932
- Kirk, Roger E. (1995): *Experimental Design: Procedures for the Behavioral Sciences*. 3. Auflage. Pacific Grove: Brooks/Cole Publishing
- Kuhfeld, Warren F./Tobias, Randall D./Garrat, Mark (1994): Efficient Experimental Design with Marketing Research Applications. In: *Journal of Marketing Research* 21: 545–557
- Liebig, Stefan/Mau, Steffen (2002): Einstellungen zur sozialen Mindestsicherung. Ein Vorschlag zur differenzierten Erfassung normativer Urteile. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 54: 109–134
- Lodge, Milton (1981): *Magnitude Scaling. Quantitative Measurement of Opinions*. Beverly Hills, CA: Sage
- Rossi, Peter H./Anderson, Andy B. (1982): The Factorial Survey Approach: An Introduction. In: Rossi, Peter H./Nock, Steven L. (Hg.): *Measuring Social Judgments: The Factorial Survey Approach*. Beverly Hills, CA: Sage, 15–67
- Sauer, Carsten/Auspurg, Katrin/Hinz, Thomas/Liebig, Stefan (2011): The Application of Factorial Surveys in General Population Samples: The Effect of Respondent Age and Education on Response Times and Response Consistency. In: *Survey Research Methods* 5: 89–102

- Shamon, Hawal/Dülmer, Hermann/Giza, Adam (2022): The Factorial Survey: The Impact of the Presentation and of the Answer Format of Vignettes on Answer Behavior and Processing Time. In: *Sociological Methods & Research* 51: 396–438
- Snijders, Tom A. B./Bosker, Roel J. (2012): *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Analysis*. 2. Auflage. Los Angeles: Sage
- Steiner, Peter M./Atzmüller, Christiane (2006): Experimentelle Vignettendesigns in faktoriellen Surveys. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 58: 117–146
- Taylor, Brian J. (2006): Factorial Surveys: Using Vignettes to Study Professional Judgement. In: *British Journal of Social Work* 36: 1187–1207
- Wallander, Lisa (2009): 25 Years of Factorial Surveys in Sociology: A Review. In: *Social Science Research* 38: 505–520

Hermann Dülmer ist Privatdozent an der Universität zu Köln. *Ausgewählte Publikationen:* Measurement Nonequivalence Using Multilevel Structural Equation Modeling: The Case of Attitudes Toward Citizenship Rights, in: *Sociological Methods & Research* 47, 4 (zusammen mit Eldad Davidov, Jan Ciecuch, Anabel Kunz, Daniel Seddig und Peter Schmidt, 2018); The Factorial Survey: Design Selection and its Impact on Reliability and Internal Validity, in: *Sociological Methods & Research* 45, 2 (2016); Experimental Plans in Factorial Surveys: Random or Quota Design?, in: *Sociological Methods & Research* 35, 3 (2007). *Webseite:* www.iss-wiso.uni-koeln.de/de/institut/personen/d/hermann-duelmer. *Kontaktadresse:* hduelmer@uni-koeln.de.