

Kapitel 7. Vorgehensweise bei der Auswertung

A. Verwendetes Statistikprogramm

Für die empirische Auswertung wurde das Statistikprogramm SPSS Statistics, Version 18 verwendet.⁶⁵⁰ Dazu war es zunächst erforderlich, die Variablen einschließlich aller ihrer Eigenschaften wie etwa Name, Variablentyp oder Skalenniveau zu definieren. Insbesondere mussten bei nominal skalierten Variablen die möglichen Werte codiert werden.⁶⁵¹ Nach Fertigstellung der Eingabemaske konnten die Variablen aus den Fragebögen in die Datenbank eingegeben werden.⁶⁵² Alle nachfolgend aufgeführten Daten wurden mittels SPSS berechnet.

B. Bildung von Indizes

Wie oben bereits ausgeführt, können die wenigsten interessierenden Variablen unmittelbar empirisch ermittelt werden, sodass nur der Weg bleibt, sich diesen durch empirisch feststellbare Indikatoren zu nähern, von denen angenommen wird, dass deren Vorhandensein auf die interessierenden Sachverhalte schließen lässt.⁶⁵³ Verschiedene theoretische Konstrukte lassen sich jedoch durch einen einzelnen Indikator nur unzureichend erfassen, sodass es sich empfiehlt, mehrere Indikatoren zu einem Index zusammenzufassen. Die Verwendung von Indizes erhöht die Chance, Messungengenauigkeiten zu verringern und so den gemeinten „wahren“ Sachverhalt besser abzubilden.⁶⁵⁴ Bei im Prinzip nur sehr ungenau messbaren Sachverhalten, wie etwa Einstellungen, ist davon auszugehen, dass der Mittelwert der Werte mehrerer gleichwertiger Messungen dem „wahren“ Wert näher kommt als das Ergebnis einer einzigen Messung.⁶⁵⁵ Nun ist der „wahre“ Wert in der Regel nicht bekannt, sondern soll erst – näherungsweise – ermittelt werden, sodass die Reliabilität eines Index nicht mittels eines Vergleichs von Index- und „wahrem“ Wert überprüft werden kann.⁶⁵⁶ Um dennoch annehmen zu können, dass ein verwendeter Index die interessierende Variable hinreichend zuverlässig abbildet, müssen bei der Konstruktion des Index bestimmte Mindeststandards erfüllt werden. Um die Einhaltung der Mindestvoraussetzungen zu überprüfen, unterzieht man die zunächst für geeignet gehaltenen Einzelindikatoren (Items) verschiedenen Verfahren der

650 Die Version 18 wurde unter dem Namen PASW Statistics vertrieben.

651 Ein Beispiel hierfür ist die Variable „Bundesland“; vgl. zu den Skalentypen oben, B. II. 2.

652 Hierbei wurde die Verfasserin von einer studentischen Hilfskraft unterstützt.

653 Vgl. oben, Kapitel 6, B. II. 1.

654 Vgl. *Kromrey*, *Empirische Sozialforschung*, S. 169.

655 Vgl. *Janssen / Laatz*, *Statistische Datenanalyse mit SPSS*, S. 585.

656 Vgl. *Janssen / Laatz*, *Statistische Datenanalyse mit SPSS*, S. 586.

Itemanalyse. Bei der Konstruktion der in dieser Arbeit verwendeten Indizes wurden folgende Verfahren angewendet.

I. Test auf Itemschwierigkeit

Ein wichtiges Qualitätskriterium bei der Auswahl von Indikatoren ist deren Differenzierungsfähigkeit. Dahinter steht die Überlegung, dass etwa der Zustimmungsgang zu einem Statement wenig aussagekräftig ist, wenn alle Befragten diesem zustimmen oder dieses ablehnen. Ein solches Item wäre kaum hilfreich, um die interessierende Variable zu messen.⁶⁵⁷ Ein Maß für die Differenzierungsfähigkeit ist die sogenannte Itemschwierigkeit. Sie gibt an, welcher Anteil der Befragungspersonen einem bestimmten Item zustimmt. Schwierige Items werden nur von wenigen Personen bejaht, umgekehrt sind die Zustimmungsraten bei leichten Items relativ hoch.⁶⁵⁸ Bei mehrstufigen Items lässt sich die Schwierigkeit als der Anteil der erreichten an der maximal erreichbaren Punktzahl ausdrücken.⁶⁵⁹ Im Allgemeinen sind Itemschwierigkeiten im mittleren Bereich zu bevorzugen, hier ist es üblich, den mittleren Bereich zwischen 0,2 und 0,8 anzusiedeln.⁶⁶⁰ Ideal wäre es, wenn der Durchschnitt der Item-Mittelwerte bei 0,5 der maximal erreichbaren Punktzahl läge, denn bei mittlerer Itemschwierigkeit erreicht die Itemvarianz ihr Maximum, die Differenzierungsfähigkeit ist am größten.⁶⁶¹ Für die hier verwendeten Items bedeutet dies folgendes: Alle Items beziehen sich auf die oben dargestellte Skala, weisen also Werte zwischen 0 und 6 auf.⁶⁶² Dementsprechend sind für die Itemkonstruktion Items wünschenswert, die einen Mittelwert zwischen 1,2 und 4,8 (also zwischen 0,2 und 0,8 der maximal erreichbaren Punktzahl von 6) aufweisen. Dementsprechend muss auch der Durchschnitt der Item-Mittelwerte für den jeweiligen Index in diesem Bereich liegen, idealerweise läge er nah am Wert 3 (also 0,5 der maximal erreichbaren Punktzahl von 6). Der mittlere Item-Mittelwert liegt bei allen hier verwendeten Indizes im Bereich zwischen 1,2 und 4,8. Auch die einzelnen Item-Mittelwerte erfüllen diese Voraussetzung nahezu alle. Lediglich bei einem Index liegt der höchste Item-Mittelwert mit 4,83 knapp über einem Schwierigkeitswert von 0,8.⁶⁶³

657 Vgl. *Moosbrugger / Kelava*, Testtheorie und Fragebogenkonstruktion, S. 76.

658 Vgl. *Bortz / Döring*, Forschungsmethoden und Evaluation, S. 218.

659 Vgl. *Bortz / Döring*, Forschungsmethoden und Evaluation, S. 219.

660 Vgl. *Bortz / Döring*, Forschungsmethoden und Evaluation, S. 219.

661 Vgl. *Moosbrugger / Kelava*, Testtheorie und Fragebogenkonstruktion, S. 83.

662 Vgl. zur Skala oben, Kapitel 6, B. II. 2.

663 Zu den genauen Werten vgl. die Auswertungs-Kapitel 9 - 13.

II. Test auf Trennschärfe

Auch bei der sogenannten Itemtrennschärfe geht es darum, wie gut ein einzelnes Item in der Lage ist, zwischen den Befragungspersonen zu unterscheiden, also zwischen Personen mit einer positiven und solchen mit einer negativen Einstellung zu einem gegebenen Objekt zu diskriminieren.⁶⁶⁴ Die Trennschärfe gibt an, wie stark die Differenzierung des jeweiligen Items mit der Differenzierung der zusammengefassten übrigen Items übereinstimmt.⁶⁶⁵ Ein gängiges Verfahren zur Ermittlung der Trennschärfe ist die Ermittlung der Item-zu-Rest-Korrelation, also der Korrelation zwischen jedem einzelnen Item-Wert und den jeweiligen gemittelten restlichen Item-Werten.⁶⁶⁶ Es existieren keine einheitlichen Vorgaben dazu, wie hoch der Item-zu-Rest-Korrelationskoeffizient mindestens sein sollte, teils wird eine Untergrenze von 0,2 angenommen,⁶⁶⁷ teils auch von 0,4.⁶⁶⁸ Allgemein kann ein Trennschärfekoeffizient von 0,5 oder höher als hoch eingestuft werden, was dafür spricht, das entsprechende Item zu verwenden.⁶⁶⁹ Die in dieser Arbeit verwendeten Indizes bestehen bis auf einen Fall alle nur aus Items, die Trennschärfekoeffizienten von 0,5 oder darüber aufweisen; bei einem Index beträgt die niedrigste Item-zu-Rest-Korrelation 0,449.⁶⁷⁰

III. Test auf Homogenität

Wenn die einzelnen Indikatoren Operationalisierungen desselben Konstrukts darstellen, so muss sich dies darin zeigen, dass sie untereinander korrelieren, die Höhe dieser sogenannten Item-Interkorrelation ergibt die Homogenität eines Index.⁶⁷¹ Bei eindimensionalen Messinstrumenten, also bei Indizes, deren Einzelindikatoren letztlich dasselbe theoretische Konstrukt messen sollen, sind hohe Homogenitäten erstrebenswert.⁶⁷² Weist ein Item eine auffallend geringere Item-Interkorrelation auf als die anderen Items, dann deutet dies darauf hin, dass dieses Item nicht nur die Zieldimension

664 Vgl. Rohwer / Pötter, Datenkonstruktion, S. 104.

665 Vgl. Moosbrugger / Kelava, Testtheorie und Fragebogenkonstruktion, S. 84.

666 Vgl. Bortz / Döring, Forschungsmethoden und Evaluation, S. 219; Asendorpf / Neyer, Psychologie der Persönlichkeit, S. 53f., 92.

667 Vgl. Asendorpf / Neyer, Psychologie der Persönlichkeit, S. 54.

668 Vgl. Diekmann, Empirische Sozialforschung, S. 246, die Ausführungen beziehen sich jedoch auf die Item-zu-Summenscore-Korrelation, die höher ausfällt als die Item-zu-Rest-Korrelation, da das getestete Item selbst Teil des Summenscores ist.

669 Vgl. Bortz / Döring, Forschungsmethoden und Evaluation, S. 220.

670 Vgl. Index „Beurteilung der eigenen Subjektstellung durch die Klägerin / den Kläger“, unten, Kapitel 11, B.

671 Vgl. Bortz / Döring, Forschungsmethoden und Evaluation, S. 220.

672 Vgl. Bortz / Döring, Forschungsmethoden und Evaluation, S. 220.