2.5 Skalierungsverfahren

2.5.1 Skalierungsverfahren I: Notation und probabilistische Grundstruktur

Methoden zur Konstruktion von Messinstrumenten:
Ermittle individuelle Ausprägungen einer latenten Eigenschaft mit Hilfe einer 'Batterie' von Fragen ('Items')

Notation

- Vorstellung: Universum von
  * Personen und von
  * Stimuli, Items, Aufgaben, etc.
- Personenfähigkeit, Einstellungsgrad, etc. latentes Merkmal Γ
- Ebenso: Schwierigkeit der Items$_i$: latentes Merkmal B (hier:'großes β')

- Ziehe zufällige Stichprobe von Personen $i, i = 1, \ldots, n$, und von Items $j, j = 1, \ldots, k$, (Vorsicht: Oft werden **I**tems auch mit i indiziert und Personen mit j)
- Γ_i: Personenfähigkeit, Einstellungsgrad etc. der $i-$ten Person in der Stichprobe, Ausprägung γ_i
- B_j: Itemschwierigkeit des $j-$ten Items, Ausprägung β_j
- Datenmatrix $(C_i^{(j)})_{i=1,\ldots,n;j=1,\ldots,k}$; $C_i^{(j)}$ als Antwort der Person i auf Frage/Item j

- Zur Ermittlung von $(\gamma_i)_{i=1,\dots,n}$ und $(\beta_j)_{j=1,\dots,k}$ sehe diese Größen als Parameter eines statistischen Modells für die Daten $(C_i^{(j)})_{i=1,\dots,n;j=1,\dots,k}$, d. h. modelliere Wahrscheinlichkeitsverteilung von $C_i^{(j)}$ als Funktion von γ_i und β_j

$$\mathcal{L}(C_i^{(j)}) = h(\gamma_i, \beta_j), \qquad i = 1, \dots, n; j = 1, \dots, k$$

- Es macht Sinn, von einer <u>probabilistischen Testtheorie</u> zu sprechen, wobei in der Literatur dieser Begriff nur für kategoriale $C_i^{(j)}$ gebräuchlich ist.

- Typischerweise wird <u>'lokale stochastische Unabhängigkeit'</u> vorausgesetzt:
  Gegeben Γ_i und B_j, $i = 1, \dots, n; j = 1, \dots, k$, sind die einzelnen $(C_i^{(j)})_{i=1,\dots,n;j=1,\dots,k}$ stochastisch unabhängig.

2.5.2 Skalierungsverfahren II: Metrische Beobachtungen

Ist $(C_i^{(j)})_{i=1,\ldots,n;j=1,\ldots,k}$ metrisch, so ist ein natürlicher Ansatz in Verallgemeinerung des Grundmodells der klassischen Testtheorie der lineare Ansatz

$$\mathbb{E}(C_i^{(j)}|\Gamma_i = \gamma_i, B_j = \beta_j) = \gamma_i - \beta_j.$$

Für Kleinste-Quadrate-Methode vorstellen:

$$C_i^{(j)} = \sum_{\ell=1}^{n} \gamma_\ell X_{\ell,i} + \sum_{p=1}^{k} \tilde{\beta}_p X_{n+p,j}$$

für die $\tilde{n} = n \cdot k$ Beobachtungen mit

$$X_{\ell,i} = \begin{cases} 1 & \ell = i \\ 0 & \text{sonst} \end{cases} \quad 1 \leq \ell \leq n \qquad \text{und} \qquad X_{n+p,j} \begin{cases} 1 & p = j \\ 0 & \text{sonst} \end{cases} 1 \leq p \leq k\,.$$

Momentenmethode liefert üblichen Summenscore:

2.5.3 Skalierungsverfahren III: Binäre Beobachtungen

Situation $C_i^{(j)}$ binär, $C_i^{(j)} = 1$ [bzw. $C_i^{(j)} = 0$], Person i hat Aufgabe j richtig [bzw. falsch] beantwortet, oder Aussage j zugestimmt [abgelehnt].

Modelliert wird $p(C_i^{(j)} = C_i^{(j)} | \Gamma_i = \gamma_i, B_j = \beta_j)$. Wegen der lokalen stochastischen Unabhängigkeit gilt mit:

$$\mathbf{C} = (C_i^{(j)}), i = 1, \ldots, n; j = 1, \ldots k,$$

$$(\mathbf{\Gamma}, \mathbf{B}) = (\Gamma_i, B_j), i = 1, \ldots, n; j = 1, \ldots k,$$

und entsprechenden Ausdrücken für die Ausprägungen

$$p\left(\mathbf{C} = \mathbf{t} | \mathbf{\Gamma} = \gamma, \mathbf{B} = \beta\right) = \prod_{i=1}^{n} \prod_{j=1}^{k} p\left(C_i^{(j)} = C_i^{(j)} | \Gamma_i = \gamma_i, B_j = \beta_j\right)$$

Itemcharakteristische Funktionen für binäre Items

Wieder: Vergleich zwischen Stärke des Subjekt und Stärke des Stimuli (bzw. Item). Die Veranschaulichung erfolgt mithilfe sogenannter ICCs (*item characteristic curve*, traceline oder Itemcharakteristik);

- Die Grundform der ICCs wird in einigen Verfahren als bekannt (z.B. logistische Kurve im Rasch-Modell (s. u.)) angenommen.

- Zeichne $p(C_i^{(j)} = C_i^{(j)} | \Gamma_i = \gamma_i, B_j = \beta_j)$ als Funktion von γ_i für verschiedene Werte von β_j, d. h.

  Ausprägung der latenten Variablen wird auf der Abszisse abgetragen.

  Lösungswahrscheinlichkeit bzw. Zustimmungswahrscheinlichkeit auf der Ordinate

- Interpretation monotoner Verlauf: Mit steigendem Wert der latenten Eigenschaft steigt die Lösungs- oder Zustimmungswahrscheinlichkeit.

* Je fähiger die Person, desto höher ist die Lösungswahrscheinlichkeit,
* bzw. je extremer die Einstellung der Person, desto höher Zustimmungswahrscheinlichkeit.

- Trennschärfe erkennt man an der Steigung der ICCs (verlaufen die Items parallel, so haben sie die gleiche Trennschärfe.)

- Die sog. *Item Response Theorie* (IRT) umfasst viele Skalierungsverfahren, das wohl bekannteste ist das *Rasch-Modell*.

Dort setzt man:

$$P(C_i^{(j)} = C_i^{(j)} | \Gamma_i = \gamma_i, B_j = \beta_j) = \frac{\exp[(\gamma_i - \beta_j) \cdot C_i^{(j)}]}{1 + \exp(\gamma_i - \beta_j)}$$

Auch in der PISA Studie werden das Rasch-Modell und einige Erweiterungen verwendet, um die Vorbereitung von 15-jährigen Schülern auf die Wissensgesellschaft zu messen.

Vorgehen:

1. Schätzung der Aufgabenschwierigkeiten

2. Schätzung der Personenparameter

- Nichtstandard-Fall im Maximum-Likelihood-Kontext

- Zur Schätzung der Parameter kann z.B. die sog. *bedingte Maximum-Likelihood Schätzung* oder die sog. marginale Likelihood-Schätzung verwendet werden.

- Wenn das Rasch-Modell gilt, entsprechen die Personenparameter (mindestens) einer Intervallskala.

- Wenn das Rasch-Modell gilt, dann ist die geschätzte Fähigkeit oder Einstellung eine monotone Transformation der Summe der gelösten Items. Diese bilden dann eine ordinale Skala.

- Zentraler Vorteil der IRT Modelle ist, dass die Annahmen explizit angegeben werden können und Tests zur empirischen Überprüfng der Annahmen existieren.

Weitere Skalierungsverfahren

- Mehrdimensionale Rasch-Modelle
- Birnbaum Modell (auch 2 PL-Modell genannt)
- 3 PL-Modell
- Partial-Credit Modell, Rating-Scale-Modell, Graded-Response Modell
- Im weiteren Sinn auch Paarvergleichmodellierung :

2.5.4 Skalierungsverfahren IV: Modellabweichungen, Itembereinigung

- ’Differential Item Functioning’ ; Aufgabenschwierigkeit hängt von externen Faktoren ab, eine Aufabe ist also nicht gleich schwer für alle Teilnehmer

- Manche Items nicht ’trennscharf’, kein/geringer Einfluss von individueller Fähigkeit

Likert-Skala für die Situation aus 2.5.2

- Likert-Skala bei 'als metrisch angenommen Größen: Statements von „stimme voll zu“, „stimme eher zu“, „teils/teils“, „stimme eher nicht zu“, „stimme nicht zu“

- übliches Vorgehen:
  * „Ratings“ werden aufsummiert (▷ Intervallskala vorausgesetzt!?)
  * Prüfung der Trennschärfe (wie gut werden hohe Eigenschaftswerte und niedrige Eigenschaftswerte durch die Kategorien separiert) und der Reliabilität (Cronbachs α, s. o.)
  * ausreichend reliable und trennscharfe Items bleiben, aus diesen wird Summe oder Mittelwert gebildet

- Probleme
  * Eindimensionalität häufig verletzt
  * kein Nachweis eines Skalenniveaus
  * abhängig von der Stichprobe
  * Verfahren ist in gewisser Weise zirkulär.

3 Indikatoren und Indizes: weitere Beispiele, Konstruktionsprinzipien und Analysetechniken

Literatur

Grundlegende Aspekte:

- von der Lippe, P. (1996 ^{5}): *Wirtschaftsstatistik*. UTB (Lucius & Lucius).
- Schaich, E. & Schweitzer, W. (1995): *Ausgewählte Methoden der Wirtschaftsstatistik*. Verlag Franz Vahlen.
- Toutenburg, H. & Heumann, C. (2009^{7}): *Deskriptive Statistik*. Springer.
- Rinne, H. (1996^{2}): *Wirtschafts- und Bevölkerungsstatistik*. Oldenbourg Verlag. Insbesondere Kapitel 13.
- Winker, P. (2010^{3}): *Empirische Wirtschaftsforschung und Ökonometrie*. Springer. Insbesondere Kapitel 1, 3, 4 und 10.
  Volltext-Download für LMU-Angehörige.

Weitere verwendete Literatur

- Abberger, K., Nierhaus, W. (2007): Das ifo Geschäftsklima: Ein zuverlässiger Frühindikator der Konjunktur. *ifo Schnelldienst* 5/2007. (http://www.cesifo-group.de/link/ifosd_2007_5_3.pdf, aufgerufen am 15.11.2016 (auch open access))

- Bechtold, S., Elbel, G., Hannappel, H.-P. (2005): Messung der wahrgenommenen Inflation in Deutschland: Die Ermittlung der Kaufhäufigkeiten durch das Statistische Bundesamt. *Wirtschaft und Statistik* 9/2005, 989-998. (https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/Monatsausgaben/WistaSeptember05.pdf?__blob=publicationFile, aufgerufen am 15.11.2016 (auch open access))

- Brachinger, H. W. (2005): Der Euro als Teuro? Die wahrgenommene Inflation in Deutschland. *Wirtschaft und Statistik* 9/2005, 999-1013. (https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/Monatsausgaben/WistaSeptember05.pdf?__blob=publicationFile, aufgerufen am 15.11.2016 (auch open access))

- Linz, S., Eckert. G. (2002): Zur Einführung hedonischer Methoden in der Preisstatistik. *Wirtschaft und Statistik* 2002/10; 857-863. (`https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/Monatsausgaben/WistaOktober02.pdf?__blob=publicationFile`, aufgerufen am 15.11.2016 (auch open access))

- Hoffmann, J.; Leifer, H.-A.; Lorenz, A. (2005): Index der wahrgenommenen Inflation oder Verbraucherumfragen? Zu einem Ansatz von Hans Wolfgang Brachinger. *Wirtschaftsdienst* 85/11, 706-714. (`http://www.wirtschaftsdienst.eu/downloads/getfile.php?id=451`, aufgerufen am 15.11.2016 (auch open access))

- Inter-Secretariat Working Group on Price Statistics (IWGPS, Hg.): (2011): *Intersecretariat Working Group on Price Statistics – Terms of Reference.* (`www.ilo.org/public/english/bureau/stat/guides/cpi/textor.pdf`, aufgerufen am 15.11.2016)

- Linz, S., Dexheimer, V., Kathe, A. (2003): Hedonische Preismessung bei Gebrauchtwagen. *Wirtschaft und Statistik* 2003/6, 538-542. (`https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/Monatsausgaben/WistaJuni03.pdf?__blob=publicationFile`, aufgerufen am 15.11.2016 (auch open access))

In diesem Kapitel werden zunächst aus mehreren Größen zusammengesetzte Indikatoren und daraus abgeleitete Indizes zur Messung eines komplexen Konstrukts genauer betrachtet.

- Verhältniszahlen
- Preis- und Mengenindizes
- Konjunkturindikatoren

Eine große Rolle spielen dabei zeitliche Aspekte.

3.1 Verhältniszahlen

Eine Verhältniszahl ist eine Kennzahl, die durch den Quotienten zweier statistischer Größen gebildet wird.

Man kann drei Arten von Verhältniszahlen unterscheiden:

1. Gliederungszahlen/ Quoten
2. Beziehungszahlen
3. Messzahlen/ Maßzahlen

Verhältniszahlen werden häufig noch „*gegliedert*“, also für bestimmte Subgruppen getrennt ermittelt (z.B. nach Geschlecht, Alter)

3.1.1 Gliederungszahlen/ Quoten

Gliederungszahlen bzw. Quoten drücken Anteile aus, d.h. der Zähler ist jeweils ein Teil des Nenners. Damit liegen Gliederungszahlen immer zwischen 0 und 1 bzw., wenn in $\%$ ausgedrückt, zwischen 0 und 100.

Beispiele:

- Betreuungsquote der unter 3-Jährigen:

$$\frac{\text{Anzahl Kinder unter 3 Jahre in Kindertagesbetreuung}}{\text{Anzahl Kinder unter 3 Jahre}}$$

  `https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2016/09/PD16_345_225.html`, aufgerufen am 15.11.2016.

- Frauenquote eines Unternehmens

- Arbeitslosenquote (nach BA)
(`www.statistik.arbeitsagentur.de`, aufgerufen am 15.11.2016):

$$\frac{\text{Anzahl der gemeldeten Arbeitslosen}}{\text{zivile Erwerbstätige} + \text{Arbeitslose}}$$

- Erwerbslosenquote (nach ILO-Standard)
(`https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/Arbeitsm`
`Erwerbslosigkeit/Erwerbslosigkeit.html`, aufgerufen am 15.11.2016):

$$\frac{\text{Anzahl der Erwerbslosen}}{\text{Erwerbstätige} + \text{Erwerbslose}}$$

„Arbeitslos sind nach dem Sozialgesetzbuch Personen, die vorübergehend nicht in einem Beschäftigungsverhältnis stehen, das 15 Wochenstunden und mehr umfasst, eine versicherungspflichtige Beschäftigung von mindestens 15 Wochenstunden suchen und dabei den Vermittlungsbemühungen der Agenturen für Arbeit bzw. der Träger der Grundsicherung zur Verfügung stehen und sich dort persönlich arbeitslos gemeldet haben." (Quelle: `www.statistik.arbeitsagentur.de`, aufgerufen am 15.11.2016 Bundesagentur für Arbeit)

„Die ILO-Definition von Erwerbslosigkeit wird [. . .] konkretisiert auf nicht erwerbstätige Personen von 15 bis 74 Jahren, die in den vier Wochen vor der Befragung aktiv nach einer Tätigkeit gesucht haben und eine solche innerhalb von zwei Wochen aufnehmen könnten." (Quelle: `https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/Arbeitsmarkt/Erwerbslosigkeit/Erwerbslosigkeit.html`, aufgerufen am 15.11.2016 Statistisches Bundesamt)

Arbeitslosenquote versus Erwerbslosenquote (Quellen: https://www.destatis.de/DE/Publikationen/Datenreport/Datenreport.html, aufgerufen am 15.11.2016 Datenreport 2016, https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmw Arbeitsmarkt/Arbeitsmarkt.html, aufgerufen am 15.11.2016 Stat. B.-Amt):

Jahr	Arbeitslosen-quote in %	Erwerbslosen-quote in %
2003	10,5	9,2
2004	10,5	9,7
2005	11,7	10,5
2006	10,8	9,8
2007	9,0	8,3
2008	7,8	7,2
2009	8,1	7,4
2010	7,7	6,8
2011	7,1	5,7
2012	6,8	5,3
2013	6,9	4,9
2014	6,7	4,7
2015	6,4	4,3

3.1.2 Beziehungszahlen

Bei Beziehungszahlen sind Zählergröße und Nennergröße gleichrangig, betreffen aber jeweils unterschiedliche Tatbestände. (Wertebereich: $\mathbb{R}$, typischerweise $\mathbb{R}_0^+$)

Beispiele:

- Anspannungsindex:

$$\frac{\text{registrierte Arbeitslose}}{\text{gemeldete offene Stellen}}$$

- Staatsverschuldung in Relation zur Wirtschaftsleistung:

$$\frac{\text{Bruttoverschuldung des Staates}}{\text{Brutto-Inlands-Produkt (BIP) des Staates}}$$

Das BIP gibt den Gesamtwert aller innerhalb eines bestimmten Zeitraumes (i.d.R. ein Jahr) in einem Land produzierten Waren und Dienstleistungen an, die für den Endverbrauch bestimmt sind. (Siehe z.B. `https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/VGR/Inlandsprodukt/Inlandsprodukt.html`, aufgerufen am 15.11.2016 Statistisches Bundesamt)

3.1.3 Messzahlen/ Maßzahlen

Messzahlen bzw. Maßzahlen drücken eine interessierende Variable in Bezug auf einen Basiswert aus. Zählergröße und Nennergröße sind gleichrangig und betreffen beide gleichartige Tatbestände. (Wertebereich: $\mathbb{R}$, typischerweise $\mathbb{R}_0^+$)

Beispiele:

- Preismesszahl zur Basisperiode 0:

$$\frac{\text{Preis eines Gutes zum Zeitpunkt } t}{\text{Preis eines Gutes zum Zeitpunkt } 0}$$

- aus der Zeitreihe von Preisen $p_0, p_1, \ldots, p_t, \ldots$ kann z.B. die folgende Zeitreihe von Preismesszahlen gebildet werden:

$$\frac{p_1}{p_0}, \frac{p_2}{p_0}, \ldots, \frac{p_t}{p_0}, \ldots$$

- dient als Basis für Preisindizes, nachfolgend detaillierter betrachtet