5.3 Konzeptspezifikation, Operationalisierung, Messung

5.3.1 Begriffe (Konzepte), Begriffsintension und -extension

- Zwei Begriffsarten:
  * logische (und, wenn,...)
  * empirische (außerlogische):
- Intension (Inhalt): Menge der Merkmale, die für Vorliegen des Begriffs gegeben sein müssen.
- Extension (Umfang): Menge aller Objekte, die die Intension erfüllen (kann leer sein).
- Exkurs: Fuzzy Sets

Dispositionsbegriff, Operationalisierung *„Solche Begriffe bezeichnen Eigenschaften, die sich nur unter bestimmten Bedingungen ermitteln lassen. So zeigt Zucker seine Disposition „löslich“ nur, wenn man ihn ins Wasser gibt. Der Dispositionsbegriff „Löslichkeit“ läßt sich also nur durch die zusätzliche Angabe bestimmter Bedingungen bzw. Situationen operationalisieren. Die überwiegende Mehrzahl der sozialwissenschaftlich relevanten theoretischen Begriffe dürfte zur Klasse der Dispositionsmerkmale gehören. [...] Hier müssen mit Hilfe von Fragen oder andere Methoden (T.A.) solche Situationen, in denen sich dispositionale Eigenschaften zeigen, generiert werden.“* (Schnell, Hill, Esser: 67)

Typische Dispositionsmerkmale in den Sozialwissenschaften sind:

- Politische Einstellung
- Schicht
- Religiösität
- Rechtsextremismus

Dispositionsbegriff – latente Variablen

Wenn eine Variable nicht immer direkt beobachtbar ist, gilt der Dispositionsbegriff. In der Psychologie wird bei nicht beobachtbaren Größen von latenten Eigenschaften (latent trait im metrischen Fall – z.B. Intelligenz – oder latent class im kategorialen Fall – z.B. schizophren) gesprochen. Da man im Regelfall während der Erhebung keine Situation herstellen kann, in der sich die Größe direkt zeigt, werden als Messsituation oft Fragen gestellt.

- Fragen zu Indikator(en)
- Fragebatterie, Itembatterie
- Datenerhebung nach jeweiligem Messverfahren

Korrespondenzproblem und Basissätze

„So ist die empirische Basis der objektiven Wissenschaft nichts „Absolutes“; die Wissenschaft baut nicht auf Felsengrund. Es ist eher Sumpfland. . . “ Popper (1976; p. 75) zitiert nach Schnell, Hill, Esser (2005; p. 82)

Operationalisierung Die Festlegung von Korrespondenzregeln, mit deren Hilfe Konstrukte und Indikatoren in Beziehung gesetzt werden, wird insbesondere in der Sozialwissenschaft als $Operationalisierung$ bezeichnet. Vor allem in der amtlichen Statistik wird die Frage der Passung von Konstrukten und Indikatoren als $Adäquation$ problematisiert.

- Typischerweise operationalistische Indikatorenbildung
  * Direkte Definition des Begriffs über Messvorschrift (operationale Definition von Begriffen)
  * T.A: Physikalische Einheiten: Ein Ampere ist die Stärke eines zeitlich unveränderlichen elektrischen Stromes, der, durch zwei im Vakuum parallel im Abstand 1 Meter voneinander angeordnete, geradlinige, unendlich lange Leiter von vernachlässigbar

kleinem, kreisförmigem Querschnitt fließend, zwischen diesen Leitern pro 1 Meter Leiterlänge die Kraft $2 \cdot 10^{-7}$ Newton hervorrufen würde.

* Z.B. Intelligenz ist das, was ein Intelligenztest misst

* Forschungsleistung: Mit den Impact-Factor der Zeitschrift gewichtete Anzahl von Publikationen

* Problem: Theorienvergleich nur möglich, wenn gleiche Operationalisierung verwendet wurde.

5.3.2 Messen in der empirischen Sozialforschung

Skalenniveaus und zulässige Verfahren werden hier natürlich nicht mehr besprochen.

Messung

-

„*Eine Messung im Sinne der Meßtheorie liegt vor, wenn (=def.) ein Isomorphismus oder ein Homomorphismus zwischen einem empirischen und einem numerischen Relativ existiert.* “ (Diekmann 247);

Skala

Das Grundmodell der klassischen Testtheorie Testtheorie im Sinne der Psychologie gemeint!

$$\underset{\substack{\uparrow \\ \text{gemessener Wert}}}{T} = \underset{\substack{\uparrow \\ \text{wahrer Wert}}}{\theta} + \underset{\substack{\uparrow \\ \text{Messfehler}}}{\delta}$$

$\mathbb{E}(\delta)$ und $\mathrm{Var}(\delta)$ existieren.

1. Kein systematischer Messfehler, d.h. es gilt: $\mathbb{E}(\delta) = 0$.

2. θ und δ sind unkorreliert (oder unabhängig)

3. Messfehler bei verschiedenen Variablen sind unkorreliert (bzw. von einander unabhängig)

4. Messfehler korrelieren auch nicht mit wahren Werten anderer Variablen (bzw. sind davon unabhängig)

5.3.3 Gütekriterien: Objektivität, Reliabilität, Validität

Eine Messung als Homomorphismus kann zunächst relativ willkürlich und inhaltlich sinnlos sein. Bei der Berurteilung der Güte einer Messung unterscheidet man typischerweise drei Aspekte:

i) Objektivität:

   Grad der Unabhängigkeit der Messung von Einflüssen außerhalb der befragten bzw. untersuchten Person, Einheit

   * Durchführungsobjektivität:

   * Auswertungsobjektivität:

   * Interpretationsobjektivität:

ii) Reliabilität (Zuverlässigkeit)
In welchem Ausmaß führt eine wiederholte Messung zu demselben Ergebnis?

iii) Validität (Gültigkeit)
Grad der Genauigkeit, mit ein Verfahren oder eine Messung das misst, was es messen soll

Beurteilung der Reliabilität

- Maßzahl für die Reliabilität:
  „Strukturelles Modell“: θ selbst zufällig! (damit hat θ selbst eine Varianz)
  Sind die entsprechenden Varianzen > 0, so definiert man die Reliabilität des Indikators groß für die Variable θ bei Gültigkeit des Grundmodells der klassischen Testtheorie als:

$$Rel(\theta, T) := \frac{Var(\theta)}{Var(T)} =$$

  Zusammenhang zur Korrelation

- „Schätzung“ der Reliabilität
  * $Validierungsdaten$:
  * Test-Retest-Methode:

* Paralleltestmethode:

* Split-Half-Methoden zur Erzeugung paralleler Tests:

* Cronbachs Alpha (berücksichtigt die verschiedenen Aufteilungsmöglichkeiten)
  Normierte Form:

$$\alpha = \frac{n \cdot \bar{\rho}}{1 + \bar{\rho}(n-1)}$$

Beurteilung der Validität

- Inhaltsvalidität
- Kriteriumsvalidität
- Konstruktvalidität
- *Theoretische Validität:* $\rho(T, \theta)$, zusätzlich $\mathbb{E}(\delta) = 0$
- *Inhaltsvalidität*: Alle Dimensionen des Konstrukts erfasst – und nur diese!
- *Empirische Validität:*
  Korrelation zu beobachtbarer Variable, die als Kriterium dient

- *Kriteriumsvalidität*: Hoher Zusammenhang zwischen Messwerten und einem anderen gemessenem Kriterium („externes Kriterium“)
  * Prädiktive Validität:
  * Konkurrente Validität:
- *Konstruktvalidität:*

5.3.4 Indexbildung, Skalierungsverfahren

Indexbildung

- Zusammenfassung mehrerer Indikatoren eines Konstrukts zu einer Kennzahl. (vgl. Kap. 2.2: kollektive Kennzahl, die durch Kobination verschiedener Größen gebildet wird.)

- Oft sind Konstrukte mehrdimensional

* Werden die Dimensionen jeweils durch Indikator erfasst?

* Welche Dimensionen fließen ein?

* Wie werden sie kombiniert?
  * Jeder Ausprägungskombination wird ein Wert zugeordnet
  * Zuordnungsregel: additiver Index
  * Zuordnungsregel: multiplikativer Index

* Zuordnungsregel: gewichteter Index $\sum_{j=1}^{k} g_j \cdot t_j$, wobei t_j mit $j = 1, \ldots, k$ die Items bzw. Indikatoren widerspiegeln und g_j die (festgesetzten) Gewichte darstellen

– Die Verknüpfungen durch Rechenoperationen als solche setzen ein höheres Skaleniveau voraus. Daher Verwendung von Indizes in vielen Situationen kritisch zu hinterfragen.

– Index kontinuierlicher Variablen

Skalierungsverfahren Methoden zur Konstruktion von Messinstrumenten

Vorbereitung der Skalierungsverfahren Notation

- Personen i mit $i = 1, \ldots, n$
- Stimuli, Items, Aufgaben j mit $j = 1, \ldots, k$
- Personenfähigkeit, Einstellung etc. als latente Variable θ
- „Fähigkeit“ der Person i ist θ_i
- Aufgabenparameter, Schwierigkeit des Items, Extremheit des Items β
- Schwierigkeit der Aufgabe j ist β_j
- Antwort auf eine dichotome Aufgabe (Fähigkeit) oder eine Zustimmungsfrage (ja oder nein, 0/1).

Antwort der Person i auf
Item j wird mit U_{ij} bezeichnet

$U_{ij} = 0$ Nicht-Lösung, Ablehnung

$U_{ij} = 1$ Lösung, Zustimmung

* Modelliert wird oft $P(U_{ij} = U_{ij}|\theta_i, \beta_j)$

Exkurs: Einige Regeln zur Formulierung von Items

- eindimensionale Items
- Frage nach gegenwärtigem Zustand
- keine Tatsachenbeschreiben, Suggestion
- Items sollten den Wertebereich, indem die Befragten liegen abdecken
- einfache, klare, kurze, verständliche Struktur, keine Mehrdeutigkeit, keine doppelte Verneinung
- Antwortkategorien erschöpfend

Likert-Skala

- Ziel: Messe Eigenschaft der Person aufgrund einer großen Anzahl von Items
- Annahmen: Eindimensionalität, Aussagen zu Statements hängen nur mit der Eigenschaft zusammen
- übliches Vorgehen:
  * Statements von „stimme voll zu“, „stimme eher zu“, „teils/teils“, „stimme eher nicht zu“, „stimme nicht zu“
  * sogenannte „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)

  * 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.

Itemcharakteristische Funktionen für binäre Items

Neuer Aspekt: Vergleich zwischen Subjekt und 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) angenommen.
- Latente Variable wird auf der Abszisse abgetragen.
- Lösungswahrscheinlichkeit bzw. Zustimmungswahrscheinlichkeit auf der Ordinate
- Interpretation monotoner Verlauf: Mit steigendem Wert der latenten Eigenschaft steigt Lösungs- oder Zustimmungswahrscheinlichkeit.
  * Je fähiger die Person, desto höher Lösungswahrscheinlichkeit.
  * 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.)

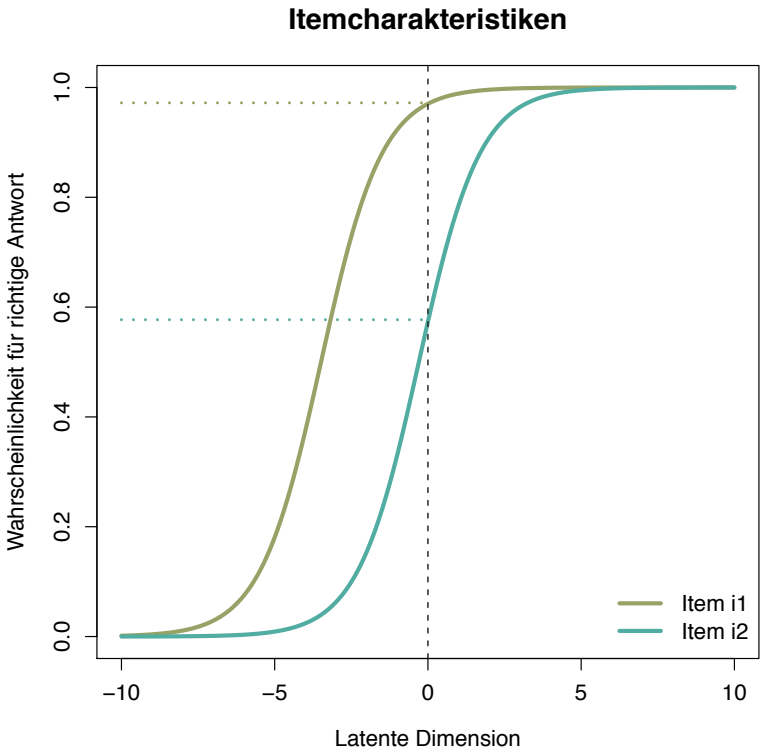
Rasch-Skala und Rasch-Modell

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

Ziel ist wieder die Schätzung eindimensionaler latenter Personeneigenschaften wie Fähigkeiten (z.B.: Mathekompetenz von Schülern, Lesekompetenz) oder Einstellungen.

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.

- Ziel: Schätze latente Eigenschaft der Person z.B. Mathefähigkeit (PISA) oder Einstellung; Wenn das Modell gilt, erhält man eine Messung auf mindestens Intervallskalenniveau (genauer: Differenzenskala.)


Itemcharakteristiken
Wahrscheinlichkeit für richtige Antwort
0.0
0.2
0.4
0.6
0.8
1.0
−10
−5
0
5
10
Latente Dimension
Item i1
Item i2

Modell:

$$P\left((U_{ij} = u_{ij}|\theta_i, \beta_j)_{i=1,\ldots,n,j=1,\ldots k}\right) = \prod_{i=1}^{n}\prod_{j=1}^{k} P\left(U_{ij} = u_{ij}|\theta_i, \beta_j\right)$$

mit

$$P(U_{ij} = u_{ij} \,|\, \theta_i, \beta_j) = \frac{\exp[(\theta_i - \beta_j) \cdot u_{ij}]}{1 + \exp(\theta_i - \beta_j)}$$

ICCs im Rasch-Modell

- Modell:

$$\mathbb{P}\left(U_{ij} = u_{ij} \,|\, \theta_i, \beta_j\right) = \frac{\exp[(\theta_i - \beta_j) \cdot u_{ij}]}{1 + \exp(\theta_i - \beta_j)}$$

- 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

5.4 Erhebungsarten

- Befragung
- Beobachtung
- Inhaltsanalyse
- Nicht-reaktive Verfahren

5.4.1 Befragung

Arten der Befragung

- face-to-face („persönlich“)
- telefonische Befragung
- schriftliche Befragung
- CATI, CAPI: Computer Assisted Telefone Interview, - Personal Interview

Charakteristika der persönlichen Befragung (Befragung als soziale Situation)

- Reaktivität:
- Interviewsituation:
- Interviewereffekte:
- Strukturierungsgrad:
- neutrale Interviewtechnik:
- Befragung setzt auf:
  1. Kooperation im Regelfall
  2. Norm der Aufrichtigkeit
  3. gemeinsame Sprache

Beachten sollte man (je nach Möglichkeiten)

- neutrale Interviewtechnik
- Retrospektivfragen
- heikle Fragen
- Incentives

- **Befragtenmerkmale:** Ja-Sager (Akquieszenz), Meinungslose, Pseudo-Opinions, Soziale Erwünschtheit, Response Sets

- **Interview(er)merkmale:** Situation, Interviewer

- **Fragemerkmale:** z.B. Dimension der Fragen, Reihenfolge, Antwortmöglichkeiten

5.4.2 Beobachtung i.e.S.

Arten und Charakteristika der Beobachtung

- Feld versus Labor
- teilnehmend versus nicht-teilnehmend
- offen versus verdeckt
- unstrukturiert versus strukturiert

5.4.3 Inhaltsanalyse

5.4.4 Nicht-reaktive Verfahren

5.5 Zur statistischen Modellierung „defizitärer Daten“

Statistische Methoden zur Behandlung unvermeidlicher Fehler: **5.5.1 Messfehlermodelle (siehe Übung)**

Regressionsmodelle zwischen dem Grundmodell der klassischen Testtheorie gehorchenden latenten Variablen.

5.5.2 Fehlklassifikation

Univariater Fall

Ausgangssituation im binären Fall:

- θ wahre binäre Variable

- T beobachtete binäre Variable

- Defizitätsmodell

$$
\begin{aligned}
P(T=1|\theta=1) &=: \pi_{11} \quad \text{Sensitivität} \\
P(T=0|\theta=0) &=: \pi_{00} \quad \text{Spezifität} \\
P(T=0|\theta=1) &= 1-\pi_{11} =: \pi_{01} \\
P(T=1|\theta=0) &= 1-\pi_{00} =: \pi_{10}
\end{aligned}
$$

- Fehlklassifikationsmatrix

$$
\pi = \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix}
$$

Auf der Diagonale: Wahrscheinlichkeiten konkreter Klassifikationen.

Auswirkungen der Fehlklassifikation

- gesucht $P(\theta=1) =: p$

- naiver Schätzer:

$$\hat{p}^{-} = \frac{1}{n} \sum_{i=1}^{n} T_i$$

-

$$\begin{aligned} P(T=1) &= P(t=1|\theta=1) \cdot P(\theta=1) \\ &\quad + P(t=1|\theta=0) \cdot P(\theta=0) \\ &= \pi_{11} \cdot P(\theta=1) + \pi_{10} \cdot P(\theta=0) \end{aligned}$$

- Ausmaß des Bias

$$\begin{aligned} P(T=1) - P(\theta=1) &= \pi_{11} \cdot P(\theta=1) + \pi_{10} \cdot P(\theta=0) \\ &\quad - (\pi_{11} + \pi_{01}) \cdot P(\theta=1) \\ &= \pi_{10} \cdot P(\theta=0) - \pi_{01} \cdot P(\theta=1) \end{aligned}$$

- $Bias = 0$, falls $P(\theta = 1)$ und $\pi_{00} = \pi_{11}$

- positiver wie negativer Bias möglich

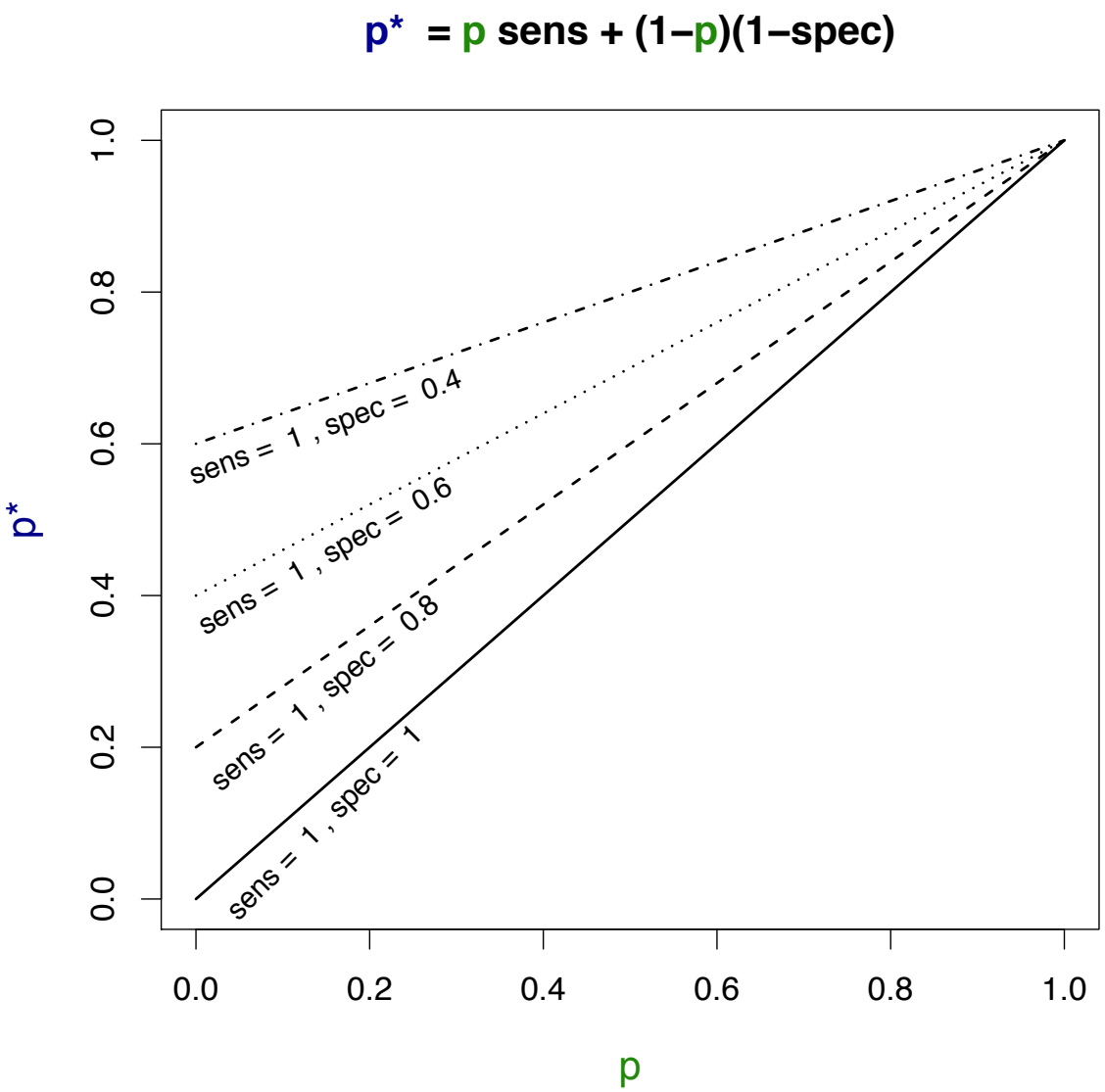
Korrektur

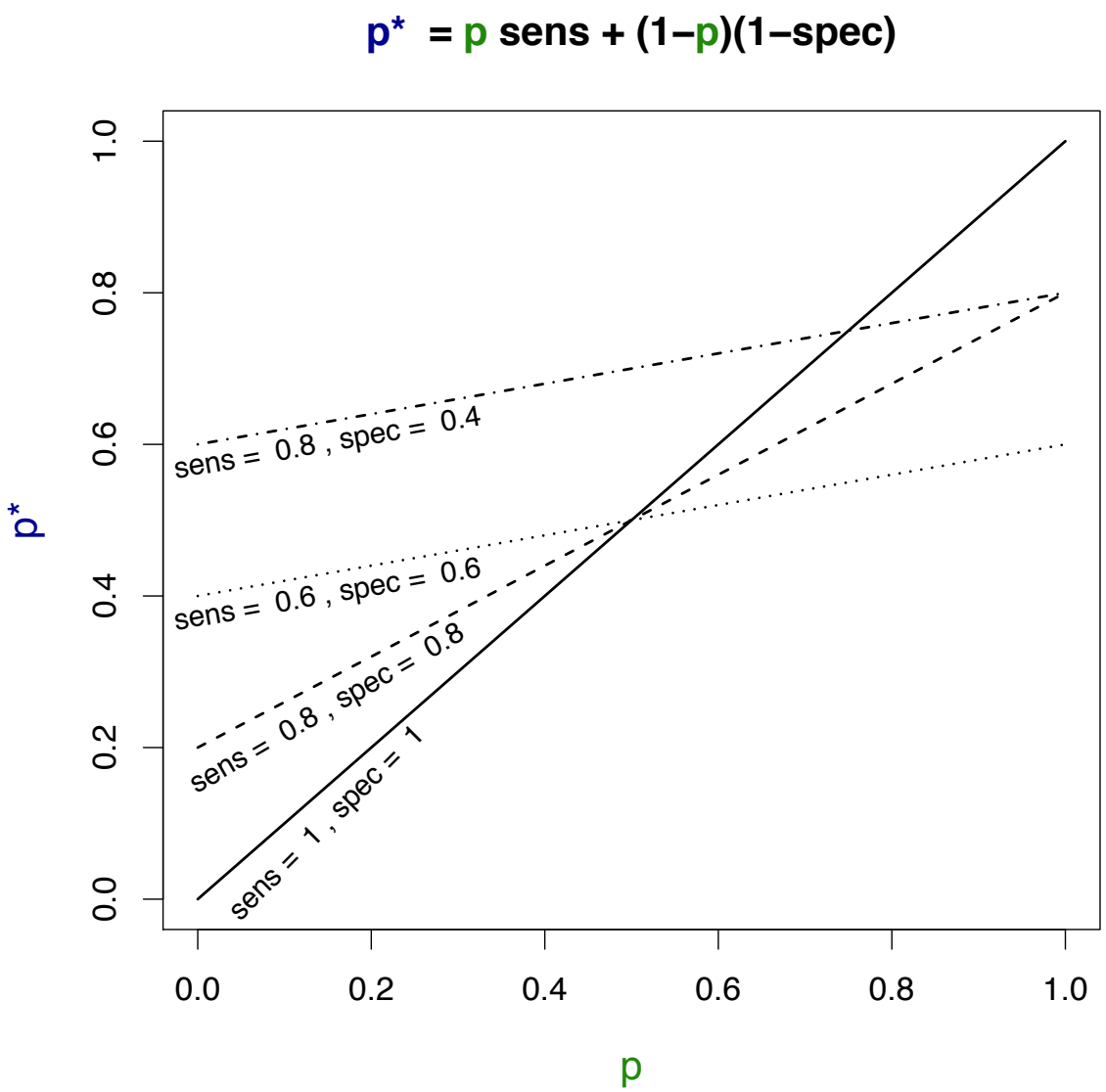
- $P(T = 1)$ konsistent schätzbar
- Löse Gleichung

$$\begin{aligned} P(T=1) &= \pi_{11} \cdot P(\theta = 1) + \pi_{10} \cdot P(\theta = 0) \\ &= \pi_{11} \cdot P(\theta = 1) + \underbrace{\pi_{10}}_{1-\pi_{00}} \cdot (1 - P(\theta = 1)) \end{aligned}$$

  nach $P(\theta = 1)$ auf

- $P(\theta = 1) = \dfrac{P(T = 1) - \pi_{10}}{\pi_{11} + \pi_{00} - 1}$
- Voraussetzungen
  π_{11}, π_{00} bekannt
  $\pi_{11} + \pi_{00} > 1$ (z.B. bei $\pi_{11} > 1/2$, $\pi_{11} > 1/2$; Klassifikation soll besser sein als fälliges klassifizieren))


p* = p sens + (1−p)(1−spec)
p*
p
0.0
0.2
0.4
0.6
0.8
1.0
sens = 1 , spec = 0.4
sens = 1 , spec = 0.6
sens = 1 , spec = 0.8
sens = 1 , spec = 1


p* = p sens + (1−p)(1−spec)
p*
p
0.0
0.2
0.4
0.6
0.8
1.0
sens = 0.8 , spec = 0.4
sens = 0.6 , spec = 0.6
sens = 0.8 , spec = 0.8
sens = 1 , spec = 1

Multinomialer Fall

- Fehlklassifikationswahrscheinlichkeiten

$$\pi_{ij} = P(T = i|\theta = j)$$

- Fehlklassifikationsmatrix Π (bekannt oder konsistent geschätzt)

- $$\begin{pmatrix} P(T=1) \\ P(T=2) \\ \vdots \\ P(T=k) \end{pmatrix} = \Pi \cdot \begin{pmatrix} P(\theta=1) \\ P(\theta=2) \\ \vdots \\ P(\theta=k) \end{pmatrix}$$

- „Matrixmethode“

$$\begin{pmatrix} P(\theta = 1) \\ P(\theta = 2) \\ \vdots \\ P(\theta = k) \end{pmatrix} = \Pi^{-1} \cdot \begin{pmatrix} P(T = 1) \\ P(T = 2) \\ \vdots \\ P(T = k) \end{pmatrix}$$

- nicht immer zielführend, die Schätzungen mancher Komponenten nicht im Intervall $[0; 1]$ liegen müssen.

Erweiterungen

- Bivariater Fall, insbesondere 2×2 Tafeln

- Regressionsmodelle

  * Fehlklassifikationin binärer abhängiger Variable:
    Hausmann et al (Journal of Econometrics, 1998), Neuhaus (Biometrika, 1999)
    Likelihood bei „bekannter“ Fehlklassifikationsmatrix

  * Erweiterung SIMEX-Verfahren: Küchenhoff & Lesaffre (2006, Biometrics)

5.5.3 Fehlende Daten

- Little, R.J.A., Rubin, D.B. (2002): Statistical Analysis with Missing Data, 2nd edition, New York: John Wiley

- Spieß, M. (2009): Missing Data: Analyse von Daten mit fehlenden Werten, Lit Verlag

Grundlegende Begriffe

- unit-nonresponse:
  * keine Information über Fall,
  * vollständige Verweigerung, allgemein
  * „non coverage“,
  * Ausschöpfungsquote

- item-nonresponse: nur Werte bestimmter Variablen fehlen, hier im folgenden

- *Complete Case Analysis:*
  * Analysiere nur die Fälle, bei denen zu allen Variablen die Ausprägungen vorliegen
  * wenn Fehlen nicht zufällig $\Rightarrow$ Verzerrung, z.B. Zensierung
  * auf alle Fälle Effizienzverlust
  * eventuell bleiben nur wenige Fälle übrig

- Available Case Analysis
  benutze jeweils alle Fälle, die bezüglich der jeweils betrachteten Frage vollständig sind

- Imputation
  „Fülle Datanmatrix auf“
  verschiedene Verfahren

- komplexe Schätzverfahren:
  * EM-Algorithmus
  * Bayesianische Methoden
  * Gewichtung (Horvitz-Thompson)

- Fehlendmechanismen bei partiell fehlendem Response Y und Kovariable X

  * betrachte Indikatorvariable $\Delta_i = \begin{cases} 1 & Y_i \text{ beobachtet} \\ 0 & Y_i \text{ nicht beobachtet} \end{cases}$
  * und die Wahrscheinlichkeit $P(\Delta_i = 1)$
  * unterscheide die folgenden Fälle:
    Δ_i hängt
    * $[\alpha)$] weder von X noch von Y
    * $[\beta)$] von X, aber nicht von Y

  * $[\gamma)$] von Y, aber nicht von X
  * $[\delta)$] von Y und von X

  ab.

* bei $\alpha)$ spricht man von Missing completely at random MCAR
  (Beobachtete Daten sind eine echte Zufallsstichprobe aller Daten)
* bei $\beta)$ spricht man von Missing at random: MAR
  Beobachtete Daten sind in jeder bezüglich x gebildeten Klasse eine Zufallsstichprobe
* $\gamma)$ und $\delta)$ Hier fehlen die Daten nicht zufällig; Man spricht von: Missing not at random (MNAR)

- Viele Verfahren, mit fehlenden Daten umzugehen, benötigen MAR
- Was tun bei MCAR?
  * Defizitätsprozess modellieren
  * MAR ist nicht testbar

- „Vorsichtige Analyse fehlender Daten“, Sensitivitätsanalyse
  * Ökonometrie: insbesondere Manski (2003): Partial Identification of Probability Distributions, Springer
  * Biometrie: systematische Sensitivitätsanalyse (Vansteelandt, Goetghebeur, Kenward, Molenberghts. Statistica Sinica, 2006.)
  * Manski's „Law of Decreasing Credibility“
    Die Glaubwürdigkeit statistischer Aussagen sinkt mit der Stärke der Annahmen, auf denen sie beruhen.
  * Gib den Anspruch auf, jede statistische Analyse müsse eine eindeutige Antwort geben
  * Betrachte die Menge $aller$ mit den Daten verträglichen Modelle
  * $Partielle\ \ Identifikation$ statt Punktidentifikation z.B. „Intervallwertige Punktschätzer“ für Parameter oft für substanzwissenschaftlich relevante Aussagen ausreichend.

* unschärfere, aber dafür glaubwürdigere Aussagen
* insbesondere wird deutlich, wie stark gewisse Annahmen die substanzwissenschaftlichen Folgerungen prägen.
* aktueller Forschungsgegenstand