

## **Stichwörter Empirische Forschungsmethoden (Skript Beller):**

Was ist eine Hypothese ? => falsifizierbar

„Vermutete Antwort auf eine Frage über deren Wahrheit man noch nichts weiss“

Arten Variablen kontinuierlich und diskret (dichotom und Polytom)

Manifeste und latente Variablen

Unterscheidung von uV und aV

Moderatorvariablen (Kontroll und Störvariablen)

Konfundiert (uV korreliert systematisch mit Störvariablen)

Grundpositionen:

- Realismus
- Idealismus
- Empirismus
- Rationalismus
- Konstruktivismus
- Pragmatismus
- Konventionalismus
- Skeptizismus

Ablauf einer empirischen Untersuchung

- Fragestellung
- Inhaltliche Hypothese
- Variablen Bestimmung (uVs und aVs)
- Versuchsplan erstellen
- Operationalisierung
- Inhaltliche Hypothese in Statistische Hypothese umwandeln
- Stichprobe ziehen
- Datenerhebung
- Datenauswertung
- Beibehalten oder Verwerfen von  $H_0/H_1$
- Rückschluss auf inhaltliche Hypothese
- Erkenntnisgewinn

## **Datenerhebungsmethoden**

**Messen** => Abbildung eines empirischen Relativs auf ein numerisches Relativ

Repräsentationsproblem (Abbildung auf die richtige Skala)

Eindeutigkeitsproblem (Erlaubte Transformationen)

Bedeutsamkeitsproblem (Bedeutung der Werte nach der Transformation)

Skalenniveaus

- Nominalskala
- Ordinalskala
- Intervallskala
- Verhältnisskala

## **Operationalisierung:**

### **Zählen**

Kategoriensystem (Genauigkeit, Exklusivität, Exhaustivität)

## Urteilen

Direkte Rangordnung  
paarweiser Vergleich

Konsistenz (Transitivität erfüllt?)  
Konkordanz (Urteilsgüte Übereinstimmung verschiedenr Jurymitglieder)

Ratingskalen (Anzahl Stufenm Unipolar Bipolar, Numerisch, verbal, symbolisch, zwang zur Tendenz,

Urteilsfehler:

- Boden und Deckeneffekte
- Auswirkungen ungleicher Merkmalsverteilungen
- Halo-Effekt
- Milde Härte Fehler
- Tendenz zur Mitte
- Rater-Ratee Interaktion
- Primacy Recency Effekt

## Testen

Leistungs und Persönlichkeitstests

### Klassische Testtheorie

5 Axiome:  $X=T+E$ , Messfehler im Durchschnitt Null, Meßfehler unabhängig vom wahren Wert, Meßfehler ist unabhängig von anderen merkmalen der person, Meßfehler verschiedener testanwendungen sind unabhängig voneinander

Testentwicklung:  
Inhalt des Testes,  
Aufgabenpool,  
Itemanalyse (Itemschwierigkeit, Trennschärfe, Homogenität)

Gütekriterien (Objektiv, Reliabel, Valide)

- Objektivität (Durchführungsobjektivität, Auswertungsobjektivität, Interpretationsobjektivität)
- Reliabilität (Retest Reliabilität, Parallelttest Reliabilität, Testhalbierung Reliabilität)
- Validität (Inhalts Validität, Kriteriums Validität, Konstruktvalidität)

Normtabelle erstellen

### Probabilistische Testtheorie

Lösungswahrscheinlichkeit der Items und Itemcharakteristik

Dichotomisches Logistisches Modell (Rasch)

5 Annahmen

- Endliche Menge von Items
- Homogen
- Itemcharakteristiken monoton steigend
- lokale stochastische Unabhängigkeit
- Es ist egal wecle Aufgaben gelöst wurden die Anzahl der gelösten ist entscheidend

Testverfälschungen

- Raten
- Selbstdarstellung (z.B. soz. Erwünschtheit)
- Antworttendenzen (Ja-Sage Tendenz)

## **Befragen**

Interviewvarianten: Informationsermittelnde und Informationsvermittelnd

Einzelinterview, Gruppeninterview, Tandeminterview, Boardinterview (Hearing)

Standardisierung der Fragen (Nicht, Voll und Teilstandardisiert)

Standardisierung der Antworten (Offen Antworten, vorgegebene Antworten (Multiple Choice))

Gute Fragen: einfach, eindeutig, nicht suggestiv, nicht zu allgemein, auf einen Sachverhalt gerichtet, bringen den Befragten nicht in Verlegenheit, pot. beantwortbar, angemessene Antwortvorgabe (z.B. bei MC), keine negativ formulierten Fragen

Durchführungsregeln:

- Sprachtempo, Lautstärke
- nicht unterbrechen bei spontanen berichten
- nie mehrere Fragen gleichzeitig
- Kritik aufnehmen
- non verbales Verhalten beobachten
- Störungen und Ablenkungen vermeiden

Mögliche Probleme:

- Nicht erreichbarkeit
- verweigern der Teilnahme
- Interviewereffekte
- ablehnung bestimmter Fragen
- Antwortverfälschungen (Ja-Sage etc.)

## **Beobachten**

Teilnehmende Beobachtung

Nicht teilnehmende Beobachtung

selbst und Fremdbeobachtung

offene und verdeckte Beobachtung

Mögliche Probleme:

- Überforderung des Beobachters
- Beobachtung vs Interpretations dessen (Fehlinterpretation)
- unvollständiges Kategoriensystem
- Reaktivität (wissen dass man beobachtet wird)

## **Deskriptive Statistik**

Darstellung von Daten (Grafische, tabellarische, Rohdatentabelle)

**Statistische Kennwerte:**

Modalwert (Mo) (Wert der am häufigsten Vorkommt)

Median (Md) (Wert der die Verteilung halbiert)  
Arithmetisches Mittel (m)

### **Masse der Streuung:**

Variationsbreite (Range) (Differenz zwischen größtem und kleinstem vorkommenden Wert)  
Average Deviation (AD) (wie Varianz nur ohne Quadrat dafür aber Betrag)  
Varianz ( $S^2$ ) (Abweichung zum Mittelwert, quadriert, aufsummiert und durch die Anzahl dividiert)  
Standardabweichung (S)

Vergleichbarkeit herstellen durch z-Transformation

### **Suggestive Datenrepräsentation**

- Illusion der Präzision (z.B. Alter)
- Verwechslung von Anzahl und Anteil (Hund der Eier legt)
- Wachstumsturbo (Wachstum vom Wachstum in Prozent)
- Mittelwert ist nicht gleich Mittelwert (Streuung wird vergessen, zB Temperaturen)
- Unzulässige Gruppierung von Daten (Tote und Verletzte)
- Daten auf der Streckbank (DAX)
- von Trends und Trends (Nur Ausschnitt z.B. eines positiven Trends ohne den vorherigen negativen Trend)
- Zahlenverhältnisse und Piktogramme (Kantenlänge bei 2 Dimensionen verdoppeln bedeutet Vervierfachung)

### **Populationsbeschreibende Untersuchungen**

Stichprobenauswahlverfahren (Repräsentativität)

- Zufallsstichprobe
- Geschichtete Stichprobe (zB nur Herzpatienten oder Gesellschaftsschicht)
- Klumpenstichprobe (Schulkassen)
- Ad-hoc Stichprobe (nach Verfügbarkeit)

Punktschätzung: Schätzen eines Populationskennwertes durch eine Stichprobe (Meist Mittelwert oder Standardabweichung)

Stichprobenkennwerte (z.B. Würfeln mit zwei Würfeln und Schnitt berechnen)

Zentrales Grenzwerttheorem: Stichprobe geht mit zunehmendem Umfang in Normalverteilung über ( $n > 30$  bedeutet hinreichend Normalverteilung)

Standardfehler des Mittelwertes (Abweichung der Stichprobenmittelwerte um den Populationsmittelwert, aber Populationsmittelwert nicht verfügbar, deshalb nimmt man  $S^2$ )

### **Gütekriterien für die Punktschätzung:**

- Erwartungstreue (wenn das arithmetische Mittel  $k$  der Stichprobe dem der Population entspricht ( $K$ ))
- Konsistenz (Annäherung an Population bei steigendem Stichprobenumfang, dann Konsistent)
- Effizienz (Wenn es den Kennwert besser schätzt als andere verfügbare Tests, und die Streuung der Kennwerteverteilung kleiner ist)
- Suffizienz (Alle Informationen wurden ausgeschöpft, z.B. Median vs. arithmetisches Mittel)

### **Intervallschätzung**

Konfidenzintervall für das arithmetische Mittel (Bereich in dem der Populationsmittel zu einer Wahrscheinlichkeit liegt)

## **Prinzip des Signifikanztests**

Vergleich von Hypothesen ( $H_0$  und  $H_1$ )

Arten von Hypothesen: Unterschieds und Zusammenhangshypothesen

gerichtete und ungerichtet  
spezifische und unspezifisch

Überführung der inhaltlichen Hypothese in eine statistische Hypothese

Statistische Entscheidung=> Alpha und Betafehler

Je größer die Effektgröße (also je steiler die Glockenkurve) desto kleiner der Alpha Fehler  
Wird der Standardfehler kleiner (also die Kurve wird auch steiler) dann wird der Alpha Fehler auch kleiner  
Alpha und Betafehler verhalten sich gegenläufig

Damit alpha Bereich und Beta Bereich sich nicht überschneiden oder einen Bereich nicht abdecken und somit eine Entscheidung über  $H_0$  oder  $H_1$  unmöglich machen, benötigt man die Poweranalyse.

Poweranalyse (1-beta, oder Gegenwahrscheinlichkeit zum Betafehler)

Man kontrolliert damit den Betafehler

Die Poweranalyse besteht aus 4 Komponenten (alpha Fehler, Betafehler, Effektgröße und Stichprobenumfang) Mit 3en kann man den 4ten berechnen.

Meist wird der optimale Stichprobenumfang berechnet oft auch die Power des Tests (1-Beta) bzw. der Betafehler

## **Hypothesenprüfende Untersuchungen**

### **Vergleich eines Stichprobenwerts mit einem Populationswert**

**z-Test:** Vergleich eines StichprobenMITTELwertes mit einem PopulationsMITTELwert

Signifikanzprüfung: z-Wert Berechnen und aus Tabelle vergleichen mit entsprechendem ALPHA-Niveau

**1-dimensionaler  $\chi^2$  Test:** Vergleich der Häufigkeiten eines k-fach gestuften Merkmals

Signifikanzprüfung:  $\chi$  Wert berechnen und aus Tabelle ( $\chi^2$ -Verteilung) unter Berücksichtigung der Freiheitsgrade und des Alpha-Niveaus ablesen.

Die optimale Stichprobengröße erhält man mit der Poweranalyse

### **Testverfahren für Unterschiedshypothesen**

#### **Das zwei Gruppen- (between-subject-) Design**

eine zweifach gestufte uV (zwei Gruppen) und nur eine aV

interne Validität (keine Störvariablen aV direkt auf uV zurückzuführen)  
externe Validität (Generalisierbarkeit)

Personengebundene Störvariablen kontrollierbar durch: Konstanthalten, parallelisieren, als zusätzlichen uV nehmen oder rechnerisch kontrollieren

Situationsbedingte Störvariablen möglichst eliminieren

**4-Felder Chi<sup>2</sup> Test:** Vergleich der Häufigkeiten zweier 2-fach gestufter Merkmale analog zum 1-dimensionalen, andere Formel zur Berechnung

für k und l fach gestufte Designs gibt es einen k X l Chi<sup>2</sup>-Test

### **t-test**

Vergleich zweier Stichprobenmittelwerte aus unabhängigen Werten

kleine Stichproben t-verteilt (bis n=50) sonst Normalverteilt

kleine Stichproben => population bezgl. Merkmal Normalverteilt  
Varianzen Homogen  
Stichproben unabhängig

Gibt auch t-test für abhängige Gruppen

### **Mehr Gruppen-Between Subject Design (Ungerichtete Hypothese)**

zur Auswertung die Chi<sup>2</sup> Verfahren nehmen (bei nominalskaliert), einfaktorielle Varianzanalyse bei Intervallskaliert und Verhältnisskaliert.

### **Einfaktorielle Varianzanalyse (ungerichtet)**

Bestimmung der Gesamtvarianz  
Bestimmung der Treatmentvarianz  
Bestimmung der Fehlervarianz  
=> F-Verteilung

### **Faktorielle Designs**

Mehrere UVariablen die sich gleichzeitig auf einen AV auswirken  
Einfachstes: Zweifaktorielles (k X l), Zwei Faktoren mit k bzw l Stufen  
=> Haupteffekte, Interaktionseffekte  
=> Auswertung zB mit k mal l Chi<sup>2</sup> Test oder zweifaktorielle Varianzanalyse

### **Zweifaktorielle Varianzanalyse**

=> Messung von Haupteffekt A und B und Interaktionseffekt A X B  
Ablesbar im Diagramm (Gegenläufige Graphen kann man nicht interpretieren)

Mehrfaktorielle Varianzanalyse auch möglich

### **Testverfahren für Zusammenhangshypothesen**

Lediglich 1-Gruppen Design zur Messung von Zusammenhängen zwischen mehreren aVs

Zwei aVs => Korrelationsrechnung

**Produkt Moment Korrelation**, r (wenn beide aVs Intervallskaliert)  
positive Kovariation (0 bis +1), negative Kovariation (-1 bis 0)

r zwischen -1 und +1

je näher r bei 0 desto geringer der gemessene Effekt

Signifikanzprüfung mittel t-Test

Die Stichprobe muss repräsentativ sein, sonst gibt es Verzerrungen d.h. unter bzw übersvhätzung der Korrelation, Ausreisser verzerren das Ergebnis auch

**pHi Koeffizient (bei 2 zweistufigen Nominalskalen, also zB männlich und weiblich)**  
ähnlich wie 4-Felder Chi<sup>2</sup> Test ( $\text{Chi}^2 = \text{Phi}^2 \text{ mal } n$ )

Phi ist ein MAass für den Zusammenhang zwischen der aVs  
Chi benötigt man für den Signifikanztest

Nur ungerichtete Hypothesen da Nominalskaliert

## **Der Untersuchungsbericht**

Inhalte eines Untersuchungsberichtes

- Einleitung
- Untersuchungsmethode (Info über Stichproben, Material, Design, Durchführung)
- Ergebnisse und statistische Auswertung
- Diskussion
- Literaturangaben