

# **Statistische Tests in der Mehrsprachigkeitsforschung**

## **Aufgaben, Anforderungen, Probleme.**

**Florian Frötscher und Demet Özçetin**

florian.froetscher@uni-hamburg.de

SFB 538 Mehrsprachigkeit  
Max-Brauer-Allee 60  
22765 Hamburg



Universität  
Hamburg



März 2007

# Überblick

## 1. **Erklärung einiger Grundbegriffe**

1.1 Bereiche der Statistik

1.2 Hypothesentestung

1.3 Variablen

1.4 Skalenniveau

## 2. **Statistische Verfahren**

2.1 Entscheidungsbäume

2.2 QMwiki

## 3. **Ein Beispiel**

## 4. **Schwierigkeiten bei linguistischen Daten**

# 1.1 Bereiche der Statistik

- Deskriptive Statistik:

Methoden zur Beschreibung von Daten – Tabellen, Grafiken, einzelne Kennwerte

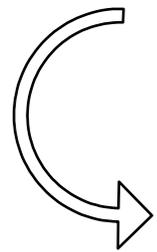
- Inferenzstatistik, schließende St.:

Überprüfung der Richtigkeit von formulierten Hypothesen aufgrund von empirischen Daten

- Explorative Statistik:

(oft auch als Teil der Inferenzst. gesehen)

# 1.2 Hypothesentestung



H<sub>0</sub>: Nullhypothese

H<sub>1</sub>: Forschungshypothese = Alternativhypothese

# 1.2 Hypothesentestung

## **Signifikanz**

Die Unterschiede oder Zusammenhänge in einer Untersuchung können als signifikant bezeichnet werden, wenn sie nur mit einer gewissen maximalen Wahrscheinlichkeit (als Signifikanzniveau, meist 5%, festgelegt) zufällig zustande gekommen sein können

aber: S. ist von verschiedenen Einflussfaktoren abhängig  
→ ist ein signifikantes Ergebnis auch inhaltlich relevant?

# 1.3 Variablen

## **unabhängige Variable (uV)**

- Prädiktor, erklärende Variable

## **abhängige Variable (aV)**

- Kriterium, Zielvariable

In einem Experiment wird untersucht, wie sich systematisch variierte uV auf die Ausprägungen der aV auswirken.

# 1.4 Skalenniveau

Warum eigentlich ein Skalen- oder Messniveau?

**„Messen“ ist die Zuordnung von Symbolen (Zahlen) zu beobachtbaren Merkmalen von Objekten nach definierten Regeln.**

Wie werden beobachtete Zusammenhänge abgebildet?

Welche Beziehungen bestehen zwischen den Ausprägungen der entstandenen Variablen?

Welche Operationen und Aussagen sind mit der Variablen möglich?

# 1.4 Skalenniveau

## **Qualitative Daten - Kategorial**

### **Nominalskala:**

unterschiedliche Zahlen sind unterschiedliche Merkmalsausprägungen

Beispiel: Geschlecht, Nationalitäten

zugelassene Operationen: Häufigkeiten auszählen und vergleichen; keine direkten „mehr-weniger“ Aussagen möglich, keine Mittelwerte o.ä.

# 1.4 Skalenniveau

## Quantitative Daten - Kategorial

### Ordinalskala:

Zahlen drücken eine Rangfolge aus, ohne Aussage über Abstände zwischen den Ausprägungen

Beispiel: Präferenzreihenfolge, Zieleinlauf

zugelassene Operationen: Median, „größer-kleiner“  
Aussagen möglich, keine Mittelwerte

# 1.4 Skalenniveau

## **Quantitative Daten - Kardinalskalen**

### **Intervallskala:**

Wie Ordinalskala, nur dass sich die Abstände zwischen den Klassen begründen lassen

Beispiel: Temperatur in Celsius, IQ-Werte

zugelassene Operationen: zusätzlich Mittelwerte, Streuungen

# 1.4 Skalenniveau

## **Quantitative Daten - Kardinalskalen**

### **Verhältnisskala:**

Wie Intervallskala, zusätzlich natürlicher Nullpunkt

Beispiel: cgs-System

zugelassene Operationen: z.B. „doppelt so...“, „halb so...“

# Überblick

1. **Erklärung einiger Grundbegriffe**
  - 1.1 Bereiche der Statistik
  - 1.2 Hypothesentestung
  - 1.3 Variablen
  - 1.4 Skalenniveau
2. **Statistische Verfahren**
  - 2.1 Entscheidungsbäume
  - 2.2 QMwiki
3. **Ein Beispiel**
4. **Schwierigkeiten bei linguistischen Daten**

# Überblick

1. **Erklärung einiger Grundbegriffe**
  - 1.1 Bereiche der Statistik
  - 1.2 Hypothesentestung
  - 1.3 Variablen
  - 1.4 Skalenniveau
2. **Statistische Verfahren**
  - 2.1 Entscheidungsbäume
  - 2.2 QMwiki
3. **Ein Beispiel**
4. **Schwierigkeiten bei linguistischen Daten**

# 2. Statistische Verfahren

Forschungsfrage

» statistische Fragestellung

+ Daten

---

geeignetes statistisches Verfahren

# 2.1 Entscheidungsbaum

**Überblick über einige statistische Verfahren:**  
der statistische Entscheidungsbaum

(nach Universität Bern

<http://www.vislabs.ch/Lehre/EST/est.html>)

# Hypothesen prüfen

# Hypothesen entdecken

## Unterschiede

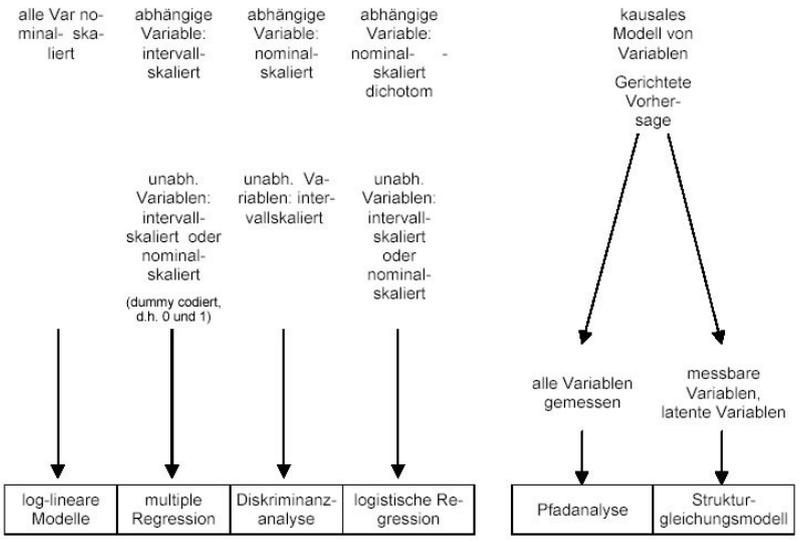
## Zusammenhänge Abhängigkeit

gegeben: mehrere Variablen, Korrelationsmatrix  
gegeben: Ähnlichkeit zwischen Personen und Objekten  
gegeben: mehrere Variablen Skalenniveau von uV, aV beliebig

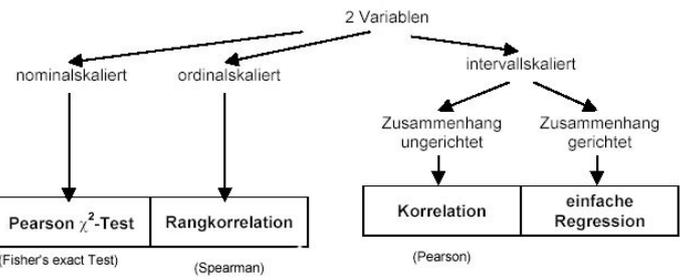
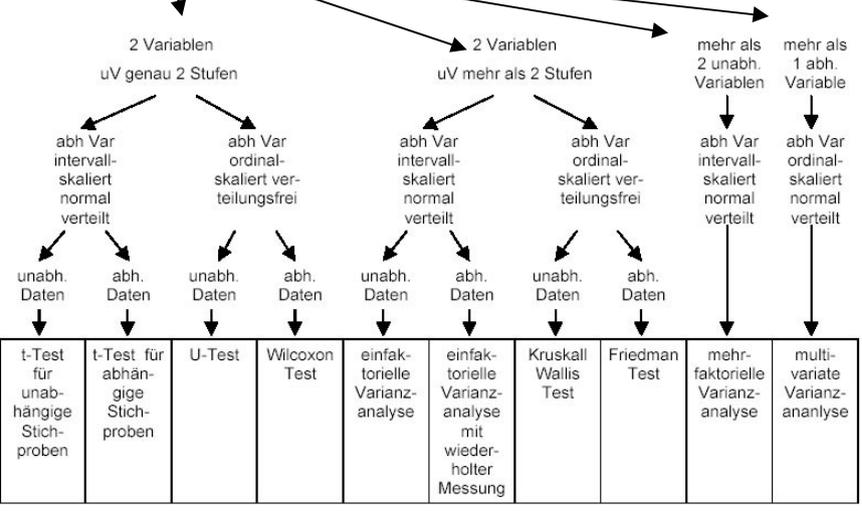
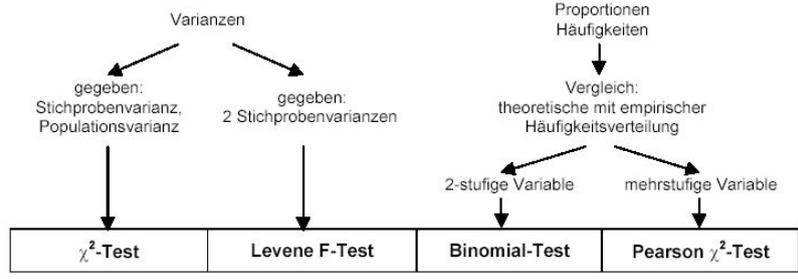
Ziel: Bündelung, Reduktion der Variablen  
Ziel: Gruppierung der Personen oder Objekte  
Ziel: Modell für abhängige Variable  
(Lösung für Diskriminanzanalyse, logistische Regression, Clusteranalyse)

Reduktion auf Skalen  
**Faktorenanalyse**  
**Clusteranalyse**  
**AnswerTree**

mehr als 2 Variablen



**Zentrale Tendenz**  
Mittelwerte;  
Mediane



HINWEIS

# STATISTISCHER ENTSCHEIDUNGSBAUM

## Hypothesen prüfen

### Unterschiede

#### Varianz

- geg.: Stichproben- und Populationsvarianz
- geg.: 2 Stichprobenvarianzen

#### Zentrale Tendenz

- 2 Variablen, uV = 2 Stufen
  - aV intervallskaliert, NV
    - unabh. Stichproben
    - abh. Stichproben
  - aV ordinalskaliert, verteilungsfrei
    - unabh. Stichproben
    - abh. Stichproben
- 2 Variablen, uV > 2 Stufen
  - aV intervallskaliert, NV
    - unabh. Stichproben
    - abh. Stichproben
  - aV ordinalskaliert, verteilungsfrei
    - unabh. Stichproben
    - abh. Stichproben
- mehr als eine uV
  - aV intervallskaliert, NV
- mehr als eine aV
  - aV ordinalskaliert, verteilungsfrei

#### Häufigkeiten/Proportionen Vergleich: theoretische mit empirischer Verteilung

- 2-stufige Variable
  - theoretisch: entspr. der emp. Verteilung
- mehrstufige Variable
  - theoretisch: Normalverteilung

### Zusammenhang/Abhängigkeiten

#### 2 Variablen

- nominalskaliert
- ordinalskaliert
- intervallskaliert
  - Zusammenhang ungerichtet
  - Zusammenhang gerichtet

#### >2 unabhängige Variablen

- alle Var. nominalskaliert
- aV: intervallskaliert  
uVs: intervall- oder nominalskaliert (dummy-kodiert)
- aV: nominalskaliert  
uVs: intervallskaliert
- aV: nominal, dichotom  
uVs: intervall- oder nominalskaliert
- kausales Modell von Vorhersagen  
gerichtete Vorhersage

## Hypothesen generieren Strukturen entdecken

gegeben: mehrere Variablen, Korrelationsmatrix  
Ziel: Bündelung, Reduktion der Variablen

gegeben: Ähnlichkeit zwischen Personen und Objekten  
Ziel: Gruppierung der Personen oder Objekte

gegeben: mehrere Variablen  
Skalenniveau uV, aV beliebig  
Ziel: Modell für abhängige Variable  
(Lösung für DA, IR, CLA)

# 2.1 Entscheidungsbaum

Eine Auswahl der zu beurteilenden Fragen:  
Hypothesen prüfen oder Hypothesen entdecken?

# 2.1 Entscheidungsbaum

Eine Auswahl der zu beurteilenden Fragen:

**Hypothesen prüfen** oder Hypothesen entdecken?

- Unterschiede oder Zusammenhänge?
- Unterschiede in... (zentraler Tendenz, Häufigkeiten,...)?
- Anzahl (und Stufen) der Variablen (uV/aV)?
- Skalenniveau der Variablen?
- Größe der Stichprobe?

# 2.1 Entscheidungsbaum

Eine Auswahl der zu beurteilenden Fragen:

**Hypothesen prüfen** oder Hypothesen entdecken?

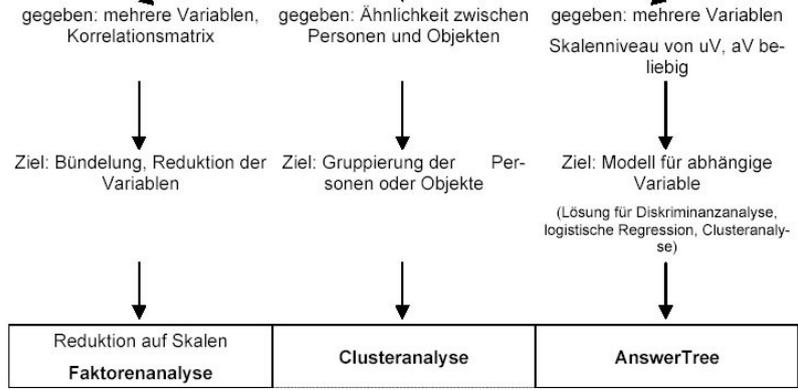
- **Unterschiede** oder Zusammenhänge?
- Unterschiede in... (**zentraler Tendenz**, Häufigkeiten,...)?
- Anzahl (und Stufen) der Variablen (uV/aV)?
- Skalenniveau der aV?
- Größe der Stichprobe?

# Hypothesen prüfen

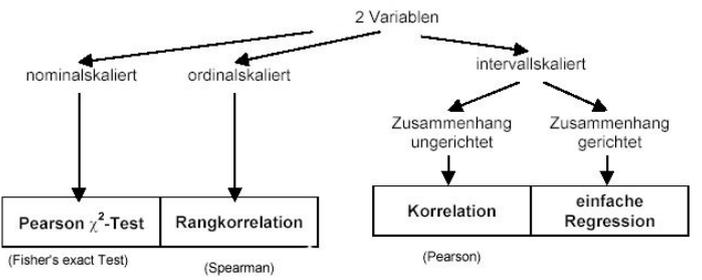
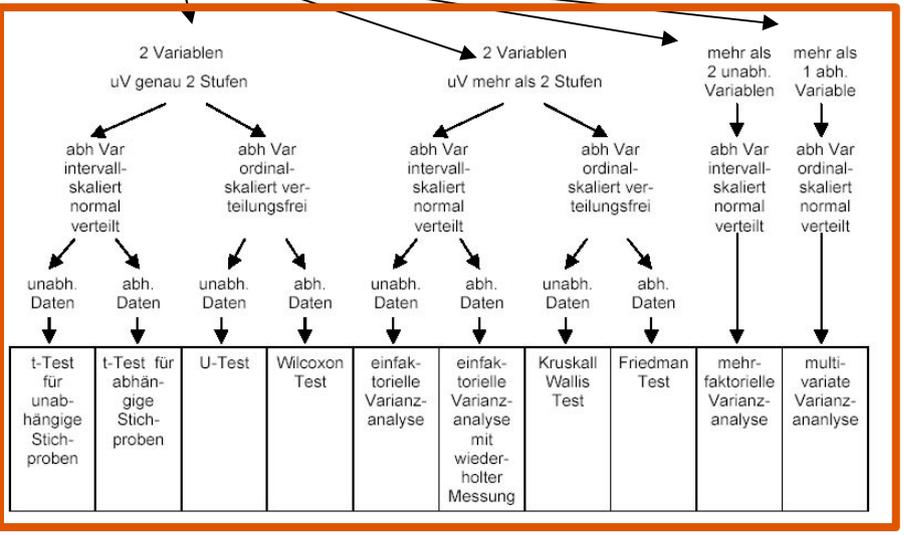
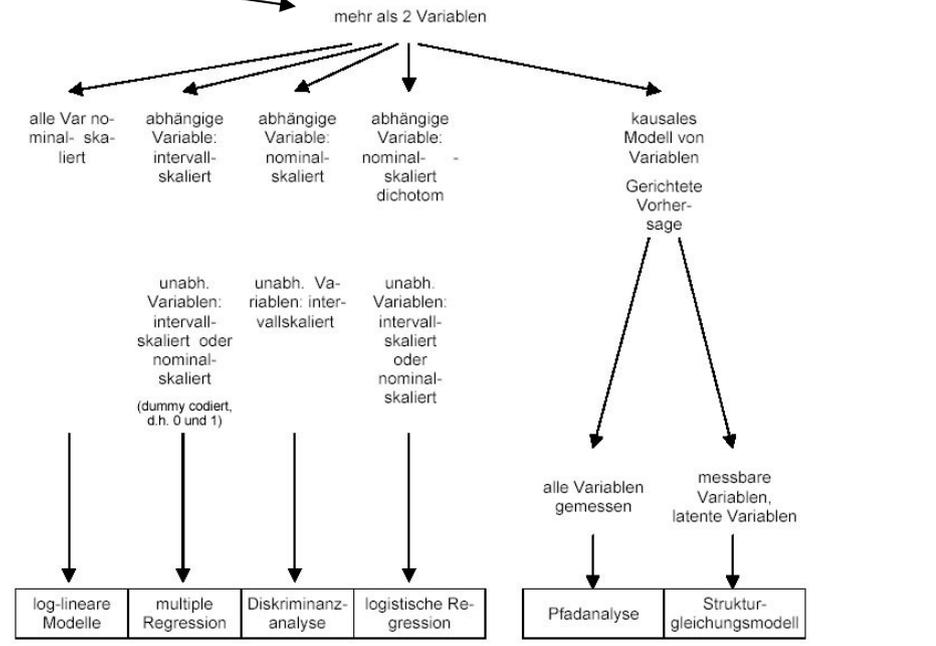
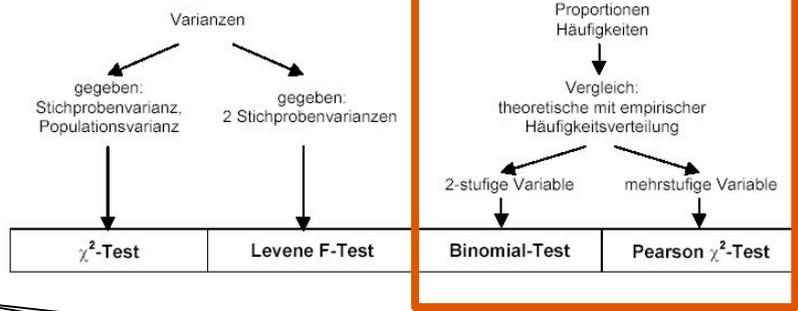
# Hypothesen entdecken

## Unterschiede

## Zusammenhänge Abhängigkeit



**Zentrale Tendenz**  
Mittelwerte;  
Mediane



# 2.1 Entscheidungsbaum

Eine Auswahl der zu beurteilenden Fragen:

Hypothesen prüfen oder **Hypothesen entdecken?**

- mehrere mindestens intervallskalierte Variablen können mit Hilfe von Faktoren- und Reliabilitätsanalyse gebündelt und reduziert werden
- Ähnlichkeiten (Distanzen) zwischen Objekten und Personen können mittels Clusteranalyse in Gruppen gebündelt werden

# 2.1 Entscheidungsbaum

Wichtig: Testvoraussetzungen beachten

Sonst ist Interpretation der Ergebnisse nicht sinnvoll und nicht zulässig

Viele linguistische Daten sind für die geläufigen statistischen Verfahren nicht geeignet.

# 2.1 Entscheidungsbaum

**Problem:** Existierende Übersichten...

- beinhalten nicht Verfahren, die für den linguistischen Bereich weiter entwickelt wurden
- sind nicht wirklich übersichtlich
- sind statisch, nicht erweiterbar
- sind schwer verständlich

## 2.2 OMwiki

Projektbereichsübergreifendes Arbeitsprojekt (Pbü-Projekt):

**Quantitative Methoden in der Mehrsprachigkeit**

## 2.2 OMwiki

### **Ziel / Inhalte:**

Systematisierende Darstellung unterschiedlicher mathematischer Verfahren

- Anwendungsfeld
- Anwendungsvoraussetzungen
- Ggf. Softwarewerkzeuge
- Linguistische Beispiele und Beispielstudien

## 2.2 OMwiki

### **Darstellungsmedium:** Web/Wiki

- Leichte Online-Verfügbarkeit
- Hypertext-Technik: Flexible Darstellungsform für vernetzte Inhalte / Baumstrukturen
- Leichte und langfristige Bearbeitung im Team möglich

## 2.2 OMwiki

### **Stand der Arbeit:**

- Wiki-Grundgerüst angelegt mit einigen speziellen Erweiterungen zur Darstellung der Baumstruktur
- Wichtigste grundlegende Verfahren mit Entscheidungswegen eingefügt
- Beschreibungen und Beispiele müssen noch vervollständigt werden
- im Laufe dieses Jahres soll das Wiki breiter nutzbar sein



# Hauptseite

## QMwiki

Ein Wiki zu Quantitativen Methoden in der Linguistik

herzlich willkommen.

[\[bearbeiten\]](#)

### Navigation

- **Entscheidungsbaum**: Index statistischer Verfahren, gegliedert nach einer Klassifikation typischer Fragestellungen
- **Mindmap** [↗](#): Dynamisches Schaubild derselben Klassifikation, wird nur in unregelmäßigen Abständen aktualisiert
- **alphabetischer Index** der wichtigsten statistischen Verfahren
- **Glossar**
- **Literaturliste**
- **Anwendungs- und Bearbeitungshinweise** finden sich in der **QMwiki:Hilfe**

Erstellt und in Arbeit durch die **PBÜ ARBEITSGRUPPE QUANTITATIVE METHODEN** des Sonderforschungsbereichs Mehrsprachigkeit (SFB 538 [↗](#)) der Universität Hamburg [↗](#)

Mitarbeit: Jasmine Bennöhr - Martin Elsig - Florian Frötscher - Lukas Pietsch - Tülay Selcuk

**Kontakt:** [QMwiki@uni-hamburg.de](mailto:QMwiki@uni-hamburg.de)

#### Navigation

- [Hauptseite](#)
- [Aktuelle Ereignisse](#)
- [Letzte Änderungen](#)
- [Zufälliger Artikel](#)
- [Hilfe](#)
- [Literatur](#)
- [Mindmap](#)
- [Glossar](#)

#### Suche

[Los](#)

[Suche](#)

#### Werkzeuge

- [Was zeigt hierhin](#)
- [Verlinkte Seiten](#)
- [Hochladen](#)
- [Spezialseiten](#)
- [Druckversion](#)
- [Permanentlink](#)

## 2.2 QMwiki

### Navigation

---

- **Entscheidungsbaum**: Index statistischer Verfahren, gegliedert nach einer
  - **Mindmap** : Dynamisches Schaubild derselben Klassifikation, wird nur in u
  - **alphabetischer Index** der wichtigsten statistischen Verfahren
  - **Glossar**
  - **Literaturliste**
- 
- Anwendungs- und Bearbeitungshinweise finden sich in der **QMwiki:Hilfe**

HINWEIS

# STATISTISCHER ENTSCHEIDUNGSBAUM

## Hypothesen prüfen

### Unterschiede

#### Varianz

- geg.: Stichproben- und Populationsvarianz
- geg.: 2 Stichprobenvarianzen

#### Zentrale Tendenz

- 2 Variablen, uV = 2 Stufen
  - aV intervallskaliert, NV
    - unabh. Stichproben
    - abh. Stichproben
  - aV ordinalskaliert, verteilungsfrei
    - unabh. Stichproben
    - abh. Stichproben
- 2 Variablen, uV > 2 Stufen
  - aV intervallskaliert, NV
    - unabh. Stichproben
    - abh. Stichproben
  - aV ordinalskaliert, verteilungsfrei
    - unabh. Stichproben
    - abh. Stichproben
- mehr als eine uV
  - aV intervallskaliert, NV
- mehr als eine aV
  - aV ordinalskaliert, verteilungsfrei

#### Häufigkeiten/Proportionen

- Vergleich: theoretische mit empirischer Verteilung
  - 2-stufige Variable
    - theoretisch: entspr. der emp. Verteilung
  - mehrstufige Variable
    - theoretisch: Normalverteilung

### Zusammenhang/Abhängigkeiten

#### 2 Variablen

- nominalskaliert
- ordinalskaliert
- intervallskaliert
  - Zusammenhang ungerichtet
  - Zusammenhang gerichtet

#### >2 unabhängige Variablen

- alle Var. nominalskaliert
- aV: intervallskaliert  
uVs: intervall- oder nominalskaliert (dummy-kodiert)
- aV: nominalskaliert  
uVs: intervallskaliert
- aV: nominal, dichotom  
uVs: intervall- oder nominalskaliert
- kausales Modell von Vorhersagen  
gerichtete Vorhersage

## Hypothesen generieren Strukturen entdecken

gegeben: mehrere Variablen, Korrelationsmatrix  
Ziel: Bündelung, Reduktion der Variablen

gegeben: Ähnlichkeit zwischen Personen und Objekten  
Ziel: Gruppierung der Personen oder Objekte

gegeben: mehrere Variablen  
Skalenniveau uV, aV beliebig  
Ziel: Modell für abhängige Variable  
(Lösung für DA, IR, CLA)

# Überblick

1. **Erklärung einiger Grundbegriffe**
  - 1.1 Bereiche der Statistik
  - 1.2 Hypothesentestung
  - 1.3 Variablen
  - 1.4 Skalenniveau
2. **Statistische Verfahren**
  - 2.1 Entscheidungsbäume
  - 2.2 QMwiki
3. **Ein Beispiel**
4. **Schwierigkeiten bei linguistischen Daten**

# Überblick

1. **Erklärung einiger Grundbegriffe**
  - 1.1 Bereiche der Statistik
  - 1.2 Hypothesentestung
  - 1.3 Variablen
  - 1.4 Skalenniveau
2. **Statistische Verfahren**
  - 2.1 Entscheidungsbäume
  - 2.2 QMwiki
3. **Ein Beispiel**
4. **Schwierigkeiten bei linguistischen Daten**

# 3 , Beispiel

Ein Beispiel (G. Doetjes)

- Sprachverständnis zwischen Schweden und Dänen und Effekte 9 Varianten sprecherseitiger Anpassungen
- Ca. 450 Probanden aufgeteilt in 2x9 Gruppen verschiedener Anpassungsbedingungen
- Ergebnisse eines Textverständnistests

# 3 . Beispiel

- 2 uVs      uV1: L1, 2-fach gestuft  
                  uV2: Sprachbedingung, 9-fach gestuft
- 1 aV      Textverständnistest, verhältnisskaliert
- Hypothese: Unterschied



# Hauptseite

## QMwiki

Ein Wiki zu Quantitativen Methoden in der Linguistik

herzlich willkommen.

[\[bearbeiten\]](#)

### Navigation

- **Entscheidungsbaum:** Index statistischer Verfahren, gegliedert nach einer Klassifikation typischer Fragestellungen
- **Mindmap** [↗](#): Dynamisches Schaubild derselben Klassifikation, wird nur in unregelmäßigen Abständen aktualisiert
- **alphabetischer Index** der wichtigsten statistischen Verfahren
- **Glossar**
- **Literaturliste**
- **Anwendungs- und Bearbeitungshinweise** finden sich in der [QMwiki:Hilfe](#)

Erstellt und in Arbeit durch die **PBÜ ARBEITSGRUPPE QUANTITATIVE METHODEN** des Sonderforschungsbereichs Mehrsprachigkeit (SFB 538 [↗](#)) der Universität Hamburg [↗](#)

Mitarbeit: Jasmine Bennöhr - Martin Elsig - Florian Frötscher - Lukas Pietsch - Tülay Selcuk

**Kontakt:** [QMwiki@uni-hamburg.de](mailto:QMwiki@uni-hamburg.de)

#### Navigation

- [Hauptseite](#)
- [Aktuelle Ereignisse](#)
- [Letzte Änderungen](#)
- [Zufälliger Artikel](#)
- [Hilfe](#)
- [Literatur](#)
- [Mindmap](#)
- [Glossar](#)

#### Suche

[Los](#)

[Suche](#)

#### Werkzeuge

- [Was zeigt hierhin](#)
- [Verlinkte Seiten](#)
- [Hochladen](#)
- [Spezialseiten](#)
- [Druckversion](#)
- [Permanentlink](#)



# Baum

Dies ist ein Index, der eine Anzahl üblicher statistischer Verfahren für unterschiedliche Zwecke anhand eines Entscheidungsbaums erschließt. Sie können den für Ihre Zwecke geeigneten Test finden, indem Sie anhand der Baumstruktur die Art Ihrer Fragestellungen und Daten schrittweise genauer eingrenzen.

Statistische Verfahren lassen sich generell danach einteilen, ob sie bestehende Hypothesen über eine Grundgesamtheit überprüfen (*konfirmatorische Statistik*) oder ob sie Hypothesen (Modelle) über eine Grundgesamtheit generieren (*explorative Statistik*).

## Hypothesen prüfen

Das statistische Test soll eine bereits gefasste Hypothese überprüfen.

[Detaillierte Erklärung](#)

[Baum erweitern](#)

## Hypothesen generieren

Das statistische Verfahren soll neue Hypothesen aus den Daten ableiten.

► [Detaillierte Erklärung](#) ▼ [Baum erweitern](#)

### Navigation

- [Hauptseite](#)
- [Aktuelle Ereignisse](#)
- [Letzte Änderungen](#)
- [Zufälliger Artikel](#)
- [Hilfe](#)
- [Literatur](#)
- [Mindmap](#)
- [Glossar](#)

### Suche

Los

Suche

### Werkzeuge

- [Was zeigt hierhin](#)
- [Verlinkte Seiten](#)
- [Hochladen](#)
- [Spezialseiten](#)
- [Druckversion](#)
- [Permanentlink](#)



## Navigation

- [Hauptseite](#)
- [Aktuelle Ereignisse](#)
- [Letzte Änderungen](#)
- [Zufälliger Artikel](#)
- [Hilfe](#)
- [Literatur](#)
- [Mindmap](#)
- [Glossar](#)

## Suche

## Werkzeuge

- [Was zeigt hierhin](#)
- [Verlinkte Seiten](#)
- [Hochladen](#)
- [Spezialseiten](#)
- [Druckversion](#)
- [Permanentlink](#)

## Erläuterungen

Beim "Hypothesen prüfen" werden zuerst Eigenschaften einer Population (Grundgesamtheit) postuliert, und dann überprüft man, inwieweit die postulierten Eigenschaften der Population durch die Stichproben-Daten bestätigt werden können ([Bortz, 1999](#)).

Bei diesem confirmatorischen Vorgehen müssen vor der Datenerhebung die zu testenden Hypothesen, das Testverfahren und auch das Signifikanzniveau festgelegt werden ([Neuhäuser, 2000](#)).

Synonyme: Induktive Statistik, schliessende Statistik, confirmatorische Statistik, Inferenzstatistik ([Rönz & Strohe, 1994](#)).

## Beispiele

Ist der durchschnittliche Lernerfolg bei Schulkindern, die mit einer neuen Lehrmethode unterrichtet wurden besser als derjenige von Schulkindern, die mit einer konventionellen Lehrmethode unterrichtet wurden? ([Bortz, 1999](#))

Sind in einem Korpus von amerikanischem Englisch die Sätze von Regierungstexten länger als die anderen Texte? ([Oakes, 1998](#))

## Weitere Schritte

### Unterschiede

Der Test soll prüfen, ob zwischen zwei Stichproben Unterschiede bestehen.

- ▶ [Detaillierte Erklärung](#) ▼ [Baum erweitern](#)

### Zusammenhänge

Der Test soll prüfen, ob es zwischen Variablen Zusammenhänge/Abhängigkeiten gibt.

- ▶ [Detaillierte Erklärung](#) ▼ [Baum erweitern](#)

- Was zeigt hierhin
- Verlinkte Seiten
- Hochladen
- Spezialseiten
- Druckversion
- Permanentlink

Ist der durchschnittliche Lernerfolg bei Schulkindern, die mit einer neuen Lehrmethode unterrichtet wurden besser als derjenige von Schulkindern, die mit einer konventionellen Lehrmethode unterrichtet wurden? ([Bortz, 1999](#))

Sind in einem Korpus von amerikanischem Englisch die Sätze von Regierungstexten länger als die anderen Texte? ([Oakes, 1998](#))

## Weitere Schritte

### Unterschiede

Der Test soll prüfen, ob zwischen zwei Stichproben Unterschiede bestehen.

► [Detaillierte Erklärung](#) ▲ [Ausblenden](#)

#### Zentrale Tendenz

Stichproben sollen auf ihre zentrale Tendenz hin untersucht werden.

► [Detaillierte Erklärung](#) ▼ [Baum erweitern](#)

#### Varianzen

Stichproben sollen anhand ihrer Varianzen verglichen werden.

► [Detaillierte Erklärung](#) ▼ [Baum erweitern](#)

#### Häufigkeiten/Proportionen

Es soll eine erwartete mit einer beobachteten Häufigkeit verglichen werden.

► [Detaillierte Erklärung](#) ▼ [Baum erweitern](#)

### Zusammenhänge

Der Test soll prüfen, ob es zwischen Variablen Zusammenhänge/Abhängigkeiten gibt.

► [Detaillierte Erklärung](#) ▼ [Baum erweitern](#)

- Was zeigt hierhin
- Verlinkte Seiten
- Hochladen
- Spezialseiten
- Druckversion
- Permanentlink

Ist der durchschnittliche Lernerfolg bei Schulkindern, die mit einer neuen Lehrmethode unterrichtet wurden besser als derjenige von Schulkindern, die mit einer konventionellen Lehrmethode unterrichtet wurden? (Bortz, 1999)

Sind in einem Korpus von amerikanischem Englisch die Sätze von Regierungstexten länger als die anderen Texte? (Oakes, 1998)

## Weitere Schritte

---

### Unterschiede

Der Test soll prüfen, ob zwischen zwei Stichproben Unterschiede bestehen.

▶ [Detaillierte Erklärung](#) ▲ [Ausblenden](#)

#### Zentrale Tendenz

Stichproben sollen auf ihre zentrale Tendenz hin untersucht werden.

▶ [Detaillierte Erklärung](#) ▲ [Ausblenden](#)

#### 2 Variablen, unabhängige Variable mit 2 Stufen

Es gibt zwei Variablen, die unabhängige Variable besitzt 2 Stufen.

▶ [Detaillierte Erklärung](#) ▼ [Baum erweitern](#)

#### Mehr als 2 Stufen

Die unabhängige Variable hat mehr als 2 Stufen

▶ [Detaillierte Erklärung](#) ▼ [Baum erweitern](#)

#### Mehr als eine unabhängige Variable

Es gibt mehr als eine unabhängige Variable

▶ [Detaillierte Erklärung](#) ▲ [Ausblenden](#)

#### Abhängige Variable intervallskaliert

Die abhängige Variable ist intervallskaliert und normalverteilt.

▶ [Detaillierte Erklärung](#) ▶ [Mehrfaktorielle ANOVA](#)

#### Mehr als eine abhängige Variable

Es gibt mehr als eine abhängige Variable

#### Navigation

- [Hauptseite](#)
- [Aktuelle Ereignisse](#)
- [Letzte Änderungen](#)
- [Zufälliger Artikel](#)
- [Hilfe](#)
- [Literatur](#)
- [Mindmap](#)
- [Glossar](#)

#### Suche

#### Werkzeuge

- [Was zeigt hierhin](#)
- [Verlinkte Seiten](#)
- [Hochladen](#)
- [Spezialseiten](#)
- [Druckversion](#)
- [Permanenlink](#)

## Beschreibung des Verfahrens

Die **Mehrfaktorielle Varianzanalyse** (*multifactorial analysis of variance*) überprüft, wie eine mindestens intervallskalierte abhängige Variable von zwei oder mehr unabhängigen Variablen (Faktoren) beeinflusst wird. Dabei wird berücksichtigt, dass nicht nur die einzelnen Faktoren für sich eine Auswirkung auf die aV haben können, sondern auch die Faktoren kombiniert im Sinne einer Interaktion (Wechselwirkung).

Einzelvergleiche können im Anschluss zeigen, zwischen welchen Faktorstufen im Einzelnen signifikante Unterschiede bestehen.

## statistische Voraussetzungen

- **eine** aV, intervallskaliert und (annähernd) normalverteilt
- **mehrere** nominalskalierte uVs, die unterschiedliche Anzahlen von Stufen haben können
- ähnliche Stichprobengrößen aus unabhängigen Erhebungen und gleiche Varianzen für jede Faktorstufenkombination

Bei ähnlich großen Stichproben mit einem Umfang von mindestens 10 Objekten pro Faktorstufenkombination fallen Verletzungen der Voraussetzungen Normalverteilung und Varianzhomogenität weniger ins Gewicht und können dann vernachlässigt werden. Falls Zweifel bestehen, dass eine normalverteilte Grundgesamtheit vorliegt, lässt sich dies mit statistischen Tests überprüfen (Bortz, 1999).

## Anwendung

SPSS: Analysieren > Allgemeines Lineares Modell > Univariat

## Beispiele

### allgemeine Beispiele

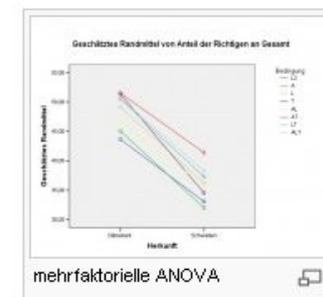
Welchen Einfluss haben Körpergröße, Händigkeit und Trainingsmethode (= 3 uV) auf die Leistung (= aV) (gemessen an der Anzahl Treffer) bei BasketballspielerInnen?

### Beispiele aus der Linguistik

Welchen Einfluss auf das Grammatikalitätsurteil bei W-Fragen mit vorangestelltem W-Element haben der Einschub verschiedener intervenierender Elemente nach der W-Phrase (Faktor A, 3-stufig) bzw. Subjekt- oder Objekteigenschaft dieses Elements (Faktor B, 2-stufig; Adli, 2004)?

Welchen Einfluss auf das Grammatikalitätsurteil in Fragesätzen des Französischen haben die syntaktische Bewegung (Faktor A, 2-stufig) verschiedener Elemente in unterschiedlichen syntaktischen Konstruktionen (Faktor B, 2-stufig; Adli, 2004).

Gibt es in der Kommunikation zwischen Dänen und Schweden einen Unterschied im gegenseitigen Verständnis, wenn die Sprecher ihr Verhalten auf verschiedene Weisen anpassen? Unabhängige Variablen: 9 Bedingungen von Anpassungsverhalten (Ausssprache, Tempo, Lexikon - einzeln und kombiniert) - abhängige Variable: Ergebnis eines Textverständnistests (Doetjes, in Vorbereitung)



## Verwandte Verfahren

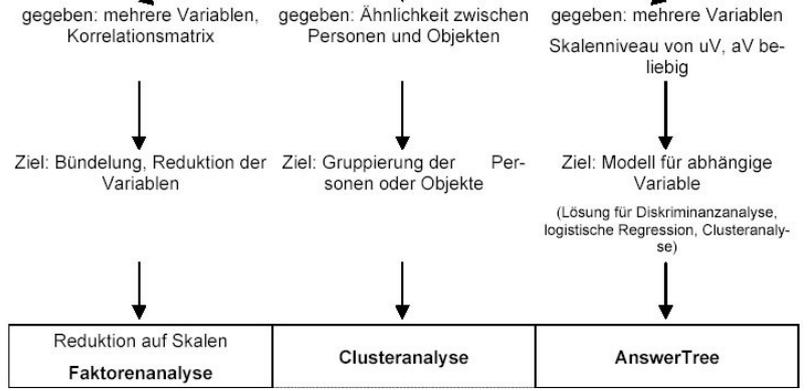
- **einfaktorielle ANOVA**: einfaktorielle Varianzanalyse, wenn also nur Gruppen verglichen werden, die sich aus den Stufen einer uV bilden
- **einfaktorielle ANOVA mit Messwiederholung**: einfaktorielle Varianzanalyse, wobei die Gruppen voneinander abhängig sind, z.B. durch wiederholte Messungen an der gleichen Stichprobe
- **MANOVA**: multivariate Varianzanalyse, wenn Stichproben nicht anhand einer, sondern mehrerer aVs verglichen werden

# Hypothesen prüfen

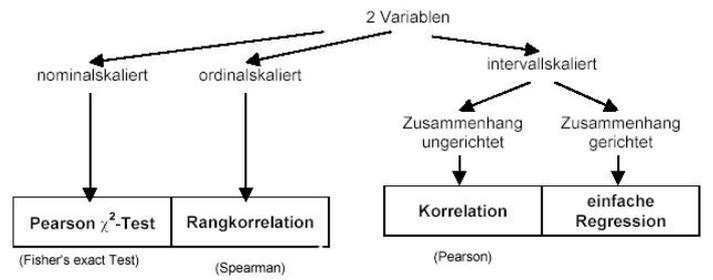
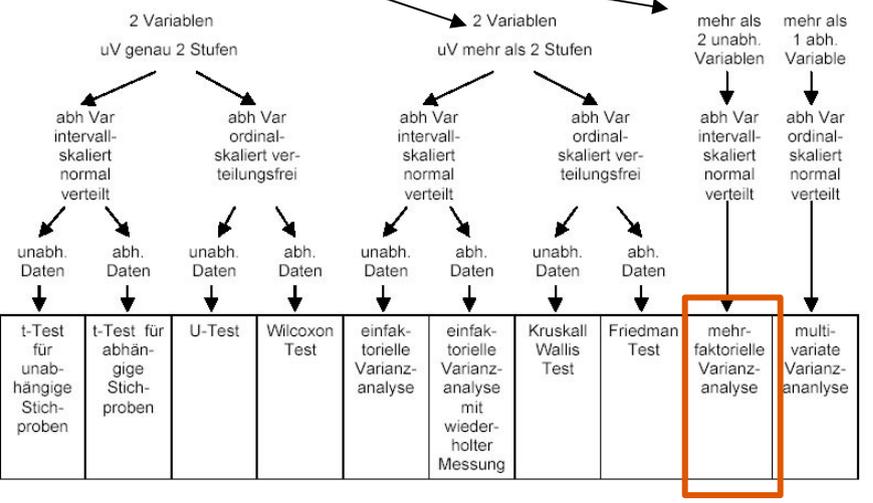
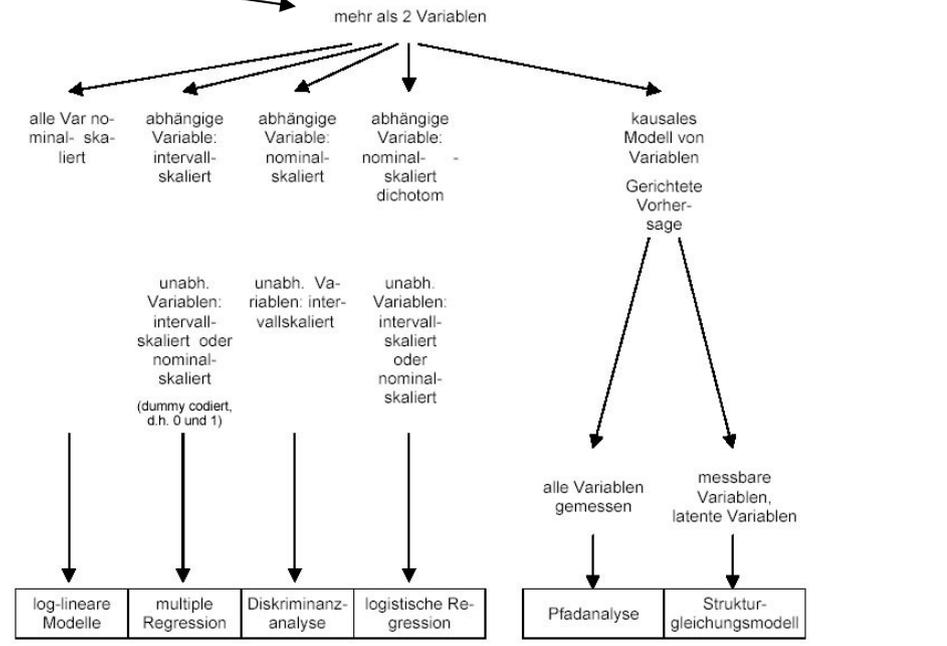
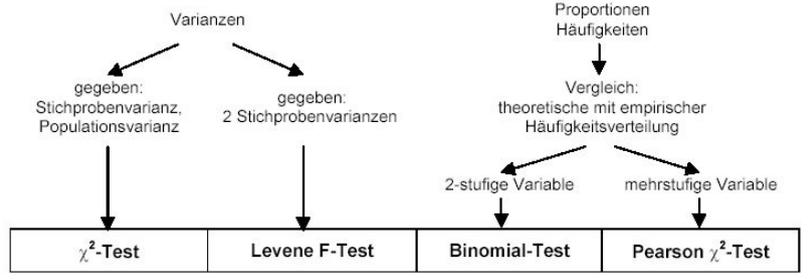
# Hypothesen entdecken

## Unterschiede

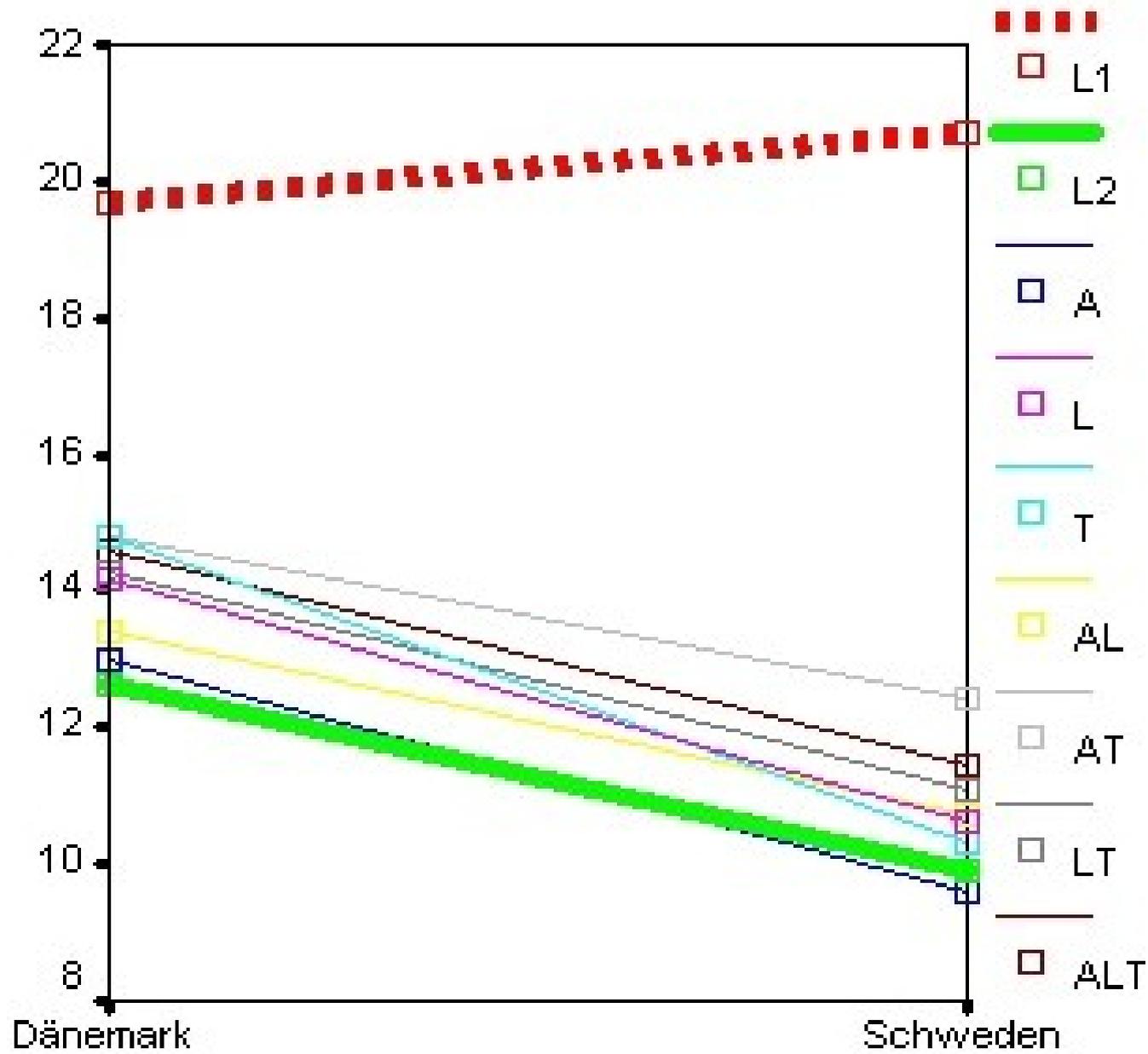
## Zusammenhänge Abhängigkeit



**Zentrale Tendenz**  
Mittelwerte;  
Mediane



# 3 . Beispiel: Auswertung



# Überblick

1. **Erklärung einiger Grundbegriffe**
  - 1.1 Hypothesentestung
  - 1.2 Signifikanz
  - 1.3 Skalenniveau
  - 1.4 Variablen
2. **Statistische Verfahren**
  - 2.1 Entscheidungsbäume
  - 2.2 QMwiki
3. **Ein Beispiel**
4. **Schwierigkeiten bei linguistischen Daten**

# Überblick

1. **Erklärung einiger Grundbegriffe**
  - 1.1 Hypothesentestung
  - 1.2 Signifikanz
  - 1.3 Skalenniveau
  - 1.4 Variablen
2. **Statistische Verfahren**
  - 2.1 Entscheidungsbäume
  - 2.2 QMwiki
3. **Ein Beispiel**
4. **Schwierigkeiten bei linguistischen Daten**

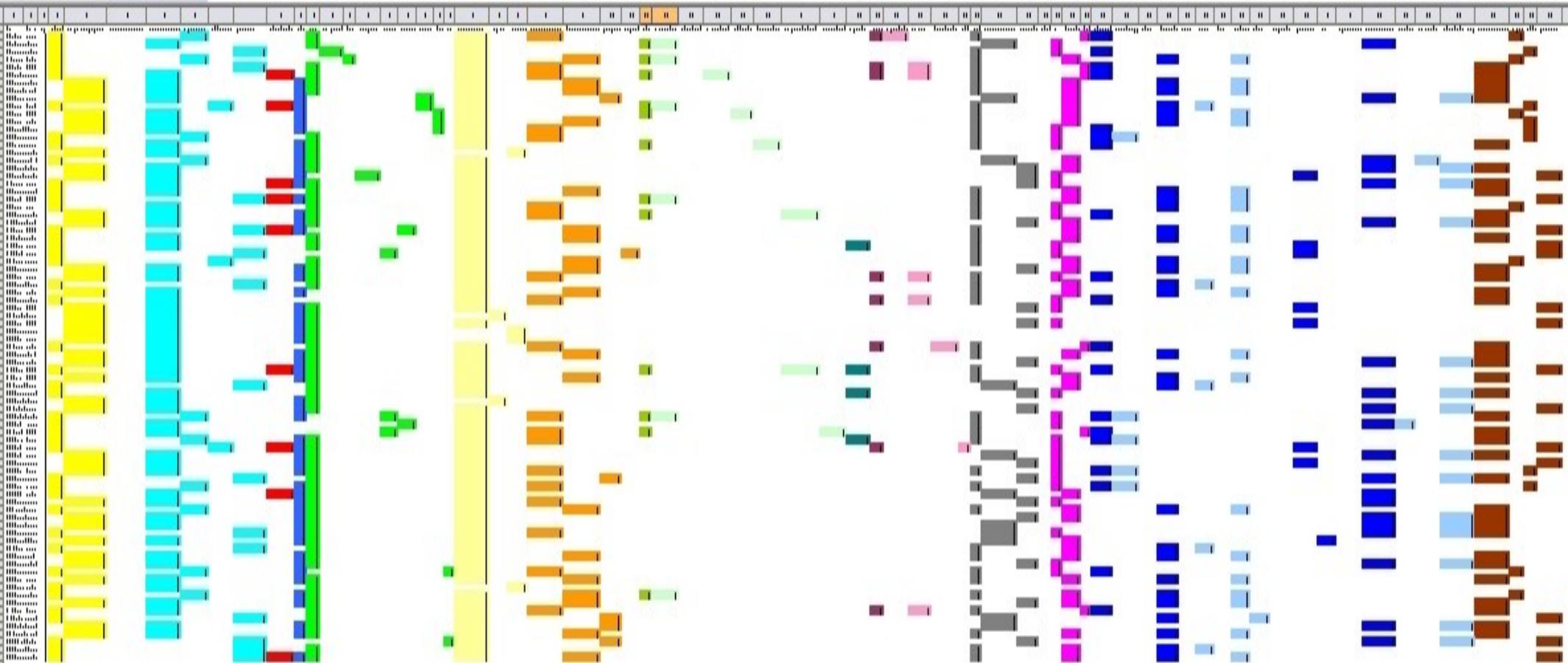
# 4. Linguistische Daten

## Warum Schwierigkeiten?

- Daten von sehr wenigen Personen
- Korpusdaten mit vielen seltenen und wenigen häufigen Variablen
  - » besondere Verfahren nötig

# 4. Beispiel Korpusdaten

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Text	Comment	0-ing	ing	ing-progress	unklar	subject	extraposed-s	subject-predi	adverbial	part-of-np	complement	post-predica	noun-postm	sentential-rel	temporal-rel	instrumental	cause-
2	1. Today, we	agent: implizi	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0
3	2. [[Building]] on this sur		0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0



44	201. # The : NUN: Tempc	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
45	209. We ml DARAUf: ka	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
46	210. We took the top 100	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
47	215. # 'e continued to d	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0
48	216. We will be delivering	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0

Tabelle1 / Tabelle2 / Tabelle3 /

# 4. Beispiel Korpusdaten

ing-Typ		Auspr	miss						
G-J	semantische Relation	4	0%	Beschreibung des englischen -ing					
K	eigenes Subjekt	1	98%						
L	Explizite Konnexion	1	95%						
M-U	ing-Verbtyp	9	3%						
V-X	Übersetzungstyp	3	0%						
	V			Übersetzungsmerk					
	I			Y-AF	Merkmal d. Übersetzungsstruktur (1)	7		46%	
				AG	Merkmal: Adverb (2)	1		85%	
					AH-AQ	Typ adverb	10	85%	
				AR	Merkmal 3_temporaldeixis	1		95%	
				AS	Merkmal 4_verweis	1		90%	
					AT-AZ	Typ verweis	7	90%	
				BA-BD	Merkmal 5_direktes äquiv	4		9%	
				BE-BG	Merkmal 6_infostuktur	3		15%	
				BH-CB	Merkmal 7_syntaktstruktur	5		7%	
					BH	BI-BJ	Independent	1	94%
					BK	BL-BQ	Dependent	6	65%
					BR		split_sentence	1	91%
					BS		Matrixsatz	1	99%
					BT		single_sent	1	100%
					BU	BV-CB	no_equivalent	6	79%
				CC-CE	Merkmal 8_semantische Relation_D	3		46%	

# 4. Beispiel Korpusdaten

Aufgabenstellung statistisch gesehen:

(a) Gruppen von Fällen (Sätzen) zu finden, die sich in der englischen Version bedeutsam voneinander unterscheiden, und zu untersuchen, ob sich diese Gruppen im Deutschen ebenfalls unterscheiden.

Bzw. stellt sich die Frage, ob es

(b) ein multivariates Verfahren gibt, das diese Abhängigkeit mehrerer kategorialer aVs von mehreren kategorialen uVs abbilden kann.

# 4. Beispiel Korpusdaten

Spezielle Probleme:

- Große Anzahl von Variablen, die eine sehr geringe Auftrittshäufigkeit haben
- Hierarchische Abhängigkeit einzelner Variablen untereinander
- Naheliegende Verfahren können nicht ohne weiteres angewendet werden (Clusteranalyse, Faktorenanalyse)

# 4. Linguistische Daten

Noch einige ungelöste Aufgaben –

Recherche nach Verfahren, die auf diese Art von Daten und linguistische Fragestellungen spezialisiert sind

AnswerTree

Logit-Loglineare Modelle

Hauptkomponentenanalyse für kategoriale Daten

Variable Rules Analysis (GoldVarb, R-Varb)

Randomization Tests (Trotter)

q.e.d.