

Werkzeuge der empirischen Forschung

Wolfgang Kössler

Institut für Informatik, Humboldt-Universität zu Berlin
SS2008

14. April 2008

Übersicht

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

1 Einleitung

Einleitung

Statistik und Wahrscheinlichkeitsrechnung

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Stochastik

- befasst sich mit zufälligen Erscheinungen
Häufigkeit, Wahrscheinlichkeit und Zufall
grch: Kunst des geschickten Vermutens
- Teilgebiete
 - Wahrscheinlichkeitsrechnung
 - Statistik

Einleitung

Statistik und Wahrscheinlichkeitsrechnung

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Stochastik

- befasst sich mit zufälligen Erscheinungen
Häufigkeit, Wahrscheinlichkeit und Zufall
grch: Kunst des geschickten Vermutens
- Teilgebiete
 - Wahrscheinlichkeitsrechnung
 - Statistik

Wahrscheinlichkeitsrechnung

gegebene Grundgesamtheit (Verteilung) →
Aussagen über Realisierungen einer Zufallsvariablen
treffen.

Einleitung

Statistik

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Statistik

- Gesamtheit aller Methoden zur Analyse zufallsbehafteter Datenmengen
- Gegeben: (Besondere) zufallsbehaftete Datenmengen
- Gesucht: (Allgemeine) Aussagen über die zugrundeliegende Grundgesamtheit
- Teilgebiete:
 - Beschreibende oder Deskriptive Statistik
 - Induktive Statistik
 - Explorative oder Hypothesen-generierende Statistik (data mining)

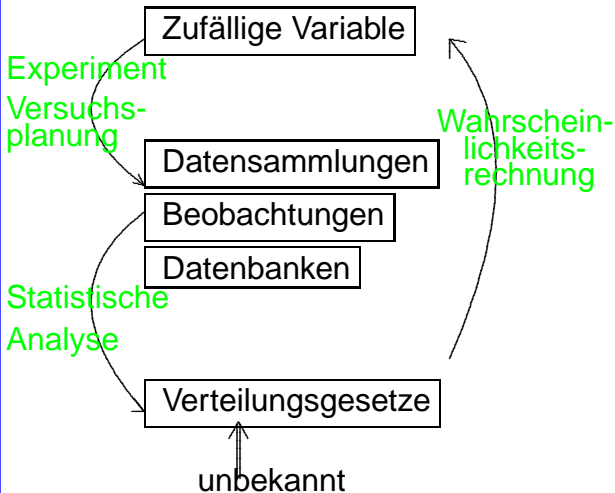
Einleitung

Überblick: Statistik

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung



Einleitung

Beschreibene Statistik

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Beschreibene Statistik

- statistische Maßzahlen: Mittelwerte, Streuungen, Quantile, ...
- Box-Blots
- Stamm-und Blattdiagramme
- Balkendiagramme
- Zusammenhangsmaße
- Punktediagramme (Scatterplots)

Einleitung

Schließende Statistik

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Schließende Statistik

- Vergleich von Behandlungen, Grundgesamtheiten, Effekten
→ t-Test, Wilcoxon-Test, ANOVA, Kruskal-Wallis-Test, Friedman-Test
- Ursache-Wirkungsanalysen, Vorhersagen, Bestimmen funktionaler Beziehungen, Trendbestimmungen
→ lineare, nichtlineare Regression
→ Kurvenschätzung
→ logistische Regression
→ Korrelation und Unabhängigkeit

Einleitung

Schließende Statistik

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Schließende Statistik

- **Klassifikation**
 - Clusteranalyse
 - Hauptkomponentenanalyse
 - Faktorenanalyse
 - Diskriminanzanalyse
- **weitere Verfahren**
 - Lebensdaueranalyse (Zuverlässigkeit)
 - Qualitätskontrolle
 - Zeitreihenanalyse

Einleitung

Vergleich von Behandlungen, Grundgesamtheiten, Effekten

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Vergleich von Behandlungen, Grundgesamtheiten, Effekten

- Einstichprobenproblem
Messungen sollen mit einem vorgegebenen Wert verglichen werden:
- Zweistichprobenproblem
 - Vergleich zweier unabhängiger Stichproben
 - Vergleich zweier abhängiger Stichproben
- Vergleich mehrerer unabhängiger Stichproben
- Vergleich mehrerer abhängiger Stichproben

Einleitung

Ein- und Zweistichprobenproblem

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Eine Stichprobe

- Banknoten: vorgegebene Länge eingehalten?

→ **Einstichproben t-Test, Signed-Wilcoxon-Test**

Einleitung

Ein- und Zweistichprobenproblem

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Eine Stichprobe

- Banknoten: vorgegebene Länge eingehalten?

→ **Einstichproben t-Test, Signed-Wilcoxon-Test**

Abhängige und Unabhängige Stichproben

- Vergleich zweier unabhängiger Stichproben
 - echte - gefälschte Banknoten
 - Schädel aus verschiedenen Gegenden Tibets

→ **t-Test, Wilcoxon-Test**

- Vergleich zweier abhängiger Stichproben
Länge des Scheines oben und unten

→ **Einstichproben t-Test,
Vorzeichen-Wilcoxon-Test**

Einleitung

Vergleich von Behandlungen, Grundgesamtheiten, Effekten

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Abhängige und Unabhängige Stichproben

- Vergleich mehrerer unabhängiger Stichproben:
Ägypt. Schädel: mehrere Grundgesamtheiten,
Epochen
→ **ANOVA, Kruskal-Wallis-Test**
- Vergleich mehrerer abhängiger Stichproben
Blutdruck von Patienten an mehreren
aufeinanderfolgenden Tagen,
(Faktoren: Patient, Tag)
Preisrichter beim Synchronschwimmen
→ **2 fakt. Varianzanalyse, Friedman-Test**

Einleitung

Ursache - Wirkungsanalysen

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Ursache - Wirkungsanalysen

- Ursache - Wirkungsanalysen
 - Zusammenhangsanalyse
 - Bestimmen funktionaler Beziehungen
 - Trends, Vorhersagen
 - Beispiele:
 - Bluthochdruck - Rauchgewohnheiten
 - Blutdruck - Proteinuria
 - Größe - Gewicht
 - Sterblichkeit - Wasserhärte
- **Lineare, Nichtlineare und Nichtparametrische Regression**
- **Korrelation**

Einleitung

Klassifikation

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Klassifikation

- Auffinden von Gruppen in Daten
→ **Clusteranalyse**
- Individuen sollen einer von vorgegebenen Klassen zugeordnet werden
→ **Diskriminanzanalyse**
- Datensatz hat Variablen, die mehr oder weniger voneinander abhängen.
Welche Struktur besteht zwischen den Variablen?
→ **Hauptkomponentenanalyse**
→ **Faktorenanalyse**

Einleitung

Literatur

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Literatur (1)

Dufner, Jensen, Schumacher (1992). Statistik mit SAS, Teubner.

Falk, Becker, Marohn (1995). Angewandte Statistik mit SAS, Springer.

Graf, Ortseifen (1995). Statistische und grafische Datenanalyse mit SAS, Spektrum akademischer Verlag Heidelberg.

Krämer, Schoffer, Tschiersch (2004). Datenanalyse mit SAS, Springer.

SAS-Online Dokumentation, SAS-Handbücher

Einleitung

Literatur

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Literatur (2)

Hartung (1993). Statistik, Lehr- und Handbuch, Oldenbourg.

Sachs (1999). Angewandte Statistik, Springer.

Muche, Habel, Rohlmann (2000). Medizinische Statistik mit SAS Analyst, Springer.

Graf, Bundschuh, Kruse (1993). Effektives Arbeiten mit SAS, Wissenschaftsverlag.

Gogolok, Schuemer, Ströhlein (1990). Datenverarbeitung und statistische Auswertung mit SAS, Fischer

Einleitung

Literatur

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Literatur (3)

Steinhausen, Zörkendörfer (1992).

Informationsverarbeitung und Datenanalyse mit
dem Programmsystem SAS, Oldenbourg.

Göttsche (1992). SAS-kompakt, Fischer.

Einleitung

Statistik Software

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Statistik-Software

- SAS
 - sehr umfangreich, universell
 - weit verbreitet
- SPSS
 - umfangreich
 - Anwendung vor allem in Biowiss., Medizin, Sozialwiss.
- SYSTAT
 - ähnlich wie SPSS
 - sehr gut
- BMDP
 - umfangreich
- S, S⁺, R
 - funktionale Sprachen
 - R: frei verfügbar

STATGRAPHICS, XPLORE, MATHEMATICA ...

Einleitung

Starten und Beenden von SAS

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Account für Mathepool beantragen

- Die Software ist im Mathepool R. 2.212 und R. 2.213 installiert.
- vor erster Benutzung Nutzerkennzeichen beantragen

`https://www.math.hu-berlin.de/account`

- Passwort abholen (Dr. Gehne, 2.2.05)

Einleitung

Starten und Beenden von SAS

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Starten und Beenden von SAS

- Starten von SAS
 1. Sitzungsart KDE anmelden
 2. Behelfsfenster- **Konsole** starten
 3. beim Windows-Server einloggen:
`rdesktop -f idun`
Passwort angeben;
log on to: **localmath** (nicht: idun)
 4. Start von SAS: **All Programs > SAS-System > The SAS-System 9.1 (English)**
- Beenden der Sitzung
All Desktop exit > Logoff > Abmelden

Einleitung

Allgemeine Struktur von SAS

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

SAS-Fenster

- Nach dem Starten erscheinen 3 Fenster
 - Log-Fenster
 - Editor-Fenster
 - Output-Fenster (verdeckt)
- weitere Fenster:
 - Results: Ergebnisse aus der Sitzung
 - Grafik-Fenster (gegebenfalls)
 - Hilfen

Einleitung

Allgemeine Struktur von SAS

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Hilfen

- help > SAS Help and Documentation
- SAS Products
- BASE SAS
 - > SAS Language Concepts
 - > Data Step Concepts
 - > SAS STAT
 - > SAS STAT User's Guide

Einleitung

Allgemeine Struktur eines SAS-Programms

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Aufbau einer SAS-file

DATA

PROC

DATA

PROC

PROC

...

- DATA-Schritte:
 - Erstellen der SAS-Dateien
 - Einlesen, Erstellen, Modifikation der Daten
- PROC-Schritte:
Auswertung der Dateien

Einleitung

Daten

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Daten

Ausgangspunkt sind die Daten, die für die Analyse relevant sind. Die Struktur der Daten hat die folgende allgemeine Form: x_{ij}

| Objekte | Merkmale | | | | | | |
|---------|----------|---|---|----|----------|----|---|
| | 1 | 2 | 3 | .. | j | .. | p |
| 1 | | | | | | | |
| 2 | | | | | | | |
| 3 | | | | | | | |
| .. | | | | | | | |
| i | | | | | x_{ij} | | |
| .. | | | | | | | |
| N | | | | | | | |

Wert oder
Ausprägung
des Merkmals j
am Objekt i

Einleitung

Daten

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Daten

p : Anzahl der Merkmale

N : Gesamtanzahl der einbezogenen Objekte
(Individuen)

| Objekte | Merkmale | | | | | | |
|---------|----------|---|---|----|----------|----|-----|
| | 1 | 2 | 3 | .. | j | .. | p |
| 1 | | | | | | | |
| 2 | | | | | | | |
| 3 | | | | | | | |
| .. | | | | | | | |
| i | | | | | x_{ij} | | |
| .. | | | | | | | |
| N | | | | | | | |

Qualität des
Datenmaterials
wird im Wesent-
lichen durch die
Auswahl der
Objekte aus einer
größeren
Grundgesamtheit
bestimmt.

Einleitung

Daten

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Beispiele

- Objekte: **Patienten einer Klinik**
Merkmale: Alter, Geschlecht, Krankheiten
- Objekte: **Bäckereien in einer bestimmten Region**
Merkmale: Anzahl der Beschäftigten,
Geräteausstattung, Umsatz, Produktpalette
- Objekte: **Banknoten**
Merkmale: Längenparameter

Einleitung

Daten

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Datenmatrix

- Zeilen: Individuen, Objekte, Beobachtungen
- Spalten: Merkmalsausprägungen, -werte, -realisierungen

| Banknote | Merkmale | | | | | | |
|----------|----------|------|-------|----|---|----|----|
| | laenge | oben | unten | .. | j | .. | gr |
| 1 | | | | | | | |
| 2 | | | | | | | |
| 3 | | | | | | | |
| .. | | | | | | | |
| i | | | | | | | |
| .. | | | | | | | |
| N | | | | | | | |

x_{ij}

Einleitung

Daten

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Merkmale

- Definition: **Merkmale** sind Zufallsvariablen, die für jedes Individuum (Objekt) eine bestimmte Realisierung (Merkmalsausprägung) haben.
- Stetige Merkmale: laenge, oben
- Diskrete Merkmale: gr (Gruppe)

| Banknote | Merkmale | | | | | | |
|----------|----------|------|-------|----|---|----|----|
| | laenge | oben | unten | .. | j | .. | gr |
| 1 | | | | | | | |
| 2 | | | | | | | |
| .. | | | | | | | |