

## Seminarübersicht - Quantitative Datenanalyse

### 1. Sitzung (24.04): Kursüberblick - Datenanalyse im Forschungsprozess

- Geschichte der Statistik; Datenanalyse im Forschungsprozess; Deskriptive und induktive Statistik;
- Einführung in R und RStudio;
- Skalenniveau (Äquivalenz-, Ordnungs- und Abstandsrelation)

Literatur:

1. Field, Andrew (2012). *Discovering Statistics using R*. Kapitel 1, S. 1-31.

### 2. Sitzung (08.05): Datensatzexploration

- Datenquellen;
- Datensatz und Untersuchungseinheit, das Tidy Data-Konzept;
- Erste Schritte der Datenmanipulation mit R und *dplyr*

Literatur:

1. Luhmann, Maike (2015). *R für Einsteiger: Einführung in die Statistiksoftware für die Sozialwissenschaften*. Kapitel 2 und 3, S.16-32.
2. DataCamp *Introduction to R*. [www.datacamp.com/courses/free-introduction-to-r](http://www.datacamp.com/courses/free-introduction-to-r)

### 3. Sitzung (15.05): Lineare Regression

- Grundlagen der statistischen Modellierung;
- Grundlegende Begriffe: *Effektstärke, Richtung, Signifikanz*

Literatur:

1. Lewis-Beck, Colin and Michael S. Lewis-Beck (2016). *Applied regression: an introduction, Second edition*. Kapitel 1, S. 1-21.
2. DataCamp *Data Manipulation in R with dplyr*. [www.datacamp.com/courses/dplyr-data-manipulation-r-tutorial](http://www.datacamp.com/courses/dplyr-data-manipulation-r-tutorial)

### 4. Sitzung (22.05): Lineare Regression II

- Methode der kleinsten Quadrate
- Gütemaß der Regression
- Gewichtung

Literatur:

1. Wolf, Christof und Henning Best (2010). *Lineare Regressionsanalyse* in Wolf/Best: Handbuch der sozialwissenschaftlichen Datenanalyse. S. 607-638.
2. Kaplan, David (2015). *Data Computing*. Kapitel 1 bis 3, S. 1-31.

### 5. Sitzung (29.05): Deskriptive Statistik

- Variablen und Ausprägungen;
- Diskrete und stetige Merkmale;
- Absolute, relative und kummulierte Häufigkeiten;
- Grafische Darstellung: Scatterplot, Barplot, Linechart, Histogram, Density Chart, Boxplot

Literatur:

1. Bortz, Jürgen und Christof Schuster (2010). *Statistik für Human- und Sozialwissenschaftler*. Kapitel 3, S.39-48.
2. Benninghaus, Hans (2002). *Deskriptive Statistik*. Kapitel 2 bis 3.2 S. 20-3.

## 6. Sitzung (12.06): Maße der zentralen Tendenz

- Maße der zentralen Tendenz (Modus, Median, Arithmetisches Mittel),
- Varianz, Standardabweichung, Schiefe; Wölbung;
- Verteilung, insb. Normalverteilung

Literatur:

1. Bortz, Jürgen und Christof Schuster (2010). *Statistik für Human- und Sozialwissenschaftler*. Kapitel 2, S.25-38.
2. Benninghaus, Hans (2002). *Deskriptive Statistik*. Kapitel 3.3 S. 36-58.

## 7. Sitzung (19.06): Stochastik

- z-Transformation,
- Zentraler Grenzwertsatz
- Beispiele: *The Lady Tasting Tea*, *How much of the earth is covered with water?*

## 8. Sitzung (26.06): Multiple Regression

- *Auswertung der Hausarbeit*
- Mehrstufiger Modellaufbau
- Interaktionsterme
- Modellinterpretation

Literatur:

1. Luhmann, Maike (2015). *R für Einsteiger: Einführung in die Statistiksoftware für die Sozialwissenschaften*. Kapitel 16, S.221-242.

## 9. Sitzung (03.07): Unterschiedshypothesen

- Einführung in den t-Test

Literatur:

1. Luhmann, Maike (2015). *R für Einsteiger: Einführung in die Statistiksoftware für die Sozialwissenschaften*. Kapitel 13, S.-220.
2. Field, Andrew (2012). *Discovering Statistics using R*. Kapitel 9, S. 359-397.

## 10. Sitzung (10.07): Offene Sitzung und Fragestunde zur Klausur

## 11. Sitzung (17.07): Klausur

Im Laufe des Kurses besteht die Möglichkeit im Rahmen einer Hausarbeit ein kleinere statistische Analyse anzufertigen:

- Ausgabe der Hausarbeit: 12. Juni 2017
- Abgabe der Hausarbeit: 22. Juni 2017

## Begriffsübersicht

---

	Englisch	Deutsch
<b>Design</b>	Population Sample	Grundgesamtheit Stichprobe
<b>Variablen</b>	Independent (predictor) variable Dependent (outcome) variable Confounding variable	Unabhängige (erklärende) Variable Abhängige (erklärte) Variable Dritt-/Störvariable
<b>Modelle</b>	Linear Model Fit Degrees of freedom Sum of squared errors Sum of deviances Standard error (of the mean) Central limit theorem	Lineares Modell Anpassungsgüte Freiheitsgrade Residuenquadratsumme Abweichungssumme Standardfehler Zentrales Grenzwerttheorem
<b>Verteilungen</b>	Frequency distribution Probability distribution Normal distribution Range Interquartile range Skew Kurtosis	Häufigkeitsverteilung Wahrscheinlichkeitsverteilung Normalverteilung Variationsbreite Interquartilbereich Schiefe Wölbung

---

## Notation

Symbol	Entsprechung
$i$	Laufindex der Untersuchungseinheiten $e$
$n_i$	Absolute Häufigkeit
$f_i$	Relative Häufigkeit
$F_i$	Kumulierte Häufigkeit
$f_i^*$	Häufigkeitsdichte
$\Delta x_i$	Klassenbreite
$k_i$	Klassengrenzen
$Mo$	Modalwert
$Md = x_{50\%}$	Median
$\bar{x}$	Arithmetisches Mittel (der Stichprobe)
$\mu$	Arithmetisches Mittel (der Grundgesamtheit)
$s^2$	(Stichproben-) Varianz
$\sigma^2$	Varianz (der Grundgesamtheit)
$s$	Standardabweichung (der Stichprobe)
$\sigma$	Standardabweichung (der Grundgesamtheit)
$IQR$	Interquartilsbereich
$Q_1 = x_{25\%}$	<i>Obere</i> Grenze des <i>untersten/ersten Quartils</i>
$z$	z-Werte/z-Scores
$\hat{b}_0$	Schätzer für den Schnittpunkt der Regressionsgerade mit der y-Achse (Intercept)
$\hat{b}_1$	Schätzer für die Steigung der Regressionsgerade (Slope)