

K 3.7 Explorative Datenanalyse (EDA)

3.7.1 EDA Ziele

- Für eine erste Durchsicht eines Datensatzes
 - Um Datenqualität zu untersuchen (und Data Cleaning durchzuführen)
 - Um Annahmen über die Datenstruktur zu überprüfen
 - Um vorgegebene Hypothesen anzuschauen
- Um neue Ideen zu entwickeln und neue Einsichten zu gewinnen
- Um Ergebnisse aus Modellen
 - zu untersuchen, zu verstehen, zu ergänzen
- Für alle Phasen einer Analyse (John Tukey, 1962)

3.7.2 Unterschiede zwischen Statistik und EDA

- Quantitative Information v. Qualitative Information
- Präzise Antworten auf eingeschränkte Fragen v. unscharfe Antworten auf allgemeine Fragen
- Modellgesteuert v. Datengesteuert
- Tatsachen v. Einsichten
- Mathematische Modelle v. Heuristiken und Graphiken

3.7.3 EDA Beispiele

- Strukturdaten für deutsche Wahlkreise 2005
- Titanic
- Web Marketing