

Statistik I — bis jetzt

K 1 Einführung

K 2 Beschreibende Statistik

K 3 Graphiken

und jetzt

K 4 Schätzen

und jetzt zurück zu Graphiken

K 3.5 Interaktive Graphiken

K 3.5 Interaktive Graphiken

3.5.1 Beispiel: Die Patienten von Dr. H. Shipman

- Daten aus dem offiziellen Untersuchungsbericht
- Tode von 508 Patienten zwischen 1973 und 1998
- Alter, Name, Todestag, Todesort, Entscheidung der juristischen Untersuchung
- (keine Information zur Todeszeit oder ob der Arzt anwesend war)

3.5.2 Interaktive Graphik und EDA

EDA (Explorative Datenanalyse) bedeutet die Generierung von Hypothesen — nicht die Überprüfung.

IG ist (idealerweise) die direkte Manipulation von statistischen Objekten in Graphiken. Statistische Objekte sind, z.B. Datenpunkte, Säulen in Säulendiagrammen oder Histogrammen, Achsen. Sie werden aus graphischen Objekten (z.B. Pixeln oder Linien) erstellt.

IG muss schnell, flexibel und “forgiving” (verzeihend) sein.

Für IG können Sie die Software MONDRIAN
<http://stats.math.uni-augsburg.de/Mondrian/>
von Martin Theus verwenden.

3.5.3 Hauptmerkmale von Interaktive Graphik

- Abfrage
Welche Werte werden dargestellt?
Default und erweiterte Abfragen.
- Selektion und Linking (Verknüpfungen)
Ausgewählte Fälle werden überall hervorgehoben.
(1) Bei Histogrammen und Säulendiagrammen sieht man #(selektierte aus X), bei Spineplots sieht man $P(\text{selektiert}|X)$.
(2) Gegeben $X = x_j$ selektiert, wird $X = x_j|Y = y_k$ in der $Y = y_k$ Säule gezeigt.
- Umformatierung
Fenstergröße ändern, Seitenverhältnis ändern, Punkte vergrößern, neuskalieren (Minimum, Maximum, Binbreite), gemeinsame Skalierung, sortieren, zoomen, ...
- Multiple Ansichten
Es gibt keine optimale Darstellung. Verschiedene Darstellungen liefern verschiedene Einsichten.

3.5.4 Interaktivität und statistische Graphiken

- Säulendiagramme (und Spineplots)
- Histogramme (und Spinogramme)
- Boxplots
- Streudiagramme