

Einführung in die Statistik

HAUSÜBUNG 2 - Explorative Datenanalyse, erkennen von Eigenschaften und Visualisierung von unterschiedlichen Variablentypen

1. Explorieren und visualisieren Sie die Variablen "Fertility", "Agriculture", "Education", "Catholic" und "Infant. Mortality" aus dem R Datensatz *swiss* des R package *utils*. betrachten Sie vorerst jede Variable als separate Stichprobe für eindimensionale Exploration (ziehen Sie die Bedeutung der Variablen im Sachkontext in Betracht). Für jede Variable:

- wählen Sie sinnvolle Schätzer für Lokation, Variation, Schiefe and Gewicht in den Rändern.
- Geben Sie dem Nutzer die Möglichkeit zwischen unterschiedlichen graphischen Darstellungen zu wechseln. Erklären Sie die Zusammenhänge und Eigenschaften der Daten, die sich aus diesen Visualisierungen erkennen lassen.
 - Sind die Daten symmetrisch/schief?
 - Haben die Daten schwere Ränder?
 - Bieten Sie robuste und nichtrobuste Lagemaße und Skalenmaße im Vergleich oder zur Auswahl an.
 - Sind die Daten (approximativ) normalverteilt?

Was lässt sich über die Zusammenhänge zwischen den Variablen aussagen? (Tipp: scatterplot matrix.)

2. Explorieren und visualisieren Sie die Variablen "Population", "Income", "Illiteracy", "Life.Exp", "Murder", "HS Grade" und "Frost" aus dem R Datensatz *state.x77*. Betrachten Sie vorerst jede Variable als separate Stichprobe für eindimensionale Exploration (ziehen Sie die Bedeutung der Variablen im Sachkontext in Betracht). Für jede Variable:

- wählen Sie sinnvolle Schätzer für Lokation, Variation, Schiefe and Gewicht in den Rändern.
- Geben Sie dem Nutzer die Möglichkeit zwischen unterschiedlichen graphischen Darstellungen zu wechseln. Erklären Sie die Zusammenhänge und Eigenschaften der Daten, die sich aus diesen Visualisierungen erkennen lassen.
 - Sind die Daten symmetrisch/schief?
 - Haben die Daten schwere Ränder?
 - Bieten Sie robuste und nichtrobuste Lagemaße und Skalenmaße im Vergleich oder zur Auswahl an.
 - Sind die Daten (approximativ) normalverteilt?

Was lässt sich über die Zusammenhänge zwischen den Variablen aussagen? (Tipp: scatterplot matrix.)

3. Explorieren und visualisieren Sie den Datensatz "LakeHuron" mit und ohne Berücksichtigung des Zeitreihenaspektes.

- wählen Sie sinnvolle Schätzer für Lokation, Variation, Schiefe and Gewicht in den Rändern.
- Geben Sie dem Nutzer die Möglichkeit zwischen unterschiedlichen graphischen Darstellungen zu wechseln. Erklären Sie die Zusammenhänge und Eigenschaften der Daten, die sich aus diesen Visualisierungen erkennen lassen.
 - Sind die Daten symmetrisch/schief?
 - Haben die Daten schwere Ränder?
 - Bieten Sie robuste und nichtrobuste Lagemaße und Skalenmaße im Vergleich oder zur Auswahl an.
 - Sind die Daten (approximativ) normalverteilt?

4. Explorieren und visualisieren Sie den Datensatz Titanic. Wie beeinflussen Alter, Geschlecht und Klasse das Überleben? Finden Sie, wo sich Simpson's Paradoxon zeigt und begründen Sie, woher dieser Effekt kommt.