

# 12

## Fehlende Werte

Fehlende Werte sind ein großes Problem, da das Fehlen einen Grund haben könnte. Ist das Fehlen nicht vollständig zufällig, werden Schätzer wie Mittelwert etc. verzerrt.

## **Missing Completely at Random (MCAR)**

Das Fehlen tritt völlig zufällig auf

z.B. zufällige Störung beim Übertragen der Daten

## **Missing at Random (MAR)**

Das Fehlen hängt nicht von der Variable selbst, aber von anderen Variablen ab

z.B. "vergisst" einer der behandelnden Ärzte, den Raucherstatus zu erfassen

## **Missing not at Random (MNAR)**

Das Fehlen hängt von der Variable selber ab

z.B. Gehaltsangabe wird nicht gemacht bei niedrigem Einkommen

Wir haben zwei Möglichkeiten, mit fehlenden Werten umzugehen:

- Löschen
  - nur zulässig bei wenigen fehlenden Werten bzw. wenn das Fehlen vollständig zufällig ist (MCAR)
- Imputation
  - **Single Imputation:** Einsetzen eines Wertes, z.B. des Mittelwerts oder Vorhersage des Werts anhand von anderen Variablen (z.B. lineare Regression). Problem ist, dass die Standardabweichung reduziert wird.
  - **Multiple Imputation:** Erstellen von mehreren Datensätzen mit jeweils single imputation + Zufall. Damit wird die Standardabweichung nicht verzerrt. Das Verfahren ist jedoch aufwändig.