

Lösung zu Kapitel 2: Beispiel 1

Beschreiben Sie in eigenen Worten die folgenden Begriffe:

- Population
- Stichprobe
- Beobachtungseinheit
- Inferenzstatistik
- Variable
- Merkmalsausprägung
- Messung
- kategoriale Daten
- metrische Daten
- Skalenniveau

- Die **Population** ist die Menge aller möglichen Beobachtungseinheiten, über die man eine Aussage treffen will. Sie wird auch Grundgesamtheit genannt. z.B. 10000 Dentisten und Zahnärzte, welche CARIDEX im ersten Jahr verwenden würden
- Die **Stichprobe** ist die nach bestimmten Kriterien ausgewählte Teilmenge (ein Ausschnitt) aus der Grundgesamtheit. Sie soll ein möglichst repräsentatives Abbild der Grundgesamtheit sein. z.B. 400 (zufällig) ausgewählte Dentisten und Zahnärzte aus der Grundgesamtheit
- Die **Beobachtungseinheit** ist ein Individuum oder Objekt, an dem etwas beobachtet oder es befragt wird. z.B. ein Zahnarzt, der CARIDEX im ersten Jahr verwenden würde
- Soll von einer Stichprobe auf die Grundgesamtheit geschlossen werden, verwendet man Methoden der **Inferenzstatistik** (auch inferenzielle oder schließende Statistik). Es wird dabei ermittelt, in welchem Bereich die Werte in der Grundgesamtheit (mit einer bestimmten Wahrscheinlichkeit) liegen. Solche Hochrechnungen der Werte aus der Stichprobe auf die Grundgesamtheit werden auch Schätzungen genannt. Um vernünftige Prognosen abgeben zu können muss aus den Daten ein statistisches Modell abgeleitet werden, das bestimmte Sachverhalte geeignet abbildet und Beziehungen zwischen ihnen zu erklären hilft.
- Die **Variable** ist ein bestimmtes Merkmal (bzw. Eigenschaft) von Beobachtungseinheiten, das man erheben will. Eine Variable kann je nach Beobachtungseinheit verschiedene Merkmalsausprägungen (Werte) annehmen. z.B. die Anzahl der behandelten Zähne in einem Jahr
- **Merkmalsausprägungen** sind die beobachteten bzw. erhobenen Werte der Beobachtungseinheiten. z.B. 7, 3, 5, . . .
- Bei der **Messung** wird einem beobachteten Tatbestand eine Zahl zugeordnet. Dabei sollen die Beziehungen zwischen den beobachteten Phänomenen durch die Beziehungen zwischen den zugeordneten Zahlen widerspiegelt werden.
- Erfolgt bei der Messung einer Variablen die Einteilung der Beobachtungseinheiten durch Klassifikation in zwei oder mehreren Kategorien (Gruppen, Klassen), so spricht man von **kategorialen Daten**. Kategoriale Daten sind Bezeichnungen bzw. Kategorienamen. Die beobachtbare Beziehung zwischen diesen Kategorien ist, dass sie unterschiedlich sind.
Handelt es sich um ungeordnete Kategorien, können durch die Kategorisierung nur Unterschiede beschrieben werden. Die Messung der Variable erfolgt auf einer Nominalskala. z.B. Geschlecht . . . 0 =männlich ; 1 =weiblich
Handelt es sich um geordnete Kategorien - d.h. besteht zwischen den Kategorien noch eine größer-kleiner- oder mehr-weniger-Beziehung - kann man die Kategorien in eine Rangreihe

bringen. Die Messung der Variable erfolgt auf einer Ordinal- oder Rangskala. z.B. Schulnote . . . 1 ist besser als 2; 2 ist besser als 3; 3 ist besser als 4; 4 ist besser als 5

- Handelt es sich um **metrische Daten**, kommt das Ergebnis einer Messung durch eine Art Zählen zustande. Metrische Daten sind numerisch und können nur Zahlen sein. Die Messung der Variable erfolgt auf einer Intervall- oder Verhältnisskala. Die beobachtbare Beziehung zwischen diesen Merkmalsausprägungen ist, dass sie unterschiedlich sind UND man sie in eine sinnvolle Reihung bringen kann UND die Unterschiede (Differenzen) zwischen 2 Merkmalsausprägungen gleich oder ungleich bzw. kleiner oder größer sein können.
- Das **Skalenniveau** beschreibt den Informationsgehalt einer Skala. Die wenigste Information steckt in nominal skalierten Daten, während die meiste Information bei ratio(nal) skalierten Daten zu finden ist. Man kann Daten so transformieren, dass sie auch mittels einer Skala mit geringerem Informationsgehalt beschrieben werden, aber nicht umgekehrt.