

Statistik

Teil 1

Ulrich Schrader
mail@ulrich-schrader.de

Grundgesamtheit – Stichprobe

Oft interessiert einen generalisierte Aussagen über etwa das Dekubitusrisiko aller Diabetes-Patienten zu machen. Diese stellen dann die sogenannte *Grundgesamtheit (Population)* dar. Aus praktischen Überlegungen heraus kann man in der Regel allerdings nie alle Beobachtungseinheiten der Grundgesamtheit untersuchen, sondern man muss sich auf eine Teilmenge, eine *Stichprobe* beschränken.

Beobachtungseinheit – Merkmalsträger

Das einzelne Objekt der Untersuchung oder Studie wird als Beobachtungseinheit oder Merkmalsträger bezeichnet. Aufgabe der Untersuchung ist es festgelegte Merkmale dieser Beobachtungseinheit zu messen oder zu bestimmen.

Merkmal (Variable)

Ein Merkmal (Eine Variable) ist eine Eigenschaft einer Beobachtungseinheit, die gemessen oder anderweitig bestimmt werden kann. Merkmale können etwa bestimmte Symptome oder einzelne Vitalparameter sein.

Merkmalsausprägung

Die Werte, die ein Merkmal annehmen kann, werden als Merkmalsausprägungen bezeichnet. Merkmalsausprägungen können etwa die Gradeinteilungen eines Dekubitus sein, wenn das Merkmal „Schweregrad des Dekubitus“ ist.

Qualitativ – Quantitativ

- Es wird zwischen qualitativen und quantitativen Merkmalen unterschieden.
- *Qualitative Merkmale* sind Merkmale, die sich nicht durch Zahlen direkt erfassen lassen. Beispiele: Blutgruppe, Haarfarbe, Beruf und Ausbildung.
- *Quantitative Merkmale* sind dagegen durch Zahlen bestimmbar. Die Merkmale lassen sich durch wiegen, messen, abzählen usw. bestimmen.

Qualitativ Nominal – Ordinal

- Qualitative Merkmale bei denen eine Rangfolge der Merkmalsausprägungen festgelegt werden kann werden als *ordinale Merkmale* bezeichnet.
- Beispiel: „größter Ausbildungsabschluss“. Die Merkmalsausprägungen „Hauptschulabschluss“, „Realschulabschluss“, „Gymnasialabschluss“, „Hochschulabschluss“ lassen sich in eine Rangfolge bringen. Ein sinnvoller Abstand zwischen den einzelnen Merkmalsausprägungen lässt sich aber nicht bestimmen.
- Ist eine solche Rangfolge nicht sinnvoll möglich, so wird von einem *nominalen Merkmal* gesprochen.

Quantitativ Diskret - Stetig

- Quantitative Merkmale werden dabei weiter in quantitativ diskrete und quantitativ stetige Merkmale unterschieden.
- *Diskrete Merkmale* können nur bestimmte Werte annehmen. So kann das diskrete Merkmal Kinderzahl nur ganze positive Zahlen annehmen.
- Im Gegensatz dazu kann die Körpergröße in einem bestimmten sinnvollen Bereich jeden beliebigen Wert annehmen. Die Körpergröße ist ein *stetiges Merkmal*

Empirische Kenngrößen

dienen der zusammenfassenden Beschreibung der Ergebnisse einer Stichprobe.

$$x_1, x_2, \dots, x_n$$

seien n Beobachtungen eines quantitativen Merkmals X .

Üblich: n für Anzahl der Beobachtungen

Lagemaße – Lokalisationsmaße (1)

- Minimum – kleinster beobachteter Wert

$$x_{\min}$$

- Maximum – größter beobachteter Wert

$$x_{\max}$$

- Modalwert - häufigst auftretende Werte

$$x_{\text{mod}}$$

(bei mehreren häufigsten Werten kann es auch mehrere Modalwerte geben.)

Lagemaße – Lokalisationsmaße (2)

- Arithmetisches Mittel (Mittelwert, Durchschnitt)

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

Beobachtet: 1, 3, 2, 5, 5, 2

$$\bar{x} = \frac{1}{6} \cdot (1 + 3 + 2 + 5 + 5 + 2) = \frac{18}{6} = 3$$

Problem: Ausreißer, einzelne sehr große Werte verschieben das Mittel

Lagemaße – Lokalisationsmaße (3)

- (Empirischer) Median

Seien $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$

der Größe nach angeordnete Beobachtungen

x_1, x_2, \dots, x_n

$$\tilde{x} = \begin{cases} \frac{1}{2} (x_{(n/2)} + x_{(n/2+1)}) & \text{falls } n \text{ gerade} \\ x_{(n+1/2)} & \text{falls } n \text{ ungerade} \end{cases}$$

Median - Beispiel

- Ungerade Anzahl von Beobachtungen

1,1,2,4,5 6 6,7,8,8,9

- Gerade Anzahl von Beobachtungen

1,1,2,4,5 6,7,8,8,9
5,5

Streuungs- oder Dispersionsmaße (1)

- Spannweite oder Variationsbreite

$$x_{\max} - x_{\min} = x_{(n)} - x_{(1)}$$

(Breite des Intervalls, in das alle beobachteten Werte passen)

Streuungs- oder Dispersionsmaße (2)

- Empirische Varianz

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \cdot \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \cdot \left(\sum_{i=1}^n x_i \right)^2 \right]$$

- Empirische Standardabweichung

$$s = \sqrt{s^2}$$

Summenzeichen

Endwert
 \sum Regel für die Bildung der Summanden
 Startwert

$$\sum_{i=1}^5 i^2 = 1 + 4 + 9 + 16 + 25$$

$$\sum_{i=1}^5 x_i = x_1 + x_2 + x_3 + x_4 + x_5$$

Summe der ersten
 fünf beobachteten
 Werte