schema-F

Grundlagen der Wirtschaftsmathematik und Statistik

Deskriptive Statistik
Erläuterungen

Lernmaterial zum Modul - 40601 - der Fernuniversität Hagen

2 Lageparameter

2.1 Einfache Lageparameter aus einer gegebenen Messreihe ablesen

Erklärung

Die Hauptaufgabe der deskriptiven Statistik besteht darin, aus einer Menge an erhobenen Daten Informationen zu filtern. Die wichtigsten Fragen sind dabei zunächst:

- In welcher Größenordnung liegen die gemessenen Daten?
- Gibt es eine zentrale Tendenz?

Die Antwort auf diese Fragen liefern die sogenannten *Lageparameter*. Hier einige kurze, einführende Bemerkungen. Die Berechnung wird in den Beispielen klar.

1. Die Spannweite
Die Spannweite kann bei metrisch skalierten Merkmalen angegeben werden. Sie gibt Auskunft darüber, in welchem Bereich Werte gemessen wurden. Sie ergibt sich aus der Differenz zwischen dem größten und dem kleinsten gemessenen Wert.

2. Der Median
Der Median gibt die zentrale Tendenz der gemessenen Daten an. Er steht in der Mitte der aufsteigend sortierten Messreihe. Er kann, um überhaupt sortieren zu können, nur bei mindestens ordinal skalierten Merkmalen berechnet werden.
Hat man eine gerade Anzahl von Messungen, so bildet man das arithmetische Mittel der beiden zentral stehenden Werte.

3. Der Modalwert
Der Modalwert (oder Modus) ist gleich demjenigen Wert, der am häufigsten gemessen wurde. Dieser Parameter kann für jedes Skalenniveau angegeben werden.

Im Folgenden werden die Parameter stets für metrisch skalierte Merkmale berechnet.

Beispielaufgabe A

Wir betrachten die folgende Messreihe

$$6, \quad 6, \quad 3, \quad 4, \quad 2, \quad 5$$

Um die Lageparameter ablesen zu können, sortieren wir diese zunächst der Größe nach:

$$2, \quad 3, \quad 4, \quad 5, \quad 6, \quad 6$$

Dann markieren wir die größte und kleinste Merkmalsausprägung:

$$2, \quad 3, \quad 4, \quad 5, \quad 6, \quad 6$$

Die Spannweite ist gleich der Differenz aus diesen beiden Werten:

$$x_{\text{max}} - x_{\text{min}} = 6 - 2 = 4$$

Die Messreihe hat eine gerade Anzahl an Elementen, nämlich 6. Also betrachten wir zur Bestimmung des Medians die beiden zentral stehenden Werte:

$$2, \quad 3, \quad 4, \quad 5, \quad 6, \quad 6$$

Der Median soll „in der Mitte" von 4 und 5 stehen. Eine sinnvolle Wahl ist es, deren Mittelwert $x_{\text{med}} = \frac{4+5}{2} = 4,5$ zu betrachten.

Achtung! Hier muss man immer darauf achten, ob der Mittelwert sinnvoll interpretiert werden kann.

Der Modus schließlich, ist diejenige Merkmalsausprägung, die am häufigsten vorkommt: $x_{\text{mod}} = 6$.

$$\underbrace{2,}_{1\text{ mal}} \quad \underbrace{3,}_{1\text{ mal}} \quad \underbrace{4,}_{1\text{ mal}} \quad \underbrace{5,}_{1\text{ mal}} \quad \underbrace{6, \quad 6}_{2\text{ mal}}$$

Beispielaufgabe B

Wir betrachten die folgende Messreihe

$$8, \quad 5, \quad 8, \quad 4, \quad 8, \quad 5, \quad 4, \quad 4, \quad 6, \quad 3$$

und sortieren:

$$3, \quad 4, \quad 4, \quad 4, \quad 5, \quad 5, \quad 6, \quad 8, \quad 8, \quad 8$$

Wir markieren den größten und kleinsten Wert, sowie die beiden zentral stehenden Werte und bestimmen die Häufigkeiten der Werte

$$\underbrace{3,}_{1\text{ mal}} \quad \underbrace{4, \quad 4, \quad 4,}_{3\text{ mal}} \quad \underbrace{5, \quad 5,}_{2\text{ mal}} \quad \underbrace{6,}_{1\text{ mal}} \quad \underbrace{8, \quad 8, \quad 8}_{3\text{ mal}}$$

Die Spannweite beträgt also $x_{\text{max}} - x_{\text{min}} = 8 - 3 = 5$ und der Median ist $x_{\text{med}} = 5$.

In diesem Beispiel gibt es zwei Modalwerte, nämlich $x_{\text{mod1}} = 4$ und $x_{\text{mod2}} = 8$, da beide Werte mit der größten Häufigkeit auftreten.

2.2 Eine Häufigkeitstabelle aus einer Messreihe erstellen

Erklärung

In oben behandelten Beispielen hat man aus Gründen der Darstellung sehr kleine Messreihen mit wenigen Merkmalsausprägungen betrachtet. In der Anwendung ist dies jedoch fast nie der Fall, dort behandelt man oft Messreihen mit mehreren Tausend Merkmalsausprägungen. Diese stets vollständig zu verarbeiten ist sehr unhandlich und nicht zweckmäßig.

Es zeigt sich, dass es zur Beschreibung einer Messreihe genügt, die *unterschiedlichen* gemessenen Merkmalsausprägungen, gemeinsam mit ihren *Häufigkeiten* zu speichern. Eine Tabelle, die diese Daten enthält, nennt man *Häufigkeitstabelle*. Aus ihr kann man alle relevanten Parameter ablesen, wie in den folgenden Abschnitten gezeigt wird.[2]

Bei den Häufigkeiten unterscheidet man noch zwischen *absoluter* und *relativer* Häufigkeit. Erstere gibt an, wie oft ein Wert in der Messreihe vorkommt. Zweitere gibt an, welchen Anteil dieser Wert an der Menge der gemessenen Werte hat. Mit der relativen Häufigkeit lassen sich auch Messreihen unterschiedlicher Größe miteinander vergleichen. Bezeichnet h die absolute Häufigkeit eines Wertes in der Messreihe der Länge n, so lässt sich die relative Häufigkeit, genannt f, mit der Formel $f = {}^{h}/_{n}$ berechnen.

Beispielaufgabe

Wir betrachten die Messreihe

$$4,\ 4,\ 4,\ 7,\ 6,\ 5,\ 7,\ 6,\ 3,\ 4,\ 3,\ 4,\ 4,\ 4,\ 7,\ 7,\ 7,\ 7,\ 6,\ 4,\ 5,\ 3,\ 6,\ 6,\ 7$$

Es wurden fünf unterschiedliche Werte gemessen: 3, 4, 5, 6 und 7. Die Tabelle hat also diese als erste Zeile

x_i	3	4	5	6	7

Wir markieren die entsprechenden Merkmalsausprägungen

$$4,\ 4,\ 4,\ 7,\ 6,\ 5,\ 7,\ 6,\ 3,\ 4,\ 3,\ 4,\ 4,\ 4,\ 7,\ 7,\ 7,\ 7,\ 6,\ 4,\ 5,\ 3,\ 6,\ 6,\ 7$$

Nun stellen wir die Tabelle mit den entsprechenden absoluten Häufigkeiten auf:

x_i	3	4	5	6	7
$h(x_i)$	3	8	2	5	7

[2] Liegen sehr viele unterschiedliche Merkmalsausprägungen vor, so kann man diese noch in sogenannte *Klassen* zusammenfassen. Darauf wird im Abschnitt 4.1 kurz eingegangen.

Die Länge der Messreihe beträgt $n = 25$. Wir können also die relativen Häufigkeiten wie folgt berechnen

$$f(3) = \frac{h(3)}{n} = \frac{3}{25} = 0,12, \quad f(4) = \frac{h(4)}{n} = \frac{8}{25} = 0,32 \quad \text{usw.}$$

Das führt zu der folgenden Tabelle:

x_i	3	4	5	6	7
$h(x_i)$	3	8	2	5	7
$f(x_i)$	$0,12$	$0,32$	$0,08$	$0,20$	$0,28$