"KANN MAN AUF DIE (N+1)/2 REGEL BEIM SCHÄTZEN DES MEDIANS BEI KLASSIERTEN DATEN VERZICHTEN?"

von MARY ROUNCEFIELD

Originaltitel in "Teaching Statistics" Vol. 10 (1988), Nr. 1: Can We Abandon (N+1)/2 when Estimating the Media for Grouped Data?

Übersetzung und Bearbeitung: G. Scheu, Karlsruhe

Zusammenfassung: Die Autorin zeigt, daß die (N+1)/2 Regel bei klassierten Daten zur Bestimmung des Medians nicht benötigt wird. Sie leitet an Hand eines Beispiels ein einfaches Verfahren zur numerischen oder grafischen Bestimmung des Medians her.

ZDM-Klassifikation: K43, K40

Nachdem die Autorin IAN COOK's Artikel "Estimation of the Median from Grouped Data", veröffentlicht in Teaching Statistics, Januar 1987 gelesen hat, wollte sie gerne sein Argument bereits eine Ebene tiefer benutzen.

Sie ist dafür, daß bei nicht klassierten Daten, die bei einer Stichprobe mit N Beobachtungswerten (Stichprobenwerten) entstehen, mit der (N+1)/2 Regel gerechnet wird, um die Lage des Medians zu bestimmen. Bei einer ungeraden Anzahl von Werten wird damit die Lage des Medians bei dem Stichprobenwert in der Mitte der Reihe bestimmt. Bei einer geraden Anzahl von Werten wird damit die Lage des Medians zwischen zwei Stichprobenwerten der Reihe bestimmt.

Beispiel: Bei 10 Stichprobenwerten ist (N+1)/2 gleich 5.5, d.h. der Median ist zwischen dem fünften und sechsten Stichprobenwert. Bei klassierten Daten jedoch, die in einer Häufigkeitstabelle dargestellt sind, kann die N/2 Regel zur Bestimmung der Lage des Medians klar und einfach erläutert werden.

Hier ein typisches Beispiel aus einem Schulbuch. Die folgende Tabelle enthält die Größen von zwanzig elfjährigen Kindern (mathematisch gerundet und gemessen in cm).

Größe in cm	Häufigkeiten	obere Grenze des Intervalls	Summenhäufigkeiten
121-130	2	x < 130.5	2
131-140	4	x < 140.5	6
141-150	7	x < 150.5	13
151-160	5	x < 160.5	18
161-170	2	x < 170.5	20

Wenn die individuellen Größen der zwanzig Kinder gegeben wären, dann wäre der Median gleich dem Mittelwert der Größen des 10-ten und des 11-ten Kindes. Da diese Daten hier nicht mehr bekannt sind, muß der Median gleich dem Mittelwert der geschätzten Größen des 10-ten und 11-ten Kindes sein. Zuerst muß das Klassenintervall gesucht werden, das die Größen der beiden Kinder enthält. In der Spalte der Summenhäufigkeiten wird das Intervall gesucht, bei dem die Summe zuerst größer als 10 bzw. 11 wird. In dieser Spalte gibt es 13 Kinder mit Größen bis zu 150.5 cm. Deshalb muß der Median zwischen 140.5 cm und 150.5 cm liegen. In diesem Klassenintervall gibt es sieben Kinder.

Da die einzelnen Daten in der Klasse nicht bekannt sind, wird die Klassenintervallänge von 10 cm in 7 gleiche Teile aufgeteilt (in der Annahme, daß eine gleichmäßige Verteilung der Größen in diesem Intervall vorliegt). Das erste Kind in diesem Intervall ist das 7-te Kind und das letzte Kind in diesem Intervall ist das 13-te Kind dieser Gruppe mit zwanzig Kindern.

```
  <---      7 Kinder gibt es in diesem Intervall       --->
                                     !
     7       8       9      10       !  11      12      13
  !-------!-------!-------!-------!-------!-------!-------!
140.5 cm                             !                  150.5 cm
                                  Median
```

Die Lage des Medians ist genau zwischen dem 10-ten und 11-ten Teilintervall und er liegt deshalb auf der Grenze zwischen beiden. Diese Grenze ist die obere Schranke für die geschätzte Größe des 10-ten Kindes und deshalb wird der Median an der Stelle N/2 gefunden.
Die N/2 Regel gilt auch für eine ungerade Anzahl von Stichprobenwerten und ist leichter zu verstehen, als die (N+1)/2 Regel.

Ohne jede schwerfällige Formel kann der Median durch lineare Interpolation mit Hilfe des folgenden Verfahrens bestimmt werden.

1. Bestimme das Klassenintervall, das den Median enthält (140.5 cm - 150.5 cm im obigen Beispiel).
2. Wie groß ist die Intervallänge? (10 cm)
3. Wieviele Stichprobenwerte liegen in diesem Intervall? (7)
4. Teile das Intervall in die erforderliche Anzahl von Teilintervallen (10 cm : 7 = 1.43 cm)
5. Zeichne ein Diagramm zur Bestimmung der Lage des Medians (am Ende des 10-ten Teilintervalls).
6. Wieviele Teilintervalle sind vor dem Median? (4)
7. Berechne den Median wie folgt:
   Median = untere Grenze des Klassenintervalls + Anzahl der Längen der Teilintervalle
   (Median = 140.5 cm + (4 x 1.43 cm) = 140.5 cm + 5.72 cm
   = 146.22 cm)

Die gleichen Überlegungen können auch bei grafischen Verfahren benutzt werden. Wenn eine gleichmäßige Verteilung für jedes Klassenintervall vorausgesetzt wird, kann ein Polygon für die Summenhäufigkeit gezeichnet werden, und die "magische Linie" liegt bei N/2. Die Lösung ist unabhängig davon, ob vom Beginn oder vom Ende der Verteilung aus gearbeitet wird. IAN COOK hat gezeigt, daß dies im Falle der (N+1)/2 Regel nicht der Fall ist. Analog werden die Quartile an den Stellen N/4 und 3N/4 bei klassierten Daten gefunden. Das vorgestellte

Verfahren ist nicht nur leichter zu behalten, sondern es ist auch logisch korrekt und funktioniert sowohl bei rechnerischen als auch zeichnerischen Verfahren.

Also wie ist es, können wir auf die (N+1)/2 Regel bei klassierten Daten verzichten?

Literatur

COOK, I.: Estimation of the Median from Grouped Data. Teaching Statistics (9, 1, 26-29), 1987.