7.3.2 Modalwert

Eine Ausprägung a eines Merkmals X ist ein *Modalwert* (oder *Modus*) der Stichprobe $x_1, ..., x_n$, wenn es keine Ausprägung gibt, die in der Stichprobe häufiger vorkommt, d.h. wenn für die Häufigkeitsverteilung H_n der Stichprobe gilt:

$$H_n(a) \geq H_n(x_i) \quad \text{für alle } 1 \leq i \leq n.$$

Modalwerte sind Lagemaße, aber im Allgemeinen nicht eindeutig. Falls doch, werden sie mit x_{mod} bezeichnet und die Häufigkeitsverteilung heißt *unimodal*. Modalwerte lassen sich auch bei Stichproben von nominalen Merkmalen bestimmen.

7.3.3 Quantile

Zu $x_1, ..., x_n$ bezeichne $x_{(1)}, ..., x_{(n)}$ die *geordnete Stichprobe*, d.h. es gilt

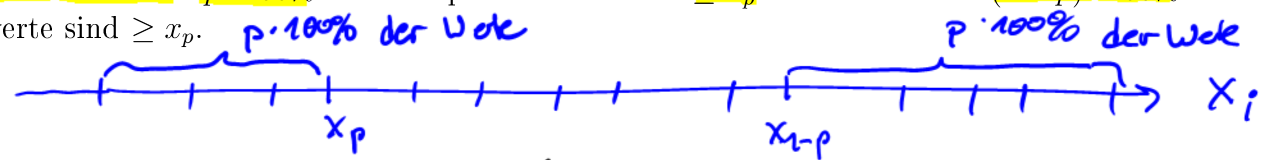
$$\{x_1, ..., x_n\} = \{x_{(1)}, ..., x_{(n)}\} \text{ und } x_{(1)} \leq ... \leq x_{(n)}$$

Zur Beschreibung der Struktur einer Stichprobe zieht man dann sogenannte *Quantile* heran:

Definition
Sei $p \in [0, 1]$. Eine Ausprägung $\boxed{x_p}$ heißt **p-Quantil** der Stichprobe $x_1, ..., x_n$, wenn für die Häufigkeitsverteilung h_n gilt:

$$\sum_{\substack{1 \leq i \leq s \\ a_i \leq x_p}} h_n(a_i) \geq p \quad \text{und} \quad \sum_{\substack{1 \leq i \leq s \\ a_i \geq x_p}} h_n(a_i) \geq 1 - p, \qquad (\star)$$

d.h. mindestens $p \cdot 100\%$ der Stichprobenwerte sind $\leq x_p$ und mindestens $(1 - p) \cdot 100\%$ der Stichprobenwerte sind $\geq x_p$.



i	1	2	3	4	5
x_i	4	5	1	2	6
$x_{(i)}$	1	2	4	5	6

$x_{(3)} = 4$, $x_3 = 1$

$n = 5$, $h_5(1) = \frac{1}{5}$

$p = \frac{1}{2}$, $x_{\frac{1}{2}} = 4 = x_1$ ist $\frac{1}{2}$-Quantil:

$$\sum_{a_i \leq x_{\frac{1}{2}} = 4} h_5(a_i) = h_5(1) + h_5(2) + h_5(4) = \frac{3}{5} \geq \frac{1}{2}$$

$$\sum_{a_i \geq x_{\frac{1}{2}} = 4} h_5(a_i) = \frac{3}{5} \geq \frac{1}{2}$$

Bemerkung

- Für $p = \frac{1}{2}$ heißt das p-Quantil $x_{0,5}$ *Median* (oder *Zentralwert*),

  d.h. mindestens 50 % der Stichprobenwerte sind größer gleich und mindestens 50 % sind kleiner gleich $x_{0,5}$.
- Das $\frac{1}{4}$-Quantil $x_{0,25}$ heißt *unteres Quartil*, das $\frac{3}{4}$-Quantil $x_{0,75}$ *oberes Quartil*.
- $x_{j \cdot 0,1}$ heißt *j-tes Dezil*.
- Im Fall $p = 0$ erhält man den kleinsten Wert, das *Minimum* $x_{min} = x_{(1)}$ der Stichprobe, im Fall $p = 1$ den größten Wert, das *Maximum* $x_{max} = x_{(n)}$.

Alle p-Quantile sind Lagemaße. Verglichen mit dem arithmetischen Mittel ist der Median robuster (d.h. weniger anfällig) gegenüber „Ausreißern".

Satz 7.3.3 (Berechnung des p-Quantils). *Sei $0 < p < 1$. Eine Ausprägung a ist genau dann ein p-Quantil der Stichprobe $x_1, ..., x_n$, wenn gilt:*

$$a = x_{(\lceil np \rceil)}, \quad \textit{falls } np \notin \mathbb{N} \quad \textit{bzw.} \quad x_{(np)} \leq a \leq x_{(np+1)}, \quad \textit{falls } np \in \mathbb{N}, \qquad (\star\star).$$

Hierbei steht $\lceil x \rceil$ für die kleinste ganze Zahl größer oder gleich x.

← obere Gaussklammer

Bsp $\lceil 2{,}1 \rceil = 3$, $\lceil -0{,}9 \rceil = 0$

Insbesondere gilt für den Median $x_{0,5}$ somit

$$x_{0,5} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \textit{falls } n \textit{ ungerade,} \\ a, & \textit{falls } n \textit{ gerade und } a \textit{ eine Ausprägung} \\ & \textit{mit } x_{\left(\frac{n}{2}\right)} \leq a \leq x_{\left(\frac{n}{2}+1\right)} \textit{ ist.} \end{cases}$$

Einschub 7.3.4. ...

1)

i	1	2	3	4	5	6	7	8	9
x_i	-	-	-	-	-	-	-	-	-
$x_{(i)}$	2	4	7	8	10	13	15	18	20

$n = 9$

$p = \frac{1}{4}$: $\lceil np \rceil = \lceil 9 \cdot \frac{1}{4} \rceil = \lceil 2{,}25 \rceil = 3 \overset{\text{Satz}}{\Rightarrow} x_{(3)} = x_{1/4} = 7$

$p = 3/4$: $\lceil np \rceil = \lceil 9 \cdot 3/4 \rceil = \lceil \frac{27}{4} \rceil = 7 \overset{\text{Satz}}{\Rightarrow} x_{(7)} = x_{3/4} = 15$

$p = \frac{1}{2}$: $\lceil 9 \cdot \frac{1}{2} \rceil = \lceil 4{,}5 \rceil = 5 \Rightarrow x_{1/2} = x_{(5)} = 10$

2) In 1) nur bis $i = 8$, also $n = 8$

$p = \frac{1}{2}$: $np = 8 \cdot \frac{1}{2} = 4 \in \mathbb{N} \overset{\text{Satz}}{\Rightarrow} x_{(np)} = x_{(4)} \leq a \leq x_{(np+1)} = x_{(5)}$

$\Leftrightarrow 8 \leq a \leq 10$, d.h. jede Ausprägung a, die $8 \leq a \leq 10$ erfüllt ist ein $x_{\frac{1}{2}}$ - Quantil, also sowohl 8 als auch 10 ist ein Median (also nicht eindeutig).

Daher gibt es folgende Konvention: wähle

$a = \frac{1}{2}\left(x_{(np)} + x_{(np+1)}\right)$ hier dann $a = 9$

7.4 Streumaße

Definition 7.4.1. Sei $x_1, ...x_n$ eine Stichprobe eines quantitativen Merkmals X. Eine Funktion

$$s : \mathbb{R}^n \to \mathbb{R}, \qquad (x_1, ...x_n) \mapsto s(x_1, ...x_n)$$

heißt *Streumaß*, wenn für alle $a \in \mathbb{R}$ und alle $(x_1, ...x_n) \in \mathbb{R}^n$ gilt:

$$s(x_1 + a, ..., x_n + a) = s(x_1, ...x_n),$$

d.h. durch Verschieben der Stichprobe ändert sich die Streuung nicht. Oftmals fordert man für ein Streumaß s auch noch, dass $s(0, 0, ..., 0) = 0$ gilt.

Beispiele 7.4.2.

- *Spannweite*: $x_{max} - x_{min} = x_{(n)} - x_{(1)}$
- *p-Quantilsabstand*, $p \in (0, \frac{1}{2})$: $x_{1-p} - x_p =$ Differenz von $(1-p)$- und p-Quantil; im Fall $p = \frac{1}{4}$ heißt der p-Quantilsabstand kurz *Quartilsabstand*.
- *Medianabweichung*: Median der Abstände $|x_1 - x_{0,5}|, |x_2 - x_{0,5}|, ..., |x_n - x_{0,5}|$ der Stichprobenwerte vom Median $x_{0,5}$
- *Mittlere absolute Abweichung vom Median*: $\tilde{s} = \frac{1}{n} \sum_{i=1}^{n} |x_i - x_{0,5}|$
- *Mittlere quadratische Abweichung vom arithmetischen Mittel* (oder *Varianz*):

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

- *Standardabweichung*: $s = \sqrt{s^2}$

Es handelt sich hier in allen Fällen um Streumaße, da zur Berechung stets Differenzen von Stichprobenwerten und Lagemaßen bzw. von zwei Lagemaßen verwendet werden. Bei Verschiebung der Stichprobe um a, heben sich die Verschiebungen somit gegenseitig auf.

Satz 7.4.3 (Berechnung der Varianz, Steinersche Formel)**.** *Die Varianz s^2 einer Stichprobe $x_1, ..., x_n$ lässt sich wie folgt berechnen:*

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

Beweis
Mithilfe der Binomischen Formel gilt:

$$\begin{aligned} s^2 &= \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 = \frac{1}{n} \sum_{i=1}^{n} \left(x_i^2 - 2x_i\overline{x} + \overline{x}^2 \right) = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \frac{2}{n}\overline{x} \sum_{i=1}^{n} x_i + \frac{1}{n} \cdot n\overline{x}^2 \\ &= \frac{1}{n} \sum_{i=1}^{n} x_i^2 - 2\overline{x}^2 + \overline{x}^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \overline{x}^2 \end{aligned}$$

□

□

Beispiele 7.4.4. Urliste $n = 13$

$x_1 = 2, 1, 3, 1, 3, 5, 5, 5, 5, 5, 3, 5, 5 = x_{13}$

x_4

geordnet

$x_{(1)} = 1, 1, 2, 3, 3, 3, 5, 5, 5, 5, 5, 5, 5 = x_{(13)}$

$x_{(4)}$ $x_{(7)}$

Median $x_{0,5} = x_{(\lceil 13 \cdot 0{,}5 \rceil)} = x_{(7)} = 5$

20% Quartilabstand $x_{0,2} = x_{(\lceil 13 \cdot 0{,}2 \rceil)} = x_{(\lceil 2{,}6 \rceil)} = x_{(3)} = 2$

$x_{0,8} = x_{(\lceil 13 \cdot 0{,}8 \rceil)} = x_{(11)} = 5$

$x_{1-p} - x_p = x_{0,8} - x_{0,2} = 5 - 2 = 3$

Medianabweichung $x_{0,5} = 5$

Abstände zum Median $|x_{(1)} - 5|, \ldots, |x_{(13)} - 5|$ ergibt
4, 4, 3, 2, 2, 2, 0, 0, 0, 0, 0, 0, 0

geordnet also

0, 0, 0, 0, 0, 0, 0, 2, 2, 2, 3, 4, 4

Medianabweichung (wie oben aus $\lceil 13 \cdot 0{,}5 \rceil$)

Mittlere absolute Abweichung vom Median $x_{0,5} = 5$

$$\tilde{s} = \frac{1}{13}\left(|x_1 - x_{0,5}| + \cdots + |x_{13} - x_{0,5}| \right) = \frac{17}{13}$$

$n = 13$

Mittelwert $\bar{x} = \frac{1}{13}\Big(\underbrace{1+1}_{=1 \cdot H_{13}(1)} + 2 + \underbrace{3+3+3}_{=3 \cdot H_{13}(3)} +$

$+ \underbrace{5+5+5+5+5+5+5}_{=5 \cdot H_{13}(5)} \Big) = \frac{1}{13}\Big(1 \cdot \overbrace{H_{13}(1)}^{2} +$

$2 \cdot H_{13}(2) + 3 \cdot H_{13}(3) + 5 \cdot H_{13}(5) \Big) = \frac{48}{13}$ ← Tim

Varianz (Steiner)

$$s^2 = \frac{1}{13}\left(1^2 + 1^2 + 2^2 + 3^2 + 3^2 + 3^2 + 5^2 + \cdots + 5^2 \right) - \bar{x}^2$$
$$= \frac{1}{13}\left(1^2 \cdot H_{13}(1) + 2^2 \cdot H_{13}(2) + 3^2 \cdot H_{13}(3) + 5^2 H_{13}(5) \right) - \bar{x}^2$$

$$= \frac{1}{13}\left(2+4+27+7\cdot 25\right) - \left(\frac{48}{13}\right)^2 = \cancel{2} \dots$$

<u>Vorteil</u> <u>Steiner bei Berechnung</u>:

$$\left.\begin{array}{ll} \text{Mittelwert} & x_i^2 \\ \text{Mittelwert} & x_i \end{array}\right\} \xRightarrow[\text{Satz 7.4.4}]{\text{im}} \text{alle } H_n(a_i) \text{ gleich}$$