Median

Eine Liste von Datenwerten wird der Größe nach geordnet.

Der Median (Zentralwert) teilt diese geordnete Liste der Datenwerte in ***eine untere und eine obere Teilliste***.

$$\overline{x_M} = \begin{cases} x_{\frac{n+1}{2}} & \text{für n ungerade} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) & \text{für n gerade} \end{cases}$$

=> Bei einer *<u>ungeraden Zahl</u>* von Datenwerten entspricht der Median dem mittleren Wert.

*Beim Teilen in untere und obere Teilliste zur Ermittlung der **Quartile** wird dann dieser Wert (=Median) nicht weiter berücksichtigt.*

Quantil(e) & Quartil(e)

Vorgehensweise: Die beiden Teillisten weiter zerlegt.

- Das untere Quartil (auch: 1.Quartil q_1) ist der Median der unteren Teilliste.
- Das obere Quartil (auch: 3.Quartil q_3) ist der Median der oberen Teilliste.
- Der Median (Zentralwert) entspricht dann dem mittleren Quartil (auch: 2.Quartil q_2).

Diese Lageparameter sind für ordinal- und metrisch-skalierte Merkmale bestimmbar. Sie beziehen sich auf die hierarchisch (nach der Größe der Werte der Merkmalsausprägungen) geordneten Daten.

Sind die Daten nach der Größe geordnet, so kann man die Daten unterteilen, in dem man eine Stelle auf der Skala der Merkmalsausprägungen bestimmt, vor der (die Stelle selbst eingeschlossen) mind. p · 100 % und nach der (die Stelle selbst eingeschlossen) mind. (1 − p) · % der Daten liegen. Im Grunde genommen kann man solche Stellen durch Auszählen bestimmen.

Quantile

<u>Definition:</u>

Als p-Quantil x_p bezeichnet man diejenige reelle Zahl (Merkmalsausprägung), für die gilt:

***mindestens p · 100 %** der metrisch-skalierten (ordinalskalierten) Merkmalsausprägungen der Daten sind kleiner oder gleich x_p und*

***mindestens (1 − p) · 100 %** der Merkmalsausprägungen der Daten sind größer oder gleich x_p.*

<u>Satz:</u> Seien $x_1, x_2, \ldots, x_n$ die erhobenen und metrisch skalierten Merkmalsausprägungen in einer Stichprobe mit Umfang n, dann gilt für 0 < p < 1:

$$\overline{x_p} = \begin{cases} x_{[n\cdot p]+1} & \text{für } (\mathrm{n}\cdot\mathrm{p}) \text{ nicht ganzzahlig} \\ \frac{1}{2}\left(x_{n\cdot p} + x_{n\cdot p+1}\right) & \text{für } (\mathrm{n}\cdot\mathrm{p}) \text{ ganzzahlig} \end{cases}$$

Anmerkungen:

• Sind die Merkmalsausprägungen ***<u>ordinalskaliert</u>***, so umfasst das p-Quantil bei n · p nicht ganzzahlig *statt des Mittelwertes zwischen zwei Merkmalsausprägungen* **beide Merkmalsausprägungen**, d.h. das p-Quantil besteht als aus einer ***zweielementigen Menge***.

• Die [x] in der Formel bezeichnet die Gauß-Klammer. Ihr Wert ist der ganzzahlige Anteil der Zahl x. Beispielsweise gilt [3, 45] = 3 oder [18, 999] = 18.

<u>Beispiel:</u> *Man hat einen Datensatz mit 15 Daten und ordnet diese nach ihrer Größe.*

Wo liegt die Stelle, für die gilt ...

*... **mindestens 25 Prozent** (p = 0, 25) der Daten (die Stelle selbst eingeschlossen) liegen vor dieser Stelle?*

*... **mindestens 75 Prozent** ((1 − p) = 0, 75) der Daten (die Stelle selbst eingeschlossen) liegen hinter dieser Stelle?*

In diesem Beispiel ergibt 25 % von 15 einen Dezimalbruch: 0, 25 · 15 = 3,75. Da man die Daten nicht teilen kann, kann man sagen, dass das vierte Datum die gesuchte Stelle ist, denn vor dieser Stelle (die Stelle selbst eingeschlossen) liegen vier Daten, also etwas mehr als 25 % der Daten. Nach dieser Stelle (die Stelle selbst eingeschlossen) liegen 12 Daten, was etwas mehr als 75 % des Datensatzes entspricht.

Mathematisiert man dieses Prinzip zur Bestimmung von Lageparametern einer Häufigkeitsverteilung, so erhält man allgemein die Quantile und spezieller für p = 0,25 die Quartile q_1 ,für 2p = 0,5 den Median (= Quartile q_2) und für (1-p) = 0,75 die Quartile q_3.

Die Idee besteht also darin, die Datenwerte in vier Klassen aufzuteilen:

- Ein Viertel der Werte liegt unterhalb des unteren Quartils q_1.
- Ein Viertel der Werte liegt zwischen dem unteren Quartil q_1 und dem Median.
- Ein Viertel der Werte liegt zwischen dem Median und dem oberen Quartil q_3.
- Ein Viertel der Werte liegt oberhalb des oberen Quartils q_3.
- Im mittleren Bereich zwischen dem unterem und oberen Quartil liegen 50% der Datenwerte, die sogenannten „mittleren 50%“ der Datenwerte.

Beispiel n = 15:

2	3	3	4	8	10	11	14	14	16	17	18	20	20	24
			q_1				$\overline{x}_M$				q_3			

Beispiele zur Quartileberechnung:

Fall 1: n*p ist ganzzahlig

Daten:	2,1	3,4	7,3	8,9	9,3	10,1	11,2	11,9	→ n = 8

Berechnung des 1. Quartils:

1. Bestimmung des unteren und oberen Datenpunktes:
   oberer Datenpunkt:
   n· 0,25
   8· 0,25 =2 → Stelle 2 → Wert der Stelle 2: 3,4
   unterer Datenpunkt:
   n· 0,25+1
   8· 0,25 +1=3 → Stelle 3 → Wert der Stelle 5: 7,3

2. Berechnung von Q1:
   (Wert des unteren Datenpunktes + Wert des oberen Datenpunktes)/2
   (3,4 + 7,3)/2=5,35
   → Q1: 5,35

Fall 2: n*p ist nicht ganzzahlig

Beispiel:

Daten:	2,1	3,4	7,3	8,9	9,3	10,1	11,2	→ n = 7

Berechnung des 1. Quartils:

n = 0,25 => 7 * 0,25 = 1,75 => aufrunden => Stelle 2

=> Wert der Stelle 2: 3,4 => Q1: 3,4

Beispiel 1:

Daten:	1	3	4	7	9	10	12

n=7

7*0,5=3,5	Median:	7
7*0,25=1,75	Q1:	3
7*0,75=5,25	Q3:	10

(Andere Maßzahlen: Mittelwert: 6,57; Varianz: 16,29; Standardabweichung: 4,04)

Beispiel 2:

Daten:	0	5	2	7	8	1

!!!ordnen!!!

n=6

6*0,5=3	Median:	3,5
6*0,25=1,5	Q1:	1
6*0,75=4,5	Q3:	7

(Andere Maßzahlen: Mittelwert: 3,84; Varianz: 10,97; Standardabweichung: 3,31)

Beispiel 3:

Daten:	1	2	3	4	5	6	7	8	9	10

n=10

10*0,5=5	Median:	5,5
10*0,25=2,5	Q1:	3
10*0,75=7,5	Q3:	8

(Andere Maßzahlen: Mittelwert: 5,5; Varianz: 9,17; Standardabweichung: 3,03)

Beispiel 4:

Daten:	9	3	10	16	1	7	12	4

n=8

8*0,5=4	Median:	8
8*0,25=2	Q1:	3,5
8*0,75=6	Q3:	11

(Andere Maßzahlen: Mittelwert: 7,75; Varianz: 25,07; Standardabweichung: 5,01)

Beispiel 5:

Daten:	1	5	5	8	10

n=5

5*0,5=2,5	Median:	5
5*0,25=1,25	Q1:	5
5*0,75=3,75	Q3:	8

(Andere Maßzahlen: Mittelwert: 5,8; Varianz: 11,7; Standardabweichung: 3,42)