

## Zur Exkursion „Statistik mit dem Computer“

Beim Zeichnen von Boxplots (mit Excel) ist das Sortieren der Daten und das Bestimmen der Quartile „durch Ablesen“ umständlich. Daher greift man auf die verfügbaren Befehle zurück. Dabei können kleine Abweichungen auftreten, was aber für die Aussagen, die die Diagramme machen, irrelevant ist. Wie Excel intern die Quartile berechnet, liest man nach unter <http://www.learn-line.nrw.de/angebote/eda/medio/tipps/excel-quartile.htm>

**Nach welchem Verfahren berechnet Excel eigentlich Quartile?**  
**Hinter die Kulissen von Excel geschaut**

**Wo ist die Mitte?** Der Median ist in der Literatur durchgängig definiert als der "Wert in der Mitte" zwischen Minimum und Maximum eines Datensatzes, wobei die Mitte durch *Abzählen der Daten in sortierter Reihenfolge* gefunden wird. Liegt die Mitte zwischen zwei Daten (bei gerader Anzahl), so wird deren arithmetischer Mittelwert angenommen.

84 - Maximum  
73  
71  
64 - ob. Quartil (67,5)  
62  
56 - Median  
42  
42  
34 - unt. Quartil (38)  
23  
16 - Minimum

Ähnlich sind die Quartile definiert: Das obere Quartil ist der Wert "in der Mitte" zwischen Median und Maximum; entsprechend ist das untere Quartil der Wert "in der Mitte" zwischen Median und Minimum. Die genaue Berechnung erfolgt bei einzelnen Programmen jedoch durchaus unterschiedlich. Während z. B. Medass-Light tatsächlich den mittleren Wert bzw. den Mittelwert der Nachbarwerte ausgibt, wichten andere Programme wie z. B. Excel hier die benachbarten Werte, so dass die Berechnung nicht mehr so einfach nachzuvollziehen ist.

Bei der Beispielliste links sind sich Excel und Medass-Light noch einig, fügt man jedoch ein weiteres Element - z. B. 11 - hinzu, so kommen die beiden Programme bei den Quartilen zu unterschiedlichen Werten: Bei Medass-Light bleibt erwartungsgemäß das obere Quartil unverändert bei 67,5 und das untere Quartil wird mit 28,5 berechnet - dem Mittelwert von 23 und 34. Excel dagegen liefert die Werte 65,75 und 31,25!

Diese Werte liegen zwar auch *zwischen* den benachbarten Listenwerten, sind aber offenbar nicht deren arithmetisches Mittel.

Um den Algorithmus von Excel verstehen zu können, sei im Folgenden **n** die **Anzahl der Listenelemente** und **X<sub>i</sub>** der **Wert an der Stelle i**. (Die Werte sind aufsteigend sortiert!) Außerdem benötigen wir noch die Größen **Position** und **Gewicht**.

**unteres Quartil:** Für das untere Quartil definieren wir

$$\text{Rang} := 0,25 \cdot (n-1) + 1, \text{ Index} = \text{GANZZAHL}(\text{Rang}) \text{ und } \text{Gewicht} := \text{Rang} - \text{Index}$$

Excel rechnet nun mit

$$\text{Unteres Quartil} := X_{\text{Index}} + \text{Gewicht} \cdot (X_{\text{Index}+1} - X_{\text{Index}})$$

Am Beispiel der Liste 11, 16, 23, 34, 42, 42, 56, 62, 64, 71, 73, 84 ist demnach

$$n = 12, \text{ Rang} = 0,25 \cdot 11 + 1 = 3,75, \text{ Index} = 3 \text{ und } \text{Gewicht} = 3,75 - 3 = 0,75$$
$$\text{Unteres Quartil} = X_3 + 0,75 \cdot (X_4 - X_3) = 23 + 0,75 \cdot (34 - 23) = 31,25.$$

**oberes Quartil:** Das obere Quartil wird fast genauso berechnet; lediglich beim Rang wird die 0,25 durch 0,75 ersetzt:

$$\text{Rang} := 0,75 \cdot (n-1) + 1, \text{ Index} = \text{GANZZAHL}(\text{Rang}) \text{ und } \text{Gewicht} := \text{Rang} - \text{Index}$$

Excel berechnet wieder

$$\text{Oberes Quartil} := X_{\text{Index}} + \text{Gewicht} \cdot (X_{\text{Index}+1} - X_{\text{Index}}).$$

Für unser Beispiel bedeutet dies:

$$\text{Rang} = 0,75 \cdot 11 + 1 = 9,25, \text{ Index} = 9 \text{ und } \text{Gewicht} = 9,25 - 9 = 0,25$$
$$\text{Oberes Quartil} = X_9 + 0,25 \cdot (X_{10} - X_9) = 64 + 0,25 \cdot (71 - 64) = 65,75.$$

**beliebige Quantile:** Die zu den Quartilen benachbarten Listenwerte werden also unterschiedlich stark gewichtet. Das Verfahren lässt sich sofort auf beliebige Quantile übertragen, wenn man bei der Berechnung des Ranges statt 0,25 bzw. 0,75 jeweils die entsprechenden Prozentsätze der Quantile einsetzt.