

Zahlenreihen visualisieren mit Boxplots



Papa, was ist ein Boxplot?

Wie lange dauert und was kostet dies und jenes im Durchschnitt, höchstens, mindestens und am wahrscheinlichsten? Statistik ist das halbe Leben, in Form von Zahlen und in Form von Grafiken. Stefan, kannst du dazu eine Aufgabe stellen?

dnpCode: A1007DojoAufgabe

Wer übt, gewinnt

In jeder dotnetpro finden Sie eine Übungsaufgabe von Stefan Lieser, die in maximal drei Stunden zu lösen sein sollte. Wer die Zeit investiert, gewinnt in jedem Fall – wenn auch keine materiellen Dinge, so doch Erfahrung und Wissen.

Es gilt:

- Falsche Lösungen gibt es nicht. Es gibt möglicherweise elegantere, kürzere oder schnellere Lösungen, aber keine falschen.
- Wichtig ist, dass Sie reflektieren, was Sie gemacht haben. Das können Sie, indem Sie Ihre Lösung mit der vergleichen, die Sie eine Ausgabe später in dotnetpro finden.

Übung macht den Meister. Also – los geht's. Aber Sie wollten doch nicht etwa sofort Visual Studio starten ...



Haben Sie sich auch schon mal gefragt, wie Sie eine Zahlenreihe anschaulich darstellen können? Eine übersichtliche Form bieten die sogenannten Boxplots [1]. Darauf hat mich meine Tochter gebracht. Sie musste Boxplots als Hausaufgabe in Mathe zeichnen. 7. Klasse! Da werden Sie sich als gestandener Softwareentwickler doch nicht wegducken wollen, oder?

Boxplots dienen dazu, die Verteilung der Werte zu visualisieren. Dazu werden neben dem kleinsten und dem größten Wert auch die sogenannten Quartile visualisiert. Die Zahlenreihe wird in vier Bereiche unterteilt. Sind diese Bereiche gleich groß, bedeutet das, dass die Werte in der Zahlenreihe gleichmäßig verteilt sind. Dazu ein Beispiel. Wenn Sie sich fragen, ob der Pizzadienst um die Ecke immer gleich lange braucht, um die Pizza anzuliefern, oder ob es auch schon mal große Ausreißer gibt, können Sie die Werte mit einem Boxplot schön visualisieren. Ich habe nicht gemessen, aber die Werte könnten so aussehen:

18, 24, 19, 19, 20, 25, 24, 18, 24, 17

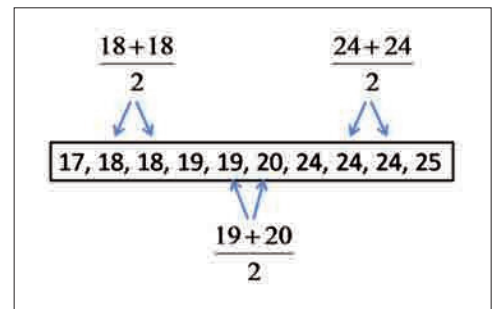
Aus den reinen Zahlen wird man nicht sofort etwas erkennen können. Helfen würde schon mal der Mittelwert. Wie der berechnet wird, ist jedem sofort klar. Aber wie sieht es mit dem Median aus? Erinnern Sie sich noch?

Zur Berechnung des Medians müssen die Werte zunächst sortiert werden. Dann nimmt man einfach den mittleren Wert. Wenn die Anzahl der Werte gerade ist, nimmt man die beiden mittleren Werte und bildet daraus den Mittelwert. Im obigen Beispiel sind es zehn Werte. Nach dem Sortieren sieht die Zahlenreihe wie folgt aus:

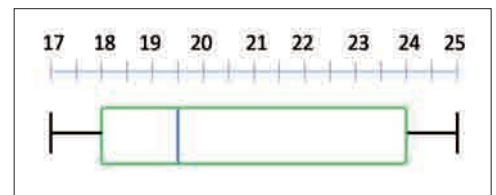
17, 18, 18, 19, 19, 20, 24, 24, 24, 25

Die beiden mittleren Werte sind 19 und 20. Der Mittelwert aus diesen ist $(19 + 20) / 2 = 19,5$.

Beim unteren und oberen Quartil geht es sinngemäß, wie Abbildung 1 zeigt. Aus den Werten wird dann ein Boxplot erstellt, in dem Minimum und Maximum die untere bzw. obere Begrenzung bilden. Dazwischen werden die beiden Quartile sowie der Median eingezeichnet, fertig



[Abb. 1] Die Ausgangs-Zahlenreihe.



[Abb. 2] Grafische Darstellung mit einem Boxplot.

ist der Boxplot. Abbildung 2 zeigt das Ergebnis. Aufgabe des dojos ist es, ein Control zu entwickeln, das einen Boxplot darstellt. Ob Sie das mit Windows Forms, WPF oder Silverlight lösen, ist egal. Selbst eine Ausgabe auf der Konsole wäre reizvoll.

Die Logik von der Umsetzung trennen

Bedenken Sie, dass eine saubere Trennung der Belange wichtig ist. Die Logik des Boxplots sollte unabhängig sein von den technischen Details des Controls. Das hilft in jedem Fall beim automatisierten Testen. Ferner schaffen Sie durch diese Trennung die Grundlage dafür, dass Sie das Control später auch einmal in einer anderen Technologie realisieren können.

Denken Sie auch darüber nach, wie die Schnittstellen der beteiligten Funktionseinheiten am besten aussehen. Sollte das Control die komplette Zahlenreihe erhalten? Oder direkt die Quartile und die anderen benötigten Werte? Viele Fragen, nächsten Monat gibt es hier wieder Antworten. Bis dahin, viel Spaß mit Boxplots! [ml]

[1] <http://de.wikipedia.org/wiki/Boxplot>