

Voraussetzung und verwandte Themen

Für diese Beschreibungen sind Grundlagen der Statistik und insbesondere der statistischen Verteilungen vorteilhaft. Weiterführendes Thema ist

www.versuchsmethoden.de/Normalverteilung.pdf

www.versuchsmethoden.de/tTest.pdf

Einführung

Der Boxplot, auch Box-Whisker-Plot genannt, ist eine spezielle Darstellung der Häufigkeitsverteilung. Ein Boxplot soll schnell einen Eindruck darüber vermitteln, in welchem Bereich die Daten liegen und wie sie sich verteilen.

Ziel und Nutzen

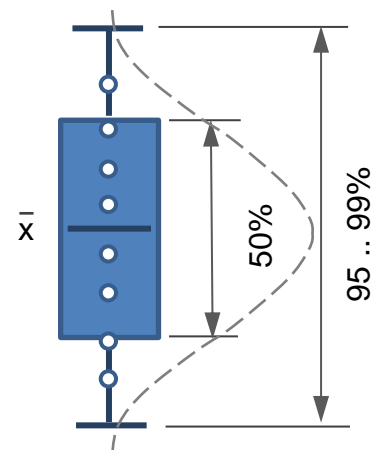
Der große Vorteil ist hierbei, dass mehrere Datenreihen als Boxplots nebeneinander sehr komprimiert darstellbar sind, während ein Histogramm die gesamte Diagrammbreite für nur eine Datenreihe benötigt.

Grundlagen

Grundsätzlich besteht ein Boxplot aus einem Rechteck, Box genannt, und zwei Verlängerungen mit begrenzenden Linien, meist Whisker oder Antennen genannt. Es gibt unterschiedliche Verfahren diese Bereiche zu bestimmen:

1. Normalverteilung

Aus den Daten werden Mittelwert und Standardabweichung berechnet. Mit diesen Kennwerten lassen sich die Bereiche von 50% und den zu definierenden äußeren Whisker für z.B. 95% oder 99% festlegen. Diese Methode liefert gute Schätzwerte für den Box und die Whisker und ist nicht so stark von der zufälligen Lage einzelner Punkte abhängig.

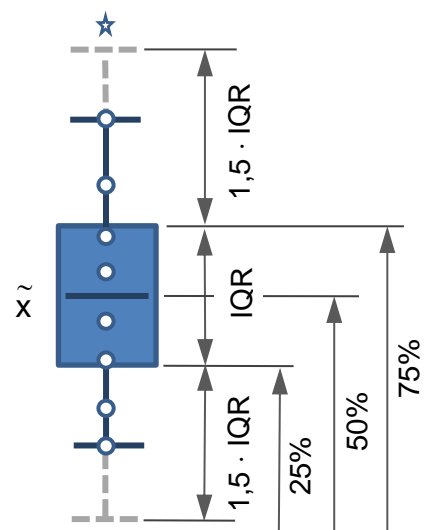


2. Verteilungsunabhängig

Aus den Häufigkeiten der sortierten Daten werden die Bereiche direkt bestimmt. Die Häufigkeit jedes i -ten Datenpunktes wird durch $H=i/(n+1)$ definiert. Bestimmte Literaturquellen verwenden $H=i/n$, wodurch die Bereiche etwas anders liegen.

In der Mitte des Boxplots befindet sich der Median (verteilungsunabhängiger Zentralwert). Der Box, in dem sich 50% aller Werte befinden, nennt man auch den Interquartil-Range, kurz IQR. Die Whisker liegen $1,5 \cdot \text{IQR}$ ober- und unterhalb des 50%-Bereiches. Es wird jedoch nicht genau der Bereich von $\pm 1,5 \cdot \text{IQR}$ dargestellt, sondern nur bis zum nächst innerhalb liegendem Werte.

Liegen Werte außerhalb $\pm 1,5 \cdot \text{IQR}$ werden diese als „milde Ausreißer“ betrachtet und als Stern markiert.



Anwendung in Visual-XSel 14

www.crgraph.de

	A	B	C	D	E	F	G	H	I
1	Gew	Krfst	Zyl	Hubr	KW	Achse	Beschl	Verbr	
2	1340	Benz		4	1599	90	3,64	10,7	5,8
3	1350	Benz		4	1995	105	3,39	8,7	5,9
4	1375	Benz		4	1995	125	3,73	7,7	6,4
5	1460	Benz		6	2996	195	3,46	6,0	8,3
6	1395	Dies		4	1995	105	3,07	8,9	4,5
7	1450	Dies		4	1995	130	2,56	7,4	4,8
8	1495	Dies		4	1995	150	2,81	6,9	5,2
9	1670	Benz		4	1995	125	3,64	9,1	6,6
10	1730	Benz		6	2996	160	3,15	7,6	7,6
11	1780	Benz		6	2996	200	3,15	6,4	7,7

Visual-XSel - Diagrammtypen - Diagramm 1

Favoriten | Standard | Statistik | Spezial | Tabelle

Linie Netz Säulen Pareto **Boxplot** Histogramm Normal-Vert. Log-Normal Weibull Nur Text

Favoriten bearbeiten : Zuerst mit rechter Maustaste hier auf ein Symbol klicken

Reset Fav.

Optionen:

- Diagrammrahmen
- Hintergrundfarbe/-bild
- 3D-Effekte
- Schatteneffekt

Breite: Weitere Einstell...

Beispiel + erwartete Datenaufteilung

	A	B	C	D	E
1	A	B	C		
2	3	5	4		
3	6	7	6		
4	7	8	7		
5	7	9	8		
6	8	10	9		
7					
8					

Datenauswahl

Datenspalten (Doppelklick)

[A] Gew
[D] Hubr
[E] KW
[F] Achse
[G] Beschl

Darzustellende Daten

[H] Verbr

Merkmale für Gruppierung

[B] Krfst - [Kategorie] **[C] Zyl**

Benz 4
Dies 6
8

Mehrfachdiagramme

OK Schließen Hilfe

Gibt es mehrere Datenspalten (siehe \Beispieldaten\Beispiel_Verbrauch.vxt), so erscheint eine Auswahlbox, welche als Boxplot angezeigt und welche evtl. als Gruppierung verwendet werden sollen.

Boxplot

Für den Fall, dass mindestens eine Datenreihe nicht normalverteilt ist, wird automatisch die Vorgabe für die Bestimmung der Bereiche auf Datenhäufigkeit umgeschaltet.

Ist die Gruppengröße unter Darstellung größer als 1, so haben jeweils mehrere Boxplots die gleiche Farbe.

Unter „Test auf Gleichheit“ gibt es drei stat. Hypothesentests zur Auswahl. Sind die Daten normalverteilt, wird der bekannte t-Test verwendet. Beim Tukey – Simultantest werden alle Datenreihen gemeinsam getestet. Grundsätzlich wird aber bei allen Tests vorausgesetzt, dass die Datenreihen unabhängig voneinander sind.

Da hier einigen Datenreihen nicht normalverteilt sind, ist nur der u-Test zu empfehlen. Zwischen dem ersten und letzten Boxplot ist ein Pfeil dargestellt, der die Hypothese auf gleichen Median „bestätigt“.

Wo rote Linien vorhanden sind, wird die Hypothese abgelehnt.

