



Klasseneinteilung

Häufig stecken wir bestimmte Menschen in Schubladen, weil es unser Leben erleichtert. Dass dies nicht unbedingt gut ist, wurde hoffentlich im Ethik- oder Reliunterricht hinreichend besprochen. In der Statistik jedoch ist das anders. Hier ist es oft das einzige Mittel besonders großen Datensätzen Herr zu werden. Dass dies zu Fehlern führen kann, nimmt man billigend in Kauf. Was tut man nicht alles für die Übersichtlichkeit.

Beispiel

Betrachten wir ein Beispiel, bei dem wir fast schon wieder ethische Bedenken haben. Wir kategorisieren Menschen in Untergewicht, Normalgewicht, Übergewicht und Adipositas (Fettleibigkeit). Dies kann man mit Hilfe des sogenannten BMI (body mass index) machen. Er berechnet sich:

$BMI = \frac{m}{l^2}$, wobei m die Masse in kg und l die Körpergröße in m ist.

Menschen mit einem BMI von weniger als 18,5 haben Untergewicht, bis 25 ist Normalgewicht, bis 30 Übergewicht und ab 30 Adipositas.

Es ergeben sich mathematisch also folgende Klassen (Intervalle):

$[0; 18,5[$ $[18,5; 25[$ $[25; 30[$ und $[30; [$

Betrachten wir nun eine Messreihe der BMI von 10 Schülern: 203; 175; 314; 235; 273; 257; 194; 328; 229; 183

Nun ordnen wir die Liste und teilen sie in Klassen ein:

BMI	0-18,5	18,5-25	25-30	> 30
Ergebnisse	17,5; 18,3	19,4; 20,3; 22,9; 23,5	25,7; 27,3	31,4; 32,8
Absolute Anzahl	2	4	2	2



Klasseneinteilung

Um größere Datensätze zu kategorisieren, kann man sie in Klassen einteilen. Klassen sind offene oder geschlossene Intervalle zum Teil unterschiedlicher Größe.





Histogramme

Eine weitere graphische Darstellung von Daten sind Histogramme. Sie ähneln dem Säulendiagramm, berücksichtigen jedoch die Klassenbreite.

Beispiel

Ein Paketdienstleister teilt die zu versendenden Pakete in drei Klassen ein. Es wird jeweils die Summe aus der längsten und der kürzesten Kante gebildet. Paketgröße S haben Pakete mit einer Summe von 50cm, M von 70cm und L bis 1m. (Größere Pakete werden nicht versandt.)

Ein Paketshop hat folgenden Tagesversand an Paketen mit einer Kantenlängensumme in cm.

47; 57; 95; 43; 56; 74; 21; 45; 63; 36; 57; 61; 83; 55; 55; 67; 33; 57

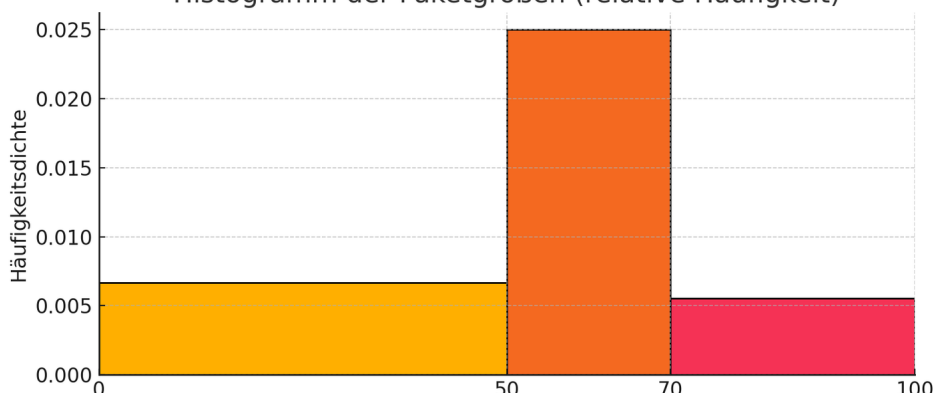
Hier die Zahlenfolge sortiert von klein nach groß:

21 ; 33 ; 36 ; 43 ; 45 ; 47 ; 55 ; 55 ; 56 ; 57 ; 57 ; 57 ; 61 ; 63 ; 67 ; 74 ; 83 ; 95

Ein Klasseneinteilung sieht dann wie folgt aus:

Größe	S	M	L
Intervall	0cm-50cm	50cm-70cm	70cm-100cm
Klassenbreite	50	20	30
Absolute Häufigkeit	6	9	3
Relative Häufigkeit	$\frac{6}{18} = \frac{1}{3}$	$\frac{9}{18} = \frac{1}{2}$	$\frac{3}{18} = \frac{1}{6}$
Häufigkeitsdichte = $\frac{\text{Relative Häufigkeit}}{\text{Klassenbreite}}$	$\frac{1}{3} : 50 = \frac{1}{150} = 0,0067$	$\frac{1}{2} : 20 = \frac{1}{40} = 0,025$	$\frac{1}{6} : 30 = \frac{1}{180} = 0,006$

Histogramm der Paketgrößen (relative Häufigkeit)





Quartile

Im letzten Kapitel haben wir festgestellt, dass der Median häufig der Wert ist, der eine Messreihe am besten charakterisiert, da er unauffällig gegen Ausreißer ist. Dennoch ist es gerade bei umfangreichen Messreihen nicht sinnvoll, diese nur durch einen einzigen Lageparameter zu charakterisieren. Deshalb hat man sich überlegt, jeweils den Median der unteren und der oberen Hälfte der geordneten Liste auch noch anzugeben. Man nennt sie oberes bzw. unteres Quartil (lat.: quartum oder vierte Teil). Der Median und die beiden Quartile teilen also die Messreihe nun in vier gleiche Teile.



Quartil

Der Median der unteren oder oberen Hälfte einer geordneten Liste, heißt unteres Quartil q_u bzw. oberes Quartil q_o . Das untere Quartil lässt sich wie folgt berechnen:

Berechne $0,25 * n$

1. Fall: Ist das Ergebnis keine natürliche Zahl, so runde auf. DU erhältst die Zahl i . $q_u = x_i$

2. Fall: Ist das Ergebnis eine natürliche Zahl, so ist $i = 0,25 * n$

$$q_u = \frac{x_i + x_{i+1}}{2}$$

Beispiel

Betrachten wir eine Messreihe mit $n = 7$ Werten, so ergibt sich für das untere Quartil:

$$0,25 * n = 0,25 * 7 = 1,75$$

$$i = 2$$

$$q_u = x_2$$

Besitzt eine Messreihe $n = 9$ Werte, so ergibt sich:

$$0,25 * n = 0,25 * 9 = 2,25$$

$$i = 3$$

$$q_u = x_3$$

Wir runden immer auf!

Liegt hingegen eine Messreihe mit $n = 8$ Werten vor, so rechnen wir:

$$0,25 * n = 0,25 * 8 = 2$$

$$i = 2$$

$$i + 1 = 3$$

$$q_u = \frac{x_2 + x_3}{2}$$





INPUT: Klasseneinteilung und Histogramme

Mathematik Statistik R 9

Boxplots

Nun haben wir so viele Parameter kennengelernt. Da kann man leicht die Übersicht verlieren. Deshalb gibt es die sogenannten Boxplots. Dabei handelt es sich um Diagramme, die die Lageparameter aus dem letzten Kapitel übersichtlich darstellen. Minimum, Maximum, Median und die beiden Quartile werden für das Erstellen eines Boxplots benötigt. Machen wir uns also an die Arbeit.

Beispiel

Bei den Olympischen Spielen 2012 in London ergab sich folgende Rangliste beim Kugelstoßen der Männer:

Teilnehmer	3	4	5	6	Result
Tomasz Majewski	21.87	X	21.72	21.89	21.89
David Storl	21.46	X	X	X	21.86
Reese Hoffa	21.23	21.11	19.53	X	21.23
4 Christian Cantwell	X	X	20.65	21.19	21.19
5 Dylan Armstrong	20.74	X	X	20.34	20.93
6 German Lauro	20.84	20.34	20.65	X	20.84
7 Asmir Kolasinac	X	20.54	20.46	X	20.71
8 Pavel Lyzhyn	X	19.93	20.04	X	20.69

Wir sortieren natürlich von klein nach groß. So ergibt sich: 20,69; 20,71; 20,84; 20,93; 21,19; 21,23; 21,86; 21,89

Die Liste hat acht Werte, sodass sich für die Lageparameter folgendes ergibt:

$$z = \frac{x_4 + x_5}{2} = \frac{20,93 + 21,19}{2} = 21,06$$

$$x_{min} = 20,69$$

$$q_u = \frac{x_2 + x_3}{2} = \frac{20,71 + 20,84}{2} = 20,78$$

$$q_o = \frac{x_6 + x_7}{2} = \frac{21,86 + 20,93}{2} = 21,55$$

Nun können wir den Boxplot erstellen.





INPUT: Klasseneinteilung und Histogramme

Mathematik Statistik R 9



Boxplot

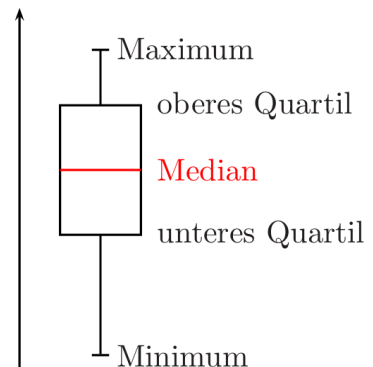
Der Boxplot ist die grafische Darstellung der wesentlichen Informationen einer Messreihe. Er besteht aus einer Box und sogenannten Whiskers (engl.: Barthaare, Schnurrhaare). Die Box reicht vom unteren bis zum oberen Quartil. Ihre Breite ist frei wählbar. Der Median wird mit einer waagrechten roten Strecke innerhalb der Box gekennzeichnet. Die Whiskers erstrecken sich vom unteren Quartil zum Minimum und vom oberen Quartil zu Maximum. Eine vertikale Achse gibt wie beim Koordinatensystem die Skalierung an.

Boxplot - einfach erklärt | Mathematik | Lehrerschmidt

Was ist ein Boxplot? Wie unterscheidet sich ein Boxplot von einem anderen Diagramm? Wie erstellt man ein Boxplot?



YouTube-Video



Um den Boxplot für unsere Kugelstoßer zu zeichnen, wählen wir für 0,1m auf dem Rasen 5mm, also ein Kästchen auf dem Papier. Außerdem nehmen wir auch nur den Ausschnitt, der uns wirklich interessiert. Wir fangen dementsprechend nicht bei 0m, sondern bei 20,6m an.

