

3.5 Beschreibende Statistik

Inhaltsverzeichnis

1 Die Darstellung von Daten

1.1 Das Kreisdiagramm

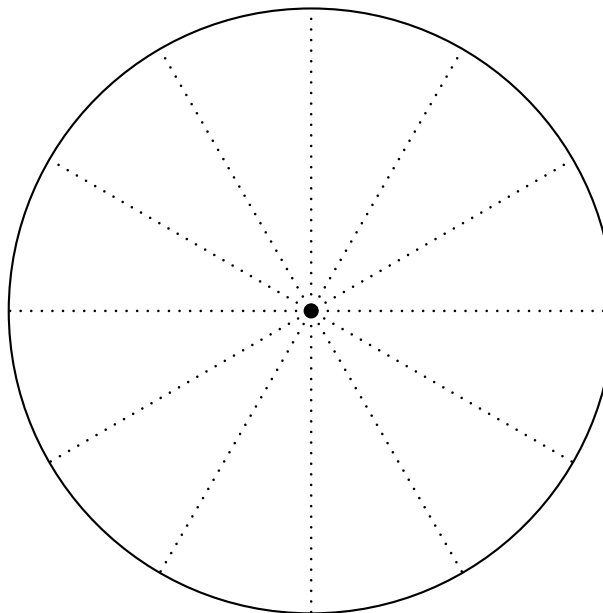
Wir beginnen mit einem Beispiel, welches uns durch das ganze Skript hindurch begleiten wird (Standardbeispiel):

Eine Klasse aus 12 Schülern hat bei einer Prüfung folgende Noten geschrieben:

2.9	3.7	4.9	3.3	4.1	5.1	5.4	4.1	4.2	5.7	5.7	3.7
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Fülle die untenstehende Tabelle aus und stelle Deine Ergebnisse nachher mit Hilfe des Kreises dar.

Note	1-2	2.1-3.0	3.1-4.0	4.1-5.0	5.1-6.0
Anzahl					



Die Liste mit den Noten (oder allgemein mit den Daten) heisst **Urliste**, der Kreis mit den eingezeichneten Daten ist ein **Kreisdiagramm**.

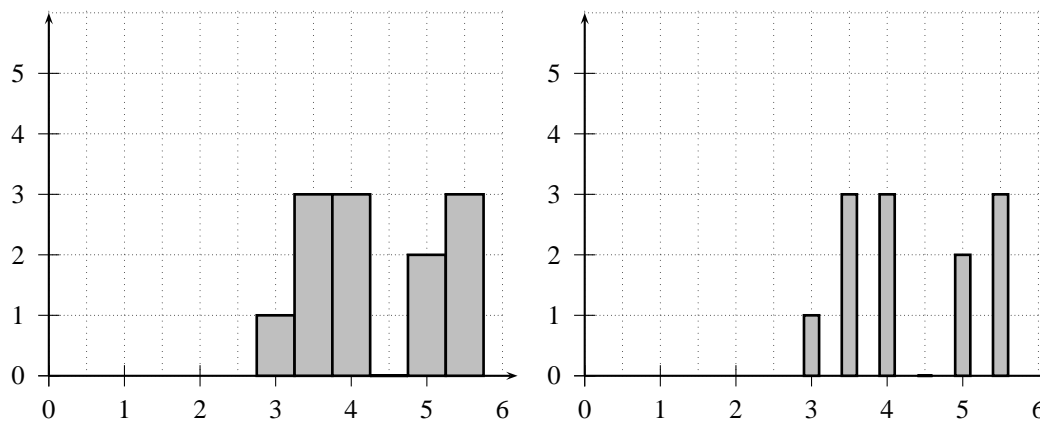
1.2 Histogramm

Es ist auch möglich, die Noten mit einem sogenannten Histogramm darzustellen. Dazu runden wir die erzielten Noten zuerst auf halbe Noten (so können wir die Daten einfacher darstellen). Wir erhalten folgende gerundete Noten:

3	3.5	5	3.5	4	5	5.5	4	4	5.5	5.5	3.5
---	-----	---	-----	---	---	-----	---	---	-----	-----	-----

Wir haben also 1 mal die Note 3, 3 mal die Note 3.5, 3 mal die Note 4, 2 mal die Note 5 und 3 mal die Note 5.5.

Das Histogramm sieht folgendermassen aus:



Die Unterteilung des Histogramms (die Breite der Balken) kann unterschiedlich erfolgen.

Übung

1. Ein Würfel wird zweimal geworfen. Wir betrachten dabei die Augensumme der beiden Zahlen. Fülle die Tabelle aus und stelle die Daten mit einem Histogramm dar.

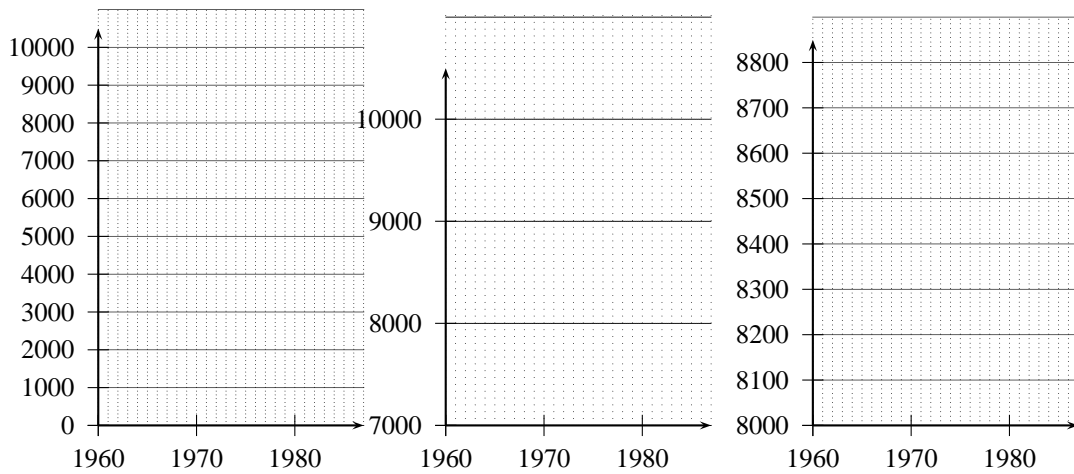
AS	2	3	4	5	6	7	8	9	10	11	12
Anz. Fälle	1	2									

1.3 Daten verfälschen

Einführungsbeispiel Zwischen 1962 und 1984 wurde der Zehnkampf-Weltrekord 13 Mal verbessert (Ab 1984 wurde eine neue Bewertungsskala eingeführt). Die Punktzahlen lauten folgendermassen:

1963	1967	1972	1975	1976	1982	1984
8206	8319	8454	8524	8618	8704	8798

Zeichne diese Werte in die untenstehenden drei Tabellen ein (jeweils von Neuem).



Frage: Was stellst Du fest, wenn Du die oberen drei Diagramme vergleichst ?

Übungen

2. Die Zahlen in der untenstehenden Tabelle geben jeweils den Benzin-Durchschnittsverbrauch auf 100km in Litern an.
 - Die mittlere Zeile: Durchschnittsbenzinverbrauch aller in Europa zugelassenen Dieselfahrzeuge.
 - Die unterste Zeile: Durchschnittsbenzinverbrauch eines bestimmten Dieselmotors.

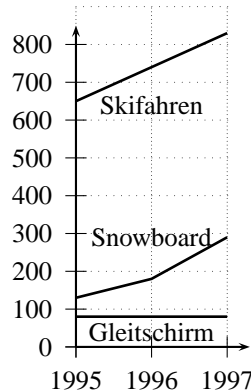
1984	1986	1988	1990	1992	1994	1996	1998	2000
6.6	6.3	6.2	6.2	6.3	6.3	6.2	6.2	6.3
6.5	5.9	5.8	5.7	5.7	5.7	5.6	5.7	5.7

Die Daten sollen so dargestellt werden, dass möglichst viel Werbung für das Modell der untersten Zeile gemacht wird.

1.4 Schlüsse ziehen aus Daten

Übung

3. Im Beobachter wurde einmal folgendes Diagramm zum Thema Heli-Rettungseinsätze gedruckt:



Was meinst Du zu der folgenden Aussage: „Von allen drei Sportarten ist Skifahren am gefährlichsten“.

4. Kommentiere die Aussagen in den folgenden zwei Artikeln:

2 Der Mittelwert

Mit dem **Mittelwert** \bar{X} wollen wir die Leistung der gesamten Klasse mit einer Zahl ausdrücken. Beispielsweise ist der Notendurchschnitt aus 4 Noten ein Mittelwert. Er berechnet sich mit folgendem Schema:

- Alle Noten aufsummieren.
- Die erhaltene Zahl durch die Anzahl Noten dividieren.

Frage: Wie gross ist der Mittelwert für die Daten des Einführungsbeispiels ?

[4.40]

Formal wird der Mittelwert folgendermassen notiert:

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Übungen

5. Ein fairer Würfel wird 10 Mal geworfen, die geworfenen Zahlen sind 1,2,2,3,3,2,5,3,1,6. Berechne den Mittelwert. [2.8]
6. Eine Klasse mit 20 Schülern hat eine Prüfung in Mathematik geschrieben und einen Notendurchschnitt von 4.20 erzielt. Berechne den Notendurchschnitt, wenn
 - a) zwei Schüler anstatt eine 3 nun plötzlich eine 5 erhalten, nachdem sie erfolgreich reklamiert haben. [4.40]
 - b) der Schüler mit der Note 1 nicht berücksichtigt wird. [4.37]
7. Warum muss man vorsichtig sein bei der Interpretation des Mittelwertes ?
8. Bei einer Querschnittsprüfung hat eine Klasse A einen Notendurchschnitt von 4.2, während die Klasse B einen Notendurchschnitt von 3.8 hat. Was meinst Du zu der Aussage: „Das Niveau der Klasse A ist höher als das Niveau der Klasse B“?

3 Der Median

Der Median ist „die Zahl, die in der Mitte steht“ bezogen auf die Grössenordnung. Er wird mit folgendem Vorgehen bestimmt:

- Daten der Grösse nach ordnen.
- – Ist die Anzahl Daten ungerade, dann wird einfach die Zahl in der Mitte genommen.
– Ist die Anzahl der Daten gerade, dann wird der Mittelwert der beiden in der Mitte liegenden Daten genommen.

Aufgabe: Berechne den Median für die Daten des Einführungsbeispiels. [4.15]

Übung

9. Überlege: Welchen Vorteil hat der Median gegenüber dem Mittelwert, welchen Nachteil ?
10. Berechne den Median der folgenden beiden Datenreihen. [4.4.5]
 - a) 1,3,3,3,4,5,8,9,9
 - b) 9,8,2,4,2,2,8,5

4 Die Varianz und die Standardabweichung

Wir versuchen nun mit einer Zahl auszudrücken, wie weit die Daten „auseinanderliegen“ (Streuung). Dies geschieht folgendermassen:

- Berechne den Mittelwert.
- Berechne für jede Note die Differenz zum Mittelwert.
- Quadriere alle Differenzen.
- Summiere alle Quadrate auf und dividiere durch $(n-1)$ (in unserem Fall die Anzahl Noten minus 1).

Die Zahl, die dabei herauskommt, nennen wir **Varianz**. Beachte, dass es auch andere Wege gibt, um die Streuung der Daten mit einer Zahl auszudrücken! Dass durch $(n-1)$ dividiert wird, hat technische Gründe, auf die wir nicht näher eingehen.

Aufgabe: Berechne die Varianz für die Daten des Einführungsbeispiels. [0.89]

Symbolisch wird die Varianz folgendermassen notiert:

$$\text{Var}(X) = \frac{(x_1 - \bar{X})^2 + (x_2 - \bar{X})^2 + \dots + (x_n - \bar{X})^2}{n - 1}$$

Eine weitere Grösse ist die Standardabweichung $\sigma(X)$, die der Vollständigkeit halber noch erwähnt wird. Sie ist einfach die Wurzel der Varianz.

$$\sigma(X) = \sqrt{\text{Var}(X)}$$

Frage: Wie gross ist die Standardabweichung des Einführungsbeispiels? [0.94]

Übung

11. Überlege: Warum werden die Differenzen quadriert?
12. In einem Bienenverein wird die Anzahl der Bienenschwärme gemeldet, die pro 10 Tage eingefangen worden sind. Berechne aus diesen Daten die Varianz und die Standardabweichung. [46.77, 6.84]

April	April	April	Mai	Mai	Mai	Juni	Juni	Juni	Juli
1.-10.	11.-20.	21.-30.	1.-10.	11.-20.	21.-31.	1.-10.	11.-20.	21.-30.	1.-10.
4	3	14	22	11	5	6	3	0	1

zusätzliche Antworten

7. Er ist empfindlich gegenüber Ausreißern, die den Mittelwert verfälschen können.
8. Man müsste die Hintergründe untersuchen, aufgrund des Notenschnittes kann noch kein Schluss gezogen werden. Vielleicht sind in der Klasse B zwei Schüler, die im Sommer sowieso gehen wollen und eine 1 geschrieben haben,... . Solche Schüler beeinflussen eben den Mittelwert.
9. Ausreisser spielen keine Rolle, es werden aber nicht alle Daten berücksichtigt.
11. Damit sich positive und negative Zahlen nicht gegenseitig aufheben. Man könnte zum Beispiel auch die Beträge nehmen und addieren (wir würden so einfach eine andere Schätzgrösse erhalten).