

***Statistische Grundbegriffe
für die Sekundarstufe 1***

Teil 1:

Statistische Erhebungen, Erfassen von Merkmalen
Absolute und relative Häufigkeiten
Mittelwerte

Datei Nr. 31001

Friedrich Buckel

Stand 14. Januar 2020

INTERNETBIBLIOTHEK FÜR SCHULMATHEMATIK

<https://mathe-cd.de>

Inhalt

1	Statistische Erhebungen	
1.1	Was will man in Umfragen erfahren?	3
1.2	Erfassen von Merkmalen – Grundbegriffe	5
1.3	Rechnen mit absoluten und relativen Häufigkeiten	7
	Stabilisierung der relativen Häufigkeiten	11
1.4	Klassierung von stetigen Merkmalen	13
1.5	Trainingsaufgaben 1 bis 10	15
1.6	Aufgaben mit mindestens und höchstens	17
	Kumulierte relative Häufigkeit	17
	Trainingsaufgaben 11, 12, 13	19
1.7	Schaubilder von relativen Häufigkeiten	21
	Trainingsaufgaben 14, 15	27
2	Kennzahlen von Häufigkeitsverteilungen	
2.1	Arithmetisches Mittel mit den absoluten Häufigkeiten berechnen.	28
	Verkürzte Mittelwertberechnung	29
2.2	Arithmetisches Mittel trickreich berechnen	32
	Trainingsaufgabe 16	33
2.3	Arithmetisches Mittel mit den relativen Häufigkeiten berechnen.	34
2.4	Arithmetisches Mittel bei einer Klasseneinteilung berechnen	37
2.5	Zentralwert – Median	38
3	Lösungen	41 - 49

Dieser Text ist sowohl Lesebuch wie auch Übungsbuch, so dass man die wichtigen Fähigkeiten dazu trainieren kann.

Jedes Schulbuch hat dazu andere Aufgaben. Doch der Hintergrund und die Berechnungsverfahren sind überall dieselben.

1 Was will man in Umfragen erfahren?

Es gibt die verschiedensten Gründe, sich mit Daten zu beschäftigen. Hier einige Beispiele dazu:

- (1) Politiker und auch ganz „normale“ Leute interessieren sich dafür, wie die Bürger ihres Landes mit der Politik ihrer Regierung zufrieden sind. Angestellte von Umfrageinstituten rufen dann zuhause an und bitten um 3 Minuten Mitarbeit. Sie stellen Fragen wie:
- Welche Partei würden Sie wählen, wenn heute Wahl wäre?
 - Sind Sie mit der Arbeit der Bundeskanzlerin zufrieden? (Noten 1 bis 6).
 - Ist es richtig, dass die Bevölkerung nun erst mit 67 in Rente gehen kann? Ja – nein.
 - Wie hoch sollte Ihrer Meinung nach das Kindergeld pro Kind sein: 100, 200 oder 300 €?
 - Wie viele Kinder haben Sie?
 - In welchem Bereich liegt Ihr Familieneinkommen monatlich:
 - Bei höchstens 1000 €,
 - Mehr als 1000 € aber höchstens 2000 €,
 - Mehr als 2000 € aber höchstens 4000 €,
 - Mehr als 4000 €.
 - Sollte es auf der Autobahn eine Geschwindigkeitsbegrenzung geben? (ja – nein – egal)
 - Wenn Sie in g) mit ja geantwortet haben, wo sollte Ihrer Meinung nach die Grenze sein?

So unterschiedlich alle diese Fragen sind. Sie haben alle etwas gemeinsam. Sie wollen eine Auskunft über ein **MERKMAL**.

Für jedes Merkmal gibt es Werte, die zur Auswahl stehen. Diese nennt man die **Merkmalsausprägungen**.

- In
- heißt das Merkmal **Partei** (die man wählen würde). Mögliche Ausprägungen dieses Merkmals sind derzeit: CDU, SPD, FDP, Grüne, Linke usw.
 - Hier ist das Merkmal die **Zufriedenheit** mit Frau Merkel. Als mögliche Werte sind Schulnoten angesetzt, also 1 bis 6.
 - Bei der Frage nach dem Merkmal **Rente** wird verlangt, dass man sich für ja oder nein entscheidet.
 - Für das Merkmal **Kindergeld** wurden 3 Beträge zur Auswahl gestellt.
 - Das Merkmal **Kinderzahl** bezieht sich auf die Zahlen 0, 1, 2, 3 usw.
 - Um das Merkmal **Einkommen** sinnvoll erfassen zu können, wurden sogenannte Einkommensklassen festgelegt. Dazu später mehr.
 - Für das Merkmal **Geschwindigkeitsbegrenzung** stehen drei „Werte“ zur Auswahl: ja, nein, egal.
 - Das Merkmal **Geschwindigkeitsobergrenze** ist ganz problematisch, weil hier jeder seine beliebige Zahl nennen darf.

Man ahnt schon, dass es sich hier um verschiedenartige Merkmale handelt. Manche sind durch Zahlen erfassbar, andere wieder nicht. Man hat dafür Fachbegriffe eingeführt:

MERKE: Bei einem **quantitativen Merkmal** sind die Werte, die dieses Merkmal annehmen kann, **Zahlen**.

Bei einem **qualitativen Merkmal** sind die Werte **Begriffe oder Namen**, aber **keine Zahlen**.

Zu den **qualitativen Merkmalen** gehören Begriffe wie Partei, Name, Haarfarbe, Staatsangehörigkeit, Religion, Familienstand aber auch Begriffe oder Fragen, auf die man mit ja oder nein oder gelegentlich auch mit einer Zahl antwortet, wie Notwendigkeit der Geschwindigkeitsbegrenzung, Zufriedenheit.

Bei den **quantitativen Merkmalen** gibt es zwei verschiedene Typen:

Es gibt **Merkmale, bei denen nur bestimmte Zahlenwerte auftreten können**, etwa bei der Kinderzahl, hier sind die Zahlen 0, 1, 2, 3 usw. möglich, nicht aber 2,6 Kinder oder 1,52 Kinder. Auch die Frage nach dem Kindergeld hat nur drei Werte zugelassen: 100 €, 200 € oder 300 €. Andere Merkmalwerte tauchen in dieser Umfrage nicht auf. Die Zufriedenheit mit der Arbeit der Bundeskanzlerin gehört auch dazu, denn man sollte nur zwischen den Zahlen 1, 2, 3, 4, 5, oder 6 auswählen. Solche Merkmale heißen **diskrete** Merkmale.

Merkmale, bei denen „eigentlich jeder“ Zwischenwert denkbar ist, nennt man **stetige** Merkmale. Dazu gehört alles, was man durch Messen bestimmt, etwa Temperatur, Geschwindigkeit, Längen, Benzin- oder Stromverbrauch, Lautstärke und auch der Verdienst.

Man muss diese Beispiele etwas kritischer betrachten.

Beginnen wir beim Verdienst. Weil die kleinste Geldeinheit bei uns Cent ist, gibt es doch keine beliebigen Zwischenwerte. Denn wenn Herr Dups 1245,13 € im Monat verdient und Frau Haps genau 1 Cent mehr, also 1245,14 €, dann gibt es keine Zwischenwerte zu diesen beiden. Also sind Geldbeträge „eigentlich“ keine stetigen Merkmale. Man rechnet sie dennoch dazu, weil man solche kleinen Geldbeträge wie halbe Cent usw. gar nicht beachtet.

Ähnlich ist es auch bei jeder Messung. Man kann nie beliebige fein messen. Wenn jemand die Längen von Bohnenkernen bestimmt, dann misst er vielleicht sehr genau 21,2 mm und bei der anderen Bohne 21,3 mm. Weil er mit seinem Messgerät nicht genauer messen kann, kann er keine Bohne mit der Länge 21,134 mm finden. Also fallen auch hier die Zwischenwerte weg.

Dasselbe gilt für Geschwindigkeitsmessungen. Ein Radarmessgerät hält vielleicht $87,5 \frac{\text{km}}{\text{h}}$ fest, es kann auch $87,4 \frac{\text{km}}{\text{h}}$ messen, nicht aber $87,48 \frac{\text{km}}{\text{h}}$!

Also kann man bei stetigen Merkmalen aus ganz praktischen Gründen nie beliebige Zwischenwerte feststellen. Dennoch heißen sie stetig, eben weil Zwischenwerte im Rahmen der Möglichkeiten existieren.

1.2 Erfassen von Merkmalen - Grundbegriffe

Jedes **Merkmal** hat verschiedene **Ausprägungen**. Die zugrunde liegende Menge der Merkmalsträger, bei denen also die Merkmale untersucht werden, nennt man die **Grundgesamtheit**.

Weil es zu teuer ist, **alle** Leute zu befragen, führt man immer eine **Stichprobe** durch. Damit meint man, dass eine bestimmte Auswahl der Bevölkerung befragt wird. Dies gilt auch bei anderen Erhebungen, etwa beim Messen einer Größe (Länge, Temperatur usw.).

Weil man Ergebnisse haben möchte, die **aussagekräftig** sind, sollte man stets eine **repräsentative Umfrage** durchführen. Repräsentativ heißt, dass in der Stichprobe auch alle Ausprägungen etwa ihrem Anteil nach vertreten sein sollten. Wenn man etwa nach politischen Ansichten fragt, sollte man nicht nur Leute befragen, die Unternehmer sind, oder nur Leute, die arbeitslos sind. Sonst bekommt man ein einseitiges Umfrageergebnis, das nicht für die ganze Bevölkerung repräsentativ ist.

Übersicht: Merkmale		
qualitativ	quantitativ	
(nicht messbar, etwa sortiert nach Namen)	(zählbar - messbar)	
z.B. Partei, Name, Haarfarbe, Staatszugehörigkeit, Schultyp, Religion, Familienstand.	diskretes M. z. B. Kinderzahl, Klassenarbeitsnote, Schülerzahlen	stetiges M. z. B. Längen, Geschwindigkeiten, Benzinverbrauch, Stromverbrauch, Lautstärke, Verdienst (!)

Die Erfassung von Merkmalen geschieht in der Regel mit einer Art **Strichliste**, mit der man dann die Anzahlen abzählen kann.

Beispiel 1

In der Goethe-Schule wurden die Konfessionen der Kinder statistisch erfasst. Von den 428 Kindern der Goethe-Schule erhielt man folgendes Ergebnis:

Evang.	198
Röm.-kath.	45
Moslem	11
Sonstige Konf.	8
Ohne Konf.	166

Die Zahlen, die zu den Religionen gehören, sind die **absoluten Häufigkeiten**.

Ihre Summe ist der **Umfang n der Stichprobe**: $n = 198 + 45 + 11 + 8 + 166 = 428$!

Beispiel 2

Zwei Schüler unterhalten sich über dieses Ergebnis. Klaus sagt: In meiner Schule (der Goethe-Schule) sind 11 Moslems, sein Freund Mohammed prahlt: „In meiner Schule (der Schiller-Schule) sind es mehr, nämlich 14.

Oberflächlich betrachtet hat Mohammed Recht. Wir schauen uns jetzt die Statistik an, die für seine Schule erstellt worden ist. Die Tabelle gibt sie wieder:

Evang.	255
Röm.-kath.	108
Moslem	14
Sonstige Konf.	20
Ohne Konf.	211

Diese Umfrage hat den Umfang $n = 255 + 108 + 14 + 20 + 211 = 608$.

Einerseits ist also der Anteil der Moslems höher, aber andererseits hat diese Schule auch mehr Schüler.

Wir sollten daher ausrechnen, wie hoch der Bruchteil der Moslems an beiden Schulen ist.

Anteil der Moslems an der Goethe-Schule: $h_1 = \frac{11}{422} \approx 0,026$ (gerundet)

Anteil der Moslems an der Schiller-Schule: $h_2 = \frac{14}{608} \approx 0,023$ (gerundet)

Wer schon gelernt hat, wie man diese Bruchteile in Prozent umrechnet, kann sagen:

An der Goethe-Schule sind 2,6 % der Schüler Moslems,
an der Schiller-Schule sind es 2,3 %..

Zum Vergleichen von Umfragen braucht man die relativen Häufigkeiten, denn sie zeigen, dass der Anteil der Moslems an der Goethe-Schule trotz kleinerer absoluter Häufigkeit größer ist!

1.3 Rechnen mit absoluten und relativen Häufigkeiten

Beispiel 3

5 Schüler der Klasse 7b testen eine Cent-Münze. Sie wollen herausfinden, ob diese gleichmäßig fällt. Sie beschließen, dass jeder seine Münze 1000-mal wirft.

Die Zahl $n = 1000$ ist jetzt der **Umfang der Stichprobe**, die somit aus 1000 Würfen besteht. In einer Tabelle notieren sie, wie oft die Zahl 1 oben ist bzw. das Wappen.

Schüler	Absolute Häufigkeiten		Relative Häufigkeiten	
	Zahl	Wappen	Zahl	Wappen
Klaus	486	514	0,486	0,514
Max	527	473	0,527	0,473
Lisa	561	439	0,561	0,439
Jussuf	445	555	0,445	0,555
Sarah	732	268	0,732	0,268

Zur Erinnerung:

Die absoluten Häufigkeiten geben an, wie oft tatsächlich die Zahl 1 bzw. das Wappen nach oben gezeigt hat.

Die relativen Häufigkeiten sind die Bruchteile, die man bekommt, wenn man die absoluten Häufigkeiten durch den Umfang der Stichprobe teilt.

Formel für die relativen Häufigkeiten:

Bezeichnet man die absoluten Häufigkeiten bei Klaus mit $n_1 = 486$ und $n_2 = 514$, dann berechnet man die zugehörigen relativen Häufigkeiten auf diese Weise:

$$h_1 = \frac{n_1}{n} = \frac{486}{1000} = 0,486 \quad \text{bzw.} \quad h_2 = \frac{n_2}{n} = \frac{514}{1000} = 0,514$$

Diese relativen Häufigkeiten kann man auch in Prozent angeben.

Da „Prozent“ ja Hundertstel heißt, muss man einfach in der Dezimalzahl, welche die relative Häufigkeit angibt, das Komma um zwei Stellen nach rechts verschieben. Dann erhält man die prozentualen Häufigkeiten zu

Aus $h_1 = 0,486$ wird $p_1 = 48,6\%$ und aus $h_2 = 0,514$ wird $p_2 = 51,4\%$

**Prozentuale Häufigkeiten sind also relative Häufigkeiten
in anderer Darstellung!**

Kontrolliere jetzt bitte, ob die Berechnung der relativen Häufigkeiten der anderen Schüler richtig ist.

Es gibt drei Kontrollmöglichkeiten für solche Tabellen:

1. Die Summe der absoluten Häufigkeiten muss gleich groß wie der Umfang der Stichprobe sein.
2. Die Summe der relativen Häufigkeiten muss 1 sein.
3. Die Summe der prozentualen Häufigkeiten muss 100% sein.

Bemerkungen dazu:

Der 1. Punkt ist klar, denn alle Möglichkeiten zusammen ergeben die Anzahl der Würfe.

Dass die Summe der relativen Häufigkeiten immer 1 sein muss, das zeigt eine einfache Berechnung, die wir für den Schüler Klaus durchführen wollen:

$$h_1 + h_2 = \frac{n_1}{n} + \frac{n_2}{n} = \frac{486}{1000} + \frac{514}{1000} = \frac{486 + 514}{1000} = \frac{1000}{1000} = 1$$

Oder ganz allgemein :

$$h_1 + h_2 + \dots = \frac{n_1}{n} + \frac{n_2}{n} + \dots = \frac{n_1 + n_2 + \dots}{n} = \frac{n}{n} = 1$$

Schließlich die prozentualen Häufigkeiten:

Sie sind ja nichts anderes als relative Häufigkeiten in Prozentangabe.

Deren Summe ist 1, und das sind $\frac{100}{100} = 100\%$!

Wir wollen uns nachträglich noch das [Ergebnis des Münz-Tests](#) anschauen.

Niemand kann erwarten, dass Wappen und Zahl genau gleich oft vorkommen. Aber die Natur wirkt hier ausgleichend. Je länger man eine Münze wirft, desto näher kommt die relative Häufigkeit dem Wert 0,5 bzw. 50%. Dies gilt natürlich nur, wenn sie ganz gleichmäßig gearbeitet ist.

Bei den Münzen der ersten vier Schüler scheint es so zu sein, dass es sich um so genannte **ideale Münzen** handelt, denn ihre relativen Häufigkeiten liegen nach 1000 Würfeln in der Nähe von 50 %

Aber die Münze von Sarah scheint leicht verbogen zu sein, so dass sie nach dem Werfen eine bevorzugte Lage einnimmt: Mit der Häufigkeit von 73,2 % lag nämlich die Zahl 1 oben!

Übungsaufgabe 1

Die Schüler der Klassen 5 bis 10 einer Internatsschule werden in einer Tabelle erfasst. Um die Zahlen zu verstehen, muss man wissen, dass in den unteren Klassen noch nicht so viele Schüler im Internat wohnen, wie in den oberen Klassen.

Jahrgangsstufe	Schülerzahlen	Relative Häufigkeiten
5	16	
6	24	
7	28	
8	29	
9	35	
10	48	

Die Schulleitung will nun wissen, wie die prozentualen Häufigkeiten der Schüler der einzelnen Klassenstufen sind.

Berechne diese Werte!

Übungsaufgabe 2

Ein Mathelehrer gibt in seiner Klasse einen Test zurück und hat sich folgende Notenliste notiert:

Note	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6
n_i	2	1	4	8	5	1	4	2	2	0	1
h_i											
p_i											

Berechne die relativen Häufigkeiten h_1, h_2 usw. (man schreibt dafür allgemein h_i) und die prozentualen Häufigkeiten p_i .

Übungsaufgabe 3

Im Biologieunterricht werden Bohnenkerne der Länge nach sortiert. Die Tabelle gibt das Ergebnis wieder, wobei die Länge in mm abgegeben ist.

Länge	15	16	17	18	19	20	21	22
Anzahl	10	12	18	28	20	17	13	2
h_i								
p_i								

Berechne auch die relativen und die prozentualen Häufigkeiten.