Mittelwerte und ihr Vergleich

Ein t-Test zum Vergleich von zwei arithmetischen Mittelwerten

von Norbert Kessel

Wirkt eine Behandlung oder erscheint es nur so?

Bei vielen naturwissenschaftlichen Versuchen muss die Wirkung einer Behandlung beurteilt werden. Im klassischen Fall in der Land- und Forstwirtschaft beispielsweise die Wirkung einer Düngung oder der Einfluss des Klimas auf das Wachstum von Pflanzen.

Für das folgende Beispiel wurden die Sprosslängen von jungen Fichten gemessen. Es handelt sich insgesamt um 18 Pflanzen, die nach der Anzucht aufgeteilt wurden: Neun Pflanzen wurden in Norddeutschland, die anderen neun in Süddeutschland in ein Beet gepflanzt und deren Sprosslängen nach Ablauf von drei Jahren gemessen.

Hier sind die Sprosslängen aufgeführt:

Norddeutschland	Süddeutschland
20	19
22	21
27	24
15	28
17	26
19	19
17	25
21	27
23	23

Die Daten stammen aus einem Biometrie-Praktikum der Universität Freiburg (Institut für Forstliche Biometrie, mit freundlicher Genehmigung von Prof. Dr. D. Pelz).

Die Fragestellung

Entwickeln sich die Pflanzen im Norden anders als im Süden? Hierzu werden als erstes die beiden arithmetischen Mittelwerte berechnt und die miteinander verglichen. Dazu kann man einen Taschenrechner verwenden:

	Norddeutschland	Süddeutschland
arithmetischer Mittelwert	20,1	23,6

Aufgrund dieser beiden Mittelwerte können bereits Vermutungen angestellt werden: die Pflanzen in Süddeutschland sind größer als die aus Norddeutschland.

Allerdings wäre diese Aussage nicht sehr fundiert, denn wir wissen noch nichts über die Streuung der Werte. Ist die Streuung zu groß, dann sind die Unterschiede nicht aussagekräftig, die Unterschiede in den Sproßlängen wären zufällig entstanden. Um zu sicheren Aussagen zu kommen, nutzt man seit mehr als 100 Jahren den t-Test, der für den Vergleich von arithmetischen Mittelwerten entwickelt wurde – interessanterweise für das Bier der Guinness-Brauerei in Dublin. Dort musste die Qualität der Gerste beurteilt werden, die in großen Mengen eingekauft wurde. Einen t-Test kann man zwar auch mit einem Taschenrechner berechnen, aber eleganter geht es mit der Software ‚R'.

R ist eine Statistik-Software, die an Universitäten entwickelt wurde. Man kann sie gratis aus dem Internet herunterladen, sie ist für alle gängigen Betriebssysteme erhältlich. Nähere Informationen finden sich bei www.r-project.org.

Die Erfassung der Daten

Damit die Daten ausgewertet werden können, müssen sie zuerst in R erfasst werden. Dazu gibt man sie folgendermaßen ein:

```
nord<-c(20,22,27,15,17,19,17,21,23)
sued<-c(19,21,24,28,26,19,25,27,23)
```

Zur Erläuterung: Man legt zwei Variablennamen fest (die Namen sind frei wählbar, hier: ‚nord' und ‚sued') und speichert mit der Funktion ‚c()' (=combine, verbinden) die in Klammern angegebenen Daten darin ab.

Sind viele Werte zu verarbeiten, speichert man diese besser in XLS- oder TXT-/CSV-Dateien). Zum Lesen dieser Dateien gibt es andere Funktionen.

Zur Kontrolle kann man die Daten anzeigen lassen (den Variablennamen eingeben und die Enter-Taste drücken). Mit der Funktion mean() kann man den arithmetischen Mittelwert anzeigen lassen:

```
mean(nord)
mean(sued)
```

Alternativ kann man eine kleine zusammenfassende Statistik anzeigen lassen, die Informationen zu den Datenreihen liefert:

```
summary(nord)
summary(sued)
```

Die Voraussetzungen für den t-Test

Für den t-Test werden in der Literatur sechs Voraussetzungen genannt, eine davon (normal verteilte Werte) ist so wichtig, dass sie hier geprüft werden soll. Hierzu wird ein Test mit dem Namen ‚Shapiro-Wilk' verwendet. Er liefert einen sogenannten ‚p-Wert' (‚p' steht für ‚probability', Wahrscheinlichkeit).

Der Aufruf des Tests ist denkbar einfach:

```
shapiro.test(nord)
shapiro.test(sued)
```

Beide p-Werte sind hier größer als die übliche Irrtumswahrscheinlichkeit von 0,05. Deshalb kann man daraus folgern, dass es sich um normal verteilte Werte handelt.

Der t-Test zur Beantwortung der Frage

Der eigentliche t-Test wird mit der folgenden Anweisung gerechnet:

```
t.test(nord, sued)
```

Als Ergebnis erhält man eine Reihe von Ergebnissen; die wichtigste Information steht bei ‚p-value' (fett hervorgehoben):

```
Welch Two Sample t-test
data:  nord and sued
t = -2.0924, df = 15.855, p-value = 0.05285
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.93676732  0.04787843
sample estimates:
mean of x mean of y
 20.11111  23.55556
```

Dort findet sich der Wert 0.05285. Der liegt über der üblichen Irrtumswahrscheinlichkeit von 0,05 und bedeutet, dass die Unterschiede zwischen den Mittelwerten statistisch nicht (!) signifikant sind.

Die Antwort auf die Frage

Die Unterschiede zwischen den beiden arithmetischen Mittelwerten der zwei Datenreihen sind mithin zufälliger Natur, die Pflanzen in Süddeutschland entwickelten sich nicht besser als die in Norddeutschland. Der erste Eindruck beim Vergleich der Mittelwerte war also falsch.

Der Zeitbedarf für die Dateneingabe und -auswertung lag bei unter zwei Minuten, das ist bemerkenswert.

Grundgesamtheit, Stichprobe

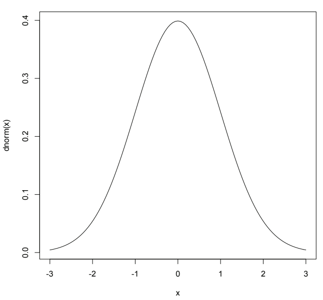
Üblicherweise interessiert man sich für die Eigenschaften einer Grundgesamtheit. Da man aber nicht alle Objekte dieser Grundgesamtheit untersuchen kann (oder möchte), nimmt man eine Stichprobe daraus. Die mit der Stichprobe gewonnenen Erkenntnisse werden dann auf die Grundgesamtheit übertragen.
Dabei ist die Frage nach dem Umfang einer Stichprobe von großer Bedeutung. Eine feste Zahl gibt es nicht, aber es leuchtet ein, dass Schlussfolgerungen, die aus zu kleine Stichproben gewonnen werden, nicht sinnvollerweise verallgemeinert werden dürfen.

Mittelwerte

Es gibt verschiedene Arten von Mittelwerten, der bekannteste ist sicherlich der arithmetische Mittelwert, der fast immer dann berechnet wird, wenn es um Längen, Durchmesser oder Gewichte geht.

Normalverteilung

Viele Messwerte, die man in der Natur erhebt (die Höhe von Bäumen, das Gewicht von Kartoffeln) folgen der sogenannten Normalverteilung: das heißt, es gibt eine Häufung in der Mitte und zu den Extremwerten hin nimmt die Anzahl an Messwerten ab. Früher hat man das untersucht, indem man Klassen festgelegt und Strichlisten angefertigt hat. Grafisch dargestellt, entstand dann eine Häufigkeitskurve, die mehr oder weniger stark der Normalverteilung ähnelte.
Liegt Normalverteilung vor (hier im Aufsatz wurde das mit einem Test untersucht), dann steht einem ein großes Instrumentarium zur Auswertung zur Verfügung: t-Test und die ebenfalls mit arithmetischen Mittelwerten arbeitende Varianzanalyse sind sicher die wichtigsten Vertreter. Liegt keine Normalverteilung vor, dann kann man – eventuell nach einer Transformation der Werte – mit anderen Tests die Daten auswerten.
Mit dem Begriff Normalverteilung eng verbunden ist der berühmte Mathematiker Gauss, der diesen Sachverhalt beschrieben hat. Ihm zu Ehren war auf dem alten 10-DM-Schein die Kurve der Normalverteilung und er selbst abgebildet.
Für jemanden, der schon mal mit Kurvenlinealen hantiert hat, wird es verblüffend sein, wie mit nur einer Anweisung in der Sprache ‚R' eine solche Kurve angefertigt werden kann. Die Anweisung `curve(dnorm, -3, 3)` erzeugt eine solche Kurve.



Die Streuung von Messwerten: Standardabweichung und Varianz

Ein Gefühl für die Streuung hat jeder von uns, aber wie wird sie berechnet? Um die Streuung von Messwerten und deren Berechnung zu veranschaulichen, sollen hier zwei Datenreihen verwendet werden, es könnte sich beispielsweise um das Gewicht von Kartoffeln handeln.

erste Datenreihe: 20g, 20g, 20g
zweite Datenreihe: 10g, 20g, 30 g

In der ersten Datenreihe sind die Kartoffeln alle gleich schwer. offensichtlich ist die Streuung 0.
Für die zweite Datenreihe muss zuerst der arithmetische Mittelwert bestimmt werden (20 g). Anschließend muss die Differenz des Gewichts jeder Kartoffel zum Mittelwert berechnet werden. Da es hierbei zu negativen Vorzeichen kommt (dritte Spalte), werdem die Werte quadriert (vierte Spalte). Die folgende Tabelle fasst das zusammen.

Messwert	arithm. Mittelwert	Differenz zum Mittelwert	Quadrate der Differenzen
10	20	-10	100
20	20	0	0
30	20	+10	100
			200

Die Summe aller Werte in der vierten Spalte nennt man die *Summe der Abweichungsquadrate*, das ist ein wichtiger Begriff in der Statistik.
Diese Summe wird nun dividiert durch die Anzahl der Messwerte abzüglich 1, mathematisch ausgedrückt sieht das folgendermaßen aus:

$$\frac{\textit{Summe der Abweichungsquadrate}}{n-1} = \frac{200}{2} = 100$$

Aus dem Wert 100 muss nun noch die Quadratwurzel gezogen werden, es ergibt sich eine Abweichung von 10 g. Somit können wir die eben gestellte Frage (wie groß ist die Streuung in der ersten Datenreihe) beantworten: die Standardabweichung ist 10 g, anders ausgedrückt: um 10 g weichen die Kartoffeln vom Mittelwert ab.
Die allgemeine Formel zur Berechnung der Standardabweichung (die mit ‚s' bezeichnet wird) sieht folgendermaßen aus:

$$s = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n-1}}$$

Berechnet man nicht die Quadratwurzel, dann erhält man die Varianz, diese wird mit ‚s^2' bezeichnet.