Bedeutung einiger häufig gebrauchter statistischer Kennzahlen und Begriffe und ihre Interpretation

Fränzi Korner

1. Kennzahlen für die Beschreibung von Daten

n (Anzahl, engl. number)

Anzahl der unabhängigen Beobachtungen, die Daten geliefert haben. Grundsätzlich gilt: Je mehr Beobachtungen zu einer Aussage beigetragen haben, desto breiter abgestützt ist sie. Unabhängige Beobachtungen werden an Einheiten gemacht, von denen jede das gemessene Merkmal als seine Eigenschaft zeigt. Es gibt keine offensichtlichen Gründe für Ähnlichkeiten zwischen den Einheiten. Wenn zum Beispiel die Gesangsaktivität von Tannenmeisen 10-mal am gleichen Individuum gemessen wird, dann sind die einzelnen Beobachtungen nicht unabhängig, weil sie von demselben Individuum stammen. Misst man die Gesangsaktivität aber an 10 verschiedenen (nicht verwandten) Individuen, so dürfen die Beobachtungen als unabhängig betrachtet werden.

Achtung! Je grösser das n, desto eher wird ein gegebener Effekt signifikant. Dies kann dazu führen, dass biologisch irrelevante Effekte (sehr kleine Unterschiede, sehr schwache Trends) statistisch signifikant werden. Deshalb muss unbedingt die Effektgrösse (Grösse des Unterschiedes, Steigung des Trends) betrachtet und auf seine biologische Relevanz hin diskutiert werden.

Mittelwert (m, engl. mean)

Der Mittelwert zeigt den Schwerpunkt von Messwerten an. Er beträgt

$$m = \frac{\sum_{i}^{n} x_i}{n} \qquad (1)$$

Dabei steht x_i für die i-te Beobachtung und n für die Anzahl Beobachtungen. Der Mittelwert trifft nur dann den Schwerpunkt der Messwerte, wenn die einzelnen Werte symmetrisch um den Mittelwert verteilt sind. Bei asymmetrischen Verteilungen oder wenn Extremwerte vorhanden sind, wird empfohlen, anstatt den Mittelwert den Median zu verwenden.

Median oder Zentralwert (med, engl. median)

Der Median ist der mittlere Wert einer Stichprobe von Messwerten. Man erhält den Median, indem man die Stichprobe in aufsteigender Reihenfolge sortiert und in der Mitte teilt. Der Wert der Beobachtung, die zwischen den beiden Hälften liegt, wird als Median bezeichnet. Bei gerader Anzahl Beobachtungen liegen zwei Werte in der Mitte. Dann liefert der Durchschnitt dieser zwei Werte den Median. Der Median beschreibt den Schwerpunkt von asymmetrisch verteilten Beobachtungen besser als der Mittelwert und ist gegenüber Extremwerten, so genannten Ausreissern, robust.

Varianz (var, engl. variance)

Ein Mass für die Streuung von Beobachtungen um ihren Mittelwert. Die Varianz ist die Summe aller Abweichungen jeder Beobachtung vom Mittelwert dividiert durch die um 1 verringerte Stichprobengrösse:

$$\text{var} = \frac{\sum_{i}^{n} x_i - m}{n-1} \qquad (2)$$

Dabei steht x_i für die i-te Beobachtung, m für den Mittelwert (1) und n für die Anzahl Beobachtungen. Die Varianz ist schwierig zu interpretieren. Sie wird hauptsächlich als Grundlage für die Berechnung weiterer statistischer Kennzahlen verwendet, z.B. Standardabweichung oder Standardfehler. Grundsätzlich gilt, je grösser die Varianz desto stärker streuen die Beobachtungen um den Mittelwert.

Standardabweichung (sd, engl. standard deviation)

Die Standardabweichung ist ebenfalls ein Mass für die Streuung von Beobachtungen. Sie wird als die Wurzel aus der Varianz (2) berechnet

$$\mathrm{sd} = \sqrt{\mathrm{var}} \qquad (3)$$

Die Standardabweichung lässt sich einfach interpretieren, falls die Beobachtungen aus einer Normalverteilung (s. unten) stammen: Im Intervall Mittelwert minus Standardabweichung und Mittelwert plus Standardabweichung ([m – sd, m + sd]) liegen ca. 68 % der Beobachtungen, im Intervall Mittelwert minus zweimal die Standardabweichung und Mittelwert plus zweimal die Standardabweichung ([m – 2 × sd, m + 2 × sd]) liegen ca. 95 % der Beobachtungen.

Standardfehler (se, engl. standard error)

Mass für die Unsicherheit einer berechneten Kennzahl (z.B. Mittelwert). Kennzahlen, die aufgrund von Stichproben berechnet werden, sind Schätzungen für den «wahren» Wert, den man erhalten hätte, wenn man sämtliche Individuen oder Objekte aus derselben Grundgesamtheit gemessen hätte. Der Standardfehler wird berechnet, indem man die Standardabweichung (3) durch die Wurzel aus der Stichprobengrösse teilt:

$$\mathrm{se} = \frac{sd}{\sqrt{n}} \qquad (4)$$

Der Standardfehler zeigt an, wie stark die Mittelwerte von zufälligen anderen Stichproben um den erhobenen Mittelwert streuen. Er lässt sich ebenfalls einfach interpretieren: Im Intervall Mittelwert minus Standardfehler und Mittelwert plus Standardfehler ([m – se, m + se]) liegen ca. 68 %, im Intervall Mittelwert minus zweimal Standardfehler und Mittelwert plus zweimal Standardfehler ([m – 2 × se, m + 2 × se]) liegen ca. 95 % der Mittelwerte.

Der Standardfehler wird angegeben, wenn man die Unsicherheit des Mittelwertes oder einer anderen Kennzahl angeben möchte. Mit der Standardabweichung zeigt man, wie stark die Beobachtungen um den Mittelwert streuen.

Normalverteilung (engl. normal distribution)

Beobachtungen (Messwerte) verteilen sich um ihren Mittelwert. Die meisten liegen nahe am Mittelwert. Die Dichte der Beobachtungen nimmt mit zunehmender Distanz zum Mittelwert ab. Nimmt die Dichte der Beobachtungen nach beiden Seiten hin symmetrisch entsprechend einer Glockenkurve ab (Abb. 1), so spricht man von einer Normalverteilung. Für viele Berechnungen, zum Beispiel Varianz, Standardabweichung und Standardfehler, wird vorausgesetzt, dass die Beobachtungen normal verteilt sind.

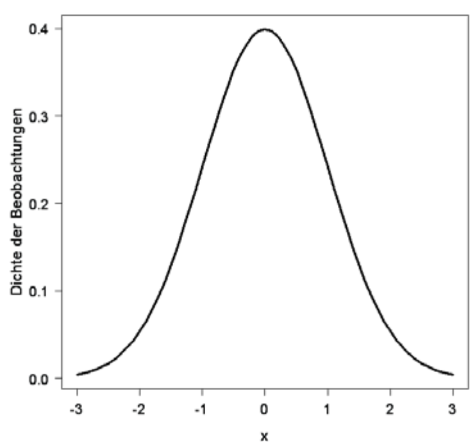

Abb. 1. Normalverteilung mit dem Mittelwert m = 0 und der Standardabweichung sd = 1.

2. Kennzahlen aus der schliessenden Statistik

FG (Freiheitsgrade) oder df (engl. degrees of freedom)

Anzahl unabhängiger Beobachtungen in einem Datensatz minus Anzahl für eine statistische Aussage berechneter Parameter.

$$FG = n - k \qquad (5)$$

k steht für Anzahl berechneter Parameter, zum Beispiel k = 1 für die Berechnung des Mittelwertes, wenn getestet werden sollte, ob sich der Mittelwert von einer bestimmten Zahl unterscheidet. FG ist ein statistisch-technisches Mass für die Stichprobengrösse, das für die Berechnung des p-Wertes (s. unten) gebraucht wird.

p (Wahrscheinlichkeit, engl. probability)

Eine der am häufigsten gebrauchten statistischen Kennzahlen ist der p-Wert. Er ist die Wahrscheinlichkeit einer beobachteten Datenlage unter der Nullhypothese. Die Nullhypothese ist die Annahme, dass in den Daten ein bestimmter Effekt nicht vorhanden ist (kein Unterschied zwischen Gruppen, kein Trend oder kein Zusammenhang zwischen Variablen). Wenn p genügend klein ist, dann darf angenommen werden, dass ein Effekt (Unterschied, Trend oder Zusammenhang) vorhanden ist. Man spricht im Allgemeinen von einem signifikanten Effekt, wenn $p < 0{,}05$ (oder $p < 5\,\%$) ist. Dieses Signifikanzniveau von 0,05 ist willkürlich gewählt. Gerade so gut könnte man 0,001 oder 0,1 wählen. Allerdings wurde der Grenzwert von 0,05 in den empirischen Wissenschaften zur Konvention. Meist wird die Stärke der Signifikanz mit Sternen markiert, wobei

*	bedeutet	$0{,}01 < p \leq 0{,}05$	knapp signifikant
**	bedeutet	$0{,}001 < p \leq 0{,}01$	signifikant
***	bedeutet	$p \leq 0{,}001$	hoch signifikant

Liegt p zwischen 0,05 und 0,1, wird im Allgemeinen von einem Trend gesprochen.

Signifikant heisst also, dass ein in den Daten sichtbarer Effekt höchstens in 5 % der Fälle durch Zufall hätte zustande kommen können.

Achtung! Der p-Wert zeigt in 5 % der Fälle einen signifikanten Effekt an, der nur durch Zufall entstanden ist. Das betrifft jeden zwanzigsten statistischen Test. Wichtig ist es deshalb, die Resultate kritisch zu betrachten und mit anderen, ähnlichen Studien zu vergleichen. Man bedenke dabei, dass nicht-signifikante Resultate leider meist nicht publiziert werden.

t, F, Chi^2 und U-Werte

t, F, Chi^2 und U-Werte sind standardisierte Masse, die für die Berechnung des p-Wertes gebraucht werden.

Man kann sie in vielen Fällen folgendermassen interpretieren:

t (Teststatistik aus dem Student's t-Test): Je grösser t, desto stärker ist die Evidenz für die Existenz eines Unterschieds zwischen zwei Mittelwerten.

F (Teststatistik aus der Varianzanalyse): Je grösser F, desto stärker ist die Evidenz für die Existenz eines Einflusses einer erklärenden Variablen auf die Zielvariable.

Chi^2 (Teststatistik aus dem Chi^2-Test): Je grösser Chi^2, desto stärker ist die Evidenz für die Existenz eines Unterschieds zwischen zwei Häufigkeitsverteilungen. Chi^2 wird oft mit dem griechischen Buchstaben χ^2 geschrieben.

U (Teststatistik aus dem Wilcoxon Rangsummentest): Je kleiner U, desto eher unterscheiden sich die Messwerte zwischen zwei Gruppen.

3. Mathematische Modelle

Modell

Mathematische Beschreibung eines Sachverhaltes meist in der Form y ~ f(x) («y ist eine Funktion von x»), wobei y die Zielvariable und x die erklärende Variable ist. Es können eine oder mehrere erklärende Variablen im Modell enthalten sein. Modelle können an in der Natur erhobene Daten angepasst, und es kann das am besten passende Modell gesucht werden.

Parameter

Meist ein fester, aber unbekannter Wert, der Bestandteil eines Modells ist, zum Beispiel der Achsenabschnitt a oder die Steigung b im Modell $y \sim a + b \times x$. Die Parameter werden mit den Daten geschätzt.

Likelihood (engl. Wahrscheinlichkeit)

Wahrscheinlichkeit für beobachtete Daten unter der Annahme, die Daten stammten aus einem gegebenen Modell (s. oben). Die Likelihood wird hauptsächlich für die Schätzung der Parameter eines Modells gebraucht. Dabei wird der zu schätzende Parameter variiert und jener Wert ausgewählt, der die grösste Wahrscheinlichkeit für das Auftreten der beobachteten Daten mit sich bringt (Maximum-Likelihood-Schätzung). Die Formel für die Berechnung der Likelihood sieht je nach Modell unterschiedlich aus. Die Berechnung der Maximalstelle erfolgt entweder analytisch über das Null-Setzen der ersten Ableitung oder durch Iterationsverfahren. Meist werden die Wahrscheinlichkeiten vorgängig logarithmiert, um die Rechnung zu vereinfachen. Interessierte Leser finden in Stahel (2000) eine ausführliche Beschreibung der Maximum-Likelihood-Methode.

AIC (Akaike information criterion)

Mit dem AIC-Wert werden verschiedene Modelle verglichen, die denselben Datensatz beschreiben. Er kann auch negativ sein. Je kleiner der AIC-Wert, desto besser ist ein Modell.

So wird der AIC-Wert berechnet:

$$\mathrm{AIC} = -2 * \log(Likelihood) + 2 * k \qquad (6)$$

wobei k die Anzahl Parameter im Modell ist. Er baut also auf der Likelihood (s. oben) auf. Da die Likelihood mit negativem Vorzeichen in die Berechnung des AIC's eingeht, haben die wahrscheinlicheren Modelle die kleineren AIC-Werte. Je mehr Parameter ein Modell enthält, desto genauer können die Daten vorausgesagt werden. Ein gutes Modell soll mit möglichst wenigen Parametern die Daten möglichst gut voraussagen. Deshalb wird im AIC-Kriterium für jeden Parameter noch ein Betrag dazugezählt.

4. Literatur

Einführungen in die Statistik (deutsch)

KÖHLER, W., G. SCHACHTEL & P. VOLESKE (1996): Biostatistik. Springer, Berlin.
Kurze und bündige Einführung in die statistische Datenanalyse. Wenig theoretischer Hintergrund, dafür einfache Kochbuchrezepte für die Durchführung von statistischen Tests. Praktisch.

STAHEL, W. A. (2000): Statistische Datenanalyse. Eine Einführung für Naturwissenschaftler. Vieweg, Braunschweig.
Gibt eine ausführliche Einführung in die Statistik. Geht hauptsächlich auf die wahrscheinlichkeitstheoretischen Grundlagen der Statistik ein. Theoretisch.

Einführungen in die Statistik (englisch)

DALGAARD, P. (2002): Introductory statistics with R. Springer, Berlin.
Einführung in die wichtigsten statistischen Verfahren und konkrete Anleitungen für deren Ausführung auf dem Statistikprogramm R (R kann gratis von www.r-project.org heruntergeladen werden). Praktisch.

HOWELL, D. C. (1995): Fundamental statistics for the behavioral sciences. Wadsworth, Belmont.
Leicht verständliche und gründliche Einführung in die Statistik. Mit Übungen.

Klassiker in der Biostatistik

SIEGEL, S. & CASTELLAN, N. J. (1988): Non-parametric statistics for the behavioral sciences. McGraw-Hill, New York.
Anleitungen für die klassischen nicht-parametrischen Methoden.

SOKAL, R. R. & ROHLF, F. J. (1995): Biometry. Freeman, New York.
Das Nachschlagewerk für jeden Biostatistiker.

ZAR, J. H. (1996): Biostatistical analysis. Prentice Hall, London.
Ein zweites Nachschlagewerk.

26. Mai 2006, Fränzi Korner,

oikostat – Biostatistische Analysen und Beratung, Ausserdorf 43, CH–6318 Ettiswil, fraenzi.korner@oikostat.ch, www.oikostat.ch