

Einfache statistische Testverfahren in der Biologie

Maria-Luise Hubert

Hubert, M.-L. 2008: **Simple methods for statistical testing**. Ber. Vogelwarte Hiddensee 18: 59 - 60. *This is a brief introduction in the use of simple statistical tests. Preconditions and practical application of the t-test and the χ^2 -test are explained and how those tests can be performed by the use of Excel® (MS Windows®).*

Vorbemerkungen

In der Biologie sind oft Vergleiche von Mittelwerten erforderlich. So wird z.B. gefragt, ob sich die Flügel-längen oder andere Körpermaße artspezifisch unterscheiden oder das Verhalten zweier Arten deutlich voneinander abweicht. Dem aufmerksamen Beobachter fallen die Unterschiede auf - aber auch, dass es Überschneidungen der Messbereiche gibt. Will man seine Beobachtungen veröffentlichen, wird sofort nach der Signifikanz gefragt, d.h., ob der beobachtete Unterschied nicht rein zufällig ist. Schließlich ist die Variabilität eine wichtige Eigenschaft lebender Organismen. Was ist zu tun?

Es gibt umfangreiche Statistik-Pakete wie SPSS, mit denen man zahllose Testverfahren durchführen kann – welche sind für den Anwender sinnvoll? Wie sind sie durchzuführen? Welche Voraussetzungen sind notwendig? Alles Fragen, die vom Statistik-Laien kaum zu beantworten sind. Wenn die Anforderungen nicht zu hoch sind, genügen oft wenige leicht durchzuführende Verfahren. Zwei sollen hier erläutert werden mit ein paar Hinweisen, wie man sie nutzen kann.

Der t-Test mit Excel

Wenn es sich um Körpermaße von Tieren handelt, deren Mittelwerte verglichen werden sollen, bietet sich der t-Test in seinen verschiedenen Formen an. Dieses Testverfahren findet man auch in Tabellenkalkulationen wie z.B. Excel. Voraussetzung des t-Testes ist näherungsweise Normalverteilung, die bei den genannten Maßen meist gegeben ist.

Für den t-Test sind die zu vergleichenden Messwerte in zwei Spalten einzutragen. Häufig wird man eine paarweise Zuordnung nicht vornehmen können. Falls aber doch eine genaue Paar-Zuordnung möglich ist, kann der folgende F-Test übersprungen werden. Grundsätzlich stellt sich aber zunächst die Frage, ob sich die Varianzen beider Messwertserien (Datenbereiche) signifikant voneinander unterscheiden, und man muss die Funktion „F-Test“ in einer

leeren Zelle aufrufen. Als „Matrix 1“ bzw. „Matrix 2“ markiert man jeweils die Datenbereiche und erhält einen Zahlenwert als Ergebnis. Dieser Wert ist die Wahrscheinlichkeit, einen Fehler zu begehen, wenn man behauptet, dass sich – in diesem Fall – die Varianzen wesentlich, d.h. signifikant voneinander unterscheiden. 5% oder ein Wert von 0,05 wird in der Biologie gewöhnlich als obere Grenze der Irrtumswahrscheinlichkeit α gewählt. Liegt der berechnete Wert darüber, so ist die Irrtumswahrscheinlichkeit zu hoch und die Aussage lautet, dass es keinen signifikanten Unterschied der Varianzen gibt.

Nun erfolgt der Aufruf der Funktion „t-Test“ in einer leeren Zelle. Man markiert die zu vergleichenden Zellbereiche für Matrix 1 und 2, nutzt meist den 2-seitigen Test und muss den Typ auswählen. Handelt es sich um gepaarte Werte – z.B. wenn man Paare untersucht hat - so wird für „Typ“ die 1 gewählt. Sind die Varianzen nur unwesentlich verschieden ($\alpha > 0,05$), so ist der Typ 2 (homoskedastisch) zu wählen, Typ 3 gilt für heteroskedastische Messreihen, also solche, deren Varianzen sich signifikant unterscheiden. Das Ergebnis ist wieder ein Zahlenwert, der wie beim F-Test zu beurteilen ist: Ist er kleiner oder gleich 0,05, also 5%, so lautet die Aussage: „Die Mittelwerte unterscheiden sich signifikant (wesentlich) voneinander mit einer Irrtumswahrscheinlichkeit 5%“. Ist er größer als 0,05, so gibt es keine signifikanten Unterschiede der Mittelwerte.

Der χ^2 -Test für die Vier-Felder-Tafel

Der t-Test ist nicht anwendbar, wenn es sich um Auszählungen handelt, d.h. das Ergebnis der Untersuchung nicht in Messwerten besteht, sondern in Häufigkeiten. Hier liegen keine quantitativen Unterschiede vor, sondern man hat qualitative Daten, etwa die Anzahl der Männchen bzw. der Weibchen oder zweier Arten in zwei verschiedenen Territorien. Auch hier wird für Aussagen wie „Art A bevorzugt Feuchtgebiete, Art B trockene Standorte“ eine Signifikanz gefordert. Diese Fragestellung kann mit dem χ^2 -Test bearbeitet werden.

Im einfachsten Fall hat man wie oben vier Häufigkeiten: Die Häufigkeiten von Art A im trockenen bzw. Feucht-Bereich ebenso wie für Art B. Dafür wendet man den 4-Felder-Test wie in der folgenden Tabelle an: a, b, c und d sind die ausgezählten Häufigkeiten, ja und nein etwa für die Frage „Feuchtgebiet ja oder nein“ (Merkmal 2),

Merkmal2 → Merkmal1 ↓	ja	nein	Σ
-	a	b	a+b=n ₁
+	c	d	c+d=n ₂
Σ	a+c	b+d	a+b+c+d=n

Merkmal 1 ist - für Art A und + für Art B. In den Summenfeldern erscheinen die Randsummen, n ist die Gesamtzahl aller beobachteten Individuen. Der 2-Wert errechnet sich nach der folgenden Formel:

$$\hat{\chi}^2 = \frac{n * (a * d - b * c)^2}{(a + b) * (c + d) * (a + c) * (b + d)}$$

Für den errechnete Wert gilt die 2-Tafel für 1 Freiheitsgrad:

FG \ α	0,10	0,05	0,025	0,01	0,001	0,0001
1	2,7055	3,8415	5,0239	6,6349	10,8276	15,1367

Ist der berechnete χ^2 -Wert größer oder gleich dem Tafelwert – meist also 3,8415 für $\alpha = 0,05$ – so unterscheiden sich die Arten in ihrem Verhalten signifikant; ist er kleiner als dieser Tafelwert und damit $\alpha > 0,05$, so gibt es keine nachweisbaren Unterschiede. Bei kleinem n₁ oder n₂ (≤ 6) wird mit (n - 1) anstatt mit n multipliziert:

$$\hat{\chi}^2 = \frac{(n - 1) * (a * d - b * c)^2}{(a + b) * (c + d) * (a + c) * (b + d)}$$

Liegt n zwischen 20 und 60, ist die Kontinuitätskorrektur nach Yates nötig:

$$\hat{\chi}^2 = \frac{n * (|a * d - b * c| - n/2)^2}{(a + b) * (c + d) * (a + c) * (b + d)}$$

Der Test erfolgt stets wie oben beschrieben.

Sollen mehr als vier Häufigkeiten verglichen werden, lautet die Formel:

$$\hat{\chi}^2 = \sum \frac{(B - E)^2}{E} = \sum \left(\frac{B^2}{E} \right) - n$$

mit B für die einzelnen beobachteten und E für die jeweils zugehörigen erwarteten Häufigkeiten. Jedoch sind diese Aussagen umso schwieriger zu interpretieren, je mehr Gruppen gebildet werden. Das Ergebnis der Mehr-Felder-Tafel ist ohnehin nur die Aussage: „Es gibt irgendwo signifikante Unterschiede“ oder eben nicht. Günstiger ist es, seine Daten gleich in einer Vier-Felder-Tafel zusammenzufassen. Excel bietet den χ^2 -Test ebenfalls an. Man muss jedoch für seine Durchführung zuerst zu den beobachteten Häufigkeiten die jeweils zugehörige erwartete Häufigkeit berechnen. Der hier beschriebene Vier-Felder-Test ist mindestens ebenso leicht sogar mit dem Taschenrechner durchzuführen.

Adresse der Autorin:

Dr. Maria-Luise Hubert
Sebastian-Bach-Str. 25
18069 Rostock
mar1lu9@web.de

ZOBODAT - www.zobodat.at

Zoologisch-Botanische Datenbank/Zoological-Botanical Database

Digitale Literatur/Digital Literature

Zeitschrift/Journal: [Berichte aus der Vogelwarte Hiddensee](#)

Jahr/Year: 2007

Band/Volume: [2007_18](#)

Autor(en)/Author(s): Hubert Maria-Luise

Artikel/Article: [Einfache statische Testverfahren in der Biologie 53-54](#)