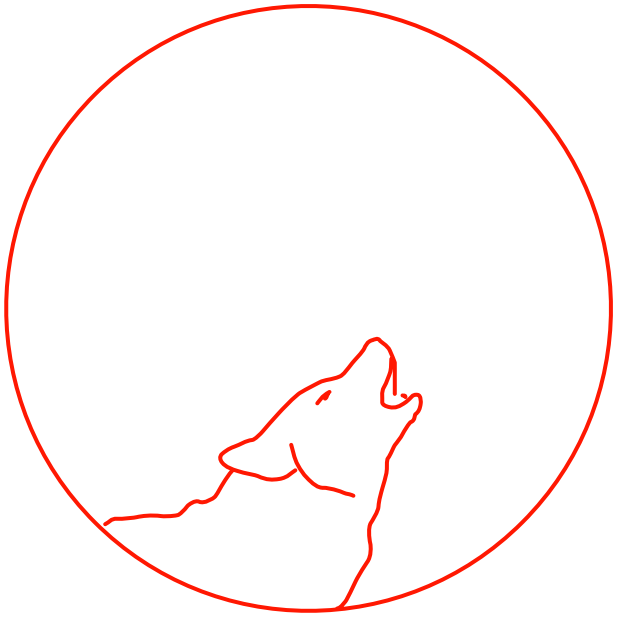
Hans Walser

Mathematik 2 für Naturwissenschaften



Modul 207

Testen von Hypothesen 1

Lernumgebung

UNI
BASEL

Inhalt

1 Ist der Würfel gefälscht? 1
2 Cola gegen Cola 1
3 Statistik der Geburten 3
4 99%-Vertrauensintervall 4
5 Wie heißt der Mörder? 5
6 95%-Vertrauensintervall 5
7 Akkus 6
8 Akkus zum zweiten 8
9 Wer ist der Schwerste? 9
10 Wer ist der Größte? 10
11 Biologen gegen Pharmazeuten 11
12 Hochsprung 12
13 Länge von Kuckuckseiern 13
14 Neue Kartoffelsorte 15
15 Noten in Mathe 16
16 Fische 18

Modul 207 für die Lehrveranstaltung *Mathematik 2 für Naturwissenschaften*

Sommer 2006 Probeversion
Sommer 2007 Ergänzungen und Korrekturen
Frühjahr 2008 Ergänzungen
Frühjahr 2009 Keine Änderung
Frühjahr 2010 Erweiterung
Frühjahr 2011 Fehlerkorrektur und Erweiterung
Frühjahr 2014 Überarbeitung. Kürzung

last modified: 25. März 2014

Hans Walser
Mathematisches Institut, Rheinsprung 21, 4051 Basel
www.walser-h-m.ch/hans

1 Ist der Würfel gefälscht?

H_0: $p(\textit{Augenzahl "sechs"}) = \frac{1}{6}$

H_1: $p(\text{Augenzahl "sechs"}) < \frac{1}{6}$ oder $p(\text{Augenzahl "sechs"}) > \frac{1}{6}$

Es soll $n = 600$ mal gewürfelt werden.

a) Der Verwerfungsbereich für H_0 sei $\{0,\ldots,80\} \cup \{120,\ldots,600\}$ (Zweiseitiger Test). Wie gross ist die Wahrscheinlichkeit α für einen Fehler erster Art?

b) Wählen Sie den Verwerfungsbereich möglichst groß, aber so, dass $\alpha < 5\%$.

c) Wählen Sie den Verwerfungsbereich möglichst groß, aber so, dass $\alpha < 1\%$.

Ergebnis

a) $\alpha = 3.236\%$

b) Verwerfungsbereich $\{0,\ldots,81\} \cup \{119,\ldots,600\}$

c) Verwerfungsbereich $\{0,\ldots,75\} \cup \{125,\ldots,600\}$

2 Cola gegen Cola

A lucky guess is never merely luck.

[Austen 1816, p. 11]

Vladimir behauptet, er könne mit einer Sicherheit von 70% C-Cola von P-Cola unterscheiden.

Annina stellt mit Vladimir folgenden Test an: Sie stellt ihm 20 Gläser Cola zur Bestimmung vor. Falls er mindestens 15 Gläser richtig identifiziert, glaubt sie seiner Behauptung.

Wie groß ist die Wahrscheinlichkeit eines Fehlers erster Art bei diesem Test?

Wie groß ist die Wahrscheinlichkeit eines Fehlers zweiter Art bei diesem Test?

[Austen 1816] Austen, Jane: Emma. Penguin Popular Classics 1994. ISBN 0-14-062010-9

Bearbeitung

Nullhypothese H_0: Vladimir rät bloß. Seine Trefferwahrscheinlichkeit im Einzelversuch ist also $p = \frac{1}{2}$.

Die Wahrscheinlichkeit α, dass er gleichwohl den Test besteht, ist:

$$\alpha = P_{20}(k \geq 15) = \sum_{k=15}^{20} \binom{20}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{20-k}$$

Um die Tabelle verwenden zu können, müssen wir das umformen:

$$\alpha = P_{20}(k \geq 15) = \sum_{k=15}^{20} \binom{20}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{20-k}$$

$$= 1 - \sum_{k=0}^{14} \binom{20}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{20-k} = 1 - 0.979 = 0.021 \approx 2\%$$

Die Wahrscheinlichkeit eines Fehlers erster Art ist also etwa 2%.

Für die Berechnung der Wahrscheinlichkeit eines Fehlers zweiter Art gehen wir von $p = 70\% = 0.7$ aus und berechnen die Wahrscheinlichkeit, dass Vladimir, obwohl er nicht bloß rät, beim Test durchfällt:

$$\beta = P_{20}(k \leq 14) = \sum_{k=0}^{14} \binom{20}{k} 0.7^k 0.3^{20-k}$$

Dummerweise ist $p = 0.7$ nicht in der Tabelle.

Wir müssen entweder umrechnen oder logisch denken.

Zuerst umrechnen, und zwar so, dass wir die Tabelle für $p = 1 - 0.7 = 0.3$ anwenden können:

$$\beta = P_{20}(k \leq 14) = \sum_{k=0}^{14} \binom{20}{k} 0.7^k 0.3^{20-k}$$

$$= \binom{20}{0} 0.7^0 0.3^{20} + \cdots + \binom{20}{14} 0.7^{14} 0.3^6$$

$$= \binom{20}{14} 0.7^{14} 0.3^6 + \cdots + \binom{20}{0} 0.7^0 0.3^{20}$$

$$= \binom{20}{6} 0.7^{14} 0.3^6 + \cdots + \binom{20}{20} 0.7^0 0.3^{20}$$

$$= \binom{20}{6} 0.3^6 0.7^{14} + \cdots + \binom{20}{20} 0.3^{20} 0.7^0$$

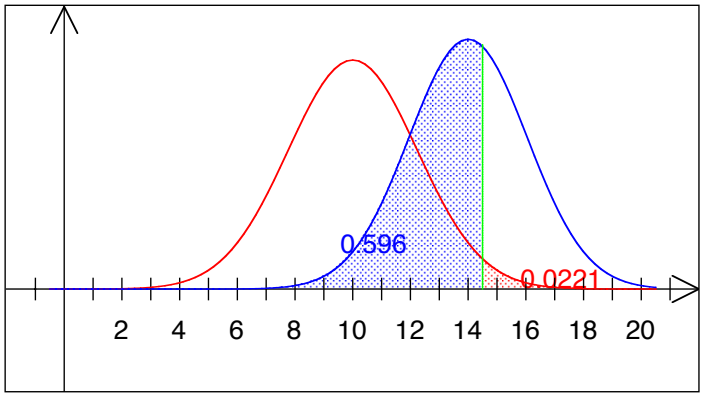
$$= \sum_{k=6}^{20} \binom{20}{k} 0.3^k 0.7^{20-k} = 1 - \sum_{k=0}^{5} \binom{20}{k} 0.3^k 0.7^{20-k} = 1 - 0.416 = 0.584 \approx 58\%$$

Logisch denken: Vladimir fällt durch, wenn er höchstens 14 Erfolge erzielt. Das heißt, dass er dann mindestens sechs Misserfolge einstecken muss. Die Wahrscheinlichkeit dafür ist:

$$\beta = \sum_{k=6}^{20} \binom{20}{k} 0.3^k 0.7^{20-k} = 1 - \sum_{k=0}^{5} \binom{20}{k} 0.3^k 0.7^{20-k} = 1 - 0.416 = 0.584 \approx 58\%$$

Die Wahrscheinlichkeit eines Fehlers zweiter Art ist also etwa 58%.

Die Figur zeigt die Situation approximiert durch die Normalverteilung. Rot entspricht der Nullhypothese mit $p = \frac{1}{2}$, blau der Hypothese mit $p = 70\% = 0.7$.



Approximation durch Normalverteilung

3 Statistik der Geburten

Zeitungsnotiz: In Middletown wurden im Mai 60 Knaben, aber nur 35 Mädchen geboren.

Die Wahrscheinlichkeit für eine Knabengeburt beträgt $p = 0.514$. Ist die oben stehende Meldung eine echte Sensation?

Wie groß ist die Wahrscheinlichkeit, dass die beobachtete Anzahl Knabengeburten um so viel oder mehr wie in der Zeitungsnotiz von der erwarteten Anzahl abweicht.

Ergebnis

2.852% (zweiseitig). Also doch eine kleine Sensation.

Bearbeitung

Wir arbeiten mit der Normalverteilung. Es ist:

$$n = 95,\ p = 0.514,\ \mu = np = 48.83,\ \sigma = \sqrt{np(1-p)} \approx 4.8715$$

Für die obere Grenze $60 - \frac{1}{2}$ erhalten wir standardisiert:

$$u_b = \frac{60 - \frac{1}{2} - \mu}{\sigma} \approx 2.1903$$

Die Tabelle liefert:

$$\Phi(u_b) = \Phi(2.19) = 0.98574$$

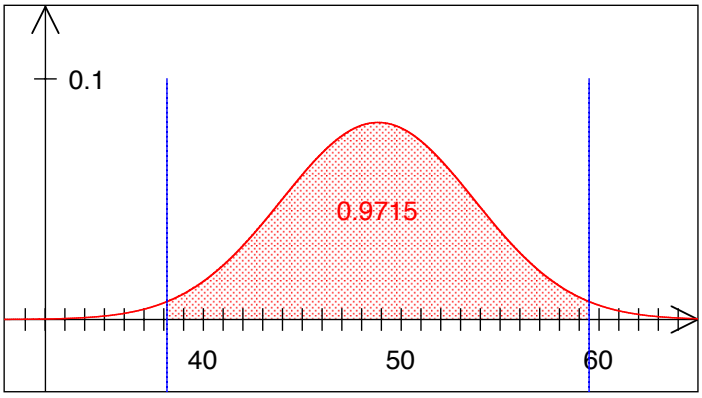
Die Wahrscheinlichkeit, dass 60 oder mehr Knaben geboren werden, ist:

$$1 - \Phi(2.19) = 1 - 0.98574 = 0.01426$$

Entsprechend die Wahrscheinlichkeit für eine Abweichung nach unten. Somit das Resultat:

$$2\left(1 - \Phi(2.19)\right) = 0.02852 = 2.852\%$$

In der folgenden Grafik ist das der Anteil außerhalb des roten Bereiches.



97.15% im roten Bereich

4 99%-Vertrauensintervall

Wir haben folgende acht Messwerte:

4.4	5.8	3.7	9.2	4.1	3.8	5.3	3.7

Für diese Messwerte gilt:

$$\bar{x} = 5, \quad s_x = SD_x = 1.8655, \quad SE_x = \frac{s_x}{\sqrt{8}} = 0.6595$$

Von wo bis wo geht das zu diesen Daten gehörende 99% Vertrauensintervall?

Bevor Sie losrechnen: Die Messwerte sind dieselben wie im Beispiel in der Vorlesung. Statt $\alpha = 5\%$ haben wir hier aber $\alpha = 1\%$. Wir wollen also eine kleinere Irrtumswahrscheinlichkeit, positiv formuliert: Mehr Sicherheit. Wird dadurch das Vertrauensintervall größer oder kleiner?

Bearbeitung

In unserem Beispiel ist $\nu = n - 1 = 8 - 1 = 7$.

Relevanter Ausschnitt aus der Tabelle:

FG	Irrtumswahrscheinlichkeit α für den zweiseitigen Test								
ν	0.50	0.20	0.10	0.05	0.02	0.01	0.002	0.001	0.0001
6	0.718	1.440	1.943	2.441	3.143	3.707	5.208	5.959	9.082
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785	5.408	7.885
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501	5.041	7.120
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297	4.781	6.594
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144	4.587	6.211

Schranken der *t*-Verteilung

Es ist $t_{krit} = 3.499$; der kritische Wert ist größer als im alten Beispiel. Die Latte wird höher gelegt.

99% Vertrauensintervall:

$$\left[\bar{x} - t_{1\%,7} \cdot SE_x, \bar{x} + t_{1\%,7} \cdot SE_x\right] = \left[5 - 3.499 \cdot 0.6595, 5 + 3.499 \cdot 0.6595\right]$$
$$= \left[2.6924, 7.3075\right]$$

Das Vertrauensintervall ist länger geworden. Wir sind vorsichtiger mit dem Ablehnen der Nullhypothese.

Anders formuliert: Wenn wir ein kleineres Vertrauensintervall wählen, wird die Irrtumswahrscheinlichkeit größer.

5 Wie heißt der Mörder?

Auf der Suche nach dem Mörder

Miss Marple sagt, der Mörder sei mit 95% Wahrscheinlichkeit in der Liste {Jim, Jack, Joe, Jennifer, John} enthalten. Beim selben Mordfall sagt aber Sherlock Holmes, der Mörder sei mit 95% Wahrscheinlichkeit in der Liste {Jury, Jack, Jan, Jennifer, John} enthalten. Maigret schließlich gibt die Liste {Jury, Jack, Jennifer, Janine} an.

Wie heißt der Mörder?

Bearbeitung

Das können wir nicht sagen. Er könnte auch Julius heißen.

6 95%-Vertrauensintervall

Wir haben zwei Stichproben, welche auf derselben Grundgesamtheit basieren.

a)

3.5	3.5	3.8	3.2	2.9	3.5

b)

3.6	3.3	3.4	3.7

Geben Sie zu beiden Fällen ein 95%-Vertrauensintervall für den Mittelwert μ der gemeinsamen Grundgesamtheit.

Bearbeitung

a)

						Mittelwert	SD	SE
3.5	3.5	3.8	3.2	2.9	3.5	3.4	0.3098	0.1265

Es ist $\nu = n - 1 = 6 - 1 = 5$.

Relevanter Ausschnitt aus der Tabelle:

FG	Irrtumswahrscheinlichkeit α für den zweiseitigen Test								
ν	0.50	0.20	0.10	0.05	0.02	0.01	0.002	0.001	0.0001
1	1.000	3.078	6.3114	12.706	31.821	63.65	318.30	636.61	6366.1
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327	31.598	99.992
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214	12.924	28.000
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173	8.610	15.544
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893	6.869	11.178

Schranken der *t*-Verteilung

95%-Vertrauensintervall:

$$\left[\bar{x} - t_{5\%,5} \cdot SE_x, \bar{x} + t_{5\%,5} \cdot SE_x\right] = \left[3.4 - 2.571 \cdot 0.1265, 3.4 + 2.571 \cdot 0.1265\right]$$
$$= \left[3.075, 3.725\right]$$

b)

				Mittelwert	SD	SE
3.6	3.3	3.4	3.7	3.5	0.1826	0.0913

Es ist $\nu = n - 1 = 4 - 1 = 3$.

Relevanter Ausschnitt aus der Tabelle:

FG	Irrtumswahrscheinlichkeit α für den zweiseitigen Test								
ν	0.50	0.20	0.10	0.05	0.02	0.01	0.002	0.001	0.0001
1	1.000	3.078	6.3114	12.706	31.821	63.65	318.30	636.61	6366.1
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327	31.598	99.992
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214	12.924	28.000
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173	8.610	15.544
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893	6.869	11.178

Schranken der *t*-Verteilung

95%-Vertrauensintervall:

$$\left[\bar{x} - t_{5\%,3} \cdot SE_x, \bar{x} + t_{5\%,3} \cdot SE_x\right] = \left[3.5 - 3.182 \cdot 0.0913,\ 3.5 + 3.182 \cdot 0.0913\right]$$
$$= \left[3.209,\ 3.791\right]$$

7 Akkus

Beträgt die durchschnittliche Laufzeit eines Akkus wirklich mindestens 3.2 Stunden, wie der Händler behauptet?

Wir testen auf dem Signifikanzniveau 5%. Dazu werden bei 10 Akkus die Laufzeiten gemessen. Es ergeben sich die Laufzeiten:

Akku Nr	1	2	3	4	5	6	7	8	9	10
Laufzeit	3.1	3.2	2.5	4	3.8	2.6	3.5	3.8	3.2	2.9

Bearbeitung

Wegen „*mindestens* 3.2 Stunden“ testen wir einseitig.

Wir erhalten zunächst:

Akku Nr	1	2	3	4	5	6	7	8	9	10	Mittelwert	SD
Laufzeit	3.1	3.2	2.5	4	3.8	2.6	3.5	3.8	3.2	2.9	3.26	0.512510163

Mittelwert	SD	SE	Testgrösse t_Exp
3.26	0.512510163	0.162069944	0.370210531

Für diese Messwerte gilt also:

$$\bar{x} = 3.26, \quad s = SD = 0.512510163, \quad SE = \frac{s}{\sqrt{10}} = 0.162069944$$

Nullhypothese: Der Mittelwert 3.26 liegt rein zufällig oberhalb der vom Hersteller angegebenen Grenze 3.2.

Für die Testgröße erhalten wir:

$$t_{\text{Exp}} = \frac{|\bar{x} - 3.2|}{SE} = \frac{3.26 - 3.2}{0.162069944} = \frac{0.06}{0.162069944} = 0.370210531$$

Wir haben den Freiheitsgrad $\nu = 10 - 1 = 9$.

Tabelle, relevanter Ausschnitt:

FG	Irrtumswahrscheinlichkeit α für den zweiseitigen Test								
ν	0.50	0.20	0.10	0.05	0.02	0.01	0.002	0.001	0.0001
6	0.718	1.440	1.943	2.441	3.143	3.707	5.208	5.959	9.082
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785	5.408	7.885
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501	5.041	7.120
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297	4.781	6.594
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144	4.587	6.211
FG	0.25	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	0.00005
ν	Irrtumswahrscheinlichkeit α für den einseitigen Test								

t-Verteilung

Die Tabelle liefert den kritischen Wert 1.833. Da unsere Testgröße $t_{\text{Exp}} = 0.370210531$ kleiner ist, müssen wir die Nullhypothese beibehalten.

Bemerkung 1: Aus der Sicht des Konsumenten macht es natürlich keinen Sinn, bei den gegebenen Daten mit $\bar{x} = 3.26$ an den Angaben des Herstellers (mindestens 3.2) zu zweifeln. Anders sieht die Sache aus der Sicht des Händlers aus. Falls es sich zeigen sollte (was sich in unserem Beispiel aber nicht gezeigt hat), dass $\bar{x} = 3.26$ bei der Nullhypothese 3.2 sehr unwahrscheinlich ist, also die Nullhypothese zu verwerfen ist, könnte der Händler eine höhere mittlere Laufzeit versprechen.

Bemerkung 2: Excel liefert:

TVERT(0.370210531;9;1) = 0.359893981

Dabei ist 0.370210531 die oben berechnete Testgröße, 9 der Freiheitsgrad und 1 bedeutet, dass einseitig getestet wird. Die Ausgabe 0.359893981 bedeutet, dass die Wahrscheinlichkeit eines so „extremen“ oder noch extremeren Ergebnisses für die Testgröße nur ca. 36% beträgt. Das ist nicht umwerfend. Die Nullhypothese kann nicht verworfen werden.

8 Akkus zum zweiten

Beträgt die durchschnittliche Laufzeit eines Akkus wirklich mindestens 3.2 Stunden, wie der Händler behauptet?

Wir testen auf dem Signifikanzniveau 5%. Dazu werden bei 10 Akkus die Laufzeiten gemessen. Es ergeben sich die Laufzeiten:

Akku Nr	1	2	3	4	5	6	7	8	9	10
Laufzeit	3.1	3.5	2.6	4.1	3.8	3.6	3.9	3.8	3.7	2.9

Bearbeitung

Wegen „*mindestens* 3.2 Stunden“ testen wir einseitig.

Wir erhalten zunächst:

Akku Nr	1	2	3	4	5	6	7	8	9	10	Mittelwert	SD
Laufzeit	3.1	3.5	2.6	4.1	3.8	3.6	3.9	3.8	3.7	2.9	3.5	0.48074017

Mittelwert	SD	SE	Testgrösse t_Exp
3.5	0.48074017	0.15202339	1.973380543

Für diese Messwerte gilt also:

$$\bar{x} = 3.5, \quad s = SD = 0.48074017, \quad SE = \frac{s}{\sqrt{10}} = 0.15202339$$

Nullhypothese: Der Mittelwert 3.5 liegt rein zufällig oberhalb der vom Hersteller angegebenen Grenze 3.2.

Für die Testgröße erhalten wir:

$$t_{\text{Exp}} = \frac{|\bar{x} - 3.2|}{SE} = \frac{3.5 - 3.2}{0.162069944} = \frac{0.3}{0.15202339} = 1.973380543$$

Wir haben den Freiheitsgrad $\nu = 10 - 1 = 9$.

Tabelle, relevanter Ausschnitt:

FG	Irrtumswahrscheinlichkeit α für den zweiseitigen Test								
ν	0.50	0.20	0.10	0.05	0.02	0.01	0.002	0.001	0.0001
6	0.718	1.440	1.943	2.441	3.143	3.707	5.208	5.959	9.082
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785	5.408	7.885
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501	5.041	7.120
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297	4.781	6.594
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144	4.587	6.211
FG	0.25	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	0.00005
ν	Irrtumswahrscheinlichkeit α für den einseitigen Test								

t-Verteilung

Die Tabelle liefert den kritischen Wert 1.833. Da unsere Testgröße $t_{\text{Exp}} = 1.973380543$ größer ist, können wir die Nullhypothese verwerfen.

Bemerkung 1: Aus der Sicht des Konsumenten macht es natürlich keinen Sinn, bei den gegebenen Daten mit $\bar{x} = 3.5$ an den Angaben des Händlers (mindestens 3.2) zu zweifeln. Anders sieht die Sache aus der Sicht des Herstellers aus. Falls es sich wie in unserem Beispiel zeigt, dass $\bar{x} = 3.5$ bei der Nullhypothese 3.2 sehr unwahrscheinlich ist, kann die Nullhypothese verworfen werden. Das heißt, der Händler kann mit gutem Gewissen eine höhere mittlere Laufzeit versprechen.

Bemerkung 2: Excel liefert:

TVERT(1.973380543;9;1) = 0.039953182

Dabei ist 1.973380543 die oben berechnete Testgröße, 9 der Freiheitsgrad und 1 bedeutet, dass einseitig getestet wird. Die Ausgabe 0.039953182 bedeutet, dass die Wahrscheinlichkeit eines so extremen oder noch extremeren Ergebnisses für die Testgröße nur 3.9953% beträgt. Es ist also $\alpha < 5\%$.

9 Wer ist der Schwerste?

Im Frühjahrsemester 2008 wurde in der Lehrveranstaltung *Mathematik für Naturwissenschaften* an der Uni Basel bei $n_x = 85$ Studenten ein Durchschnittsgewicht $\bar{x} = 74.084$ kg mit einer Standardabweichung $s_x = 10.447$ kg festgestellt. Bei $n_y = 90$ Studentinnen ergab sich ein Durchschnittsgewicht $\bar{y} = 56.944$ kg und eine Standardabweichung $s_y = 7.193$ kg.

Wir testen die Nullhypothese, dass Studenten und Studentinnen gleich schwer sind, gegen die Alternative, dass Studenten schwerer sind. Wir haben also einen *einseitigen* Test. Als Signifikanzniveau wählen wir 1 Promille.

Bearbeitung

Testgröße:

$$t = (\bar{x} - \bar{y})\sqrt{\frac{n_x n_y}{n_x + n_y}}\sqrt{\frac{n_x + n_y - 2}{s_x^2(n_x - 1) + s_y^2(n_y - 1)}} = 12.701$$

Und nun gehen wir in die *t*-Tabelle. Wir haben $\nu = n_x + n_y - 2$ Freiheitsgrade. In unserem Fall ist $\nu = n_x + n_y - 2 = 85 + 90 - 2 = 173$. Das ist nicht in der Tabelle, aber $\nu = 200$ tut es auch. Weil wir einseitig testen, müssen wir unten in die Tabelle einsteigen.

Tabelle, relevanter Ausschnitt:

100	0.677	1.290	1.660	1.984	2.364	2.626	3.114	3.390	4.053
200	0.676	1.286	1.653	1.972	2.345	2.601	3.131	3.340	3.970
500	0.675	1.283	1.648	1.965	2.334	2.586	3.107	3.310	3.92.2
∞	0.674	1.226	1.645	1.960	2.326	2.575	3.090	3.290	3.890
FG	0.25	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	0.00005
ν	Irrtumswahrscheinlichkeit α für den einseitigen Test								

Schranken der *t*-Verteilung

Somit erhalten wir auf dem Signifikanzniveau 0.001 den kritischen Schrankenwert $t_{krit} \approx 3.131$. Dies ist deutlich kleiner als unsere berechnete Testgröße $t \approx 12.701$. Wir können also die Nullhypothese verwerfen. Die Studenten sind schwerer.

10 Wer ist der Größte?

Im Sommersemester 2006 wurde an der Uni Basel bei $n_x = 90$ Studenten eine Durchschnittsgröße $\bar{x} = 1.8059$ m mit einer Standardabweichung $s_x = 0.0752$ m festgestellt. Bei $n_y = 110$ Studentinnen ergab sich eine Durchschnittsgröße $\bar{y} = 1.6854$ m und eine Standardabweichung $s_y = 0.0604$ m.

Wir testen die Nullhypothese, dass Studenten und Studentinnen gleich groß sind, gegen die Alternative, dass Studenten größer sind. Wir haben also einen *einseitigen* Test. Als Signifikanzniveau wählen wir 1%.

Bearbeitung

Testgröße:

$$t = (\bar{x} - \bar{y})\sqrt{\frac{n_x n_y}{n_x + n_y}}\sqrt{\frac{n_x + n_y - 2}{s_x^2(n_x - 1) + s_y^2(n_y - 1)}} = 12.5682$$

Und nun gehen wir in die *t*-Tabelle. Wir haben $\nu = n_x + n_y - 2$ Freiheitsgrade. In unserem Fall ist $\nu = n_x + n_y - 2 = 90 + 110 - 2 = 198$. Das ist nicht in der Tabelle, aber $\nu = 200$ tut es auch. Weil wir einseitig testen, müssen wir unten in die Tabelle einsteigen.

Tabelle, relevanter Ausschnitt:

100	0.677	1.290	1.660	1.984	2.364	2.626	3.114	3.390	4.053
200	0.676	1.286	1.653	1.972	2.345	2.601	3.131	3.340	3.970
500	0.675	1.283	1.648	1.965	2.334	2.586	3.107	3.310	3.92.2
∞	0.674	1.226	1.645	1.960	2.326	2.575	3.090	3.290	3.890
FG	0.25	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	0.00005
ν	Irrtumswahrscheinlichkeit α für den einseitigen Test								

Schranken der *t*-Verteilung

Somit erhalten wir auf dem Signifikanzniveau 1% den kritischen Schrankenwert $t_{krit} \approx 2.345$. Dies ist deutlich kleiner als unsere berechnete Testgröße $t \approx 12.5682$. Wir können also die Nullhypothese verwerfen. Die Studenten sind größer oder zumindest länger.

11 Biologen gegen Pharmazeuten

Im Frühjahrssemester 2008 wurden in der Lehrveranstaltung Mathematik 2 für Naturwissenschaften an der Uni Basel bei $n_x = 36$ männlichen Biologie-Studierenden eine Durchschnittsgröße $\bar{x} = 1.811\,\text{m}$ mit einer Standardabweichung $s_x = 0.058\,\text{m}$ festgestellt. Bei $n_y = 22$ männlichen Pharmazie-Studierenden ergab sich eine Durchschnittsgröße $\bar{y} = 1.822\,\text{m}$ und eine Standardabweichung $s_y = 0.119\,\text{m}$. Ist die Abweichung signifikant auf dem Signifikanzniveau 1 Promille? (Zweiseitiger Test).

Bearbeitung

Testgröße:

$$t_{Exp} = |\bar{x} - \bar{y}| \sqrt{\frac{n_x n_y}{n_x + n_y}} \sqrt{\frac{n_x + n_y - 2}{s_x^2 (n_x - 1) + s_y^2 (n_y - 1)}} = 0.4721151540$$

Und nun gehen wir in die *t*-Tabelle. Wir haben $\nu = n_x + n_y - 2$ Freiheitsgrade. In unserem Fall ist $\nu = n_x + n_y - 2 = 36 + 22 - 2 = 56$. Das ist nicht in der Tabelle, aber $\nu = 60$ tut es auch. Weil wir zweiseitig testen, müssen wir oben in die Tabelle einsteigen.

Studentsche *t*-Verteilung

FG	Irrtumswahrscheinlichkeit α für den zweiseitigen Test								
ν	0.50	0.20	0.10	0.05	0.02	0.01	0.002	0.001	0.0001
35	0.682	1.306	1.690	2.030	2.438	2.724	3.340	3.591	4.389
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307	3.551	4.321
45	0.680	1.301	1.679	2.014	2.412	2.690	3.281	3.520	4.269
50	0.679	1.299	1.676	2.009	2.403	2.678	3.261	3.496	4.228
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232	3.460	4.169
70	0.678	1.294	1.667	1.994	2.381	2.648	3.211	3.435	4.127
80	0.678	1.292	1.664	1.990	2.374	2.639	3.195	3.416	4.096
90	0.677	1.291	1.662	1.987	2.368	2.632	3.183	3.402	4.072
100	0.677	1.290	1.660	1.984	2.364	2.626	3.114	3.390	4.053
	0.25	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	0.00005
FG	Irrtumswahrscheinlichkeit α für den einseitigen Test								
ν									

Somit erhalten wir auf dem Signifikanzniveau 1 Promille den kritischen Schrankenwert $t_{krit} \approx 3.460$. Dies ist größer als unsere berechnete Testgröße $t_{Exp} \approx 0.472$. Wir müssen also die Nullhypothese beibehalten.

12 Hochsprung

Wir untersuchen die Hochsprungleistungen zweier Sportvereine. Dabei kennen wir folgende Werte:

$$n_x = 24 \qquad \bar{x} = 151\text{cm} \qquad s_x = 8.98\text{cm}$$
$$n_y = 18 \qquad \bar{y} = 155.5\text{cm} \qquad s_y = 10.10\text{cm}$$

Signifikanzniveau 5%.

Bearbeitung

$\alpha = 5\%$. Zweiseitiger Test.

Testgröße:

$$t_{Exp} = (\bar{x} - \bar{y})\sqrt{\frac{n_x n_y}{n_x + n_y}}\sqrt{\frac{n_x + n_y - 2}{s_x^2(n_x - 1) + s_y^2(n_y - 1)}} = 4.5 \cdot \sqrt{\frac{24 \cdot 18}{42}}\sqrt{\frac{40}{8.98^2 \cdot 23 + 10.10^2 \cdot 17}} \approx 1.5236$$

Freiheitsgrad:

$$\nu = n_x + n_y - 2 = 24 + 18 - 2 = 40$$

Tabelle (relevanter Ausschnitt):

FG	Irrtumswahrscheinlichkeit α für den zweiseitigen Test								
v	0.50	0.20	0.10	0.05	0.02	0.01	0.002	0.001	0.0001
35	0.682	1.306	1.690	2.030	2.438	2.724	3.340	3.591	4.389
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307	3.551	4.321
45	0.680	1.301	1.679	2.014	2.412	2.690	3.281	3.520	4.269
50	0.679	1.299	1.676	2.009	2.403	2.678	3.261	3.496	4.228
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232	3.460	4.169
FG	0.25	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	0.00005
v	Irrtumswahrscheinlichkeit α für den einseitigen Test								

Schranken der *t*-Verteilung

Wegen $t_{Exp} \approx 1.5236 < 2.021$ muss die Nullhypothese beibehalten werden. Die beiden Sportvereine unterscheiden sich nicht im Hochsprung.

13 Länge von Kuckuckseiern

Kuckuck

Für 25 Kuckuckseier ergaben sich beim Ausmessen folgende Längen [mm]:

aus Zaunkönigsnestern	aus Rohrsängernestern
22.2 19.9 21.6 21.0 22.1 21.1 22.4 21.1 20.4 21.0 22.1 20.1 20.9 21.3 21.1	23.3 22.1 22.3 21.3 21.7 21.7 22.7 22.0 23.0 22.9
Mittelwert: 21.22 mm	Mittelwert: 22.24 mm

Unterscheidet der Kuckuck zwischen Zaunkönigsnestern und Rohrsängernestern?

Wir vergleichen die beiden Mittelwerte (zweiseitig, $\alpha = 5\ \%$).

Bearbeitung

Z	1	19.9
Z	1	22.2
Z	1	21.6
Z	1	21
Z	1	22.1
Z	1	21.1
Z	1	22.4
Z	1	21.1
Z	1	20.4
Z	1	21
Z	1	22.1
Z	1	20.1
Z	1	20.9
Z	1	21.3
Z	1	21.1
Anzahl	15	
Mittelwert		21.22
SD		0.754226188

R	1	23.3
R	1	22.1
R	1	22.3
R	1	21.3
R	1	21.7
R	1	21.7
R	1	22
R	1	22.1
R	1	23
R	1	22.9
Anzahl	10	
Mittelwert		22.24
SD		0.641525959

Testgröße t	3.50785396

Zaunkönigsnester: $n_x = 15$ $\overline{x} = 21.22$ mm $s_x = 0.75422619$ mm

Rohrsängernester: $n_y = 10$ $\overline{y} = 22.24$ mm $s_y = 0.64152596$ mm

Testgröße:

$$t_{\text{Exp}} = (\overline{x} - \overline{y})\sqrt{\frac{n_x n_y}{n_x + n_y}}\sqrt{\frac{n_x + n_y - 2}{s_x^2(n_x - 1) + s_y^2(n_y - 1)}}$$

$$= 1.02 \cdot \sqrt{\frac{15 \cdot 10}{25}}\sqrt{\frac{23}{0.75422619^2 \cdot 14 + 0.64152596^2 \cdot 9}} \approx 3.507853954$$

Freiheitsgrad:

$$\nu = n_x + n_y - 2 = 15 + 10 - 2 = 23$$

$\alpha = 5\%$. Zweiseitiger Test.

Tabelle (relevanter Ausschnitt):

FG	Irrtumswahrscheinlichkeit α für den zweiseitigen Test								
v	0.50	0.20	0.10	0.05	0.02	0.01	0.002	0.001	0.0001
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527	3.819	4.784
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505	3.792	4.736
23	01685	1.319	1.714	2.069	2.500	2.807	3.485	3.767	4.693
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467	3.745	4.654
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450	3.725	4.619
FG	0.25	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	0.00005
v	Irrtumswahrscheinlichkeit α für den einseitigen Test								

Schranken der *t*-Verteilung

Wegen $t_{Exp} \approx 3.5078 > 2.069$ ist die Nullhypothese zu verwerfen. Der Kuckuck macht einen Unterschied zwischen Zaunkönigs- und Rohrsängernestern.

14 Neue Kartoffelsorte

Eine neue Kartoffelsorte wird mit einer Standardsorte verglichen, indem beide Sorten an je 6 Standorten angebaut werden.

Dabei ergeben sich Erträge nach folgender Tabelle:

	Erträge	
Standorte	Neue Sorte	Standardsorte
1	102	115
2	125	119
3	87	94
4	113	129
5	98	119
6	109	128

a) Welcher Signifikanz-Test ist für einen Vergleich geeignet? Begründen Sie Ihre Wahl.

b) Wie lautet das Test-Resultat (Signifikanzniveau 5%)

Bearbeitung

a) Die Standorte führen zu einer Paarung der beiden Kartoffelsorten. Wir verwenden den *t*-Test für gepaarte Stichproben.

b) Rechnungen

i	Neue Sorte	Standardsorte	Differenz
1	182	195	-13
2	205	199	6
3	167	174	-7
4	193	209	-16
5	178	199	-21
6	189	208	-19
Mittelwert	185.6666667	197.3333333	-11.666667
Streuung	13.10979278	12.6912043	9.95322393
Testgröße t	-2.8711682		

Nullhypothese: Beide Kartoffelsorten sind gleich gut. Die Testgröße ist negativ; die neue Weizensorte scheint ein Flop zu sein. Wir arbeiten mit dem Betrag 2.8711682 der Testgröße weiter.

Die Tabelle

FG	Irrtumswahrscheinlichkeit α für den zweiseitigen Test				
	0.20	0.10	0.05	0.02	0.01
1	3.078	6.3114	12.706	31.821	63.65
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032

liefert für den Freiheitsgrad 5 und das Signifikanzniveau 5% den Schrankenwert 2.571. Wegen 2.8711682 > 2.571 muss die Nullhypothese verworfen werden. Die neue Kartoffelsorte ist signifikant schlechter (was den Ertrag betrifft).

15 Noten in Mathe

Mathematik 1 für Naturwissenschaften, Prüfungen vom Fr, 22. Januar 2010

Es ergaben sich folgende Notendurchschnitte.

Studienrichtung	Anzahl	Notendurchschnitt	Standardabweichung
Biologie	153	4.460	1.085
Chemie	47	4.609	1.187
Geowissenschaften	63	4.474	1.015
Pharmazie	94	4.694	1.043

Sind die Abweichungen der Notendurchschnitte signifikant?

Bearbeitung

Beispiel

Wir wählen die Extrema, nämlich Biologie (x) und Pharmazie (y). Hier sind auch die Anzahlen am größten, so dass die Statistik am aussagekräftigsten ist. Wir testen die Nullhypothese, dass Biologie-Studierende gleich gut sind in Mathe wie Pharmaziestudierende, gegen die Alternative, dass Pharmaziestudierende besser sind. Als Signifikanzniveau wählen wir 1%.

$$t_{\text{Exp}} = \frac{|\bar{x}-\bar{y}|}{SE_{\bar{x}-\bar{y}}} = |\bar{x}-\bar{y}|\sqrt{\frac{n_x n_y}{n_x+n_y}}\sqrt{\frac{n_x+n_y-2}{s_x^2(n_x-1)+s_y^2(n_y-1)}} = 1.670$$

Wir haben $\nu = n_x + n_y - 2 = 245$.

Tabelle, relevanter Ausschnitt:

200	0.676	1.286	1.653	1.972	2.345	2.601	3.131	3.340	3.970
500	0.675	1.283	1.648	1.965	2.334	2.586	3.107	3.310	3.92.2
∞	0.674	1.226	1.645	1.960	2.326	2.575	3.090	3.290	3.890
FG	0.25	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	0.00005
ν	Irrtumswahrscheinlichkeit α für den einseitigen Test								

Schranken der *t*-Verteilung

Somit erhalten wir auf dem Signifikanzniveau 1% den kritischen Schrankenwert $t_{\text{krit}} \approx 2.334$. Dies ist größer als unsere berechnete Testgröße $t \approx 1.670$. Wir müssen also die Nullhypothese beibehalten. Pharmaziestudierende sind *nicht* besser in Mathe also Biologiestudierende.

Weitere Beispiele

Bio vs Chemie:

$t_{\text{Exp}} = 0.805$, Freiheitsgrade $\nu = 198$. Kritischer Schrankenwert = 2.345. Nullhypothese beibehalten.

Bio vs Geo:

$t_{\text{Exp}} = 0.088$, Freiheitsgrade $\nu = 214$. Kritischer Schrankenwert = 2.345. Nullhypothese beibehalten.

Chemie vs Geo:

$t_{\text{Exp}} = 0.642$, Freiheitsgrade $\nu = 108$. Kritischer Schrankenwert = 2.364. Nullhypothese beibehalten.

Chemie vs Pharma:

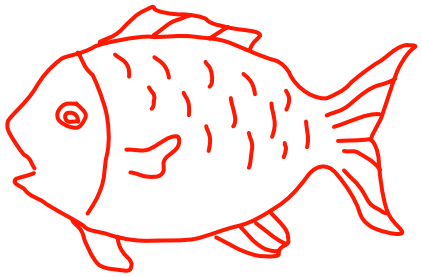
$t_{Exp} = 0.435$, Freiheitsgrade $\nu = 139$. Kritischer Schrankenwert = 2.355. Nullhypothese beibehalten.

Geo vs Pharma:

$t_{Exp} = 1.309$, Freiheitsgrade $\nu = 155$. Kritischer Schrankenwert = 2.354. Nullhypothese beibehalten.

Es ergeben sich keine signifikanten Unterschiede zwischen den verschiedenen Studienrichtungen.

16 Fische



Fisch

Brauchen Fische in fließendem Wasser mehr Sauerstoff als in stehendem Wasser?

Wasser	Sauerstoffverbrauch									
fließend	113	127	149	134	112	120	119	102	101	131
stehend	90	157	88	74	100	92	76	99		

Wir prüfen zuerst mit dem *F*-Test, ob die Varianzen signifikant unterschiedlich sind ($\alpha = 5\%$). Anschließend wenden wir uns der Hauptfrage zu ($\alpha = 5\%$).

Bearbeitung

Wasser	Sauerstoffverbrauch										Mittelwert	SD
fließend	113	127	149	134	112	120	119	102	101	131	120.8	14.9056
stehend	90	157	88	74	100	92	76	99			97	26.0165
Varianzquotienten-Test												
Testgröße F 3.0465												

Für den Varianzenquotienten-Test müssen wir die zweite Datenzeile für *x* nehmen. Es ist dann:

$$n_x = 8 \qquad \overline{x} = 97 \qquad s_x = 26.0165$$
$$n_y = 10 \qquad \overline{y} = 120.8 \qquad s_y = 14.9056$$

Wir erhalten für die Testgröße:

$$F_{\text{Exp}} = \frac{s_x^2}{s_y^2} \approx 3.0465$$

Nun gehen wir in die Tabelle für $\alpha = 5\%$. Die Freiheitsgrade für beide Stichproben sind $8-1=7$ und $10-1=9$.

F-Tabelle (relevanter Ausschnitt):

Freiheitsgrade für den Nenner (kleinere Varianz)	Freiheitsgrade für den Zähler (größere Varianz)									
	1	2	3	4	5	6	7	8	9	10
1	161	199	216	225	230	234	237	239	241	242
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98

Schranken der *F*-Verteilung für das Signifikanzniveau 5%

Aus der Tabelle erhalten wir in unserem Beispiel den kritischen Schrankenwert $F_{\text{krit}} = 3.29$. Dieser ist größer als $F_{\text{Exp}} \approx 3.0465$; wir können also die Nullhypothese der Varianzengleichheit beibehalten und übergehen zum nächsten Schritt:

Wir testen mit der *t*-Verteilung die Nullhypothese $\mu_x = \mu_y$ gegen die Alternative $\mu_x < \mu_y$. (x entspricht dem stehenden Wasser). Wir testen also einseitig. Es sei wieder $\alpha = 5\%$.

Testgröße:

$$t_{\text{Exp}} = (\overline{x} - \overline{y})\sqrt{\frac{n_x n_y}{n_x + n_y}}\sqrt{\frac{n_x + n_y - 2}{s_x^2(n_x - 1) + s_y^2(n_y - 1)}} \approx 2.44508$$

Freiheitsgrade:

$$\nu = n_x + n_y - 2 = 16$$

In der *t*-Tabelle müssen wir (einseitiger Test) unten einsteigen. Relevanter Tabellenausschnitt:

FG	Irrtumswahrscheinlichkeit α für den zweiseitigen Test								
v	0.50	0.20	0.10	0.05	0.02	0.01	0.002	0.001	0.0001
1	1.000	3.078	6.3114	12.706	31.821	63.65	318.30	636.61	6366.1
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327	31.598	99.992
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214	12.924	28.000
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173	8.610	15.544
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893	6.869	11.178
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686	4.015	5.134
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646	3.965	5.044
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610	3.922	4.966
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579	3.883	4.897
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552	3.850	4.837
∞	0.674	1.226	1.645	1.960	2.326	2.575	3.090	3.290	3.890
FG	0.25	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	0.00005
v	Irrtumswahrscheinlichkeit α für den einseitigen Test								

Schranken der *t*-Verteilung

Wir erhalten auf dem Signifikanzniveau 5% den kritischen Schrankenwert $t_{krit} \approx 1.746$. Dies ist deutlich kleiner als unsere berechnet Testgröße $t_{Exp} \approx 2.44508$. Wir können also die Nullhypothese verwerfen. Fische brauchen im fließendem Wasser mehr Sauerstoff als in stehendem.