Statistik mit Red Bull - Pharmaforschung im Klassenraum

Einleitung

In Statistikbüchern „wimmelt" es von Aufgaben, in denen mit Hilfe fingierter Daten Medikamente oder medizinische Behandlungsmethoden auf ihre Wirksamkeit hin „untersucht" werden sollen. Vermutlich gibt es von Flensburg bis Füssen aber keine Schulklasse, die solch eine Untersuchung tatsächlich einmal durchgeführt hat. Der folgende Beitrag möchte „Lust machen" es einmal auszuprobieren. Sie brauchen eine Doppelstunde, eine Kiste Energy-Drink, vielleicht auch eine Packung Traubenzucker … und Statistik bekommt - auch Jahre nach dem Ende der Schulzeit einen ungewohnt süßen - sinnstiftenden - Nachgeschmack.

Welche statistischen Methoden man bei der Auswertung einsetzt (oder noch besser: an diesem Beispiel in sinnstiftendem Kontext erarbeitet), hängt von der Jahrgangsstufe ab, in der man experimentiert. Lassen Sie die Dosen „knacken" und stürzen Sie sich in die Statistik der Pharmaforschung!

1 Die Projektidee

Die Initiative „Hart am Limit" (HALT, http://www.halt-projekt.de) gegen „Alkohol-Kampftrinken" hat sich bundesweit einen Namen gemacht. Warum sollte man nicht versuchen, auch im Mathematikunterricht den negativen Einfluss von Alkohol auf die Reaktionsfähigkeit zu untersuchen? Schließlich sind Bremsweg und Reaktionszeiten (in Abhängigkeit von der Geschwindigkeit) ein beliebtes Thema in den Klassen 8 und 9. Nur „schade", dass Alkoholkonsum in der Schulzeit auf nahe liegende Schwierigkeiten stößt, auch wenn er im Dienste der Wissenschaft erfolgt.
So wurde die Idee geboren, das genaue Gegenteil zu untersuchen: Doping mit Energy-Drinks. Denn offensichtlich glauben nicht nur Formel1-Piloten wie Sebastian Vettel, sondern auch Leistungssportler und Klausuren schreibende Schüler an die beflügelnde Wirkung von Koffeinbrausen. Die Versuchsdurchführung ist denkbar einfach:

2 Versuch (Doppelstunde)

__08.10 Uhr__
die Computer sind hochgefahren und startklar für den ersten Reaktionstest… 200 Daten speichern (christa-1.xls)

__08.25 Uhr__
zwei Esslöffel Traubenzucker mit Energiebrause herunterspülen …Wirkung abwarten, Daten inspizieren

__08.40 Uhr__
zweiter Reaktionstest, 200 Daten speichern (christa-2.xls)

__09.10 Uhr__
dritter Reaktionstest („Langzeitwirkung") 200 Daten speichern (christa-3.xls)

__09.30 Uhr__
Computer sind heruntergefahren

Abb. 1: Das Doping…


Abb. 2: … scheint zu wirken

a) Bitten Sie ihre Schüler, morgens keine Muntermacher zu konsumieren und in „nüchternem" Zustand ihre Reaktionszeiten zu messen, indem sie 200mal mit der Maus auf ein zufällig am Bildschirm auftauchenden Button klicken (das Excel-Reaktionszeiten Messprogramm, Abb. 3, ist im Downloadpool und bei www.riemer-koeln.de erhältlich). Das dauert ca. 10 Minuten. Jeder speichert seine 200 Daten ab. Um anonym zu bleiben, verwendet man am besten einen Codenamen (wie z. B. den Vornamen der Oma). Die Datei heißt dann z. B. Christa-1.xls
b) Anschließend bekommt jeder Versuchsteilnehmer zwei Esslöffel DextroPur Traubenzucker und spült das Pulver mit der Dose Energiebrause hinunter (Abb. 1).
c) Nach 15 Minuten wird die zweite Messreihe aufgenommen und abgespeichert (Christa-2.xls).
d) Nach weiteren 20 Minuten, in denen man die zuvor aufgenommen Daten inspizieren kann, folgt die dritte Messreihe (Christa-3.xls) zur Untersuchung der „Langzeit-Wirkung".

3 Ergebnisse - beschreibende Statistik

Abbildung 6 zeigt ein typisches Resultat für den Schüler mit dem Codenamen Christa. Das arithmetische Mittel der Reaktionszeit sinkt hier tatsächlich nach dem Doping von 1012ms auf 863ms (Verbesserung um 15%) und nach der 20-minütigen Wartepause weiter auf 820ms (Verbesserung um weitere 5%). Die Verbesserung der Reaktionszeit ist auch am Absinken des Medians abzulesen, wobei der Median stets unter dem arithmetischen Mittel liegt. Das liegt daran, dass die glockenförmige Verteilung der Reaktionszeiten (Abb. 5) Ausreißer nach oben („Schlafmütze“) aber nicht nach unten haben kann. Der Median ist im Gegensatz zum arithmetischen Mittel Ausreißern gegenüber unempfindlich. Diese leichte Unsymmetrie spiegelt sich auch in den Boxplots (Abb. 6) wider.

kritische Anmerkung:

Meist kommt der Einwand (Hypothese), dass die Verbesserung der Reaktion möglicherweise nur ein Übungseffekt ist, nicht aber auf der „medizinischen Wirkung“ des Doping beruht. Dieser Einwand wird wie folgt entkräftet: Man zerlegt jede Serie aus 200 Reaktionszeiten in zwei 100er Serien (Christa-1.xls wird in Christa-1a.xls und Christa-1b.xls zerlegt). Man erhält dann drei „Zwillingspärchen“, deren Mittelwerte Abbildung. 7 veranschaulicht. Bei Vorliegen eines Übungseffektes müssten die Mittelwerte der jeweils ersten 100 Reaktionszeiten deutlich unter denjenigen der ersten 100 Reaktionszeiten liegen. Das ist aber nicht der Fall, wie Abbildung. 7 und noch deutlicher Abbildung 8 (15 Schüler zusammen) zeigen.

Info: Doppelt-Blindtest- denn „Der Glaube versetzt Berge“

Professionelle Untersuchungen sind als „Doppelt-Blindtests“ angelegt: Dabei wissen weder Versuchsleiter noch Probanden („beide sind blind“), ob sie den Wirkstoff (hier Koffein/Taurin) oder ein wirkungsloses Placebo erhalten. Damit soll die psychologische Wirkung ausgeschlossen werden, die allein aus „dem Gefühl“ resultiert, ein wirksames „Medikament“ bekommen zu haben. Das auch Placebos wirken können kommt im Sprichwort zum Ausdruck: „Der Glaube versetzt Berge“. Da koffeinfreies RedBull nicht angeboten wird … und koffeinfreie Cola sich von normaler Cola geschmacklich unterscheiden lässt, ist ein „doppelt-blind“ Design hier nicht möglich. Der reale Versuch schärft aber durch Nachdenken über Fehlereinflüsse ein kritisches Bewusstsein für sachgerechte Interpretation von Versuchsergebnissen. Und auch das gehört - neben einer Kenntnis von Verfahren - zur statistischen Allgemeinbildung.

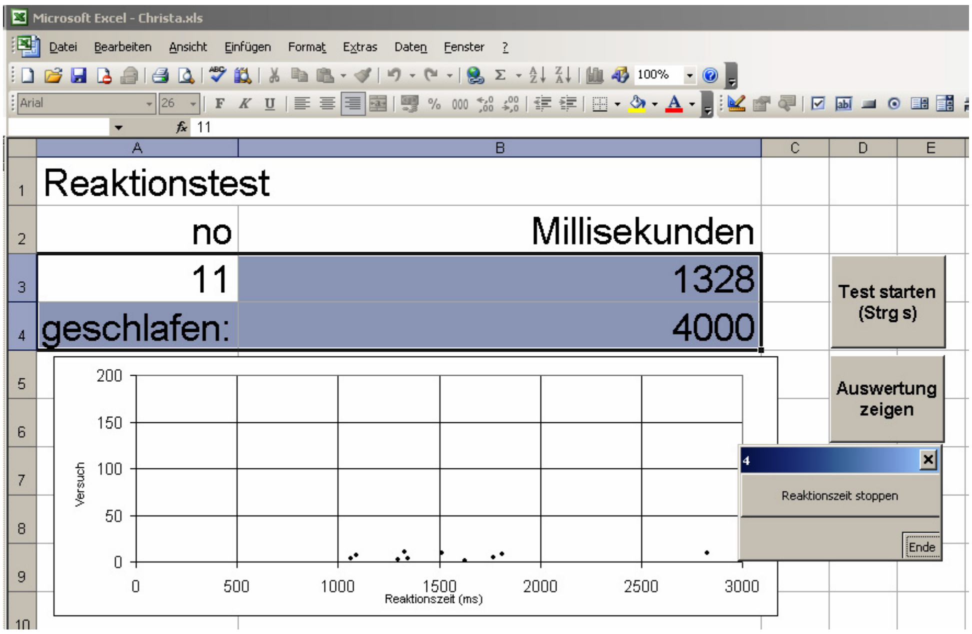


Abb. 3: das Reaktionszeiten-Messprogramm (Excel). Man muss mit der Maus möglichst schnell auf den an einer Zufallsposition erscheinen Knopf „Reaktionszeit stoppen“ klicken. Nach 200 Versuchen beendet man das Experiment durch Klick auf „Ende“ und erhält eine erste statistische Auswertung mit Kenngrößen, Boxplot, Histogramm und Trendanalyse.

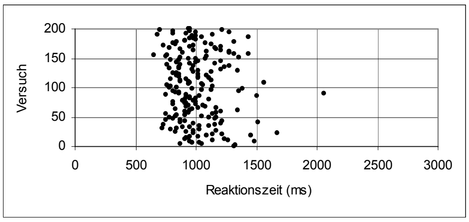


Abb. 4: Während des Tests werden die Reaktionszeiten als Punke einer Wolke, die nach oben hin wächst, sichtbar

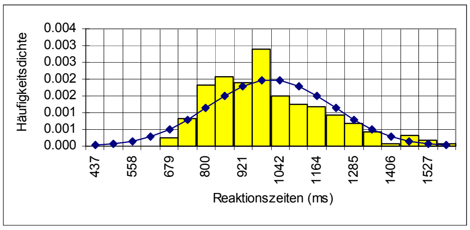
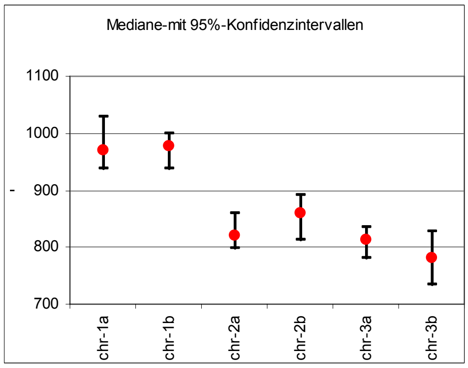


Abb. 5: … und am Ende in der Auswertung als Histogramm (und auch als Boxplot) präsentiert

Ergebnisse des Probanden mit Codename „Christa“ aus drei Serien mit je 200 Reaktionszeiten…

	christa-1	christa-2	christa-3
Maximum	2063	1562	1516
oberes Quartil	1125	957	922
Mittelwert	1012	863	820
Median	969	844	812
unteres Quartil	875	734	688
minimum	656	578	329

werden durch Zerlegung sechs 100er-Serien.



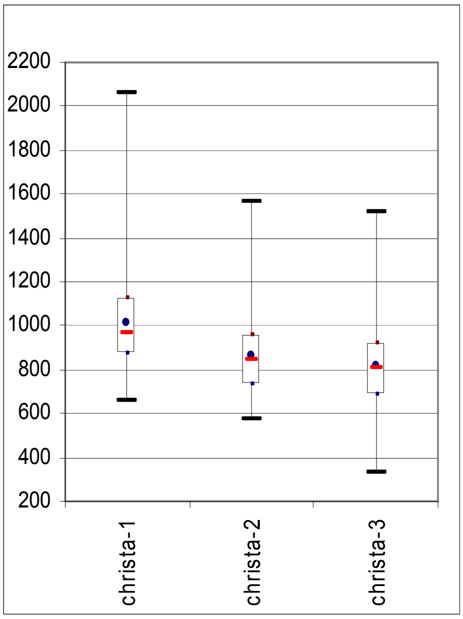


Abb. 6: die Mittelwerte der Reaktionszeiten (blaue Punkte) sinken nach dem Doping deutlich, wobei der Median (roter Strich) fast immer unter dem arithmetischen Mittel liegt

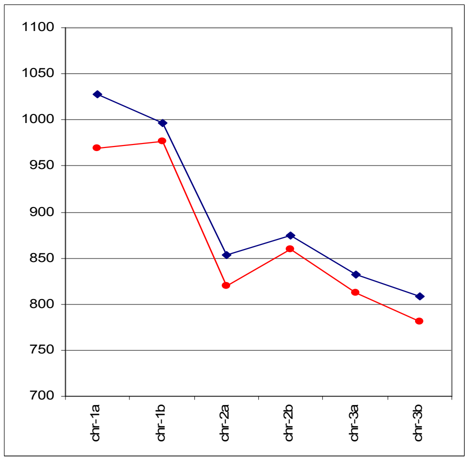


Abb. 7: Median (rot) und arithmetisches Mittel (blau) der Reaktionszeiten. Die drei 200er Serien wurden in 2x100 unterteilt. Man erkennt: Die Verringerung der Reaktionszeiten fußt nicht auf einem Übungseffekt.

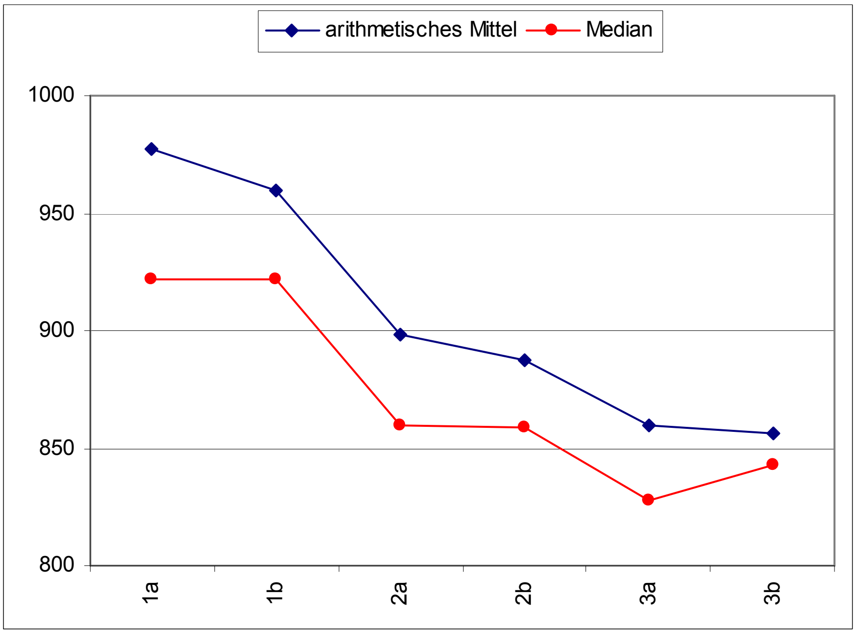


Abb. 8: Gruppenergebnisse - in „Zwillingspaare“ aufgeteilten Reaktionszeiten von 15 Schülern. Jeder der sechs Mediane (rot)/arithmetischen Mittelwerte (blau) basiert auf 1500 Reaktionszeiten.

Ergebnis (beschreibende Statistik)
Durch Doping mit zwei Esslöffeln DextroPur und einer Dose RedBull verbessert sich die Reaktionszeit eines Jugendlichen um durchschnittlich ca. 10%. Diese Verbesserung beruht nicht auf einem Übungseffekt.

4 beurteilende Statistik

In der Sekundarstufe wird man sich die Möglichkeit nicht entgehen lassen, bei der „Doping-Studie“ die Standard - Werkzeuge beurteilender Statistik zu erproben, vielleicht auch „neue“ kennen zu lernen - wie das Konfidenzintervall für den Median (wodurch eine Brücke geschlagen wird zwischen „beschreibender Boxplot-Statistik“ und beurteilender Statistik, den Rangsummentest - oder den t-Test. Auch die Tatsache, dass diese Werkzeuge in Tabellenkalkulationsprogrammen und graphischen Taschenrechnern (bisher meist unentdeckt) schlummern, macht neugierig.

4.1 Vorzeichentest: einfacher geht's nicht

	chr-1	chr-3	>		
median	969	812		Anzahl +	162
1	1313	1063	+	Anzahl -	38
2	969	718	+	Summe	200
3	1328	1047	+	Anteil +	0.81
4	875	641	+	Anteil -	0.19
5	1046	750	+		
195	953	781	+		
196	813	1344	-		
197	703	1297	-		
198	1219	843	+		
199	953	579	+		
200	937	766	+		

Abb. 9 Wirkung der Brause:
Vergleich von Christa-1 und Christa-3 mit Vorzeichentest (200 Datenpaare)

	chr-1a	chr-1b			
	969	976.5		Anzahl +	53
1	1313	781	+	Anzahl -	47
2	969	797	+	Summe	100
3	1328	1140	+	Anteil +	0.53
4	875	766	+	Anteil -	0.47
5	1046	1062	-		
95	1125	953	+		
96	1047	813	+		
97	1391	703	+		
98	968	1219	-		
99	782	953	-		
100	844	937	-		

Abb. 10: Übungseffekt?
Vergleich von Christa-1a und Christa-1b mit Vorzeichentest (100 Datenpaare).

Mit dem Vorzeichentest prüft man, ob zwei Datensätzen die gleiche Wahrscheinlichkeitsverteilung zugrunde liegt wie folgt: Man subtrahiert den Messwert des zweiten (Christa-3) von dem entsprechenden Messwert des ersten Datensatzes (Christa-1) und notiert das Vorzeichen (+/-) der Differenz. Gleiche Reaktionszeiten werden einfach überschlagen. Hätte das Doping keinen Einfluss, wären die Vorzeichen B(n;0,5)-verteilt wie die Seiten einer fairen Münze. Bei dem in Abb. 9 abgebildeten Datensatz mit n=200 Paaren hatte man 162 mal „+“. Mit

$\mu = 0,5 \cdot 200 = 100$ und $\sigma = \sqrt{200 \cdot 0,5 \cdot 0,5} \approx 7,07$ ergibt die Sigmaregel, dass die Anzahl der „+“ mit 95% Wahrscheinlichkeit im Intervall $[\mu - 1,96\sigma; \mu + 1,96\sigma] = [87;113]$ liegen müsste.

Da die beobachtete Anzahl 162 weit außerhalb liegt, kann man die Hypothese, das Doping habe keinen Einfluss auf die Reaktionszeit für „Christa“ hoch signifikant zurückweisen. Der „P-Wert“, also die Wahrscheinlichkeit, dass man bei Gültigkeit der Hypothese einen noch höheren Testwert erhielte ist mit $1 - P(38 < X < 162) = 1 \cdot 10^{-15}$ verschwindend klein.
Wenn man die beiden Datensätze Christa-1a und Christa-1b („nüchtern“) mit je 100 Reaktionszeiten vergleicht, ergibt sich mit 53 mal „+“ kein signifikanter Unterschied. Da man beim Vorzeichentest auf viele Informationen, die im Datensatz stecken, verzichtet, ist er nicht sehr trennscharf: Die Hypothese der Gleichheit beider Verteilungen wird tendenziell zu häufig akzeptiert.

4.2 Test und Vertrauensintervall für den Median (Fortsetzung des Vorzeichentests)

Mit der Idee des Vorzeichentests testet man auch Hypothesen über den Median ν einer Wahrscheinlichkeitsverteilung, bestimmt Konfidenzintervalle für ν … und schlägt so eine Brücke zwischen der beschreibenden und der beurteilenden Statistik.

Der Gedankengang wird am Datensatz Christa-1a vom Umfang n=100 erläutert.
Der aus diesem Datensatz geschätzte Median $\hat{\nu}$ hat hier den Wert 969ms. Natürlich wird dieser Schätzwert von Versuch zu Versuch schwanken und zwar um den theoretischen (aber unbekannten) Median ν der Wahrscheinlichkeitsverteilung. Wir fragen: „Ist die Hypothese ν =930ms mit dem beobachteten Datensatz auf dem 5% Signifikanzniveau vereinbar?“
Antwort: Wenn der Median der theoretischen Reaktionszeitenverteilung den Wert ν =930ms hätte, wäre bei vielen Wiederholungen des Reaktionstests die *Anzahl k* der Reaktionszeiten unter 930ms *B(100;1/2) - binomialverteilt*, denn jede einzelne Reaktionszeit liegt nach Definition mit Wahrscheinlichkeit ½ unter dem angenommen Median. Dann müsste k nach der Sigmaregel mit 95%iger Wahrscheinlichkeit im 1,96σ-Intervall um den Erwartungswert 50 liegen: $k \in [50 - 1,96\sigma; 50 + 1,96\sigma] \approx [40;60]$.
Tatsächlich liegen aber nur 38 Reaktionszeiten unter 930ms und die Hypothese muss (knapp) verworfen werden.

Alle Hypothesen, die man auf diese Weise nicht verwerfen kann, bilden das 95% Konfidenzintervall für ν. Wenn man die 100 Reaktionszeiten aufsteigend sortiert, dann liegt das 95% Konfidenzintervall zwischen der 40ten und der zur 60ten Reaktionszeit. Es gilt: $I_{0,95} = [t_{40}; t_{60}] = [938;1016]$.

Für n=200 (wie z. B. bei dem kompletten Datensatz Chista-1 (mit σ=7,07) müsste für k analog mit ca. 95%iger Wahrscheinlichkeit gelten $k \in [100 - 1,96\sigma; 100 + 1,96\sigma] \approx [86;114]$. Das 95%-Konfidenzintervall für den Median ist also $I_{0,95} = [t_{86}; t_{114}] = [953;1000]$.

Das Ablesen der Grenzen des Kofidenzintervalls aus den sortierten Datenlisten ist mit dem PC kein Problem, man kann sich die Grenzen aber auch ohne vorher zu sortieren als die 0,4 / 0,6-Quantile (n=100) bzw. 0,43 / 0,57 - Quantile (n=200) ausgeben lassen.

Mitunter wird empfohlen, das 95%-Konfidenzintervall des Medians aus den Quartilabstand (Q3-Q1), also der Länge der Box im Boxplot zu schätzen nach der Faustregel: $I_{0,95} \approx [\hat{\nu} - 1,58(Q_3 - Q_1)/\sqrt{n}; \hat{\nu} + 1,58(Q_3 - Q_1)/\sqrt{n}]$ (Sachs S. 228 - Abschnitt [314]).
Diese Faustregel liefert für die 100 Datensätze von Christa-1a $I_{0,95}$=[930;1009] und für die 200 Datensätze von Christa-1 $I_{0,95}$=[941;997] im Vergleich mit den exakten Konfidenzintervallen recht brauchbare Näherungswerte.

4.3 Vierfeldertafeln: Mood's Median-Test

Da Vierfeldertafeln derzeit im Stochastikunterricht sehr populär sind, sei am Rande auch Moods Test auf die Gleichheit zweier Mediane erwähnt: Man fasst die zweimal 200 Reaktionszeiten ohne Doping und mit Doping (Christa-1 und Christa-3)

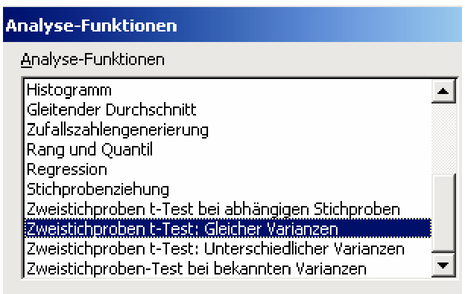
906	chr-1	chr-3	61.45
<med	60	138	198
>med	135	58	193
	195	196	391

Abb. 11:
Vierfeldertafel mit Chi-Quadrat-Testwert

zu einem Satz mit 400 Reaktionszeiten zusammen und bestimmt den gemeinsamen Median (hier 906). Dann zählt man, wie viele Reaktionszeiten von Christa-1 unter und wie viele über dem gemeinsamen Median liegen, ebenso bei Chista-3. Die Ergebnisse werden in den beiden Spalten der Vierfeldertafel notiert. Die Tatsache, dass bei Christa-1 „fast alle" Reaktionszeiten über und bei Christa-3 „fast alle" Reaktionszeiten unter dem gemeinsamen Median liegen spricht für sich. Wer den Unterschied statistisch sauber belegen möchte, verwendet den Chi-Quadrat- Unabhängigkeitstest [Riemer 1990, Lambacher-Schweizer 2011]. Hier ergibt sich der Testwert 61,45, der weit über der kritischen Grenze 3,8 liegt.
Wer - statt nur die Anzahl der Daten über oder unter dem Median zu zählen - auch deren Rangzahlen berücksichtigen, also mehr von der Information nutzten möchte die in den Messwerten steckt, dem sei der Mann-Whitney Test empfohlen. Um ihn zu verstehen, muss man allerdings die Normalverteilung kennen. Der neugierige Leser sei auf [Sachs, S. 380 f] verwiesen.

4.4 t-Test auf Gleichheit zweier Mittelwerte

Der t-Test ist „DAS Referenzwerkzeug", um zu prüfen, ob die Erwartungswerte μ_1, μ_2 zweier normalverteilter Zufallsgrößen (mit gleicher Standardabweichung σ) übereinstimmen. Da man „die volle Information", also die Messwerte selber und nicht nur Vorzeichen von Differenzen oder Rangzahlen nutzt - und auch von der Annahme einer Normalverteilung Gebrauch macht, ist dieser Test trennschärfer als der Vorzeichentest. Zunächst wird am Beispiel von „Christas" Reaktionszeiten erläutert, wie man z. B. in Excel mit diesem Test arbeiten kann.



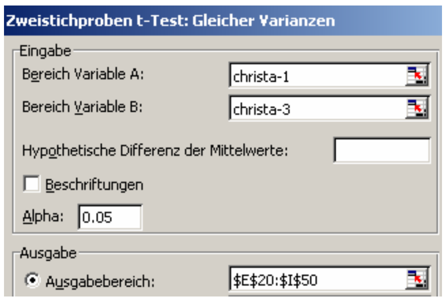


Abb. 12: t-Test als Excel-Analysefunktion

Zweistichproben t-Test unter der Annahme gleicher Varianzen

	Christa-1	*Christa-3*
Mittelwert	1012	820
Varianz	40694	33383
Beobachtungen	200	200
Gepoolte Varianz	37038	
Hypothetische Differenz der Mittelwerte	0	
Freiheitsgrade (df)	398	
t-Statistik	9.96	
P(T<=t) einseitig	2.57E-21	
Kritischer t-Wert bei einseitigem t-Test	1.65	
P(T<=t) zweiseitig	5.13E-21	
Kritischer t-Wert bei zweiseitigem t-Test	1.97	

Zweistichproben t-Test unter der Annahme gleicher Varianzen

	Christa-1a	*Christa-1b*
Mittelwert	1028	996
Varianz	47591	33711
Beobachtungen	100	100
Gepoolte Varianz	40651	
Hypothetische Differenz der Mittelwerte	0	
Freiheitsgrade (df)	198	
t-Statistik	1.10	
P(T<=t) einseitig	1.36E-01	
Kritischer t-Wert bei einseitigem t-Test	1.65	
P(T<=t) zweiseitig	2.73E-01	
Kritischer t-Wert bei zweiseitigem t-Test	1.97	

Abb. 13: t-Test - Ergebnisse

Zunächst wählt man unter Extras - Analysefunktionen die Option Zweistichproben t-Test: gleiche Varianzen (Abb. 12) und gibt im sich öffnenden Fenster die zu vergleichenden Datenbereiche christa-1 (200 Reaktionszeiten nüchtern) und christa-3 (200 Reaktionszeiten - Langzeitdoping) sowie das gewünschte Signifikanzniveau ein. Im Ausgabebereich (Abb. 13) werden die mittleren Reaktionszeiten, deren Varianzen und auch der (bei unterschiedlichen Stichprobenumfängen gewichtete) Mittelwert der Varianzen als „gepoolte Varianz" ausgegeben. Wenn der Testwert („t-Statistik", hier 9,96) über dem kritischen t-Wert (hier 1,97) liegt, muss man die Hypothese auf dem angegebenen Signifikanzniveau (hier 5%) verwerfen. Die Unterschiede zwischen den ungedopten und den gedopten Reaktionszeiten sind beim zweiseitigen Test wegen 9,69>1,97 signifikant. Die Wahrscheinlichkeit, dass man bei Gültigkeit der Hypothese (μ_1=μ_2) einen noch größeren Testwert erhielte - Statistiker bezeichnen diese Wahrscheinlichkeit als P-Wert - ist mit P=2,57*10^{-21} praktisch 0. Er ist deutlich kleiner als der entsprechende P-Wert beim Vorzeichentest. Wenn man zum Vergleich die Mittelwerte der jeweils 100 Reaktionszeiten am Anfang und am Ende der Messung „nüchtern" miteinander vergleicht, ergeben sich dagegen keine signifikanten Unterschiede - genau wie beim Vorzeichentest.

4.5 Hintergrund des t-Tests: Arbeiten mit der Normalverteilung

Man geht von der Modellannahme aus, dass die Reaktionszeiten eines Probanden normal verteilt sind: Ohne Doping nach $N(\mu_1;\sigma)$ und mit Doping nach $N(\mu_2;\sigma)$ mit jeweils der gleichen (aber unbekannten) Standardabweichung σ. Dann ist der Mittelwert aus 200 Reaktionszeiten ohne Doping $N(\mu_1;\frac{\sigma}{\sqrt{200}})$ - und derjenige aus 200 Reaktionszeiten mit Doping $N(\mu_2;\frac{\sigma}{\sqrt{200}})$ - normal verteilt. Die Differenz der Mittelwerte $\bar{x}_2-\bar{x}_1$ ist dann nach $N(\mu_2-\mu_1;\sqrt{\frac{\sigma^2}{200}+\frac{\sigma^2}{200}})=N(\mu_2-\mu_1;\frac{\sigma}{\sqrt{100}})$ - normalverteilt

[Lambacher-Schweizer 2011 S. 406].

Wenn beide Erwartungswerte gleich wären - und davon geht man bei der zu prüfenden Hypothese aus - dann wäre die Differenz $\bar{x}_2-\bar{x}_1$ also $N(0;\frac{\sigma}{\sqrt{100}})$ - normal verteilt. Die normierte Testgröße $t=\frac{(\bar{x}_2-\bar{x}_1)\sqrt{100}}{\sigma}$ wäre N(0;1) standard-normalverteilt und müsste nach der Sigmaregel mit 95%iger Wahrscheinlichkeit zwischen -1,96 und +1,96 (bzw. bei einseitiger Fragestellung mit 95% Wahrscheinlichkeit unter 1,65) liegen.

Da die Standardabweichung σ nicht bekannt ist, schätzt man sie durch die aus beiden Teil-Stichproben gemittelte Stichprobenstandardabweichung $\sigma\approx s=\sqrt{37038}\approx 192{,}45$ und erhält den Testwert $t=\frac{(1012-820)\cdot 10}{192{,}45}=9{,}69$, der in Abb. 13 als „Test-Statistik“ auftaucht.

Dadurch, dass man im Testwert t die theoretische Standardabweichung σ durch die Stichprobenstandardabweichung s ersetzt, ändert sich die Verteilung von t „ein klitzekleines bisschen“. Aus der Standard-Normalverteilung mit Varianz 1 wird die die Student'sche t-Verteilung (hier mit 398 Freiheitsgraden), die auch glockenförmig und symmetrisch zu 0 ist, aber mit 398/396=1,005 eine unmerklich höhere Varianz besitzt. Auch die t-Verteilung mit 198 Freiheitsgraden und der Varianz 198/196=1,007, die beim Vergleich der 100 Reaktionszeiten Christa-1a und Christa-1b zum Einsatz kommt, unterscheidet sich praktisch nicht von der Standard-Normalverteilung. Für praktische Zwecke kann man bei den hier verwendeten Stichprobenumfängen die t-Verteilung bedenkenlos durch die Normalverteilung ersetzen.
Schließlich kommt es beim „Verstehen stochastischer Aussagen“ sehr viel mehr auf die Interpretation von Testgrößen an, als auf technische Details.

Die t-Verteilung wurde von William Sealy Gosset, einem Angestellten der irischen Guinness-Brauerei entdeckt. Da sein Arbeitgeber die Veröffentlichung nicht gestattete, veröffentlichte Gosset sie 1908 unter dem Pseudonym Student. Seitdem ist sie in der Statistik als die Studentsche t-Verteilung bekannt. Die t-Verteilung mit n Freiheitsgraden hat die

Dichte $t_n(x)=\frac{c}{(1+\frac{x^2}{n})^{\frac{n+1}{2}}}$

mit dem Erwartungswert μ=0 und der Standardabweichung $\sigma=\sqrt{\frac{n}{n-2}}$ (n>2)

Je größer n desto ähnlicher wird sie der Standard-Normalverteilung N(0,1). C ist eine Konstante, die dafür sorgt, dass die Fläche unter dem, Funktionsgraphen 1 wird.

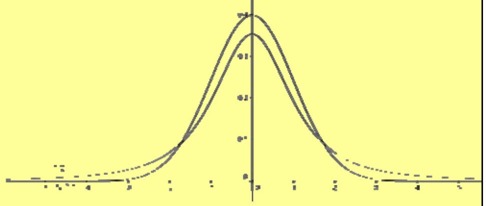


Abb. 14 Dichte der t-Verteilung mit 2 Freiheitsgraden im Vergleich mit der Standard-Normalverteilung

Ergebnis:
Die Verbesserung der Reaktionszeiten durch Doping mit Energiebrausen erweist sich als statistisch *höchst signifikant*. Sie ist damit kein Produkt bloßer Zufallsschwankungen. Das hat sich auch in zwei weiteren Klassen (9 und 11) bestätigt.
Dass sie von manch einem Vereinssportler auch für *relevant* gehalten wird wird, belegen Interviews und auch leere Aludosen in der Umkleide eines Basketballvereins….

Literatur
Sachs, Lothar: Angewandte Statistik: Berlin, Heidelberg - Springer, 9. Aufl. 1999.
Riemer, Wolfgang: Der Chi-Quadrat-Unabhängigkeitstest: MU 5-1990, 50-66.
Lambacher-Schweizer: Qualifikationsphase NRW 735401, Stuttgart - Klett 2011.

Dr. Wolfgang Riemer
Studienseminar (ZfsL) Köln
August-Bebel-Str. 80
50259 Pulheim
w.riemer@arcor.de
www.wiemer-koeln.de

Formel 1 - Sebastian Vettel

fährt für Red Bull … Denn

Red Bull trinken … ist wie Umsteigen vom Fahrrad auf aufs Auto…