

0. Einführender Text: ein quickly prototyping

Mit der zunehmenden Computerunterstützung bei der Anwendung statistischer Verfahren haben sich zugleich die Schwierigkeiten beim Einsatz der Statistik verschoben. Vor der Anwendung von Computerprogrammen bestand die Hauptschwierigkeit darin, für ein bestehendes Problem die entsprechende rechnerische Lösung erst selbst von Hand durchführen zu müssen. Heutzutage gibt es dagegen kaum ein statistisches Problem, wofür nicht bereits eine fertig programmierte Lösung vorliegt. Für den Nutzer von Statistiksoftware stellt sich dafür umso dringender die Frage, welche als Programm bereits realisierte Lösung denn nun für das jeweils vorliegende statistische Problem geeignet sei.

Dieser Verlagerung der Schwierigkeiten beim Einsatz von Statistik muss auch eine Einführung in die Statistik Rechnung tragen: Nicht die Vermittlung einzelner Rechenschritte hat im Vordergrund zu stehen (auch keine formalen Ableitungsregeln) - Statistik ist kein Flickwerk arithmetischer Details -, sondern die Vermittlung der *theoretischen Zusammenhänge*. Dabei sind insbesondere folgende Fragen zu berücksichtigen: Welche Wahrscheinlichkeitsschlüsse liegen prinzipiell einer statistischen Analyse zugrunde? Welche statistischen Verfahren sind bei welchen empirisch erhobenen Daten anzuwenden?

Die *schließende* Statistik oder auch Inferenzstatistik ist Thema dieses Sommersemesters. Sie baut auf die im letzten Semester behandelte deskriptive Statistik auf. Während letztere sich auf eine Zusammenfassung und Darstellung der empirisch erhobenen Daten beschränkt, geht es in ersterer um die Frage, mit welcher Wahrscheinlichkeit die untersuchten Daten mit einer aufgrund einer Theorie formulierten Hypothese verträglich sind.

Aber auch die Erarbeitung der inneren Logik statistischer Entscheidungsfindung kann nicht ohne die Vermittlung mathematischer Detailkenntnisse erfolgen. Um nun nicht vor lauter Bäumen den Wald aus den Blick zu verlieren, sei gleich am Beginn der eigentliche Hauptzweck an zwei prototypischen Beispielen - in freilich grob vereinfachender Form - vorweggenommen. Eine solche Vorgehensweise wird im EDV-Bereich gelegentlich als "quickly prototyping" bezeichnet. Darunter versteht man die Erprobung eines Großprojekts am Beispiel eines Miniaturprojekts, mit dessen Hilfe Chancen und Risiken des Großprojekts ohne großen Zeitaufwand abgeschätzt werden kann. Unser Großprojekt ist im konkreten Fall die Einführung in die Inferenzstatistik. Die beiden folgenden prototypischen Beispiele sind unsere - aus Gründen der Didaktik und Einfachheit - stilisierten Miniaturprojekte, an denen grob abgeschätzt werden soll, worauf die Inferenzstatistik schlussendlich abzielt.

1. Beispiel: der Falschspieler

Man stelle sich einen gewöhnlichen Würfel vor, wie er beispielsweise beim bekannten Spiel Mensch-ärgere-dich-nicht verwendet wird. Ein solcher Würfel hat 6 Augenzahlen. Wurde er nicht manipuliert (indem beispielsweise eine seiner 6 Seiten angebohrt und mit einem Bleigewicht ausgestattet wurde), so hat jede der 6 Augenzahlen die gleiche Chance, bei einmaligem Würfeln aufzutreten. Es gibt keinen physikalischen (oder anderen) Grund, der das Würfeln einer bestimmten Augenzahl *wahrscheinlicher* macht als das einer anderen Zahl. Ganz nebenbei und vorläufig lassen sich an diesem Beispiel zwei zentrale Begriffe der

Wahrscheinlichkeitstheorie einführen: Unter dem Würfeln einer bestimmten Augenzahl - beispielsweise einer "6" - versteht man in der Wahrscheinlichkeitslehre ein *Ereignis*. Ein Ereignis ist, fürs erste betrachtet, nichts anderes als ein bestimmtes Resultat, das wir bei einmaligem Würfeln bekommen können. Beim Würfelspiel können wir sechs verschiedene Resultate (nämlich die Augenzahlen von 1 bis 6) erhalten. Diese sechs verschiedenen Resultate bezeichnet man in der Wahrscheinlichkeitslehre als *Ereignisraum*. Aus diesem Ereignisraum kann bei einmaligem Würfeln immer nur jeweils eines der insgesamt 6 möglichen Ereignisse auftreten (Wir können unmöglich gleichzeitig eine 1 oder eine 6 würfeln). Da alle 6 Ereignisse bei einem normalen Würfel gleich wahrscheinlich sind, beträgt die Wahrscheinlichkeit, eine bestimmte Augenzahl zu würfeln genau ein Sechstel (die Chance, eine 6 zu bekommen, steht 1 : 6).

Bevor wir uns mit einem konkreten Beispiel befassen, stellen wir noch folgende allgemeine Überlegung an: Statt eines Würfels hätten wir ein Zahlenschloss, wie es bei Fahrrädern verwendet wird. Unser Zahlenschloss bestünde im einfachsten Falle aus einem einzigen Ring mit insgesamt sechs Einstellmöglichkeiten. In diesem Falle wäre die Wahrscheinlichkeit, per Zufall die richtige Einstellung zu erraten, gleich einem Sechstel.

Stellen wir uns nun weiters vor, unser Zahlenschloss bestünde aus zwei Ringen und jeder Ring hätte wiederum sechs Wahlmöglichkeiten. Überlegen wir uns, wie viele Einstellmöglichkeiten ein solches Zahlenschloss insgesamt hat. Dreht man den rechten Ring auf eine bestimmte Zahl (beispielsweise auf die Zahl 1), so gibt es 6 verschiedene Möglichkeiten, den linken Ring einzustellen. Bei jeder Einstellung des rechten Ringes können wir also den linken Ring 6 mal verstellen. Nachdem aber auch der rechte Ring insgesamt 6-mal verstellt werden kann, haben wir bei unserem fiktiven Zahlenschloss insgesamt 6×6 verschiedene Einstellmöglichkeiten. Die möglichen Einstellungen reichen von "11", "21", "31", ... bis "66". Dies ergibt 36 Wahlmöglichkeiten. Man kann an diesen Überlegungen ohne Schwierigkeiten erkennen: Die Wahrscheinlichkeit, bei zwei Ringen mit je 6 Einstellmöglichkeiten rein zufällig die richtige Einstellung des Zahlenschlosses zu erraten, beträgt $1 : 6^2$ (= 1 : 36). Bei drei Ringen beträgt diese Wahrscheinlichkeit $1 : 6^3$.

Übertragen wir diese Überlegungen auf das Beispiel eines Würfels, so ergibt sich daraus das folgende. Wollen wir beispielsweise wissen, wie wahrscheinlich es ist, 10-mal hintereinander eine 6 zu würfeln, dann brauchen wir uns lediglich ein Zahlenschloss mit 10 Ringen vorzustellen, wobei jeder Ring 6 Einstellmöglichkeiten hat. Diese Wahrscheinlichkeit beträgt $1 : 6^{10}$!

Nun ist diese Wahrscheinlichkeit äußerst gering. Dennoch haben wir mit der Berechnung dieser Wahrscheinlichkeit ($1 : 6^{10}$) *genau* jene Wahrscheinlichkeit ausgerechnet, mit einem fairen (nicht manipulierten) Würfel 10 mal hintereinander die Augenzahl 6 zu würfeln. Theoretisch lässt sich auf die gleiche Weise auch die Wahrscheinlichkeit berechnen, 1000 mal hintereinander die Augenzahl 6 zu bekommen ($1 : 6^{1000}$).

Kommen wir vor dem Hintergrund dieser abstrakten Überlegungen zu einem konkreten Beispiel. Mit dem Beispiel soll die zentrale Logik von Wahrscheinlichkeitsschlüssen, wie sie in der Inferenzstatistik generell zum Tragen kommt, auf einen möglichst knappen Punkt gebracht werden.

Nehmen wir an, wir wären Besitzer eines Spielsalons und folgender, freilich äußerst unwahrscheinlicher Fall, würde eintreten. Einer der anwesenden Spieler würde tatsächlich 1000-mal hintereinander eine 6 würfeln. Stellen wir uns weiters vor, wir ließen sowohl den

Würfel genau untersuchen und würden auch den Spieler unter Beobachtung stellen, ohne dabei irgendwelche Unregelmäßigkeiten feststellen zu können. Vermutlich würde man den Spieler mit Lokalverbot belegen, und zwar aus dem einfachen Grund, weil die Wahrscheinlichkeit, ohne besondere Tricks (Manipulieren des Würfels und ähnliches) 1000-mal eine 6 zu würfeln, verschwindend gering ist. Der Hinauswurf des Spielers erfolgte also nicht deshalb, weil er etwa beim Falschspiel ertappt wurde, sondern aufgrund von Wahrscheinlichkeitsüberlegungen. Dennoch spräche bei unserer - auf einem Wahrscheinlichkeitsschluss beruhenden - Entscheidung (Hinauswurf des Spielers) auch etwas *zugunsten* des Spielers, nämlich eben die Wahrscheinlichkeit $1 : 6^{1000}$! Diese Zahl ist aber so klein, dass wir uns - obwohl gegen den Spieler keine konkreten Verdachtsmomente vorliegen - *gegen* die Unschuld des Spielers entscheiden würden.

Dieses künstliche Beispiel enthält - wenngleich in zugespitzter und äußerst knapper Form - die wesentlichen Grundprinzipien, die in der statistischen Entscheidungsfindung zum Tragen kommen. Man kann sich die schließende Statistik als eine Art *Indizienprozess* vorstellen, bei dem zunächst die Wahrscheinlichkeit für die *Unschuldsbehauptung* ausgerechnet wird. Wir nehmen in dem Indizienprozess zunächst also die Position des Verteidigers ein. Dieser geht von der Unschuld des Angeklagten aus. In unserem konkreten Beispiel (dem potentiellen Falschspieler) gehen wir davon aus, dass alle Verdachtsmomente rein zufälliger Natur sind. Diese Position des Verteidigers bezeichnet man in der Statistik als die *Nullhypothese*. Im Unterschied dazu geht die *Alternativhypothese* davon aus, dass die Verdachtsmomente zu stark sind, um noch sinnvoll von einem Zufall sprechen zu können.

Was immer zuerst berechnet wird, ist die Wahrscheinlichkeit für die Unschuldsbehauptung, also für die Nullhypothese. In unserem konkreten Beispiel ist dies die Wahrscheinlichkeit, bei einem nicht manipulierten Würfel und ohne Tricks 1000 mal eine 6 zu würfeln, wie bereits mehrfach erwähnt, $1 : 6^{1000}$.

Ist diese Wahrscheinlichkeit (nochmals gesagt: die Wahrscheinlichkeit für die Nullhypothese) sehr gering, so entscheidet man sich *gegen* die Nullhypothese. In unserem konkreten Beispiel bedeutet diese Entscheidung, dass wir den Spieler mit Lokalverbot belegen. Nachdem aber immer noch eine, wenngleich geringe, Wahrscheinlichkeit zugunsten der Nullhypothese spricht (letztere ist in unserem Indizienprozess ja nur eine *Hypothese*, deren Wahrscheinlichkeit wir berechnen), ist diese Wahrscheinlichkeit zugleich jene Wahrscheinlichkeit, mit der wir uns *irrtümlicherweise* gegen die Nullhypothese entscheiden. Es besteht ja immerhin die (allerdings sehr schwache) Chance, dass jemand zufällig 1000-mal eine 6 würfelt. Die Wahrscheinlichkeit, aufgrund derer wir uns gegen die Nullhypothese entscheiden, wird in der Statistik daher auch als *Irrtumswahrscheinlichkeit* bezeichnet.

2. Beispiel: der Einfluß eines bestimmten Medikamentes auf das Reaktionsvermögen im Straßenverkehr

In einer psychologischen Untersuchung interessiert uns in der Regel nicht das Ergebnis eines Würfelspiels, unsere empirischen Daten sind dahingegen beobachtete Eigenschaften von

Versuchspersonen, die auf ganz bestimmte, im Rahmen einer Theorie aufgestellte Hypothesen hin untersucht werden.

Im folgenden soll gezeigt werden, wie die - am Beispiel eines Würfelspiels demonstrierten - Wahrscheinlichkeitsaussagen auch auf typische Fragestellungen übertragen werden können, wie wir sie in der Psychologie antreffen.

Betrachten wir hierzu das folgende Beispiel: In einem Experiment soll herausgefunden werden, ob ein bestimmtes Medikament das Reaktionsvermögen im Straßenverkehr beeinträchtigt. Eine mögliche, wenngleich wenig Zielführende Methode, um dies festzustellen, bestünde darin, bei *allen* Verkehrsteilnehmern nachstehenden Test durchzuführen. Zuerst messen wir das Reaktionsvermögen (in ms) der Verkehrsteilnehmer *vor* Verabreichung des Medikaments und später, *nach* Verabreichung des Medikaments, wiederholen wir die Messung. Unterscheiden sich die Reaktionszeiten vor und nach Verabreichung des Medikaments voneinander, so könnte man sagen: Das Medikament beeinflusst das Verhalten im Straßenverkehr.

Offensichtlich ist ein derartiges Experiment in der Praxis nicht durchführbar. Wir können unmöglich *alle* Verkehrsteilnehmer für ein bestimmtes Experiment erfassen. Tatsächlich können wir immer nur eine *Zufallsstichprobe* der Verkehrsteilnehmer erheben. Statt also alle Verkehrsteilnehmer zu untersuchen, greifen wir rein zufällig eine kleine Anzahl von Testpersonen (beispielsweise 30) aus allen potentiellen Verkehrsteilnehmern heraus und führen mit diesen unser Experiment durch. Bei 30 Personen messen wir also die Reaktionszeit vor (Zeitpunkt T1) und nach (Zeitpunkt T2) Verabreichung des Medikaments. Die Ergebnisse dieser beiden Messungen über alle Versuchspersonen lassen sich etwa wie folgt zusammenfassen:

	T1	T2
Hans Meier	15	30
Franz Huber	20	45
.	.	.
.	.	.
.	.	.

usw.

Vergleicht man die beiden Zahlenkolonnen, so stellt sich fürs erste die Frage, wie wir zunächst nur für diese eine Stichprobe, bestehend aus den 30 Probanden, feststellen können, ob zwischen den beiden Reaktionszeiten überhaupt ein Unterschied besteht und, falls sich ein solcher ermitteln lässt, welcher Größenordnung er angehört. Wir brauchen also einen *statistischen Kennwert*, um für diese eine Stichprobe generell den Unterschied in den beiden Reaktionszeiten abschätzen zu können. Unser Interesse richtet sich schließlich nicht auf Unterschiede von Reaktionszeiten zwischen den einzelnen Probanden der Stichprobe. So wollen wir nicht wissen, ob beispielsweise Hans Meier verschiedene Reaktionszeiten hat, wir wollen dahingegen wissen, ob für alle Personen in der Stichprobe ein gemeinsames Unterschiedsmaß berechnet werden kann. Die hier interessierende Frage ist schließlich: Beeinflusst das Medikament die Reaktionszeit im Straßenverkehr, und zwar ungeachtet vom speziellen Verhalten einer einzelnen Person?

Stellen wir uns folgende Vorgehensweise vor. Zuerst berechnen wir getrennt für jeden einzelnen Probanden den Unterschied in der Reaktionszeit vor und nach Verabreichung des Medikaments. Für jedes Messwertpaar bekommen wir auf diese Weise eine bestimmte Differenz. Diese Differenz beträgt z.B. bei Hans Meier -15 ms, bei Franz Huber -25 ms usw. Bei 30 Versuchspersonen erhalten wir auf diese Weise 30 verschiedene Differenzen. Nachdem wir nicht an den Unterschieden in den Reaktionszeiten der einzelnen Testpersonen interessiert sind, berechnen wir anschließend den durchschnittlichen Wert der Differenzen, den so genannten Mittelwert. Dieser ist die Summe der einzelnen Differenzen dividiert durch die Anzahl der Versuchspersonen. Im Kontext der folgenden Überlegung gehen wir von einem angenommenen Mittelwert der Differenzen von ca. -18 ms aus.

Laut Annahme handelt es sich bei den 30 untersuchten Personen nun aber lediglich um eine *Stichprobe* aus allen potentiell vorhandenen Verkehrsteilnehmern. Wir können daher von vornherein nicht wissen, ob sich bei *allen* Verkehrsteilnehmern ebenfalls ein derartiger Unterschied im Reaktionsvermögen feststellen lässt oder ob unsere durchschnittliche Abweichung von -18 ms nur auf Zufallsschwankungen zurückzuführen ist.

Um diese Frage zu klären, gehen wir von ganz analogen Überlegungen aus, wie sie bereits beim Falschspieler beschrieben wurden. Zunächst gehen wir von der Unschuldbehauptung aus und dann berechnen wir die Wahrscheinlichkeit für die Unschuldbehauptung.

Im vorliegenden Beispiel stellen sich somit zwei Fragen: 1) Was ist hier unter der Unschuldbehauptung zu verstehen? 2) Wie können wir die Wahrscheinlichkeit für diese Unschuldbehauptung berechnen?

Zur ersten Frage: Unter der Unschuldbehauptung verstehen wir im vorliegenden Fall die Annahme, dass der berechnete Unterschied in den Reaktionszeiten lediglich auf einen Zufall bei der Auswahl der Versuchspersonen zurückzuführen ist. Diese Annahme bezeichnet man, wie bereits erwähnt, auch als die *Nullhypothese*. Dieser Hypothese zufolge hat das Medikament *generell* keinen Einfluss auf die Reaktionszeit der Verkehrsteilnehmer. Würde man also *alle* Verkehrsteilnehmer in Betracht ziehen können, so erhielte man eine durchschnittliche Differenz von 0 ms. Lediglich unsere per Zufall ausgewählten Personen weisen eine derartige durchschnittliche Differenz auf.

Zur zweiten Frage: Um nun die Wahrscheinlichkeit für die Nullhypothese berechnen zu können, gehen wir von folgender Überlegung aus. Nehmen wir an, tausende von Psychologen würden das gleiche Experiment durchführen. Um den Einfluss eines bestimmten Medikaments auf das Reaktionsvermögen zu ermitteln, würden sie jeweils 30 zufällig ausgewählte Personen untersuchen. Jede dieser psychologischen Untersuchungen würde also eine andere Zufallsstichprobe verwenden und käme daher auch rein zufällig zu einer etwas anderen durchschnittlichen Abweichung in den Reaktionszeiten (eine Untersuchung käme zu einer durchschnittlichen Abweichung von -5 ms, eine andere von -2 ms, eine weitere von +5 ms usw.).

Denken wir nun daran, dass wir bei unserer Fragestellung zunächst von der *Nullhypothese* ausgehen. Gemäß dieser Hypothese hat das Medikament, bezogen auf alle Verkehrsteilnehmer, keinen Einfluss auf das Reaktionsvermögen. Aus diesem Grunde dürfte sich in der Gesamtpopulation aller Verkehrsteilnehmer auch kein Unterschied in den Reaktionszeiten feststellen lassen. Folgen wir der Nullhypothese, so sind die in den einzelnen psychologischen Untersuchungen errechneten durchschnittlichen Abweichungen lediglich auf Zufallsschwankungen zurückzuführen. Für alle Verkehrsteilnehmer erwarten wir, wie bereits

erwähnt, eine Differenz von 0 ms. Ziehen wir aus dieser Gesamtpopulation nun Zufallsstichproben, so werden die durchschnittlichen Differenzen der Reaktionszeiten in diesen Stichproben nicht exakt 0 ergeben, sondern mehr oder weniger von dieser Differenz abweichen. Unter Geltung der Nullhypothese (die Differenz in der Gesamtpopulation ist 0) werden größere Abweichungen von 0 eher selten und kleinere Abweichungen eher häufig zu erwarten sein.

In unserer angenommenen Stichprobe haben wir nun tatsächlich eine durchschnittliche Abweichung von - 18 ms errechnet. Ein solches Resultat dürfte unter Geltung der Nullhypothese eher selten vorkommen - vielleicht mit einer ebenso geringen Häufigkeit wie jene, bei einem Würfelexperiment 1000-mal hintereinander eine 6 zu erhalten.

Um herauszufinden, wie sich unser tatsächliches empirisches Ergebnis - eine durchschnittliche Abweichung von - 18 ms in der Stichprobe - mit der Nullhypothese (wir erwarten in der Population eine durchschnittliche Abweichung von 0) verträgt, sind wir an der *Wahrscheinlichkeit* interessiert, unter Geltung der Nullhypothese rein zufällig ein derartiges Ergebnis zu bekommen. Um diese Wahrscheinlichkeit berechnen zu können, muss die *Verteilung* der durchschnittlichen Abweichungen aller Zufallsstichproben unter Geltung der Nullhypothese bekannt sein. Denn nur in diesem Falle können wir die Frage, wie häufig bzw. selten die durchschnittliche Abweichung einer ganz bestimmten Stichprobe vorkommt, quantifizieren. Unter ganz bestimmten theoretischen Voraussetzungen (Annahmen bzw. Hypothesen) in der Inferenzstatistik (die sich erst später genau darstellen lassen) läßt sich diese Verteilung aller durchschnittlichen Abweichungen mathematisch exakt beschreiben. Da es in diesem einführenden Teil nur darum geht, das wesentliche Prinzip des inferenzstatistischen Schließens prototypisch vorzustellen, gehen wir davon aus, diese Verteilung sei bekannt. Unter dieser Annahme könnten wir die Wahrscheinlichkeit berechnen, unter Geltung der Nullhypothese eine durchschnittliche Abweichung von - 18 ms zu erhalten. Die konkrete Durchführung dieser Berechnung erfordert freilich genaue Kenntnisse der Grundlagen der Inferenzstatistik, die wir uns erst schrittweise aneignen können.

Um diese Wahrscheinlichkeit dennoch an dieser Stelle möglichst praxisnahe darzustellen, wenden wir als Behelfsmittel ein Computerprogramm an, das uns diese Wahrscheinlichkeit als bereits fertiges Resultat ausgibt. Aber auch ein derartiges Computerprogramm liefert uns Ergebnisse, deren Interpretation erst nach dem Studium der Inferenzstatistik möglich sind. Um dieses Problem zu umgehen, müssen wir die Ergebnisse fürs erste unter einem ganz bestimmten Blickwinkel interpretieren. Am besten versetzen wir uns bei der Lektüre der folgenden Ausgabe des Computerprogramms SPSS in die Rolle einer Person, die an einer Swatchuhr, an der neben Stunden- und Minutenzeiger das ganze Räderwerk sichtbar ist, die Zeit abliest. Ähnlich wie wir uns im Falle der Swatchuhr beim Ablesen der Zeit nicht um das Räderwerk kümmern, konzentrieren wir uns im nachstehenden Ausdruck lediglich auf die Angabe der Wahrscheinlichkeit. Alle anderen Angaben sind im Kontext dieser einführenden Darstellung ein zu vernachlässigender Hintergrund.

Nun zu dem Beispiel:

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
V1	Reaktionszeit vor Medikament 30	,114	,550	22,4333	7,229	1,320
V2	Reaktionszeit nach Medikament			40,0333	8,385	1,531

Paired Differences						
Mean	SD	SE of Mean	t-value	df	2-tail Sig	
-17,6000	10,431	1,904	-9,24	29	,000	
95% CI (-21,495; -13,705)						

In dem angeführten Beispiel wurde bei 30 Probanden die durchschnittliche Abweichung in der Reaktionszeit berechnet. Diese beträgt hier -17,6. Die Wahrscheinlichkeit, eine derartige durchschnittliche Abweichung unter Geltung der Nullhypothese zu erhalten, wird von SPSS mit 0.000 angegeben (tatsächlich ist der Wert nicht exakt gleich Null; Ziffern, die aber erst in weiteren Nachkommastellen als den angegebenen aufscheinen, werden im Ausdruck jedoch abgeschnitten). Um dieses Ergebnis interpretieren zu können, gehen wir wiederum von ganz analogen Überlegungen aus wie im vorherigen Falle des Falschspielers. Zunächst sollten wir daran denken, dass die Nullhypothese, derzufolge in der Gesamtpopulation kein Unterschied zwischen den beiden Reaktionszeiten besteht, nur eine rein *theoretische Annahme* (eben eine Hypothese) ist, die es zuallererst zu überprüfen gilt. Als empirische Daten für diese Hypothese stehen uns die nach dem Zufallsprinzip ausgewählten 30 Probanden zur Verfügung. Ist die Wahrscheinlichkeit, unter Geltung der Nullhypothese zu einer derartigen durchschnittlichen Abweichung von -17,6 zu kommen, eher gering, so spricht dies *gegen* unsere theoretische Annahme. Man kann dies auch so ausdrücken: Unsere theoretische Annahme wird durch das empirische Resultat nicht unterstützt. Wir werden im vorliegenden konkreten Falle die Nullhypothese verwerfen und davon ausgehen, dass das Medikament tatsächlich einen Einfluss auf das Reaktionsvermögen hat. Nachdem aber diese geringe Wahrscheinlichkeit, aufgrund derer wir die Nullhypothese im vorliegenden Falle verwerfen, zugleich die Wahrscheinlichkeit ist, unter Geltung der Nullhypothese eine derartige Abweichung zu erhalten, wird sie auch als *Irrtumswahrscheinlichkeit* (d.h. als die Wahrscheinlichkeit, mit der wir uns *irrtümlicherweise* gegen die Nullhypothese entscheiden) bezeichnet.

Das angegebene Beispiel enthält, in freilich äußerst knapper Darstellung, die wesentlichen Entscheidungsregeln, wie sie in der schließenden Statistik zur Anwendung kommen. Wie das Ergebnis des Programms SPSS gezeigt hat, ist ein Kernpunkt der statistischen Tests die ausgegebene Wahrscheinlichkeit. Sie ist sozusagen die entscheidende 'Kennzahl' in dem jeweils ausgegebenen Test, auf die wir uns stützen, wenn wir uns für oder gegen die Nullhypothese entscheiden.

Aus dem Beispiel sollte allerdings auch hervorgehen, dass am Anfang jeder statistischen Analyse die *Auswahl* des für die jeweilige Untersuchung angemessenen Tests steht. Diese Auswahl hängt unter anderem von der *Qualität* der Messdaten ab. Wie das Beispiel gezeigt hat, gehen wir zunächst von einem bestimmten *statistischen Kennwert* in der Stichprobe aus (im angegebenen Beispiel war dies die durchschnittliche Abweichung der Reaktionszeiten),

um im Anschluss daran zu fragen, mit welcher Wahrscheinlichkeit dieser Kennwert (unser empirisch gewonnenes Resultat) sich mit einer vorher formulierten theoretischen Annahme (Nullhypothese) verträgt. Je nach Datenqualität lassen sich nun aber verschiedene statistische Kennwerte berechnen. Nicht in jedem Falle ist der Mittelwert ein geeignetes Mittel zur Beschreibung der Daten. Diese Frage wurde in Zusammenhang mit der im letzten Semester behandelten Messtheorie angesprochen.