



„Faszinierend“*. Menschen halten zu lange an gelernten Schemata fest und blenden neues Wissen aus. Heutiges Beispiel ist die festverankerte Bedeutung der statistischen ‚Signifikanz‘ in der Marktforschung.

“What’s wrong with significance testing? Well, among many other things, it does not tell us what we want to know, and, out of desperation, we nevertheless believe in that it does!”¹

Tatsächlich, wo ein starkes Bedürfnis ist, gibt es auch ein Angebot. Jeder Marketingverantwortliche, der auf seine Marktforschungsergebnisse schaut überlegt, welches Ergebnis denn nun reiner Zufall ist und welche bedeutsam. Ich kenne das aus meinen vielen Jahren im Marketing. Das Statistikprogramm SPSS berechnet, welche Werte bei 95 prozentiger Sicherheit ‚signifikant‘ sind. Nur leider ist das nicht die gewünschte Bedeutsamkeit! Viele Experten wissen das, aber haben aufgegeben, gegen das Sicherheitsbedürfnis zu argumentieren.

Sind Sie neugierig geworden? Da ich selbst kein Statistiker bin, versuche ich die komplizierte Materie so bildhaft wie möglich zu beschreiben. Die echten Experten unter uns mögen mir die eine oder andere Vereinfachung verzeihen, die Kernaussagen sind auf jeden Fall *bedeutsam*.

Wie sind eigentlich p-Werte, t-Tests und Wahrscheinlichkeitsberechnungen auf die Welt gekommen? Die Kurzgeschichte könnte so aussehen. Wissenschaftler haben schon vor Jahrhunderten begonnen, Phänomene vor allem in der Natur zu beobachten. Wie sich die Wolken entwickeln, wie sich Pflanzen entwickeln abhängig von klimatischen Einflüssen, wie sich Zellen teilen, auch in der Petrischale. Immer mehr Datenpunkte kamen dazu und den beobachteten Phänomenen war es egal, ob die Wissenschaftler zuschauten oder gerade schliefen. Die Prozesse liefen weiter. Irgendwann wurden Muster erkannt und das versetzte die Forscher in die Lage Modelle zu erstellen und vorherzusagen, wann genau die Sonne aufgehen wird, wo die Wolken sich abregnen werden – abhängig von Bergen und Gewässern, wie schnell sich die Zellen teilen werden – abhängig von Tem-

peraturen und Luftfeuchtigkeit. Und sie konnten die Wahrscheinlichkeit für die richtige Vorhersage für diese natürlichen Phänomene aufstellen. Sie brauchten ab sofort nur relativ kleine Datenreihen und konnten ihre Erkenntnisse auf große Phänomene hochrechnen. Sie lernten, wie groß die Datenreihen mindestens sein mussten, um eine bestimmte Sicherheit für ihre Vorhersagen zu erreichen. Das ist die Geburtsgeschichte der Signifikanzbestimmung – extrem verkürzt. Und die Geschichte geht weiter.

Nachdem die Naturwissenschaftler nun ein so treffliches Modell in den Händen hielten, wollten die Verhaltenswissenschaftler nicht zurückstehen und wandten dieses Modell auch auf ihren Datensätzen an. Obwohl menschliches Verhalten nicht nur kompliziert sondern zusätzlich komplex ist. Denn es folgt nicht biologisch oder physikalisch festgelegten Programmen mit geringen Abweichungen. Und enthält auch keine einfach bestimmbaren Messgrößen. Einstellungen, Emotionen und Erinnerungen sind keine determinierten Größen und nicht konstant, sondern unterliegen den vielfältigsten situativen Einflüssen innerhalb und außerhalb des Menschen. Der glühende Ärger über eine Beleidigung beeinflusst uns unbewusst die folgenden Stunden. Zudem haben Menschen oft keinen guten Einblick in ihre Gefühle und Verhaltensweisen und geben damit verzerrte Antworten. Schlimmer noch, während in der Natur die Messergebnisse nicht durch die Beobachtung oder das Material des Messapparates beeinflusst werden, ist das bei Menschen sehr wohl der Fall. Wie man die Fragen formuliert, wie man die Skalen festlegt, wie sympathisch der Marktforscher ist, auf welchem Weg man die Befragung erhebt, Studio-Telefon-Online, alles hat einen Einfluss auf das Messergebnis.²



In der Marktforschungspraxis kommt eine weitere Verschärfung dazu: Im Gegensatz zu den Naturwissenschaftlern erheben wir meist keine Zufallsstichproben. Denn sinnvoller Weise konzentrieren wir uns auf die interessierenden Zielgruppen. Damit haben wir aber keine Zufallsstichprobe mehr und damit auch keinen Zufallsfehler! Trotzdem möchte sich verständlicher Weise der Marketingverantwortliche sicher sein, dass er ein relevantes Ergebnis erkennt – und schaut doch auf die Signifikanzen. „Ist der Schritt von einem Mittelwert von 3,6 auf 3,8 signifikant?“ Eine sehr vernünftige Frage, aber die Antwort prüft eine andere Frage. Vereinfacht wird geklärt, ob Sie mit einer kleinen oder großen Stichprobe gearbeitet haben. Bei 100 Fällen ist die Antwort vermutlich nein, bei 1000 Fällen fast mit Sicherheit ja. Das führt dann dazu, dass in einem Tabellenband einer großen Stichprobe sehr viele Daten als ‚signifikant‘ markiert werden. Aber das ist nicht das gleiche, wie Bedeutsamkeit für das Verständnis von Verbraucher und Marke! Die meisten dieser als signifikant markierten Messgrößen haben keine Bedeutung für die Markenarbeit!

Der Signifikanztest sagt uns also vor allem, ob wir viele oder wenige Befragte hatten. Großartig, das wussten wir vorher. Wo stecken aber in den Daten die für das wirtschaftliche Ergebnis bedeutsamen Unterschiede/Ergebnisse? Was wir uns fragen sollten, ist: Wie groß ist der *Effekt*, mit dem wir es zu tun haben? Das aber misst Signifikanz nicht. Ein besseres statistisches Maß ist die Methode der Effektstärkemessung von Cohen (1988)^{1,7}. Allerdings kann man dieser Methode vorwerfen, dass sie Stichprobengrößen nicht berücksichtigt. (Größere Stichproben reduzieren weißes Rauschen)

Es geht einfacher. Die kanadische Statistikbloggerin Annie Pettit (Love Stats Blog und Chefreditorin der kanadischen Marktforschungszeitschrift)³ empfiehlt, die gewünschte Effektstärke auf den wichtigen Dimensionen selbst festzulegen, im Sinne eines bedeutsamen Action Standards: ‚Wir führen das Produkt nur ein, wenn die Hinwendung in der Kernzielgruppe von 3,6 auf mindestens 3,9 steigt.‘³

Verlässt man sich aber auf die immer wieder ausgewiesene Signifikanz, dann werden bei großen Stichproben marginale Unterschiede, die „im realen Leben“ keinerlei Bedeutung haben, als signifikant/bedeutsam angesehen. Und bedeutsame Unterschiede in Studien mit wenigen Befragten als irrelevant abgetan. Für die gewünschte Sicherheit durch die Signifikanzbestimmung bezahlt man im Marketing öfters einen hohen Preis.

Die Kritik an der Verwendung der Signifikanztests außerhalb der Naturwissenschaften lässt inzwischen nichts mehr an Deutlichkeit zu wünschen übrig. So schreibt Deirdre McCloskey (übersetzt von mir) „Der Fortschritt der Ökonomiewissenschaften ist durch die verbreitete Praxis der Signifikanztests stark beschädigt worden. Man kann nichts von dem glauben, was da heraus kommt. Nicht ein Wort. (!) Es ist alles Unsinn, weshalb zukünftige Generationen von Ökonomen alle (Experimente) noch einmal machen werden müssen. Das meiste, was in den besten Journalen der Ökonomie veröffentlicht wird, ist unwissenschaftlicher Blödsinn. Ich finde das unsagbar traurig. Alle meine Freunde, meine lieben, lieben Freunde in der Ökonomie haben ihre Zeit verschwendet.“⁴ Oder nochmals deutlicher in ihrem letzten Buch mit Stephen Ziliak:⁵ „If null-hypothesis significance testing is as idiotic as we and other critics have so long believed, how on earth has it survived?“

John Campbell, der heute als Psychologe an der University of Minnesota in Minneapolis forscht, beklagte die Zustände schon 1982, als er Editor des "Journals of Applied Psychology" war: „Es ist praktisch unmöglich, die Autoren von ihren p-Werten abzubringen. Und je mehr Nullen nach dem Komma stehen, umso unerbittlicher hängen sie daran.“⁶

Noch immer werden Signifikanztests an den verhaltenswissenschaftlichen Unis gelehrt, Studienergebnisse so interpretiert und in renommierten Fachjournalen so wiedergegeben. Oder wie Max Planck gesagt hat, neues Wissen setzt sich nicht dadurch durch, dass es das alte Wissen verdrängt,



sondern dadurch, dass das alte Wissen mit seinen Besitzern ausstirbt.

Kein schöner Gedanke. Eine aktuelle Parallele: 40 Jahre wurde Cholesterin in Lebensmitteln verteuert, Eigelb für die Verstopfung unserer Arterien dämonisiert. Trotz vieler gegenläufiger Daten. An Unis wird es immer noch so gelehrt. Erst jetzt hat sich die amerikanische Food and Drug Administration durchgerungen, den Mythos der Beeinflussung des Cholesterinspiegels durch Cholesterin aus der Nahrung als unhaltbar festzustellen. Wie lange wird es dauern, bis die Bevölkerung und die Ärzte das alte Wissen überprägt haben? Nach einer amerikanischen Studie dauert es im Mittel 17 Jahre (!) bis neue wissenschaftliche Erkenntnisse bei den meisten behandelnden Ärzten ankommen.

Lassen Sie uns stärker auf unsere Expertise zum menschlichen Verhalten und dem Markenerleben besinnen und uns nicht von der Stichprobengröße abhängigen ‚Signifikanz‘ auf falsche Fährten locken. Schließlich müssen wir ja nicht der Lernkurve von Max Planck folgen.

*Mr. Spock. Leider verstarb unser Vorzeige-Homo oeconomicus und Vulkanier Leonard Nimoy am 27.02.2015 im Alter von 83 Jahren an COPD.

Literatur:

1. Cohen, J (1992). "A power primer". Psychological Bulletin 112 (1): 155–159.
2. <https://lovestats.wordpress.com/tag/significance-testing/>
3. <https://lovestats.wordpress.com/tag/significance-testing/>
4. McCloskey, D. (2002): The Secret Sins of Economics, Chicago, IL.
5. Ziliak, S. and McCloskey, D.N. (2008): The cult of statistical significance: how the standard error costs us jobs, justice, and lives. Ann Arbor (MI): The University of Michigan Press, xxiii, 322 pp.
6. Campbell, John P. (1982): Editorial: Some Remarks From the Outgoing Editor. Journal of Applied Psychology, Vol 67(6), Dec, 691-700.
7. <http://www.psychometrica.de/effektstaerke.html>.



**Feedback, Anregungen, Kritik
zu diesem Artikel an:**
<mailto:braincandy@ka-brandresearch.com>

Der Autor

Ralph Ohnemus, CEO. Seit 2001 Vorstand und Hauptanteilseigner von K&A BrandResearch. Vorher war er 15 Jahre Kunde von K&A BrandResearch. Nationale und internationale Marketing- und Vertriebserfahrung in Senior Management Positionen, darunter FMCG, Mode, Medien und Telekommunikation – zuletzt als SVP Consumer Sales verantwortlich für Marketing, Vertrieb und Filialketten bei Viag Interkom O₂.

Kontakt: <mailto:r.ohnemus@ka-brandresearch.com>

