



Statistische Verfahren zur Minimierung der Tierzahlen in der Antragserstellung

Matthias Schmidt

RUHR-UNIVERSITÄT BOCHUM

Statistische Verfahren zur Minimierung der Tierzahlen

Matthias Schmidt

Obesity Surgery 31, 2021

Obesity Surgery (2021) 31:437–438
<https://doi.org/10.1007/s11695-020-04470-y>



BRIEF COMMUNICATION



Lies, Damn Lies, and Bad Statistics?

Italo Braghetto¹ · Manuel Figueroa¹

Published online: 24 October 2020
 © Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

A wide spectrum of research such as experimental, randomized trials, cohort or epidemiological studies, technical or control case reports, systematic reviews, and meta-analyses has resulted in a huge amount of publications. These studies and publications may be subject to errors due to poor application of statistical tests, which can lead to misinformation, misinterpretation, and erroneous conclusions, sometimes even considered as lies. In this article, some ideas about this issue are discussed in order to adopt new directions in the future and thus avoid lies and bad statistics.

Keywords Publications · Statistic

A wide spectrum of research such as experimental, randomized trials, cohort or epidemiological studies, technical reports, or control/case reports, systematic reviews, and meta-analyses results in a huge amount of publications that require statistical support in order to be validated. Different requirements such as materials, number of patients included, methodology, and of course, statistical analyses are needed in order to achieve the desired results. Thus, knowledge, expertise, capability, and availability to perform these statistical analyses are needed. However, these are not always readily available and hence, we must ask for help.

Surgeons have ideas, experience, knowledge, casuistry, and frequently, new questions. These questions lead to new hypotheses, which must then be proved through different methods of investigation. Currently, there is an enormous quantity of publications and a hypertrophic production of papers due to the “publish or perish” imperative. This occurs because different and controversial points of view arise that are not always deemed acceptable. Some are considered lies and even worse, as “damn lies.” Generally speaking, surgeons are motivated through personal interest to answer questions, share results, and experiences; however, confusion or mistakes do occasionally appear. Although they may be due to poor methodology or poor statistical application, I believe that

surgeons are reliable and responsible and that these do not occur for ill-intended reasons. So, what are the possible causes? What are the underlying problems? Lack of expertise? Lack of knowledge about statistical analyses?

We must trust our colleagues and believe that there are no intended lies, just honest mistakes. Why do these mistakes occur? There are probably many causes such as poor hypotheses, conceptual flaws, poor experimental design, wrong methodology, wrong performance, subject variability, heterogeneity, incorrect statistical planning, or simply a lack of expertise and knowledge. These mistakes could conduce to a lack of reproducibility of the results that have been observed. The solution is to put together a good planning group and a work team. The data management must consist of pure and critical observation of results, strict revision, analysis and correct interpretation. It must be based on scientific interest and always motivated by scientific goals. It must never be based solely on personal interest and it should avoid bias (personal ego: to be the “first”). By following these guidelines, the possibility of data manipulation greatly decreases and the concept of “team work” becomes an essential.

Today, surgeons and clinicians are under heavy pressure to publish, and clinical publications have become fertile grounds for lies. In 2002, a survey revealed that approximately 60% of 192 authors of clinical practice guidelines had financial connections with the companies. Publications promoting new devices and new procedures are often observed and proof of their advantages and results is mandatory. Very good results are almost always published; however, the question arises as to whether or not they are true, because there is a lack of communication of adverse effects or aftermath and complications. There are many examples.

✉ Italo Braghetto
 italobraghetto@gmail.com

¹ Department of Surgery, Hospital “Dr José J. Aguirre”, Faculty of Medicine, University of Chile, Santos Dumont 999, Santiago, Chile

Nature Reviews Neuroscience 14, 2013

ANALYSIS

Power failure: why small sample size undermines the reliability of neuroscience

Katherine S. Button^{1,2}, John P. A. Ioannidis³, Claire Mokrysz¹, Brian A. Nosek⁴, Jonathan Flint⁵, Emma S. J. Robinson⁶ and Marcus R. Munafó¹

Abstract | A study with low statistical power has a reduced chance of detecting a true effect, but it is less well appreciated that low power also reduces the likelihood that a statistically significant result reflects a true effect. Here, we show that the average statistical power of studies in the neurosciences is very low. The consequences of this include overestimates of effect size and low reproducibility of results. There are also ethical dimensions to this problem, as unreliable research is inefficient and wasteful. Improving reproducibility in neuroscience is a key priority and requires attention to well-established but often ignored methodological principles.

It has been claimed and demonstrated that many (and possibly most) of the conclusions drawn from biomedical research are probably false¹. A central cause for this important problem is that researchers must publish in order to succeed, and publishing is a highly competitive enterprise, with certain kinds of findings more likely to be published than others. Research that produces novel results, statistically significant results (that is, typically $p < 0.05$) and seemingly ‘clean’ results is more likely to be published^{2,3}. As a consequence, researchers have strong incentives to engage in research practices that make their findings publishable quickly, even if those practices reduce the likelihood that the findings reflect a true (that is, non-null) effect⁴. Such practices include using flexible study designs and flexible statistical analyses and running small studies with low statistical power^{1,5}. A simulation of genetic association studies showed that a typical dataset would generate at least one false positive result almost 97% of the time⁶, and two efforts to replicate promising findings in biomedicine reveal replication rates of 25% or less^{7,8}. Given that these publishing biases are pervasive across scientific practice, it is possible that false positives heavily contaminate the neuroscience literature as well, and this problem may affect at least as much, if not even more so, the most prominent journals^{9,10}.

Here, we focus on one major aspect of the problem: low statistical power. The relationship between study power and the veracity of the resulting finding is under-appreciated. Low statistical power (because of

low sample size of studies, small effects or both) negatively affects the likelihood that a nominally statistically significant finding actually reflects a true effect. We discuss the problems that arise when low-powered research designs are pervasive. In general, these problems can be divided into two categories. The first concerns problems that are mathematically expected to arise even if the research conducted is otherwise perfect: in other words, when there are no biases that tend to create statistically significant (that is, ‘positive’) results that are spurious. The second category concerns problems that reflect biases that tend to co-occur with studies of low power or that become worse in small, underpowered studies. We next empirically show that statistical power is typically low in the field of neuroscience by using evidence from a range of subfields within the neuroscience literature. We illustrate that low statistical power is an endemic problem in neuroscience and discuss the implications of this for interpreting the results of individual studies.

Low power in the absence of other biases

Three main problems contribute to producing unreliable findings in studies with low power, even when all other research practices are ideal. They are: the low probability of finding true effects; the low positive predictive value (PPV; see BOX 1 for definitions of key statistical terms) when an effect is claimed; and an exaggerated estimate of the magnitude of the effect when a true effect is discovered. Here, we discuss these problems in more detail.

¹School of Experimental Psychology, University of Bristol, Bristol, BS8 1TU, UK.

²School of Social and Community Medicine, University of Bristol, Bristol, BS8 2BN, UK.

³Stanford University School of Medicine, Stanford, California 94305, USA.

⁴Department of Psychology, University of Virginia, Charlottesville, Virginia 22904, USA.

⁵Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK.

⁶School of Physiology and Pharmacology, University of Bristol, Bristol, BS8 1TD, UK. Correspondence to M.R.M.

e-mail: marcus.munaf@bristol.ac.uk
 doi:10.1038/nrn3475
 Published online 10 April 2013
 Corrected online 15 April 2013

Warum biometrische Planung?

- Ziel sind „wissenschaftlich solide“ und „statistisch gesicherte“ Ergebnisse.
- Versuchsplanung nach bester wissenschaftlicher Praxis
 - geeignete Methoden
 - verlässliche Datenerfassung
 - sinnvolle Kontrollen
- Test des Versuchsergebnisses auf „statistische Signifikanz“
 - „einmal ist keinmal“
 - Wiederholungen sind nötig.

Warum biometrische Planung?

- Test des Versuchsergebnisses auf „statistische Signifikanz“
 - „einmal ist keinmal“
 - Wiederholungen sind nötig.
 - Wie viele Wiederholungen?
 - „So viele wie möglich!“, sagt der Statistiker. – Aber 3R!
 - „Aller guten Dinge sind drei“ – „All good things come in threes“ ?
 - „... wurden mindestens 4 (/ 6 / 10) Wiederholungen durchgeführt.“
- Gibt es eine allgemein gültige Mindestanzahl?



<https://www.youtube.com/watch?v=bU7lluL9PgE>

Dasselbe Problem – verschiedene Sprachen

- Einige wichtige Begriffe und Parameter
 1. Wahrscheinlichkeit

Wahrscheinlichkeit

- “Kopf oder Zahl”
 - einfachstes statistische Problem, denn es gibt nur zwei gleich häufige Ergebnisse
 - Die Wahrscheinlichkeit ist jeweils $\frac{1}{2}$.
 - Die Wahrscheinlichkeit, zweimal hintereinander “Zahl” zu erhalten, ist also $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.
 - Das kann man nutzen um zu überprüfen, ob die Münze gezinkt ist.
 - Sechsmal aufeinanderfolgend dieselbe Seite erhält man in $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{32}$ Fällen, also bei fünf Wiederholungen der zuerst geworfenen Seite (~3%).
 - Die Münze ist vermutlich gezinkt, wenn bei sechs aufeinander folgenden Würfeln immer dieselbe Seite erscheint.
 - Sind also $n = 6$ Wiederholungen notwendig (ausreichend), um ein statistisch signifikantes Ergebnis zu erzielen?



Wahrscheinlichkeit

- Würfeln
 - Ist der Würfel manipuliert?
 - Ja, wenn wiederholt dieselbe Augenzahl erscheint.
 - Wie oft wiederholt?
 - häufiger als 5% aller Zufälle
 - Jede Seite des Würfels erscheint mit der Wahrscheinlichkeit $\frac{1}{6}$.
 - Zweimalige Wiederholung der vorher geworfenen Zahl in $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$ der Fälle (das entspricht ca. 2,8%).
 - Der Würfel ist manipuliert, wenn dreimal hintereinander dieselbe Zahl erscheint.



Dasselbe Problem – verschiedene Sprachen

- Einige wichtige Begriffe und Parameter
 1. Wahrscheinlichkeit
 2. Irrtumswahrscheinlichkeit (Fehlerwahrscheinlichkeit)

Irrtumswahrscheinlichkeit

- Anzahl der Wiederholungen hängt ab von der Häufigkeit (= der Wahrscheinlichkeit) der erwarteten Beobachtung.
- Je geringer die Wahrscheinlichkeit, desto geringer die notwendige Anzahl der Wiederholungen.
- Je größer die Anzahl der Wiederholungen, umso (statistisch) sicherer ist die getroffene Aussage.
- Die Wahrscheinlichkeit, immer wieder dasselbe Ergebnis zu erhalten,
 - ist 0,002% (2 von 100.000) nach 6-maligem Würfeln,
 - ist 0,003% (3 von 100.000) nach 15-maligem Münzenwerfen.
- Es bleibt immer eine (statistische) Chance für ein Fehlurteil.
- Aber: die Wahrscheinlichkeit für ein Fehlurteil ist bekannt.

Dasselbe Problem – verschiedene Sprachen

- Einige wichtige Begriffe und Parameter
 1. Wahrscheinlichkeit
 2. Irrtumswahrscheinlichkeit (Fehlerwahrscheinlichkeit)
 3. relevante Differenz

Relevante Differenz

- Wie groß muss der Unterschied der gemessenen Größe zwischen Kontroll- und Versuchsgruppe sein, damit er als biologisch/medizinisch relevant angesehen werden kann?

Die Statistik weiß nichts über Biologie oder Medizin.

Mathematisch bedeutet „Das Versuchsergebnis zeigt keinen Unterschied zwischen Kontroll- und Versuchsgruppe“:

$$\bar{x}_{Kontrolle} - \bar{x}_{Versuch} = 0$$

und bestätigt damit die „Nullhypothese“: Es besteht kein Unterschied!

Aber: Dieses Ergebnis wird vermutlich nicht so eindeutig auftreten.

Relevante Differenz

- Wie groß muss der Unterschied der gemessenen Größe zwischen Kontroll- und Versuchsgruppe sein, damit er als biologisch/medizinisch relevant angesehen werden kann?

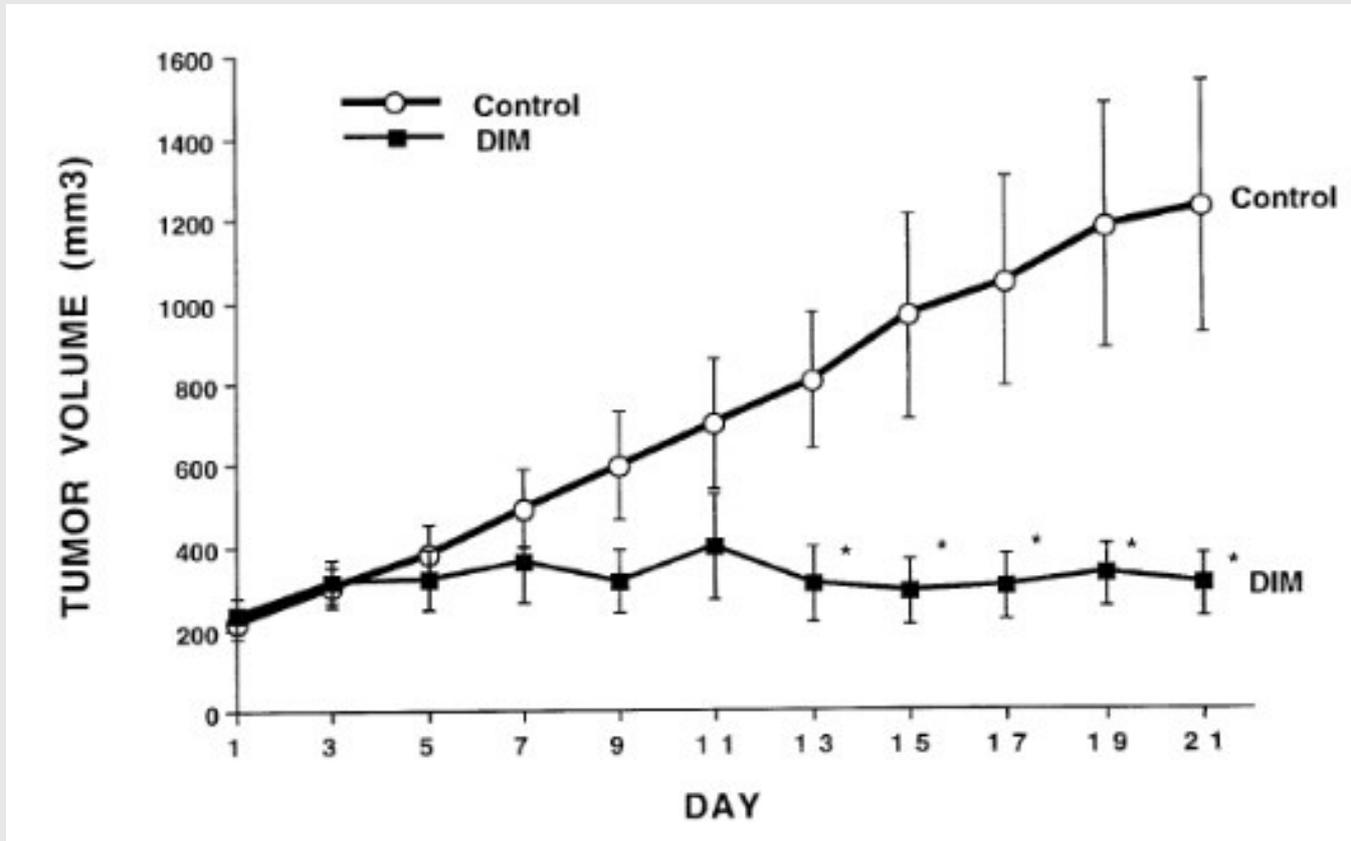
Die Statistik weiß nichts über Biologie oder Medizin.

Mathematisch bedeutet „Das Versuchsergebnis zeigt keinen Unterschied zwischen Kontroll- und Versuchsgruppe“:

$$\bar{x}_{Kontrolle} - \bar{x}_{Versuch} = 0$$

Wie groß darf die Abweichung von „0“ sein?

Relevante Differenz



Chen et al. Carcinogenesis, 19:1631, 1998

Relevante Differenz

- Wie groß muss der Unterschied der gemessenen Größe zwischen Kontroll- und Versuchsgruppe sein, damit er als biologisch/medizinisch relevant angesehen werden darf?

Die Statistik weiß nichts über Biologie oder Medizin.

Mathematisch bedeutet „Das Versuchsergebnis zeigt keinen Unterschied zwischen Kontroll- und Versuchsgruppe“:

kein Unterschied in den Mittelwerten!

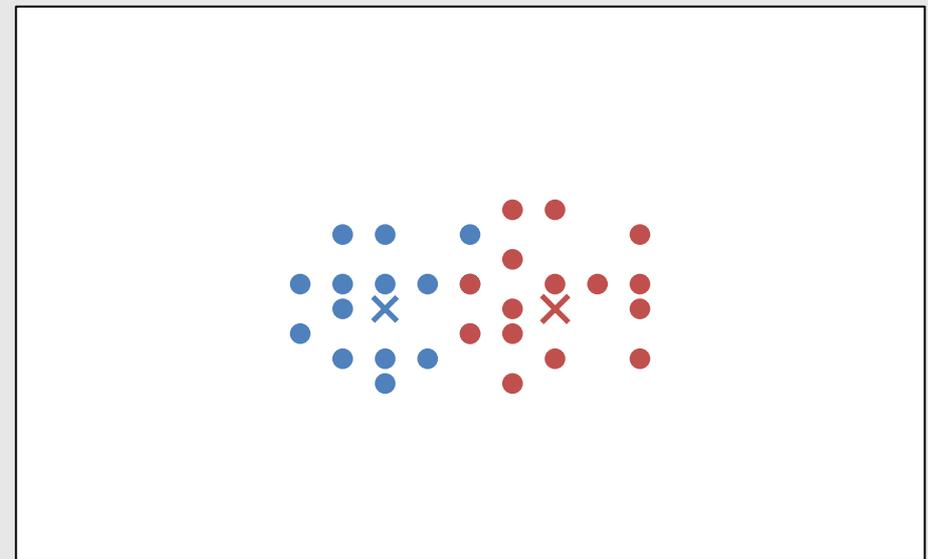
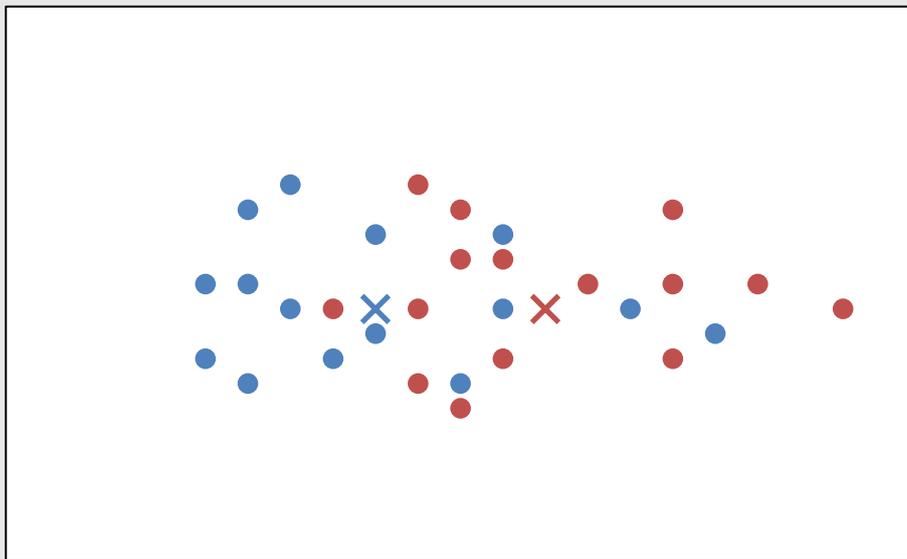
Jeder (noch so kleine) Unterschied kann auf statistische Signifikanz getestet werden, ggfs. aber nur mit großen Versuchsgruppen - also auf Kosten mindestens eines der drei „R“.

Dasselbe Problem – verschiedene Sprachen

- Einige wichtige Begriffe und Parameter
 1. Wahrscheinlichkeit
 2. Irrtumswahrscheinlichkeit (Fehlerwahrscheinlichkeit)
 3. relevante Differenz
 4. Varianz (Streuung)

Varianz - Streuung

- „Die Varianz ist definiert als die mittlere quadrierte Abweichung der Beobachtungen zum Mittelwert.“
- Maß für die Unterschiede eines Messwertes **innerhalb** einer Gruppe



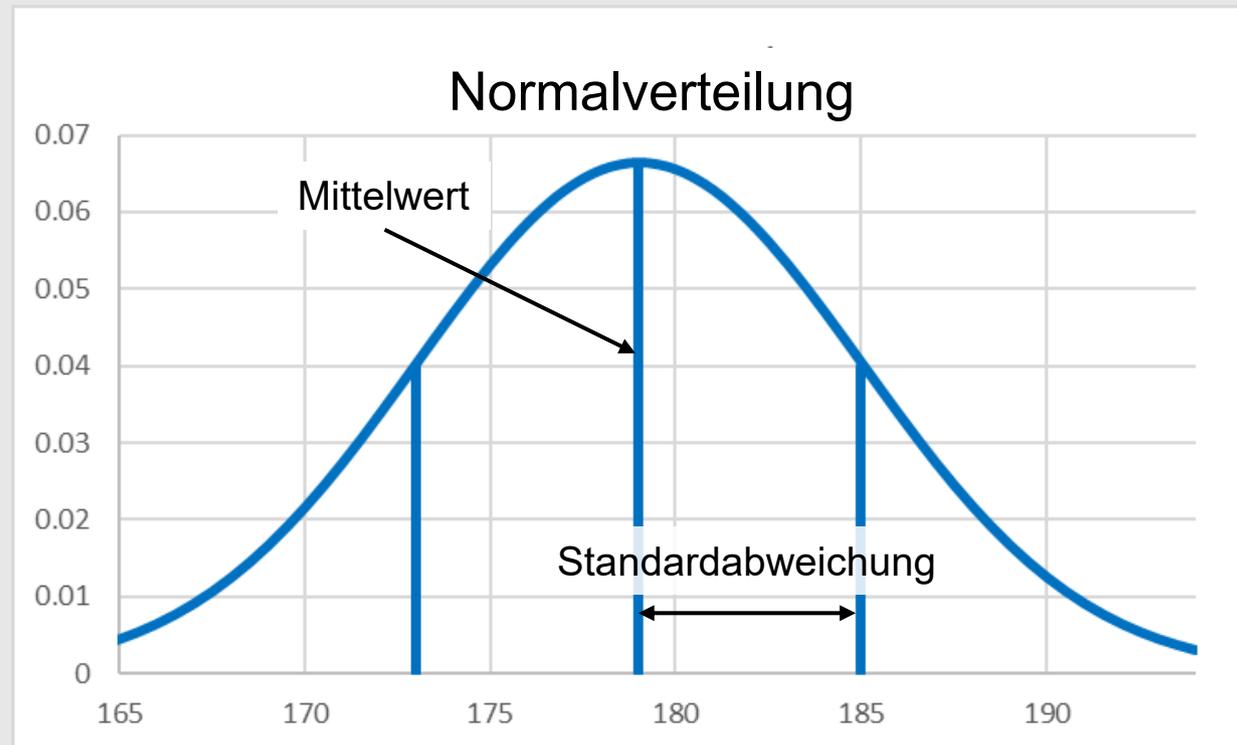
Streuung

- Standardisierung zur Verminderung der Streuung
 - Genotyp (→ Wurfgeschwister), Geschlecht, Alter
 - Haltungs- und Umgebungsbedingungen
 - Versuchsablauf
 - fixe Tageszeiten der Messwerterfassung, wenn zirkadiane Rhythmik Messwerte beeinflussen kann
 - Experimentatoren
 - Messungen werden immer vom selben Experimentator durchgeführt
 - Randomisierung: Tiere völlig zufällig auf die Gruppen verteilen
 - gewährleistet Strukturgleichheit, d. h. Unterschiede können auf Behandlung zurückgeführt werden.

Dasselbe Problem – verschiedene Sprachen

- Einige wichtige Begriffe und Parameter
 1. Wahrscheinlichkeit
 2. Irrtumswahrscheinlichkeit (Fehlerwahrscheinlichkeit)
 3. relevante Differenz
 4. Varianz (Streuung)
 5. Verteilung

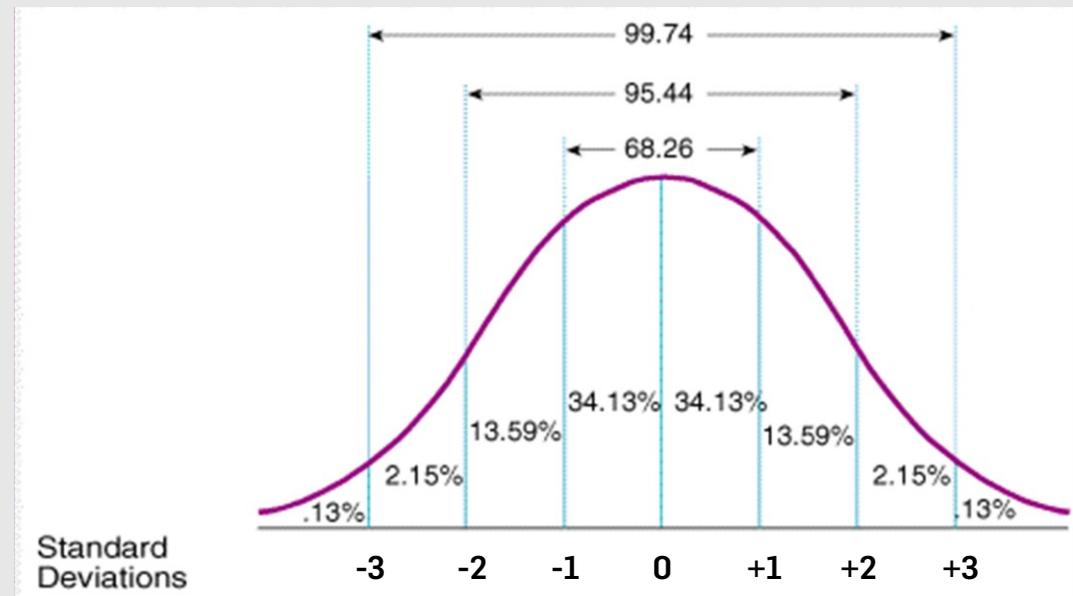
Normalverteilung



Charakteristische Größen der Normalverteilung sind der Mittelwert \bar{x} und die Standardabweichung s .

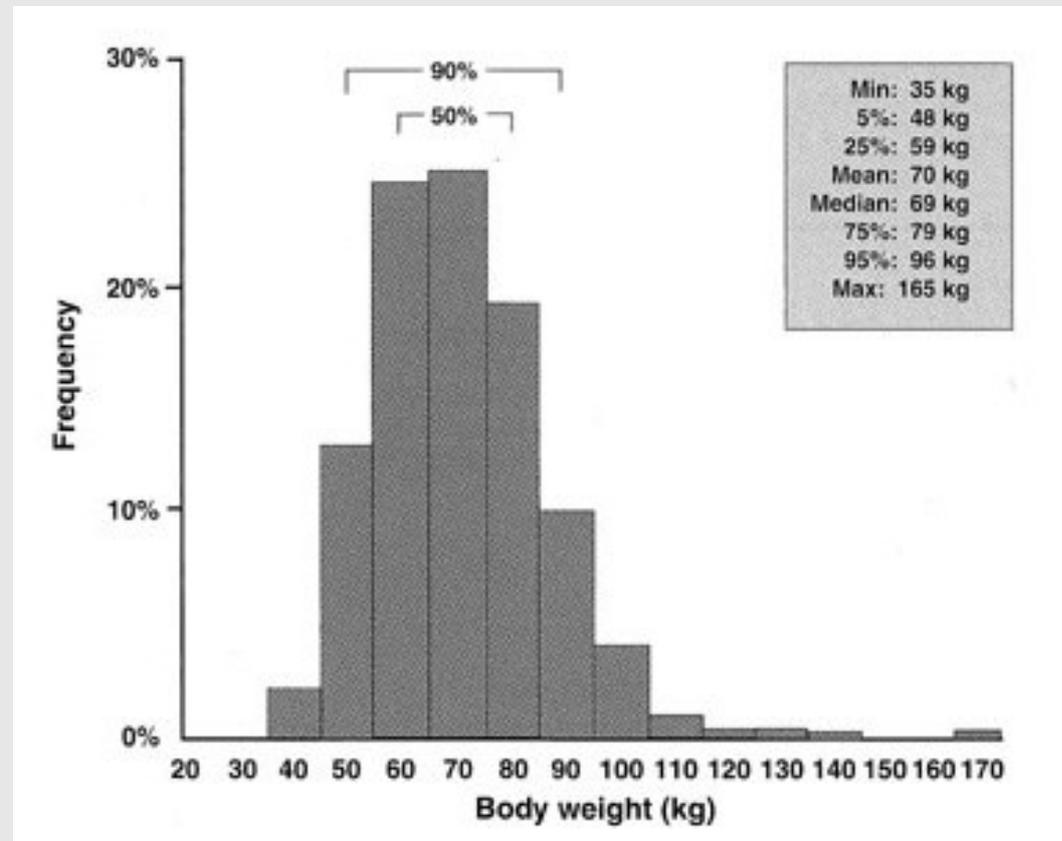
Normalverteilung – Konfidenzintervall

- Sind Mittelwert und Standardabweichung einer Normalverteilung bekannt, dann lassen sich daraus eine Vielzahl von Vorhersagen treffen:
- Häufigkeit (= Wahrscheinlichkeit), dass ein Wert innerhalb eines (beliebigen) Intervalls liegt.
- Häufigkeit (= Wahrscheinlichkeit), mit der ein vorgegebener Wert unter- oder überschritten wird.
- Vor allem die letzte Vorhersage eignet sich für statistische Testverfahren.



Schiefe Verteilung

- Auch wenn biologische Messwerte meist nicht genau einer Normalverteilung folgen, kann man diese Beziehungen auf symmetrische Verteilungen mit einem Maximum meist mit genügend hoher Genauigkeit anwenden.
- Nicht normalverteilt sind in der Regel biologische Größen, die nicht unter einen Minimalwert fallen können (z.B. die nicht negativ sein können).

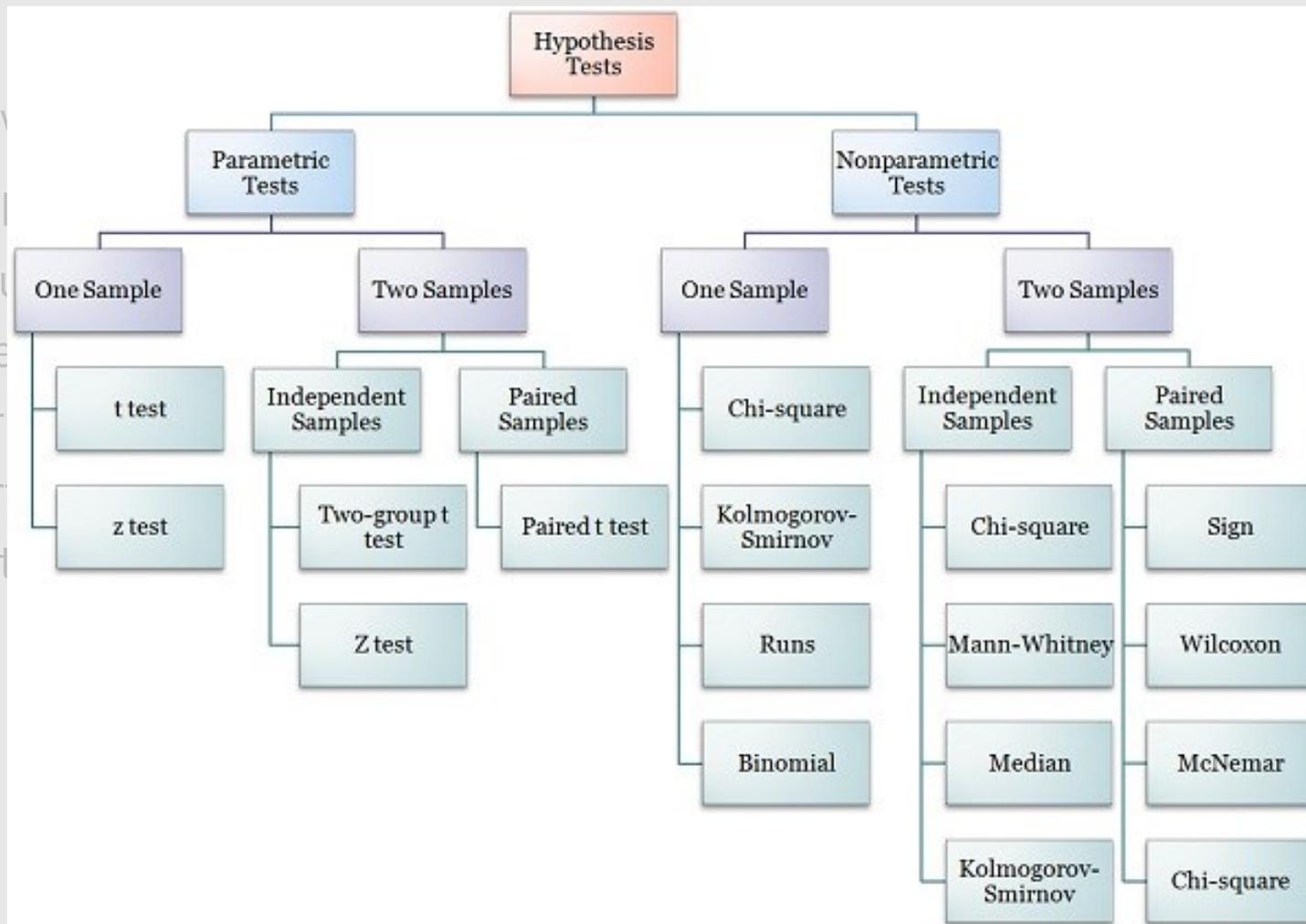


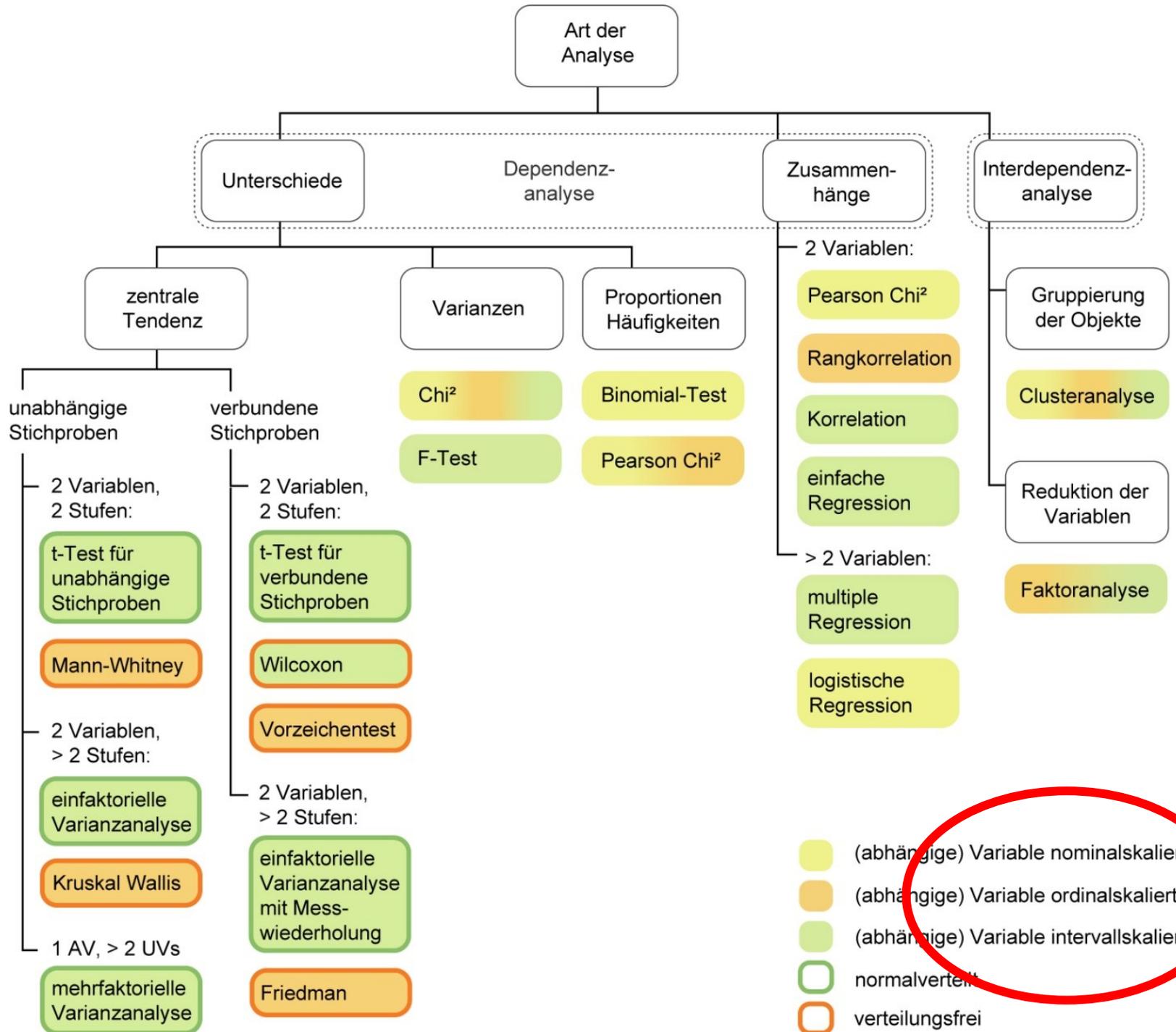
Dasselbe Problem – verschiedene Sprachen

- Einige wichtige Begriffe und Parameter
 1. Wahrscheinlichkeit
 2. Irrtumswahrscheinlichkeit (Fehlerwahrscheinlichkeit)
 3. relevante Differenz
 4. Varianz (Streuung)
 5. Verteilung
 6. statistisches Testverfahren

Dasselbe Problem – verschiedene Sprachen

- Einige v
- 1. Wal
- 2. Irrtu
- 3. rele
- 4. Var
- 5. Ver
- 6. stat



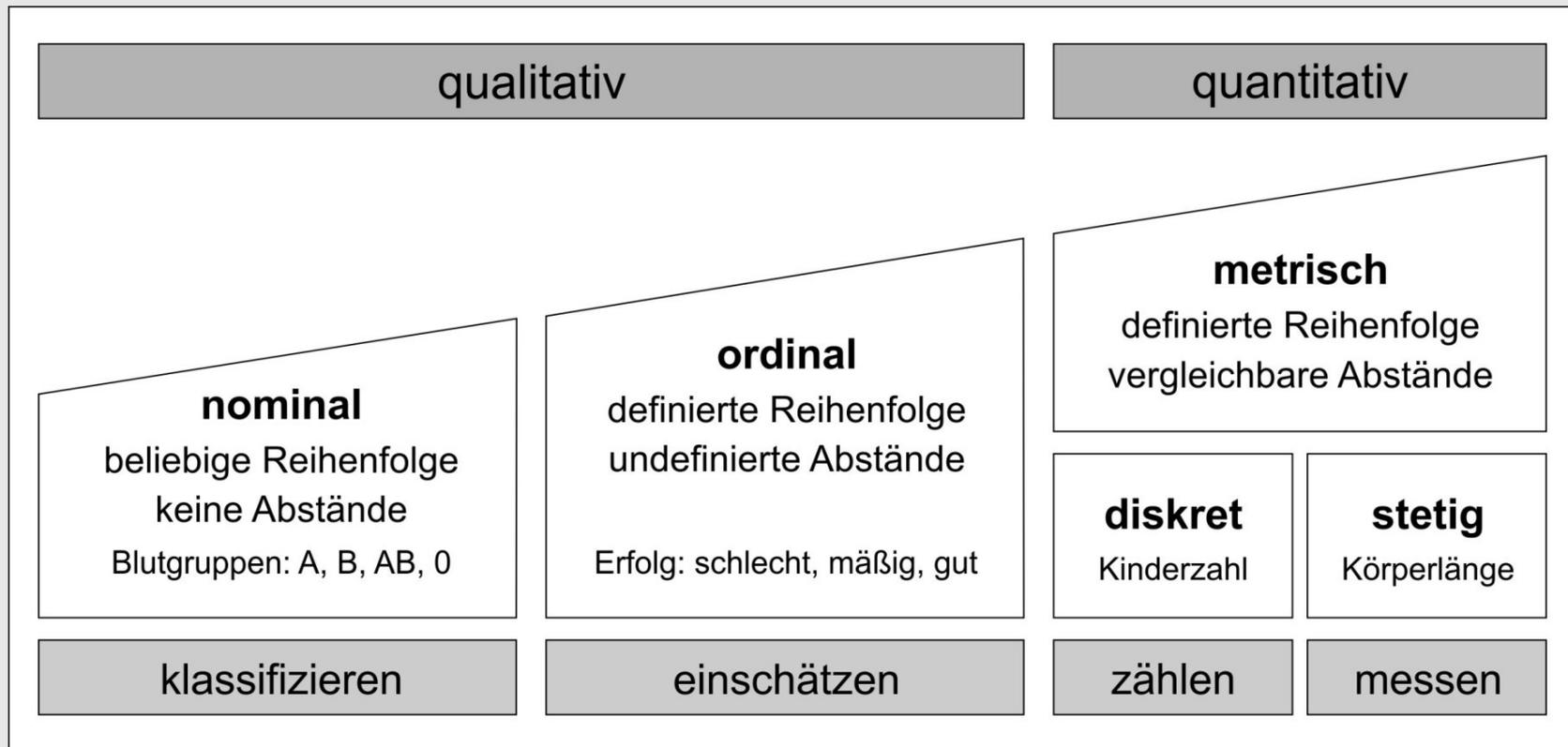


Quelle: <http://www.methodenberatung.uzh.ch/datenanalyse.html>

Dasselbe Problem – verschiedene Sprachen

- Einige wichtige Begriffe und Parameter
 1. Wahrscheinlichkeit
 2. Irrtumswahrscheinlichkeit (Fehlerwahrscheinlichkeit)
 3. relevante Differenz
 4. Varianz (Streuung)
 5. Verteilung
 6. statistisches Testverfahren
 7. Skalierung

Skalierung des Messwertes



Statistische Verfahren zur Minimierung der Tierzahlen

Matthias Schmidt

Skala der Messwerte	Art des Versuchs			
	2 Stichproben		mehr als 2 Stichproben	
	unabhängig (mit verschiedenen Individuen)	verbunden (vor und nach einer Therapie am selben Individuum)	unabhängig (mit verschiedenen Individuen)	verbunden (mehrere Therapien an denselben Individuen)
metrisch	t-Test für unverbundene Stichproben	t-Test für verbundene Stichproben	Varianzanalyse	Varianzanalyse für mehrfache Messungen
ordinal	Mann-Whitney-Rang-Test	Wilcoxon-Rang-Test	Kruskal-Wallis-Test	Friedmann-Test
nominal	Chi-Quadrat-Test für Kontingenztafeln	McNemar-Test	Chi-Quadrat-Test für Kontingenztafeln	Q-Test nach Cochran

Schritte zu einer zuverlässigen biometrischen Planung

1. Übersetzen der (biologischen/medizinischen) Fragestellung in eine für statistische Methoden geeignete Form.

Schritte zu einer zuverlässigen biometrischen Planung

1. Übersetzen der (biologischen/medizinischen) Fragestellung in eine für statistische Methoden geeignete Form.
2. Wie wird die Wirkung (einer Behandlung, einer genetischen Veränderung, ...) bestimmt?
Welche Messgröße wird erfasst?
Welche (mathematischen) Eigenschaften hat die Messgröße (metrisch, ordinal, nominal)?
 - danach entscheidet sich, welches statistische Verfahren angewendet werden kann/darf.

Schritte zu einer zuverlässigen biometrischen Planung

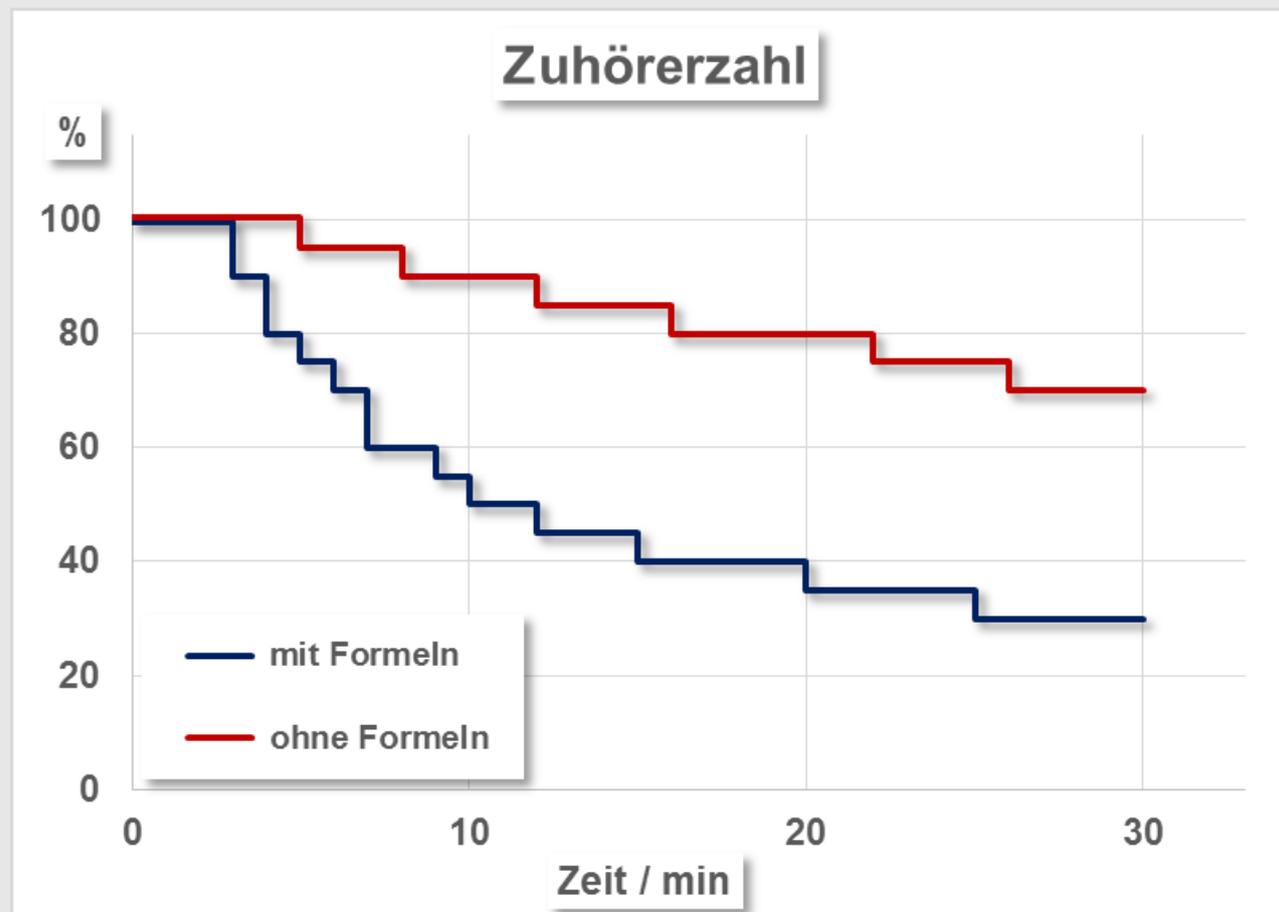
1. Übersetzen der (biologischen/medizinischen) Fragestellung in eine für statistische Methoden geeignete Form.
2. Wie wird die Wirkung bestimmt?
3. Wie groß muss der Unterschied des Messparameters zwischen Kontroll- und Versuchsgruppe sein, damit man von einem Effekt sprechen kann?
4. Wie groß ist die Streuung der Messgröße (unter Kontroll- und Versuchsbedingungen)?

Ohne Kenntnis der Streuung der Messgröße ist die biometrische Planung nicht möglich.

Statistische Verfahren zur Minimierung der Tierzahlen

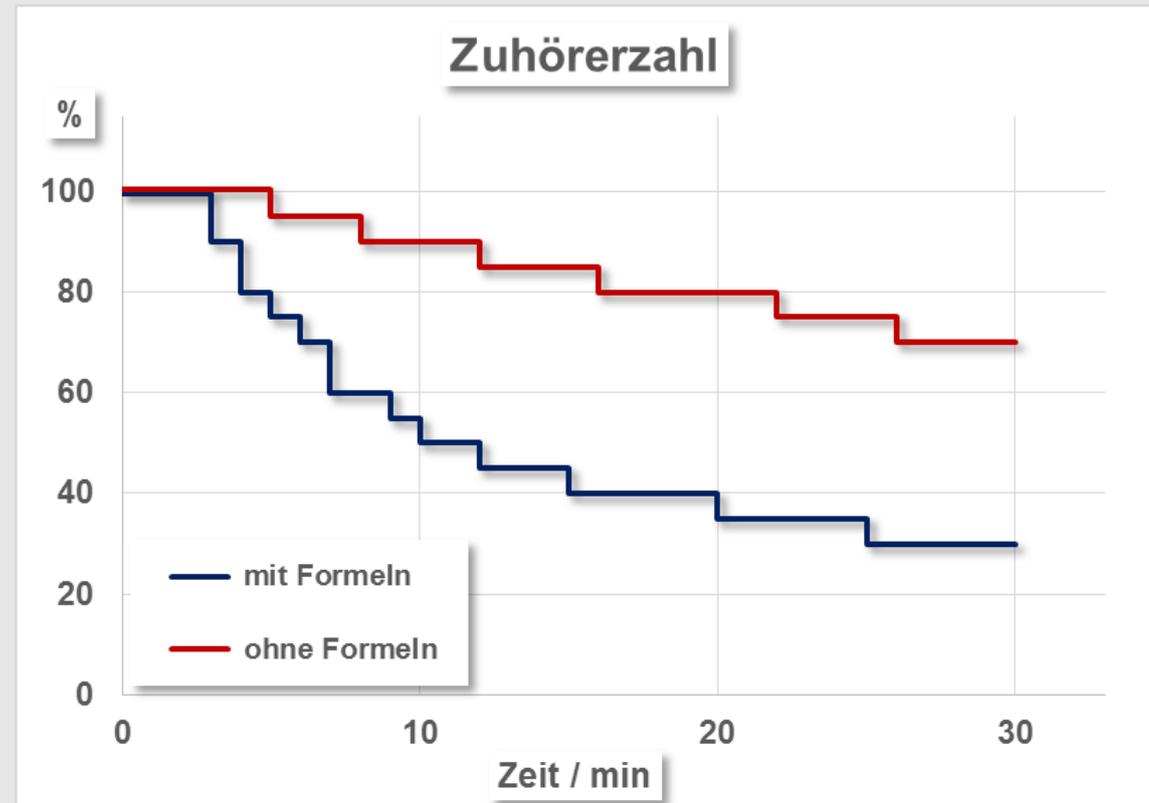
Matthias Schmidt

„Einfluss mathematischer Formeln auf das Zuhörerinteresse“



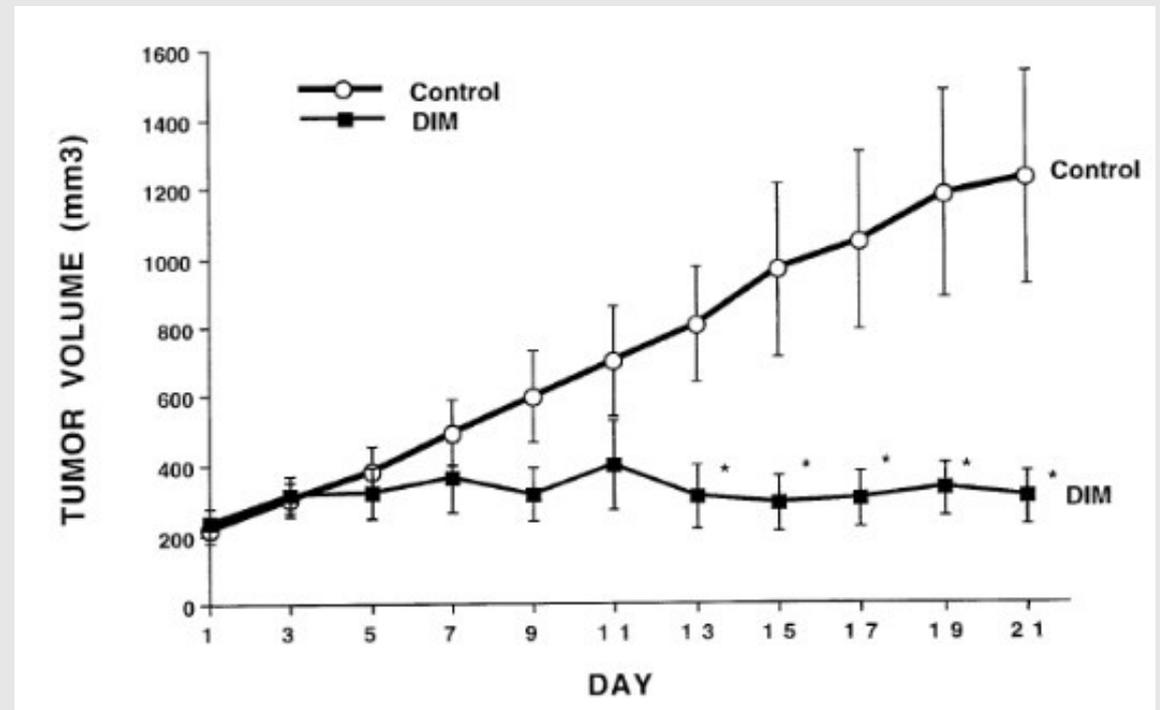
Schritte zu einer zuverlässigen biometrischen Planung

- Übersetzen der Fragestellung in eine für statistische Methoden geeignete Form.
- Ist der Zuhörerschwund unabhängig davon, ob mathematische Formeln präsentiert werden?



Schritte zu einer zuverlässigen biometrischen Planung

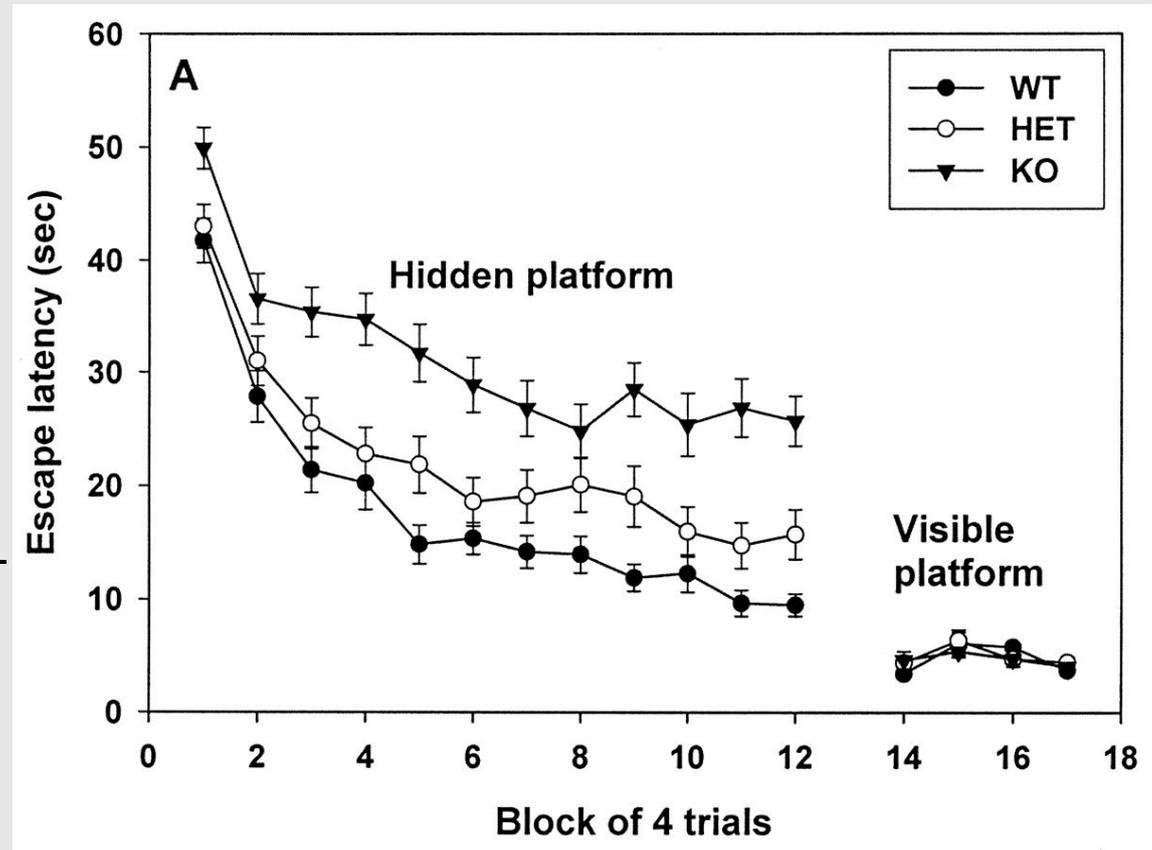
- Übersetzen der Fragestellung in eine für statistische Methoden geeignete Form.
- Ist das verringerte Tumorwachstum unabhängig vom eingesetzten Medikament?



Chen et al. Carcinogenesis, 19:1631, 1998

Schritte zu einer zuverlässigen biometrischen Planung

- Übersetzen der Fragestellung in eine für statistische Methoden geeignete Form.
- Ist der Lernerfolg der genetisch veränderten Mauslinie nur zufällig schlechter als der der Wildtyp-Tiere?



Pak et al. PNAS 97:11232, 2010

Schritte zu einer zuverlässigen biometrischen Planung

- Das Antragsformular fragt die Angaben zur biometrischen Planung einzeln ab.

Angaben zur biometrischen Planung im Antrag

- Das Antragsformular fragt die Angaben zur biometrischen Planung einzeln ab.

1.1.5.2	Vorgesehene Anzahl und Begründung für die Anzahl der Tiere einschließlich Angaben zur biometrischen Planung (§ 31 Abs. 1 Satz 2 Nr. 1c TierSchVersV)	
	- ggf. Anlage „Statistisches Gutachten“ beifügen - - ggf. Anlage Formblatt „Angaben zur biometrischen Planung“ verwenden -	
	Tierart	Gesamtanzahl
	Versuchs- und Kontrollgruppen	
	Erläuterungen:	
	Angaben zur biometrischen/statistischen Planung	
	Erläuterungen:	

Angaben zur biometrischen Planung im Antrag

- Das Antragsformular fragt die Angaben zur biometrischen Planung einzeln ab.

1.1.5.2 Vorgesehene Anzahl und Begründung für die Anzahl der Tiere einschließlich Angaben zur biometrischen Planung (§ 31 Abs. 1 Satz 2 Nr. 1a TierSchVersV)

Die Planung der Tierzahl pro Versuchs- bzw. Kontrollgruppe sollte unter Einsatz biometrischer Verfahren vorgenommen werden. Die Gruppengröße muss so gewählt bzw. errechnet werden, dass die Power groß genug ist, um den erwarteten Mittelwertunterschied bei einem möglichst kleinen Fehler 1. Art bestimmen zu können.

Erläuterungen:

Angaben zur biometrischen/statistischen Planung	
Erläuterungen:	

Angaben zur biometrischen Planung im Antrag

- Das Antragsformular fragt die Angaben zur biometrischen Planung einzeln ab.

1.1.5.2	<p>Vorgesehene Anzahl und Begründung für die Anzahl der Tiere einschließlich Angaben zur biometrischen Planung (§ 31 Abs. 1 Satz 2 Nr. 1c TierSchVersV)</p> <p>- ggf. Anlage „Statistisches Gutachten“ beifügen -</p> <p>- ggf. Anlage Formblatt „Angaben zur biometrischen Planung“ verwenden -</p>								
	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 50%; padding: 2px;">Tierart</th> <th style="width: 50%; padding: 2px;">Gesamtanzahl</th> </tr> </thead> <tbody> <tr> <td style="height: 20px;"></td> <td></td> </tr> <tr> <td style="height: 20px;"></td> <td></td> </tr> <tr> <td style="height: 20px;"></td> <td></td> </tr> </tbody> </table>	Tierart	Gesamtanzahl						
Tierart	Gesamtanzahl								
	<p>Erläuterungen:</p>								
	<p>Angaben zur biometrischen/statistischen Planung</p>								
	<p>Erläuterungen:</p>								

Hier erklären Sie Ihre statistische Vorgehensweise zur Gruppengrößenberechnung wie oben beschrieben.

Angaben zur biometrischen Planung im Antrag

- Das Antragsformular fragt die Angaben zur biometrischen Planung einzeln ab.

1. Hauptzielgröße

- Anzahl der Zuhörer
- Tumorgröße
- Zeit zum Erreichen der Plattform
- subjektive Schmerzintensität („Grimace Scale“)
- Erkrankungsgrad (EAE „score“; Metastasierung)
- Geschlecht
- Blutgruppe

Angaben zur biometrischen/statistischen Planung
Erläuterungen:
Hauptzielgröße(n):

Angaben zur biometrischen Planung im Antrag

- Das Antragsformular fragt die Angaben zur biometrischen Planung einzeln ab.

1. Hauptzielgröße

- Anzahl der Zuhörer
- Tumorgröße
- Zeit zum Erreichen der Plattform

Angaben zur biometrischen/statistischen Planung	
Erläuterungen:	} metrische Größen: T-Test, ANOVA
Hauptzielgröße(n):	

- subjektive Schmerzintensität („Grimace Scale“)
 - Erkrankungsgrad (EAE „score“; Metastasierung)
 - Geschlecht
 - Blutgruppe
- } nominale Größen: χ^2 -Test, Fisher-Yates-Test
- } ordinale Größen: U-Test, Kruskal-Wallis-Test

Angaben zur biometrischen Planung im Antrag

- Das Antragsformular fragt die Angaben zur biometrischen Planung einzeln ab.
- Hauptzielgröße / Nebenzielgröße(n)
 - Studientyp

Nebenzielgröße(n):		
Studientyp		
<input type="checkbox"/>	a)	Orientierungsstudie
<input type="checkbox"/>	b)	Vergleichsstudie
Es werden folgende biometrische Verfahren zur Auswertung eingesetzt:		

Studientyp

Orientierungsstudie (Beobachtungsstudie)

Wissenschaftler greift **nicht aktiv** ins Geschehen ein;
ohnedies vorhandene Merkmale werden erhoben;
in der Regel **keine Rückschlüsse auf Kausalität** möglich, nur Aussagen über Assoziationen

Auswertung

beschreibend (deskriptiv) oder **Hypothesen generierend (explorativ)**

Beispiele

die meisten epidemiologischen Studien
Ökosystemstudien
(Retrospektive) Studien in einer Klinik

Vergleichsstudie (experimentelle Studie)

Wissenschaftler **beeinflusst Merkmale aktiv**, er interveniert;
die „Behandlung“ wird kontrolliert;
herausfinden, **ob Behandlung** bei den Probanden/Tieren **einen Effekt verursacht**

Auswertung

konfirmatorische Auswertung, *a priori* aufgestellte Hypothesen werden geprüft

Beispiele

fast alle tierexperimentellen Studien
klinische Prüfungen zur Erprobung neuer Therapien (es gilt das AMG)

Angaben zur biometrischen Planung im Antrag

- Antragsformular fragt die Angaben zur biometrischen Planung einzeln ab.

1. Hauptzielgröße

2. Studientyp

3. Statistisches Verfahren zur Auswertung

→ Skala der Messgröße

→ Anzahl der Gruppen

Angaben zur biometrischen/statistischen Planung		
Erläuterungen:		
Hauptzielgröße(n):		
Nebenzielgröße(n):		
Studientyp		
<input type="checkbox"/>	a)	Orientierungsstudie
<input type="checkbox"/>	b)	Vergleichsstudie
Es werden folgende biometrische Verfahren zur Auswertung eingesetzt:		

Angaben zur biometrischen Planung im Antrag

- Antragsformular fragt die Angaben zur biometrischen Planung einzeln ab.

1. Hauptzielgröße

2. Studientyp

3. Statistisches Verfahren zur Auswertung

→ Skala der Messgröße

→ Anzahl der Gruppen

Mess-Skala	Versuchsdesign	
	2 Gruppen mit verschiedenen Individuen	mehr als 2 Gruppen mit verschiedenen Individuen
Metrisch	t-Test	Varianzanalyse (ANOVA)
Ordinal	Mann-Whitney-Rang-Test	Kruskal-Wallis-Test
Nominal	χ^2 -Test	χ^2 -Test

Angaben zur biometrischen Planung im Antrag

- Antragsformular fragt die Angaben zur biometrischen Planung einzeln ab.

3. Statistisches
Verfahren zur
Auswertung

Die vorgesehene Tierzahl und Gruppengröße ist notwendig zur statistischen Absicherung mit	
–	einer Wahrscheinlichkeit für den Fehler 1. Art von
	0,05 ($\alpha = 5\%$, $p = 0,05$)

4. Fehler 1. Art

gibt an, mit welcher Wahrscheinlichkeit das Ergebnis zufällig entstanden sein könnte.

Effekt einer Behandlung wird angenommen, obwohl er in Wirklichkeit zufällig entstanden ist → „falsch positives“ Ergebnis

5% wird akzeptiert

Signifikanzniveau → p-Wert

Aussagekraft von p-Werten



“The p-value was never intended to be a substitute for scientific reasoning.”

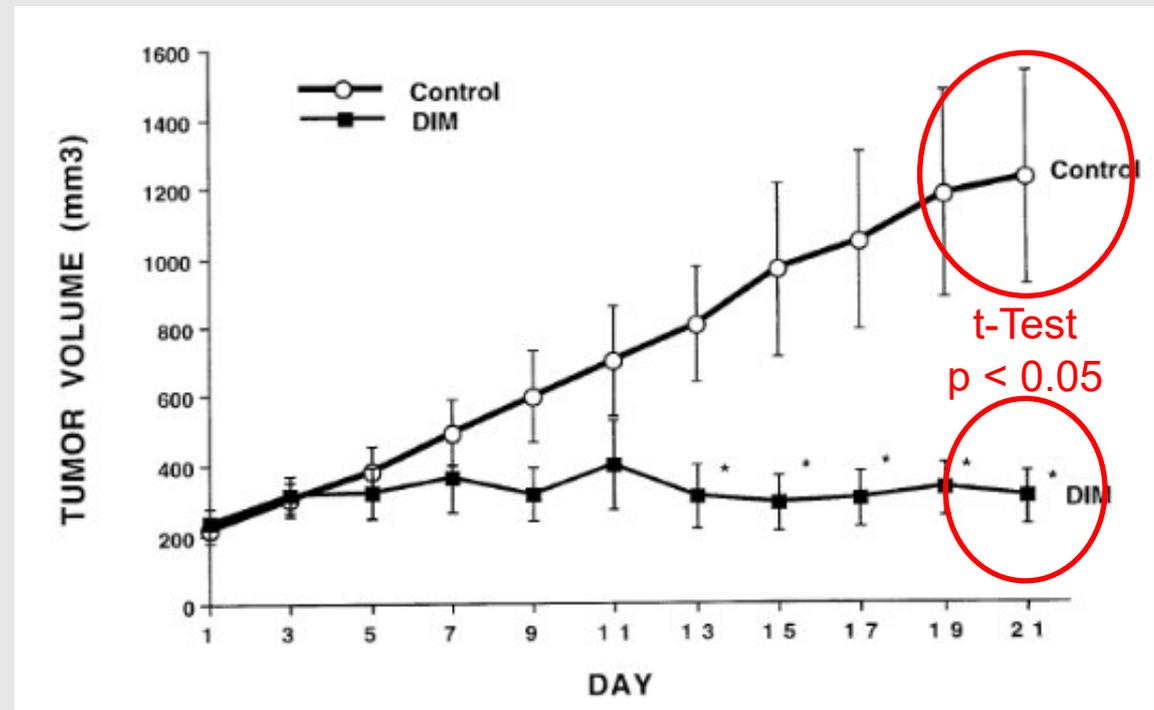
Ron Wasserstein,
Executive Director,
American Statistical Association

Aussagekraft von p-Werten

Die Wirkung einer DIM-Behandlung auf die TumorgroÙe ist eindeutig: während Tumore unter Kontrollbedingungen kontinuierlich wachsen, bleiben Tumore in behandelten Tieren unverändert.

Eine statistische Absicherung dieses Ergebnisses ist wohl wissenschaftlich nicht nötig.

p ist sicherlich deutlich kleiner als 0,05 (n=10).

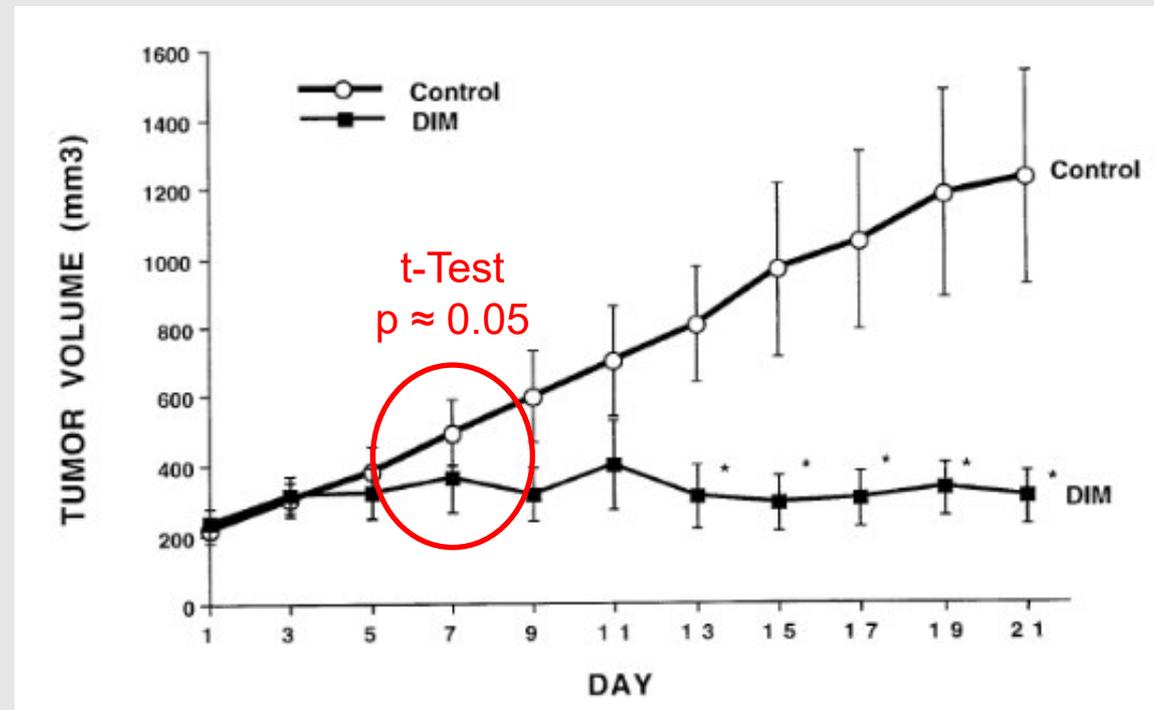


Chen et al. Carcinogenesis, 19:1631, 1998

Aussagekraft von p-Werten

Die Wirkung einer DIM-Behandlung auf die Tumorgröße ist eindeutig: während Tumore unter Kontrollbedingungen kontinuierlich wachsen, bleiben Tumore in behandelten Tieren unverändert.

Eine statistische Absicherung dieses Ergebnisses ist wohl wissenschaftlich nicht nötig.



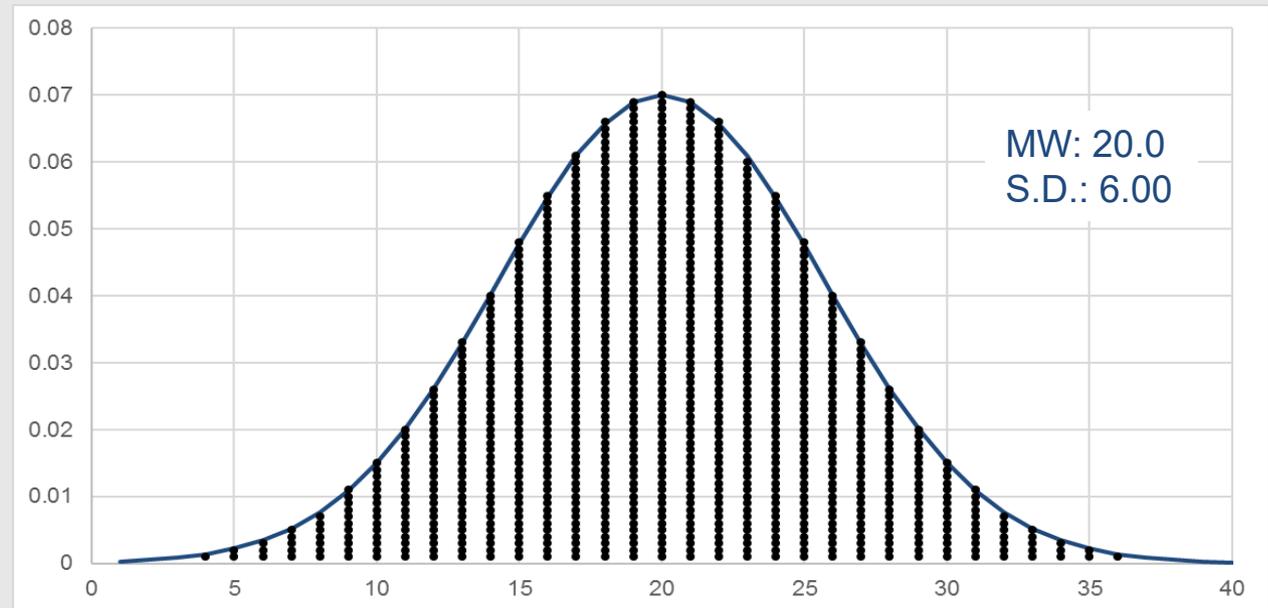
Chen et al. Carcinogenesis, 19:1631, 1998

Zu einem früheren Zeitpunkt liegt p sehr nahe am Schwellenwert von 0,05.

Aussagekraft von p-Werten

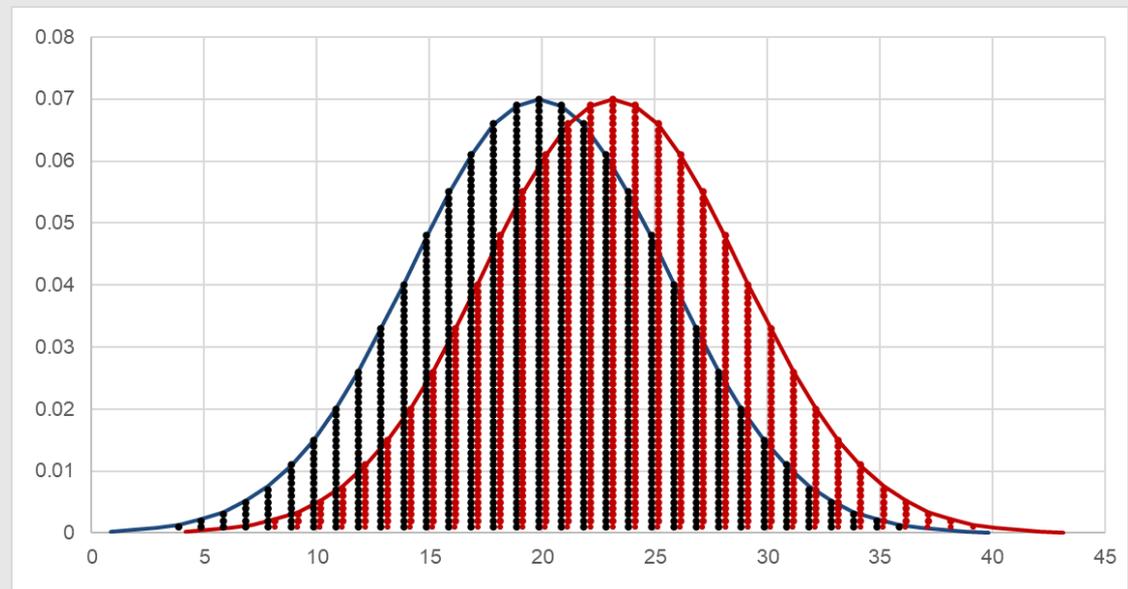
Die Zuverlässigkeit von p-Werten ist von mehreren Faktoren abhängig:

- Größe des Unterschieds
- Streuung innerhalb der Gruppen
- Größe der Stichprobe



Aussagekraft von p-Werten

Test-Nr.	n=6	n=10	n=20	n=30
1	0.06	0.22	0.04	0.02
2	0.56	0.32	0.16	0.07
3	0.46	0.14	0.08	0.05
4	0.11	0.05	0.03	0.08
5	0.35	0.22	0.12	0.03



Die Zuverlässigkeit von p-Werten ist von mehreren Faktoren abhängig:

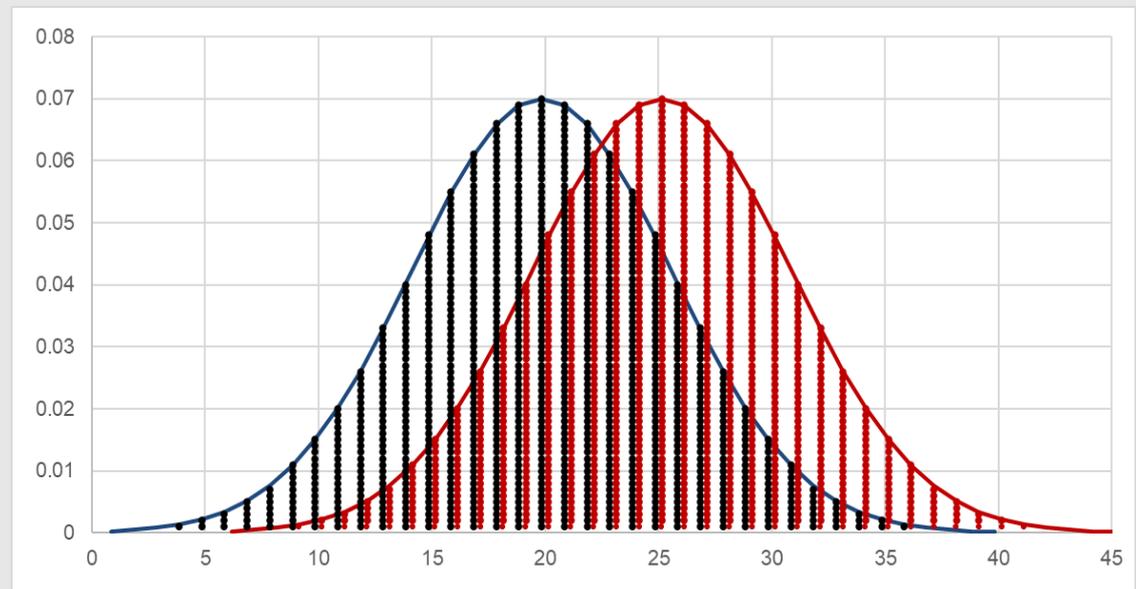
- Größe des Unterschieds:
- Streuung innerhalb der Gruppen
- Größe der Stichprobe

$$\text{Mittelwert}_{\text{rot}} = \text{Mittelwert}_{\text{schwarz}} + 3$$

$$\text{Streuung}_{\text{rot}} = \text{Streuung}_{\text{schwarz}} = \pm 6$$

Aussagekraft von p-Werten

Test-Nr.	n=6	n=10	n=20	n=30
1	0.04809	0.16698	0.01227	0.00154
2	0.22607	0.05571	0.01983	0.00069
3	0.03936	0.06703	0.00336	0.00043
4	0.17058	0.07457	0.01625	0.00145
5	0.18042	0.03882	0.00501	0.00234



Die Zuverlässigkeit von p-Werten ist von mehreren Faktoren abhängig:

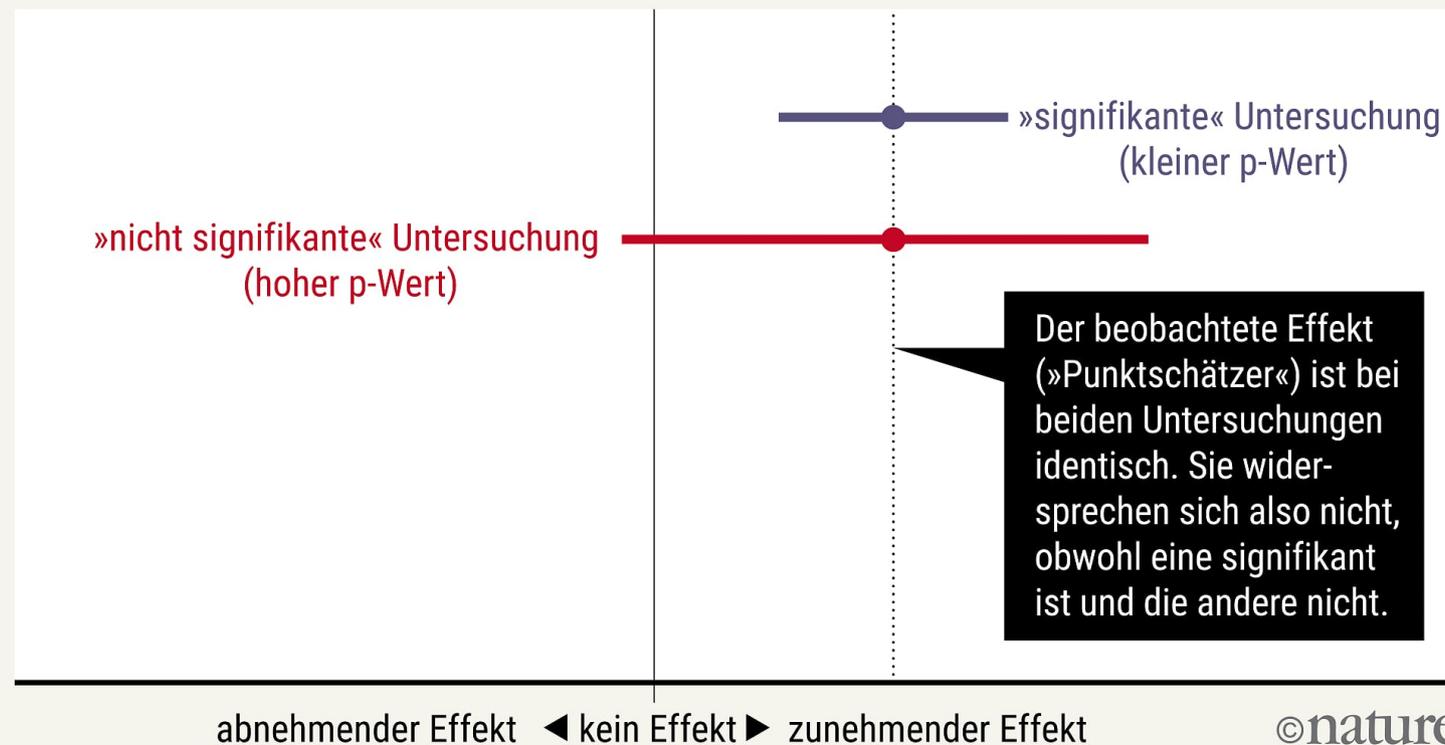
- Größe des Unterschieds:
- Streuung innerhalb der Gruppen
- Größe der Stichprobe

$$\text{Mittelwert}_{\text{rot}} = \text{Mittelwert}_{\text{schwarz}} + 5$$

$$\text{Streuung}_{\text{rot}} = \text{Streuung}_{\text{schwarz}} = \pm 6$$

Vorsicht vor falschen Schlussfolgerungen

Untersuchungen, die derzeit als »statistisch signifikant« oder »statistisch nicht signifikant« eingestuft werden, müssen deshalb keineswegs zueinander im Widerspruch stehen. Solche Einordnungen können dazu führen, dass real vorhandene Effekte übersehen werden.

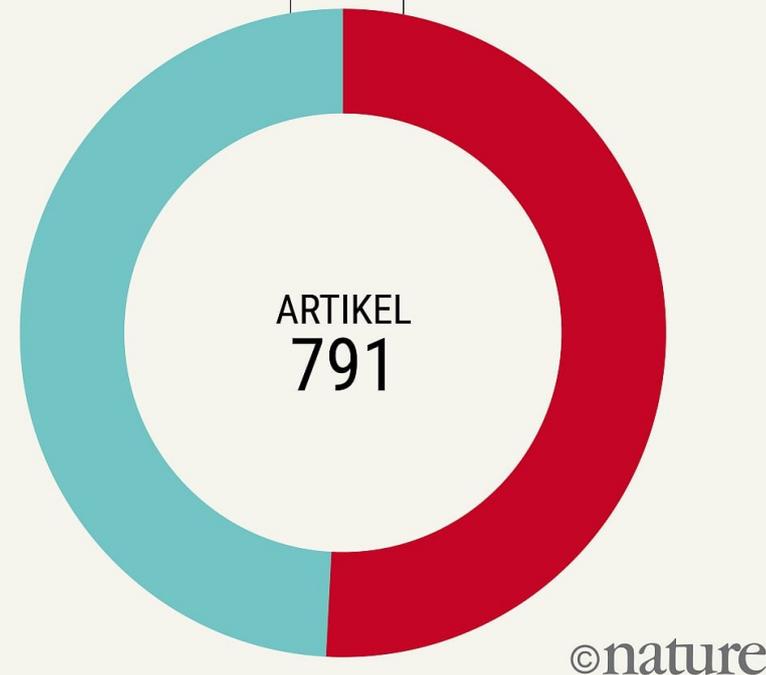


Falsche Interpretationen

Die Analyse von 791 Artikeln in fünf Fachzeitschriften hat gezeigt, dass in etwa der Hälfte fälschlich aus einem nicht signifikanten Ergebnis gefolgert wird, es gäbe den untersuchten Effekt nicht.

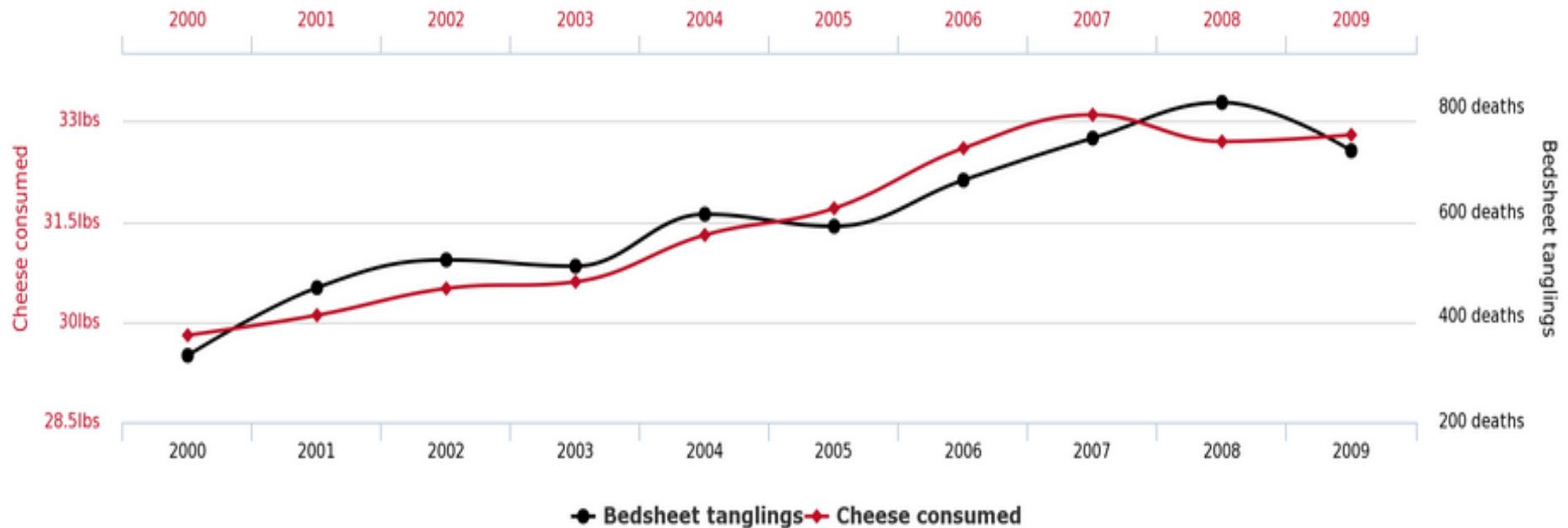
angemessen interpretiert
49%

falsch interpretiert
51%



© Nature; Amrhein, V. et al.: Retire statistical significance. Nature 567, 2019; nach Daten von: Schatz, P. et al.: Arch. Clin. Neuropsychol. 20, 2005; Fidler, F. et al.: Conserv. Biol. 20, 2006; Hoekstra, R. et al.: Psychon. Bull. Rev. 13, 2006; Bernardi, F. et al.: Eur. Sociol. Rev. 33, 2017; dt. Bearbeitung: Spektrum der Wissenschaft (Ausschnitt)

Pro-Kopf-Verbrauch von Käse
korreliert mit
Anzahl der Strangulationen durch Bettlaken



tylervigen.com

„p-hacking“

- auch bekannt als *“data-dredging”*, *“snooping”*, *“significance-chasing”*, *“double-dipping”*
- durch nachträgliches Verändern der Datenanalyse oder der Analyse-Bedingungen (bewusst oder unbewusst) wird versucht, ein vorgegebenes Signifikanzniveau zu erreichen
 - Wiederholung einer Messreihe, bis statistische Signifikanz erreicht ist
 - verschiedene statistische Methoden bei der Auswertung verwenden, aber nur diejenigen zu berichten, die statistisch signifikant wurden
 - Abbrechen der Messwerterhebung, wenn statistische Signifikanz erreicht ist
 - Nicht-Berücksichtigung von „Ausreißern“
 - Reduktion des Skalen-Niveaus (z.B. Altersklassen statt Alter)
 - Aggregieren von zwei verschiedenen Variablen in eine, beispielsweise durch die Bildung einer Relativgröße

Aussagekraft von p-Werten

Marital satisfaction and break-ups differ across on-line and off-line meeting venues (PNAS, 110 (25) 10135-10140, 2013).

Marital discord is costly to children, families, and communities. The advent of the Internet, social networking, and on-line dating has affected how people meet future spouses, but little is known about the prevalence or outcomes of these marriages or the demographics of those involved. We addressed these questions in a nationally representative sample of 19131 respondents who married between 2005 and 2012. Results indicate that more than one-third of marriages in America now begin on-line. In addition, **marriages that began on-line, when compared with those that began through traditional off-line venues, were slightly less likely to result in a marital break-up (separation or divorce) and were associated with slightly higher marital satisfaction among those respondents who remained married.** Demographic differences were identified between respondents who met their spouse through on-line vs. traditional off-line venues, but **the findings for marital break-up and marital satisfaction remained significant after statistically controlling for these differences.**

Aussagekraft von p-Werten

Marital satisfaction and break-ups differ across on-line and off-line meeting venues

PNAS, 110 (25) 10135-10140, 2013.

The percentage of marital break-ups was lower for respondents who met their spouse on-line (5.96%) than off-line (7.67%; $p < 0.002$).

Analyses indicated that currently married respondents who met their spouse on-line reported higher marital satisfaction ($M = 5.64$, $SE = 0.02$, $n = 5349$) than currently married respondents who met their spouse off-line ($M = 5.48$, $SE = 0.01$, $n = 12253$; mean difference = 0.18, $F[1, 17,601] = 46.67$, $p < 0.001$).

“These data suggest that the Internet may be altering the dynamics and outcomes of marriage itself”

Aussagekraft von p-Werten

Marital satisfaction and break-ups differ across on-line and off-line meeting venues

PNAS, 110 (25) 10135-10140, 2013.

The percentage of marital break-ups was lower for respondents who met their spouse on-line (5.96%) than off-line (7.67%; $p < 0.002$).

Analyses indicated that currently married respondents who met their spouse on-line reported higher marital satisfaction ($M = 5.64$, $SE = 0.02$, $n = 5349$) than currently married respondents who met their spouse off-line ($M = 5.48$, $SE = 0.01$, $n = 12253$; mean difference = 0.18, $F[1, 17,601] = 46.67$, $p < 0.001$). (auf einer Skala von 1 bis 7)

“These data suggest that the Internet may be altering the dynamics and outcomes of marriage itself”

“We should be asking, 'How much of an effect is there?', not, 'Is there an effect?'”

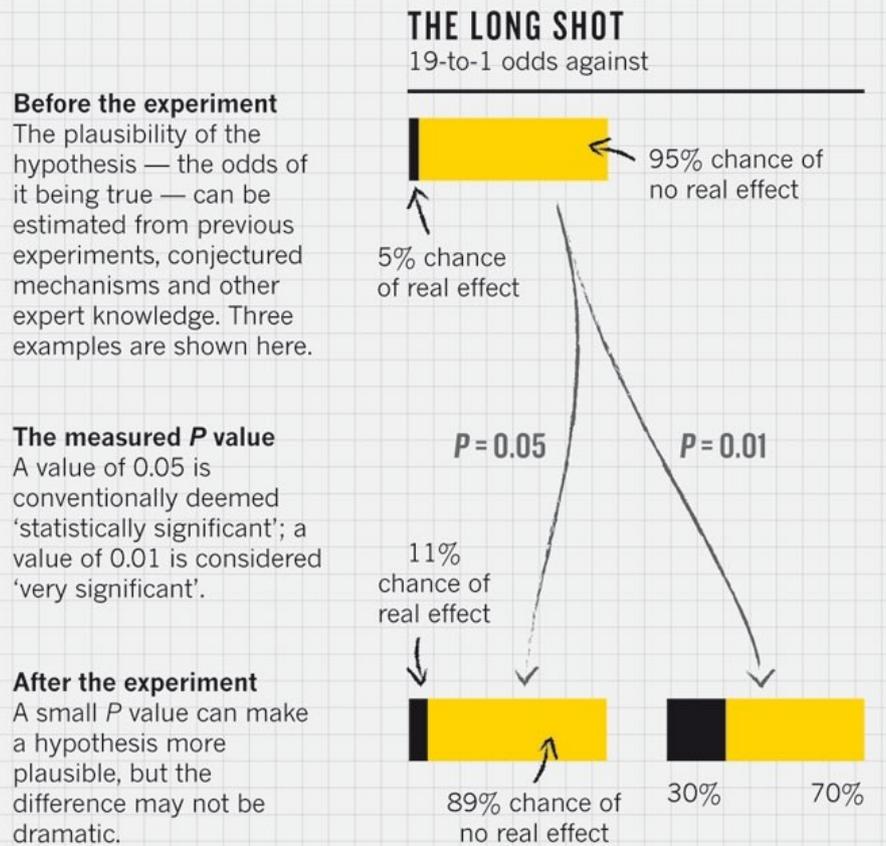
Statistische Verfahren zur Minimierung der Tierzahlen

Matthias Schmidt

PROBABLE CAUSE

A *P* value measures whether an observed result can be attributed to chance. But it cannot answer a researcher's real question: what are the odds that a hypothesis is correct? Those odds depend on how strong the result was and, most importantly, on how plausible the hypothesis is in the first place.

■ Chance of real effect
 ■ Chance of no real effect



- Die Aussagekraft eines „statistisch signifikanten“ Ergebnisses wird außerdem dadurch bestimmt, wie realistisch die getestete Hypothese ist.
- Bei einer wenig realistischen Hypothese ist auch ein statistisch signifikantes Ergebnis nicht unbedingt ein starkes Argument für deren Richtigkeit.

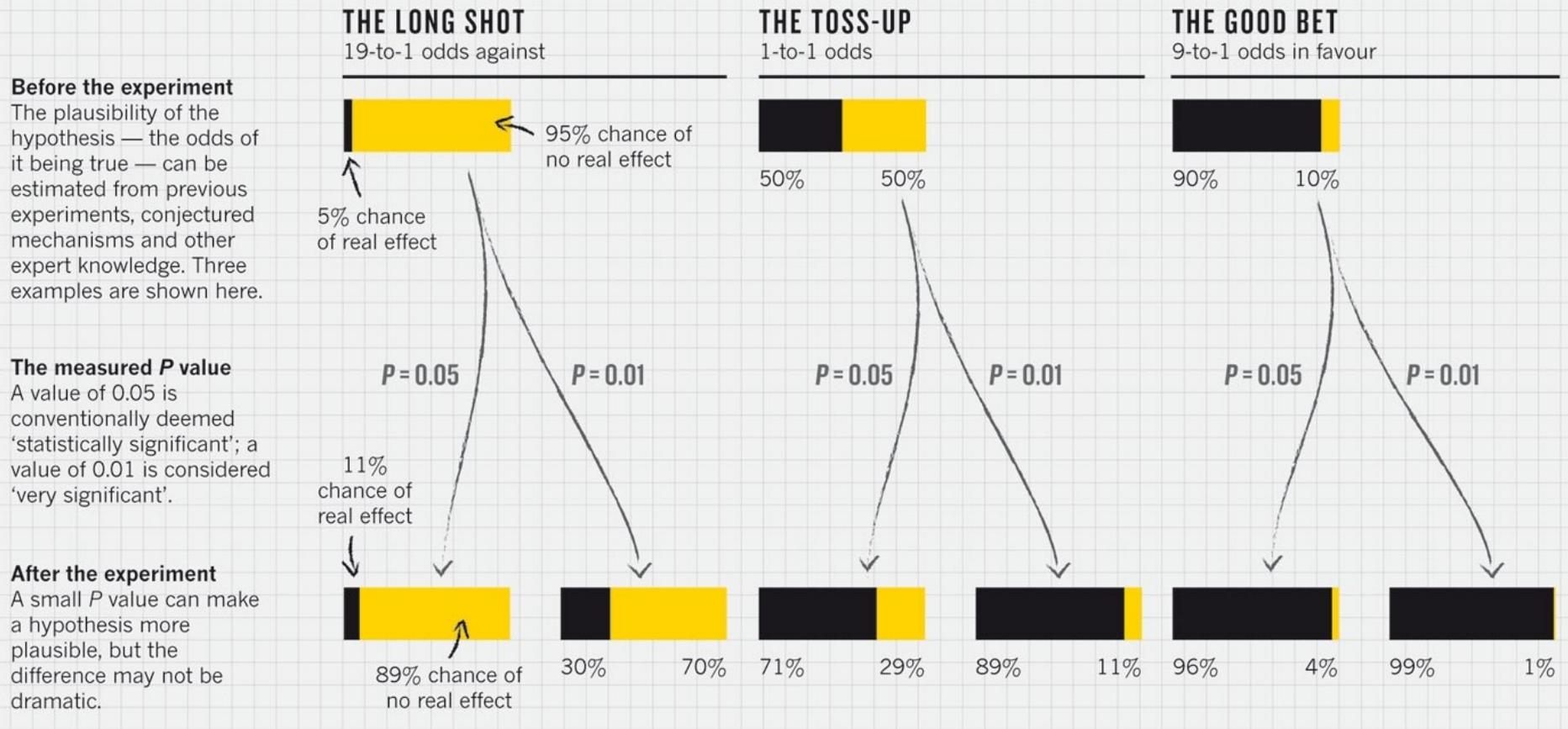
Statistische Verfahren zur Minimierung der Tierzahlen

Matthias Schmidt

PROBABLE CAUSE

A *P* value measures whether an observed result can be attributed to chance. But it cannot answer a researcher's real question: what are the odds that a hypothesis is correct? Those odds depend on how strong the result was and, most importantly, on how plausible the hypothesis is in the first place.

■ Chance of real effect
 ■ Chance of no real effect



Aus: Nuzzo R (2014) Scientific method: Statistical errors, Nature 506:150–152

Statistische Verfahren zur Minimierung der Tierzahlen

Matthias Schmidt

PROBABLE CAUSE

A *P* value measures whether an observed result can be attributed to chance. But it cannot answer a researcher's real question: what are the odds that a hypothesis is correct? Those odds depend on how strong the result was and, most importantly, on how plausible the hypothesis is in the first place.

■ Chance of real effect
■ Chance of no real effect

THE LONG SHOT

19-to-1 odds against

THE TOSS-UP

1-to-1 odds

THE GOOD BET

9-to-1 odds in favour

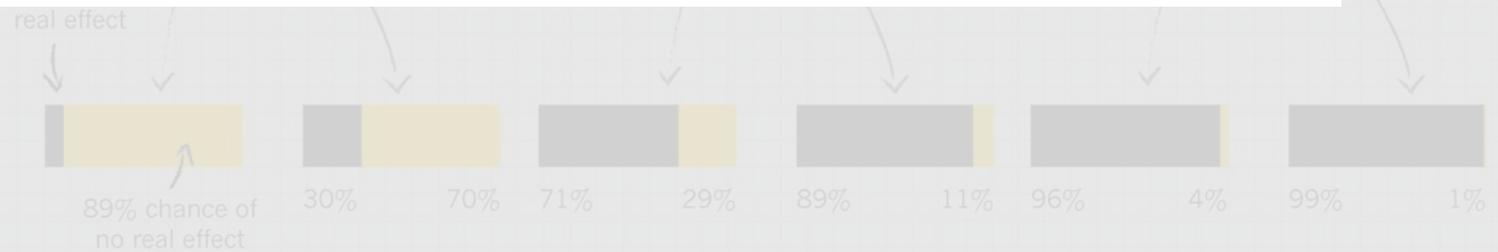
Before the experiment
The plausibility of a hypothesis — its being true — is estimated from experiments, mechanisms, expert knowledge, and other examples are

The measure
A value of 0.05 is conventional, 'statistically significant', a value of 0.01 is 'very significant'

“The more implausible the hypothesis — telepathy, aliens, homeopathy — the greater the chance that an exciting finding is a false alarm, no matter what the *p* value is.”

After the experiment

A small *P* value can make a hypothesis more plausible, but the difference may not be dramatic.



Aussagekraft von p-Werten

- p-Werte geben keine Auskunft über Plausibilität, Vorhandensein, Wahrhaftigkeit oder wissenschaftliche Bedeutung einer Beobachtung oder eines Effekts.
- Die Bezeichnung „statistisch signifikant“ ist kein Maß dafür, dass eine Beobachtung oder ein Effekt besonders glaubwürdig, real, richtig oder wissenschaftlich bedeutsam ist.
- Ebenso bedeutet „nicht signifikant“ nicht, dass eine Beobachtung oder ein Effekt unglaubwürdig, nicht vorhanden, falsch oder wissenschaftlich unwichtig ist.

Wasserstein et al. (2019) Moving to a World Beyond " $p < 0.05$ ". Am Stat 73, S1, 1-19

Angaben zur biometrischen Planung im Antrag

- Antragsformular fragt die Angaben zur biometrischen Planung einzeln ab.

3. Statistisches Verfahren zur Auswertung

4. Fehler 1. Art

5. Fehler 2. Art

Die vorgesehene Tierzahl und Gruppengröße ist notwendig zur statistischen Absicherung mit	
–	einer Wahrscheinlichkeit für den Fehler 1. Art von
	0,05
–	einer Wahrscheinlichkeit für den Fehler 2. Art von
	0,20 ($\beta = 20\%$, daraus Power $(1 - \beta) = 0,80$)

gibt an, mit

welcher Wahrscheinlichkeit ein tatsächlicher Effekt übersehen wird, ein Ergebnis also fälschlicherweise als zufällig angesehen wird.

→ „falsch negatives“ Ergebnis

20% wird akzeptiert

Statistische Power

- In der Regel soll mit Hilfe des statistischen Tests nachgewiesen werden, dass die Nullhypothese verworfen werden kann, weil das bedeutet, dass die Behandlung (/das Medikament/die genetische Veränderung/...) eine signifikante Wirkung hat.
- Der Unterschied zwischen Kontroll- und der Behandlungsgruppe(n) beschreibt den **Effekt** der Behandlung/des Medikaments, die Wahrscheinlichkeit, einen vorhandenen Effekt zu erkennen, beschreibt die **statistische Power**.
- Vor allem bei Experimenten mit vielen Bedingungen und Wirkungen („-omics“) können viele signifikante Ergebnisse dennoch falsch sein.

Angaben zur biometrischen Planung im Antrag

- Antragsformular fragt die Angaben zur biometrischen Planung einzeln ab.

4. Fehler 1. Art

5. Fehler 2. Art

6. Differenz

Wie groß soll die Abweichung der Messwerte der Versuchsgruppe(n) vom

Referenzwert der Kontrollgruppe sein, damit von einem biologisch oder medizinisch relevantem Effekt gesprochen werden kann?

Die vorgesehene Tierzahl und Gruppengröße ist notwendig zur statistischen Absicherung mit	
–	einer Wahrscheinlichkeit für den Fehler 1. Art von
	0,05 ($\alpha = 5\%$, $p = 0,05$)
–	einer Wahrscheinlichkeit für den Fehler 2. Art von
	0,20 (Power 0,80)
–	einer biologisch relevanten Differenz

Angaben zur biometrischen Planung im Antrag

- Antragsformular fragt die Angaben zur biometrischen Planung einzeln ab.

4. Fehler 1. Art

5. Fehler 2. Art

6. Differenz

7. Varianz oder Effektstärke

Wie groß ist die Streuung der Messgröße

(unter Kontroll- und Versuchsbedingungen)??

Ohne Kenntnis der Streuung der Messgröße ist die biometrische Planung auf der Basis von konkreten Zahlen nicht möglich.

Die vorgesehene Tierzahl und Gruppengröße ist notwendig zur statistischen Absicherung mit	
–	einer Wahrscheinlichkeit für den Fehler 1. Art von
	0,05 ($\alpha = 5\%$, $p = 0,05$)
–	einer Wahrscheinlichkeit für den Fehler 2. Art von
	0,20 (Power 0,80)
–	einer biologisch relevanten Differenz
	zwischen Kontroll- und Versuchsgruppe(n)
–	Varianz oder Effektstärke (mit Angabe des genutzten Parameters, z.B. Effektstärke nach Cohen)

Angaben zur biometrischen Planung im Antrag

- Antragsformular fragt die Angaben zur biometrischen Planung einzeln ab.

- 4. Fehler 1. Art
- 5. Fehler 2. Art
- 6. Differenz
- 7. Varianz oder Effektstärke (Verhältnis von Differenz zu Streuung)

Die vorgesehene Tierzahl und Gruppengröße ist notwendig zur statistischen Absicherung mit	
–	einer Wahrscheinlichkeit für den Fehler 1. Art von
	0,05 ($\alpha = 5\%$, $p = 0,05$)
–	einer Wahrscheinlichkeit für den Fehler 2. Art von
	0,20 (Power 0,80)
–	einer biologisch relevanten Differenz
	zwischen Kontroll- und Versuchsgruppe(n)
–	Varianz oder Effektstärke (mit Angabe des genutzten Parameters, z.B. Effektstärke nach Cohen)
	Verhältnis von Differenz zu Streuung

- Die biometrische Berechnung der Gruppengrößen kann nun mit Hilfe der hier angegebenen Parameter erfolgen.

Fallzahlplanung durch Poweranalyse

- Statt Differenz und Varianz kann auch die Effektstärke für die biometrische Planung herangezogen werden.
- Wie deutlich müssen sich zwei Gruppen in einer untersuchten Variable unterscheiden, damit dieser Unterschied wissenschaftlich relevant ist?
- Die Effektstärke hängt vom Abstand der Gruppenmittelwerte und von der Varianz der beiden Gruppen ab.
- Sie beurteilt die Frage der „praktischen Bedeutsamkeit“ von beobachteten Unterschieden zwischen zwei oder mehreren Versuchs- und Kontrollgruppen.

Fallzahlplanung durch Poweranalyse

- Die Effektstärke beurteilt die Frage der „praktischen Bedeutsamkeit“ von beobachteten Unterschieden zwischen zwei oder mehreren Versuchs- und Kontrollgruppen.

Anmerkung:
Die Berechnung der Effektstärke nach

- Wie deutlich müssen sich zwei Gruppen in einer untersuchten Variable unterscheiden, damit dieser Unterschied wissenschaftlich relevant ist?
- $$\text{Effektstärke} = \frac{\text{Differenz zwischen den Gruppen}}{\text{Standardabweichung}}$$

- Die Effektstärke hängt vom Abstand der Gruppenmittelwerte und von der Varianz ab.
gilt in dieser einfachen Weise nur für den t-Test. Für andere statistische Tests (ANOVA, U-Test, Kruskal-Wallis-Test, usw.) wird die Effektstärke anders berechnet. Die Effektstärken aus dem t-Test dürfen also nicht für andere Tests übernommen werden.

Fallzahlplanung durch Poweranalyse

- Die Effektstärke beurteilt die Frage der „praktischen Bedeutsamkeit“ von beobachteten Unterschieden zwischen zwei oder mehreren Versuchs- und Kontrollgruppen
- Aber: die Effektstärke d für den t-Test kann für andere statistische Tests umgerechnet werden

Effect Size	d	1
d		1
r		0,442
η^2		0,2
f		0,5
Odds Ratio		6,1337
Number Needed to Treat (NNT)		1,9212

https://www.psychometrica.de/effect_size.html#transform

Fallzahlplanung durch Poweranalyse

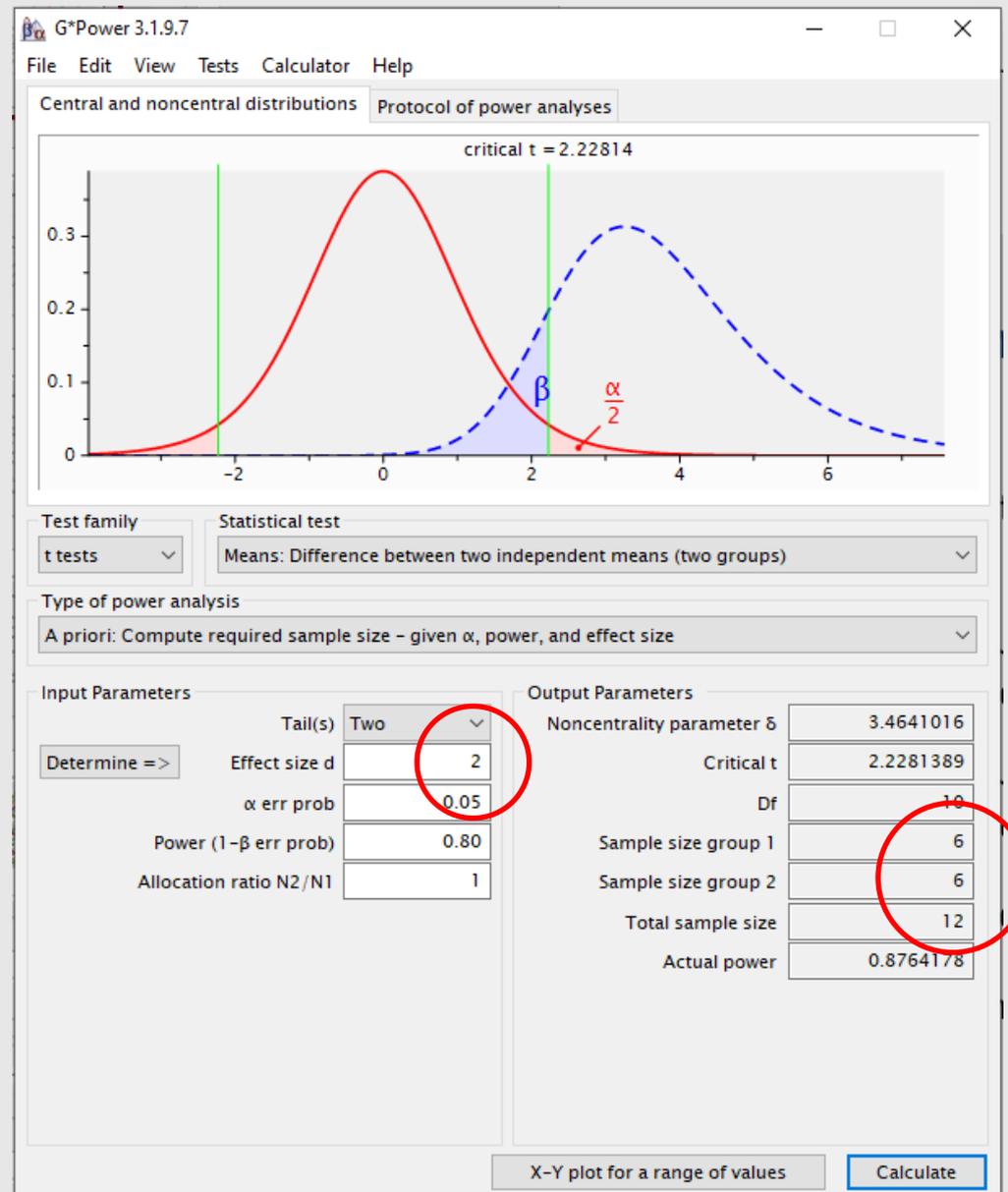
- Im Genehmigungsantrag muss dargelegt werden, warum die beantragte Tierzahl die wissenschaftlich unerlässliche Mindestanzahl ist, die ein statistisch abgesichertes Ergebnis liefert und gleichzeitig die Vorgabe der 3R erfüllt.
- Die Poweranalyse erfüllt diese Forderung, denn mit Hilfe der Parameter
 - Wahrscheinlichkeit für den Fehler 1. Art, α
 - Wahrscheinlichkeit für den Fehler 2. Art, β bzw. Power ($= 1 - \beta$)
 - Biologisch relevante Differenz
 - Varianz oder Effektstärke

lässt sich die Frage beantworten:

Wie groß muss die Stichprobe mindestens sein, damit bei vorgegebenem Signifikanzniveau α eine gegebene Effektstärke mit einer Wahrscheinlichkeit von $(1 - \beta)$ auch erkannt wird?

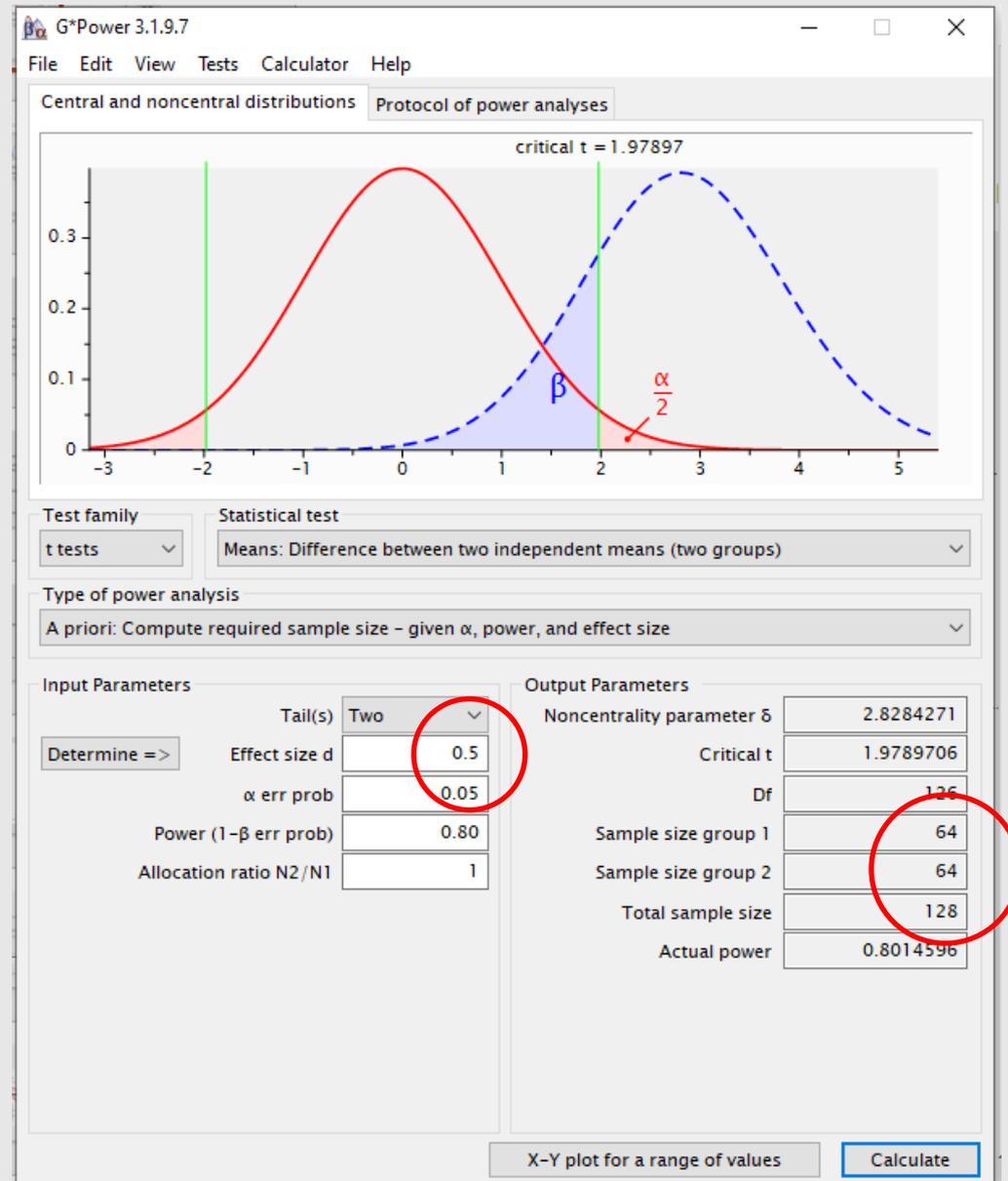
Fallzahlplanung durch Poweranalyse

- Effektstärke bestimmt die Gruppengröße
- große Differenz, geringe Streuung (= Effektstärke hoch) → kleine Gruppen



Fallzahlplanung durch Poweranalyse

- Effektstärke bestimmt die Gruppengröße
 - große Differenz, geringe Streuung (= Effektstärke hoch) → kleine Gruppen
 - kleine Differenz, hohe Streuung (= Effektstärke niedrig) → große Gruppen
- relevante Differenz vorgeben
- Streuung vermindern



Fallzahlplanung durch Poweranalyse

- Im Genehmigungsantrag muss dargelegt werden, warum die beantragte Tierzahl die wissenschaftlich unerlässliche Mindestanzahl ist, die ein statistisch abgesichertes Ergebnis liefert und gleichzeitig die Vorgabe der 3R erfüllt.
Anmerkung:
• Von der ermittelten Gruppengröße sollte nicht abgewichen werden.
– Kleinere Gruppengrößen (um Tiere einzusparen) vermindern die statistische Aussagekraft:
– Wahrscheinlichkeit für den Fehler 2. Art, β bzw. Power ($= 1 - \beta$)
– Biologisch relevante Differenz
– Varianz oder Effektstärke
– lässt sich die Frage beantworten:
Wie groß muss die Stichprobe mindestens sein, damit bei vorgegebenem Signifikanzniveau α eine gegebene Effektstärke mit einer Wahrscheinlichkeit von $(1 - \beta)$ auch erkannt wird?

Fallzahlplanung durch Poweranalyse

- Der wesentliche Vorteil der Poweranalyse ist:
 - Hat man die optimale Gruppengröße vorher bestimmt und führt nun die Datenerhebung und Auswertung durch, kann das erhaltene Ergebnis immer eindeutig interpretiert werden:
 - (i) Es liegt ein signifikantes Testresultat vor:
Die Alternativhypothese H_1 gilt mit einer Fehlerwahrscheinlichkeit von höchstens α .
 - (ii) Es liegt ein nichtsignifikantes Testresultat vor:
Die Nullhypothese H_0 gilt mit einer Fehlerwahrscheinlichkeit von höchstens β .

Fehler bei der statistischen Aussage

- Fehlinterpretation von p-Werten (Signifikanzen):

p-Werte sind kein objektives Maß für die Richtigkeit einer Hypothese:

„Absence of evidence is no evidence of absence.“

- Das Signifikanzniveau ist eine willkürliche Größe, die aus Tradition gilt, nicht aufgrund wissenschaftlicher Gesetzmäßigkeiten.
- Ausreißer nicht einfach verwerfen (trotz *p-hacking*)
 - Analyse mit und ohne Ausreißer durchführen
 - Nicht-parametrische Verfahren anwenden
- prozentuale Angaben zu Absolutwerten in Beziehung setzen
(eine 50%ige Steigerung eines sehr kleinen Effektes ist u.U. immer noch ein sehr kleiner Effekt)

Statistische Verfahren zur Minimierung der Tierzahlen

Matthias Schmidt

- Kontakt für Rückfragen und Kommentare:

Matthias Schmidt

Tierschutzbeauftragter der Ruhr-Universität Bochum

Matthias.Schmidt@ruhr-uni-bochum.de