

Statistik I

Einheit 11: Effektstärke und Power

23.01.2024 | Prof. Dr. Stephan Goerigk

Effektstärke und Powerberechnung

Was ist eine Effektstärke

- Indikator für die Stärke eines Treatments oder Phänomens, ähnlich wie der p-Wert ein Indikator für die Wahrscheinlichkeit eines beobachteten Wertes ist.
- Effektstärken sind ein wichtiger Faktor in empirischen Studien
- können verwendet werden, um die Stichprobengröße für künftige Studien zu bestimmen und die Stärke des Effekts über mehrere Studien hinweg zu vergleichen.
- Sind standardisierte Maße (unabhängig von der Einheit der Variablen)

Effektstärken können in zwei Kategorien eingeteilt werden:

- d-Familie, die Unterschiede zwischen Gruppen betrachtet
- r-Familie, die den Zusammenhang zwischen Daten misst.

Effektstärke und Powerberechnung

Effektstärke und Power

- Eine größere Teilnehmerzahl in einem Experiment führt zu einer höheren Wahrscheinlichkeit, dass die Null-Hypothese abgelehnt wird und das Ergebnis statistisch signifikant ist.
- Je eher dies geschieht, desto größer ist die Power des statistischen Verfahrens.
- Ein Testverfahren mit einer hohen Power hat eine geringe Wahrscheinlichkeit, einen Effekt zu übersehen, auch wenn dieser klein ist.
- Andererseits wird bei einem Verfahren mit einer hohen Teststärke jeder relativ kleine Effekt schnell signifikant, auch wenn er keine praktische Bedeutung hat.

→ Unter Statistikern existiert daher auch die Meinung nicht mehr nur p-Werte, sondern verstärkt Effektstärken zu betrachten, um den Nutzen von Ergebnissen einzuordnen

Effektstärke und Powerberechnung

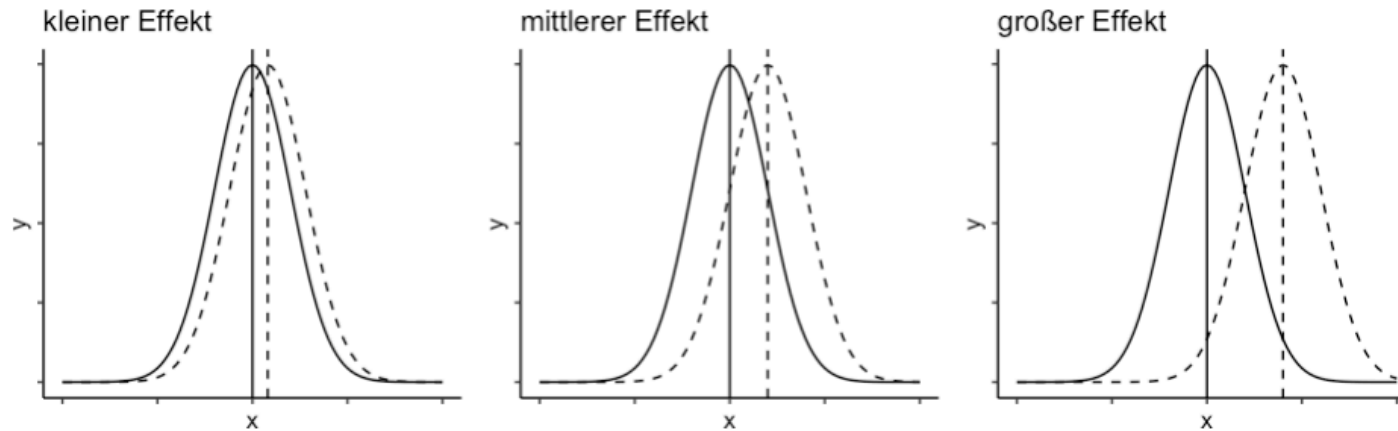
Cohen's d

- Cohen's d ist die bekannteste und gängigste Methode zur Bestimmung der Effektstärke.
- Es wird verwendet, um den Effekt von Studien zu vergleichen, auch wenn die abhängige Variable unterschiedlich gemessen wird.
- Es ähnelt genau der Formel für die z-Standardisierung und wird daher von einigen Autoren auch als standardisierte Effektstärke bezeichnet.
- Kann alle reelle Zahlen von $-\infty$ bis $+\infty$ annehmen
- Das Vorzeichen gibt zwar die Richtung des Effekts an, spielt allerdings für die Interpretation der Stärke des Effekts keine Rolle

$$d = \frac{\mu_1 - \mu_2}{\sigma}$$

Effektstärke und Powerberechnung

Cohen's d



- Für Cohen's d definiert sich die Effektstärke über die Mittelwertsdifferenz
- diese wird relativ zur Streuung angegeben

Effektstärke und Powerberechnung

Cohen's d - Probleme der Standardabweichung

- Es gibt ein Problem bezüglich der Definition der Standardabweichung in Cohen's Originalarbeit (Cohen, 1988).
- In der Publikation wird nicht angegeben, wie die Standardabweichung genau definiert werden soll.
- Cohen schreibt, dass man entweder eine der beiden Standardabweichungen für σ verwenden kann, da angenommen wird, dass beide gleich sind.
- Andere Forscher definieren Cohen's d expliziter und nutzen den Mittelwert beider Varianzen (gepoolte Standardabweichung) als Wurzel.
- manche Autoren verwenden die Notation d_s für Cohen's d mit gepoolter Standardabweichung

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$$

t-Test

Cohen's d - Beispiel (bekannt von t-Test Vorlesung)

- Eine Psychologiestudentin möchte Versuchspersonen nach Teilnahme an ihrem Experiment mit einem kleinen Geldbetrag vergüten.
- Sie möchte wissen, ob in Abhängigkeit der Höhe des Betrags Unterschiede in der Zufriedenheit der Teilnehmer:innen bestehen.
- Höhere Werte zeigen höhere Zufriedenheit an.
- Ihre Ergebnisse stehen in der nachfolgenden Tabelle:

Gruppe 1: erhält 10€	Gruppe 2: erhält 50€
4	6
2	4
3	7
5	8
4	8

- $\bar{x}_{G1} = 3.6$
- $\bar{x}_{G2} = 6.6$
- $s_{G1}^2 = 1.3$
- $s_{G2}^2 = 2.8$

t-Test

Cohen's d - Beispiel (bekannt von t-Test Vorlesung)

- $\bar{x}_{G1} = 3.6$
- $\bar{x}_{G2} = 6.6$
- $s_{G1}^2 = 1.3$
- $s_{G2}^2 = 2.8$

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$$
$$d = \frac{3.6 - 6.6}{\sqrt{\frac{1.3 + 2.8}{2}}} = \frac{3}{1.43} = 2.10$$

Nach Cohen liegt ein großer Effekt vor (s.u.).

Effektstärke und Powerberechnung

Hedge's g (oft auch einfach nur d)

- Eine Variation von Cohen's d ist Hedge's g, welche ähnlich berechnet wird.
- Der Unterschied besteht darin, dass bei der Berechnung von Hedge's g die Varianzen nicht nur miteinander verbunden, sondern auch durch die Bessel-Korrektur (n-1) korrigiert werden.
- Diese Korrektur reduziert insbesondere bei kleinen Stichproben ($n < 20$) den Fehler der Schätzung.

$$g = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1) \cdot s_1^2 + (n_2-1) \cdot s_2^2}{n_1+n_2-2}}}$$

Effektstärke und Powerberechnung

Glass's Δ

- Glass's delta gehört zu der gleichen Familie, wie die bereits vorgestellten Formeln von Cohen und Hedge.
- Glass's Δ wurde ursprünglich für experimentelle Studien entwickelt. Anstatt eine gepoolte oder Besselkorrigierte Standardabweichung zu verwenden, wird einfach die Standardabweichung der Kontrollgruppe verwendet.
- Die Idee dahinter ist, dass die Kontrollgruppe kein Treatment bekommen hat und daher die Standardabweichung der Grundgesamtheit eher wiedergibt.

$$\Delta = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{control}}$$

Effektstärke und Powerberechnung

Cohen's d berechnet aus der t-Statistik

Zwischen dem t-Wert bei unabhängigen Stichproben und Cohen's d besteht ein Zusammenhang

Somit können wir einfache Art und Weise Cohen's d aus dem t-Wert berechnen

Dies ist insbesondere bei Meta-Analyse nützlich, wenn wir keine Originaldaten, sondern die Ergebnisse aus bereits gerechneten Hypothesentests vorliegen haben

$$d = \frac{2t}{\sqrt{df}}$$

Effektstärke und Powerberechnung

Unterschiedliche Bezeichnung für Maße der Effektstärke

- Vorsicht: Unterschiedliche Bezeichnung für Maße der Effektstärke!
- Es gibt eine hohe Verwirrung bei Autoren wissenschaftlicher Literatur bezüglich der Verwendung von Terminologie für bestimmte Formeln (Ellis, 2010).
- Die Verwendung der Bezeichnungen ist oft inkonsistent und widersprüchlich.
- Manche Autoren bezeichnen etwas als "g", was von Hedge und Olkin als "d" bezeichnet wurde und umgekehrt.
- Dies liegt daran, dass viele dieser Effektstärken-Maße ihren Ursprung in den 1980er Jahren haben, als "g" das Standardmaß der Effektstärke war, das von Cohen erfunden und von Glass verfeinert wurde (g steht für Glass).
- Allerdings wird seitdem "g" synonym mit der Formel von Hedge verwendet - nicht mit Glass's "g". (Deswegen heißt es auch Hedge's "g" und nicht Hedge's "h", es wurde nach Glass benannt, obwohl es von Hedge entwickelt wurde.)

Effektstärke und Powerberechnung

Interpretation von d

- Die drei wichtigsten Indizes der d-Familie (Cohen's, Glass's Δ und Hedge's g) vermitteln Informationen über die Größe des Effekts bezogen auf die Standardabweichung.
- Ein Wert von .50 entspricht einer halben Standardabweichung, während ein Wert von .75 dreiviertel einer Standardabweichung entspricht.
- Je höher der Effektstärkenwert, desto größer ist der Effekt. Effektstärken sind standardisiert, was bedeutet, dass sie über verschiedene Studien hinweg verglichen werden können.
- Wenn zwei Studien jeweils eine Effektstärke von .40 angeben, haben sie Effekte in gleicher Größe gefunden.

Effektstärke und Powerberechnung

Interpretation von d - Faustregeln

- Cohen (1988) lieferte auch gleich eine Faustregel zur Interpretation der Ergebnisse aus seiner Formel.
- Demnach fängt ein kleiner Effekt ab 0.2 an, ein mittlerer ab 0.5 und ein großer ab 0.8.
- Hier ist zu beachten, dass Cohen's d durchaus auch negativ werden kann.
- Dies bedeutet lediglich, dass der Effekt in die umgekehrte Richtung geht.
- Die Faustregeln hierfür bleiben die gleichen.
- Die Formel nach Cohen ist allerdings anders zu interpretieren als der Korrelationskoeffizient: sie kann auch Werte größer 1 oder kleiner -1 einnehmen (Cohen's d ist definiert von $-\infty$ bis $+\infty$). In diesem Fall handelt es sich einfach nur um einen ausgesprochen großen Effekt.

Effektstärke	d
Kleiner Effekt	≥ 0.2
Mittlerer Effekt	≥ 0.5
Großer Effekt	≥ 0.8

Effektstärke und Powerberechnung

Interpretation von d - Vorsicht bei Faustregeln

- Es ist bequem, die Cohen-Regeln zur Interpretation der Effektstärke zu verwenden, aber es gibt Situationen, in denen kleine Effektstärken dennoch große Auswirkungen haben können.
- Thompson (2007) empfiehlt, in diesen Fällen nicht die Regeln von Cohen zu nutzen, sondern stattdessen direkt und explizit die berechnete Effektstärke mit Effektstärken aus vergleichbaren Studien zu vergleichen.
- Als Beispiel nennt er folgende Interventionen
 - schwer zu behandelnde Krankheiten wie Essstörungen
 - Veränderungen, die sich nur allmählich über einen längeren Zeitraum auswirken, wie eine Intervention in einem Kindergarten, die zwar zum damaligen Zeitpunkt nur geringe Auswirkungen hatte, aber sich später positiv auf das gesamte Leben der Kinder auswirken wird.

Effektstärke und Powerberechnung

Korrelationskoeffizienten

- r quantifiziert die Stärke und Richtung des Zusammenhangs zwischen zwei Variablen.
- Im Gegensatz zu Cohen's d , ist der Wertebereich von r beschränkt.
- r kann Werte zwischen -1 und 1 annehmen.
 - Ein Wert von 1 bedeutet eine perfekte positive Beziehung zwischen zwei Variablen
 - Ein Wert von -1 bedeutet eine perfekte negative Beziehung zwischen zwei Variablen
 - 0 hingegen bedeutet, dass gar keine Beziehung zwischen den Variablen besteht.

WICHTIG:

- Der Korrelationskoeffizient r und darauf basierende Maße sind standardisiert.
- Das bedeutet, dass ein Effekt, der in r angegeben wurde, mit jedem anderen verglichen werden kann.
- r kann demnach direkt als Effektstärke genutzt werden

r	Interpretation
>0.1	kleiner Effekt
>0.3	mittlerer Effekt
>0.5	großer Effekt

Effektstärke und Powerberechnung

Cohen's d in den Korrelationskoeffizienten r umrechnen

- McGrath und Meyer (2006) publizierten eine Formel zum Umrechnen von d in den Korrelationskoeffizienten r , bei ungleicher Größe der Gruppen n_1 und n_2 :

$$r = \sqrt{\frac{d^2}{d^2 + \frac{(n_1+n_2)^2 - 2 \cdot (n_1+n_2)}{n_1 \cdot n_2}}}$$

Effektstärke und Powerberechnung

Aufgeklärte Varianz

- Der Beziehung zwischen zwei Variablen wird oft in Form der aufgeklärten Varianz, auch als gemeinsame Varianz bekannt, beschrieben.
- Maße für aufgeklärte Varianz weisen oft eine Potenz von 2 auf (²).
- Beispielsweise kann die aufgeklärte Varianz bei einer Korrelation aus dem Korrelationskoeffizienten berechnet werden, indem dieser quadriert wird: R^2 .
 - Dies wird auch als Determinationskoeffizient bezeichnet.
 - Wenn zwei Variablen beispielsweise eine Korrelation von -.40 haben, beträgt der Determinationskoeffizient $(-.40)^2 = .16$.
- Eine aufgeklärte Varianz von .16 bedeutet, dass 16% der Streuung durch die Beziehung zwischen den beiden Variablen erklärbar sind.
- Je mehr Varianz die Variablen gemeinsam haben, desto stärker kovariieren sie und desto stärker ist der Effekt sowie die aufgeklärte Varianz.
- Maße für die aufgeklärte Varianz bewegen sich zwischen 0 und 1.

Visualisierung Aufgeklärte Varianz

Effektstärke und Powerberechnung

Referenzwerte verschiedener Maße der Effektstärke - Referenzwerte von Cohen, abgewandelt nach Ellis (2010).

Testart	Maß der Effektstärke	Klein	Mittel	Groß
Vergleich zweier unabhängiger Stichproben	d, Δ , Hedge's g	≥ 0.20	≥ 0.50	≥ 0.80
Vergleich von zwei Korrelationen	q	0.1	0.3	0.5
Korrelationskoeffizient	r	≥ 0.10	≥ 0.30	≥ 0.50
Determinationskoeffizient	r^2	0.01	0.09	0.25
Kreuztabelle	Cohen's w, ϕ , Cramer's V, C	≥ 0.10	≥ 0.30	≥ 0.50
ANOVA	f	0.1	0.25	0.4
η^2_{part}	η^2_{part}	0.01	0.06	0.14
Multiple Regression	R^2	0.02	0.13	0.26
	Cohen's f^2	0.02	0.15	0.35

Effektstärke und Powerberechnung

Statistische Power (Teststärke)

- Statistische Power ist die Wahrscheinlichkeit, dass ein Effekt entdeckt wird, wenn der Effekt auch tatsächlich existiert.
- (Statistische) Power wird definiert als die Wahrscheinlichkeit, korrekterweise eine falsche Nullhypothese zurückzuweisen.
- Im Deutschen wird auch oft der Begriff Teststärke verwendet, der Synonym mit statistischer Power ist.
- Wenn die statistische Power hoch ist, sinkt die Wahrscheinlichkeit, einen Typ-II-Fehler zu begehen oder festzustellen, dass es keinen Effekt gibt, wenn es tatsächlich einen gibt.
- $\text{Power} = 1 - \beta$ (β = die Wahrscheinlichkeit ist, einen Fehler 2. Art zu begehen)
- Statistische Power ist eine Funktion der Effektstärke, der Stichprobengröße (N), des verwendeten statistischen Verfahrens und des Signifikanzniveaus (Wahrscheinlichkeit einen Fehler 1. Art zu begehen).
- In den letzten 20 Jahren hat die statistische Power und ihre Bedeutung für wissenschaftliche Studien mehr Aufmerksamkeit erhalten, und mehr Journals, Stakeholder und Reviewer verlangen von Forschern Power-Analysen.

Effektstärke und Powerberechnung

Poweranalyse

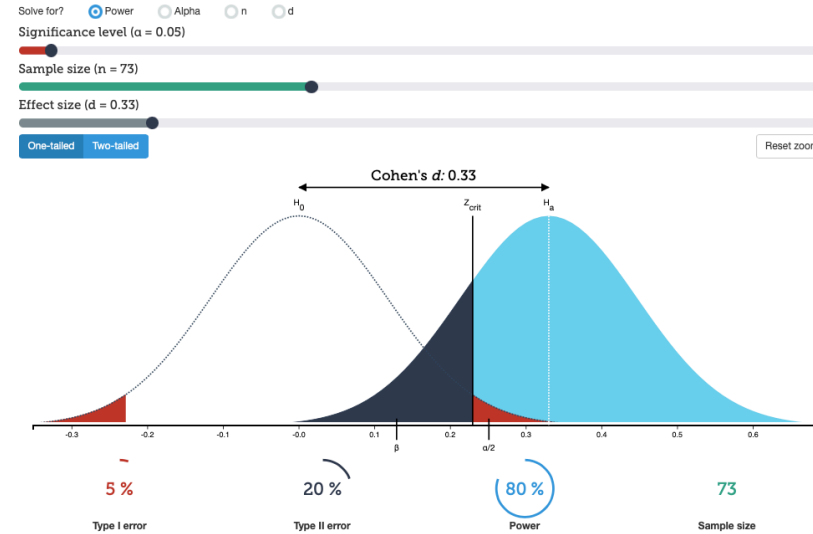
- Poweranalyse ist eine Methode zur Schätzung der Stichprobengröße in psychologischen Studien, um eine ausreichende statistische Power zu gewährleisten.
- Sollte vor der Durchführung einer Studie durchgeführt werden, um eine angemessene Stichprobengröße zu bestimmen und die Güte der Studie zu verbessern.
- Es bestimmt, wie groß die Stichprobe sein muss, um bei angenommener Effektstärke signifikante Ergebnisse zu erzielen und das Risiko eines Type-II-Fehlers zu minimieren.
- Folgende Faktoren werden bei einer Poweranalyse immer berücksichtigt:
 - Signifikanzniveau (i.d.R. $\alpha = .05$)
 - gewünschte Power $1 - \beta$ (i.d.R. mindestens 0.8)
 - vermutete **Effektstärke**

→ Zur Planung einer neuen Studie, sollte eine angenommene Effektstärke mit berücksichtigt werden

→ Problem: Effektstärke aus Vorstudien (oder Pilotstudien) nicht immer verfügbar (manchmal muss einfach geschätzt werden)

Effektstärke und Powerberechnung

Poweranalyse



Visualisierung Poweranalyse

Effektstärke und Powerberechnung

Poweranalyse

- Eine Poweranalyse wird in der Regel computersisiert berechnet
- Für gängige statistische Tests wird of folgende Software eingesetzt
 - Das Programm G*Power
 - frei verfügbar von der Universität Düsseldorf ([Link](#))
 - Das R Paket "pwr"
- Komplexere Verfahren wie Mehrebenenmodelle oder multivariate Verfahren (z.B. Strukturgleichungsmodelle) erfordern Simulation (z.B. Monte-Carlo Simulation)

Effektstärke und Powerberechnung

Poweranalyse - Berechnung mit G*Power

