

Häufige Missverständnisse über den p-Wert

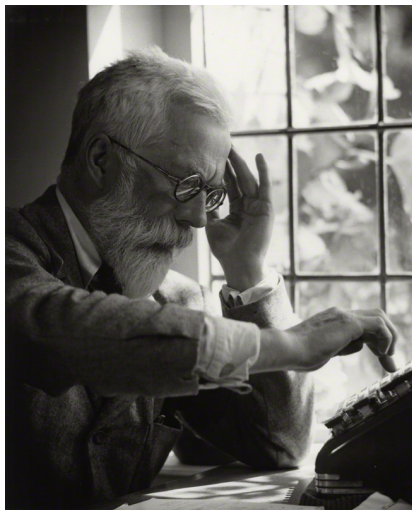
Dr. med. Andreas Mock, MSc, MPhil

Nationales Centrum für Tumorerkrankungen (NCT) Heidelberg

28.11.2017



Der Erfinder des p-Werts



Sir Ronald Fisher (1890-1962)
Gonville & Caius College, Cambridge

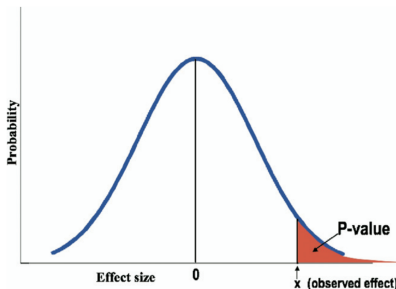
„Personally, the writer prefers to set a low standard of significance at the 5 percent point [...] A scientific fact should be regarded as experimentally established only if a properly designed experiment rarely fails to give this level of significance.“

Statistical Methods for Research
Workers, 1926

Definition des p-Wertes

Wahrscheinlichkeit das gleiche Stichprobenergebnis oder ein noch extremeres zu erhalten, wenn die Nullhypothese wahr ist.

Algebraische Definition: $P(X \geq x \mid H_0)$ wobei X eine Zufallsvariable und x der beobachtete Wert in den Daten ist



Goodman, 2008

#1 | Wenn $p < 0.05$, ist die Nullhypothese nur in 5% wahr

Dies ist die wohl **häufigste Fehlinterpretation** des p-Wertes.

Der p-Wert wird unter der Annahme berechnet, dass die Nullhypothese zutrifft ($P(\text{Daten} \mid H_0)$), er kann daher nicht gleichzeitig die Wahrscheinlichkeit sein, dass die Nullhypothese zutrifft ($P(H_0 \mid \text{Daten})$).

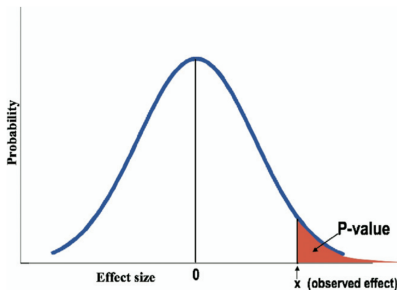
Beispiel: Die Wahrscheinlichkeit drei Mal hinter einander Kopf beim Münzwurf zu erhalten ist $p=0.125$. Dies bedeutet jedoch nicht, dass die Wahrscheinlichkeit, dass die Münze fair ist nur 12.5% beträgt.

#2 | $p > 0.05$ bedeutet, dass es keinen Unterschied zwischen den Gruppen gibt

Eine nicht signifikante Differenz bedeutet bloß, dass die beobachteten **Daten konsistent mit der Nullhypothese** sind und nicht, dass die Nullhypothese wahrscheinlicher ist.

#3 | $p=0.06$ ist substantiell schlechter als $p=0.04$

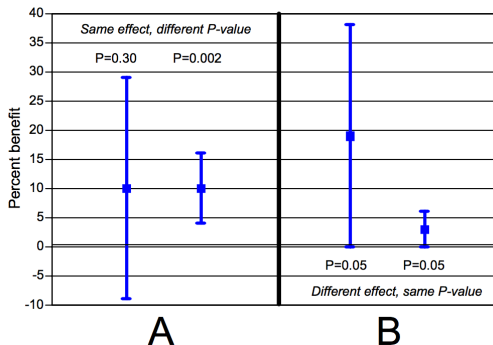
Fisher hat den p-Wert als **kontinuierliche Variable** eingeführt um abzuschätzen, ob ein Ergebnis es Wert ist weiter untersucht zu werden. Die magische p-Wert Grenze von 0.05 ist **völlig arbiträr**. p-Werte von 0.04 und 0.06 sind sehr ähnliche Wahrscheinlichkeiten!



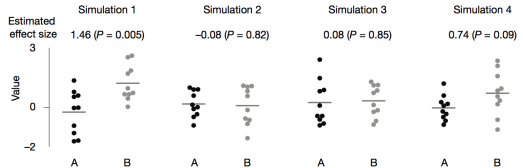
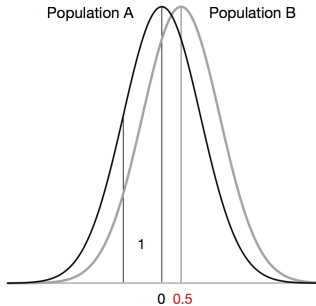
Goodman, 2008

#4 | Studien mit gleichem p-Wert zeigen eine ähnlich starke Effektgrösse

Der folgende Plot zeigt, dass dies nicht zutrifft. Der gleiche p-Wert kann einen völlig anderen Effekt indizieren (Fig. B). Umgekehrt, kann es einen identischen Effekt bei unterschiedlichem p-Wert geben (Fig. A):



#5 | $p=0.05$ bedeutet, dass man bei Wiederholung des Experiments in 5% ein nicht signifikantes Ergebnis erhält



Halsey et al., 2015

Nur bei einer **großen Effektgröße bzw. Power** (i.e. Gruppengröße) sind p-Werte bei Wiederholung des Experiments mit einer anderen Stichprobe reproduzierbar!

- ▶ **A Dirty Dozen: Twelve P-Value Misconceptions**
Goodman, S
Semin Hematol. 2008 Jul;45(3):135-40.
- ▶ **The fickle P value generates irreproducible results**
Halsey LG, Curran-Everett D, Vowler SL & Drummond GB
Nat Methods. 2015 Mar;12(3):179-85.