

(13) Power-Analyse: die Magie der Statistik – Oder: Der Unterschied zwischen Signifikanz und Relevanz

Normalerweise ist der Durchschnittsbürger und Durchschnittswissenschaftler zufrieden, wenn er hört, ein Forschungsergebnis sei „statistisch signifikant“ gewesen. Wir meinen dann landläufig: die Hypothese, mit der man an die Forschung ging, sei belegt, das Faktum, das man untersucht bewiesen. Und umgekehrt, findet man kein signifikantes Ergebnis, glaubt man, das in Frage stehende Phänomen sei nicht gefunden, also nicht existent. Deswegen glaubt z.B. der Durchschnittsarzt, -journalist und -bürger die Bioresonanz sei als unwirksam belegt und Homöopathie ist Placebo, und halb Amerika nimmt Lipidsenker zur Primärprävention von Herzinfarkt, weil man glaubt das sei eine wissenschaftlich bewiesene Tatsache.

Ich will in diesem Kapitel ein paar von diesen Meinungen genauer unter die Lupe nehmen und zeigen, warum sie entstanden sind und die Frage stellen, wie berechtigt sie sind. Es wird sich herausstellen: es hat mit dem zu tun, was ich die Magie der Statistik nenne. Das ist die Frage, wie mächtig ein statistischer Test ist. Die hängt zusammen mit der Frage, wie groß der Effekt ist, den wir untersuchen. Und davon hängt ab, wie groß die Stichprobe ist, die wir benötigen, um den Effekt wirklich statistisch sichtbar zu machen, oder ein signifikantes Ergebnis zu erhalten. Anders ausgedrückt: Wenn es einen systematischen Effekt gibt, egal wie groß er ist, dann lässt er sich mit einer Untersuchung auch belegen, vorausgesetzt, wir haben genügend Ressourcen.

Die Frage, die sich jeder Leser einer wissenschaftlichen Untersuchung stellen sollte ist nicht: Ist eine Studie signifikant? Sondern: Ist der gezeigte Effekt, egal ob signifikant oder nicht, klinisch und systematisch von Bedeutung? Wenn er dann auch noch signifikant ist, können wir von einer wissenschaftlichen Bestätigung ausgehen. Wenn er nicht signifikant ist, müssen wir uns die Frage stellen: war die Größe der Studie geeignet, den Effekt zu finden? bzw. umgekehrt: wie groß müsste eine Studie sein, um einen Effekt von der gefundenen Größenordnung mit einigermaßen zufriedenstellender Sicherheit statistisch absichern zu können? Das ist die Essenz der Power-Analyse, um die es jetzt geht.

Wir haben es also in jeder wissenschaftlichen Untersuchung mit dem Spiel von insgesamt vier Größen zu tun, die voneinander abhängen wie die Teile eines filigranen Mobile. Wenn wir eines verändern, verändern sich alle anderen auch. Das wären:

1. Der Fehler erster Art oder der alpha-Fehler.
2. Der Fehler zweiter Art oder der beta-Fehler.
3. Die Größe des Effekts, oder die Effektgröße.
4. Die Größe der Studie oder die Anzahl von untersuchten Personen (im Falle von klinischen oder diagnostischen Studien) bzw. die Anzahl der Beobachtungen.

Aufgrund des Umfangs dieses Kapitels ist dieses in einem PDF zusammengefasst, lesen Sie bitte hier weiter:

(13) Power-Analyse: die Magie der Statistik – Oder: Der Unterschied zwischen Signifikanz und Relevanz

Walach Methodenlehre-Teil13 Poweranalyse

