t-Test für abhängige Stichproben

AUSGANGSSITUATION

Ziel ist - wie beim „t-Test unabhängig" - zwei Stichproben zu vergleichen. Der wesentliche Unterschied besteht darin, dass hier die Messwerte paarweise zusammengefasst werden können.

VORAUSSETZUNGEN

Wie beim Test für unabhängige Stichproben muss bei den Messungen mindenstens Intervallskala vorausgesetzt werden.

Die Stichproben müssen selbstverständlich *abhängig* sein. Deshalb lassen sich jeweils ein Wert der einen und ein Wert der anderen Stichprobe zu einem „Messwertepaar" verbinden. Bei diesem Test wird (siehe unten) nur mit den *Differenzen* der gepaarten Messwerte gerechnet.

Diese Differenzen müssen Normalverteilung aufweisen.

Varianzhomogenität muss hier nicht überprüft werden.

ABHÄNGIGKEIT DER STICHPROBEN

Zwei Stichproben heißen abhängig (auch: verbunden, korreliert), wenn je zwei Messwerte (x_{i1}, x_{i2}, also einer aus jeder Gruppe) paarweise verbunden sind. Typisch für abhängige Stichproben ist eine positive Korrelation zwischen den Messwertefolgen.

Der Zweck der Verwendung abhängiger Stichproben liegt darin, dass verhindert wird, dass versehentlich einer der Gruppen von vornherein tendenziell „bessere" Vpn zugeordnet werden als der anderen, was bei rein zufälliger Aufteilung passieren kann.

Außerdem ist bei abhängigen Versuchsanordnungen die Chance größer, einen vorhandenen Unterschied als signifikant zu erkennen, d.h. sie haben größere *Macht* als die entsprechenden unabhängigen Designs.

Näheres zur Macht von Tests, sowie zu Fehlern 1. und 2. Art ist im folgenden Abschnitt nachzulesen.

Es gibt verschiedene Methoden, abhängige Stichproben herzustellen. Die gebräuchlichsten unter ihnen sind:

Messwiederholung

Jede Vpn wird zwei Tests unterzogen, einmal unter Kontroll-, dann unter Versuchsbedingung.

Nachteil:
Häufig kommt es zu Lern- oder Ermüdungseffekten, die das Ergebnis des zweiten Testdurchgangs verfälschen.

Geschwister- oder Ehepaare

und dergleichen. Ein Spezialfall sind Zwillingsstudien.

Nachteil:
Dies ist nur in speziellen Ausgangssituationen sinnvoll. In vielen Bereichen haben Geschwister- und Ehepaare keine ähnlichen, sondern sogar entgegengesetzte Fähigkeiten und Interessen.

Eineiige Zwillingspaare weisen i.A. stärkere Ähnlichkeit auf als normale Geschwister, sind also für viele abhängige Versuchsanordnungen prädestiniert. Allerdings ist es äußerst schwierig, genügend Paare für Zwillingsstudien zu gewinnen.

Parallelisierung

Aufgrund von (meist durch einen Vortest ermittelten) Vorinformationen werden Paare von Vpn gebildet, die sich hinsichtlich relevanter Merkmale gleichen oder zumindest möglichst ähnlich sind. Diese werden zufällig auf die beiden Gruppen aufgeteilt, sodass einer der „Versuchszwillinge“ unter Kontroll-, der andere unter Versuchsbedingung getestet wird.

Nachteil:
Parallelisierung durch Vortests ist meist relativ aufwändig. Oft gibt es einfachere Wege, die Macht eines Tests zu erhöhen.

Vorgehen

Das Vorgehen unterscheidet sich dahingehend vom Test für unabhängige Stichproben, dass bei jedem Paar die Differenz der verbundenen Messwerte ($d_i = x_{i1} - x_{i2}$) gebildet wird. Ob ein signifikanter Unterschied zwischen den Stichproben besteht wird anhand des Mittelwerts ($\bar{d}$) dieser Differenzen überprüft.

Aufstellen der Hypothesen

$H_0 \ldots$ kein Versuchseffekt

$H_1 \ldots$ Versuchseffekt vorhanden (bzw. positiver/negativer Effekt vorhanden)

Üblicherweise wird der Versuchseffekt als α bezeichnet. (Dies ist für uns unangenehm, da Verwechslungsgefahr mit der Irrtumswahrscheinlichkeit α besteht.) Die Hypothesen können damit folgendermaßen formal angeschrieben werden:

$H_0 \ldots \alpha = 0$

$H_1 \ldots \alpha \neq 0$ (bzw. $\alpha > 0$ oder $\alpha < 0$)

Berechnen der Prüfgröße

Die Zufallsvariable

$$T = \frac{\bar{D}}{\hat{\sigma}_{\bar{D}}} = \frac{\bar{D}}{\frac{\hat{\sigma}_D}{\sqrt{n}}}$$

ist t-verteilt mit $df = n - 1$.

Die Prüfgröße stellt eine Realisierung dieser Variablen dar:

$$t = \frac{\bar{d}}{\frac{\hat{\sigma}_D}{\sqrt{n}}} = \frac{\frac{\sum_i d_i}{n}}{\frac{s_D}{\sqrt{n}}} = \frac{\sum_i d_i}{s_D \cdot \sqrt{n}}$$

Vergleich mit dem kritischen Wert

Ob der Unterschied zwischen den Stichproben signifikant ist und somit ein Versuchseffekt besteht, zeigt wie üblich der Vergleich mit dem kritischen t-Wert für ein- bzw. zweiseitige Testung, Irrtumswahrscheinlichkeit α und $df = n - 1$, der der Tabelle zu entnehmen ist.