

Methoden der Sportwissenschaft 2

Arbeitsblatt 2.6: Die Varianzanalyse als Methode der Wahl bei Experimenten (aktualisiert am 04.01.2007)

Einleitung

Bei Experimenten möchte man herausbekommen, ob eine bestimmte Maßnahme (Treatment) effektiv ist. Beispielsweise will man herausfinden, ob Sit-Ups am Schrägbrett eine effektive Trainingsform für die Bauchmuskulatur darstellen. Eventuell möchte man sogar prüfen, ob Unterschiede zwischen verschiedenen Trainingsübungen im Hinblick auf die Effektivität bestehen.

Ein Experiment zur Verbesserung der Maximalkraft der Bauchmuskulatur im Wintersemester 2001/2002 (vgl. Wydra, 2002) erbrachte folgendes Ergebnis (siehe Abbildung 1). Verglichen wurden die Effektivität von Sit-Ups am Schrägbrett mit einer Kontrollgruppe.

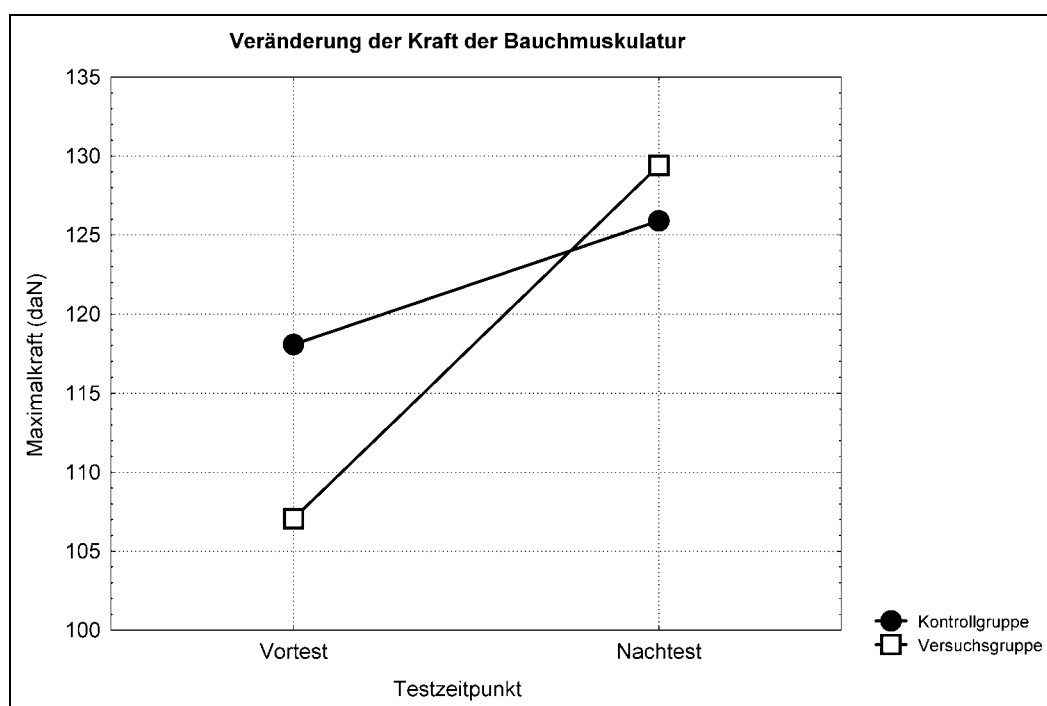


Abbildung 1: Ergebnis eines Trainingsexperiments zur Verbesserung der Kraft der Bauchmuskulatur.

Aus der Abbildung kann man ersehen, dass es sowohl in der Versuchs- als auch in der Kontrollgruppen offensichtlich zu einer Verbesserung gekommen ist. Es stellt sich aber die entscheidende Frage, ob die Verbesserung signifikant ist und ob sich die Experimentalgruppe stärker verbessert hat als die Kontrollgruppe. Denn nur das wäre ein Beleg für die Effektivität des Trainingsprogramms.

Mit den normalen statistischen Verfahren (t-Test, Wilcoxon- und U-Test) stößt man sehr schnell an Grenzen (siehe Abbildung 2). Man kann zwar die Unterschiede zwischen den Gruppen bzw. zwischen den Testzeitpunkten herausarbeiten, aber nicht die Frage beantworten, ob sich die Experimentalgruppe stärker verbessert hat als die Kontrollgruppe. Um solche Untersuchungsergebnisse statistisch abzusichern, hat sich die Varianzanalyse als Methode der Wahl etabliert.

	Vortest		Nachtest
Gruppe 1	normalverteilt	Wilcoxon-Test	nicht normalverteilt
	t-Test für unabhängige Stichproben		U-Test
Gruppe 2	normalverteilt	t-Test für gepaarte (abhängige) Stichproben	normalverteilt

Abbildung 2: Entscheidungsdiagramm zur Wahl der klassischen Verfahren der Inferenzstatistik

Zur Varianzanalyse

Im Allgemeinen besteht der Zweck der Varianzanalyse (ANOVA) darin, die Signifikanz von Mittelwertdifferenzen zu testen. Werden nur zwei Mittelwerte verglichen, liefert die Varianzanalyse dieselben Ergebnisse wie der t-Test für unabhängige Stichproben (beim Vergleich von zwei verschiedenen - unabhängigen Gruppen) oder der t-Test für gepaarte Stichproben (beim Vergleich von Vor- und Nachtestergebnissen). Es erscheint auf den ersten Blick möglicherweise seltsam, dass ein Verfahren zum Vergleich von Mittelwerten Varianzanalyse genannt wird. Diese Bezeichnung leitet sich jedoch aus der Tatsache ab, dass beim Testen der statistischen Signifikanz von Mittelwertdifferenzen eigentlich Varianzen (Streuungen) verglichen (d. h. analysiert) werden.

Die Welt ist naturgemäß kompliziert und multivariat. Fälle, in denen nur eine einzige Variable ein Phänomen vollständig erklärt, sind selten. Im obigen Beispiel sind zwei verschiedene Faktoren zu berücksichtigen, die einen Einfluss auf die Ausprägung der beobachteten Kraftwerte haben:

- der Testzeitpunkt (Vortest bzw. Nachtest)
- die Zugehörigkeit zu einer der beiden Gruppen (Versuchs- bzw. Kontrollgruppe).

In der Varianzanalyse kann **jeder einzelne Faktor isoliert** getestet werden, während der Einfluss aller anderen Faktoren ausgeblendet wird. Zum Nachweis eines signifikanten Effekts werden weniger Beobachtungen benötigt als bei einem einfachen t-Test. Das ist der eigentliche Grund dafür, warum die Varianzanalyse im statistischen Sinne mächtiger als der einfache t-Test ist.

Es gibt einen weiteren Vorteil der Varianzanalyse gegenüber einfachen t-Tests: Die Varianzanalyse ermöglicht die **Aufdeckung von Interaktionseffekten** bzw. Wechselwirkungen zwischen Variablen, d. h. dass ein Effekt durch einen anderen Effekt modifiziert wird. In unserem Beispiel kann die Frage beantwortet werden, ob sich die Experimentalgruppe stärker verbessert hat als die Kontrollgruppe.

Ein konkretes Beispiel

Eine für das obige Beispiel mit WinStat gerechnete Varianzanalyse mit Messwiederholung erbrachte folgendes Ergebnis.

Tabelle 1: Ergebnis der Varianzanalyse

Wiederholte Messungen

Messvariable: BAUCH1

BAUCH2

gruppiert nach: VGKG

	Quadrat-summe	Freiheits- grade	mittlere QS	F	p
Zwischen Variablen	5698,000385	1	5698,000385	33,72409155	4,35608E-07
Zwischen Fällen	135812,5776	50	2716,251551	16,07636185	2,44032E-18
Zwischen Gruppen	366,3978142	1	366,3978142	0,134890973	0,714963841
Interaktion	1364,352415	1	1364,352415	8,07503381	0,006478674
Fehler	8447,9672	50	168,959344		
Gesamt	151689,2954	103	1472,711606		

Das Ergebnis ist wie folgt zu beschreiben, wobei die Zeile „Zwischen den Fällen“ vernachlässigt werden kann, denn uns interessieren die Unterschiede zwischen den verschiedenen Gruppen und Messzeitpunkten:

1. Es bestehen hochsignifikante ($F=33,7$; $p<0,001$) Unterschiede zwischen den Variablen *Bauch1* und *Bauch2*. D. h. es kommt - über beide Gruppen gemittelt - zu einer hochsignifikanten Verbesserung der Kraft.
Beachte die Darstellung des p-Wertes: 4,35608E-07. Das „E-07“ bedeutet, dass das Komma um sieben Stellen nach links verschoben werden muss. Daraus resultiert der Wert $p=0,000\ 000\ 244\ 032$
2. Es bestehen keine signifikanten Unterschiede zwischen den Gruppen (der Versuchs- und der Kontrollgruppe), wenn man die Vor- und Nachttestergebnisse mittelt.
3. Es besteht eine sehr signifikante ($F=8,1$; $p<0,007$) Interaktion zwischen den beiden Hauptfaktoren (Variablen = Testzeitpunkt und Gruppenzugehörigkeit), d. h. die Verbesserung der Kraft fällt in der Kontrollgruppe kleiner aus als in der Versuchsgruppe.

Bestehen signifikante Unterschiede zwischen den Variablen oder zwischen den Gruppen, muss mit weitergehenden statistischen Verfahren geprüft werden, wo die Unterschiede begründet sind.

- Für den Fall signifikanter Unterschiede zwischen den Variablen (Messzeitpunkten) kann ein t-Test für gepaarte Stichproben oder bei mehr als zwei Messzeitpunkten eine Varianzanalyse mit Messwiederholung ohne Angabe einer Gruppierungsvariable gerechnet werden. Die Varianzanalyse führt bei nur zwei Messzeitpunkten zum gleichen Ergebnis wie der t-Test für gepaarte Stichproben.
- Für den Fall signifikanter Unterschiede zwischen den Gruppen kann ein t-Test für unabhängige (bzw. bei mehr als zwei Gruppen eine einfache Varianzanalyse) jeweils getrennt für die verschiedenen Messzeitpunkte gerechnet werden, die dann bei mehr als zwei Gruppen auch multiple Vergleiche zulässt.

Literatur

Statistica (2003). *Auszug aus dem elektronischen Handbuch des Statistikprogramms Statistica 6.1*. Tulsa, Ok: StatSoft, Inc. (siehe Materialien zu Methoden der Sportwissenschaft <http://www.uni-saarland.de/fak5/sportpaed/>)

Wydra, G. (2002). Sit-Ups am Schrägbrett: Untersuchungen zur Effektivität als Bauchmuskeltraining und zur Verkürzung der Hüftbeugemuskulatur. *Deutsche Zeitschrift für Sportmedizin*, 53, 285 - 290.