


Wann vergleicht man Mittelwerte mittels Varianzanalyse?

Ulrike Held

Horten-Zentrum, UniversitätsSpital, Zürich

Die Varianzanalyse, auch ANOVA (*Analysis of Variance*) genannt, stellt eine der am häufigsten verwendeten statistischen Techniken dar und wird auch im Bereich der medizinischen Forschung oft verwendet. Man prüft, ob zwei Vorgänge, die häufig gemeinsam auftreten oder die bei Veränderung des Umfelds ähnlich variieren, einen (kausalen) Zusammenhang haben. Möchte man z.B. untersuchen, ob sich die Hämoglobinwerte von Patienten mit unterschiedlicher Form von Sichelzellanämie unterscheiden, kann man dies mit Hilfe der Varianzanalyse tun. Da wir in diesem Beispiel nur eine erklärende Variable, nämlich den Typus der Sichelzellanämie, haben, spricht man von einfaktorieller (*one-way*) Varianzanalyse. Nähme man noch das Geschlecht als weiteren Faktor hinzu, bräuhete man eine zweifaktorielle (*two-way*) ANOVA. Die Varianzanalyse stellt einen Spezialfall der multiplen linearen Regression dar, die bereits zu Beginn dieser Artikelserie erläutert wurde.

In unserem Beispiel haben wir Hämoglobin-Messungen von 41 Patienten mit Sichelzellanämie (Tab. 1 ). Die Daten stammen aus einer im Jahr 1981 im *BMJ* veröffentlichten Studie von Anionwu et al. (in Kirkwood und Sterne, siehe empfohlene Literatur, p. 82).

Nun möchten wir also untersuchen, ob die mittleren Hämoglobinwerte sich in den drei Gruppen der Sichelzellanämie-Patienten unterscheiden. Dabei wird untersucht, welcher Anteil der Gesamtvariabilität der Zielgrösse (Hämoglobinwert) sich durch Unterschiede zwischen den Gruppen (Typ) erklären lässt. Bei der einfaktoriellen Varianzanalyse wird hierfür die sog. Summe der Abweichungen im Quadrat


$$SQA = \sum_{i=1}^n (x_i - \bar{x})^2 \quad i = 1, \dots, n$$

berechnet und anschliessend in zwei Komponenten zerlegt: (1.) die Summe der quadrierten Abweichungen durch Unterschiede *zwischen den Gruppenmittelwerten* und (2.) die Summe der quadrierten Abweichungen durch Unterschiede *zwischen den Beobachtungen* innerhalb jeder Gruppe. Man testet nun die *Nullhypothese*, dass kein Unterschied zwischen den Gruppenmittelwerten besteht, indem man (vereinfacht gesagt) die Variabilität *zwischen* den Gruppen und die Variabilität *innerhalb* der Gruppen vergleicht. Ist die Gruppenzugehörigkeit von Bedeutung, dann ist die Variabilität zwischen den Gruppen deutlich grösser als diejenige innerhalb der Gruppen.

Die Fragestellung für unser Beispiel ist also, ob Unterschiede in Hämoglobin-Niveaus zwischen den drei Sichelzelltyp-Gruppen bestehen. Nach Anwendung der ANOVA kommt in diesem Fall heraus, dass die Variation zwischen den Gruppen (= 49,945) viel grösser ist als die Variation innerhalb der Gruppen (= 0,999), was sich dann auch in einem sehr kleinen p-Wert widerspiegelt (p < 0,001). Wir können also davon ausgehen, dass die

Steady-state-Hämoglobin-Niveaus sich unterscheiden, je nachdem, an welchem Typ von Sichelzellanämie ein Patient erkrankt ist. Man kann noch die Gruppenmittelwerte berechnen und stellt fest, dass der Hämoglobinmittelwert in der Hb-SS-Gruppe am geringsten ist (8,713 g/dl), in der Hb-S/β-Thalassämie-Gruppe liegt der Mittelwert bei 10,630 g/dl, und am höchsten ist der Mittelwert in der Hb-SC-Gruppe mit 12,300 g/dl.

Bei der Varianzanalyse gibt es zwei wichtige zugrunde liegende Annahmen, die immer kontrolliert werden sollten. Die eine Annahme beinhaltet, dass die Zielgrösse gruppenweise *normalverteilt* sein muss, die andere sagt aus, dass die *Standardabweichung* der Individuen in allen Gruppen gleich sein sollte. Abweichungen von der Normalverteilungsannahme wiegen weniger schwer als Abweichungen von der Annahme der gleichen Standardabweichungen.

In unserem Beispiel kontrollieren wir die Annahme der Normalverteilung mit Hilfe eines sog. Quantil-Quantil-Plots (auch QQ-Plot), von statistischen Tests auf Normalverteilung raten wir generell eher ab. Abbildung 1  zeigt, wie die beobachteten Quantile der Zielgrösse mit den theoretischen Quantilen einer Normalverteilung übereinstimmen.

Man kann anhand von Abbildung 1 erkennen, dass die Annahme der gruppenweisen Normalverteilung der Hämoglobinwerte gut zu rechtfertigen ist, da die Punkte relativ genau auf einer Geraden liegen. Die geschätzten Standardabweichungen der drei Gruppen betragen 0,844 (Hb-SS), 1,284 (Hb-S/β-Thalassämie) und 0,942 (Hb-SC). Wir können somit davon ausgehen, dass sie sich ähnlich sind. Falls sich die Standardabweichungen zu stark unterscheiden sollten, gibt es Methoden, damit umzugehen. Diese könnten z.B. eine Transformation der Zielgrösse sein.

In diesem Beispiel haben wir einen Fall von einfaktorieller Varianzanalyse vorgestellt. Falls man mehrfaktorielle Varianzanalysen berechnen möchte, sollte darauf geachtet werden, dass die Anzahl der einzelnen Individuen pro Gruppe nicht zu klein wird.

Zu Beginn dieses Artikels hatten wir angedeutet, dass die Varianzanalyse ein Spezialfall der im ersten Artikel dieser Serie beschriebenen linearen Regressionsanalyse ist. Darauf möchten wir noch etwas näher eingehen. Das lineare Regressionsmodell wird verwendet, um einen Zusammenhang zwischen einem z.B. stetigen Einflussfaktor und der Zielgrösse herzustellen und den Effekt der Einflussgrösse zu schätzen. Hat man stattdessen einen kategoriellen Einflussfaktor mit typischerweise mindestens drei Kategorien (bei zwei Kategorien kann man z.B. auch einen t-Test verwenden), dann kann man dies mit Hilfe eines linearen Regressionsmodells angehen, aber eben auch mittels Varianzanalyse. Ausschlaggebend sollte sein, ob man eher an der Erklärung von Variabilität (Varianzanalyse) interes-



Ulrike Held

Die Autorin erklärt, dass sie keine Interessenkonflikte im Zusammenhang mit diesem Beitrag hat.

Tabelle 1. Hämoglobin-Messungen (steady state).

| Typ der Sichelzellanämie | Hämoglobin (g/dl) |
|--------------------------|--|
| Hb-SS | 7,2; 7,7; 8,0; 8,1; 8,3; 8,4; 8,4; 8,5; 8,6; 8,7; 9,1; 9,1; 9,1; 9,8; 10,1; 10,3 |
| Hb-S/β-Thalassämie | 8,1; 9,2; 10,0; 10,4; 10,6; 10,9; 11,1; 11,9; 12,0; 12,1 |
| Hb-SC | 10,7; 11,3; 11,5; 11,6; 11,7; 11,8; 12,0; 12,1; 12,3; 12,6; 12,6; 13,3; 13,3; 13,8; 13,9 |

siert ist oder an der Schätzung eines Effekts (lineares Regressionsmodell). Man könnte sich fragen, warum die Varianzanalyse einen so hohen Stellenwert in der medizinischen Forschung hat, wenn man doch alle Varianzanalyseprobleme auch als Regressionsprobleme auffassen kann. Die Erklärung kommt vermutlich aus der historischen Entwicklung der beiden Methoden. Während die Varianzanalyse bereits in den 1920er Jahren von R. A. Fisher vorgeschlagen wurde (und seither als solche verwendet wird), kamen einfache lineare sowie generalisierte Regressionsmodelle erst einige Zeit später «in Mode», was bestimmt auch mit der Erleichterung der Berechnungen durch die Erfindung des Computers einherging. Somit werden Fragestellungen, in denen ein Faktor mit mehr als zwei verschiedenen Kategorien bezüglich einer Zielgrösse untersucht werden soll, in der medizinischen Forschung fast immer eher als Varianzanalyseproblem denn als Regressionsproblem angesehen.

Glossar

Normalverteilung

Man spricht von normal verteilten Daten, wenn die grafische Darstellung einer Glockenkurve ähnelt. Diese wird auch Gauss-Verteilung genannt.

Nullhypothese

Eine Nullhypothese, oder H_0 , in der Statistik formuliert man als eine Aussage, von der man zeigen möchte, dass sie nicht stimmt. Unter Einbeziehung der Daten und eines statistischen Modells berechnet man eine Teststatistik, die dann für oder gegen diese Nullhypothese spricht. Spricht

die Teststatistik stark genug gegen die Nullhypothese, kann man diese verwerfen und davon ausgehen, dass die Alternativhypothese mit hoher Wahrscheinlichkeit richtig ist.

Quantil

Ein bestimmter Teil einer nach Grösse sortierten Datenmenge wird Quantil genannt. Es wird durch ein Quantil festgelegt, wie viele Werte einer Verteilung ober- oder unterhalb liegen. Bekannteste Quantile sind Quartile (Datensatz wird in vier gleich grosse Teile aufgeteilt), Quintile (fünf gleich grosse Teile) oder Perzentile (Hundertstel).

QQ-Plot

Mit einem QQ-Plot (die Abkürzung steht für Quantil-Quantil-Plot) werden die Quantile zweier Variablen gegeneinander abgetragen. Der QQ-Plot dient dazu, die Verteilung der beiden Variablen zu vergleichen. Häufig wird er dazu verwendet, die beobachteten (Stichproben-)Quantile mit denen einer theoretischen Verteilung (z.B. Normalverteilung) zu vergleichen und somit zu untersuchen, ob die Daten so verteilt sind wie die hypothetische Verteilung.

Standardabweichung

Die Standardabweichung ist ein Mass für die Streuung einer Messreihe. Sie wird als Wurzel aus der Varianz berechnet und hat dadurch gegenüber der Varianz einen Vorteil in der Interpretierbarkeit, denn sie liegt auf der gleichen Skala wie die Beobachtungen selbst.

Korrespondenz:

Dr. rer. nat. Ulrike Held
 Horten-Zentrum
 UniversitätsSpital Zürich
 Postfach Nord
 CH-8091 Zürich
Ulrike.Held@usz.ch

Empfohlene Literatur

- Held L, Rufibach C, Seifert B. Einführung in die Biostatistik. 4. Auflage. Zürich: Abteilung Biostatistik, Institut für Sozial- und Präventivmedizin der Universität Zürich; Juli 2009. <http://www.biostat.uzh.ch>.
- Hüsler J, Zimmermann H. Statistische Prinzipien für medizinische Projekte. 4. Auflage. Bern: Huber-Verlag; 2006.
- Kirkwood B, Sterne J. Essential Medical Statistics. 2. Auflage. Massachussets: Blackwell Science; 2003.
- Kreienbrock L, Schach S. Epidemiologische Methoden. 4. Auflage. München: Elsevier-Verlag, 2005.
- R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2008. ISBN 3-900051-07-0; URL <http://www.R-project.org>.

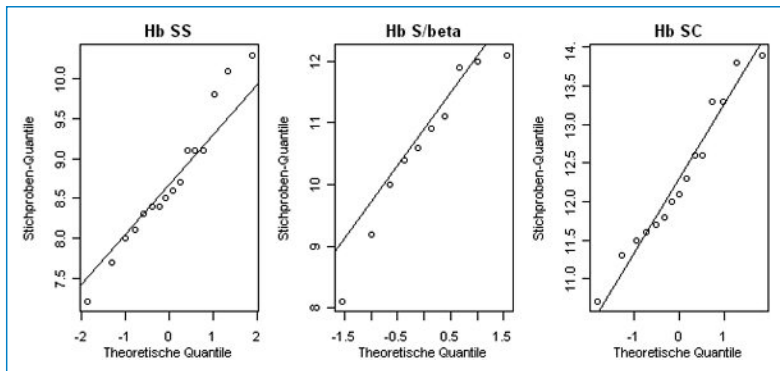


Abbildung 1
 QQ-Plot der Hämoglobinwerte nach Sichelzelltypus.