Einfache Varianzanalyse für unabhängige Stichproben

VARIANZANALYSE

Die Varianzanalyse ist das dem t-Test entsprechende Mittel zum Vergleich mehrerer ($k \geq 2$) Stichprobenmittelwerte. Sie wird hier mit *VA* abgekürzt, oft trifft man aber auch auf die englischsprachige Abkürzung *ANOVA* (Analysis of Variance).

Voraussetzungen

1.) Intervallskaleneigenschaft der Daten
Das strenge Prüfen des Skalenniveaus übersteigt unsere Möglichkeiten. Dennoch muss man vor jeder Anwendung der VA überlegen, ob bei der Erhebung der Messwerte diese Voraussetzung verletzt wurde.

2.) Normalverteilung (in jeder Stichprobe)
Die Überprüfung auf NV wird in Abschnitt *6.1. Anpassungstests* behandelt.

3.) Homogenität der Varianzen
Leider können mittels F-Test nur zwei Varianzen auf Homogenität getestet werden. Entsprechende Methoden für drei oder mehr Stichproben sind der *Bartlett*- und der *Cochran*-Test, die im folgenden Abschnitt vorgestellt werden.

Prinzip

Die VA funktioniert folgendermaßen:

Die Varianzen müssen (s. oben) in den untersuchten Stichproben homogen sein. D.h. man kann davon ausgehen, dass die Streuung *innerhalb* aller Stichproben ein bestimmtes Ausmaß nicht übersteigt. *Zwischen* den einzelnen Stichproben gibt es natürlich auch Abweichungen. Es gilt herauszufinden, ob sie signifikant größer ausfallen als die Abweichungen innerhalb der Stichproben, oder ob es sich lediglich um Zufallsschwankungen handelt.

Man benötigt Schätzungen für die beiden Varianzen: Die *Varianz Innerhalb* der Stichproben und die *Varianz Zwischen* den Stichproben. Diese werden mittels F-Test verglichen. Ein signifikantes Ergebnis bedeutet, dass zwischen den Stichproben überzufällige Unterschiede bestehen.

Aufstellen der Hypothesen

Davon ausgehend, dass jede Stichprobe einer bestimmten NV entstammt ($N(\mu_1,\sigma_1), N(\mu_2,\sigma_2), ..., N(\mu_k,\sigma_k)$), werden die Hypothesen folgendermaßen formuliert:

$H_0 \ldots \mu_1 = \mu_2 = \ldots = \mu_k$

Da die Varianzen als gleich vorausgesetzt werden, bedeutet dies, dass alle Stichproben der selben NV entstammen.

$H_1 \ldots \exists(\mu_j,\mu_l) : \mu_j \neq \mu_l$

Das bedeutet: Es gibt mindestens ein Paar von Mittelwerten, die nicht gleich sind. Anders formuliert besagt die H_1, dass mindestens eine Stichprobe einer anderen Normalverteilung entstammt.

Anmerkung:
Manchmal wird die Varianzhomogenität ($\sigma_1 = \sigma_2 = \ldots = \sigma_k$) beiden Hypothesen angefügt.

Berechnen der Varianzschätzungen

Um die besagten Varianzen schätzen zu können, werden zuerst QI und QZ berechnet:

$QI \ldots$ „Quadratsumme Innerhalb“ $\ldots$ Summe der Abweichungsquadrate innerhalb der Stichproben, also zwischen den einzelnen Werten einer Stichprobe und dem jeweiligen Stichprobenmittelwert

$QZ \ldots$ „Quadratsumme Zwischen“ $\ldots$ Summe der Abweichungsquadrate zwischen den Stichproben, also zwischen dem jeweils ersten Wert jeder Stichprobe und dem Durchschnitt der ersten Stichprobenwerte, analog für die jeweils zweiten, dritten, ... n-ten Stichprobenwerte

Als Schätzung der Varianzen dienen die mittleren Abweichungsquadrate MQI und MQZ. Um diese zu berechnen, werden QI und QZ durch die Anzahl der entsprechenden Freiheitsgrade dividiert:

Varianz Innerhalb ... $df = N - k$

$$\Rightarrow \hat{\sigma}_0^2 = MQI = \frac{QI}{N-k}$$

Varianz Zwischen ... $df = k - 1$

$$\Rightarrow \hat{\sigma}_1^2 = MQZ = \frac{QZ}{k-1}$$

N ... Gesamtzahl aller Messwerte
k ... Anzahl der Stichproben

QT

Um sich die oft aufwändige Berechnung von QI zu erleichtern, berechnet man die „totale (Abweichungs-)Quadratsumme“ QT. Sie ist die Summe der Abweichungsquadrate zwischen allen Einzelwerten und dem Gesamtmittelwert und entspricht genau der Summe von Quadratsumme Innerhalb und Quadratsumme Zwischen. Also lässt sich QI leicht berechnen, wenn QT und QZ bekannt sind:

$$QT = QI + QZ$$
$$QI = QT - QZ$$

Weiters gilt:

$$df_{QT} = df_{QI} + df_{QZ}$$

also

$$df_{QT} = N - k \; + \; k - 1$$
$$df_{QT} = N - 1$$

Hier (zur besseren Übersicht in Tabellenform) die Rechenformeln:

Quelle	Abweichungsquadratsumme	df	$\hat{\sigma}^2$
total	$QT = \sum_{ij} x_{ij}^2 - \frac{(\sum_{ij} x_{ij})^2}{N}$	$N-1$	
zwischen	$QZ = \sum_j \frac{(\sum_i x_{ij})^2}{n_j} - \frac{(\sum_{ij} x_{ij})^2}{N}$	$k-1$	$\hat{\sigma}_1^2 = \frac{QZ}{k-1}$
innerhalb	$QI = QT - QZ$	$N-k$	$\hat{\sigma}_0^2 = \frac{QI}{N-k}$

Erläuterungen zu den Formeln:

Der Index j steht hier für die verschiedenen Stichproben, während i für die verschiedenen Werte innerhalb der Stichproben steht.

Der zweite Teil der Formeln für QT und QZ ist ident, muss also nur einmal berechnet werden.

$\sum_{ij} x_{ij}$ ... Summe aller Messwerte

$\sum_{ij} x_{ij}^2$ ... Summe der Quadrate aller Messwerte

$\sum_i x_{ij}$ ... Summe aller Werte einer Stichprobe (für ein fixes j)

$\frac{(\sum_i x_{ij})^2}{n_j}$ ... Die Summe wird quadriert und durch den Umfang der jeweiligen Stichprobe dividiert. Dieser Bruch wird für jedes j, also für jede Stichprobe, einzeln berechnet.

$\sum_j \frac{(\sum_i x_{ij})^2}{n_j}$ ... Summe dieser Brüche

Der F-Test

Sind die Varianzen schlussendlich unter Dach und Fach, so müssen sie für den F-Test nur noch dividiert und das Resultat mit dem kritischen Wert ($\rightarrow$ Tabelle) verglichen werden:

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_0^2}$$

Wichtig:
Da nur eine Abweichung in eine Richtung interessant ist, nämlich dass die Varianz Zwischen signifikant *größer* ist, als die Varianz Innerhalb, steht $\hat{\sigma}_1^2$ immer im Zähler und $\hat{\sigma}_0^2$ immer im Nenner des Bruchs (unabhängig davon, welcher der größere Wert ist). Wäre $\hat{\sigma}_1^2$ signifikant *kleiner* als $\hat{\sigma}_0^2$, so wären sich die einzelnen Stichproben besonders *ähnlich*, was höchstens auf mögliche Fehler im Versuchsdesign schließen lässt, eine inhaltliche Interpretation hätte aber nichts anderes als die Gültigkeit der H_0 zur Folge. Deshalb ist der Signifikanztest bei der VA stets einseitig durchzuführen.

WARUM BERECHNET MAN STATT DER VA NICHT EINFACH FÜR JEDES PAAR VON STICHPROBEN EINEN t-TEST?

Dafür gibt es verschiedene Gründe. Die beiden wichtigsten sind:

Die Zahl der notwendigen Tests steigt bei wachsender Anzahl von Stichproben stark an. Für 6 Stichproben bräuchte man bereits 15 einzelne Tests! Allgemein wären bei k Stichproben $\binom{k}{2}$ Tests durchzuführen.

Das Risiko eines α-Fehlers wird insgesamt viel zu hoch. Für $\alpha = 5\%$ und 15 Tests würde die Irrtumswahrscheinlichkeit bereits auf über 50% steigen! Allgemein gilt für m Tests: $\alpha_{ges.} = 1 - (1 - \alpha)^m$.