Varianzanalyse – ANOVA

Johannes Hain

Lehrstuhl für Mathematik VIII – Statistik

Einfaktorielle Varianzanalyse (ANOVA)

Bisher war man lediglich in der Lage, mit dem t-Test einen Mittelwertsvergleich für zwei unabhängige Stichproben durchzuführen.

Hat man nun aber mehr als zwei Stichproben vorliegen, stellt der t-Test nicht mehr die geeignete Auswertungsmöglichkeit dar. In diesem Fall muss es also noch eine andere Möglichkeit der statistischen Auswertung geben – die Varianzanalyse (**An**alyis **o**f **Va**riance).

Gegeben sind also $I > 2$ Stichproben $x_{i,1}, \ldots, x_{i,n_i}, i = 1, \ldots, I$, wobei alle auftretenden Zufallsvariablen voneinander unabhängig sind. Ferner sei der Gesamtumfang n der Stichprobe definiert durch $n := n_1 + \cdots + n_I$.

Einfaktorielle Varianzanalyse (ANOVA)

Beispiel: Vier verschiede Unterrichtsarten sollen untersucht werden. Dazu werden 32 Personen zufällig auf vier Gruppen à 8 Personen aufgeteilt. Am Ende des Kurses wird eine Abschlussprüfung durchgeführt und die Punkte jedes Teilnehmers dokumentiert:

Gruppe 1	Gruppe 2	Gruppe 3	Gruppe 4
16	16	2	5
18	12	10	8
20	10	9	8
15	14	10	11
20	18	11	1
15	15	9	9
23	12	10	5
19	13	9	9
18.25	13.75	8.75	7.00

Gibt es signifikante Unterschiede zwischen den Gruppen?

Die unabhängige Variable, die in I Kategorien vorliegt, nennt man auch **Faktor**, die einzelnen Kategorien **Faktorstufen**. Da man den Einfluss von nur einem Faktor auf die abhängige Variable untersucht, spricht man auch von einer **einfaktoriellen Varianzanalyse**.

Die zu untersuchende Nullhypothese lautet:

$$H_0 : \mu_1 = \ldots = \mu_I,$$

also dass keine Unterschiede in den Mittelwerten der I Faktorstufen vorliegen.

Einfaktorielle Varianzanalyse (ANOVA)

Um die Ergebnisse einer ANOVA verwenden zu können, müssen die folgenden drei Voraussetzungen für das obige Modell erfüllt sein:

Voraussetzungen der ANOVA

1. Die Stichproben müssen unabhängig voneinander erhoben worden sein
2. Die i-te Stichprobe ($i = 1, \ldots, I$) folgt einer $N(\mu_i, \sigma^2)$-Verteilung
3. Die Varianz ist in allen I Stichproben gleich (Varianzhomogenität)

$\Longrightarrow$ **Diese Voraussetzungen sind zu überprüfen!!**

Überprüfung der Normalverteilungsannahme bei der ANOVA

- → *Analysieren*
- → *Deskriptive Statistiken*
- → *Explorative Datenanalyse...*
- → Ziehe die zu untersuchende Variable in das Feld *Abhängige Variablen*, die Gruppierungsvariable in das Feld *Faktorenliste*
- → Wähle das Feld *Diagramme* aus und klicke dort das Feld *Normalverteilungsdiagramm mit Tests* an

Der Test zur Voraussetzung der Varianzhomogenität wird direkt bei der Durchführung der ANOVA mit ausgegeben.

Sei x_{ij} die j-te Beobachtung der i-ten Stichprobe und $\bar{x}$ das **Gesamtmittel**, sowie $\bar{x}_i$ das i-**te Gruppenmittel**. Dann gilt:

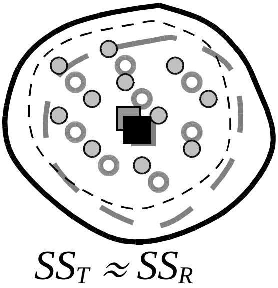
$$x_{ij} = \bar{x} + \underbrace{(\bar{x}_i - \bar{x})}_{\substack{\text{Abweichung Gruppenmittel} \\ \text{vom Gesamtmittel}}} + \underbrace{(x_{ij} - \bar{x}_i)}_{\substack{\text{Abweichung Beobachtung} \\ \text{vom Gruppenmittel}}}$$

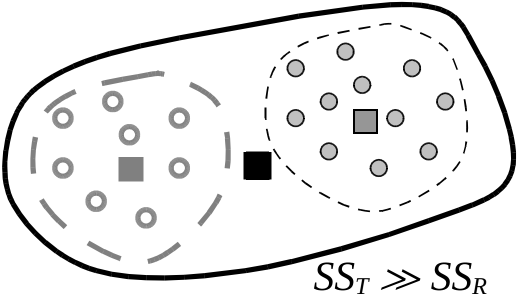
Gilt H_0 nicht, wird die Abweichung der Gruppenmittel zum Gesamtmittel hoch sein im Vergleich zur Abweichung der Beobachtungen zum Gruppenmittel.

Auf diesen Überlegungen basiert auch die Teststatistik

$$F_{0,\alpha} := \frac{\frac{1}{I-1} \cdot SS_A}{\frac{1}{n-1} \cdot SS_R} = \frac{\frac{1}{I-1} \cdot J \sum_{i=1}^{J} (\bar{x}_i - \bar{x})^2}{\frac{1}{n-1} \cdot \sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ij} - \bar{x}_i)^2}.$$

Je weiter die Mittelwerte der einzelnen Faktorstufen vom Gesamtmittel abweichen, desto größer wird der Wert für SS_A, im Vergleich zum Wert für SS_R. Unter H_0 sollte also der Quotient $\frac{SS_A}{SS_R}$ nahe bei Null liegen. Je größer SS_A wird – und somit auch je größer der Quotient wird – desto unwahrscheinlicher ist die Gültigkeit von H_0. Bei zu großen Werten von F wird H_0 verworfen.


$SS_T \approx SS_R$


$SS_T \gg SS_R$

Einfaktorielle ANOVA in SPSS

- $\rightarrow$ *Analysieren*
- $\rightarrow$ *Mittelwerte vergleichen*
- $\rightarrow$ *Einfaktorielle ANOVA...*
- $\rightarrow$ Ziehe die zu untersuchende Variable in das Feld *Abhängige Variablen* und die Gruppierungsvariable in das Feld *Faktor*
- $\rightarrow$ Klicke zusätzlich das Feld *Optionen* an und aktiviere das Feld *Test auf Homogenität der Varianzen* um die Voraussetzung der Varianzhomogenität mit dem Levene-Test zu überprüfen.

Ist der p-Wert des Levene-Tests kleiner als 0.05, wird die Voraussetzung der Varianzgleichheit in den Stichproben verworfen. In diesem Fall muss man, wie beim t-Test, auf einen bedingten Test ausweichen (Behrens-Fisher-Problem), den **Welch-Test**.

Welch-Test in SPSS

→ *Analysieren*

→ *Mittelwerte vergleichen*

→ *Einfaktorielle ANOVA...*

→ Ziehe die zu untersuchende Variable in das Feld *Abhängige Variablen* und die Gruppierungsvariable in das Feld *Faktor*

→ Klicke zusätzlich das Feld *Optionen* an und aktiviere das Feld *Welch* um den Welch-Test durchzuführen.

ANOVA – Posthoc-Analyse

Zusammenfassung ANOVA

Der durchgeführte Test (F-Test) bei der Varianzanalyse ist ein sogenannter **globaler Test** (oder auch **Omnibustest**).

Es wird also nur überprüft, ob *überhaupt* ein Unterschied zwischen den einzelnen Faktorstufen vorliegt, aber nicht *wo* eventuell vorhandene Unterschiede liegen.

$\Longrightarrow$ Wie geht es also nach der ANOVA weiter?

bzw.

$\longrightarrow$ Wie findet man bei einem signifikanten globalen Testergebnis auch noch heraus, wo genau die Unterschiede zwischen den Faktorstufen liegen?

ANOVA – Posthoc-Analyse

Die intuitive Herangehensweise an dieses Thema lautet:

$\longrightarrow$ Warum macht man nicht mit jeder Faktorkombination einen t-Tests?

Antwort:
Mit dieser Vorgehensweise steigt der **Fehler 1. Art** sehr schnell an!

Hat man z.B. 5 Faktorstufen, gibt es 10 Kombinationsmöglichkeiten. Werden diese 10 t-Tests durchgeführt steigt die Fehlerwahrscheinlichkeit auf über 40%!

ANOVA – Posthoc-Analyse

Die Lösung dieses Problems stellen die sogenannten **Posthoc-Tests** dar:

Definition: Posthoc-Tests

Posthoc-Analysen sind **paarweise Vergleichsprozeduren**, mit denen nach einem signifikanten Ergebnis des globalen Tests durch sog. **multiple Mittelwertsvergleiche** nach signifikanten Unterschieden zwischen den einzelnen Faktorstufen gesucht werden kann.

Achtung:

Posthoc-Verfahren gibt es in der Statistik sehr viele. Im Folgenden soll jedoch nur der sog. **Tukey-Test** vorgestellt werden. Der Tukey-Test ähnelt dem t-Test, hat aber die besondere Eigenschaft, dass er das Fehlerniveau konstant nahe 5% hält.

Tukey-Test SPSS

→ *Analysieren*

→ *Mittelwerte vergleichen*

→ *Einfaktorielle ANOVA...*

→ Ziehe die zu untersuchende Variable in das Feld *Abhängige Variablen* und die Gruppierungsvariable in das Feld *Faktor*

→ Klicke den Schalter *Post Hoc* an und wähle dort im Feld *Varianz-Gleichheit angenommen* die Option *Tukey* aus

Der Kruskal-Wallis-Test als nichtparametrische Altervative

Folgen die untersuchten Daten keiner Normalverteilung, stellt die ANOVA nicht das geeignete Auswertungsverfahren dar.

Eine **nichtparametrische Alternative** zur Varianzanalyse stellt der **Kruskal-Wallis-Test** dar, der kaum Voraussetzungen an das Modell fordert. Er kann als eine Verallgemeinerung des Mann-Whitney-U-Tests angesehen werden.

Genau wie der U-Test betrachtet auch der Kruskal-Wallis-Test nicht konkreten Realisierungen $x_{i,j}$ selbst, sondern nur ihre jeweiligen Ränge $R_{i,j}$.

Der Kruskal-Wallis-Test als nichtparametrische Altervative

Voraussetzungen

Gegeben sind $I > 2$ Stichproben. Die Zufallsvariablen $X_{i,j}, j = 1, \ldots, n_i$, der i-ten Stichprobe besitzen die gleiche stetige Verteilung, $i = 1, \ldots, I$. Es sei $n := n_1 + \ldots + n_I$. Ferner sind alle Zufallsvariablen unabhängig voneinander.

Es soll nun die folgende Nullhypothese untersucht werden:

H_0 : Die I Stichproben entstammen der gleichen Grundgesamtheit

Gilt H_0 dann bedeutet dies, dass dann insbesondere auch ihre Erwartungswerte übereinstimmen.

Alle Beobachtungen werden zu einer Stichprobe zusammengefasst. Diese wird der Größe nach aufsteigend angeordnet und jedem Wert $x_{i,j}$ wird sein entsprechender (zufälliger) Rang $R_{i,j}$ zugeordnet. Mit dem **Rang-Gesamtmittel** $\bar{R}$ und dem i-**ten Rang-Gruppenmittel** $\bar{R}_i$ bildet man die Teststatistik

$$SRS_A = \sum_{i=1}^{J} J(\bar{R}_i - \bar{R})^2.$$

Unterscheiden sich die Gruppen stark voneinander, ist SRS_A tendenziell groß. Je größer SRS_A, desto unwahrscheinlicher wird damit auch H_0.

Der Kruskal-Wallis-Test als nichtparametrische Altervative

Der Kruskal-Wallis-Test ist kein exakter Test – unter H_0 ist die Teststatistik

$$\frac{12}{n(n+1)} SRS_A$$

approximativ χ^2-verteilt mit $I - 1$ Freiheitsgraden.

Damit die Näherung hinreichend gute Ergebnisse liefert sollten die Stichprobenumfänge wieder groß genug sein:

Faustregeln für den Kruskal-Wallis-Test

- Falls $I = 3$: $n_1, n_2, n_3 \geq 5$.
- Falls $I \geq 4$: $n_1, \ldots, n_I \geq 4$.

Der Kruskal-Wallis-Test in SPSS

→ *Analysieren*

→ *Nichtparametrische Tests*

→ *Unabhängige Stichproben...*

→ Aktiviere das Feld *Felder*

→ Übertrage die unabhängigen Variablen in das Feld *Testfelder*

→ Übertrage die abhängige Variable in das Feld *Gruppen* und bestätige mit *Ausführen*

Bei diesem Analyseweg wird im Output-Viewer das Testergebnis im sog. **Model Viewer** angezeigt. Dort findet man auch unter „Paarweise Vergleiche“ Posthoc-Analysen.

Der Kruskal-Wallis-Test als nichtparametrische Altervative
oder . . .

Der Kruskal-Wallis-Test in SPSS

→ *Analysieren*

→ *Nichtparametrische Tests*

→ *Alte Dialogfelder*

→ *K unabhängige Stichproben...*

→ Ziehe die zu untersuchende Variable in das Feld *Testvariablen* und die Gruppierungsvariable in das Feld *Gruppenvariable*

→ Klicke den Schalter *Bereich definieren* an und bestimme dort in den Feldern *Minimum* und *Maximum* den gewünschten Untersuchungsbereich der Gruppierungsvariable

Cholesterin.sav

Der Datensatz enthält den Blut-Cholesterin-Spiegel von Probanden, die einem deutlich erhöhten Wert (>250 mg/dl) aufgefallen sind. Seit diesem Befund erfuhren die Testpersonen eine der folgenden drei Behandlungsarten: kein Medikament (0), Medikament A oder Medikament B. Aufgezeichnet werden die Cholesterinwerte einen bestimmten Zeitraum nach Gabe der Medikamente.

Verringert die Gabe eines Medikaments den Cholesterin-Wert im Vergleich zur Kontrolle? Wirkt eines der beiden Medikamente besser?

ph.sav

Der Datensatz untersucht (unter anderem) den Einfluss des Faktors Beregnung (sauer, normal, keine) auf den pH-Wert von Waldboden.
Gibt es (signifikante) Unterschiede zwischen den drei Beregnungsarten bezüglich des pH-Werts der Waldbodenproben? Falls ja, wo liegen die Unterschiede?

Wein.sav

Zur Prämierung von Weinen werden diese von Prüfern bewertet. Je höher die Bewertung, desto besser die Qualität des Weins. Es wird vermutet, dass die einzelnen Weinprüfer unterschiedliche Wertungen abgeben.
Untersuche ob es Prüfer gibt, die die Weine (signifikant) anders bewerten als andere Prüfer und identifiziere diese mit einem Posthoc-Testverfahren.