

Kapitel 5: Einfaktorielle Varianzanalyse

Durchführung einer einfaktoriellen Varianzanalyse ohne Messwiederholung	1
Berechnung der Effektstärke und der beobachteten Teststärke einer einfaktoriellen Varianzanalyse ohne Messwiederholung	5
Vergleich zwischen einer einfaktoriellen Varianzanalyse ohne Messwiederholung und einem <i>t</i> -Test für unabhängige Stichproben	7
Literatur	10

Durchführung einer einfaktoriellen Varianzanalyse ohne Messwiederholung

Dieser Abschnitt zeigt die Durchführung der in Kapitel 5 vorgestellten einfaktoriellen Varianzanalyse ohne Messwiederholung mit R. Das Beispiel im Buch beschäftigte sich mit der Frage, wie der Unterschied zwischen den drei Verarbeitungsbedingungen „strukturell“, „bildhaft“ und „emotional“ in der Erinnerungsleistung zu Stande gekommen ist. Handelt es sich um einen zufälligen Unterschied, oder gibt es Grund zu der Annahme, dass sich die drei Gruppen systematisch unterscheiden? Die Varianzanalyse vergleicht diese drei Bedingungen simultan miteinander, es ist also nur *ein* statistischer Test notwendig um zu entscheiden, ob die Unterschiede zwischen den drei Gruppen zufällig oder systematisch sind. Damit wird das Problem der α -Fehlerkumulierung vermieden (Kap. 5.1.1).

Für eine Varianzanalyse kann die Funktion `aov()` verwendet werden. Geben Sie zunächst an, für welche Variable R eine einfaktorielle Varianzanalyse durchführen soll (hier: „Gesamtzahl erinnelter Adjektive“). Außerdem müssen Sie spezifizieren, welche Gruppen oder experimentelle Bedingungen das Verfahren miteinander vergleichen soll. Diese geben Sie danach an.

Das Experiment beinhaltete insgesamt drei unterschiedliche Verarbeitungsbedingungen. Die Verarbeitungsbedingung ist für jede Versuchsperson eindeutig durch eine von drei Zahlen festgelegt. Jede Zahl steht für eine Versuchsbedingung. Mit folgendem Code können Sie in R nachvollziehen, welche Zahl für welche Verarbeitungsbedingung steht. Lesen Sie zunächst den Datensatz ein.

```
library(foreign)
beispiel <- read.spss("Beispiel.datensatz.sav",
                    to.data.frame = TRUE)
as.numeric(beispiel$bed)
```

Sie erkennen, dass die strukturelle Bedingung durch eine Eins kodiert ist. Alle Personen, die während des Experimentes die präsentierten Wörter strukturell verarbeiten sollten, haben an dieser Stelle eine Eins. Eine Zwei symbolisiert die bildhafte Verarbeitungsbedingung, eine Drei die emotionale Verarbeitungsbedingung. Vor der Durchführung der Analyse sollten Sie sich zusätzlich

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

R-Ergänzungen

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

die deskriptiven Statistiken der untersuchten Gruppen anzeigen lassen. Verwenden Sie hierfür die Funktion `describeBy()` des Pakets `psych`. Auch den Test auf Homogenität der Varianzen sollten Sie sich ausgeben lassen, denn er überprüft, ob die Voraussetzung vergleichbar großer Varianzen in den Bedingungen gewährleistet ist. Dafür benötigen Sie die Funktion `leveneTest()` des Pakets `car`. Ein Diagramm der Mittelwerte kann Ihnen die Interpretation des Mittelwertmusters erleichtern. Es liefert allerdings nur Informationen, die mit der Ausgabe der deskriptiven Statistik redundant ist und nicht die Information über die Streuungen in den Bedingungen enthält.

Der vollständige Code sieht folgendermaßen aus:

```
# Deskriptive Statistiken
library(psych)
describeBy(beispiel$ges, group = beispiel$bed)
```

```
# Test auf Homogenität der Varianzen
library(car)
leveneTest(beispiel$ges, beispiel$bed, center = mean)
```

```
# einfaktorielle Varianzanalyse
ges <- beispiel$ges
bed <- beispiel$bed
anova1 <- aov(ges ~ bed)
summary(anova1)
```

Sie erhalten folgenden Output:

```
Descriptive statistics by group
group: strukturell
  vars  n mean  sd median trimmed  mad min max range skew kurtosis  se
x1    1 50  7.2 3.16     7   7.03 2.97  1 15  14 0.45   -0.08 0.45
-----
group: bildhaft
  vars  n mean  sd median trimmed  mad min max range skew kurtosis  se
x1    1 50  11 4.14    10  10.75 4.45  5 26  21 0.9    1.51 0.59
-----
group: emotional
  vars  n mean  sd median trimmed  mad min max range skew kurtosis  se
x1    1 50 12.02 4.21    11  11.78 3.71  4 23  19 0.47   -0.11 0.59
```

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

```
Levene's Test for Homogeneity of Variance (center = mean)
```

```
      Df F value  Pr(>F)
group  2  2.5158 0.08428 .
      147
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
      Df Sum Sq Mean Sq F value    Pr(>F)
bed      2  645.2    322.6    21.59 0.00000000604 ***
Residuals 147 2197.0     14.9
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hinweis: Falls Sie keine wissenschaftliche Notation möchten, geben Sie `options(scipen = 999)` in der Konsole ein und führen Sie die Analyse erneut durch.

In der ersten Tabelle finden Sie die Angaben zur deskriptiven Statistik der verglichenen Gruppen. Die abhängige Variable („Gesamtzahl erinnerter Adjektive“) wurde im Code angegeben. So ist zum Beispiel abzulesen, dass alle drei Gruppen mit je 50 Versuchspersonen besetzt sind, und dass bei struktureller Verarbeitung 7,20, bei bildhafter 11,00 und bei emotionaler Verarbeitung 12,02 Adjektive erinnert wurden. Den Standardfehler dieser Mittelwerte finden Sie in der letzten Spalte.

Die zweite Tabelle liefert die Ergebnisse des Levene-Tests auf Varianzhomogenität. Ein signifikantes Ergebnis zeigt an, dass die Varianzen der experimentellen Bedingungen unterschiedlich groß sind und damit gegen die Voraussetzungen der Varianzanalyse verstoßen. Das Ergebnis bei dieser Analyse ist knapp nicht signifikant. Aus den R-Ergänzungen zu Kapitel 3 wissen wir, dass R das Ergebnis für den *t*-Test auf zwei verschiedene Arten berechnet, einmal als reguläre Analyse und einmal als Analyse unter Korrektur für Verletzung der Annahme der Varianzhomogenität. Dafür korrigierte R die Anzahl der Fehlerfreiheitsgrade. Ein korrigiertes Ergebnis für eine mögliche Verletzung dieser Annahme können Sie sich mit der Funktion `oneway.test()` ausgeben lassen. Hier geben Sie das Gleiche ein wie das, was oben in den Klammern von `aov()` steht. Darüber hinaus gibt es post-hoc-Verfahren, die empirische Unterschiede auch unter Verletzung dieser Voraussetzung überprüfen können (siehe unten).

Die dritte Tabelle zeigt die Ergebnisse der einfaktorischen Varianzanalyse, die bereits in Kapitel 5.2.12 ausführlich beschrieben wurden. Besonders wichtig für die Auswertung sind der *F*-Wert und die Angabe der Wahrscheinlichkeit dieses *F*-Werts unter der Nullhypothese (Spalte „Pr(>F)“).

Die Wahrscheinlichkeit, dass der *F*-Wert von $F_{(2;147)} = 21,59$ auftritt, wenn die Nullhypothese tatsächlich zutreffen würde, ist $p < 0,001$. Dieser Wert ist kleiner als ein Signifikanzniveau von $\alpha < 0,05$, und somit können wir die Nullhypothese ablehnen. Der Unterschied zwischen den drei Gruppen ist signifikant. Die errechnete Wahrscheinlichkeit ist hier sogar kleiner als 1%. In der Literatur wird ein derartiges Signifikanzniveau wie im R-Output häufig mit drei Sternen versehen (***) .

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

Allerdings testet die einfaktorielle Varianzanalyse den Unterschied zwischen allen drei Gruppen simultan. Ein signifikantes Ergebnis einer einfaktoriellen Varianzanalyse erlaubt nur die allgemeine Aussage, dass zwischen den untersuchten Gruppen ein irgendwie gearteter Unterschied vorliegt. Spezifischere Aussagen über die Unterschiede zwischen einzelnen Gruppen erfordern eine Post Hoc Analyse. Kapitel 5.4 stellt den Tukey HSD Test als eine Möglichkeit für Post Hoc Analyse vor. Dieser Test ist auch mit R durchführbar.

Sie benötigen die Funktion `glht()` des Pakets `multcomp`. Geben Sie als erstes das Objekt an, in das Sie die obige einfaktorielle Varianzanalyse gespeichert haben. Danach tragen Sie den Prädiktor und das Korrekturverfahren ein.

```
install.packages("multcomp")
library(multcomp)
posthoc1 <- glht(anova1, linfct = mcp(bed="Tukey"))
```

Das Ergebnis können Sie sich wieder mit `summary()` ausgeben lassen. Die dazugehörigen Konfidenzintervalle erhalten Sie mit `confint()`.

```
summary(posthoc1)
confint(posthoc1)
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = ges ~ bed)

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t)	
bildhaft - strukturell == 0	3.8000	0.7732	4.915	<0.0001	***
emotional - strukturell == 0	4.8200	0.7732	6.234	<0.0001	***
emotional - bildhaft == 0	1.0200	0.7732	1.319	0.387	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Adjusted p values reported -- single-step method)

Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Frieze, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

```
Fit: aov(formula = ges ~ bed)
```

```
Quantile = 2.3675
```

```
95% family-wise confidence level
```

```
Linear Hypotheses:
```

	Estimate	lwr	upr
bildhaft - strukturell == 0	3.8000	1.9695	5.6305
emotional - strukturell == 0	4.8200	2.9895	6.6505
emotional - bildhaft == 0	1.0200	-0.8105	2.8505

In der ersten Zeile wird die Verarbeitungsbedingung „bildhaft“ zunächst mit der Bedingung „strukturell“, in der nächsten Zeile wird dann „emotional“ mit „strukturell“ verglichen. Die Tabellenspalte „Estimate“ gibt die Differenz zwischen den jeweiligen Gruppen an. Aus der Tabelle der deskriptiven Statistiken (siehe oben) können Sie entnehmen, dass die Versuchspersonen in der bildhaften Verarbeitungsbedingung 11,00, und in der strukturellen Verarbeitungsbedingung 7,20 Adjektive erinnert haben. Die Differenz ist 3.8. Der Standardfehler der Differenz ist mit 0,7732 angegeben. Der Unterschied zwischen diesen beiden Gruppen ist signifikant ($p < 0,001$, Spalte „Pr(>|t|)“). R indiziert signifikante Unterschiede zwischen den einzelnen Gruppen mit den Symbolen in der Zeile „Signif. codes“. So lässt sich leicht erkennen, dass sich die Gruppe „strukturell“ signifikant von den Gruppen „bildhaft“ und „emotional“ unterscheidet, während zwischen den Gruppen „bildhaft“ und „emotional“ kein signifikanter Unterschied vorliegt ($p > 0,38$). Der Tukey HSD post-hoc-Test gilt allgemein als konservativer Test. Wenn dieser Test also einen signifikanten Unterschied zwischen zwei Gruppen anzeigt, darf man mit vergleichsweise großer Sicherheit davon ausgehen, dass dieser Unterschied auch tatsächlich existiert.

Berechnung der Effektstärke und der beobachteten Teststärke einer einfaktoriellen Varianzanalyse ohne Messwiederholung

R bietet mit dem oben vorgestellten Weg der Durchführung einer einfaktoriellen Varianzanalyse ohne Messwiederholung auch eine Möglichkeit, die Effektstärke anzeigen zu lassen. Obwohl die Berechnung der Effektstärke leicht mit dem Taschenrechner durchführbar ist (vgl. Kap. 5.3.2), soll hier ein alternativer Auswertungsweg in R vorgestellt werden, der die Angabe der Effektstärke η^2 sowie die Angabe der beobachteten Teststärke, einen empirischen Effekt dieser Größe zu finden, möglich macht. Diese Art der Berechnung führt beim Signifikanztest zu identischen Ergebnissen wie die gerade besprochene Vorgehensweise.

Lassen Sie sich das oben erstellte Objekt `anova1` mit der einfaktoriellen Varianzanalyse mit der Funktion `summary.lm()` anzeigen.

```
summary.lm(anova1)
```

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

```

Call:
aov(formula = ges ~ bed)

Residuals:
    Min       1Q   Median       3Q      Max
-8.02  -2.20  -0.60   2.00  15.00

Coefficients:
              Estimate Std. Error t value      Pr(>|t|)
(Intercept)    7.2000    0.5467  13.169 < 0.0000000000000002 ***
bedbildhaft    3.8000    0.7732   4.915    0.00000234576 ***
bedemotional   4.8200    0.7732   6.234    0.00000000456 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 3.866 on 147 degrees of freedom
Multiple R-squared:  0.227,    Adjusted R-squared:  0.2165
F-statistic: 21.59 on 2 and 147 DF,  p-value: 0.000000006035

```

In den meisten Analysen werden feste Faktoren untersucht, die sich ausschließlich auf die realisierten experimentellen Bedingungen beziehen. Zufällige Faktoren treten auf, wenn die realisierten experimentellen Bedingungen zufällig aus einer Menge aller möglichen experimentellen Bedingungen gezogen wurden, z.B. bei einer Untersuchung des Einflusses von Lärm auf die Konzentrationsfähigkeit, bei der zufällig die Lautstärken 16dB, 21dB und 83dB als experimentelle Bedingungen ausgewählt wurden (Näheres in Bortz & Schuster, 2010, Kapitel 12.2 und 12.3). In dem hier vorliegenden Fall sind die experimentellen Bedingungen „strukturell“, „bildhaft“ und „emotional“ jedoch nicht zufällig ausgewählt worden, es handelt sich deshalb um feste Faktoren.

Der untere Teil gibt das Ergebnis des Vergleichs zwischen den drei Verarbeitungsbedingungen wieder. Die Daten entsprechen der Zeile „bed“ in dem ersten vorgestellten Weg der Auswertung (`summary(anova1)`; $F_{(2,147)} = 21,59$; $p < 0,001$; siehe oben). Zusätzlich berechnet R einen empirischen Effekt auf der Stichprobenebene von Eta-Quadrat $\eta^2 = 0,227$, der dem Determinationskoeffizienten („Multiple R-squared“) entspricht. Der Faktor „Verarbeitungsbedingung“ klärt in dieser Analyse also 22,7% der Varianz der abhängigen Variable „Gesamtzahl erinnerter Adjektive“ auf.

Für die beobachtete Teststärke verwenden Sie die Funktion `power.anova.test()`. Spezifizieren Sie die Anzahl der Gruppen, die Anzahl der Personen in einer Gruppe, die Varianz zwischen den Gruppen und die Varianz innerhalb der Gruppen. Die Varianzen erhalten Sie aus der Tabelle, der oben gerechneten einfaktorischen Varianzanalyse.

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

```
power.anova.test(groups = 3, n = 50,  
                 between.var = 645.2,  
                 within.var = 2197.0)
```

Balanced one-way analysis of variance power calculation

```
groups = 3  
n = 50  
between.var = 645.2  
within.var = 2197  
sig.level = 0.05  
power = 0.9988648
```

NOTE: n is number in each group

Die Wahrscheinlichkeit, einen Effekt dieser Größe von $\eta^2 = 0,227$ mit einer Versuchspersonenzahl von 150 (50 pro Gruppe) und einem α -Niveau von 5% zu finden (falls er wirklich existiert), ist größer als 99% („power“, vgl. G*Power Ergänzungen zu diesem Kapitel).

Vergleich zwischen einer einfaktoriellen Varianzanalyse ohne Messwiederholung und einem *t*-Test für unabhängige Stichproben

Kapitel 5.3.1 beschreibt den *t*-Test als einen Spezialfall der Varianzanalyse. Der folgende Abschnitt soll diesen Zusammenhang mit Hilfe des Programms R noch einmal verdeutlichen.

Die Durchführung des *t*-Tests für den Vergleich der Verarbeitungsbedingungen „strukturell“ und „emotional“ ist bereits in den R Ergänzungen für Kapitel 3 beschrieben. Wenn eine einfaktorielle Varianzanalyse mit zwei Stufen wirklich nichts anderes als ein *t*-Test ist, dann sollten beide Berechnungen zu identischen Ergebnissen führen. Wie lässt sich dieser Vergleich in R realisieren? Für die Berechnung ist zunächst eine Selektion der zu vergleichenden Gruppen notwendig. Verwenden Sie die Funktion `subset()`. Um nur diese beiden Bedingungen zu selektieren, geben Sie folgenden Ausdruck ein.

```
strk.bild <- subset(beispiel,  
                  beispiel$bed == "strukturell"  
                  | beispiel$bed == "bildhaft")
```

In dem R Datenfenster sind nun alle Versuchspersonen, die die zu erinnernden Adjektive emotional verarbeitet haben, ausgeschlossen und werden nicht in die Analyse mit einbezogen. Das können Sie auch nachvollziehen, indem Sie sich die Häufigkeitstabelle der Variable „bed“ des neuen Objekts anzeigen lassen:

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

```
table(strk.bild$bed)
```

```
strukturell    bildhaft    emotional
           50             50             0
```

Nach der Selektion der Daten können Sie einen Vergleich der Gruppen „strukturell“ und „bildhaft“ mit einer einfaktoriellem Varianzanalyse durchführen. Folgen Sie einfach den Angaben zu Beginn dieser R Ergänzungen.

```
anova2 <- aov(strk.bild$ges ~ strk.bild$bed)
summary(anova2)
```

R liefert dann folgenden Output:

```
              Df Sum Sq Mean Sq F value    Pr(>F)
strk.bild$bed  1     361   361.0    26.6 0.0000013 ***
Residuals    98    1330    13.6
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Der R Output für diesen *t*-Test ist hier noch einmal dargestellt (vgl. R Ergänzungen zu Kapitel 3).

```
t.unab2 <- t.test(strk.bild$ges ~ strk.bild$bed,
                  var.equal = TRUE)
t.unab2
```

Two Sample t-test

```
data: strk.bild$ges by strk.bild$bed
t = -5.1575, df = 98, p-value = 0.000001305
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5.262133 -2.337867
sample estimates:
mean in group strukturell    mean in group bildhaft
                7.2                11.0
```

Zunächst fällt auf, dass die angegebene Anzahl der Nennerfreiheitsgrade ($df = 98$) für beide Rechenwege identisch ist. Dies gilt auch für die Zählerfreiheitsgrade, die bei einem *t*-Test nicht mit angegeben werden, da sie immer 1 sind. Auch die Angaben der Wahrscheinlichkeit des errechneten statistischen Kennwerts sind gleich ($p < 0,001$). Und der *F*-Wert entspricht dem quadrierten *t*-Wert: $F = t^2 \leftrightarrow 26,6 = (-5,158)^2$. Ein Vergleich der Mittelwerte zweier unabhängiger Gruppen kann also entweder mit einem *t*-Test oder einer einfaktoriellem Varianzanalyse ausgewertet werden, beide

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

R-Ergänzungen

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

Verfahren kommen zu demselben Ergebnis. Dies muss so sein, da der t -Test ein Spezialfall der einfaktoriellen Varianzanalyse für Faktoren mit zwei Stufen ist.

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

R-Ergänzungen

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

Literatur

Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Heidelberg: Springer.

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature