# Webergänzung zu Kapitel 10

## 10.1.4 Varianzanalyse (ANOVA: *analysis of variance*)

Im Kapitel 10 haben wir uns hauptsächlich mit Forschungsbeispielen beschäftigt, die nur zwei Ergebnissätze hatten (entweder werden zwei unterschiedliche Ergebnisse einer Gruppe verglichen oder zwei Gruppen wurden verglichen). Oftmals liegt jedoch mehr als eine unabhängige Variable vor oder mehr als zwei Gruppen werden verglichen, oder es gibt mehr als nur zwei Datensätze für eine Gruppe. In diesen Fällen muss eine ANOVA durchgeführt werden, eine Varianzanalyse. Auch eine ANOVA kann relativ leicht in Excel berechnet werden (auf Grund der vielen involvierten Berechnungen würde man dies nie per Hand machen, weswegen wir die Formeln auch nicht in das Online-Verzeichnis aufgenommen haben).

Wir werden im Folgenden zwei Beispiele für die Verwendung einer ANOVA besprechen. Dabei soll man wie immer vorher prüfen, ob die Voraussetzungen zur Anwendung einer ANOVA (Normalverteilung, intervallskalierte Daten, Varianzhomogenität) erfüllt sind.

### 10.1.4.1 Einfaktorielle Varianzanalyse

Nehmen wir an, wir wollen herausfinden, ob Studenten mit den Hauptfächern Geschichte, Politik und Medienwissenschaft ähnliche Englischkenntnisse besitzen.

In unserem Beispiel messen wir „Englischkenntnisse“, indem wir drei Gruppen mit je zehn Studenten einen Vokabeltest mit 50 Fragen absolvieren lassen. Wir haben eine unabhängige Variable, nämlich „Hauptfach“, mit drei Faktorstufen: Geschichte, Politik und Medienwissenschaft. Die abhängige Variable ist das Ergebnis im Vokabeltest. Die Ergebnisse unserer fiktiven Forschung finden sich in Tabelle 1.

**Tabelle 1: Ergebnisse des Vokabeltests**

| Geschichte | Politik | Medienwissen-schaft |
|---|---|---|
| 22 | 22 | 30 |
| 30 | 36 | 29 |
| 26 | 28 | 29 |
| 35 | 31 | 46 |
| 20 | 23 | 47 |
| 27 | 25 | 31 |
| 26 | 32 | 47 |
| 32 | 30 | 49 |
| 17 | 39 | 33 |
| 22 | 29 | 37 |
| $\overline{x}$ = 25,7 | $\overline{x}$ = 29,5 | $\overline{x}$ = 37,8 |

Wenn wir uns nur die Mittelwerte der drei Gruppen ansehen, stellen wir fest, dass die Medienwissenschaftsgruppe den höchsten Mittelwert (37,8) und die Geschichtsgruppe den niedrigsten Mittelwert (25,7) hat, während die Politikgruppe in der Mitte liegt (Mittelwert: 29,5). Nun stellt sich wie immer die Frage: Sind diese Unterschiede signifikant? Anders ausgedrückt: Können wir behaupten, dass die Variable „Hauptfach" einen Effekt hat? Weil wir mehr als zwei Gruppen haben, können wir keinen t-Test benutzen. Stattdessen müssen wir eine ANOVA durchführen, um diese Frage zu beantworten. Weil wir **eine** unabhängige Variable haben, ist dies eine **einfaktorielle** Varianzanalyse. (Hätten wir zwei unabhängige Variablen, würden wir eine **zweifaktorielle** Varianzanalyse durchführen, bei drei unabhängigen Variablen eine **dreifaktorielle** Varianzanalyse etc.)

Eine einfaktorielle ANOVA können wir leicht mit Excel berechnen. Die wichtigsten Werte, die wir dabei erhalten, sind die, die wir in einer Zeile „Unterschiede zwischen den Gruppen" erhalten; das sind der F-Wert sowie der p-Wert (dessen notwendiges Niveau wir auf 0,05 setzen). Wenn Sie die drei Datenreihen in eine Excel-Mappe eingeben und bei „Datenanalyse" – „Einfaktorielle Varianzanalyse" im Eingabebereich **alle** Daten eingeben, erhalten Sie ein Ergebnis wie in Tabelle 2:

**Tabelle 2: Ergebnisse einer einfaktoriellen Varianzanalyse**

| *Streuungsursache* | *Quadrat-summen* | *(df)* | *Mittlere Quadrat-summe* | *(F)* | *P-Wert* | *kritischer F-Wert* |
|---|---|---|---|---|---|---|
| Unterschiede zw. den Gruppen | 765,8 | 2 | 382,9 | 8,6716155 | 0,00123379 | 3,354130 |
| Innerhalb der Gruppen | 1192,2 | 27 | 44,155555 | | | |
| Gesamt | 1958 | 29 | | | | |

Der *F*-Wert ist 8,67, und dieser ist signifikant, wie wir aus seinem zugeordneten *p*-Wert (hier 0,0012, also kleiner als 0,05) entnehmen können. Das bedeutet, dass die Variable „Hauptfach“ in dem Sinne einen Effekt hat, dass die drei Gruppen nicht gleich sind. Vorerst sind die anderen Daten in Tabelle 4 für uns nicht relevant – bis auf zwei andere Werte, die in einer Forschungsstudie normalerweise erwähnt werden, nämlich die df-Werte (in der ersten Reihe der Tabelle finden wir dort die Zahl 2, nämlich die Anzahl unserer Gruppen minus 1, und in der zweiten Reihe die Zahl 27, die für die Anzahl der Testpersonen minus der Anzahl der Gruppen steht).[1]

Die Tabelle wird im Forschungsbericht normalerweise nicht wiedergegeben, sondern wie folgt zusammengefasst: „$F(2,27) = 8{,}67$, $p < 0{,}05$“, wobei die zwei Zahlen in den Klammern die Freiheitsgrade (*df*) angeben.

**Folgeanalyse / Post-hoc-Analyse**

Ein signifikanter F-Wert sagt uns nur, dass unsere Gruppen nicht gleich sind. Er sagt uns sogar nur, dass die Gruppe mit dem höchsten Mittelwert von der Gruppe mit dem niedrigsten signifikant verschieden ist; der signifikante F-Wert kann uns nicht sagen, ob sich alle drei Gruppen voneinander signifikant unterscheiden. Um dies herauszufinden, müssen wir eine Folgeanalyse durchführen (die auch Post-hoc-Analyse genannt wird), wie beispielsweise den Tukey-Test, den Newman-Keuls-Test oder den Scheffé-Test. Wir werden diese Analysen aufgrund der Komplexität hier nicht behandeln,[2] sondern einfach annehmen, dass eine solche Analyse bei den oben genannten Daten durchgeführt wurde mit dem Ergebnis, dass die Medienwissenschaftsgruppe (die Gruppe mit dem höchsten Mittelwert) von der Politik- und Geschichtsgruppe verschieden ist, aber die Politikgruppe sich von der Geschichtsgruppe nicht signifikant unterscheidet. Diese Ergebnisse könnte man wie folgt präsentieren:

> **Eine einfaktorielle ANOVA ergab, dass es einen Effekt von „Hauptfach“ ($F(2,27) = 8{,}67$, $p < 0{,}05$) gibt. Ein anschließend angewandter (post-hoc) Newman-Keuls-Test zeigte, dass die Medienwissenschaftsstudenten besser waren ($p < 0{,}05$) als die Geschichtsstudenten und besser ($p < 0{,}05$) als die Politikstudenten. Die Ergebnisse der beiden letztgenannten Gruppen waren jedoch nicht signifikant voneinander verschieden (s. Tabelle X):**

---

1 Bei einer ANOVA finden wir immer zwei Freiheitsgrad-Angaben, während wir zum Beispiel bei einem t-Test nur einen Wert für die Freiheitsgrade haben. Das liegt daran, dass wir bei einem t-Test wissen, dass wir nur zwei Gruppen haben; das müssen wir also nicht extra angeben. Bei Varianzanalysen hat man es mit mehr als zwei Gruppen zu tun; es muss angegeben werden, um wie viele Gruppen es sich handelt. Die Anzahl der Freiheitsgrade ist dann die Anzahl der Gruppen minus eins (hier 3 – 1 = 2). Der zweite Wert ergibt sich wie beim t-Test aus der Zahl der Versuchsteilnehmer minus der Anzahl von Gruppen (hier 30 – 3 = 27). Übrigens ergibt eine ANOVA mit nur zwei Faktorstufen dasselbe Ergebnis wie ein t-Test.

2 Post-hoc-Tests kann man mit Hilfe von Statistikprogrammen wie SPSS durchführen. Wenn Sie eine ANOVA durchführen, sollten Sie sich informieren, wie Post-hoc-Tests durchgeführt werden. Es gibt inzwischen auch sehr gute Online-Tutorien (z.B. auf Youtube), die den Hintergrund und die Schritte solcher Analysen sehr gut erklären.

**Tabelle [Nr. X]: Ergebnisse des Vokabeltests von Studenten verschiedener Hauptfächer (höchste erreichbare Punktzahl 50)**

| | Geschichte | Politik | Medienwissenschaft |
|---|---|---|---|
| Mittelwert | 25,7 | 29,5 | 37,8 |

### 10.1.4.2 Zweifaktorielle Varianzanalyse mit Messwiederholung

Nehmen wir an, wir haben die gleiche Situation wie in Beispiel 1, nur dass wir jetzt eine zweite Variable „Geschlecht" mit zwei Ebenen (männlich und weiblich, also zwei unabhängige Variablen) haben. Dies bedeutet, dass wir nun sechs statt der ursprünglichen drei Gruppen haben, da jede Gruppe in weibliche und männliche Studenten geteilt wird. Die Ergebnisse sind beispielhaft in Tabelle 3 dargestellt.

**Tabelle 3: Ergebnisse eines Englisch-Vokabeltests von sechs Gruppen**

| Geschlecht | Geschichte | Politik | Medienwissenschaft |
|---|---|---|---|
| männlich | 22 | 22 | 30 |
| männlich | 30 | 36 | 29 |
| männlich | 26 | 28 | 29 |
| männlich | 35 | 31 | 46 |
| männlich | 20 | 23 | 47 |
| weiblich | 27 | 25 | 31 |
| weiblich | 26 | 32 | 47 |
| weiblich | 32 | 30 | 49 |
| weiblich | 17 | 39 | 33 |
| weiblich | 22 | 29 | 37 |

Excel kann sogar eine zweifaktorielle ANOVA durchführen (**zwei**faktoriell, weil es nun **zwei** unabhängige Variablen gibt). Die wichtigsten Ergebnisse geben wir in Tabelle 4 wieder (wir runden hier auf drei Dezimalzahlen ab, sodass die Tabelle besser lesbar wird):

**Tabelle 4: Ergebnisse einer zweifaktoriellen Varianzanalyse**

| *Variabilität* | *Summen der Quadrate* | *df* | *mittlere Quadratsummen* | *F* | *p* |
|---|---|---|---|---|---|
| „Geschlecht" | 16,133 | 1 | 16,133 | 0,341 | 0,560 |
| „Hauptfach" | 765,8 | 2 | 382,9 | 8,089 | 0,002 |
| „Hauptfach nach Geschlecht" | 40,067 | 2 | 20,033 | 0,423 | 0,659 |
| gesamt | 1958,0 | 24 | | | |

Interpretation der Werte

Wir sehen uns nun mit 3 F-Werten konfrontiert, einem für die unabhängige Variable „Geschlecht“, einem für die unabhängige Variable „Hauptfach“ und einem für die Interaktion[3] zwischen „Geschlecht“ und „Hauptfach“. Gehen wir davon aus, wir hatten einen benötigten p-Wert von 0,05 bestimmt. Der F-Wert für „Hauptfach“ ist signifikant ($F$ = 8,089, $p$ = 0,002 – also $p$ < 0,05), die anderen beiden F-Werte sind es nicht, denn hierfür sind die p-Werte größer als 0,05. Dies bedeutet, dass die Variable „Hauptfach“ einen Effekt hat, nicht jedoch „Geschlecht“ oder „Hauptfach nach Geschlecht“. Man kann also sagen, dass (wie im ersten ANOVA-Beispiel, Kapitel 10.1.4.1) es einen Unterschied zwischen den Gruppen gibt, wenn man sie sich nur unter dem Gesichtspunkt „Hauptfach“ ansieht. Wir wissen aber noch nicht, ob alle Gruppen wirklich voneinander verschieden sind, sondern nur, dass auf jeden Fall zwei Gruppen differieren. Der F-Wert für „Geschlecht“ ist nicht signifikant, was bedeutet, dass es für die getesteten Englischkenntnisse keinen Unterschied zwischen männlichen und weiblichen Studenten gibt. Der F-Wert für die Interaktion schließlich ist auch nicht signifikant, obwohl wir nach Abb. 1 vermuten könnten,[4] dass es eine gewisse Interaktion gibt. (In diesem Fall sollte man sich bei einem Statistiker erkundigen, ob Ausreißer eine Interaktion „verstecken“.)

Weil man eine Post-hoc-Analyse nur durchführt, wenn der F-Wert signifikant ist, würden wir hier nur einen Folgetest für die Variable „Hauptfach“ vornehmen. Wir tun dies, wenn wir herausfinden wollen, ob die Geschichtsgruppen sich von den Politik- oder den Medienwissenschaftsgruppen unterscheiden (die Post-hoc-Analyse führen wir hier nicht durch, sondern gehen davon aus, dass sie nach der Hauptanalyse gemacht wurde).

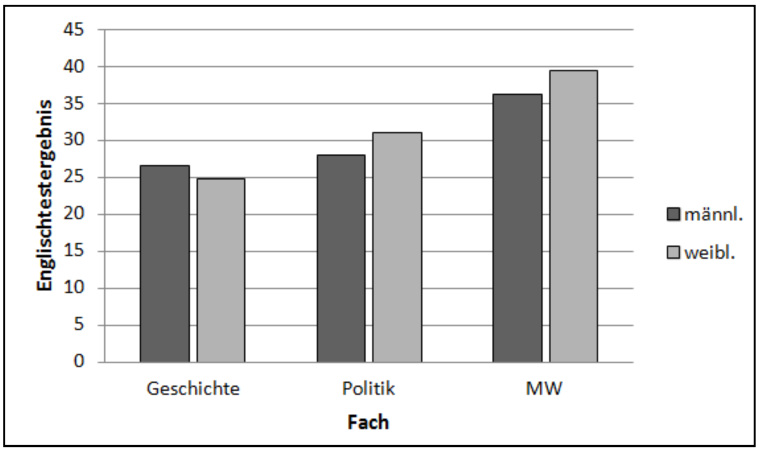


Abb. 1: Graph der Mittelwerte von sechs Gruppen

Unsere Ergebnisse können wie folgt für einen Forschungsbericht zusammengefasst werden:

[3] Der Begriff „Interaktion“ wurde kurz beschrieben in Kapitel 6.3.

[4] Das können wir daran sehen, dass die Linien nicht parallel verlaufen.

**Tabelle [Nr. X]: Ergebnisse des Englischkenntnis-Tests von männlichen und weiblichen Studenten verschiedener Hauptfächer (höchste erreichbare Punktzahl 50)**

| | Geschichte | Politik | Medienwissenschaft | Mittelwert |
|---|---|---|---|---|
| männliche Studenten | 26,6 | 28,0 | 36,2 | 30,3 |
| weibliche Studenten | 24,8 | 31,0 | 39,4 | 31,7 |
| Mittelwert | 25,7 | 29,5 | 37,8 | 31,0 |

**Eine zweifache ANOVA ergab einen Effekt von „Hauptfach" ($F$(2,24) = 8,09, $p$ < 0,05); siehe Tabelle [Nr. X]. Kein Effekt wurde für die Variable „Geschlecht" oder für eine Interaktion zwischen „Hauptfach" und „Geschlecht" gefunden. Ein anschließend durchgeführter Newman-Keuls-Test zeigte, dass die Medienwissenschaftsstudenten besser waren ($p$ < 0,05) als die Geschichtsstudenten und besser ($p$ < 0,05) als die Politikstudenten. Die Ergebnisse der beiden letztgenannten Gruppen waren jedoch nicht signifikant voneinander verschieden.**

### 10.1.4.3 Effektgröße: ANOVA und Eta zum Quadrat ($\eta^2$)

Das Konzept der *Effektgröße* wird im Kapitel 11 eingeführt und erklärt. Weil es aber im Kontext des Webkapitels sinnvoll ist, die Effektgröße direkt mit dem statistischen Test ANOVA zu besprechen, wird diese hier aufgeführt. Es empfiehlt sich jedoch zuerst die Lektüre vom Kapitel 11, bevor Sie hier weiter lesen.

Eine Maßzahl für die erklärte Varianz, d.h. ein Hinweis auf die Stärke des Effekts bei einer Varianzanalyse, ist Eta zum Quadrat ($\eta^2$). Man berechnet ihn in diesem Fall wie folgt:

$$\eta^2 = \frac{\text{Summe der Quadrate zwischen Gruppen}}{\text{Summe der Quadrate gesamt}}$$

Die *Summe der Quadrate zwischen Gruppen* und die *Summe der Quadrate gesamt* stammen aus der bereits durchgeführten ANOVA in Tabelle 2. Der entsprechende $\eta^2$-Wert ist also

$$\eta^2 = (382,9) / (1958) = 0,20$$

Wenn man diesen Wert mit 100 multipliziert, erhält man den Prozentsatz der erklärten Varianz, also in diesem Fall 20%. Um diesen Wert von $\eta^2$ zu interpretieren, schauen wir wieder in Tabelle 2 (oben) nach.

Wenn wir darüber berichten, dann zusammen mit dem F-Wert, z.B. so:

**Eine einfaktorielle ANOVA ergab, dass es einen Effekt von „Hauptfach" gab ($F$(2,27) = 8,67, $p$ < 0,01; $\eta^2$ = 0,20).**