# FAKTORIELLE VERSUCHSPLÄNE

Andreas Handl

# Inhaltsverzeichnis

# 1 Versuchsplanung

In der klassischen Qualitätskontrolle wird das Endprodukt vom Abnehmer kontrolliert. Hierdurch ist es möglich, dass ein Prozess lange Zeit unbemerkt nicht mehr unter Kontrolle ist und ein großer Teil der Produktion Ausschuss ist. Um dies zu verhindern hat Shewhart Qualitätsregelkarten entwickelt. Hierdurch bemerkt man während des Prozesses, ob noch alles unter Kontrolle ist. Die Kontrolle hat sich vom Ende des Prozesses in den laufenden Prozess verlagert. Nun wollen wir noch einen Schritt zurückgehen. Es soll der beste Prozess gewählt werden. Hierzu greift man auf Verfahren der Versuchsplanung zurück.

Das Ergebnis eines Prozesses hängt von einer Vielzahl von Einflussgrößen ab. Man nennt diese Einflussgrößen auch **Faktoren**. Wir bezeichnen Faktoren mit großen Buchstaben wie $A$ oder $B$. Die Ausprägungsmöglichkeiten $A_1, A_2, \ldots, A_I$ eines Faktors $A$ nennt man auch die **Faktorstufen**.

**Beispiel 1**
Bei der Fahrt zur Arbeit kann der Arbeitnehmer zwischen drei Strecken wählen. Er betrachtet also den Faktor $A$ mit den Faktorstufen $A_1$, $A_2$ und $A_3$. Dabei ist $A_i$ die $i$-te Strecke.

Das Ergebnis eines Prozesses beurteilen wir am Wert einer Zufallsvariablen $Y$, die wir auch Zielvariable nennen.

**Beispiel 1 (fortgesetzt)**
Der Arbeitnehmer ist an der Fahrzeit $Y$ interessiert.

Wir betrachten zunächst nur einen Faktor. Es soll untersucht werden, ob sich der Erwartungswert der Zielvariablen $Y$ auf den Faktorstufen unterscheidet. Bezeichnet man den Erwartungswert von $Y$ auf der $i$-ten Faktorstufe mit $\mu_i$, so soll getestet werden

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

gegen

$$H_1 : \mu_i \neq \mu_j$$

für mind. ein Paar $(i, j)$ mit $i \neq j$.

Zur Überprüfung der Hypothesen führen wir den Versuch auf jeder Faktorstufe durch. Hierbei ist es sinnvoll, ihn auf jeder Faktorstufe zu wiederholen. **Wiederholung** ist das *erste Prinzip der Versuchsplanung*.

Wir bezeichnen das Ergebnis der $j$-ten Wiederholung auf der $i$-ten Faktorstufe mit $y_{ij}$. dabei kann $i$ die Werte $1, 2, \ldots, I$ und $j$ die Werte $1, 2, \ldots, n_i$

annehmen. Es werden also $I$ Faktorstufen betrachtet. Außerdem kann die Anzahl der Beobachtungen auf den Faktorstufen unterschiedlich sein.

**Beispiel 1 (fortgesetzt)**
Um zu entscheiden, bei welcher Strecke die mittlere Fahrzeit geringer ist, fährt er auf jeder Strecke genau fünfmal zur Arbeit. In Tabelle 1 sind die Beobachtungen zu finden. .

Tabelle 1: Fahrzeiten eines Arbeitnehmers auf drei Strecken

| Strecke | Fahrzeit | | | | |
|---|---|---|---|---|---|
| 1 | 38 | 44 | 40 | 41 | 37 |
| 2 | 44 | 43 | 47 | 50 | 41 |
| 3 | 44 | 40 | 41 | 42 | 38 |

Es gilt also speziell $y_{23} = 47$ und $y_{32} = 40$.

Wir müssen aber noch ein weiteres Prinzip beachten. Wir wollen herausfinden, ob sich der Erwartungswert der Zielgröße auf den Faktorstufen unterscheidet. Um sicherzustellen, dass ein Unterschied in den Erwartungswerten $\mu_i$ ausschließlich an den unterschiedlichen Faktorstufen liegt, müssen wir während des versuchs alle anderen Einflussgrößen konstant halten. Würde der Arbeitnehmer nämlich alle Fahrten auf der ersten Strecke am Montag, alle Fahrten auf der zweiten Strecke am Dienstag und alle Fahrten auf der dritten Strecke am Mittwoch zurücklegen, so könnte er nicht entscheiden, ob ein Unterschied in den Erwartungswerten durch die Strecken oder durch die Wochentagen bewirkt wird.

Eine Möglichkeit zur Vermeidung dieses Problems besteht darin, die Strecken zufällig auf die Tage zu verteilen. Hierdurch soll sichergestellt werden, dass sich alle Einflussgrößen gleichmäßig auf die Faktorstufen verteilen. Diese Vorgehensweise nennt man **Randomisierung**. Das ist das *zweite Prinzip der Versuchsplanung*.

Eine andere Möglichkeit besteht darin, den Effekt des Wochentages dadurch konstant zu halten, dass jede der drei Strecken an jedem Wochentag benutzt wird. Man spricht von **Blockbildung**. Dies ist das *dritte Prinzip der Versuchsplanung*.

Wir werden im Folgenden Versuche betrachten, bei denen die ersten beiden Prinzipien beachtet werden.

# 2 Einfaktorielle Varianzanalyse

## 2.1 Die Annahmen

Um die Hypothese

$$H_0 : \mu_1 = \ldots = \mu_I \tag{1}$$

überprüfen zu können, müssen wir bestimmte Annahmen machen.
Wir gehen im Folgenden davon aus, dass die Beobachtungen $y_{ij}$ Realisationen von unabhängigen Zufallsvariablen $Y_{ij}$ sind, die mit Erwartungswert $\mu_i$, $i = 1, \ldots, I$ und Varianz $\sigma^2$ normalverteilt sind. Die Erwartungswerte auf den Faktorstufen können sich also unterscheiden, während die Varianz identisch sein muss.

## 2.2 Die ANOVA-Tabelle und der $F$-Test

Es liegt nahe zur Überprüfung der Hypothese (1) die Mittelwerte

$$\boxed{\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}} \tag{2}$$

auf den einzelnen Faktorstufen zu bestimmen und zu vergleichen.

**Beispiel 1 (fortgesetzt)**
Es gilt $\bar{y}_1 = 40$, $\bar{y}_2 = 45$ und $\bar{y}_3 = 41$.

Der Vergleich von zwei Mittelwerten $\bar{y}_1$ und $\bar{y}_2$ ist einfach. Wir bilden die Differenz $\bar{y}_2 - \bar{y}_1$ der beiden Mittelwerte. Bei mehr als zwei Faktorstufen können wir alle Paare von Faktorstufen betrachten und $\bar{y}_i$ mit $\bar{y}_j$ für $i < j$ vergleichen. Hierdurch erhalten wir aber kein globales Maß für den Vergleich aller Faktorstufen. Dieses gewinnen wir dadurch, dass wir die Mittelwerte $\bar{y}_i$, $i = 1, \ldots, I$ als eine Stichprobe auffassen und bestimmen, wie stark sie um den Mittelwert

$$\boxed{\bar{y} = \frac{1}{N} \sum_{i=1}^{I} \sum_{j=1}^{n_i} y_{ij}} \tag{3}$$

aller Beobachtungen streuen. Dabei gilt

$$N = n_1 + n_2 + \ldots + n_I$$

**Beispiel 1 (fortgesetzt)**
Es gilt $\bar{y} = 42$.

Es liegt nahe, die Streuung der Mittelwerte $\bar{y}_i$ um das Gesamtmittel $\bar{y}$ folgendermaßen zu bestimmen:

$$\sum_{i=1}^{I} (\bar{y}_i - \bar{y})^2.$$

Hierbei wird aber nicht berücksichtigt, dass die Anzahl $n_i$ der Beobachtungen auf den Faktorstufen unterschiedlich groß sein können. Eine große Anzahl $n_i$ an Beobachtungen auf einer Faktorstufe sollte ein stärkeres Gewicht erhalten als eine kleine. Wir bilden also

$$\boxed{SS_A = \sum_{i=1}^{I} n_i (\bar{y}_i - \bar{y})^2.} \tag{4}$$

Man bezeichnet $SS_A$ als *Streuung zwischen den Stufen des Faktors A*.

**Beispiel 1 (fortgesetzt)**
Es gilt

$$SS_A = 5 \cdot (40 - 42)^2 + 5 \cdot (45 - 42)^2 + 5 \cdot (41 - 42)^2 = 70$$

Wie das folgende Beispiel zeigt, ist die Größe $SS_A$ allein aber keine geeignete Teststatistik zur Überprüfung der Hypothese (1) auf Seite 6.

**Beispiel 2**
In der Tabelle 2 sind die Werte eines Merkmales auf drei Faktorstufen zu finden ist.

Tabelle 2: Fahrzeiten eines Arbeitnehmers auf drei Strecken

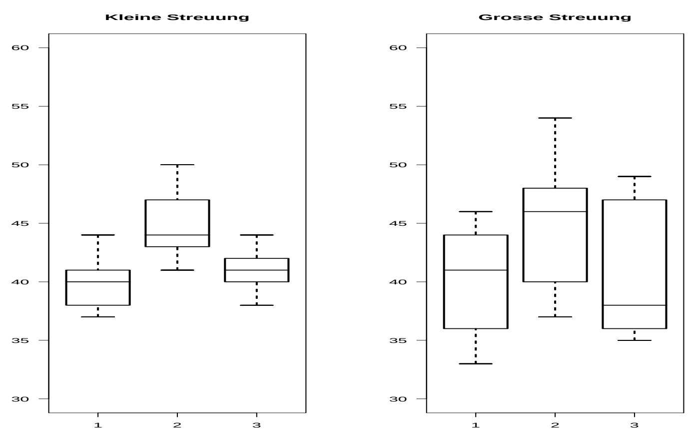
| Strecke | Fahrzeit | | | | |
|---|---|---|---|---|---|
| 1 | 41 | 33 | 44 | 36 | 46 |
| 2 | 37 | 46 | 54 | 48 | 40 |
| 3 | 47 | 49 | 38 | 36 | 35 |

Es gilt

$$\bar{y}_1 = 40, \quad \bar{y}_2 = 45, \quad \bar{y}_3 = 41, \quad \bar{y} = 42.$$

Also ist auch in beiden Tabellen der Wert von $SS_A$ identisch. Wie die Abbildung 1 zeigt, unterscheiden sich die beiden Situationen beträchtlich. Die Boxplots im ersten Bild in Abbildung 1 verdeutlichen, dass die Streuung auf den Faktorstufen klein ist, während im rechten Bild in Abbildung 1 die Streuung auf den Faktorstufen groß ist. Das linke Bild in Abbildung 1 spricht für einen Lageunterschied zwischen den Faktorstufen, während die unterschiedlichen Mittelwerte im rechten Bild in Abbbildung 1 eher durch die hohen Streuungen erklärt werden können.

Abbildung 1: Zwei Situationen




Wir müssen also neben der Streuung zwischen den Gruppen die Streuung innerhalb der Gruppen berücksichtigen. Die Streuung innerhalb der $i$-ten Gruppe messen wir durch

$$\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 . \tag{5}$$

Summieren wir (5) über alle Gruppen, so erhalten wir

$$\boxed{SS_R = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 .} \tag{6}$$

Wir nennen $SS_R$ auch **Streuung innerhalb der Gruppen** oder **unerklärte Reststreuung**

**Beispiel 1 (fortgesetzt)**
Es gilt

$$\begin{aligned} SS_R &= (38-40)^2 + (44-40)^2 + (40-40)^2 + (41-40)^2 + (37-40)^2 \\ &+ (44-45)^2 + (43-45)^2 + (47-45)^2 + (50-45)^2 + (41-45)^2 \\ &+ (44-41)^2 + (40-41)^2 + (41-41)^2 + (42-41)^2 + (38-41)^2 \\ &= 100 \end{aligned}$$

Mit

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \tag{7}$$

können wir $SS_R$auch folgendermaßen bestimmen

$$SS_R = \sum_{i=1}^{I} (n_i - 1) \cdot s_i^2$$

**Beispiel 1 (fortgesetzt)**
Es gilt

$$\begin{aligned} s_1^2 &= \frac{1}{4} \left[ (38-40)^2 + (44-40)^2 + (40-40)^2 + (41-40)^2 + (37-40)^2 \right] \\ &= 7.5 \end{aligned}$$

Analog erhalten wir $s_2^2 = 12.5$ und $s_3^2 = 5$.
Also gilt

$$SS_R = 4 \cdot 7.5 + 4 \cdot 12.5 + 4 \cdot 5 = 100$$

Die Gesamtstreuung messen wir durch:

$$\boxed{SS_T = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 .} \tag{8}$$

**Beispiel 1 (fortgesetzt)**
Es gilt

$$\begin{aligned} SS_T &= (38-42)^2 + (44-42)^2 + (40-42)^2 + (41-42)^2 + (37-42)^2 \\ &+ (44-42)^2 + (43-42)^2 + (47-42)^2 + (50-42)^2 + (41-42)^2 \\ &+ (44-42)^2 + (40-42)^2 + (41-42)^2 + (42-42)^2 + (38-42)^2 \\ &= 170 \end{aligned}$$

Im Beispiel gilt

$$\boxed{SS_T = SS_A + SS_R\,.} \tag{9}$$

Dies ist kein Zufall. Diese Beziehung gilt allgemein. Dies wird auf Seite 104 gezeigt.

Eine geeignete Teststatistik erhält man nun, indem man die mittleren Streuungen vergleicht, wobei der Mittelwert unter der Nebenbedingung bestimmt wird, wie viele der Summanden frei gewählt werden können. Die Streuung zwischen den Faktorstufen setzt sich aus $I$ Summanden zusammen, von denen aber nur $I-1$ frei gewählt werden können, da sich der Mittelwert auf der $I$-ten Faktorstufe aus

$$\bar{y}, \bar{y}_1, \ldots, \bar{y}_{I-1}$$

ergibt. Die Streuung auf den Faktorstufen setzt sich aus $N$ Summanden zusammen. Auf der $i$-ten Faktorstufe ergibt sich aber $y_{in_i}$ aus der Kenntnis von

$$y_{i1}, \ldots, y_{in_i-1}, \bar{y}_i\,.$$

Somit sind von den $N$ Summanden nur $N-I$ frei wählbar. Wir erhalten also

$$\boxed{MSS_A = \frac{SS_A}{I-1}} \tag{10}$$

und

$$\boxed{MSS_R = \frac{SS_R}{N-I}} \tag{11}$$

**Beispiel 1 (fortgesetzt)**
Es gilt $MSS_A = 70/2 = 35$ und $MSS_R = 100/12 = 8.33$.

Wir überprüfen die Hypothese

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

mit der Teststatistik

$$\boxed{F = \frac{MSS_A}{MSS_R} = \frac{\dfrac{1}{I-1}\displaystyle\sum_{i=1}^{I} n_i\,(\bar{Y}_i - \bar{Y})^2}{\dfrac{1}{N-I}\displaystyle\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2}.} \tag{12}$$

Ist die mittlere Streuung zwischen den Stichproben groß im Verhältnis zur mittleren Streuung innerhalb der Stichproben, so wird die Nullhypothese identischer Erwartungswerte abgelehnt. Unter der Nullhypothese ist die Teststatistik in Gleichung (12) mit $I-1$ und $N-I$ Freiheitsgraden $F$-verteilt.
Wir lehnen die Hypothese

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

zum Niveau $\alpha$ ab, wenn gilt

$$F \geq F_{I-1,N-I;1-\alpha}$$

Dabei $F_{I-1,N-I;1-\alpha}$ das $1-\alpha$-Quantil der $F$-Verteilung mit $I-1$ und $N-I$ Freiheitsgraden ist. Das 0.95-Quantil der $F$-Verteilung ist in Tabelle 48 auf Seite 110 zu finden.

**Beispiel 1 (fortgesetzt)**
Es gilt

$$F = \frac{35}{8.33} = 4.2\,.$$

Der Tabelle 48 auf Seite 110 entnehmen wir $F_{2,12;0.95} = 3.89$. Wir lehnen die Hypothese identischer Erwartungswerte auf den Faktorstufen also ab.

Man spricht auch vom **F-Test** und der **Varianzanalyse**, da die Teststatistik das Verhältnis von zwei Schätzern der Varianz $\sigma^2$ ist.
Die Ergebnisse einer Varianzanalyse werden in einer ANOVA-Tabelle zusammengestellt. Dabei steht ANOVA für Analysis Of Variance. Tabelle 3 zeigt den allgemeinen Aufbau einer ANOVA-Tabelle.

Tabelle 3: Allgemeiner Aufbau einer ANOVA-Tabelle

| Quelle der Variation | Quadrat-summen | Freiheits-grade | Mittlere Quadratsummen | $F$ |
|---|---|---|---|---|
| zwischen den Faktorstufen | $SS_A$ | $I-1$ | $MSS_A$ | $\frac{MSS_A}{MSS_R}$ |
| innerhalb der Faktorstufen | $SS_R$ | $N-I$ | $MSS_R$ | |
| Gesamt | $SS_T$ | $N-1$ | | |

**Beispiel 1 (fortgesetzt)**
In Tabelle 4 ist die ANOVA-Tabelle zu finden.

Tabelle 4: ANOVA-Tabelle für den Vergleich der Fahrzeit auf den 3 Strecken bei kleiner Streuung

| Quelle der Variation | Quadrat-summen | Freiheits-grade | Mittlere Quadratsummen | $F$ |
|---|---|---|---|---|
| zwischen den Faktorstufen | 70 | 2 | 35.00 | 4.2 |
| innerhalb der Faktorstufen | 100 | 12 | 8.33 | |
| Gesamt | 170 | 14 | | |

**Beispiel 2 (fortgesetzt)**
Es gilt

$$\begin{aligned} SS_R &= (41-40)^2+(33-40)^2+(44-40)^2+(36-40)^2+(46-40)^2 \\ &+ (37-45)^2+(46-45)^2+(54-45)^2+(48-45)^2+(40-45)^2 \\ &+ (47-41)^2+(49-41)^2+(38-41)^2+(36-41)^2+(35-41)^2 \\ &= 468 \end{aligned}$$

Somit erhalten wir folgende ANOVA-Tabelle.

Tabelle 5: ANOVA-Tabelle für den Vergleich der Fahrzeit auf den 3 Strecken bei großer Streuung

| Quelle der Variation | Quadrat-summen | Freiheits-grade | Mittlere Quadratsummen | $F$ |
|---|---|---|---|---|
| zwischen den Faktorstufen | 70 | 2 | 35.00 | 0.8974 |
| innerhalb der Faktorstufen | 468 | 12 | 39 | |
| Gesamt | 538 | 14 | | |

Wie zu erwarten war, lehnen wir wegen $F_{2,12;0.95} = 3.89$ die Hypothese identischer Erwartungswerte auf den Faktorstufen also nicht ab.

## 2.3 Versuche mit zwei Faktorstufen und identischen Stichprobenumfängen

### 2.3.1 Das Modell

Sehr oft betrachtet man Faktoren mit zwei Faktorstufen. Man bezeichnet die Faktorstufen mit `-` und `+`.

**Beispiel 3**
Der Arbeitnehmer kann sich zwischen zwei Strecken entscheiden. Die Strecke ist ein Faktor $A$ mit den Stufen Strecke 1, also `-`, und Strecke 2, also `+`.

Wir bezeichnen die Zielvariable auf der Faktorstufe `-` mit $Y_1$ und auf der Faktorstufe `+` mit $Y_2$. Bei zwei Faktorstufen ist man am Effekt des Faktors $A$ interessiert. Der Effekt $E_A$ des Faktors $A$ gibt an, wie sich die Zielvariable $Y$ im Mittel ändert, wenn man von der Faktorstufe `-` auf die Faktorstufe `+` übergeht. Es gilt also

$$\boxed{E_A = \mu_2 - \mu_1 \,.} \tag{13}$$

Zwei Fragen sind von Interesse:

1. Wie groß ist der Effekt des Faktors $A$?
2. Ist $E_A$ signifikant von 0 verschieden?

Zur Beantwortung beider Fragen führen wir den Versuch auf beiden Faktorstufen durch, wobei wir die ersten beiden Prinzipien der Versuchsplanung beachten. Wir beobachten also die Realisierungen $y_{ij}$ der Zufallsvariablen $Y_{ij}$. Im Folgenden gehen wir von folgender Annahme aus:

**Annahme 1**
Die $Y_{ij}$ mit $i = 1, 2$, und $j = 1, 2, \ldots, n$ sind normalverteilt ist den Parametern $\mu_i$ und $\sigma^2$.

Wir gehen also davon aus, dass die Anzahl der Beobachtungen auf jeder Faktorstufe gleich ist.

**Beispiel 3 (fortgesetzt)**
Er fährt jede Strecke fünfmal und erhält bei Strecke 1 folgende

```
38 44 40 41 37
```

und bei Strecke 2 folgende

```
44 43 47 50 41
```

Werte.

### 2.3.2 Der Schätzer des Effekts von $A$

Wenden wir uns der Beantwortung der Frage 1 auf Seite 13 zu.
Hierzu stellen wir für $i = 1, 2$ und $j = 1, \ldots, n$ folgendes Modell auf:

$$Y_{ij} = \mu_i + \epsilon_{ij} \tag{14}$$

Dabei ist $\mu_1$ der Erwartungswert von $Y$ auf der ersten und $\mu_2$ der Erwartungswert von $Y$ auf der zweiten Faktorstufe. Die $\epsilon_{ij}$ sind die Störgrößen, die alle anderen Einflussgrößen umfassen. Wir unterstellen, dass die $\epsilon_{ij}$ unabhängig und identisch normalverteilt mit $E(\epsilon_{ij}) = 0$ und $Var(\epsilon_{ij}) = \sigma^2$ sind.
Hieraus folgt, dass die $Y_{ij}$ mit $E(Y_{ij}) = \mu_i$ und $Var(Y_{ij}) = \sigma^2$ normalverteilt sind.
Im Modell (14) erfüllen die $Y_{ij}$ also die Annahme 1 auf Seite 13.
Wir schätzen $\mu_1$ und $\mu_2$ im Modell (14) nach der Methode der Kleinsten Quadrate. Wir suchen also die Werte von $\mu_1$ und $\mu_2$, für die

$$\sum_{j=1}^{n} (y_{1j} - \mu_1)^2 + \sum_{j=1}^{n} (y_{2j} - \mu_2)^2 \tag{15}$$

minimal wird.
Es gilt

$$\boxed{\hat{\mu}_i = \bar{y}_i = \frac{1}{n} \sum_{j=1}^{n} y_{ij} \,.} \tag{16}$$

Dies wird auf Seite 105 gezeigt.

**Beispiel 3 (fortgesetzt)**
Es gilt $\bar{y}_1 = 40$ und $\bar{y}_2 = 45$.

Wir schätzen den Effekt eines Faktors durch die Differenz der jeweiligen Mittelwerte auf den beiden Faktorstufen. Der geschätzte Effekt $e_A$ des Faktors $A$ gleich

$$\boxed{e_A = \bar{y}_2 - \bar{y}_1} \tag{17}$$

**Beispiel 3 (fortgesetzt)**
Der geschätzte Effekt von $A$ ist gleich

$$e_A = \bar{y}_2 - \bar{y}_1 = 45 - 40 = 5$$

Für Strecke 2 benötigt man Im Mittel 5 Minuten länger als für Strecke 1.

### 2.3.3 Der $F$-Test

Wenden wir uns der Frage 2 auf Seite 13 zu. Wir wollen also die Hypothese

$$H_0 : E_A = 0$$

überprüfen.
Wegen

$$E_A = \mu_2 - \mu_1$$

ist diese Hypothese äquivalent zu

$$H_0 : \mu_1 = \mu_2$$

Wir können also den $F$-Test aus Kapitel 2.2 anwenden. Bei zwei Faktorstufen und identischer Anzahl $n$ von Beobachtungen auf den beiden Faktorstufen vereinfacht sich die Formel von $SS_A$. Es gilt

$$\boxed{SS_A = \frac{n\,(\bar{y}_2 - \bar{y}_1)^2}{2}} \tag{18}$$

Der Beweis ist auf Seite 106 zu finden.

**Beispiel 3 (fortgesetzt)**
Wir bestimmen $SS_A$ auf zwei Arten.
Wir beginnen mit Gleichung (4) auf Seite 7. Es gilt $\bar{y} = 42.5$. Also folgt

$$SS_A = 5 \cdot (40 - 42.5)^2 + 5 \cdot (45 - 42.5)^2 = 62.5$$

Und jetzt verwenden wir die Gleichung (18).

$$SS_A = \frac{5 \cdot (45 - 40)^2}{2} = 62.5$$

Wir können jetzt auch die ANOVA-Tabelle aufstellen. Hierzu benötigen wir noch $SS_R$. Es gilt

$$\begin{aligned} SS_R &= (38-40)^2 + (44-40)^2 + (40-40)^2 + (41-40)^2 + (37-40)^2 \\ &+ (44-45)^2 + (43-45)^2 + (47-45)^2 + (50-45)^2 + (41-45)^2 \\ &= 80 \end{aligned}$$

Die ANOVA-Tabelle ist in Tabelle 6 auf Seite 16 zu finden.

Tabelle 6: ANOVA-Tabelle für den Vergleich der Fahrzeit auf den 2 Strecken

| Quelle der Variation | Quadrat-summen | Freiheits-grade | Mittlere Quadratsummen | $F$ |
|---|---|---|---|---|
| zwischen den Faktorstufen | 62.5 | 1 | 62.5 | 6.25 |
| innerhalb der Faktorstufen | 80 | 8 | 10 | |
| Gesamt | 142.5 | 9 | | |

Wegen $F_{1,8;0.95} = 5.32$ lehnen wir zum Signifikanzniveau 0.05 die Hypothese ab, dass der Effekt von $A$ gleich 0 ist.

#### 2.3.4 Der Algorithmus von Yates

Von Yates wurde ein Algorithmus vorgeschlagen, mit dem man bei einem faktoriellen Versuchsplan die Schätzer der Effekte und die Quadratsummen schnell bestimmen kann. Schauen wir uns hierzu zunächst noch einmal den Schätzer des Effekts von $A$ und $SS_A$ an.
Der Schätzer des Effekts von $A$ ist

$$\boxed{e_A = \frac{\sum_{j=1}^{n} y_{2j} - \sum_{j=1}^{n} y_{1j}}{n}} \tag{19}$$

Dies sieht man folgendermaßen

$$e_A = \bar{y}_2 - \bar{y}_1 = \frac{1}{n} \sum_{j=1}^{n} y_{2j} - \frac{1}{n} \sum_{j=1}^{n} y_{1j} = \frac{\sum_{j=1}^{n} y_{2j} - \sum_{j=1}^{n} y_{1j}}{n} .$$

Für $SS_A$ gilt

$$\boxed{SS_A = \frac{\left(\sum_{j=1}^{n} y_{2j} - \sum_{j=1}^{n} y_{1j}\right)^2}{2n}} \tag{20}$$

Der Beweis ist auf Seite 107 zu finden.
Wir sehen, dass beide Ausdrücke von

$$\boxed{K_A = \sum_{j=1}^{n} y_{2j} - \sum_{j=1}^{n} y_{1j}} \tag{21}$$

abhängen. Man nennt $K_A$ auch den **Kontrast** des Faktors $A$. Dieser ist gleich der Differenz aus der Summe der Beobachtungen auf der Faktorstufe + und der Summe der Beobachtungen auf der Faktorstufe −.
Wir schreiben im Folgenden für die Summe

$$\sum_{j=1}^{n} y_{2j}$$

der Beobachtungen auf der Faktorstufe + das Symbol $a$ und für die Summe

$$\sum_{j=1}^{n} y_{1j}$$

der Beobachtungen auf der Faktorstufe − das Symbol (1). Dabei symbolisiert das $a$, dass der Faktor $A$ auf + steht.

**Beispiel 3 (fortgesetzt)**
Es gilt

$$(1) = 38 + 44 + 40 + 41 + 37 = 200$$

und

$$a = 44 + 43 + 47 + 50 + 41 = 225$$

Mit Hilfe von (1) und $a$ können wir nun den Kontrast von $A$ bestimmen, indem wir diese in die Gleichungen (19) und (20) auf Seite 16 einsetzen. Wir erhalten

$$\boxed{K_A = a - (1)} \tag{22}$$

Es gilt

$$\boxed{e_A = \frac{K_A}{n}}$$

und

$$\boxed{SS_A = \frac{K_A^2}{2n}}$$

Mit dem **Algorithmus von Yates** kann man auch für Versuchspläne mit mehr als einem Faktor die Kontraste der Faktoren bestimmen. Wir wählen deshalb eine Form der Darstellung, die auf mehr als einen Faktor verallgemeinert werden kann. Hierzu stellen wir folgende Tabelle auf:

Tabelle 7: Die Ausgangstabelle beim Algorithmus von Yates bei einem $2^1$-Versuchsplan

| |
|---|
| (1) |
| $a$ |

Nun erzeugen wir eine weitere Spalte. Die erste Zahl in dieser Spalte ist gleich der Summe der Zahlen in der ersten Spalte. Die zweite Zahl ist die Differenz aus der zweiten Zahl und der ersten Zahl. Wir erhalten Tabelle 8 auf Seite 18.

Tabelle 8: Der erste Schritt beim Algorithmus von Yates bei einem einfaktoriellen Versuchsplan

| | |
|---|---|
| (1) | $(1) + a$ |
| $a$ | $a - (1)$ |

Bei einem Versuchsplan mit einem Faktor sind wir nach diesem Schritt fertig. Wir sehen, dass neben $A$ der Kontrast von $A$ steht.

**Beispiel 3 (fortgesetzt)**
Wir erhalten folgende Tabelle

Tabelle 9: Der erste Schritt beim Algorithmus von Yates bei einem einfaktoriellen Versuchsplan

| | | |
|---|---|---|
| (1) | 200 | 425 |
| $a$ | 225 | 25 |

Es gilt also $K_A = 25$.
Also gilt

$$e_A = \frac{25}{5} = 5$$

und

$$SS_A = \frac{25^2}{2 \cdot 5} = 62.5$$

# 3 Zweifaktorielle Varianzanalyse

Bisher sind wir davon ausgegangen, dass die Zielgröße nur von einem Faktor abhängt. In der Regel werden aber mehrere Faktoren einen Einfluss haben. Wir wollen in diesem Kapitel darstellen, wie man vorgeht, wenn zwei Faktoren in Betracht gezogen werden. Dabei werden wir davon ausgehen, dass jeder der Faktoren zwei Faktorstufen besitzt. Im Folgenden sind $A$ und $B$ Faktoren mit den Faktorstufen $A_1$ und $A_2$ beziehungsweise $B_1$ und $B_2$. Wir bezeichnen $A_1$ und $B_1$ mit - und $A_2$ und $B_2$ mit +.

**Beispiel 4**
Der Arbeitnehmer will überprüfen, ob außer der Strecke $A$ auch der Zeitpunkt $B$ der Abfahrt einen Einfluss auf die Fahrzeit hat. Dabei zieht er die Zeitpunkte $B_1$ und $B_2$ in Betracht, wobei $B_1$ die frühere Abfahrzeit bedeutet.

Wir beschäftigen uns mit **faktoriellen Versuchsplänen**. Bei diesen wird jede Kombination der Faktoren betrachtet. Hat der Faktor $A$ also die Faktorstufen - und + und der Faktor $B$ die Faktorstufen - und +, so gibt es die **Faktorstufenkombinationen**

```
- -
+ -
- +
+ +
```

Dabei steht das erste Symbol für den Faktor $A$ und das zweite für den Faktor $B$. Auf jeder Faktorstufenkombination werden $n$ Versuche durchgeführt. Dabei bezeichnen wir die Realisation, die die Zielgröße $Y$ auf der $i$-ten Faktorstufe von $A$ und der $j$-ten Faktorstufe von $B$ bei der $k$-ten Wiederholung annimmt, mit $y_{ijk}$. Der Index $i$ nimmt also den Wert 1 an, wenn $A$ auf - steht. Steht $A$ auf +, so gilt $i = 2$. Entsprechend bedeutet $j = 1$, dass $B$ auf - steht, und $j = 2$, dass $B$ auf + steht. Wir stellen die Daten in einer Tabelle dar, deren allgemeine Form in Tabelle 10 zu finden ist.

Tabelle 10: Tabelle für einen zweifaktoriellen Versuch

| $A$ | $B$ | Merkmal | | | |
|---|---|---|---|---|---|
| − | − | $y_{111}$ | $y_{112}$ | $\cdots$ | $y_{11n}$ |
| + | − | $y_{211}$ | $y_{212}$ | $\cdots$ | $y_{21n}$ |
| − | + | $y_{121}$ | $y_{122}$ | $\cdots$ | $y_{12n}$ |
| + | + | $y_{221}$ | $y_{222}$ | $\cdots$ | $y_{22n}$ |

**Beispiel 4 (fortgesetzt)**
Der Arbeitnehmer fährt auf jeder Faktorstufenkombination genau fünfmal und erhält die Daten in Tabelle

Tabelle 11: Ergebnis eines zweifaktoriellen Versuches

| $A$ | $B$ | Fahrzeit | | | | |
|---|---|---|---|---|---|---|
| − | − | 38 | 44 | 40 | 41 | 37 |
| + | − | 43 | 42 | 46 | 49 | 40 |
| − | + | 44 | 37 | 41 | 43 | 40 |
| + | + | 44 | 48 | 47 | 45 | 51 |

Wir beschäftigen uns mit zwei Modellen für $Y_{ijk}$. Diese sind Erweiterungen des einfaktoriellen Modells der Varianzanalyse. Schauen wir uns dieses also noch einmal an. Es lautet

$$Y_{ij} = \mu_i + \epsilon_{ij} \tag{23}$$

Hierin beschreibt der Index $i$ die Stufen des Faktors $A$ und der Index $j$ die Wiederholungen auf den Faktorstufen.
Es liegt nahe, das Modell der einfaktoriellen Varianzanalyse folgendermaßen auf zwei Faktoren zu erweitern:

$$Y_{ijk} = \mu_{ij} + \epsilon_{ijk}$$

Gilt $E(\epsilon_{ijk}) = 0$, so ist $\mu_{ijk}$ der Erwartungswert von $Y$ auf der Faktorstufenkombination $(i, j)$. Wir wollen $\mu_{ijk}$ in Abhängigkeit von den Faktoren $A$ und $B$ angeben. Hierzu stellen wir das Modell der einfaktoriellen Varianzanalyse in Gleichung (23) anders dar. Wir setzen

$$\begin{aligned} \mu_i &= \mu_i + \mu - \mu = \mu + \mu_i - \mu \\ &= \mu + \alpha_i \end{aligned} \tag{24}$$

mit

$$\alpha_i = \mu_i - \mu$$

Setzen wir Gleichung (24) für $\mu_i$ in Gleichung (23) ein, so erhalten wir folgendes Modell

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \tag{25}$$

Mit

$$\mu = 0.5 \cdot (\mu_1 + \mu_2)$$

gilt

$$\alpha_1 + \alpha_2 = \mu_1 - \mu + \mu_2 - \mu = 2 \cdot \frac{\mu_1 + \mu_2}{2} - 2 \cdot \mu = 0$$

Wir müssen im Modell (25) also folgende Nebenbedingung berücksichtigen

$$\alpha_1 + \alpha_2 = 0 \tag{26}$$

Wir können im Modell (25) den Effekt $E_A$ in Abhängigkeit von $\alpha_1$ und $\alpha_2$ angeben. Es gilt

$$E_A = \alpha_2 - \alpha_1 \tag{27}$$

Dies sieht man folgendermaßen:

$$E_A = \mu_2 - \mu_1 \overset{(24)}{=} \mu + \alpha_2 - (\mu + \alpha_1) = \alpha_2 - \alpha_1$$

## 3.1 Das additive Modell

### 3.1.1 Das Modell

Im Modell der einfaktoriellen Varianzanalyse können wir den Effekt von $A$ durch die $\alpha_i$ ausdrücken. Wollen wir einen weiteren Faktor $B$ betrachten, so ergänzen wir das Modell in Gleichung (23) um einen Term $\beta_j$:

$$\boxed{Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}} \tag{28}$$

Dabei gilt $i = 1, 2$, $j = 1, 2$ und $k = 1, \ldots, n$. In Analogie zum einfaktoriellen Modell müssen die Parameter $\alpha_1$, $\alpha_2$, $\beta_1$ und $\beta_2$ folgenden Nebenbedingungen gehorchen

$$\boxed{\alpha_1 + \alpha_2 = 0} \tag{29}$$

$$\boxed{\beta_1 + \beta_2 = 0} \tag{30}$$

Wir unterstellen $E(\epsilon_{ijk}) = 0$ und $Var(\epsilon_{ijk}) = \sigma^2$.
Hieraus folgt

$$\mu_{ij} = E\left(Y_{ijk}\right) = \mu_{ij} = \mu + \alpha_i + \beta_j \tag{31}$$

und

$$Var\left(Y_{ijk}\right) = \sigma^2 \tag{32}$$

Wir wollen nun im Modell (28) den Effekt $E_A$ von $A$ beschreiben. Hierzu setzen wir zuerst den Faktor $B$ auf -. Dann haben wir es mit einem einfaktoriellen Modell zu tun und wissen, wie wir den Effekt von $A$ bestimmen. Wir bilden

$$\mu_{21} - \mu_{11} \tag{33}$$

Nun setzen wir den Wert des Faktors $B$ auf + und bestimmen den Effekt von $A$:

$$\mu_{22} - \mu_{12} \tag{34}$$

Den Effekt von $A$ im zweifaktoriellen Modell definieren wir als Mittelwert des Effekts von $A$, wenn $B$ auf - steht, und des Effekts von $A$, wenn $B$ auf + steht:

$$\boxed{E_A = \frac{\mu_{21} - \mu_{11} + \mu_{22} - \mu_{12}}{2}} \tag{35}$$

Nun müssen wir uns noch fragen, ob diese Wahl sinnvoll ist. Wir setzen Gleichung (31) in (33) ein und erhalten

$$\mu_{21} - \mu_{11} = \mu + \alpha_2 + \beta_1 - (\mu + \alpha_1 + \beta_1) = \alpha_2 - \alpha_1$$

Setzen wir Gleichung (31) in (34) ein, so ergibt sich

$$\mu_{22} - \mu_{12} = \mu + \alpha_2 + \beta_2 - (\mu + \alpha_1 + \beta_2) = \alpha_2 - \alpha_1$$

Wir sehen, dass im additiven Modell der Effekt von $A$ auf beiden Faktorstufen von $B$ identisch ist. Somit ist es sinnvoll, den Effekt von $A$ wie in Gleichung (35) zu bestimmen.

Offensichtlich gilt

$$\boxed{E_A = \alpha_2 - \alpha_1} \tag{36}$$

Schauen wir uns den Effekt von $B$ an. Steht $A$ auf -, so ist der Effekt von $B$ gleich

$$\mu_{12} - \mu_{11} \tag{37}$$

Steht $A$ auf +, so ist der Effekt von $B$ gleich

$$\mu_{22} - \mu_{21} \tag{38}$$

Wir definieren den Effekt von $B$ durch

$$\boxed{E_B = \frac{\mu_{12} - \mu_{11} + \mu_{22} - \mu_{21}}{2}} \tag{39}$$

Auch der Effekt von $B$ ist im additiven Modell auf beiden Stufen von $A$ identisch. Es gilt nämlich

$$\mu_{12} - \mu_{11} = \mu + \alpha_1 + \beta_2 - (\mu + \alpha_1 + \beta_1) = \beta_2 - \beta_1$$

und

$$\mu_{22} - \mu_{21} = \mu + \alpha_2 + \beta_2 - (\mu + \alpha_2 + \beta_1) = \beta_2 - \beta_1$$

Offensichtlich gilt

$$\boxed{E_B = \beta_2 - \beta_1} \tag{40}$$

**Beispiel 5**
Es gelte $\mu_{11} = 3$, $\mu_{21} = 7$, $\mu_{12} = 5$ und $\mu_{22} = 9$.
Somit gilt

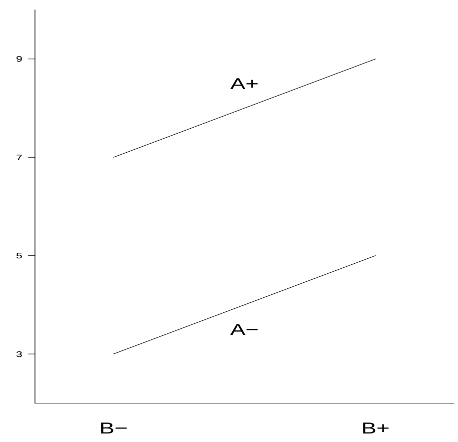
$$\mu_{22} - \mu_{12} = 9 - 5 = 4$$

und

$$\mu_{21} - \mu_{11} = 7 - 3 = 4\,.$$

Also liegt ein additives Modell vor. Abbildung 2 veranschaulicht dies.

Abbildung 2: Erwartungswerte der Faktorstufenkombinationen bei einem additiven Modell




### 3.1.2 Schätzung der Parameter und Effekte

Im Modell in Gleichung (31) auf Seite 22 bestimmen wir Schätzer der Parameter $\mu$, $\alpha_1$, $\alpha_2$, $\beta_1$ und $\beta_2$ so, dass

$$\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n}(y_{ijk} - \mu - \alpha_i - \beta_j)^2 \tag{41}$$

unter den Nebenbedingungen

$$\begin{aligned} \alpha_1 + \alpha_2 &= 0 \\ \beta_1 + \beta_2 &= 0 \end{aligned}$$

minimal wird.
Der Schätzer von $\mu$ ist

$$\boxed{\hat{\mu} = \bar{y} = \frac{1}{4n}\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n} y_{ijk}} \tag{42}$$

**Beispiel 4 (fortgesetzt)**
Es gilt $\bar{y} = 43$.

Wir schätzen $\alpha_i$ durch

$$\boxed{\hat{\alpha}_i = \bar{y}_{i\cdot} - \bar{y}} \tag{43}$$

Dabei ist $\bar{y}_{i\cdot}$ für $i = 1, 2$ der Mittelwert aller Beobachtungen, bei denen der Faktor $A$ auf der $i$-ten Stufe steht. Es gilt also

$$\boxed{\bar{y}_{i\cdot} = \frac{1}{2n} \sum_{j=1}^{2} \sum_{k=1}^{n} y_{ijk}} \tag{44}$$

**Beispiel 4 (fortgesetzt)**
Es gilt

$$\bar{y}_{1\cdot} = \frac{1}{10}(38 + 44 + 40 + 41 + 37 + 44 + 37 + 41 + 43 + 40) = 40.5\,.$$

und

$$\bar{y}_{2\cdot} = \frac{1}{10}(43 + 42 + 46 + 49 + 40 + 44 + 48 + 47 + 45 + 51) = 45.5$$

Also gilt

$$\hat{\alpha}_1 = \bar{y}_{1\cdot} - \bar{y} = 40.5 - 43 = -2.5$$

und

$$\hat{\alpha}_2 = \bar{y}_{2\cdot} - \bar{y} = 45.5 - 43 = 2.5$$

Wir schätzen $\beta_j$ durch

$$\boxed{\hat{\beta}_j = \bar{y}_{\cdot j} - \bar{y}} \tag{45}$$

Dabei ist $\bar{y}_{\cdot j}$ für $j = 1, 2$ der Mittelwert aller Beobachtungen, bei denen der Faktor $B$ auf der $j$-ten Stufe steht. Es gilt also

$$\boxed{\bar{y}_{\cdot j} = \frac{1}{2n} \sum_{i=1}^{2} \sum_{k=1}^{n} y_{ijk}\,.} \tag{46}$$

**Beispiel 4 (fortgesetzt)**
Es gilt

$$\bar{y}_{\cdot 1} = \frac{1}{10}(38 + 44 + 40 + 41 + 37 + 43 + 42 + 46 + 49 + 40) = 42\,.$$

und

$$\bar{y}_{\cdot 2} = \frac{1}{10}(44 + 37 + 41 + 43 + 40 + 44 + 48 + 47 + 45 + 51) = 44$$

Also gilt

$$\hat{\beta}_1 = \bar{y}_{\cdot 1} - \bar{y} = 42 - 43 = -1$$

und

$$\hat{\beta}_2 = \bar{y}_{\cdot 2} - \bar{y} = 44 - 43 = 1$$

Schauen wir uns nun die Schätzer von $E_A$ und $E_B$ an. Wir schätzen $E_A$, indem wir $\alpha_1$ durch $\hat{\alpha}_1$ und $\alpha_2$ durch $\hat{\alpha}_2$ aus Gleichung (43) auf Seite 25 schätzen und diese Schätzer in Gleichung (36) auf Seite 23 einsetzen

$$\boxed{e_A = \hat{\alpha}_2 - \hat{\alpha}_1 = \bar{y}_{2\cdot} - \bar{y} - (\bar{y}_{1\cdot} - \bar{y}) = \bar{y}_{2\cdot} - \bar{y}_{1\cdot}\,.}$$

Der geschätzte Effekt von $A$ gibt also an, wie sich die Zielgröße $Y$ im Mittel ändert, wenn man von der ersten Faktorstufe zur zweiten übergeht.

**Beispiel 4 (fortgesetzt)**
Der geschätzte Effekt von $A$ ist also

$$e_A = \bar{y}_{2\cdot} - \bar{y}_{1\cdot} = 45.5 - 40.5 = 5\,.$$

Auf der zweiten Strecke benötigt er im Mittel 5 Minuten länger als auf der ersten Strecke.

Wir schätzen $E_B$, indem wir $\beta_1$ durch $\hat{\beta}_1$ und $\beta_2$ durch $\hat{\beta}_2$ aus Gleichung (45) auf Seite 25 schätzen und diese Schätzer in Gleichung (40) auf Seite 23 einsetzen

$$\boxed{e_B = \hat{\beta}_2 - \hat{\beta}_1 = \bar{y}_{\cdot 2} - \bar{y} - (\bar{y}_{\cdot 1} - \bar{y}) = \bar{y}_{\cdot 2} - \bar{y}_{\cdot 1}\,.}$$

Der geschätzte Effekt von $B$ gibt also an, wie sich die Zielgröße $Y$ im Mittel ändert, wenn man von der ersten Faktorstufe zur zweiten übergeht.

**Beispiel 4 (fortgesetzt)**
Der geschätzte Effekt von $B$ ist also

$$e_B = \bar{y}_{\cdot 2} - \bar{y}_{\cdot 1} = 44 - 42 = 2\,.$$

Im Mittel benötigt er zwei Minuten länger, wenn er später losfährt.

### 3.1.3 Tests

Wir wollen testen, ob die Effekte signifikant sind. Der Test auf Signifikanz des Effekts von $A$ überprüft die Hypothese

$$H_0: \quad E_A = 0$$

und der Test auf Signifikanz des Effekts von $B$ die Hypothese

$$H_0: \quad E_B = 0\,.$$

Die Teststatistiken sind wie bei der einfaktoriellen Varianzanalyse Quotienten aus der durch den jeweiligen Faktor erklärten Streuung zur mittleren nicht durch das Modell erklärten Streuung.
Die Streuung, die nicht durch das Modell in Gleichung (25) auf Seite 21 erklärt wird, erhalten wir, indem wir in die Gleichung (41) auf Seite 24 für $\mu$, $\alpha_1$, $\alpha_2$, $\beta_1$ und $\beta_2$ die Schätzer aus den Gleichungen (42), (43) und (45) auf Seite 24 einsetzen. Wir erhalten

$$\boxed{SS_R = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y})^2} \tag{47}$$

Der Beweis ist auf Seite 107 zu finden.

**Beispiel 4 (fortgesetzt)**
Es gilt

$$\begin{aligned}
SS_R &= (38-40.5-42+43)^2 + (44-40.5-42+43)^2 \\
&+ (40-40.5-42+43)^2 + (41-40.5-42+43)^2 \\
&+ (37-40.5-42+43)^2 + (44-40.5-44+43)^2 \\
&+ (37-40.5-44+43)^2 + (41-40.5-44+43)^2 \\
&+ (43-40.5-44+43)^2 + (40-40.5-44+43)^2 \\
&+ (43-45.5-42+43)^2 + (42-45.5-42+43)^2 \\
&+ (46-45.5-42+43)^2 + (49-45.5-42+43)^2 \\
&+ (40-45.5-42+43)^2 + (44-45.5-44+43)^2 \\
&+ (48-45.5-44+43)^2 + (47-45.5-44+43)^2 \\
&+ (45-45.5-44+43)^2 + (51-45.5-44+43)^2 \\
&= 145
\end{aligned}$$

Die Anzahl der Freiheitsgrade von $SS_R$ ist $4n-3$.
Um die Hypothese

$$H_0: \quad E_A = 0$$

zu überprüfen, bestimmen wir

$$\boxed{SS_A = \sum_{i=1}^{2} 2n(\bar{y}_{i\cdot} - \bar{y})^2} \tag{48}$$

Die Anzahl der Freiheitsgrade von $SS_A$ ist 1.

**Beispiel 4 (fortgesetzt)**
Es gilt

$$SS_A \quad = \quad 2 \cdot 5 \cdot (40.5 - 43)^2 + 2 \cdot 5 \cdot (45.5 - 43)^2 = 125$$

Die Teststatistik ist

$$\boxed{F_A = \frac{SS_A}{SS_R/(4n-3)}}$$

**Beispiel 4 (fortgesetzt)**
Es gilt

$$F_A = \frac{125}{145/17} = 14.66$$

Wir lehnen $H_0$ ab, wenn gilt $F_A \geq F_{1,4n-3;1-\alpha}$. Dabei ist $F_{1,4n-3;1-\alpha}$ das $1-\alpha$-Quantil der $F$-Verteilung mit 1 und $4n-3$ Freiheitsgraden.

**Beispiel 4 (fortgesetzt)**
Es gilt $F_{1,17;0.95} = 4.45$. Wir lehnen $H_0$ also ab. Die Strecke besitzt einen signifikanten Effekt. Da wir an einer kurzen Fahrzeit interessiert sind, wählen wir die erste Strecke.

Um die Hypothese

$$H_0: \quad E_B = 0$$

zu überprüfen, bestimmen wir

$$\boxed{SS_B = \sum_{j=1}^{2} 2n(\bar{y}_{\cdot j} - \bar{y})^2} \tag{49}$$

Die Anzahl der Freiheitsgrade von $SS_B$ ist 1.

**Beispiel 4 (fortgesetzt)**
Es gilt

$$SS_B = 2 \cdot 5(42-43)^2 + 2 \cdot 5(44-43)^2 = 20$$

Die Teststatistik ist

$$\boxed{F_B = \frac{SS_B}{SS_R/(4n-3)}}$$

**Beispiel 4 (fortgesetzt)**
Es gilt

$$F_B = \frac{20}{145/17} = 2.34$$

Wir lehnen $H_0$ ab, wenn gilt $F_B \geq F_{1,4n-3;1-\alpha}$. Dabei ist $F_{1,4n-3;1-\alpha}$ das $1-\alpha$-Quantil der $F$-Verteilung mit 1 und $4n-3$ Freiheitsgraden.

**Beispiel 4 (fortgesetzt)**
Es gilt $F_{1,17;0.95} = 4.45$. Wir lehnen als $H_0$ nicht ab. die Abfahrzeit besitzt keinen signifikanten Effekt. Es ist also egal, wann wir losfahren.

### 3.1.4 Die ANOVA-Tabelle

Wir können die Informationen in einer ANOVA-Tabelle zusammenstellen. Tabelle 12 zeigt den allgemeinen Aufbau.

Tabelle 12: Allgemeiner Aufbau einer ANOVA-Tabelle im additiven Modell der zweifaktoriellen Varianzanalyse

| Quelle der Variation | Quadrat-summen | Freiheits-grade | Mittlere Quadratsummen | $F$ |
|---|---|---|---|---|
| $A$ | $SS_A$ | 1 | $MSS_A$ | $MSS_A/MSS_R$ |
| $B$ | $SS_B$ | 1 | $MSS_B$ | $MSS_B/MSS_R$ |
| Rest | $SS_R$ | $4n-3$ | $MSS_R$ | |
| Gesamt | $SS_T$ | $4n-1$ | | |

**Beispiel 4 (fortgesetzt)**
Wir erhalten die ANOVA-Tabelle in Tabelle 13.

Tabelle 13: ANOVA-Tabelle der Daten

| Quelle der Variation | Quadrat-summen | Freiheits-grade | Mittlere Quadratsummen | $F$ |
|---|---|---|---|---|
| $A$ | 125 | 1 | 125.00 | 14.66 |
| $B$ | 20 | 1 | 20.00 | 2.34 |
| Rest | 145 | 17 | 8.53 | |
| Gesamt | 290 | 19 | | |

Es gilt

$$\boxed{SS_T = SS_A + SS_B + SS_R} \tag{50}$$

Wir haben gesehen, dass es sehr mühselig ist, $SS_R$ zu berechnen. Die Bestimmung von $SS_T$ ist viel einfacher. Es gilt

$$\boxed{SS_T = \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{n} (y_{ijk} - \bar{y})^2} \tag{51}$$

Kennen wir also $SS_A$, $SS_B$ und $SS_T$, so können wir $SS_R$ mit Gleichung (50) bestimmen durch

$$SS_R = SS_T - SS_A - SS_B$$

**Beispiel 4 (fortgesetzt)**
Es gilt

$$\begin{aligned}
SS_T &= (38-43)^2 + (44-43)^2 + (40-43)^2 + (41-43)^2 + (37-43)^2 \\
&+ (44-43)^2 + (37-43)^2 + (41-43)^2 + (43-43)^2 + (40-43)^2 \\
&+ (43-43)^2 + (42-43)^2 + (46-43)^2 + (49-43)^2 + (40-43)^2 \\
&+ (44-43)^2 + (48-43)^2 + (47-43)^2 + (45-43)^2 + (51-43)^2 \\
&= 290
\end{aligned}$$

Also gilt

$$SS_R = 290 - 125 - 20 = 145$$

## 3.2 Das nichtadditive Modell

Im additiven Modell wird unterstellt, dass der Effekt des Faktors $A$ auf beiden Stufen des Faktors $B$ gleich ist. Welche Konsequenzen hat es, wenn dies nicht der Fall ist?

**Beispiel 6**
Wir betrachten einen faktoriellen Versuchsplan mit den Faktoren $A$und $B$. Es gelte

$$\mu_{11} = 3 \qquad \mu_{21} = 7 \qquad \mu_{12} = 9 \qquad \mu_{22} = 5$$

Somit ist der Effekt von $A$, wenn $B$ auf `-` steht, gleich

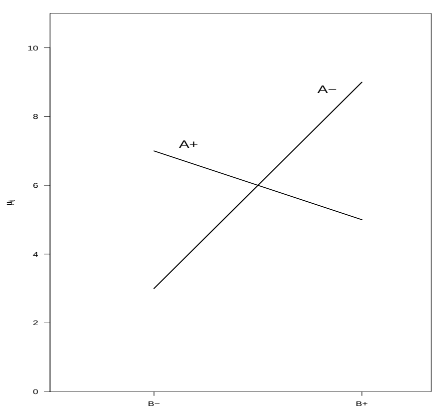
$$\mu_{21} - \mu_{11} = 7 - 3 = 4\,.$$

Der Effekt von $A$, wenn $B$ auf `B` steht, ist hingegen gleich

$$\mu_{22} - \mu_{12} = 5 - 9 = -4$$

Also liegt kein additives Modell vor. Abbildung 3 veranschaulicht dies.

Abbildung 3: Erwartungswerte der Faktorstufenkombinationen bei einem nichtadditiven Modell




Wir bestimmen den Effekt von $A$ mit Gleichung (35) auf Seite 22:

$$E_A = \frac{\mu_{21} - \mu_{11} + \mu_{22} - \mu_{12}}{2} = \frac{7 - 3 + 5 - 9}{2} = 0$$

Der Effekt von $A$ ist gleich 0. Somit hat $A$ keine Wirkung. Durch die Aggregation erhalten wir eine Aussage, die den wahren Tatbestand verdeckt. Der

Faktor $A$ besitzt durchaus einen Effekt. Dieser hängt aber von der Faktorstufe ab, auf der der Faktor $B$ steht.

Wir sagen auch, dass zwischen den Faktoren $A$ und $B$ **Interaktion** vorliegt, wenn der Effekt eines Faktors von den Faktorstufen des anderen Faktors abhängt. In diesem Fall sollte man nicht das additive Modell verwenden.

### 3.2.1 Das Modell

Beim additiven Modell bezieht sich der Term $\alpha_i$ auf den Faktor $A$ und der Term $\beta_j$ auf den Faktor $B$. Liegt Interaktion vor, so müssen wir diese im Modell berücksichtigen. Wir betrachten für $i = 1, 2$ und $j = 1, 2$ die Terme $(\alpha\beta_{ij})$ und erhalten folgendes Modell:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \tag{52}$$

Um eindeutige Schätzer der Parameter zu erhalten, benötigen wir folgende Nebenbedingungen:

$$\begin{aligned}
\alpha_1 + \alpha_2 &= 0 && (53)\\
\beta_1 + \beta_2 &= 0 && (54)\\
(\alpha\beta)_{11} + (\alpha\beta)_{12} &= 0 && (55)\\
(\alpha\beta)_{12} + (\alpha\beta)_{22} &= 0 && (56)\\
(\alpha\beta)_{11} + (\alpha\beta)_{21} &= 0 && (57)\\
(\alpha\beta)_{21} + (\alpha\beta)_{22} &= 0 && (58)
\end{aligned}$$

Wir unterstellen $E(\epsilon_{ijk}) = 0$ und $Var(\epsilon_{ijk}) = \sigma^2$.
Hieraus folgt

$$\mu_{ij} = E\left(Y_{ijk}\right) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} \tag{59}$$

und

$$Var\left(Y_{ijk}\right) = \sigma^2 \tag{60}$$

Wir wollen nun im Modell (59) den Effekt $E_A$ von $A$ beschreiben. Hierzu bestimmen wir wie im additiven Modell zuerst den Effekt $E_A$ von $A$, wenn $B$ auf - steht. Es gilt

$$\begin{aligned}
\mu_{21} - \mu_{11} &= \mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21} - \mu - \alpha_1 - \beta_1 - (\alpha\beta)_{11}\\
&= \alpha_2 - \alpha_1 + (\alpha\beta)_{21} - (\alpha\beta)_{11}
\end{aligned}$$

Nun setzen wir den Wert des Faktors $B$ auf **+** und bestimmen den Effekt von $A$:

$$\begin{aligned}
\mu_{22} - \mu_{12} &= \mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22} - \mu - \alpha_1 - \beta_2 - (\alpha\beta)_{12} \\
&= \alpha_2 - \alpha_1 + (\alpha\beta)_{22} - (\alpha\beta)_{12}
\end{aligned}$$

Wir sehen, dass der Effekt von $A$ auf den beiden Stufen von $B$ unterschiedlich ist.

Wie im additiven Modell bestimmen wir den Effekt von $A$ als Mittelwert des Effekts von $A$, wenn $B$ auf **-** steht, und des Effekts von $A$, wenn $B$ auf **+** steht:

$$\boxed{E_A = \frac{\mu_{21} - \mu_{11} + \mu_{22} - \mu_{12}}{2}}$$

Wir erhalten

$$\begin{aligned}
E_A &= \frac{\mu_{21} - \mu_{11} + \mu_{22} - \mu_{12}}{2} \\
&= \frac{\alpha_2 - \alpha_1 + (\alpha\beta)_{21} - (\alpha\beta)_{11} + \alpha_2 - \alpha_1 + (\alpha\beta)_{22} - (\alpha\beta)_{12}}{2} \\
&= \alpha_2 - \alpha_1 + \frac{(\alpha\beta)_{21} + (\alpha\beta)_{22}}{2} - \frac{(\alpha\beta)_{11} + (\alpha\beta)_{12}}{2} \\
&\overset{(58)(55)}{=} \alpha_2 - \alpha_1
\end{aligned}$$

Analog definieren wir $E_B$ wie im additiven Modell

$$\boxed{E_B = \frac{\mu_{12} - \mu_{11} + \mu_{22} - \mu_{21}}{2}}$$

und erhalten

$$E_B = \beta_2 - \beta_1$$

Neben $E_A$ und $E_B$ müssen wir im nichtadditiven Modell noch einen weiteren Effekt berücksichtigen. Dies ist der Interaktionseffekt $E_{AB}$, der den Zusammenhang zwischen den Faktoren $A$und $B$ beschreibt. Dieser gibt an, wie sich der Effekt von $A$, wenn $B$ auf **+** steht, vom Effekt von $A$, wenn $B$ auf **-** steht, unterscheidet. Wir bilden also die Differenz aus dem Effekt von $A$, wenn $B$ auf **+** steht, und dem Effekt von $A$, wenn $B$ auf **-**:

$$\boxed{E_{AB} = \frac{\mu_{22} - \mu_{12} - (\mu_{21} - \mu_{11})}{2} = \frac{\mu_{11} - \mu_{21} - \mu_{12} + \mu_{22}}{2}} \tag{61}$$

Ist der Effekt von $A$ auf beiden Stufen von $B$ gleich, so ist $E_{AB}$ gleich 0.
Wir können $E_{AB}$ auch umformen zu

$$E_{AB} = \frac{\mu_{22} - \mu_{21} - (\mu_{12} - \mu_{11})}{2}$$

In dieser Darstellung gibt $E_{AB}$ an, wie sich der Effekt von $B$, wenn $A$ auf **+** steht, vom Effekt von $B$, wenn $A$ auf **-** steht, unterscheidet.

### 3.2.2 Schätzung der Parameter und Effekte

Im Modell in Gleichung (59) auf Seite 32 bestimmen wir Schätzer der Parameter $\mu$, $\alpha_i$, $\beta_j$ und $(\alpha\beta)_{ij}$ so, dass

$$\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n}(y_{ijk} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij})^2 \tag{62}$$

unter den Nebenbedingungen in den Gleichungen (53), (54), (55), (56), (57) und (58) minimiert wird.
Der Schätzer von $\mu$, $\alpha_i$ und $\beta_j$ sind mit denen im additiven Modell identisch. Es gilt also:

$$\hat{\mu} = \bar{y} \tag{63}$$

$$\hat{\alpha}_i = \bar{y}_{i\cdot} - \bar{y} \tag{64}$$

$$\hat{\beta}_j = \bar{y}_{\cdot j} - \bar{y} \tag{65}$$

Also sind auch die Schätzer $e_A$ und $e_B$ im additiven und nichtadditiven Modell identisch. Es gilt also

$$\boxed{e_A = \bar{y}_{2\cdot} - \bar{y}_{1\cdot}} \tag{66}$$

und

$$\boxed{e_B = \bar{y}_{\cdot 2} - \bar{y}_{\cdot 1}\,.} \tag{67}$$

**Beispiel 4 (fortgesetzt)**
Es gilt also $e_A = 5$ und $e_B = 1.5$.

Der Schätzer $\widehat{(\alpha\beta)}_{ij}$ von $(\alpha\beta)_{ij}$ ist

$$\widehat{(\alpha\beta)}_{ij} = \bar{y}_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y} \tag{68}$$

mit

$$\boxed{\bar{y}_{ij} = \frac{1}{n}\sum_{k=1}^{n} y_{ijk}} \tag{69}$$

Wir schätzen den Interaktionseffekt $E_{AB}$ zwischen den Faktoren $A$ und $B$, indem wir $\mu_{ij}$ durch $\bar{y}_{ij}$ schätzen und diese Schätzer in Gleichung (61) auf Seite 34 einsetzen. Wir erhalten dann als Schätzer für $E_{AB}$:

$$\boxed{e_{AB} = \frac{\bar{y}_{11} - \bar{y}_{21} - \bar{y}_{12} + \bar{y}_{22}}{2}} \tag{70}$$

**Beispiel 4 (fortgesetzt)**
Es gilt

$$\bar{y}_{11} = \frac{38 + 44 + 40 + 41 + 37}{5} = 40$$

$$\bar{y}_{12} = \frac{44 + 37 + 41 + 43 + 40}{5} = 41$$

$$\bar{y}_{21} = \frac{43 + 42 + 46 + 49 + 40}{5} = 44$$

$$\bar{y}_{22} = \frac{44 + 48 + 47 + 45 + 51}{5} = 47$$

Also ist der geschätzte Interaktionseffekt zwischen den Faktoren $A$ und $B$ gegeben durch

$$e_{AB} = \frac{\bar{y}_{11} - \bar{y}_{21} - \bar{y}_{12} + \bar{y}_{22}}{2} = \frac{40 - 44 - 41 + 47}{2} = 1$$

Der geschätzte Effekt von $A$, wenn $B$ auf - steht, ist

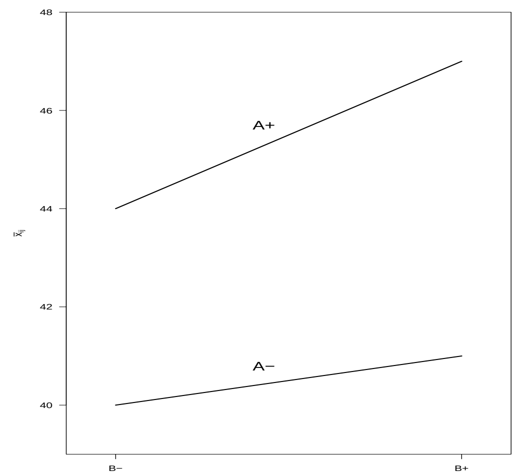
$$\bar{y}_{21} - \bar{y}_{11} = 44 - 40 = 4\,.$$

Der geschätzte Effekt von $A$, wenn $B$ auf + steht, ist

$$\bar{y}_{22} - \bar{y}_{12} = 47 - 41 = 6\,.$$

Wir können ein **Interaktionsdiagramm** erstellen. Dieses ist in Abbildung 4 zu finden.

Abbildung 4: Interaktionsdiagramm




Wir sehen, dass der geschätzte Effekt von $A$, wenn $B$ auf **+** steht, größer ist als der geschätzte Effekt von $A$, wenn $B$ auf **-** steht.

Die beiden geschätzten Effekte unterscheiden sich. Es stellt sich die Frage, ob dieser Unterschied signifikant ist.

#### 3.2.3 Tests

Nun können wir uns der Frage zuwenden, ob die Effekte signifikant sind. Wir benötigen $SS_A$, $SS_B$, $SS_{AB}$ und $SS_R$. $SS_A$ und $S_B$ sind im additiven und nichtadditiven Modell identisch. Es gilt also

$$\boxed{SS_A = \sum_{i=1}^{2} 2n(\bar{y}_{i\cdot} - \bar{y})^2}$$

und

$$\boxed{SS_B = \sum_{j=1}^{2} 2n(\bar{y}_{\cdot j} - \bar{y})^2 \,.}$$

**Beispiel 4 (fortgesetzt)**
Es gilt $SS_A = 125$ und $SS_B = 20$.

Die Streuung, die nicht durch das Modell in Gleichung (59) auf Seite 32 erklärt wird, erhalten wir, indem wir in die Gleichung (62) auf Seite 34 für $\mu$, $\alpha_i$, $\beta_j$ und $(\alpha\beta)_{ij}$ die Schätzer aus den Gleichungen (63), (64), (65) und (68)auf Seite 34 einsetzen. Es gilt

$$\boxed{SS_R = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij})^2\,.} \tag{71}$$

Der Beweis ist auf Seite 108 zu finden.

**Beispiel 4 (fortgesetzt)**
Es gilt

$$\begin{aligned}
SS_R &= (38-40)^2 + (44-40)^2 + (40-40)^2 + (41-40)^2 + (37-40)^2 \\
&+ (44-41)^2 + (37-41)^2 + (41-41)^2 + (43-41)^2 + (40-41)^2 \\
&+ (43-44)^2 + (42-44)^2 + (46-44)^2 + (49-44)^2 + (40-44)^2 \\
&+ (44-47)^2 + (48-47)^2 + (47-47)^2 + (45-47)^2 + (51-47)^2 \\
&= 140
\end{aligned}$$

Auf jeder der 4 Faktorstufenkombinationen gibt es $n$ Beobachtungen, von denen $n-1$ frei wählbar ist, wenn der Mittelwert der Faktorstufenkombination bekannt ist. Von den $4n$ Sumanden in $SS_R$ können wir also $4n-4$ frei wählen.
Zum Test auf Interaktion benötigen wir noch $SS_{AB}$. Es gilt

$$\boxed{SS_{AB} = \sum_{i=1}^{2}\sum_{j=1}^{2} n(\bar{y}_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y})^2} \tag{72}$$

Die Anzahl der Freiheitsgrade von $SS_{AB}$ ist gleich 1.

**Beispiel 4 (fortgesetzt)**
Es gilt

$$\begin{aligned}
SS_{AB} &= 5(40-40.5-42+43)^2 + 5(41-40.5-44+43)^2 \\
&+ 5(44-45.5-42+43)^2 + 5(47-45.5-44+43)^2 \\
&= 5
\end{aligned}$$

Wir überprüfen zunächst, ob Interaktion vorliegt. Wir testen also

$$H_0: \quad E_{AB} = 0$$

Die Teststatistik ist

$$\boxed{F_{AB} = \frac{SS_{AB}}{SS_R/(4n-4)}} \tag{73}$$

**Beispiel 4 (fortgesetzt)**
Es gilt

$$F_{AB} = \frac{5}{140/16} = 0.5714$$

Wir lehnen $H_0$ ab, wenn gilt $F_{AB} \geq F_{1,4n-4;1-\alpha}$. Dabei ist $F_{1,4n-4;1-\alpha}$ das $1-\alpha$-Quantil der $F$-Verteilung mit 1 und $4n-4$ Freiheitsgraden.

**Beispiel 4 (fortgesetzt)**
Es gilt $F_{1,16;0,95} = 4.49$. Wir lehnen $H_0$ nicht ab. Es liegt also keine Interaktion vor.

Lehnen wir $H_0$ ab, so spricht dies für Interaktion. Es ist in diesem Fall nicht sinnvoll nach dem Effekt von $A$ oder dem Effekt von $B$ zu fragen, da diese von den Faktorstufen des jeweils anderen Faktors abhängen.
Lehnen wir $H_0$ nicht ab, so überprüfen wir, ob der Effekt $A$ oder der Effekt von $B$ signifikant von 0 verschieden ist.
Beginnen wir mit

$$H_0: \quad E_A = 0\,.$$

Die Teststatistik ist

$$\boxed{F_A = \frac{SS_A}{SS_R/(4n-4)}} \tag{74}$$

Wir lehnen $H_0$ ab, wenn gilt $F_A \geq F_{1,4n-4;1-\alpha}$. Dabei ist $F_{1,4n-4;1-\alpha}$ das $1-\alpha$-Quantil der $F$-Verteilung mit 1 und $4n-4$ Freiheitsgraden.

**Beispiel 4 (fortgesetzt)**
Es gilt

$$F_A = \frac{125}{140/16} = 14.29$$

Wegen $F_{1,16;0,95} = 4.49$ lehnen wir $H_0$ ab. Die Strecke besitzt also einen signifikanten Einfluss auf die Fahrzeit.

Zum Überprüfen der Hypothese

$$H_0: \quad E_B = 0 \,.$$

bestimmen wir die Teststatistik

$$\boxed{F_B = \frac{SS_B}{SS_R/(4n-4)}} \tag{75}$$

Wir lehnen $H_0$ ab, wenn gilt $F_B \geq F_{1,4n-4;1-\alpha}$. Dabei ist $F_{1,4n-4;1-\alpha}$ das $1-\alpha$-Quantil der $F$-Verteilung mit 1 und $4n-4$ Freiheitsgraden.

**Beispiel 4 (fortgesetzt)**
Es gilt

$$F_B = \frac{20}{140/16} = 2.29$$

Wegen $F_{1,16;0,95} = 4.49$ lehnen wir $H_0$ nicht ab. Die Abfahrzeit besitzt also keinen signifikanten Einfluss auf die Fahrzeit.
Es liegt also keine Interaktion vor. Zwischen den Strecken gibt es einen signifikanten Unterschied, während es bei den Fahrzeiten keinen signifikanten Unterschied gibt.

### 3.2.4 Die ANOVA-Tabelle

Wir können die Informationen in einer ANOVA-Tabelle zusammenstellen. Tabelle 14 zeigt den allgemeinen Aufbau.

Tabelle 14: Allgemeiner Aufbau einer ANOVA-Tabelle der zweifaktoriellen Varianzanalyse im nichtadditiven Modell

| Quelle der Variation | Quadratsummen | Freiheitsgrade | Mittlere Quadratsummen | $F$ |
|---|---|---|---|---|
| $A$ | $SS_A$ | 1 | $MSS_A$ | $MSS_A/MSS_R$ |
| $B$ | $SS_B$ | 1 | $MSS_B$ | $MSS_B/MSS_R$ |
| $AB$ | $SS_{AB}$ | 1 | $MSS_{AB}$ | $MSS_{AB}/MSS_R$ |
| Rest | $SS_R$ | $4n-4$ | $MSS_R$ | |
| Gesamt | $SS_T$ | $4n-1$ | | |

**Beispiel 4 (fortgesetzt)**
Wir erhalten die ANOVA-Tabelle in Tabelle 15.

Tabelle 15: ANOVA-Tabelle der Fahrzeit in Abhängigkeit von der Strecke und der Abfahrzeit

| Quelle der Variation | Quadrat-summen | Freiheits-grade | Mittlere Qua-dratsummen | $F$ |
|---|---|---|---|---|
| $A$ | 125 | 1 | 125 | 14.29 |
| $B$ | 20 | 1 | 20 | 2.29 |
| $AB$ | 5 | 1 | 5 | 0.57 |
| Rest | 140 | 16 | 8.75 | |
| Gesamt | 290 | 19 | | |

### 3.2.5 Der Algorithmus von Yates

Wir hatten bei einfaktoriellen Versuchsplänen die Notation (1) und $a$ eingeführt. Wir können diese jetzt auf zweifaktorielle Versuchspläne erweitern. Taucht in einer Symbolfolge der Kleinbuchstabe eines Faktors auf, so steht dieser Faktor auf +. Ansonsten steht er auf −. Stehen alle Faktoren auf −, so verwenden wir das Symbol (1). Die Symbolfolgen stehen dann wieder für die Summe der Werte auf der jeweiligen Faktorstufenkombination. Es gilt also

$$(1) = \sum_{k=1}^{n} y_{11k} \tag{76}$$

$$a = \sum_{k=1}^{n} y_{21k} \tag{77}$$

$$b = \sum_{k=1}^{n} y_{12k} \tag{78}$$

$$ab = \sum_{k=1}^{n} y_{22k} \tag{79}$$

**Beispiel 4 (fortgesetzt)**
Es gilt

$$(1) = 200 \qquad a = 220 \qquad b = 205 \qquad ab = 235$$

Wir können auch die geschätzten Effekte in dieser Notation darstellen. Schauen wir uns zuerst die Mittelwerte auf den einzelnen Faktorstufen an. Es gilt

$$\bar{y}_{i\cdot} = \frac{1}{2n}\left(\sum_{k=1}^{n} y_{i1k} + \sum_{k=1}^{n} y_{i2k}\right)$$

und

$$\bar{y}_{\cdot j} = \frac{1}{2n}\left(\sum_{k=1}^{n} y_{1jk} + \sum_{k=1}^{n} y_{2jk}\right).$$

Also gilt

$$\bar{y}_{1\cdot} = \frac{(1)+b}{2n} \qquad \bar{y}_{2\cdot} = \frac{a+ab}{2n} \tag{80}$$

$$\bar{y}_{\cdot 1} = \frac{(1)+a}{2n} \qquad \bar{y}_{\cdot 2} = \frac{b+ab}{2n} \tag{81}$$

Der geschätze Effekt von $A$ ist

$$\boxed{e_A = \frac{a+ab-(1)-b}{2n}} \tag{82}$$

Dies sieht man folgendermaßen

$$e_A \overset{(66)}{=} \bar{y}_{2\cdot} - \bar{y}_{1\cdot} \overset{(80)}{=} \frac{a+ab-(1)-b}{2n}$$

Man nennt

$$\boxed{K_A = a+ab-(1)-b} \tag{83}$$

auch den Kontrast $K_A$ von $A$.

**Beispiel 4 (fortgesetzt)**
Es gilt $K_A = 50$. Also gilt

$$e_A = \frac{50}{10} = 5\,.$$

Der geschätze Effekt von $B$ ist

$$\boxed{e_B = \frac{b+ab-(1)-a}{2n}} \tag{84}$$

Dies sieht man folgendermaßen

$$e_B \stackrel{(67)}{=} \bar{y}_{\cdot 2} - \bar{y}_{\cdot 1} \stackrel{(81)}{=} \frac{b + ab - (1) - a}{2n}$$

Man nennt

$$\boxed{K_B = b + ab - (1) - a} \tag{85}$$

auch den Kontrast $K_B$ von $B$.

**Beispiel 4 (fortgesetzt)**
Es gilt $K_B = 20$. Also gilt

$$e_B = \frac{20}{10} = 2\,.$$

Der geschätze Effekt von $AB$ ist

$$\boxed{e_{AB} = \frac{(1) - a - b - ab}{2n}\,.} \tag{86}$$

Dies sieht man folgendermaßen

$$e_{AB} \stackrel{(70)}{=} \frac{\bar{y}_{11} - \bar{y}_{21} - \bar{y}_{12} + \bar{y}_{22}}{2} = \frac{(1) - a - b + ab}{2n}\,.$$

Hierbei haben wir benutzt, dass aus den Gleichungen (76), (77), (78) und (79) auf Seite 40 folgt

$$\bar{y}_{11} = \frac{(1)}{n} \qquad \bar{y}_{21} = \frac{a}{n} \qquad \bar{y}_{12} = \frac{b}{n} \qquad \bar{y}_{22} = \frac{ab}{n}$$

Man nennt

$$\boxed{K_{AB} = (1) - a - b + ab} \tag{87}$$

auch den Kontrast $K_{AB}$ von $AB$.

**Beispiel 4 (fortgesetzt)**
Es gilt $K_{AB} = 10$. Also gilt

$$e_{AB} = \frac{10}{10} = 1\,.$$

Wir können auch die Quadratsummen in dieser Notation darstellen. Es gilt

$$\boxed{SS_A = \frac{(a + ab - (1) - b)^2}{4n}} \tag{88}$$

Der Beweis dieser Beziehung ist auf Seite 109 zu finden.
Mit $K_A = -(1) + a - b + ab$ gilt also

$$\boxed{SS_A = \frac{K_A^2}{4n}}$$

**Beispiel 4 (fortgesetzt)**
Mit $K_A = 50$ gilt $SS_A = (50^2)/20 = 125$.

Mit $K_B = -(1) - a + b + ab$ gilt

$$\boxed{SS_B = \frac{K_B^2}{4n}}$$

**Beispiel 4 (fortgesetzt)**
Mit $K_B = 20$ gilt $SS_B = (20^2)/20 = 20$.

Mit $K_{AB} = (1) - a - b + ab$ gilt

$$\boxed{SS_{AB} = \frac{K_{AB}^2}{4n}}$$

**Beispiel 4 (fortgesetzt)**
Mit $K_{AB} = 10$ gilt $SS_{AB} = (10^2)/20 = 5$.

Wir zeigen nun, dass der Algorithmus von Yates bei einem faktoriellen Versuchsplan mit zwei Faktoren die Kontraste liefert. Hierzu stellen die Tabelle 16 auf.

Tabelle 16: Faktorstufenkombinationen bei einem zweifaktoriellen Versuchsplan

| |
|---|
| (1) |
| $a$ |
| $b$ |
| $ab$ |

**Beispiel 4 (fortgesetzt)**
Wir erhalten folgende Tabelle 17.

Tabelle 17: Faktorstufenkombinationen bei einem zweifaktoriellen Versuchsplan

| | |
|---|---|
| (1) | 200 |
| $a$ | 220 |
| $b$ | 205 |
| $ab$ | 235 |

Wir führen den Algorithmus von Yates durch. Wir beginnen mit der ersten Spalte.

1. Wir summieren die ersten beiden Zahlen dieser Spalte und schreiben sie in die erste Zeile der nächsten Spalte.

2. Wir summieren die nächsten beiden Zahlen und schreiben das Ergebnis in die zweite Zeile der nächsten Spalte.

3. Wir nehmen die nächsten beiden Zahlen, bilden die Differenz aus der unteren und der oberen und schreiben das Ergebnis in die zweite Zeile der nächsten Spalte.

4. Wir nehmen die nächsten beiden Zahlen, bilden die Differenz aus der unteren und der oberen und schreiben das Ergebnis in die zweite Zeile der nächsten Spalte.

Wir gehen mit der zweiten Spalte genau so vor wie mit der ersten. Nach diesem Schritt stehen in der letzten Spalte die Kontraste der Faktoren.

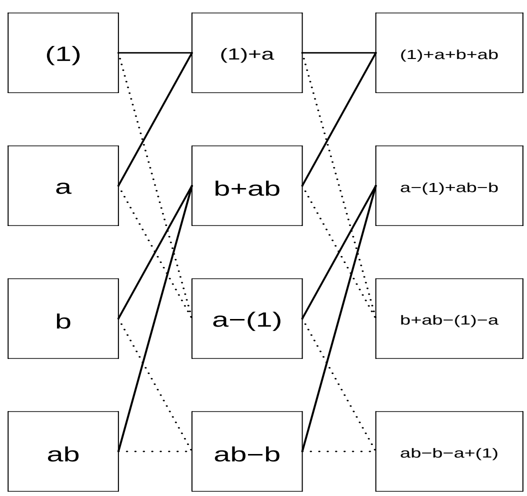
Wir erhalten also folgende Schritte

Tabelle 18: Faktorstufenkombinationen bei einem zweifaktoriellen Versuchsplan

| | | |
|---|---|---|
| $(1)$ | $(1) + a$ | $(1) + a + b + ab$ |
| $a$ | $b + ab$ | $a - (1) + ab - b$ |
| $b$ | $a - (1)$ | $b + ab - (1) - a$ |
| $ab$ | $ab - b$ | $ab - b - a + (1)$ |

Der Vergleich der Elemente in der zweiten, dritten und vierten Zeile der letzten Spalte von Tabelle 18 mit den Gleichungen (83), (85) und (87) zeigt, dass hier die Kontraste stehen. Die Abbildung 5 illustriert die Vorgehensweise beim Yates-Algorithmus. Dabei bedeutet eine durchgezogene Linie Addition und eine gestrichelte Linie Subtraktion.

Abbildung 5: Der Algorithmus von Yates



**Beispiel 4 (fortgesetzt)**
Tabelle 19 zeigt die einzelnen Schritte.

Tabelle 19: Algorithmus von Yates bei einem zweifaktoriellen Versuchsplan

| | | | |
|---|---|---|---|
| (1) | 200 | 420 | 860 |
| $a$ | 220 | 440 | 50 |
| $b$ | 205 | 20 | 20 |
| $ab$ | 235 | 30 | 10 |

## 3.3 Anmerkungen

Es stellt sich die Frage, ob man das additive oder das nichtadditive Modell als Ausgangspunkt der Analyse wählen sollte. Da Interaktion zwischen den Faktoren vorliegen kann, sollte man das nichtadditive Modell verwenden. In diesem kann man überprüfen, ob Additivität vorliegt.
Es gibt aber einen Fall, in dem man das additive Modell unterstellen sollte, wenn man die Signifikanz der Faktoren überprüfen will. Manchmal sind die Versuche so kostspielig, dass man auf jeder Faktorstufenkombination nur einen Versuch durchführne kann. Es gilt also $n = 1$. In diesem Fall ist $SS_R$ im nichtadditiven Modell gleich 0. Dies sieht man folgendermaßen.
Es gilt

$$SS_R = \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{n} (y_{ijk} - \bar{y}_{ij})^2 \,.$$

Liegt auf jeder Faktorstufenkombination nur eine Beobachtung vor, gilt $y_{ij1} = \bar{y}_{ij}$. Dies hat zur Konsequenz, dass $SS_R = 0$ gilt.
Ist $n = 1$, so kann man im nichtadditiven Modell also nicht testen. Man kann aber ein additives Modell unterstellen. In diesem Fall liefert der Algorithmus von Yates $SS_R$. Im additiven Modell gilt nämlich

$$SS_T = SS_A + SS_B + SS_R \tag{89}$$

Im nichtadditiven Modell gilt

$$SS_T = SS_A + SS_B + SS_{AB} + SS_R \tag{90}$$

In beiden Modellen sind $SS_T$, $SS_A$ und $SS_B$ identisch, falls die Daten und die Faktoren identisch sind. Ist nun $n = 1$, so gilt im nichtadditiven Modell

$SS_R = 0$. Somit ist $SS_R$ aus dem additiven Modell gleich $SS_{AB}$ aus dem nichtadditiven Modell. Wir können also für $n = 1$ im additiven Modell den Algorithmus von Yates anwenden und $SS_{AB}$ als $SS_R$ verwenden.

**Beispiel 5**
Der Arbeitnehmer fährt auf jeder Faktorstufenkombination genau einmal und erhält die Daten in Tabelle 20.

Tabelle 20: Ergebnis eines zweifaktoriellen Versuches

| $A$ | $B$ | Fahrzeit |
|---|---|---|
| − | − | 40 |
| + | − | 46 |
| − | + | 42 |
| + | + | 50 |

Wir unterstellen das additive Modell und wenden den Algorithmus von Yates an.

Tabelle 21: Algorithmus von Yates bei einem zweifaktoriellen Versuchsplan

| | | | |
|---|---|---|---|
| (1) | 40 | 86 | 178 |
| $a$ | 46 | 92 | 14 |
| $b$ | 42 | 6 | 6 |
| $ab$ | 50 | 8 | 2 |

Somit gilt

$$K_A = 14 \qquad K_B = 6 \qquad K_{AB} = 2$$

Also gilt

$$\begin{aligned} SS_A &= \frac{14^2}{4} = 49 \\ SS_B &= \frac{6^2}{4} = 9 \\ SS_{AB} &= \frac{2^2}{4} = 1 \end{aligned}$$

Also ist $SS_R = 1$ und wir erhalten die ANOVA-Tabelle in Tabelle 22.

Tabelle 22: ANOVA-Tabelle der Fahrzeit in Abhängigkeit von der Strecke und der Abfahrzeit

| Quelle der Variation | Quadrat-summen | Freiheits-grade | Mittlere Quadratsummen | $F$ |
|---|---|---|---|---|
| $A$ | 49 | 1 | 49 | 49 |
| $B$ | 9 | 1 | 9 | 9 |
| Rest | 1 | 1 | 1 | |
| Gesamt | 59 | 3 | | |

Wegen $F_{0.95;1,1} = 161.45$ ist keiner der beiden Faktoren signifikant.

Wenn wir bei zwei Faktoren ein additives Modell unterstellen, können wir nicht testen, ob Interaktion vorliegt. Wir werden im nächsten Kapitel sehen, welche Möglichkeiten bestehen, die Signifikanz aller Faktoren zu überprüfen, wenn $n = 1$ ist. Bei zwei Faktoren ist für diese Verfahren der Stichprobnumfang aber zu klein.

Bei faktoriellen Versuchsplänen wird unterstellt, dass die Störgrößen normalverteilt sind. Um diese Annahme zu überprüfen, schaut man sich die Residuen $e_{ijk}$ an. Im Modell (52) auf Seite 32 gilt

$$e_{ijk} = y_{ijk} - \bar{y}_{ij} \tag{91}$$

**Beispiel 5 (fortgesetzt)**
Es gilt

$$\bar{y}_{11} = 40 \qquad \bar{y}_{21} = 44 \qquad \bar{y}_{12} = 41 \qquad \bar{y}_{22} = 47$$

Also gilt

$$e_{111} = -2 \quad e_{112} = 4 \quad e_{113} = 0 \quad e_{114} = 1 \quad e_{115} = -3$$

$$e_{211} = -1 \quad e_{212} = -2 \quad e_{213} = 2 \quad e_{214} = 5 \quad e_{215} = -4$$

$$e_{121} = 3 \quad e_{122} = -4 \quad e_{123} = 0 \quad e_{124} = 2 \quad e_{125} = -1$$

$$e_{221} = -3 \quad e_{222} = 1 \quad e_{223} = 0 \quad e_{224} = -2 \quad e_{225} = 4$$

Es gibt eine Reihe von Möglichkeiten die Gültigkeit dieser Annahme zu überprüfen. Man kann einen Test auf Normalverteilung wie den *Kolmogorow-*

*Smirnow-Test* durchführen. Wir wollen eine graphische Darstellung betrachten, mit der man die Annahme der Normalverteilung überprüfen kann. Bei einem *Normal-Quantil-Plot* zeichnet man die geordneten Residuen $e_{(1)}, \ldots, e_{(N)}$ gegen die Quantile $\Phi^{-1}((1-0.5)/N), \ldots, \Phi^{-1}((N-0.5)/N)$ der Standardnormalverteilung. Liegt Normalverteilung vor, so sollten die Punkte um eine Gerade streuen. Hierbei bestimmt man aber nicht die KQ-Gerade, da diese nicht robust ist. Man legt die Gerade vielmehr durch das untere und obere Quartil der Punktepaare.

**Beispiel 5 (fortgesetzt)**
Die geordneten Residuen sind

```
-4 -4 -3 -3 -2 -2 -2 -1 -1  0  0  0  1  1  2  2  3  4  4  5
```

Speziell gilt $e_{(1)} = -4$, $e_{(2)} = -4$ und $e_{(20)} = 5$.
Speziell gilt

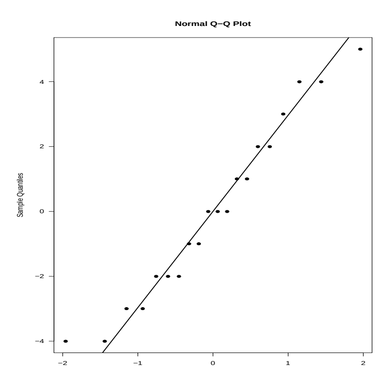
$$\Phi^{-1}((1-0.5)/20) = \Phi^{-1}(0.025) = -1.96$$

Die anderen Quantile ergeben sich analog. Alle lauten

```
-1.960 -1.440 -1.150 -0.935 -0.755
-0.598 -0.454 -0.319 -0.189 -0.063
 0.063  0.189  0.319  0.454  0.598
 0.755  0.935  1.150  1.440  1.960
```

Das untere Quartil der Residuen ist $-2$ und das obere 2. Das untere Quartil der Quantile der Normalverteilung ist $-0.6765$ und das obere $0.6765$. Wir legen die Gerade also durch die Punkte $(-2, -0.6765)$ und $(2, 0.6765)$. Abbildung 6 zeigt den Normal-Quantil-Plot mit der Geraden. Die Annahme der Normalverteilung ist gerechtfertigt.

Abbildung 6: Normal-Quantil-Plot



## 3.4 Ein Beispiel für die Auswertung eines zweifaktoriellen Versuchsplans

Wir wollen an einem Beispiel die Vorgehensweise bei der Auswertung eines zweifaktoriellen Versuchsplans illustrieren.

Die Studenten Michael Dorin und Stefan Neuhaus ließen in ihrem BI-Projekt 16 Studenten den PRESS-Test durchführen. Bei diesem müssen die Teilnehmer in 3 Minuten möglichst viele Aufgaben vom Typ $3 - 5 + 1$ lösen. Die Zielvariable ist die Anzahl der richtig gelösten Aufgaben. Es wurden zwei Faktoren $A$und $B$ betrachtet.

Der Einflussfaktor $A$ ist die Schriftgröße mit den Faktorstufen 8 pt (- und 16 pt +.

Der Faktor $B$ steht auf -, wenn während des Tests keine Musik lief. Ansonsten steht er auf +. Die Daten sind in Tabelle 23 zu finden.

Tabelle 23: Daten eines zweifaktoriellen Versuchsplans

| $A$ | $B$ | Zeit | |
|---|---|---|---|
| - | - | 42 | 39 |
| + | - | 51 | 53 |
| - | + | 48 | 43 |
| + | + | 87 | 73 |

Wir führen den Algorithmus von Yates durch.

Tabelle 24: Algorithmus von Yates bei einem zweifaktoriellen Versuchsplan

| | | | |
|---|---|---|---|
| (1) | 81 | 185 | 436 |
| $a$ | 104 | 251 | 92 |
| $b$ | 91 | 23 | 66 |
| $ab$ | 160 | 69 | 46 |

Somit gilt

$$K_A = 92 \qquad K_B = 66 \qquad K_{AB} = 46$$

Die geschätzten Effekte sind

$$e_A = \frac{K_A}{2n} = \frac{92}{4} = 23$$

$$e_B = \frac{K_B}{2n} = \frac{66}{4} = 16.5$$

$$e_{AB} = \frac{K_{AB}}{2n} = \frac{46}{4} = 11.5$$

Die Quadratsummen sind

$$SS_A = \frac{K_A^2}{4n} = \frac{92^2}{8} = 1058$$

$$SS_B = \frac{K_B^2}{4n} = \frac{66^2}{8} = 544.5$$

$$SS_{AB} = \frac{K_{AB}^2}{4n} = \frac{46^2}{8} = 264.5$$

Wir benötigen nun noch $SS_R$. Es gilt

$$\begin{aligned} SS_R &= (42-40.5)^2 + (39-40.5)^2 + (48-45.5)^2 + (43-45.5)^2 \\ &+ (51-52)^2 + (53-52)^2 + (87-80)^2 + (73-80)^2 \\ &= 117 \end{aligned}$$

Jetzt können wir die ANOVA-Tabelle aufstellen.

Tabelle 25: ANOVA-Tabelle

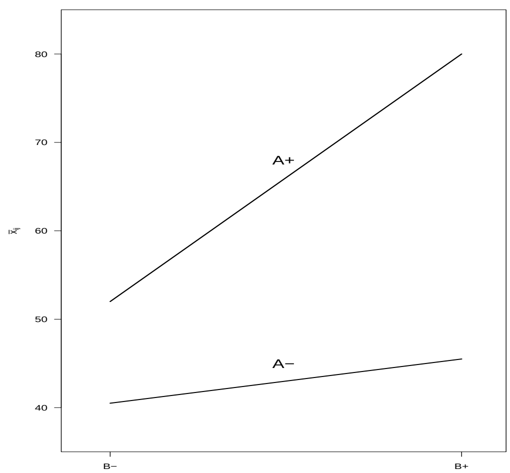
| Quelle der Variation | Quadrat-summen | Freiheits-grade | Mittlere Quadratsummen | $F$ |
|---|---|---|---|---|
| $A$ | 1058.0 | 1 | 1058.00 | 36.17 |
| $B$ | 544.5 | 1 | 544.50 | 18.62 |
| $AB$ | 264.5 | 1 | 264.50 | 9.04 |
| Rest | 117.0 | 4 | 19.25 | |
| Gesamt | 1984 | 7 | | |

Wir führen zuerst den Test auf Interaktion durch. Es gilt

$$F_{AB} = \frac{264.5}{117/4} = 9.04$$

Wegen $F_{1,4;0,95} = 7.71$ lehnen wir $H_0$ ab. Es liegt also Interaktion vor. Wir schauen uns den Interaktionsplot an. Er ist in Abbildung 7 zu finden. Ohne Musik werden mit Schriftgröße 16pt im Mittel 11.5 Aufgaben mehr bearbeitet als mit Schriftgröße 8 pt, während ohne Musik mit Schriftgröße 16pt im Mittel 34.5 Aufgaben mehr bearbeitet werden als mit Schriftgröße 8 pt.

Abbildung 7: Interaktionsplot




Um den Normal-Quantil zu erstellen, bestimmen wir die Residuen. Es gilt

$$\bar{y}_{11} = 40.5 \qquad \bar{y}_{21} = 52 \qquad \bar{y}_{12} = 42.5 \qquad \bar{y}_{22} = 80$$

Also gilt

$$e_{111} = 1.5 \quad e_{112} = -1.5$$
$$e_{211} = -1 \quad e_{212} = 1$$
$$e_{121} = -2.5 \quad e_{122} = 2.5$$
$$e_{221} = 7 \quad e_{222} = -7$$

Die geordneten Residuen sind

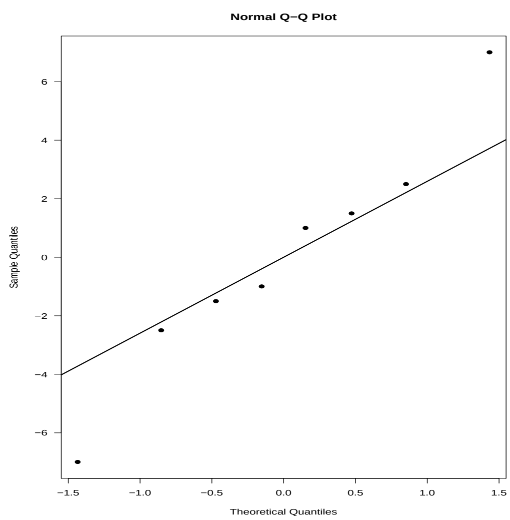
```
   -7.0  -2.5  -1.5  -1.0   1.0   1.5   2.5   7.0
```

Die Quantile der Normalverteilung sind

```
 -1.534 -0.887 -0.489 -0.157  0.157  0.489  0.887  1.534
```

Wir zeichnen die Gerade durch die Punkte $(-2, -0.688)$ und $(2, 0.688)$. Abbildung 8 zeigt den Normal-Quantil-Plot. Er deutet auf zwei Ausreißer hin.

Abbildung 8: Normal-Quantil-Plot



# 4 Versuchpläne mit $k$ Faktoren und identischer Anzahl von Versuchen auf den Faktorstufen

Im letzten Kapitel haben wir gelernt, wie man einen faktoriellen Versuchsplan analysiert, bei dem beide Faktoren jeweils zwei Faktorstufen besitzen. Nun wollen wir uns Versuchspläne mit $k$ Faktoren anschauen, bei denen jeder Faktor zwei Faktorstufen besitzt. Man spricht von einem $2^k$-Plan. Wir illustrieren die wesentlichen Aspekte am Beispiel des $2^3$-Plans.

**Beispiel 6**
In Ihrer Arbeit im BI-Projekt zu TQM im Wintersemester 2003/2004 untersuchen Christina Krause und Anika Esser, von welchen Faktoren die Wartezeit (in Sekunden) an der Kasse bei ALDI abhängt. Sie betrachten die folgenden Faktoren

A : Einkaufstag mit den Faktorstufen *Mittwoch* (-) und *Samstag* (+)

B : Uhrzeit mit den Faktorstufen *Vormittag* (-) und *Mittag* (+)

C : Filiale mit den Faktorstufen *stadtnah* (-) und *außerhalb* (+).

Auf jeder Faktorstufenkombination wurden zwei Beobachtungen gemacht. Die Daten sind in Tabelle 26 zu finden.

Tabelle 26: Daten eines $2^3$-Plans mit $n = 2$

| $A$ | $B$ | $C$ | Anzahl | |
|---|---|---|---|---|
| − | − | − | 366 | 257 |
| + | − | − | 312 | 225 |
| − | + | − | 405 | 453 |
| + | + | − | 321 | 317 |
| − | − | + | 456 | 508 |
| + | − | + | 322 | 353 |
| − | + | + | 382 | 461 |
| + | + | + | 332 | 363 |

## 4.1 Das Modell

Bei einem $2^2$-Plan bezeichnen wir die Zielgröße auf der $i$-ten Stufe des Faktors $A$ und der $j$-ten Stufe des Faktors $B$ bei der $k$-ten Wiederholung mit $Y_{ijk}$. Bei einem $2^k$-Plan benötigt man zur Beschreibung der Beobachtungen $k$ Indizes, falls $n = 1$ gilt, und $k + 1$ Indizes, falls $n$ größer als 1 ist.

**Beispiel 6 (fortgesetzt)**
Bei einem $2^3$-Plan mit $n = 1$ betrachten wir $Y_{ijk}$. Dabei bezieht sich der $r$-te Index auf die Faktorstufen des $r$-ten Faktors. Ist $n > 1$, so wählen wir $Y_{ijkl}$. Dabei beziehen sich die ersten 3 Indizes auf die drei Faktoren, während der letzte Index sich auf die $l$-te Beobachtung auf dieser Faktorstufenkombination bezieht.

Bei einem $2^k$-Plan enthält die Definitionsgleichung des Modells $2^k + 1$ Summanden.
Schauen wir uns dies für einen $2^3$-Plan mit $n$ Wiederholungen auf den Faktorstufenkombinationen an. Es gilt

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkl}$$

## 4.2 Die Effekte

### 4.2.1 Die Haupteffekte

Bei einem $2^1$-Plan gibt der Effekt des Faktors $A$ an, wie sich der Erwartungswert der Zielgröße ändert, wenn man von der Faktorstufe - zur Faktorstufe + übergeht. Es gilt also

$$E_A = \mu_2 - \mu_1$$

Bei einem $2^2$-Plan ist der Effekt des Faktors $A$ gleich dem Mittelwert aus dem Effekt des Faktors $A$, wenn der Faktor $B$ auf - steht und dem Effekt des Faktors $A$, wenn der Faktor $B$ auf + steht:

$$E_A = \frac{1}{2}\left(\underbrace{\mu_{21} - \mu_{11}}_{E_A \text{ auf } B\ \texttt{-}} + \underbrace{\mu_{22} - \mu_{12}}_{E_A \text{ auf } B\ \texttt{+}}\right)$$

Dies können wir auch folgendermaßen darstellen:

$$E_A = \underbrace{\frac{\mu_{21} + \mu_{22}}{2}}_{E(Y) \text{ auf } A\ \texttt{+}} - \underbrace{\frac{\mu_{11} + \mu_{12}}{2}}_{E(Y) \text{ auf } A\ \texttt{-}} = \mu_{2\cdot} - \mu_{1\cdot}$$

Wir können also bei einem $2^2$-Plan den Effekt des Faktors $A$ als Differenz aus dem Erwartungswert der Zielgröße, wenn $A$ auf + steht, und dem Erwartungswert der Zielgröße, wenn $A$ auf – steht, auffassen. Dies gilt auch für den Effekt des Faktors $B$.
Bei einem $2^k$-Plan definieren wir die Haupteffekte analog.
Schauen wir uns einen $2^3$-Plan an. Bei diesem gilt

$$
\begin{aligned}
E_A &= \mu_{2\cdot\cdot} - \mu_{1\cdot\cdot} \\
&= \frac{1}{4}\left(\mu_{211} + \mu_{221} + \mu_{212} + \mu_{222}\right) - \frac{1}{4}\left(\mu_{111} + \mu_{121} + \mu_{112} + \mu_{122}\right) \\
&= \frac{1}{4}\left(-\mu_{111} + \mu_{211} - \mu_{121} + \mu_{221} - \mu_{112} + \mu_{212} - \mu_{122} + \mu_{222}\right)
\end{aligned}
$$

Wir sehen, dass alle $\mu_{ijk}$ mit $i = 1$ ein negatives Vorzeichen und alle $\mu_{ijk}$ mit $i = 2$ ein positives Vorzeichen besitzen.
Analog erhalten wir dann die Effekte von $B$

$$E_B = \frac{1}{4}\left(-\mu_{111} - \mu_{211} + \mu_{121} + \mu_{221} - \mu_{112} - \mu_{212} + \mu_{122} + \mu_{222}\right)$$

und $C$

$$E_C = \frac{1}{4}\left(-\mu_{111} - \mu_{211} - \mu_{121} - \mu_{221} + \mu_{112} + \mu_{212} + \mu_{122} + \mu_{222}\right)$$

Bei $E_B$ besitzen alle $\mu_{ijk}$ mit $j = 1$ ein negatives Vorzeichen und alle $\mu_{ijk}$ mit $j = 2$ ein positives Vorzeichen, während bei $E_C$ alle $\mu_{ijk}$ mit $k = 1$ ein negatives Vorzeichen und alle $\mu_{ijk}$ mit $k = 2$ ein positives Vorzeichen besitzen.
Wir können den Effekt eines Faktors aber auch als Mittelwert des Effekts dieses Faktors auf allen Kombinationen der Faktorstufen der anderen Faktoren. Für den Faktor $A$ heißt dies bei einem $2^3$-Plan.

$$E_A = \frac{1}{4}\left(\underbrace{\mu_{211} - \mu_{111}}_{E_A \text{ auf } B\text{ -},C\text{ -}} + \underbrace{\mu_{221} - \mu_{121}}_{E_A \text{ auf } B\text{ +},C\text{ -}} + \underbrace{\mu_{212} - \mu_{112}}_{E_A \text{ auf } B\text{ -},C\text{ +}} + \underbrace{\mu_{222} - \mu_{122}}_{E_A \text{ auf } B\text{ +},C\text{ +}}\right)$$

### 4.2.2 Die Interaktionseffekte zwischen zwei Faktoren

Schauen wir uns die Interaktionseffekte zwischen zwei Faktoren an. Bei $k$ Faktoren gibt es

$$\binom{k}{2} = \frac{k \cdot (k-1)}{2}$$

Interaktionseffekte zwischen zwei Faktoren. Für $k = 3$ sind es die Interaktionseffekte $E_{AB}$, $E_{AC}$ und $E_{BC}$.
Wie können wir diese bei einem $2^3$-Plan definieren?
Schauen wir uns exemplarisch $E_{AB}$ an, wobei uns noch einmal an den $2^2$-Plan erinnern. Bei diesem vergleichen wir den Effekt von $A$, wenn $B$ auf **+** steht, mit Effekt von $A$, wenn $B$ auf **-** steht. Es gilt also

$$E_{AB} = \frac{1}{2}\left(\mu_{22} - \mu_{12} - (\mu_{21} - \mu_{11})\right)$$

Diese Definition können wir direkt auf einen $2^3$-Plan übertragen, indem wir $\mu_{ij}$ durch $\mu_{ij.}$ ersetzen. Dabei ist

$$\mu_{ij.} = \frac{1}{2}\left(\mu_{ij1} + \mu_{ij2}\right)$$

Wir erhalten dann

$$\begin{aligned} E_{AB} &= \frac{1}{2}\left(\mu_{22.} - \mu_{12.} - (\mu_{21.} - \mu_{11.})\right) \\ &= \frac{1}{4}\left(\mu_{111} - \mu_{211} - \mu_{121} + \mu_{221} + \mu_{112} - \mu_{212} - \mu_{122} + \mu_{222}\right) \end{aligned}$$

Entsprechend erhalten wir

$$E_{AC} = \frac{1}{4}\left(\mu_{111} - \mu_{211} + \mu_{121} - \mu_{221} - \mu_{112} + \mu_{212} - \mu_{122} + \mu_{222}\right)$$

und

$$E_{BC} = \frac{1}{4}\left(\mu_{111} + \mu_{211} - \mu_{121} - \mu_{221} - \mu_{112} - \mu_{212} + \mu_{122} + \mu_{222}\right)$$

### 4.2.3 Interaktionseffekte zwischen mehr als zwei Faktoren

Nun kommen wir zum Interaktionseffekt $E_{ABC}$ zwischen allen drei Faktoren. Beim Interaktionseffekt zwischen zwei Faktoren vergleichen wir den Effekt des einen Faktors auf den beiden Stufen des anderen Faktors. Beim Interaktionseffekt zwischen drei Faktoren vergleichen wir den Interaktionseffekt zwischen zwei Faktoren auf den Stufen des dritten Faktors. Wir vergleichen also den Interaktionseffekt zwischen $A$ und $B$, wenn $C$ auf + steht, mit dem Interaktionseffekt zwischen $A$ und $B$, wenn $C$ auf - steht.
Den Interaktionseffekt zwischen $A$ und $B$, wenn $C$ auf + steht, ist

$$\frac{1}{2}\left(\mu_{112}-\mu_{212}-\mu_{122}+\mu_{222}\right)$$

Den Interaktionseffekt zwischen $A$ und $B$, wenn $C$ auf - steht, ist

$$\frac{1}{2}\left(\mu_{111}-\mu_{211}-\mu_{121}+\mu_{221}\right)$$

Also ist

$$\begin{aligned} E_{ABC} &= \frac{1}{2}\left[\frac{1}{2}\left(\mu_{112}-\mu_{212}-\mu_{122}+\mu_{222}\right)-\frac{1}{2}\left(\mu_{111}-\mu_{211}-\mu_{121}+\mu_{221}\right)\right] \\ &= \frac{1}{4}\left(-\mu_{111}+\mu_{211}+\mu_{121}-\mu_{221}+\mu_{112}-\mu_{212}-\mu_{122}+\mu_{222}\right) \end{aligned}$$

Es liegt keine Interaktion zwischen allen drei Faktoren vor, wenn der Interaktionseffekt zwischen $A$ und $B$, wenn $C$ auf + steht, und der Interaktionseffekt zwischen $A$ und $B$, wenn $C$ auf - steht, identisch sind.
Wir können $E_{ABC}$ auch anders interpretieren

$$E_{ABC} = \frac{1}{2}\left[\underbrace{\frac{1}{2}\left(\mu_{211}-\mu_{221}-\mu_{212}+\mu_{222}\right)}_{E_{BC} \text{ auf } A \text{ +}}-\underbrace{\frac{1}{2}\left(\mu_{111}-\mu_{121}-\mu_{112}+\mu_{122}\right)}_{E_{BC} \text{ auf } A \text{ -}}\right]$$

$$E_{ABC} = \frac{1}{2}\left[\underbrace{\frac{1}{2}\left(\mu_{121}-\mu_{221}-\mu_{122}+\mu_{222}\right)}_{E_{AC} \text{ auf } B \text{ +}}-\underbrace{\frac{1}{2}\left(\mu_{111}-\mu_{211}-\mu_{112}+\mu_{212}\right)}_{E_{AC} \text{ auf } B \text{ -}}\right]$$

## 4.3 Schätzen und Testen

In diesem Abschnitt wollen wir uns damit beschäftigen, wie man die Effekte schätzen und ihre Signifikanz überprüfen kann. dabei liefert der Algorithmus von Yates alle Bausteine. Wir haben diesen im letzten Kapitel am Beispiel eines $2^2$-Plans ausführlich dargestellt. Nun schauen wir ihn uns für einen beliebigen $2^k$-Plan an.

### 4.3.1 Der Algorithmus von Yates

Am Beginn der Analyse steht der Algorithmus von Yates. Wir haben diesen bereits bei der ein- und zweifaktoriellen Varianzanalyse betrachtet. Schauen wir ihn uns für beliebige Versuchspläne an.

Zuerst müssen wir die Tabelle erstellen. Hierzu schreiben wir die Symbole für die $k$ Faktoren nebeneinander, beginnend mit dem Faktor $A$. Unter den Namen jedes Faktors schreiben wir bei einem $2^k$-Plan $2^k$ Symbole. Beim Faktor $A$ wechseln sich die Symbole − und + ab. Beim Faktor wechseln sich Paare von − mit Paaren von + ab. Beim Faktor $C$ wechseln sich Folgen aus vier  mit Folgen aus vier + ab. Diese Struktur wird immer weiter fortgesetzt. Beim letzten Faktor folgt einem Block aus $2^{k-1}$ Minus ein Block aus $2^{k-1}$ Plus. Tabelle 27 zeigt dies für einen $2^3$-Plan.

Tabelle 27: Vorzeichenstruktur eines $2^3$-Plans

| $A$ | $B$ | $C$ |
|---|---|---|
| − | − | − |
| + | − | − |
| − | + | − |
| + | + | − |
| − | − | + |
| + | − | + |
| − | + | + |
| + | + | + |

Zu jeder Zeile gehört eine Faktorstufenkombination. So stehen in der ersten Zeile alle Faktoren auf −, während sie in der letzten Zeile alle auf + stehen.

Wir schreiben nun in jede Zeile die Summe der Beobachtungen auf dieser Faktorstufenkombination. Dabei bezeichnen wir die Summe aller Beobachtungen mit (1), wenn alle Faktoren auf − stehen. Ansonsten bezeichnen wir

die Summe der Beobachtungen als Folge aus den Buchstaben der Faktoren, die auf + stehen. Hierbei verwenden wir Kleinbuchstaben.
Tabelle 28 zeigt dies für einen $2^3$-Plan.

Tabelle 28: Ausgangtabelle für den Yates-Algorithmus bei einem $2^3$-Plan

| $A$ | $B$ | $C$ | Summe |
|---|---|---|---|
| − | − | − | (1) |
| + | − | − | $a$ |
| − | + | − | $b$ |
| + | + | − | $ab$ |
| − | − | + | $c$ |
| + | − | + | $ac$ |
| − | + | + | $bc$ |
| + | + | + | $abc$ |

Da man weiß, zu welchen Faktorstufenkombinationen die Symbole gehören, verzichtet man in der Regel auf die Spalten mit den Faktoren.
Tabelle 29 zeigt dies für einen $2^3$-Plan.

Tabelle 29: Ausgangtabelle für den Yates-Algorithmus bei einem $2^3$-Plan

| Summe |
|---|
| (1) |
| $a$ |
| $b$ |
| $ab$ |
| $c$ |
| $ac$ |
| $bc$ |
| $abc$ |

Nun kann der Algorithmus von Yates beginnen.
Wir beginnen mit der ersten Spalte. Beginnend mit der ersten Zeile addieren wir zwei aufeinander folgende Zahlen und schreiben das Ergebnis in die nächste Spalte, wobei wir auch hier in der ersten Zeile beginnen. Haben wir

alle Paare addiert, so gehen wir wieder in die erste Zeile und subtrahieren bei jedem Paar die erste Zahl von der zweiten und füllen mit dem Ergebnis die folgende Spalte auf. Nun ist die zweite Spalte voll.
Tabelle 30 zeigt dies für einen $2^3$-Plan.

Tabelle 30: Erster Schritt des Yates-Algorithmus bei einem $2^3$-Plan

| Summe | Erster Schritt |
|---|---|
| $(1)$ | $(1) + a$ |
| $a$ | $b + ab$ |
| $b$ | $c + ac$ |
| $ab$ | $bc + abc$ |
| $c$ | $a - (1)$ |
| $ac$ | $ab - b$ |
| $bc$ | $ac - c$ |
| $abc$ | $abc - bc$ |

Die eben beschriebenen Handlungen führen wir bei einem $2^k$-Plan $k$-mal durch, wobei wir sie immer auf die neu gewonnene Spalte anwenden.
Tabelle 31 zeigt dies für einen $2^3$-Plan.

Tabelle 31: Erster Schritt des Yates-Algorithmus bei einem $2^3$-Plan

| Summe | Erster Schritt | Zweiter Schritt | Dritter Schritt |
|---|---|---|---|
| $(1)$ | $(1) + a$ | $(1) + a + b + ab$ | $(1) + a + b + ab + c + ac + bc + abc$ |
| $a$ | $b + ab$ | $c + ac + bc + abc$ | $a - (1) + ab - b + ac - c + abc - bc$ |
| $b$ | $c + ac$ | $a - (1) + ab - b$ | $b + ab - (1) - a + bc + abc - c - ac$ |
| $ab$ | $bc + abc$ | $ac - c + abc - bc$ | $ab - b - a + (1) + abc - bc - ac + c$ |
| $c$ | $a - (1)$ | $b + ab - (1) - a$ | $c + ac + bc + abc - (1) - a - b - ab$ |
| $ac$ | $ab - b$ | $bc + abc - c - ac$ | $ac - c + abc - bc - a + (1) - ab + b$ |
| $bc$ | $ac - c$ | $ab - b - a + (1)$ | $bc + abc - c - ac - b - ab + (1) + a$ |
| $abc$ | $abc - bc$ | $abc - bc - ac + c$ | $abc - bc - ac + c - ab + b + a - (1)$ |

Der Algorithmus liefert die Kontraste $K_F$ aller Faktoren beziehungsweise Faktorstufenkombinationen. Mit diesen kann man die Schätzer der Effekte und die Quadratsummen bestimmen.

**Beispiel 6 (fortgesetzt)**
Wir bilden auf jeder Faktorstufenkombination die Summe der Beobachtungen, erstellen die Tabelle und führen die einzelnen Schritte durch.

Tabelle 32: Der Algorithmus von Yates für die Wartezeit an der Kasse

| | | | | |
|---|---|---|---|---|
| (1) | 623 | 1160 | 2656 | 5833 |
| $a$ | 537 | 1496 | 3177 | -743 |
| $b$ | 858 | 1639 | -306 | 235 |
| $ab$ | 638 | 1538 | -437 | 7 |
| $c$ | 964 | -86 | 336 | 521 |
| $ac$ | 675 | -220 | -101 | -131 |
| $bc$ | 843 | -289 | -134 | -437 |
| $abc$ | 695 | -148 | 141 | 275 |

Somit gilt

$$\begin{aligned} K_A &= -743 \\ K_B &= 235 \\ K_C &= 521 \\ K_{AB} &= 7 \\ K_{AC} &= -131 \\ K_{BC} &= -437 \\ K_{ABC} &= 275 \end{aligned}$$

### 4.3.2 Schätzen der Effekte

Die Schätzer der Effekte erhält man, indem man auf jeder Faktorstufenkombination den Mittelwert bestimmt und diese Mittelwerte in die Definitionsgleichungen der Effekte einsetzt. Bei einem $2^3$-Plan ist der Schätzer $e_A$ des Effekts $E_A$ also:

$$e_A = \frac{1}{4}\left(-\bar{y}_{111} + \bar{y}_{211} - \bar{y}_{121} + \bar{y}_{221} - \bar{y}_{112} + \bar{y}_{212} - \bar{y}_{122} + \bar{y}_{222}\right)$$

und der Schätzer $e_{ABC}$ des Effekts $E_{ABC}$ also:

$$e_{ABC} = \frac{1}{4}\left(-\bar{y}_{111} + \bar{y}_{211} + \bar{y}_{121} - \bar{y}_{221} + \bar{y}_{112} - \bar{y}_{212} - \bar{y}_{122} + \bar{y}_{222}\right)$$

**Beispiel 6 (fortgesetzt)**
Es gilt

$$\begin{array}{ll} \bar{y}_{111} = 311.5 & \bar{y}_{211} = 268.5 \\ \bar{y}_{121} = 429.0 & \bar{y}_{221} = 319.0 \\ \bar{y}_{112} = 482.0 & \bar{y}_{212} = 337.5 \\ \bar{y}_{122} = 421.5 & \bar{y}_{222} = 347.5 \end{array}$$

Also gilt

$$e_A = \frac{1}{4}(-311.5+268.5-429.0+319.0-482.0+337.5-421.5+347.5) = -92.875$$

und

$$e_{ABC} = \frac{1}{4}(-311.5+268.5+429.0-319.0+482.0-337.5-421.5+347.5) = 34.375$$

Schauen wir uns diesen Effekt noch einmal genauer an. Er setzt sich aus dem Interaktionseffekt zwischen $A$ und $B$, wenn $C$ auf - steht, und dem Interaktionseffekt zwischen $A$ und $B$, wenn $C$ auf + steht, zusammen.
Der Interaktionseffekt zwischen $A$ und $B$, wenn $C$ auf - steht, ist

$$\frac{1}{2}\left(\bar{y}_{111} - \bar{y}_{211} - \bar{y}_{121} + \bar{y}_{221}\right) = \frac{1}{2}(311.5 - 268.5 - 429.0 + 319.0) = -33.5$$

Der Interaktionseffekt zwischen $A$ und $B$, wenn $C$ auf + steht, ist

$$\frac{1}{2}\left(\bar{y}_{112} - \bar{y}_{212} - \bar{y}_{122} + \bar{y}_{222}\right) = \frac{1}{2}(482.0 - 337.5 - 421.5 + 347.5) = 35.25$$

Also gilt

$$e_{ABC} = 0.5(35.25 + 33.5) = 34.375$$

Wir können den geschätzten Effekt $e_{ABC}$ auch grafisch darstellen. Hierzu zeichnen wir das Interaktionsdiagramm zwischen $A$ und $B$, wenn $C$ auf - steht, das Interaktionsdiagramm zwischen $A$ und $B$, wenn $C$ auf + steht.

**Beispiel 6 (fortgesetzt)**
Abbildung 9 zeigt das Interaktionsdiagramm zwischen $A$ und $B$, wenn $C$ auf - steht, und Abbildung 10 das Interaktionsdiagramm zwischen $A$ und $B$, wenn $C$ auf + steht.

Abbildung 9: Interaktionsdiagramm zwischen $A$ und $B$, wenn $C$ auf **-** steht.

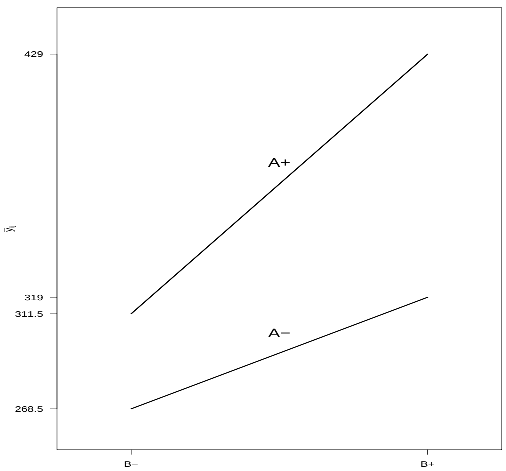
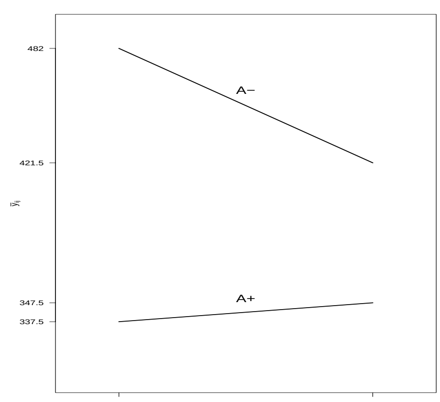



Abbildung 10: Interaktionsdiagramm zwischen $A$ und $B$, wenn $C$ auf **+** steht.




Die Abbildungen deuten darauf hin, dass $E_{AB}$ sich auf den beiden Stufen des Faktors $C$ unterscheidet. Es stellt sich die Frage, ob dieser Unterschied signifikant ist. Hiermit beschäftigen wir uns im nächsten Abschnitt.

Vorher schauen wir uns aber an, wie wir mit den Ergebnissen des Algorithmus von Yates die Effekte schätzen können.

Werden bei einem $2^k$-Plan auf jeder Faktorstufenkombination $n$ Versuche durchgeführt, so gilt

$$\boxed{e_F = \frac{K_F}{n\,2^{k-1}}} \tag{92}$$

Dabei ist $F$ entweder ein Faktor wie $A$ oder eine Interaktion wie $AB$ und $K_F$ ist der zugehörige Kontrast.

**Beispiel 6 (fortgesetzt)**
Es gilt

$$\begin{aligned}
e_A &= \frac{K_A}{4n} = \frac{-743}{8} = -92.875 \\
e_B &= \frac{K_B}{4n} = \frac{235}{8} = 29.375 \\
e_C &= \frac{K_C}{4n} = \frac{521}{8} = 65.125 \\
e_{AB} &= \frac{K_{AB}}{4n} = \frac{7}{8} = 0.875 \\
e_{AC} &= \frac{K_{AC}}{4n} = \frac{-131}{8} = -16.375 \\
e_{BC} &= \frac{K_{BC}}{4n} = \frac{-437}{8} = -54.625 \\
e_{ABC} &= \frac{K_{ABC}}{4n} = \frac{7}{8} = 34.375
\end{aligned}$$

### 4.3.3 Tests

In welcher Reihenfolge man die Signifikanz der einzelnen Effekte überprüft, wollen wir an einem $2^3$-Plan darstellen.
Wir testen zuerst die Hypothese

$$H_0 : \ E_{ABC} = 0$$

Wird diese abgelehnt, beenden wir die Testprozedur. Wird sie nicht abgelehnt, testen wir die Hypothesen

$$H_0 : \ E_{AB} = 0 \tag{93}$$

$$H_0 : \ E_{AC} = 0 \tag{94}$$

$$H_0 : \ E_{BC} = 0 \tag{95}$$

Werden mindestens zwei der Hypothesen (93), (94) oder (95) abgelehnt, so beenden wir die Testprozedur.

Wird genau eine der drei Hypothesen (93), (94) oder (95) abgelehnt, so ist der zugehörige Interaktionseffekt signifikant von 0 verschieden. Wir können dann aber immer noch testen, ob der andere Haupteffekt signifikant ist.

Wird also die Hypothese

$$H_0 : E_{BC} = 0$$

abgelehnt, so können wir immer noch testen

$$H_0 : E_A = 0$$

Wird keine der drei Hypothesen (93), (94) oder (95) abgelehnt, so testen wir

$$H_0 :\ E_A = 0$$

$$H_0 :\ E_B = 0$$

$$H_0 :\ E_C = 0$$

Um die Hypothesen überprüfen zu können, benötigen wir die Quadratsummen. Diese erhalten wir aus den Ergebnissen des Algorithmus von Yates.

Werden auf jeder Faktorstufe $n$ Beobachtungen gemacht, so gilt

$$\boxed{SS_F = \frac{K_F^2}{n\,2^k}} \tag{96}$$

**Beispiel 6 (fortgesetzt)**

Es gilt

$$\begin{aligned}
SS_A &= (-743)^2/16 = 34503 \\
SS_B &= 235^2/16 = 3452 \\
SS_C &= 521^2/16 = 16966 \\
SS_{AB} &= 7^2/16 = 3 \\
SS_{AC} &= (-131)^2/16 = 1073 \\
SS_{BC} &= (-437)^2/16 = 11936 \\
SS_{ABC} &= 275^2/16 = 4727
\end{aligned}$$

Nur $SS_R$ liefert der Algorithmus von Yates nicht. Wir bestimmen $SS_R$, indem wir von jeder Beobachtung den Mittelwert ihrer Faktorstufenkombination subtrahieren und das Ergebnis quadrieren. $SS_R$ ist die Summe all dieser quadrierten Differenzen.

**Beispiel 6 (fortgesetzt)**

$$\begin{aligned}
SS_R &= (366-311.5)^2+(257-311.5)^2\\
&+ (312-268.5)^2+(225-268.5)^2\\
&+ (405-429)^2+(453-429)^2\\
&+ (321-319)^2+(317-319)^2\\
&+ (456-482)^2+(508-482)^2\\
&+ (322-337.5)^2+(353-337.5)^2\\
&+ (382-421.5)^2+(461-421.5)^2\\
&+ (332-347.5)^2+(363-347.5)^2\\
&= 16319
\end{aligned}$$

Die Anzahl der Freiheitsgrade ist für die Quadratsumme jedes Faktors und jeder Faktorstufenkombination gleich 1 und für $SS_T$ gleich $n \cdot 2^k - 1$. Also ist sie für $SS_R$ gleich $(n-1)\cdot 2^k$. Wir sehen, dass wir für $SS_R$ keine Freiheitsgrade haben, falls $n$ gleich 1 ist. Außerdem ist $SS_R = 0$ für $n = 1$. Wir werden weiter unten zeigen, wie wir die Signifikanz aller Effekte für $n = 1$ grafisch überprüfen können.

Nun können wir die ANOVA-Tabelle aufstellen. Sie ist für einen $2^3$-Plan in Tabelle 33 zu finden.

Tabelle 33: Allgemeiner Aufbau der ANOVA-Tabelle eines $2^3$-Plans

| Quelle der Variation | Quadrat-summen | Freiheits-grade | Mittlere Quadratsummen | $F$ |
|---|---|---|---|---|
| $A$ | $SS_A$ | 1 | $MSS_A$ | $MSS_A/MSS_R$ |
| $B$ | $SS_B$ | 1 | $MSS_B$ | $MSS_B/MSS_R$ |
| $C$ | $SS_C$ | 1 | $MSS_C$ | $MSS_C/MSS_C$ |
| $AB$ | $SS_{AB}$ | 1 | $MSS_{AB}$ | $MSS_{AB}/MSS_R$ |
| $AC$ | $SS_{AC}$ | 1 | $MSS_{AC}$ | $MSS_{AC}/MSS_R$ |
| $BC$ | $SS_{BC}$ | 1 | $MSS_{BC}$ | $MSS_{BC}/MSS_R$ |
| $ABC$ | $SS_{ABC}$ | 1 | $MSS_{ABC}$ | $MSS_{ABC}/MSS_R$ |
| Rest | $SS_R$ | $8n-8$ | $MSS_R$ | |
| Gesamt | $SS_T$ | $8n-1$ | | |

**Beispiel 6 (fortgesetzt)**
Es gilt

| Quelle der Variation | Quadrat-summen | Freiheits-grade | Mittlere Qua-dratsummen | $F$ |
|---|---|---|---|---|
| $A$ | 34503 | 1 | 34503 | 16.915 |
| $B$ | 3452 | 1 | 3452 | 1.692 |
| $C$ | 16965 | 1 | 16965 | 8.317 |
| $AB$ | 3 | 1 | 3 | 0.002 |
| $AC$ | 1073 | 1 | 1073 | 0.526 |
| $BC$ | 11936 | 1 | 11936 | 5.851 |
| $ABC$ | 4727 | 1 | 4727 | 2.317 |
| Rest | 16319 | 8 | 2040 | |
| Gesamt | 88978 | 15 | | |

Wir testen die Signifikanz jedes Faktors mit

$$\boxed{F_F = \frac{MSS_F}{MSS_R}} \tag{97}$$

Die Nullhypothese $H_0$ ist, dass der Faktor keinen Effekt hat. Wenn $H_0$ zutrifft, ist $F_{Faktor}$ mit 1 und $(n-1) \cdot 2^k$ Freiheitsgraden $F$-verteilt. Wir lehnen $H_0$ ab, wenn gilt $F_{Faktor} \geq F_{1,(n-1)\cdot 2^k;1-\alpha}$.
Bei einem $2^3$-Plan testen wir zuerst die Signifikanz von $E_{ABC}$. Ist $E_{ABC}$ signifikant, so ist es nicht sinnvoll, nach den Interaktionseffekten zwischen zwei Faktoren oder den Effekten der Faktoren zu fragen.

**Beispiel 6 (fortgesetzt)**
bei allen tests ist der kritische Wert $F_{1,8;0.95} = 5.32$.
Wir testen

$$H_0 : E_{ABC} = 0$$

Wegen $F_{ABC} = 2.317 < 5.32$ lehnen wir $H_0$ nicht ab.

Ist $E_{ABC}$ nicht signifikant, so testen wir auf Signifikanz von $E_{AB}$, $E_{AC}$ und $E_{BC}$.

**Beispiel 6 (fortgesetzt)**
Wir testen

$$H_0 : E_{AB} = 0$$

Wegen $F_{AB} = 0.002 < 5.32$ lehnen wir $H_0$ nicht ab.
Wir testen

$$H_0 : E_{AC} = 0$$

Wegen $F_{AC} = 0.526 < 5.32$ lehnen wir $H_0$ nicht ab.
Wir testen

$$H_0 : E_{BC} = 0$$

Wegen $F_{BC} = 5.851 > 5.32$ lehnen wir $H_0$ ab.
Es liegt also Interaktion zwischen der Einkaufszeit und der Filiale vor.
Es gilt

$$\bar{y}_{\cdot 11} = 290 \qquad \bar{y}_{\cdot 21} = 374$$

$$\bar{y}_{\cdot 12} = 409.75 \qquad \bar{y}_{\cdot 22} = 384.5$$

In den stadtnahen Filialen wartet man im Mittel am Mittag 84 Sekunden länger als am Vormittag, während man in den Filialen, die außerhalb liegen, am Mittag im Mittel 25.25 Sekunden weniger als am Vormittag wartet.
Wir können jetzt noch auf Signifikanz des Faktors $A$ testen:

$$H_0 : E_A = 0$$

Der kritische Wert ist $F_{1,8;0.95} = 5.32$. Wegen $F_A = 16.915$ lehnen wir $H_0$ ab. Der Einkaufstag hat einen signifikanten Einfluss auf die Wartezeit, wobei die Wartezeit am Samstag im Mittel um rund 93 Sekunden kürzer ist als am Mittwoch.

Man sollte die Annahme der Normalverteilung überprüfen. Hierzu erstellt man einen Normal-Quantil-Plot der Residuen. Diese erhält man, indem man von jeder Beobachtung den Mittelwert der Beobachtungen auf der Faktorstufenkombination subtrahiert, zu der sie gehört.

**Beispiel 6 (fortgesetzt)**
Die Residuen sind

$$\begin{aligned}
e_{1111} &= 54.5 & e_{1112} &= -54.5\\
e_{2111} &= 43.5 & e_{2112} &= -43.5\\
e_{1211} &= -24 & e_{1212} &= 24\\
e_{2211} &= 2 & e_{2212} &= -2\\
e_{1121} &= -26 & e_{1121} &= 26\\
e_{2121} &= -15.5 & e_{2122} &= 15.5\\
e_{1221} &= -39.5 & e_{1222} &= 39.5\\
e_{2221} &= -15.5 & e_{2222} &= 15.5
\end{aligned}$$

Bei einem *Normal-Quantil-Plot* vergleicht man die Residuen mit ihren Erwartungswerten unter Normalverteilung. Der Erwartungswert des $i$. größten Residuums $e_{(i)}$ unter Normalverteilung ist approximativ $\Phi^{-1}((i-0.5)/N)$. Dabei ist $N = n \cdot 2^k$ und $\Phi^{-1}(p)$ die Inverse der Standardnormalteilung.
Für $i = 1, \ldots, N$ zeichnet man $e_{(i)}$ gegen $\Phi^{-1}((i-0.5)/N)$. Liegt Normalverteilung vor, so sollten die Punkte um eine Gerade streuen. Hierbei bestimmt man aber nicht die KQ-Gerade, da diese nicht robust ist. Man legt die Gerade vielmehr durch das untere und obere Quartil der Punktepaare.

**Beispiel 6 (fortgesetzt)**
Die geordneten Residuen sind

```
-54.5 -43.5 -39.5 -26.0 -24.0 -15.5 -15.5  -2.0
  2.0  15.5  15.5  24.0  26.0  39.5  43.5  54.5
```

Speziell gilt

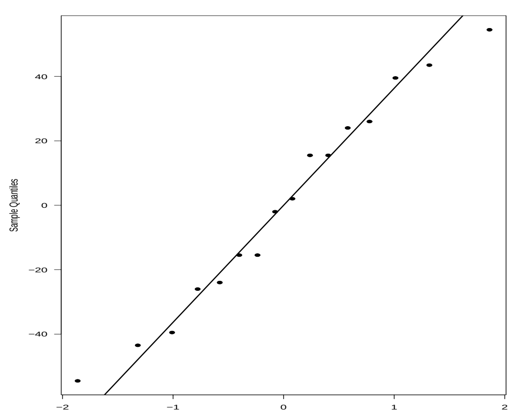
$$\Phi^{-1}((1-0.5)/16) = \Phi^{-1}(0.03125) = -1.86$$

Die anderen Quantile ergeben sich analog. Alle lauten

```
-1.86 -1.32 -1.01 -0.78 -0.58 -0.40 -0.24 -0.08
 0.08  0.24  0.40  0.58  0.78  1.01  1.32  1.86
```

Das untere Quartil der Residuen ist $-25$ und das obere 25. Das untere Quartil der Quantile der Normalverteilung ist $-0.68$ und das obere 0.68. Wir legen die Gerade also durch die Punkte $(-25, -0.68)$ und $(25, 0.68)$. Abbildung 11 zeigt den Normal-Quantil-Plot mit der Geraden. Die Annahme der Normalverteilung ist gerechtfertigt.

Abbildung 11: Normal-Quantil-Plot



## 4.4 Tests bei einem $2^k$-Plan mit $n = 1$

Für $n = 1$ ist $SS_R = 0$. Somit können wir nicht alle Effekte testen.

**Beispiel 7**
Jing Zeng und Lars Hartwig untersuchen in ihrer Projektarbeit, welche Faktoren einen Einfluss auf die Zeit haben, die man benötigt, um ein Puzzle zu vollenden.
Sie betrachten folgende Faktoren

`A`: Geschlecht mit den Faktorstufen *männlich* (-) und *weiblich* (+)

`B`: Nationalität mit den Faktorstufen *deutsch* (-) und *chinesisch* (+)

`C`: Vorlage mit den Faktorstufen *kurzfristig* (-) und *permanent* (+)

Gemessen wurde die Zeit in Sekunden. Auf jeder Faktorstufenkombination wurde eine Beobachtung gemacht. Die Daten sind in Tabelle 34 zu finden.

Tabelle 34: Daten eines $2^3$-Plans mit $n = 1$

| $A$ | $B$ | $C$ | Anzahl |
|---|---|---|---|
| - | - | - | 535 |
| + | - | - | 360 |
| - | + | - | 758 |
| + | + | - | 1497 |
| - | - | + | 592 |
| + | - | + | 316 |
| - | + | + | 1163 |
| + | + | + | 1646 |

Wir wenden den Algorithmus von Yates an.

| | | | | |
|---|---|---|---|---|
| (1) | 535 | 895 | 3150 | 6867 |
| $a$ | 360 | 2255 | 3717 | 771 |
| $b$ | 758 | 908 | 564 | 3261 |
| $ab$ | 1497 | 2809 | 207 | 1673 |
| $c$ | 592 | -175 | 1360 | 567 |
| $ac$ | 316 | 739 | 1901 | -357 |
| $bc$ | 1163 | -276 | 914 | 541 |
| $abc$ | 1646 | 483 | 759 | -155 |

Somit sind die geschätzten Effekte:

$$\begin{aligned}
e_A &= \frac{771}{4} = 192.75 \\
e_B &= \frac{3261}{4} = 815.25 \\
e_C &= \frac{567}{4} = 141.75 \\
e_{AB} &= \frac{1673}{4} = 418.25 \\
e_{AC} &= \frac{-357}{4} = -89.25 \\
e_{BC} &= \frac{541}{4} = 135.25 \\
e_{ABC} &= \frac{-155}{4} = -38.75
\end{aligned}$$

Somit gilt

$$\begin{aligned}
SS_A &= \frac{771^2}{8} = 74305.13 \\
SS_B &= \frac{3261^2}{8} = 1329265 \\
SS_C &= \frac{567^2}{8} = 40186.13 \\
SS_{AB} &= \frac{1673^2}{8} = 349866.1 \\
SS_{AC} &= \frac{(-357)^2}{8} = 15931.13 \\
SS_{BC} &= \frac{541^2}{8} = 36585.13 \\
SS_{ABC} &= \frac{(-155)^2}{8} = 3003.125
\end{aligned}$$

Es gibt aber zwei Möglichkeiten, mit denen man überprüfen kann, ob Effekte signifikant sind.

1. Wir können annehmen, dass bestimmte Interaktionseffekte nicht im Modell auftreten. Somit können wir deren Quadratsummen als $SS_R$ auffassen.

**Beispiel 7 (fortgesetzt)**
Wir unterstellen, dass $E_{ABC}$ gleich 0 ist, und erhalten folgende ANOVA-Tabelle:

| Quelle der Variation | Quadrat-summen | Freiheits-grade | Mittlere Qua-dratsummen | $F$ |
|---|---|---|---|---|
| $A$ | 74305.1 | 1 | 74305.1 | 24.74 |
| $B$ | 1329265.0 | 1 | 1329265.0 | 442.63 |
| $C$ | 40186.1 | 1 | 40186.1 | 13.38 |
| $AB$ | 349866.1 | 1 | 349866.1 | 116.50 |
| $AC$ | 15931.1 | 1 | 15931.1 | 5.30 |
| $BC$ | 36585.1 | 1 | 36585.1 | 12.18 |
| $Rest$ | 3003.1 | 1 | 3003.1 | |
| Gesamt | 1849142.0 | 7 | | |

Es gilt $F_{1,1;0.95} = 161.45$. Somit ist der Effekt von $B$ signifikant von 0 verschieden.

Wir können aber auch unterstellen, dass alle Interaktionseffekte gleich 0 sind. dann erhalten wir folgende ANOVA-Tabelle.

| Quelle der Variation | Quadrat-summen | Freiheits-grade | Mittlere Qua-dratsummen | $F$ |
|---|---|---|---|---|
| $A$ | 74305.1 | 1 | 74305.1 | 0.73 |
| $B$ | 1329265.0 | 1 | 1329265.0 | 13.12 |
| $C$ | 40186.1 | 1 | 40186.1 | 0.40 |
| $Rest$ | 405386 | 4 | 101346 | |
| Gesamt | 1849142.0 | 7 | | |

Es gilt $F_{1,4;0.95} = 7.71$. Somit ist nur der Effekt von $B$ signifikant von 0 verschieden.

Wir sehen, dass beide Vorgehensweisen das gleiche Ergebnis liefern. Dies muss aber nicht so sein.

2. Lenth (1) hat ein Verfahren vorgeschlagen, das es erlaubt, auf Signifikanz der Effekte zu testen, wenn $n$ gleich 1 ist. Schauen wir uns dieses an.

   Bei einem $2^k$-Plan werden $m = 2^k - 1$ Effekte $E_1, \ldots, E_m$ durch $e_1, \ldots, e_m$ geschätzt. Um zu überprüfen, welche der Effekte signifikant sind, sollte man nach Lenth folgendermaßen vorgehen.

   (a) Bestimme den Median $M_1$ der Absolutbeträge $|e_1|, \ldots, |e_m|$ der geschätzten Effekte $e_1, \ldots, e_m$.

   (b) Berechne

   $$s_0 = 1.5 \cdot M_1 \tag{98}$$

   (c) Bestimme den Median $M_2$ der $|e_i|$, die kleiner als $2.5 \cdot s_0$ sind.

   (d) Berechne

   $$PSE = 1.5 \cdot M_2 \tag{99}$$

   (e) Bilde

   $$ME = t_1 \cdot PSE \tag{100}$$

   Die Werte von $t_1$ in Abhängigkeit von $k$ in Tabelle 35 zu finden.

   (f) Bilde

   $$SME = t_2 \cdot PSE \tag{101}$$

   Die Werte von $t_2$ in Abhängigkeit von $k$ in Tabelle 35 zu finden.

Tabelle 35: Werte von $t_1$ und $t_2$ in Abhängigkeit von $k$

| $k$ | $t_1$ | $t_2$ |
|---|---|---|
| 3 | 2.295 | 4.891 |
| 4 | 2.140 | 4.163 |
| 5 | 2.082 | 4.030 |

Quelle: Montgomery (2), S. 255.

Bei der Überprüfung auf Signifikanz gibt es zwei Möglichkeiten.:

(a) Will man überprüfen, ob ein spezieller Effekt $E_i$ signifikant von 0 verschieden ist, so verwendet man $ME$. Wir lehnen die Hypothese

$$H_o : E_i = 0$$

ab, wenn gilt

$$|e_i| > ME$$

(b) Will man überprüfen, ob **mindestens** ein Effekt von 0 verschieden ist, verwenden wir $SME$. Alle Effekte werden als signifikant angesehen, bei denen der Absolutbetrag des geschätzten Effektes größer als $SME$ ist.

**Beispiel 7 (fortgesetzt)**
Die sortierten Absolutbeträge der Effekte sind

| $e_{ABC}$ | $e_{AC}$ | $e_{BC}$ | $e_C$ | $e_A$ | $e_{AB}$ | $e_B$ |
|---|---|---|---|---|---|---|
| 38.75 | 89.25 | 135.25 | 141.75 | 192.75 | 418.25 | 815.25 |

Somit gilt $M_1 = 141.75$.

Es gilt

$$s_0 = 1.5 \cdot M_1 = 1.5 \cdot 141.75 = 212.625$$

Die sechs kleinsten $|e_i|$ sind kleiner als $2.5 \cdot 212.625 = 531.5625$. Ihr Median ist 138.5. Also gilt $M_2 = 138.5$.

Es gilt

$$PSE = 1.5 \cdot M_2 = 1.5 \cdot 138.5 = 207.75$$

Bilde

$$ME = t_1 \cdot PSE = 2.295 \cdot 207.75 = 476.79$$

Da $|e_B| = 815.25 > 476.79$ gilt, entscheiden wir uns dafür, dass der Effekt von $B$ signifikant ist.
Hätten wir den auf $SME$ beruhenden Test angewendet, so gilt

$$SME = t_2 \cdot PSE = 4.891 \cdot 207.75 = 1016.1$$

In diesem Fall wäre kein Effekt signifikant.

**Beispiel 8**
Markus Hüls und Verena Kleine untersuchen in ihrer Projektarbeit mit dem d2-Test, welche Faktoren einen Einfluss auf die Konzentrationsfähigkeit haben. Beim d2-Test liegt den Teilnehmern ein Testbogen vor, der aus 14 Zeilen mit jeweils 47 Zeichen besteht. In einer Zeile steht eine Folge, die aus den Buchstaben `d` und `p`, über oder unten denen bis zu vier Striche sind. Es müssen alle `d` mit zwei Strichen durchgestrichen werden.
Sie betrachten folgende Faktoren

A : Musik mit den Faktorstufen *ohne* (-) und *mit* (+)

B : Geschlecht mit den Faktorstufen *männlich* (-) und *weiblich* (+)

C : Papierfarbe mit den Faktorstufen *weiß* (-) und *gelb* (+).

Auf jeder Faktorstufenkombination wurde eine Beobachtung gemacht. Die Daten sind in Tabelle 36 zu finden.

Tabelle 36: Daten eines $2^3$-Plans mit $n = 1$

| $A$ | $B$ | $C$ | Anzahl |
|---|---|---|---|
| − | − | − | 369 |
| + | − | − | 407 |
| − | + | − | 443 |
| + | + | − | 463 |
| − | − | + | 359 |
| + | − | + | 484 |
| − | + | + | 397 |
| + | + | + | 515 |

Wir wenden den Algorithmus von Yates an.

| | | | | |
|---|---|---|---|---|
| (1) | 369 | 776 | 1682 | 3437 |
| $a$ | 407 | 906 | 1755 | 301 |
| $b$ | 443 | 843 | 58 | 199 |
| $ab$ | 463 | 912 | 243 | -25 |
| $c$ | 359 | 38 | 130 | 73 |
| $ac$ | 484 | 20 | 69 | 185 |
| $bc$ | 397 | 125 | -18 | -61 |
| $abc$ | 515 | 118 | -7 | 11 |

Somit sind die geschätzten Effekte:

$$\begin{aligned}
e_A &= \frac{301}{4} = 75.25 \\
e_B &= \frac{199}{4} = 49.75 \\
e_C &= \frac{73}{4} = 18.25 \\
e_{AB} &= \frac{-25}{4} = -6.25 \\
e_{AC} &= \frac{185}{4} = 46.25 \\
e_{BC} &= \frac{-61}{4} = -15.25 \\
e_{ABC} &= \frac{11}{4} = 2.75
\end{aligned}$$

Die sortierten Absolutbeträge der Effekte sind

| $e_{ABC}$ | $e_{AB}$ | $e_{BC}$ | $e_C$ | $e_{AC}$ | $e_B$ | $e_A$ |
|---|---|---|---|---|---|---|
| 2.75 | 6.25 | 15.25 | 18.25 | 46.25 | 49.75 | 75.25 |

Somit gilt $M_1 = 18.25$.

Es gilt

$$s_0 = 1.5 \cdot M_1 = 1.5 \cdot 18.25 = 27.375$$

Die sechs kleinsten $|e_i|$ sind kleiner als $2.5 \cdot 27.375 = 68.4375$. Ihr Median ist 16.75. Also gilt $M_2 = 16.75$.

Es gilt

$$PSE = 1.5 \cdot M_2 = 1.5 \cdot 16.75 = 25.125$$

Bilde

$$ME = t_1 \cdot PSE = 2.295 \cdot 25.125 = 57.66$$

Da $|e_A| = 75.25 > 57.66$ gilt, entscheiden wir uns dafür, dass der Effekt von $A$ signifikant ist.

Hätten wir den auf $SME$ beruhenden Test angewendet, so gilt

$$SME = t_2 \cdot PSE = 4.891 \cdot 25.125 = 122.89$$

In diesem Fall wäre kein Effekt signifikant.

# 5 Fraktionelle faktorielle Versuchspläne

Bei einem $2^k$-Plan steigt mit wachsendem $k$ die Anzahl der Versuche exponentiell an. So muss man bei 6 Faktoren schon $2^6 = 64$ und bei bei 10 Faktoren schon $2^{10} = 1024$ Versuche durchführen. Ein fraktioneller Versuchsplan bietet nun die Möglichkeit, mit einer geringen Anzahl von Versuchen die Haupteffekte und einige der Interaktionseffekte zu schätzen. Hierzu muss man aber einige Annahmen machen. Wir wollen an kleinen Beispielen die Grundidee fraktioneller faktorieller Versuchspläne vermitteln.
Wir beginnen mit einem $2^3$-Plan. Tabelle 37 enthält alle Faktoren und Interaktionen zwischen Faktoren.

Tabelle 37: Alle Effekte eines $2^3$-Plans

| | $A$ | $B$ | $C$ | $AB$ | $AC$ | $BC$ | $ABC$ |
|---|---|---|---|---|---|---|---|
| (1) | − | − | − | + | + | + | − |
| $a$ | + | − | − | − | − | + | + |
| $b$ | − | + | − | − | + | − | + |
| $ab$ | + | + | − | + | − | − | − |
| $c$ | − | − | + | + | − | − | + |
| $ac$ | + | − | + | − | + | − | − |
| $bc$ | − | + | + | − | − | + | − |
| $abc$ | + | + | + | + | + | + | + |

Wir wollen nur die Hälfte der Versuche durchführen. Hierbei haben wir sehr viele Möglichkeiten. Wir wählen die Versuche aus, bei denen $ABC$ auf + steht. Es bleiben 4 Versuche übrig, die in Tabelle 38 zu finden sind.

Tabelle 38: Teilplan eines $2^3$-Plans

| | $A$ | $B$ | $C$ | $AB$ | $AC$ | $BC$ | $ABC$ |
|---|---|---|---|---|---|---|---|
| $a$ | + | − | − | − | − | + | + |
| $b$ | − | + | − | − | + | − | + |
| $c$ | − | − | + | + | − | − | + |
| $abc$ | + | + | + | + | + | + | + |

**Beispiel 9**
Jing Zeng und Lars Hartwig untersuchen in ihrer Projektarbeit, welche Faktoren einen Einfluss auf die Zeit haben, die man benötigt, um ein Puzzle zu vollenden.
Sie betrachten folgende Faktoren

A: Geschlecht mit den Faktorstufen *männlich* (-) und *weiblich* (+)
B: Nationalität mit den Faktorstufen *deutsch* (-) und *chinesisch* (+)
C: Vorlage mit den Faktorstufen *kurzfristig* (-) und *permanent* (+)

Gemessen wurde die Zeit in Sekunden. Auf jeder Faktorstufenkombination wurde eine Beobachtung gemacht. Die Daten sind in Tabelle 39 zu finden.

Tabelle 39: Daten eines $2^3$-Plans mit $n = 1$

| | $A$ | $B$ | $C$ | $AB$ | $AC$ | $BC$ | $ABC$ | Anzahl |
|---|---|---|---|---|---|---|---|---|
| (1) | − | − | − | + | + | + | − | 535 |
| $a$ | + | − | − | − | − | + | + | 360 |
| $b$ | − | + | − | − | + | − | + | 758 |
| $ab$ | + | + | − | + | − | − | − | 1497 |
| $c$ | − | − | + | + | − | − | + | 592 |
| $ac$ | + | − | + | − | + | − | − | 316 |
| $bc$ | − | + | + | − | − | + | − | 1163 |
| $abc$ | + | + | + | + | + | + | + | 1646 |

Es sollen nur vier Versuche durchgeführt werden. Wir führen nur die Versuche durch, bei denen $ABC$ auf + steht.
Wir führen also folgende Versuche durch:

1. Eine Deutsche macht das Puzzle mit kurzfristiger Vorlage
2. Ein Chinese macht das Puzzle mit kurzfristiger Vorlage
3. Ein Deutscher macht das Puzzle mit langfristiger Vorlage
4. Eine Chinesin macht das Puzzle mit langfristiger Vorlage

Die Versuche und die Daten stehen in Tabelle 40.

Tabelle 40: Daten des Teilplans eines $2^3$-Plans

| | $A$ | $B$ | $C$ | Anzahl |
|---|---|---|---|---|
| $a$ | + | − | − | 360 |
| $b$ | − | + | − | 758 |
| $c$ | − | − | + | 592 |
| $abc$ | + | + | + | 1646 |

Schauen wir uns die Spalten von Tabelle 38 auf Seite78 genauer an. In 6 Spalten stehen zwei + und zwei −. Wir können also die Effekte der entsprechenden Faktoren bzw. Interaktionen schätzen.
So ist der Schätzer $e_A$ des Effekts von $A$ gleich

$$e_A = \frac{a + abc - b - c}{2}$$

Entsprechend erhalten wir Schätzer für $E_B$

$$e_B = \frac{b + abc - a - c}{2}$$

und $E_C$

$$e_C = \frac{c + abc - a - b}{2} \, .$$

**Beispiel 9 (fortgesetzt)**
Wir erhalten folgende Schätzwerte für die Haupteffekte

$$e_A = \frac{a + abc - b - c}{2} = \frac{360 + 1646 - 758 - 592}{2} = 328$$

Entsprechend erhalten wir Schätzer für $E_B$

$$e_B = \frac{b + abc - a - c}{2} = \frac{758 + 1646 - 360 - 592}{2} = 726$$

und $E_C$

$$e_C = \frac{c + abc - a - b}{2} = \frac{592 + 1646 - 360 - 758}{2} = 560 \, .$$

Wenn wir Tabelle 38 auf Seite 78 genauer anschauen, stellen wir fest, dass jeweils zwei der ersten 6 Spalten identisch sind. So sind die Spalten $A$ und

$BC$ identisch. Mit dem Effekt von $A$ schätzen wir also gleichzeitig auch den Effekt von $BC$. genauer gesagt schätzen wir $A + BC$. Wir sagen auch, dass die beiden Effekte **vermischt** sind. Wir sehen, dass auch die Effekte $B$ und $AC$ und die Effekte $C$ und $AB$ vermischt sind.

Wenn wir annehmen, dass alle Interaktionen zwischen zwei Faktoren vernachlässigt werden können, können wir alle drei Haupteffekte schätzen. Dies ist aber nicht immer der Fall. Hätten wir die Versuche durchgeführt, bei denen in Tabelle 37 $AB$ auf + steht, so hätten wir Tabelle 41 erhalten.

Tabelle 41: Teilplan eines $2^3$-Plans

| | $A$ | $B$ | $C$ | $AB$ | $AC$ | $BC$ | $ABC$ |
|---|---|---|---|---|---|---|---|
| (1) | − | − | − | + | + | + | − |
| $ab$ | + | + | − | + | − | − | − |
| $c$ | − | − | + | + | − | − | + |
| $abc$ | + | + | + | + | + | + | + |

Wir sehen, dass hier $A$ und $B$ vermischt sind. Wir können die Effekte dieser beiden Faktoren nicht getrennt schätzen. Es ist nur möglich den Effekt von $A + B$ zu schätzen. Außerdem sind noch $C$ und $ABC$ und $AC$ und $BC$ vermischt.

Man spricht von einem **fraktionellen faktoriellem Versuchsplan** und bezeichnet diesen als $2^{3-1}$-Plan. Wir werden im Folgenden nur $2^{k-1}$-Pläne betrachten. Diese werden aus einem $2^k$-Plan dadurch gewonnen, dass man bei diesem nur die Hälfte der Versuche durchführt. Für $k > 3$ muss man dabei im Prinzip so wie bei einem $2^{3-1}$-Plan vorgehen. Da man für $k > 3$ aber leicht die Übersicht verliert, benötigt man eine Sprache, mit der man Teilpläne leicht beschreiben kann. Schauen wir uns diese am Beispiel des $2^3$-Plans an.

Wir betrachten die Spalten, die unter den Faktoren stehen, und fassen diese als Vektoren auf, die aus + und − bestehen. So gilt in Tabelle 38 auf Seite 78

$$A = \begin{pmatrix} - \\ + \\ - \\ + \end{pmatrix}$$

Wir definieren die Multiplikation der Vektoren $A$ und $B$ komponentenweise, wobei die aus der Schule bekannten Rechenregeln

$$\begin{aligned} + \cdot + &= + \\ - \cdot + &= - \\ + \cdot - &= - \\ - \cdot - &= + \end{aligned}$$

gelten mögen.
So gilt für

$$A = \begin{pmatrix} - \\ + \\ - \\ + \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} - \\ - \\ + \\ + \end{pmatrix}$$

$$A \cdot B = \begin{pmatrix} + \\ - \\ - \\ + \end{pmatrix}$$

Wir sehen, dass man in Tabelle 37 auf Seite 78 die Vorzeichenstruktur der Interaktionseffekte durch die Multiplikation der Vorzeichenstruktur der jeweiligen Haupteffekte erhält.
Besteht ein Vektor aus lauter +, so bezeichnen wir ihn mit $I$.
Die Multiplikation mit $I$ verändert einen Vektor nicht:

$$AI = IA = A\,.$$

Multiplizieren wir einen Vektor mit sich selber, so gibt es für die einzelnen Komponenten zwei Möglichkeiten. Entweder steht in der Komponente ein +. Dann steht in der entsprechenden Komponente des Produkts wegen $+ \cdot + = +$ auch ein +. Oder in der Komponente steht ein -. Dann steht in der entsprechenden Komponente des Produkts wegen $- \cdot - = +$ ein +. Für $A$ gilt also

$$AA = I$$

Um einen fraktionellen faktoriellen $2^{k-1}$-Plan aus einem $2^k$-Plan zu gewinnen, benötigen wir einen Faktor oder eine Interaktion des $2^k$-Plans, den wir Generator $G$ nennen. Wir wählen aus dem $2^k$-Plan alle Versuche aus, bei denen der Generator entweder auf + oder auf - steht.

**Beispiel 10**
Wir betrachten eine $2^3$-Plan, aus dem wir einen $2^{3-1}$-Plan gewinnen wollen. Wir wählen $ABC$ als Generator $G$.

Da alle Komponenten des Generators entweder auf - oder + stehen, gilt entweder $I = G$ oder $I = -G$.

**Beispiel 10 (fortgesetzt)**
Es gilt $I = ABC$.

Nun können wir durch einfache Algebra feststellen, welche Faktoren und Interaktionen miteinander vermischt sind. Schauen wir uns dies für das Beispiel an.

**Beispiel 10 (fortgesetzt)**
Wir multiplizieren $I = ABC$ mit $A$:

$$A = AABC = BC$$

Wir multiplizieren $I = ABC$ mit $B$:

$$B = ABBC = AC$$

Wir multiplizieren $I = ABC$ mit $C$:

$$C = ABCC = AB$$

Wählen wir hingegen als Generator $I = AB$, so gilt $A = B$, $C = ABC$ und $AC = BC$. Die Haupteffekte $A$ und $B$ sind vermischt. dies ist sicherlich nicht wünschenswert.

Mit Hilfe des Generators ist es möglich, einen $2^{k-1}$-Plan aufzustellen, ohne die Teilversuche aus der Vorzeichenstruktur des $2^k$-Plans auszuwählen.
Schauen wir uns dies am Beispiel eines $2^{3-1}$-Plans mit Generator $I = ABC$ an.
Wir schreiben zunächst die Grundstruktur eines $2^2$-Plans für die Faktoren $A$ und $B$ hin.

Tabelle 42: Erster Schritt bei der Erstellung eines $2^{3-1}$-Plans

| | $A$ | $B$ |
|---|---|---|
| (1) | - | - |
| $a$ | + | - |
| $c$ | - | + |
| $ab$ | + | + |

Nun müssen wir noch den Faktor $C$ und seine Vorzeichenstruktur hinzufügen. Diese erhalten wir aus dem Generator. Multiplizieren wir $I = ABC$ mit $C$, so erhalten wir $C = AB$. Die Vorzeichenstruktur von $C$ stimmt also mit der von $AB$ überein. Wir können also die Spalte für $C$ hinzufügen und wissen genau, welche Versuche wir durchführen müssen.

Tabelle 43: Zweiter Schritt bei der Erstellung eines $2^{3-1}$-Plan

| | $A$ | $B$ | $C$ |
|---|---|---|---|
| $c$ | − | − | + |
| $a$ | + | − | − |
| $b$ | − | + | − |
| $abc$ | + | + | + |

Da wir beim $2^{3-1}$-Plan die Struktur eines $2^2$-Plans gewählt haben, können wir den Algorithmus von Yates anwenden. Wir müssen uns nur genau überlegen, zu welchen Faktoren beziehungsweise Faktorstufenkombinationen die Kontraste gehören. In der zweiten Zeile steht der Kontrast von $A$. Da $A$ und $BC$ vermischt sind, steht hier auch $A + BC$. In der dritten Zeile steht der Kontrast von $B$.Da $B$ und $AC$ vermischt sind, steht hier also auch $B + AC$. Und in der letzten Zeile finden wir den Kontrast von $AB$. Da dieser mit $C$ vermischt, erhalten wir auch $C + AB$. Das folgende Beispiel illustriert den Sachverhalt.

**Beispiel 9 (fortgesetzt)**
Wir schreiben die Beobachtungen in der richtigen Reihenfolge und führen den Algorithmus von Yates durch.

Tabelle 44: Algorithmus von Yates bei einem $2^{3-1}$-Plan

| | | |
|---|---|---|
| 592 | 952 | 3356 |
| 360 | 2404 | 656 |
| 758 | -232 | 1452 |
| 1646 | 888 | 1120 |

Somit gilt

$$e_A = \frac{656}{2} = 328 \qquad e_B = \frac{1452}{2} = 726 \qquad e_A = \frac{1120}{2} = 560$$

Schauen wir uns noch einen $2^{4-1}$-Plan an. Als Generator wählen wir

$$I = ABCD$$

Multiplizieren wir $I = ABCD$ mit $A$, so erhalten wir $A = BCD$.
Multiplizieren wir $I = ABCD$ mit $B$, so erhalten wir $B = ACD$.
Multiplizieren wir $I = ABCD$ mit $C$, so erhalten wir $C = ABD$.
Multiplizieren wir $I = ABCD$ mit $D$, so erhalten wir $D = ABC$.
Multiplizieren wir $A = BCD$ mit $B$, so erhalten wir $AB = CD$.
Multiplizieren wir $A = BCD$ mit $C$, so erhalten wir $AC = BD$.
Multiplizieren wir $A = BCD$ mit $D$, so erhalten wir $AD = BC$.
Also sind die Haupteffekte mit Dreifachinteraktionen und die Zweifachinteraktionen mit Zweifachinteraktionen vermengt.
Das folgende Beispiel zeigt, wie man einen fraktionellen faktoriellen Versuchsplan auswertet.

**Beispiel 10**
Die Studenten Michael Dorin und Stefan Neuhaus ließen in ihrem BI-Projekt 16 Studenten den PRESS-Test durchführen. Bei diesem müssen die Teilnehmer in 3 Minuten möglichst viele Aufgaben vom Typ $3 - 5 + 1$ lösen. Die Zielvariable ist die Anzahl der richtig gelösten Aufgaben. Die Studenten betrachteten die folgenden vier Faktoren

`A` : Tageszeit mit den Faktorstufen *früh* (-) und *spät* (+)

`B` : Musik mit den Faktorstufen *ohne* (-) und *mit* (+)

`C` : Schriftgröße mit den Faktorstufen *klein* (-) und *groß* (+).

`D` : Wasser mit den Faktorstufen *ohne* (-) und *mit* (+).

Bei einem $2^4$-Plan mit $n = 1$ muss man 16 Versuche durchführen. Die Studenten hatten aber nur Zeit für 8 Versuche. Also erstellten sie einen $2^{4-1}$-Plan. Als Generator wählten sie $I = ABCD$. Es gilt also $D = ABC$. Den Teilplan erhalten sie, indem sie sich die Versuche eines $2^3$-plans hinschreiben und eine Spalte mit der Vorzeichenstruktur von $D = ABC$ hinschreiben. Dies ist zusammen mit den Daten in Tabelle 45 zu finden.

Tabelle 45: Teilplan eines eines $2^4$-Planes

| | $A$ | $B$ | $C$ | $D = ABC$ | Anzahl |
|---|---|---|---|---|---|
| (1) | − | − | − | − | 42 |
| $ad$ | + | − | − | + | 39 |
| $bd$ | − | + | − | + | 51 |
| $ab$ | + | + | − | − | 53 |
| $cd$ | − | − | + | + | 48 |
| $ac$ | + | − | + | − | 43 |
| $bc$ | − | + | + | − | 87 |
| $abcd$ | + | + | + | + | 73 |

Wir wenden jetzt den Algorithmus von Yates an. Das Ergebnis kann man Tabelle 46 entnehmen.

Tabelle 46: Der Algorithmus von Yates

| | | | | |
|---|---|---|---|---|
| (1) | 42 | 81 | 185 | 436 |
| $ad$ | 39 | 104 | 251 | -20 |
| $bd$ | 51 | 91 | -1 | 92 |
| $ab$ | 53 | 160 | -19 | -4 |
| $cd$ | 48 | -3 | 23 | 66 |
| $ac$ | 43 | 2 | 69 | -18 |
| $bc$ | 87 | -5 | 5 | 46 |
| $abc$ | 73 | -14 | -9 | -14 |

Wir haben in Tabelle die Symbolik eines $2^3$-Plans gewählt. Somit können wir ganz einfach ablesen, um welche Kontraste es sich handelt. Es gilt

$$\begin{aligned} K_A &= -20 \\ K_B &= 92 \\ K_C &= 66 \\ K_D &= -14 \\ K_{AB} &= -4 \\ K_{AC} &= -18 \end{aligned}$$

$$K_{AD} = 46$$

Wir wissen, dass $A$ und $BCD$, $B$ und $ACD$, $C$ und $ABD$, $D$ und $ABC$, $AB$ und $CD$, $AC$ und $BD$ und $AD$ und $BC$ vermengt sind. Wir können also nur $A+BCD$, $B+ACD$, $C+ABD$, $D+ABC$, $AB+CD$, $AC+BD$ und $AD+BC$ schätzen. Wenn wir unterstellen, dass die Dreifachinteraktionen vernachlässigt werden können, so können wir die Hauptefekte schätzen.
Es gilt

$$\begin{aligned} e_A &= -5 \\ e_B &= 23 \\ e_C &= 1.5 \\ e_D &= -3.5 \end{aligned}$$

Um die Signifikanz der Effekte zu schätzen, können wir entweder unterstellen, dass alle Zweifachinteraktionen vernachlässigt werden können. In diesem Fall ist die Summe deren Quadratsummen gleich $SS_R$. Wir können aber auch das Verfahren von Lenth anwenden. Wir schauen uns nur die ANOVA-Tabelle an. Wir bestimmen zuerst die Quadratsummen. Es gilt

$$\begin{aligned} SS_A &= 50 \\ SS_B &= 1058 \\ SS_C &= 544.5 \\ SS_D &= 24.5 \\ SS_{AB} &= 2 \\ SS_{AC} &= 40.5 \\ SS_{AD} &= 264.5 \end{aligned}$$

da wir unterstellen, dass alle Zweifachinteraktionen vernachlässigt werden können, gilt

$$SS_R = SS_{AB} + SS_{AC} + SS_{AD} = 307$$

Die ANOVA-Tabelle ist in Tabelle 47 zu finden.

Tabelle 47: ANOVA-Tabelle eines $2^{4-1}$-Plans

| Quelle der Variation | Quadrat-summen | Freiheits-grade | Mittlere Qua-dratsummen | $F$ |
|---|---|---|---|---|
| $A$ | 50 | 1 | 50 | 0.489 |
| $B$ | 1058 | 1 | 1058 | 10.342 |
| $C$ | 544.5 | 1 | 544.5 | 5.323 |
| $D$ | 24.5 | 1 | 24.5 | 0.240 |
| Rest | 307 | 3 | 102.3 | |
| Gesamt | 1984 | 7 | | |

Wegen $F_{0.95,1,3} = 10.13$ ist nur der Faktor $B$ signifikant. Somit hat nur die Musik einen Einfluss.

# 6 Faktorielle Versuchspläne in `R`

Schauen wir uns an, wie man einen faktoriellen Versuchsplan in R auswertet. Wir beginnen mit einem Versuchsplan mit zwei Faktoren und schauen uns das Beispiel 5 auf Seite 47 an.
Ein Arbeitnehmer will untersuchen, ob die Strecke A und der Zeitpunkt B der Abfahrt einen Einfluss auf die Fahrzeit zur Arbeit haben. Es werden also die folgenden zwei Faktoren betrachtet:

`A`: Stecke mit den Faktorstufen 1 (-) und 2 (+)
`B`: Zeitpunkt der Abfahrt mit den Faktorstufen früh (-) und spät (+)

Auf jeder Faktorstufenkombination wurde ein Versuch durchgeführt und die Anzahl der Wörter bestimmt, die aufgeschrieben wurden.
Hier sind die Daten:

```
A B Fahrzeit
- -    40
+ -    46
- +    42
+ +    50
```

Wir geben die Daten in der Reihenfolge ein, in der sie in der Tabelle stehen.

```
> Zeit<-c(40,46,42,50)
```

Nun benötigen wir Informationen über die Faktoren $A$ und $B$. Dazu geben wir die Vorzeichenstrukturen der Faktoren als Vektoren `A` und `B` ein und wandeln diese Vektoren mit der Funktion `factor` in Faktoren um.

```
> A<-factor(c(-1,1,-1,1))
> A
 [1] -1 1  -1 1
 Levels: -1 1
> B<-factor(c(-1,-1,1,1))
> B
 [1] -1 -1 1  1
 Levels: -1 1
```

Wir führen die Varianzanalyse mit der Funktion `aov` durch. Das wichtigste Argument von `aov` ist `formula`. Mit diesem Argument können wir das Modell durch eine Formel spezifizieren. Wir wollen die Variable `Zeit` durch ein additives Modell der Faktoren `A` und `B` erklären. Für das Wort erklären verwenden wir das Zeichen . Wir geben also ein.

```
> aov(Zeit~A+B)
Call:
   aov(formula = Zeit ~ A + B)

Terms:
                 A  B Residuals
Sum of Squares  49  9         1
Deg. of Freedom  1  1         1

Residual standard error: 1
Estimated effects may be unbalanced
```

Hier erhalten wir mit den Quadratsummen und den Freiheitsgraden schon sehr viele Informationen. Für die vollständige ANOVA-Tabelle müssen wir das Ergebnis der Funktion `aov` noch einer Variablen zuweisen.

```
> e<-aov(Zeit~A+B)
```

Rufen wir die Funktion `summary` mit dem Argument `e` auf, so erhalten wir die ANOVA-Tabelle:

```
> summary(e)
            Df Sum Sq Mean Sq F value  Pr(>F)
A            1     49      49      49 0.09033 .
B            1      9       9       9 0.20483
Residuals    1      1       1
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Schauen wir uns diese Tabelle genauer an.

`Df` steht für Freiheitsgrade, `Sum Sq` für Quadratsummen, `Mean Sq` für mittlere Quadratsummen und `F value` für den Wert der Teststaistik des $F$-Tests. Außerdem gibt es noch eine Spalte, die mit `Pr(>F)` überschrieben ist. Hier findet man die Überschreitungswahrscheinlichkeit des $F$-Tests. Die Überschreitungswahrscheinlichkeit heißt auch $p$-Wert. Sie ist das kleinste Signifikanzniveau, zu dem man die Nullhypothese ablehnen würde. Führt man den Test also zum Niveau 0.05 durch, so lehnt man die Nullhypothese ab, wenn die Überschreitungswahrscheinlichkeit kleiner als 0.05 ist.

Wir sehen, dass kein Effekt signifikant von 0 verschieden ist.

Man kann dem Vektor `e` aber noch weitere Informationen entnehmen. Die Residuen des Modells erhält man durch

```
> residuals(e)
   1    2    3    4 
 0.5 -0.5 -0.5  0.5 
```

Einen Normal-Quantil-Plot der Residuen gewinnt man durch

```
> qqnorm(residuals(e))
```

Die Gerade fügt man durch

```
> qqline(residuals(e))
```

hinzu.
Um die Effekte zu erhalten, muss man den Algorithmus von Yates durchführen. Hierzu gibt es keine Funktion in R. Also müssen wir einen eigene schreiben, die folgendermaßen aussieht:

```
yates<-function(v,alles=TRUE)
 { # Algorithmus von Yates fuer den Vektor v
   # ist alles gleich TRUE, so werden alle Schritte dokumentiert
   # ist alles gleich FALSE, so werden nur die Kontraste ausgegeben
    n<-length(v)
    k<-log(n,base=2)
    h<-v
    for(i in 1:k)
     { m<-matrix(v,n/2,2,byrow=T)
       v<-c(m[,1]+m[,2],m[,2]-m[,1])
       h<-cbind(h,v)
     }
    if(alles==TRUE) return(h) else return(h[,dim(h)[2]])
 }
```

Die Funktion `yates` besitzt zwei Argumente. Das Argument `v` ist gleich dem Vektor mit den Summen der Werte auf den Faktorstufenkombinationen. Mit dem zweiten Argument `alles` kann man festlegen, ob die Ergebnisse aller Schritte oder nur die Kontraste ausgegeben werden sollen. Standardmäßig werden alle Schritte ausgegeben.

```
> yates(Zeit)
      h  v   v
[1,] 40 86 178
[2,] 46 92  14
[3,] 42  6   6
```

```
[4,] 50  8   2

> yates(Zeit,alles=FALSE)
[1] 178  14   6   2
```

Wollen wir mit dem Ergebnis des Algorithmus von Yates die Effekte schätzen und die Quadratsummen bestimmen, so müssen wir es einer Variablen zuweisen und die erste Komponente entfernen.

```
> ey<-yates(Zeit,alles=FALSE)
> ey<-ey[-1]
> ey
[1] 14  6  2
```

Die geschätzten Effekte erhalten wir nun dadurch, dass wir den Vektor `ey` durch 2 dividieren

```
> ey/2
[1] 7 3 1
```

Für die Quadratsummen müssen wir ihn quadrieren und durch 4 dividieren.

```
> ey^2/4
[1] 49  9  1
```

Für den Interaktionsplot zwischen den Faktoren `A` und `B` verwenden wir die Funktion `interaction.plot`. Diese benötigt drei Argumente.
Das erste Argument ist der Faktor auf der Abszisse, das zweite Argument der zweite Faktor und das dritte Argument der Vektor mit den Daten.
Soll also `A` der Faktor auf der Abszisse sein, so geben wir ein

```
> interaction.plot(A,B,Zeit)
```

Soll hingegen der Faktor `B` auf der Abszisse stehen, so geben wir ein

```
> interaction.plot(B,A,Zeit)
```

Schauen wir uns noch das nichtadditive Modell an. Beim Aufruf verwenden wir wieder das Argument `formula`. Nun müssen wir aber neben dem beiden Faktoren noch den Interaktionseffekt $AB$ zwischen $A$ und $B$ hinzufügen. Diesen beschreiben wir durch das Zeichen :. Wir geben also ein:

```
> e<-aov(Zeit~A+B+A:B)
```

Da $n = 1$ ist, können wir nicht testen. Dies zeigt auch die ANOVA-Tabelle:

```
> summary(e)
            Df Sum Sq Mean Sq
A            1     49      49
B            1      9       9
A:B          1      1       1
```

Bisher war $n = 1$. Schauen wir uns das Beispiel auf Seite 50 an, bei dem $n = 2$ gilt.
Die Studenten Michael Dorin und Stefan Neuhaus ließen in ihrem BI-Projekt Studenten den PRESS-Test durchführen. Bei diesem müssen die Teilnehmer in 3 Minuten möglichst viele Aufgaben vom Typ $3 - 5 + 1$ lösen. Die Zielvariable ist die Anzahl der richtig gelösten Aufgaben. Es wurden zwei Faktoren $A$und $B$ betrachtet.
Der Einflussfaktor $A$ ist die Schriftgröße mit den Faktorstufen 8 pt (`-` und 16 pt `+`.
Der Faktor $B$ steht auf `-`, wenn während des Tests keine Musik lief. Ansonsten steht er auf `+`. Die Daten sind in Tabelle 23 auf Seite 50 zu finden. Hier sind sie noch einmal:

```
 A B   Zeit
 - -  42 39
 + -  51 53
 - +  48 43
 + +  87 73
```

Wir weisen die Daten Spalte für Spalte einem Vektor zu:

```
> Zeit2<-c(42,51,48,87,39,53,43,73)
```

Die Faktoren $A$ und $B$ müssen nun die Vorzeichen erhalten, die zu den entsprechenden Komponenten von `Zeit2` gehören.

```
> A<-factor(c(-1,1,-1,1,-1,1,-1,1))
> A
[1] -1 1  -1 1  -1 1  -1 1
Levels: -1 1
> B<-factor(c(-1,-1,1,1,-1,-1,1,1))
> B
[1] -1 -1 1  1  -1 -1 1  1
Levels: -1 1
```

Nun gehen wir wie bei $n = 1$ vor. Wir schauen uns das nichtadditive Modell an.

```
> e<-aov(Zeit2~A+B+A:B)
```

Die ANOVA-Tabelle 25 auf Seite 51 liefert der Aufruf von `summary`:

```
> summary(e)
            Df  Sum Sq Mean Sq F value   Pr(>F)
A            1 1058.00 1058.00 36.1709 0.003849 **
B            1  544.50  544.50 18.6154 0.012501 *
A:B          1  264.50  264.50  9.0427 0.039663 *
Residuals    4  117.00   29.25
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Den Interaktionsplot aus Abbildung 7 auf Seite 52 erhalten wir durch

```
> interaction.plot(B,A,Zeit2)
```

Und der Residuenplot 8 auf Seite 53 gelingt uns folgendermaßen:

```
> qqnorm(residuals(e))
> qqline(residuals(e))
```

Um den Algorithmus von Yates durchführen zu können, müssen wir die Summen auf den Faktorstufenkombinationen bestimmen. Hierzu erstellen wir mit der Funktion `matrix` die Matrix der Daten, wie in Tabelle 23 auf Seite 50.

```
> m<-matrix(Zeit2,4,2)
> m
     [,1] [,2]
[1,]   42   39
[2,]   51   53
[3,]   48   43
[4,]   87   73
```

Nun müssen wir die Summen der Werte in der einzelnen Zeilen bestimmen. Dies leistet die Funktion `apply`.
Sie wird aufgerufen durch

```
apply(X, MARGIN, FUN)
```

Dabei ist `X` die Matrix und `MARGIN` die Dimension, auf die Funktion `FUN` angewendet werden soll. Die erste Dimension sin die Zeilen und die zweite die Spalten. Wir geben also ein

```
> apply(m,1,sum)
[1]  81 104  91 160
```

Dieses Ergebnis verwenden wir als Argument der Funktion `yates`. Wir können also eingeben

```
> ey<-yates(apply(matrix(Zeit2,4,2),1,sum))
> ey
       h   v   v
[1,]  81 185 436
[2,] 104 251  92
[3,]  91  23  66
[4,] 160  69  46
```

Wir können nun wieder die geschätzten Effekte und die Quadratsummen bestimmen.

```
> ey<-ey[,3][-1]

> ey
[1] 92 66 46

> ey/4
[1] 23.0 16.5 11.5

> ey^2/8
[1] 1058.0  544.5  264.5
```

Wenden wir uns einem Faktoriellen Versuchsplan mit drei Faktoren und $n = 2$ zu. Hierzu schauen wir uns das Beispiel 6 auf Seite 54 an. Hier sollte
Es wurde die Wartezeit (in Sekunden) an der Kasse bei ALDI bestimmt. Dabei wurden die folgenden Faktoren betrachtet.

A : Einkaufstag mit den Faktorstufen *Mittwoch* (-) und *Samstag* (+)

B : Uhrzeit mit den Faktorstufen *Vormittag* (-) und *Mittag* (+)

C : Filiale mit den Faktorstufen *stadtnah* (-) und *außerhalb* (+).

Die Daten sind in Tabelle 26 auf Seite 54 zu finden. Hier sind sie noch einmal: Hier sind die Daten

```
A B C    Zeit
- - -  366 257
+ - -  312 255
- + -  405 453
+ + -  321 317
- - +  456 508
+ - +  322 353
- + +  382 461
+ + +  332 363
```

Wir geben die Daten in R ein.

```
> aldi<-c(366,312,405,321,456,322,382,332,
          257,225,453,317,508,353,461,363)
```

Um die Faktoren zu erzeugen, haben wir mehrere Möglichkeiten.
Wir können sie eingeben. Dabei können wir von der Funktion `rep` Gebrauch machen. Sie hat die Argument `x` und `times`. Dabei ist `x` ein Vektor der Länge `k`. Ist `times` eine Zahl, so wird der Vektor `x` `times`-mal wiederholt.

```
> rep(c(366,312,405),2)
[1] 366 312 405 366 312 405
```

Ist `times` hingegen ein Vektor der Länge `k`, so wird die $i$-te Komponente `times[i]`-mal wiederholt.

```
> rep(c(366,312,405),c(1,2,3))
[1] 366 312 312 405 405 405

Wir geben also ein:
\begin{verbatim}
> A<-factor(rep(rep(c(-1,1),4),2))
> A
 [1] -1 1  -1 1  -1 1  -1 1  -1 1  -1 1  -1 1  -1 1
Levels: -1 1

> B<-factor(rep(rep(c(-1,1),c(2,2)),4))
> B
 [1] -1 -1 1  1  -1 -1 1  1  -1 -1 1  1  -1 -1 1  1
```

```
Levels: -1 1

> C<-factor(rep(rep(c(-1,1),c(4,4)),2))
> C
 [1] -1 -1 -1 -1 1  1  1  1  -1 -1 -1 -1 1  1  1  1
Levels: -1 1
```

Wir können aber auch die Funktion `gen.factorial` im Paket `AlgDesign` verwenden, um die Faktoren zu erzeugen. Dieses Paket muss man zunächst installieren. Hierzu klickt man auf den Schalter

`Packages`

und danach auf

`Install package(s) from CRAN`

Es öffnet sich ein Fenster mit einer Liste, in der man auf `AlgDesign` klickt. Dann wird das Paket installiert. Dazu muss natürlich eine Verbindung zum Internet vorhanden sein. Nachdem man

```
> library(AlgDesign)
```

eingegeben hat, kann man die Funktion `gen.factorial` verwenden.
Diese hat vier Argumente. Das erste Argument `levels` ist die Anzahl der Faktorstufen der Faktoren und das zweite Argument `nVars` die Anzahl der Faktoren. Das Argument `factors` setzt man auf den Wert ällünd mit dem Argument `varNames` kann man den Faktoren Namen geben.
Wir geben ein

```
> m<-gen.factorial(2,3,factors="all",
                   varNames=c("Tag","Uhrzeit","Filiale"))

> m
  Tag Uhrzeit Filiale
1   1       1       1
2   2       1       1
3   1       2       1
4   2       2       1
5   1       1       2
6   2       1       2
7   1       2       2
8   2       2       2
```

Wir sehen, dass die Faktorstufen mit 1 und 2 bezeichnet werden. Wir können auf die Faktoren unter ihren Namen zugreifen, wenn wir

```
> attach(m)
```

eingegeben haben.

```
> Tag
[1] 1 2 1 2 1 2 1 2
Levels: 1 2
> Uhrzeit
[1] 1 1 2 2 1 1 2 2
Levels: 1 2
> Filiale
[1] 1 1 1 1 2 2 2 2
Levels: 1 2
```

Da $n = 2$ ist, müssen wir nun die Matrix `m` verdoppeln. Hierzu benutzen wir die Funktion `rbind`.

```
> m<-rbind(m,m)
> m
   Tag Uhrzeit Filiale
1    1       1       1
2    2       1       1
3    1       2       1
4    2       2       1
5    1       1       2
6    2       1       2
7    1       2       2
8    2       2       2
11   1       1       1
21   2       1       1
31   1       2       1
41   2       2       1
51   1       1       2
61   2       1       2
71   1       2       2
81   2       2       2
> attach(m)
```

Zur Erstellung der ANOVA-Tabelle verwenden wir wie bei einem Versuchsplan mit zwei Faktoren die Funktion `aov` verwenden. Beginnen wir mit dem additiven Modell

```
> e<-aov(aldi~Tag+Uhrzeit+Filiale)
> summary(e)
            Df Sum Sq Mean Sq F value  Pr(>F)
Tag          1  34503   34503 12.1574 0.00449 **
Uhrzeit      1   3452    3452  1.2162 0.29174
Filiale      1  16965   16965  5.9778 0.03088 *
Residuals   12  34056    2838
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Wir sehen, dass die Faktoren `Tag` und `Filiale` signifikant sind. Wir können hier natürlich auch einen Normal-Quantil-Plot der Residuen erstellen.

```
> qqnorm(residuals(e))
> qqline(residuals(e))
```

Schauen wir uns noch das nichtadditive Modell an. Bei diesem müssen wir nicht alle Faktorstufenkombinationen eingeben. Der höchste Interaktionseffekt reicht aus.

```
> e<-aov(aldi~Tag*Uhrzeit*Filiale)
> summary(e)
                    Df Sum Sq Mean Sq F value   Pr(>F)
Tag                  1  34503   34503 16.9148 0.003378 **
Uhrzeit              1   3452    3452  1.6921 0.229540
Filiale              1  16965   16965  8.3170 0.020389 *
Tag:Uhrzeit          1      3       3  0.0015 0.970041
Tag:Filiale          1   1073    1073  0.5258 0.489036
Uhrzeit:Filiale      1  11936   11936  5.8513 0.041919 *
Tag:Uhrzeit:Filiale  1   4727    4727  2.3172 0.166451
Residuals            8  16319    2040
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Wir können natürlich auch den Algorithmus von Yates mit der Funktion `yates` durchführen. Hierzu müssen wir den Vektor `aldi` in eine Matrix transfomieren und die Zeilensummen bilden.

```
> e<-yates(apply(matrix(aldi,8,2),1,sum))
> e
       h    v    v    v
[1,] 623 1160 2656 5833
```

```
[2,] 537 1496 3177 -743
[3,] 858 1639 -306  235
[4,] 638 1538 -437    7
[5,] 964  -86  336  521
[6,] 675 -220 -101 -131
[7,] 843 -289 -134 -437
[8,] 695 -148  141  275
```

Schauen wir uns nun noch ein Beispiel für $n = 1$ an.
Es soll untersucht werden, von welchen Faktoren das Erinnerungsvermögen abhängt. Hierzu wurde Personen eine Liste mit 20 dreisilbigen Wörtern vorgelegt oder vorgelesen. Danach hatten die Personen drei Minuten Zeit, alle Wörter aufzuschreiben, die sie sich gemerkt hatten. Es werden die folgenden drei Faktoren betrachtet:

- A: Geschlecht mit den Faktorstufen männlich (-) und weiblich (+)
- B: Alter mit den Faktorstufen jung (-) und alt (+)
- C: Darbietung mit den Faktorstufen akustisch (-) und gedruckt (+)

Auf jeder Faktorstufenkombination wurde ein Versuch durchgeführt und die Anzahl der Wörter bestimmt, die aufgeschrieben wurden.
Hier sind die Daten

```
A B C Anzahl
- - -    9
+ - -    9
- + -    5
+ + -    7
- - +   10
+ - +   14
- + +   12
+ + +   10
```

Wir geben die Daten in R ein. ¿ Anzahl¡-c(9,9,5,7,10,14,12,10)

```
und erzeugen die Faktoren mit der Funktion {\tt
gen.factorial}:
\begin{verbatim}
> m<-gen.factorial(2,3,factors="all",
                   varName=c("Geschlecht","Alter","Darbietung"))

> m
```

```
  Geschlecht Alter Darbietung
1          1     1          1
2          2     1          1
3          1     2          1
4          2     2          1
5          1     1          2
6          2     1          2
7          1     2          2
8          2     2          2
```

Wir können wie bei einem Versuchsplan mit zwei Faktoren die Funktion aov verwenden. Beginnen wir mit dem additiven Modell

```
> attach(m)

> e<-aov(Anzahl~Geschlecht+Alter+Darbietung)

> summary(e)
            Df Sum Sq Mean Sq F value  Pr(>F)
Geschlecht   1      2       2  0.6667 0.46005
Alter        1      8       8  2.6667 0.17781
Darbietung   1     32      32 10.6667 0.03091 *
Residuals    4     12       3
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Wir sehen, dass der Faktor `Darbietung` signifikant ist.
Unterstellen wir, dass nur der Interaktionsterm zwischen allen drei Faktoren gleich 0 ist, so geben wir ein

```
> e<-aov(Anzahl~Geschlecht*Alter*Darbietung-Geschlecht:Alter:Darbietung)
> summary(e)
                      Df    Sum Sq   Mean Sq   F value Pr(>F)
Geschlecht             1         2         2      0.25 0.7048
Alter                  1         8         8      1.00 0.5000
Darbietung             1        32        32      4.00 0.2952
Geschlecht:Alter       1         2         2      0.25 0.7048
Geschlecht:Darbietung  1 3.058e-32 3.058e-32 3.823e-33 1.0000
Alter:Darbietung       1         2         2      0.25 0.7048
Residuals              1         8         8
```

Im vollständigen Modell können wir nicht testen. Wir erhalten die Quadratsummen durch

```
> e<-aov(Anzahl~Geschlecht*Alter*Darbietung)
> summary(e)
                            Df    Sum Sq   Mean Sq
Geschlecht                   1         2         2
Alter                        1         8         8
Darbietung                   1        32        32
Geschlecht:Alter             1         2         2
Geschlecht:Darbietung        1 3.058e-32 3.058e-32
Alter:Darbietung             1         2         2
Geschlecht:Alter:Darbietung  1         8         8
```

Um die geschätzten Effekte und die Quadratsummen zu erhalten geben wir ein

```
> e<-yates(Anzahl,alles=FALSE)[-1]
> e
[1]  4 -8 -4 16  0  4 -8
```

Nun schauen wir uns das Verfahren von Lenth an. Die folgende Funktion ist nur für einen faktoriellen Versuchsplan mit drei Faktoren gedacht. Sie führt das Verfahren auf Basis des ME durch. Alle Faktoren, deren geschätzte Effekte betragsmäßig größer als ME sind, werden als signifikant angesehen.

```
lenth<-function(e)
 { # das Verfahren von LENTH bei 3 Faktoren mit dem ME
  # e ist der Vektor mit den Effekten
    eff<-c("A","B","AB","C","AC","BC","ABC")
    ea<-abs(e)
    o<-order(ea)
    cat("Die sortierten Effekte\n")
    for (i in 1:7) cat(paste(eff[o][i],ea[o][i],"\n"))
    m1<-median(ea)
    cat(paste("\nM1 =",m1,"\n"))
    s0<-1.5*m1
    cat(paste("s0 =",s0,"\n"))
    m2<-median(ea[ea<2.5*s0])
    cat(paste("M2 =",m2,"\n"))
    PSE<-1.5*m2
    cat(paste("PSE =",PSE,"\n"))
    me<-2.295*PSE
    cat(paste("ME =",me,"\n\n"))
    cat("Signifikante Faktoren:\n")
```

```
    cat(eff[ea>me],"\n")
}
```

Wir bestimmen die geschätzten Effekte und rufen die Funktion `lenth` auf:

```
> lenth(yates(Anzahl,alles=FALSE)[-1]/4)
Die sortierten Effekte
AC 0
A 1
AB 1
BC 1
B 2
ABC 2
C 4

M1 = 1
s0 = 1.5
M2 = 1
PSE = 1.5
ME = 3.4425

Signifikante Faktoren:
C
```

# 7 Beweise und Herleitungen

## 7.1 Herleitung von Gleichung (9) auf Seite 10

Aus

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$$

folgt

$$\sum_{j=1}^{n_i} y_{ij} = n_i \cdot \bar{y}_i \tag{102}$$

Also gilt

$$\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i) = \sum_{j=1}^{n_i} y_{ij} - \sum_{j=1}^{n_i} \bar{y}_i \overset{(102)}{=} n_i \, \bar{y}_i - n_i \, \bar{y}_i = 0\,. \tag{103}$$

Somit gilt

$$\begin{aligned}
SS_T &= \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i + \bar{y}_i - \bar{y})^2 \\
&= \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{I} \sum_{j=1}^{n_i} (\bar{y}_i - \bar{y})^2 + 2 \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)\,(\bar{y}_i - \bar{y}) \\
&= \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{I} n_i\,(\bar{y}_i - \bar{y})^2 + 2 \sum_{i=1}^{I} (\bar{y}_i - \bar{y}) \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i) \\
&\overset{(103)}{=} \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{I} n_i\,(\bar{y}_i - \bar{y})^2 \\
&= SS_A + SS_R\,.
\end{aligned}$$

## 7.2 Herleitung der Schätzer von $\mu_1$ und $\mu_2$ in Gleichung (15)

Wir suchen also die Werte von $\mu_1$ und $\mu_2$, für die

$$\sum_{j=1}^{n} (y_{1j} - \mu_1)^2 + \sum_{j=1}^{n} (y_{2j} - \mu_2)^2$$

minimal wird.
Wir bestimmen die ersten partiellen Ableitungen dieses Ausdrucks. Es gilt

$$\begin{aligned}\frac{\partial}{\partial \mu_1} \left( \sum_{j=1}^{n} (y_{1j} - \mu_1)^2 + \sum_{j=1}^{n} (y_{2j} - \mu_2)^2 \right) &= \sum_{j=1}^{n} \frac{\partial}{\partial \mu_1} (y_{1j} - \mu_1)^2 \\ &= \sum_{j=1}^{n} (-2)(y_{1j} - \mu_1)\end{aligned}$$

Analog erhalten wir

$$\frac{\partial}{\partial \mu_2} \left( \sum_{j=1}^{n} (y_{1j} - \mu_1)^2 + \sum_{j=1}^{n} (y_{2j} - \mu_2)^2 \right) = \sum_{j=1}^{n} (-2)(y_{2j} - \mu_2)$$

Die K-Q-Schätzer $\hat{\mu}_1$ und $\hat{\mu}_2$ von $\mu_1$ und $\mu_2$ müssen also die folgenden Bedingungen erfüllen:

$$\sum_{j=1}^{n} (y_{1j} - \hat{\mu_1}) = 0$$

und

$$\sum_{j=1}^{n} (y_{2j} - \hat{\mu_2}) = 0\,.$$

Hieraus folgt für $i = 1, 2$:

$$\hat{\mu_i} = \bar{y}_i = \frac{1}{n} \sum_{j=1}^{n} y_{ij}\,.$$

Wie man leicht nachprüft, handelt es sich um ein Minimum.

## 7.3 Der Beweis von Gleichung (18) auf Seite 15

Es gilt

$$\bar{y} = \frac{\bar{y}_1 + \bar{y}_2}{2} \tag{104}$$

Dies sieht man folgendermaßen:

$$\begin{aligned}
\bar{y} &= \frac{1}{2n} \sum_{i=1}^{2} \sum_{j=1}^{n} y_{ij} \\
&= \frac{1}{2n} \left( \sum_{j=1}^{n} y_{1j} + \sum_{j=1}^{n} y_{2j} \right) \\
&\stackrel{(102)}{=} \frac{1}{2n} \left( n\,\bar{y}_1 + n\,\bar{y}_2 \right) \\
&= \frac{\bar{y}_1 + \bar{y}_2}{2}
\end{aligned}$$

Also gilt

$$\begin{aligned}
SS_A &= n\,(\bar{y}_1 - \bar{y})^2 + n\,(\bar{y}_2 - \bar{y})^2 \\
&\stackrel{(104)}{=} n \left( \bar{y}_1 - \frac{\bar{y}_1 + \bar{y}_2}{2} \right)^2 + n \left( \bar{y}_2 - \frac{\bar{y}_1 + \bar{y}_2}{2} \right)^2 \\
&= n \left( \frac{\bar{y}_1 - \bar{y}_2}{2} \right)^2 + n \left( \frac{\bar{y}_2 - \bar{y}_1}{2} \right)^2 \\
&= n\,\frac{(\bar{y}_1 - \bar{y}_2)^2}{4} + n\,\frac{(\bar{y}_2 - \bar{y}_1)^2}{4} \\
&\stackrel{(a-b)^2=(b-a)^2}{=} n\,\frac{(\bar{y}_2 - \bar{y}_1)^2}{4} + n\,\frac{(\bar{y}_2 - \bar{y}_1)^2}{4} \\
&= \frac{n\,(\bar{y}_2 - \bar{y}_1)^2}{2}
\end{aligned} \tag{105}$$

### 7.4 Der Beweis von Gleichung (20) auf Seite 16

Es gilt

$$
\begin{aligned}
SS_A &\overset{(105)}{=} \frac{n\left(\bar{y}_2 - \bar{y}_1\right)^2}{2} \\
&= \frac{n\left(\frac{1}{n}\sum\limits_{j=1}^{n} y_{2j} - \frac{1}{n}\sum\limits_{j=1}^{n} y_{1j}\right)^2}{2} \\
&= \frac{n\left[\frac{1}{n}\left(\sum\limits_{j=1}^{n} y_{2j} - \sum\limits_{j=1}^{n} y_{1j}\right)\right]^2}{2} \\
&= \frac{\frac{n}{n^2}\left(\sum\limits_{j=1}^{n} y_{2j} - \sum\limits_{j=1}^{n} y_{1j}\right)^2}{2} \\
&= \frac{\left(\sum\limits_{j=1}^{n} y_{2j} - \sum\limits_{j=1}^{n} y_{1j}\right)^2}{2n}
\end{aligned}
$$

### 7.5 Der Beweis von Gleichung (47) auf Seite 27

Wir setzen in die Gleichung (41) auf Seite 24 für $\mu$, $\alpha_1$, $\alpha_2$, $\beta_1$ und $\beta_2$ die Schätzer aus den Gleichungen (42), (43) und (45) auf Seite 24 und formen um:

$$
\begin{aligned}
SS_R &= \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n}(y_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j)^2 \\
&= \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n}(y_{ijk} - \bar{y} - (\bar{y}_{i\cdot} - \bar{y}) - (\bar{y}_{\cdot j} - \bar{y}))^2 \\
&= \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y})^2
\end{aligned}
$$

## 7.6 Der Beweis von Gleichung (71) auf Seite 37

Wir setzen in die Gleichung (62) auf Seite 34 für $\mu$, $\alpha_i$, $\beta_j$ und $(\alpha\beta)_{ij}$ die Schätzer aus den Gleichungen (63), (64), (65) und (68)auf Seite 34 ein und formen um.

$$\begin{aligned}
SS_R &= \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n}(y_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - \widehat{(\alpha\beta)}_{ij})^2 \\
&= \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n}(y_{ijk} - \bar{y} - (\bar{y}_{i\cdot} - \bar{y}) - (\bar{y}_{\cdot j} - \bar{y}) - (\bar{y}_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}))^2 \\
&= \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n}(y_{ijk} - \bar{y} - \bar{y}_{i\cdot} + \bar{y} - \bar{y}_{\cdot j} + \bar{y} - \bar{y}_{ij} + \bar{y}_{i\cdot} + \bar{y}_{\cdot j} - \bar{y})^2 \\
&= \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij})^2
\end{aligned}$$

## 7.7 Der Beweis von Gleichung (88) auf Seite 43

Es gilt

$$\bar{y} = \frac{\bar{y}_1 + \bar{y}_2}{2}$$

Dies sieht man folgendermaßen:

$$\begin{aligned}
\bar{y} &= \frac{1}{4n} \sum_{i=1}^{2} \sum_{j=1}^{1} \sum_{k=1}^{2} y_{ijk} = \frac{1}{4n} \left( \sum_{j=1}^{1} \sum_{k=1}^{2} y_{1jk} + \sum_{j=1}^{1} \sum_{k=1}^{2} y_{2jk} \right) \\
&= \frac{1}{4n} \left(2n\, \bar{y}_{1\cdot} + 2n\, \bar{y}_{2\cdot}\right) = \frac{\bar{y}_{1\cdot} + \bar{y}_{2\cdot}}{2}
\end{aligned}$$

Also gilt

$$\begin{aligned}
SS_A &= 2n\, (\bar{y}_{1\cdot} - \bar{y})^2 + 2n\, (\bar{y}_{2\cdot} - \bar{y})^2 \\
&= 2n \left( \bar{y}_{1\cdot} - \frac{\bar{y}_1 + \bar{y}_{2\cdot}}{2} \right)^2 + 2n \left( \bar{y}_{12\cdot} - \frac{\bar{y}_1 + \bar{y}_{2\cdot}}{2} \right)^2 \\
&= 2n \left( \frac{\bar{y}_{1\cdot} - \bar{y}_{2\cdot}}{2} \right)^2 + 2n \left( \frac{\bar{y}_{2\cdot} - \bar{y}_{2\cdot}}{2} \right)^2 \\
&= n\, (\bar{y}_{2\cdot} - \bar{y}_{1\cdot})^2
\end{aligned}$$

# 8 Tabellen

Tabelle 48: Das 0.95-Quantil $F_{m,n;0.95}$ der $F$-Verteilung mit $m$ und $n$ Freiheitsgraden

| n \ m | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | 161.45 | 199.5 | 215.71 | 224.58 | 230.16 | 233.99 | 236.77 |
| 2 | 18.51 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.35 |
| 3 | 10.13 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 |
| 4 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 |
| 5 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 |
| 6 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 |
| 7 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 |
| 8 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 |
| 9 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 |
| 10 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 |
| 11 | 4.84 | 3.98 | 3.59 | 3.36 | 3.20 | 3.09 | 3.01 |
| 12 | 4.75 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.91 |
| 13 | 4.67 | 3.81 | 3.41 | 3.18 | 3.03 | 2.92 | 2.83 |
| 14 | 4.60 | 3.74 | 3.34 | 3.11 | 2.96 | 2.85 | 2.76 |
| 15 | 4.54 | 3.68 | 3.29 | 3.06 | 2.90 | 2.79 | 2.71 |
| 16 | 4.49 | 3.63 | 3.24 | 3.01 | 2.85 | 2.74 | 2.66 |
| 17 | 4.45 | 3.59 | 3.20 | 2.96 | 2.81 | 2.70 | 2.61 |
| 18 | 4.41 | 3.55 | 3.16 | 2.93 | 2.77 | 2.66 | 2.58 |
| 19 | 4.38 | 3.52 | 3.13 | 2.90 | 2.74 | 2.63 | 2.54 |
| 20 | 4.35 | 3.49 | 3.10 | 2.87 | 2.71 | 2.60 | 2.51 |
| 21 | 4.32 | 3.47 | 3.07 | 2.84 | 2.68 | 2.57 | 2.49 |
| 22 | 4.30 | 3.44 | 3.05 | 2.82 | 2.66 | 2.55 | 2.46 |
| 23 | 4.28 | 3.42 | 3.03 | 2.80 | 2.64 | 2.53 | 2.44 |
| 24 | 4.26 | 3.40 | 3.01 | 2.78 | 2.62 | 2.51 | 2.42 |
| 25 | 4.24 | 3.39 | 2.99 | 2.76 | 2.60 | 2.49 | 2.40 |
| 26 | 4.23 | 3.37 | 2.98 | 2.74 | 2.59 | 2.47 | 2.39 |
| 27 | 4.21 | 3.35 | 2.96 | 2.73 | 2.57 | 2.46 | 2.37 |
| 28 | 4.20 | 3.34 | 2.95 | 2.71 | 2.56 | 2.45 | 2.36 |
| 29 | 4.18 | 3.33 | 2.93 | 2.70 | 2.55 | 2.43 | 2.35 |
| 30 | 4.17 | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.33 |

# Literatur

[1] Lenth, Russell V. (1989): Quick and Easy Analysis of Unreplicated Factorials Technometrics, S.469-473

[2] Montgomery, D. C. (2001): Design and analysis of experiments. 5. Auflage. New York.