

# Lineare Regression II

## Varianzanalyse als multiple Regression auf Designvariablen

Das lineare Regressionsmodell setzt implizit voraus, dass nicht nur die abhängige, sondern auch die erklärenden Variablen metrisches Messniveau aufweisen.

Es ist jedoch möglich, auch nominalskalierte erklärende Variablen zu betrachten, wenn die Ausprägungen einer nominalskalierten vorher in Designvariablen umgesetzt wird.

Beispiel:

Regression der Bewertung der SPD auf die Konfession mit den Ausprägungen katholisch (1), protestantisch (2), konfessionslos (3).

Zunächst werden aus der Variable Konfession *Designvariablen* gebildet, wobei die Zahl der Designvariablen gleich der Zahl der Ausprägungen minus 1 ist. Bei der sogenannten *Dummy-Kodierung* ergibt sich folgendes Schema:

| Konfession | katholisch | protestantisch | konfessionslos |
|------------|------------|----------------|----------------|
| Dkath      | 1          | 0              | 0              |
| Dprot      | 0          | 1              | 0              |

## Varianzanalyse als multiple Regression auf Designvariablen: Dummy-Kodierung

| Konfession | katholisch | protestantisch | konfessionslos |
|------------|------------|----------------|----------------|
| Dkath      | 1          | 0              | 0              |
| Dprot      | 0          | 1              | 0              |

- Bei katholischen Befragten hat die Designvariable Dkath den Wert 1 und die Designvariable Dprot den Wert 0.
- Umgekehrt ist es bei protestantischen Befragten. Hier hat Dprot den Wert 1 und Dkath den Wert 0.
- Bei konfessionslosen Befragten weisen beiden Designvariablen Dkath und Dprot jeweils den Wert 0 auf.

Die Ausprägung einer erklärenden Variablen, für die keine eigene Designvariable gebildet wird, wird als Referenzkategorie bezeichnet.

In einer anschließenden multiplen Regression wird EvalSPD auf Dkath und Dprot regrediert.

Die folgende Analyse basiert auf einer Auswahl aus allen Fällen, wobei zufällig in jeder der drei Gruppen  $n_k = 70$  ausgewählt wurden.

## Varianzanalyse als multiple Regression auf Designvariablen: Dummy-Kodierung

| Konfession | katholisch | protestantisch | konfessionslos |
|------------|------------|----------------|----------------|
| Dkath      | 1          | 0              | 0              |
| Dprot      | 0          | 1              | 0              |

### Koeffizienten<sup>a,b</sup>

| Modell |             | Nicht standardisierte Koeffizienten |                | Standardisierte Koeffizienten | T      | Signifikanz |
|--------|-------------|-------------------------------------|----------------|-------------------------------|--------|-------------|
|        |             | B                                   | Standardfehler | Beta                          |        |             |
| 1      | (Konstante) | .529                                | .311           |                               | 1.702  | .090        |
|        | Dkat        | -.771                               | .439           | -.133                         | -1.756 | .081        |
|        | Dprot       | 1.386                               | .439           | .239                          | 3.155  | .002        |

a. Abhängige Variable: EvalSPD

b. Ausschließliche Auswahl von Fällen, bei denen group > .00

Die Regressionsfunktion lautet:

$$\hat{Y} = 0.529 - 0.771 \cdot D_{\text{kat}} + 1.386 \cdot D_{\text{prot}}$$

## Varianzanalyse als multiple Regression auf Design-Variablen: Dummy-Kodierung

| Konfession | katholisch | protestantisch | konfessionslos |
|------------|------------|----------------|----------------|
| Dkath      | 1          | 0              | 0              |
| Dprot      | 0          | 1              | 0              |

$$\hat{Y} = 0.529 - 0.771 \cdot D_{\text{kath}} + 1.386 \cdot D_{\text{prot}}$$

In Abhängigkeit von der Konfession gibt es drei unterschiedliche Vorhersagewerte:

| Konfession     | Vorhersagewert  |
|----------------|---|
| Katholisch     | $\hat{y}_{\text{Katholisch}} = b_0 + b_1 \cdot 1 + b_2 \cdot 0 = 0.529 - 0.771 = -0.242$    |
| Protestantisch | $\hat{y}_{\text{Protestantisch}} = b_0 + b_1 \cdot 0 + b_2 \cdot 1 = 0.529 + 1.386 = 1.915$ |
| Konfessionslos | $\hat{y}_{\text{Konfessionslos}} = b_0 + b_1 \cdot 0 + b_2 \cdot 0 = 0.529$                 |

Aus den Vorhersagewerten folgt für die Interpretation der Regressionskoeffizienten:

- Die Regressionskonstante  $b_0$  gibt den Vorhersagewert der abhängigen Variablen wieder, wenn die Referenzkategorie „konfessionslos“ der erklärenden Variablen realisiert wird;
- das Regressionsgewicht  $b_1$  der Designvariable  $D_{\text{kath}}$  gibt die Differenz der Vorhersage (geschätzte Mittelwertdifferenz) von katholischen Befragten zur Referenzkategorie wieder;
- das Regressionsgewicht  $b_2$  der Designvariable  $D_{\text{prot}}$  gibt die Differenz der Vorhersage (geschätzte Mittelwertdifferenz) von protestantischen Befragten zur Referenzkategorie wieder;

## Varianzanalyse als multiple Regression auf Design-Variablen: Effekt-Kodierung

| Konfession | katholisch | protestantisch | konfessionslos |
|------------|------------|----------------|----------------|
| Ekath      | 1          | 0              | -1             |
| Eprot      | 0          | 1              | -1             |

Anstelle der Dummy-Kodierung wird oft die *Effekt-Kodierung* verwendet, bei der die Designvariablen den Wert -1 aufweisen, wenn die Referenzkategorie realisiert wird.

Die Verwendung der Effekt-Kodierung ergibt folgendes Resultat:

Koeffizienten<sup>a,b</sup>

| Modell |             | Nicht standardisierte Koeffizienten |                | Standardisierte Koeffizienten | T      | Signifikanz |
|--------|-------------|-------------------------------------|----------------|-------------------------------|--------|-------------|
|        |             | B                                   | Standardfehler | Beta                          |        |             |
| 1      | (Konstante) | .733                                | .179           |                               | 4.090  | .000        |
|        | Ekath       | -.976                               | .254           | -.292                         | -3.849 | .000        |
|        | Eprot       | 1.181                               | .254           | .353                          | 4.657  | .000        |

a. Abhängige Variable: EvalSPD

b. Ausschließliche Auswahl von Fällen, bei denen group > .00

Die Regressionsfunktion lautet:

$$\hat{Y} = 0.733 - 0.976 \cdot \text{Ekath} + 1.181 \cdot \text{Eprot}$$

## Varianzanalyse als multiple Regression auf Design-Variablen: Effekt-Kodierung

| Konfession | katholisch | protestantisch | konfessionslos |
|------------|------------|----------------|----------------|
| Ekath      | 1          | 0              | -1             |
| Eprot      | 0          | 1              | -1             |

$$\hat{Y} = 0.733 - 0.976 \cdot E_{\text{kath}} + 1.181 \cdot E_{\text{prot}}$$

In Abhängigkeit von der Konfession gibt es wieder drei unterschiedliche Vorhersagewerte:

| Konfession     | Vorhersagewert  |
|----------------|---|
| Katholisch     | $\hat{y}_{\text{Katholisch}} = b_0 + b_1 \cdot 1 + b_2 \cdot 0 = 0.733 - 0.976 = -0.243$                  |
| Protestantisch | $\hat{y}_{\text{Protestantisch}} = b_0 + b_1 \cdot 0 + b_2 \cdot 1 = 0.733 + 1.181 = 1.914$               |
| Konfessionslos | $\hat{y}_{\text{Konfessionslos}} = b_0 + b_1 \cdot (-1) + b_2 \cdot (-1) = 0.733 + 0.976 - 1.181 = 0.528$ |

- Die Vorhersagewerte sind bei der Effekt-Kodierung (bis auf Abweichungen durch Rundungsfehler) identisch mit denen der Dummy-Kodierung.
- Wenn in zwei Regressionsmodellen die Zahl der Regressionskoeffizienten (Modellparameter) gleich ist und beide Modelle zum selben Ergebnis kommen, spricht man von Modellequivalenz.  
Modelle mit Effekt-Kodierung und mit Dummy-Kodierung sind also *equivalent*.

## Varianzanalyse als multiple Regression auf Design-Variablen: Effekt-Kodierung

| Konfession | katholisch | protestantisch | konfessionslos |
|------------|------------|----------------|----------------|
| Ekath      | 1          | 0              | -1             |
| Eprot      | 0          | 1              | -1             |

$$\hat{Y} = 0.733 - 0.976 \cdot \text{Ekath} + 1.181 \cdot \text{Eprot}$$

$$\hat{y}_{\text{Katholisch}} = b_0 + b_1 \cdot 1 + b_2 \cdot 0 = 0.733 - 0.976 = -0.243$$

$$\hat{y}_{\text{Protestantisch}} = b_0 + b_1 \cdot 0 + b_2 \cdot 1 = 0.733 + 1.181 = 1.914$$

$$\hat{y}_{\text{Konfessionslos}} = b_0 + b_1 \cdot (-1) + b_2 \cdot (-1) = 0.733 + 0.976 - 1.181 = 0.528$$

Für die Interpretation werden die Vorhersagewerte mit den empirischen Mittelwerten in der Stichprobe verglichen.

### Bericht

| EvalSPD            | Konfess         |                     |                     |           |
|--------------------|-----------------|---------------------|---------------------|-----------|
|                    | 1.00 katholisch | 2.00 protestantisch | 3.00 konfessionslos | Insgesamt |
| Mittelwert         | -.2429          | 1.9143              | .5286               | .7333     |
| N                  | 70              | 70                  | 70                  | 210       |
| Standardabweichung | 2.83577         | 2.16522             | 2.74375             | 2.73649   |

## Varianzanalyse als multiple Regression auf Design-Variablen: Effekt-Kodierung

$$\hat{Y} = 0.733 - 0.976 \cdot D_{\text{kat}} + 1.181 \cdot D_{\text{prot}}$$

$$\hat{y}_{\text{Katholisch}} = b_0 + b_1 \cdot 1 + b_2 \cdot 0 = 0.733 - 0.976 = -0.243$$

$$\hat{y}_{\text{Protestantisch}} = b_0 + b_1 \cdot 0 + b_2 \cdot 1 = 0.733 + 1.181 = 1.914$$

$$\hat{y}_{\text{Konfessionslos}} = b_0 + b_1 \cdot (-1) + b_2 \cdot (-1) = 0.733 + 0.976 - 1.181 = 0.528$$

### Bericht

EvalSPD

|                    | Konfess         |                     |                     | Insgesamt |
|--------------------|-----------------|---------------------|---------------------|-----------|
|                    | 1.00 katholisch | 2.00 protestantisch | 3.00 konfessionslos |           |
| Mittelwert         | -.2429          | 1.9143              | .5286               | .7333     |
| N                  | 70              | 70                  | 70                  | 210       |
| Standardabweichung | 2.83577         | 2.16522             | 2.74375             | 2.73649   |

- Die Regressionskonstante  $b_0$  gibt den Gesamtmittelwert (engl: *grand mean*) wieder;
- das Regressionsgewicht  $b_1$  der Designvariable  $E_{\text{kath}}$  gibt dann die Abweichung (Mittelwertdifferenz) von katholischen Befragten zum Gesamtmittelwert wieder;
- das Regressionsgewicht  $b_2$  der Designvariable  $E_{\text{prot}}$  gibt dann die Abweichung (Mittelwertdifferenz) von protestantischen Befragten zum Gesamtmittelwert wieder.
- Die negative Summe der beiden Regressionsgewichte ist schließlich die Abweichung von konfessionslosen Befragten zum Gesamtmittelwert.

# Varianzanalyse als multiple Regression auf Design-Variablen

Wenn allerdings die Fallzahlen in den einzelnen Gruppen unterschiedlich ausfallen, ergibt sich folgendes Bild:

**Bericht**

| EvalSPD             |            |      |                    |
|---------------------|------------|------|--------------------|
| Konfess             | Mittelwert | N    | Standardabweichung |
| 1.00 katholisch     | .2640      | 659  | 2.96434            |
| 2.00 protestantisch | 1.2424     | 817  | 2.68803            |
| 3.00 konfessionslos | .6716      | 941  | 2.87867            |
| Insgesamt           | .7534      | 2417 | 2.86471            |

Bei der Dummy-Kodierung bleibt die Interpretation gleich;  
 bei der Effekt-Kodierung gibt die Konstante nicht mehr den Gesamtmittelwert, sondern den Mittelwert der drei Gruppenmittelwerte wieder:  
 $0.726 = (.264 + 1.242 + .672) / 3$

## Dummy-Kodierung: Koeffizienten<sup>a</sup>

| Modell |             | Nicht standardisierte Koeffizienten |                | Standardisierte Koeffizienten | T      | Signifikanz |
|--------|-------------|-------------------------------------|----------------|-------------------------------|--------|-------------|
|        |             | B                                   | Standardfehler | Beta                          |        |             |
| 1      | (Konstante) | .672                                | .093           |                               | 7.255  | .000        |
|        | Dkat        | -.408                               | .144           | -.063                         | -2.826 | .005        |
|        | Dprot       | .571                                | .136           | .094                          | 4.203  | .000        |

a. Abhängige Variable: EvalSPD

| Konfession     | Vorhersagewert      |
|----------------|---------------------|
| Katholisch     | $b_0 + b_1 = 0.264$ |
| Protestantisch | $b_0 + b_2 = 1.243$ |
| Konfessionslos | $b_0 = 0.672$       |

## Effekt-Kodierung: Koeffizienten<sup>a</sup>

| Modell |             | Nicht standardisierte Koeffizienten |                | Standardisierte Koeffizienten | T      | Signifikanz |
|--------|-------------|-------------------------------------|----------------|-------------------------------|--------|-------------|
|        |             | B                                   | Standardfehler | Beta                          |        |             |
| 1      | (Konstante) | .726                                | .058           |                               | 12.435 | .000        |
|        | Ekat        | -.462                               | .087           | -.130                         | -5.339 | .000        |
|        | Eprot       | .516                                | .082           | .153                          | 6.309  | .000        |

a. Abhängige Variable: EvalSPD

| Konfession     | Vorhersagewert            |
|----------------|---------------------------|
| Katholisch     | $b_0 + b_1 = 0.264$       |
| Protestantisch | $b_0 + b_2 = 1.242$       |
| Konfessionslos | $b_0 - b_1 - b_2 = 0.672$ |

## Varianzanalyse als multiple Regression auf Design-Variablen

Wenn anstelle inhaltlicher (metrischer) Variablen ausschließlich Designvariablen als „erklärende“ Variablen (Prädiktoren) verwendet werden, spricht man davon, dass die Prädiktoren eine *Designmatrix* bilden.

Designmatrix bei Dummy-Kodierung

| Konst. | Dkath | Dprot |                      |
|--------|-------|-------|----------------------|
| 1      | 1     | 0     | bei Katholiken       |
| 1      | 1     | 0     |                      |
| ...    | ...   | ...   |                      |
| 1      | 0     | 1     | bei Protestanten     |
| 1      | 0     | 1     |                      |
| ...    | ...   | ...   |                      |
| 1      | 0     | 0     | bei Konfessionslosen |
| 1      | 0     | 0     |                      |
| ...    | ...   | ...   |                      |

Designmatrix bei Effekt-Kodierung

| Konst. | Dkath | Dprot |                      |
|--------|-------|-------|----------------------|
| 1      | 1     | 0     | bei Katholiken       |
| 1      | 1     | 0     |                      |
| ...    | ...   | ...   |                      |
| 1      | 0     | 1     | bei Protestanten     |
| 1      | 0     | 1     |                      |
| ...    | ...   | ...   |                      |
| 1      | -1    | -1    | bei Konfessionslosen |
| 1      | -1    | -1    |                      |
| ...    | ...   | ...   |                      |

Die Regressionsgewichte erfassen empirisch dann immer Kontraste, das sind hier Abweichungen von Mittelwerten.

Die Standardfehler und T-Tests prüfen, ob eine Mittelwertdifferenz signifikant ist.

## Varianzanalyse als multiple Regression auf Design-Variablen

Die Prüfung der Hypothese ob die Ausprägungen der erklärenden nominalskalierten Variablen (hier: Konfession) überhaupt Mittelwertabweichungen generiert, kann wieder über den F-Test gegen das „Konstantenmodell“ erfolgen, bei dem alle Regressionsgewichte der Designvariablen null sind.

ANOVA<sup>b</sup>

| Modell |            | Quadratsumme | df   | Mittel der Quadrate | F      | Signifikanz       |
|--------|------------|--------------|------|---------------------|--------|-------------------|
| 1      | Regression | 359.430      | 2    | 179.715             | 22.285 | .000 <sup>a</sup> |
|        | Residuen   | 19467.605    | 2414 | 8.064               |        |                   |
|        | Gesamt     | 19827.034    | 2416 |                     |        |                   |

a. Einflußvariablen : (Konstante), Dprot, Dkat

b. Abhängige Variable: EvalSPD

ANOVA<sup>b</sup>

| Modell |            | Quadratsumme | df   | Mittel der Quadrate | F      | Signifikanz       |
|--------|------------|--------------|------|---------------------|--------|-------------------|
| 1      | Regression | 359.430      | 2    | 179.715             | 22.285 | .000 <sup>a</sup> |
|        | Residuen   | 19467.605    | 2414 | 8.064               |        |                   |
|        | Gesamt     | 19827.034    | 2416 |                     |        |                   |

a. Einflußvariablen : (Konstante), Eprot, Ekst

b. Abhängige Variable: EvalSPD

Unabhängig von Dummy- oder Effekt-Kodierung führt die Varianzanalyse des gesamten Modells stets zum gleichen Ergebnis. Ursache ist die Äquivalenz der Regressionsmodelle.

Der simultane Test von Mittelwertdifferenzen wird als *Varianzanalyse* bezeichnet.

Nominalskalierte Variablen heißen in der Varianzanalyse *Faktoren*.

Im Regressionsmodell werden anstelle eines Faktors die ihm zugeordneten Designvariablen als Prädiktoren eingesetzt.

## Varianzanalyse mit mehreren Faktoren

In der Varianzanalyse wird oft der simultane Einfluss mehrerer Faktoren gleichzeitig analysiert.

Beispiel:

Neben der Konfession wird im folgenden auch die Region als 2. Faktor eingesetzt. Da es sich um eine dichotome Variable handelt, die bereits die Ausprägungen 0 (für „Westen“) und 1 (für „Osten“) aufweist, kann in der Designmatrix die Region bei Dummy-Kodierung direkt als Prädiktor verwendet werden.

| Konst. | Dkath | Dprot | Region |                                |
|--------|-------|-------|--------|--------------------------------|
| 1      | 1     | 0     | 0      | bei Katholiken im Westen       |
| ...    | ...   | ...   | ...    |                                |
| 1      | 0     | 1     | 0      | bei Protestanten im Westen     |
| ...    | ...   | ...   | ...    |                                |
| 1      | 0     | 0     | 0      | bei Konfessionslosen im Westen |
| ...    | ...   | ...   | ...    |                                |
| 1      | 1     | 0     | 1      | bei Katholiken im Osten        |
| ...    | ...   | ...   | ...    |                                |
| 1      | 0     | 1     | 1      | bei Protestanten im Osten      |
| ...    | ...   | ...   | ...    |                                |
| 1      | 0     | 0     | 1      | bei Konfessionslosen im Osten  |
| ...    | ...   | ...   | ...    |                                |

## Varianzanalyse mit mehreren Faktoren

Bei Effekt-Kodierung wird für die Region eine 1/−1-kodierte Designvariable Ereg eingesetzt, wobei der Wert −1 für die Referenzkategorie der Befragten aus den Westen verwendet wird..

| Konst. | Dkath | Dprot | Region |                                |
|--------|-------|-------|--------|--------------------------------|
| 1      | 1     | 0     | −1     | bei Katholiken im Westen       |
| ...    | ...   | ...   | ...    |                                |
| 1      | 0     | 1     | −1     | bei Protestanten im Westen     |
| ...    | ...   | ...   | ...    |                                |
| 1      | −1    | −1    | −1     | bei Konfessionslosen im Westen |
| ...    | ...   | ...   | ...    |                                |
| 1      | 1     | 0     | 1      | bei Katholiken im Osten        |
| ...    | ...   | ...   | ...    |                                |
| 1      | 0     | 1     | 1      | bei Protestanten im Osten      |
| ...    | ...   | ...   | ...    |                                |
| 1      | −1    | −1    | 1      | bei Konfessionslosen im Osten  |
| ...    | ...   | ...   | ...    |                                |

Bei der Schätzung werden zunächst wieder Modell mit gleichen Fallzahlen ( $n_{km} = 35$ ) bei allen Ausprägungskombinationen der beiden Faktoren geschätzt.

# Varianzanalyse mit mehreren Faktoren: Dummy-Kodierung

## Bericht

Mittelwert

| region Ost/West | EvalSPD         |                     |                     |           |
|-----------------|-----------------|---------------------|---------------------|-----------|
|                 | Konfess         |                     |                     |           |
|                 | 1.00 katholisch | 2.00 protestantisch | 3.00 konfessionslos | Insgesamt |
| 0 West          | -.4286          | 2.6000              | .6286               | .9333     |
| 1 Ost           | -.0571          | 1.2286              | .4286               | .5333     |
| Insgesamt       | -.2429          | 1.9143              | .5286               | .7333     |

## Koeffizienten

| Model |              | nicht standardisierte Koeffizienten |                | Standardisierte Koeffizienten | T      | Signifikanz |
|-------|--------------|-------------------------------------|----------------|-------------------------------|--------|-------------|
|       |              | B                                   | Standardfehler | Beta                          |        |             |
| 1     | (Konstante)  | .729                                | .358           |                               | 2.033  | .043        |
|       | Dkat         | -.771                               | .439           | -.133                         | -1.757 | .080        |
|       | Dprot        | 1.386                               | .439           | .239                          | 3.157  | .002        |
|       | region Ost/W | -.400                               | .358           | -.073                         | -1.116 | .266        |

| Konfession     | Reg. | Vorhersagewert            |
|----------------|------|---------------------------|
| Katholisch     | West | $b_0 + b_1 = -.042$       |
| Protestantisch | West | $b_0 + b_2 = 2.115$       |
| Konfessionslos | West | $b_0 = .729$              |
| Katholisch     | Ost  | $b_0 + b_1 + b_3 = -.442$ |
| Protestantisch | Ost  | $b_0 + b_2 + b_3 = 1.715$ |
| Konfessionslos | Ost  | $b_0 + b_3 = .329$        |

a. Abhängige Variable: EvalSPD

b. Ausschließliche Auswahl von Fällen, bei denen group > .00

Die Regressionskonstante schätzt nun (nicht ganz korrekt) den Mittelwert von Konfessionslosen im Westen.

Die Gewichte schätzen die Abweichungen von dieser Gruppe, wobei das Modell gleiche Effekte zwischen den Regionen bzw. Konfessionsgruppen unterstellt.

# Varianzanalyse mit mehreren Faktoren: Effekt-Kodierung

## Bericht

Mittelwert

| region Ost/West | EvalSPD         |                     |                     |           |
|-----------------|-----------------|---------------------|---------------------|-----------|
|                 | Konfess         |                     |                     |           |
|                 | 1.00 katholisch | 2.00 protestantisch | 3.00 konfessionslos | Insgesamt |
| 0 West          | -.4286          | 2.6000              | .6286               | .9333     |
| 1 Ost           | -.0571          | 1.2286              | .4286               | .5333     |
| Insgesamt       | -.2429          | 1.9143              | .5286               | .7333     |

## Koeffizienten<sup>a</sup>

| Modell        | Nicht standardisierte Koeffizienten | Standardfehler | Standardisierte Koeffizienten | T      | Signifikanz |
|---------------|-------------------------------------|----------------|-------------------------------|--------|-------------|
|               |                                     |                | Beta                          |        |             |
| 1 (Konstante) | .733                                | .179           |                               | 4.092  | .000        |
| Ekat          | -.976                               | .253           | -.292                         | -3.852 | .000        |
| Eprot         | 1.181                               | .253           | .353                          | 4.660  | .000        |
| Ereg          | -.200                               | .179           | -.073                         | -1.116 | .266        |

| Konfession     | Reg. | Vorhersagewert                 |
|----------------|------|--------------------------------|
| Katholisch     | West | $b_0 + b_1 - b_3 = -.043$      |
| Protestantisch | West | $b_0 + b_2 - b_3 = 2.114$      |
| Konfessionslos | West | $b_0 - b_1 - b_2 - b_3 = .728$ |
| Katholisch     | Ost  | $b_0 + b_1 + b_3 = -.443$      |
| Protestantisch | Ost  | $b_0 + b_2 + b_3 = 1.714$      |
| Konfessionslos | Ost  | $b_0 - b_1 - b_2 + b_3 = .328$ |

a. Abhängige Variable: EvalSPD

b. Ausschließliche Auswahl von Fällen, bei denen group > .00

Die Regressionskonstante schätzt wieder (korrekt) den Gesamtmittelwert.

Die Gewichte erfassen die Abweichungen von dieser Gruppe, wobei das Modell (fälschlicherweise) gleiche Effekte zwischen den Regionen bzw. Konfessionsgruppen unterstellt.

## Varianzanalyse mit mehreren Faktoren: Interaktionseffekte bei Dummy-Kodierung

Werden in der Designmatrix zusätzliche Prädiktoren aufgenommen, die die Produkte der den Faktoren zugeordneten Designvariablen sind, können *Interaktionseffekte* (erster Ordnung) geschätzt werden.

Interaktionseffekte modellieren Effekte, die spezifisch für das gleichzeitige Auftreten von Ausprägungskombinationen der Faktoren sind.

| Konst. | Dkath | Dprot | Region | Dkath×Reg. | Dkath×Reg. |                                |
|--------|-------|-------|--------|------------|------------|--------------------------------|
| 1      | 1     | 0     | 0      | 0          | 0          | bei Katholiken im Westen       |
| ...    | ...   | ...   | ...    |            |            |                                |
| 1      | 0     | 1     | 0      | 0          | 0          | bei Protestanten im Westen     |
| ...    | ...   | ...   | ...    |            |            |                                |
| 1      | 0     | 0     | 0      | 0          | 0          | bei Konfessionslosen im Westen |
| ...    | ...   | ...   | ...    |            |            |                                |
| 1      | 1     | 0     | 1      | 1          | 0          | bei Katholiken im Osten        |
| ...    | ...   | ...   | ...    |            |            |                                |
| 1      | 0     | 1     | 1      | 0          | 1          | bei Protestanten im Osten      |
| ...    | ...   | ...   | ...    |            |            |                                |
| 1      | 0     | 0     | 1      | 0          | 0          | bei Konfessionslosen im Osten  |
| ...    | ...   | ...   | ...    |            |            |                                |

# Varianzanalyse mit mehreren Faktoren: Interaktionseffekte bei Dummy-Kodierung

## Bericht

Mittelwert

| region Ost/West | EvalSPD         |                     |                     |           |
|-----------------|-----------------|---------------------|---------------------|-----------|
|                 | Konfess         |                     |                     |           |
|                 | 1.00 katholisch | 2.00 protestantisch | 3.00 konfessionslos | Insgesamt |
| 0 West          | -.4286          | 2.6000              | .6286               | .9333     |
| 1 Ost           | -.0571          | 1.2286              | .4286               | .5333     |
| Insgesamt       | -.2429          | 1.9143              | .5286               | .7333     |

## Koeffizienten<sup>a,b</sup>

| Mode            | nicht standardisierte Koeffizienten |                | standardisierte Koeffiziente | T      | Signifikanz |
|-----------------|-------------------------------------|----------------|------------------------------|--------|-------------|
|                 | B                                   | Standardfehler | Beta                         |        |             |
| 1 (Konstante)   | .629                                | .437           |                              | 1.439  | .152        |
| Dkat            | -1.057                              | .618           | -.183                        | -1.712 | .088        |
| Dprot           | 1.971                               | .618           | .340                         | 3.192  | .002        |
| region Ost/West | -.200                               | .618           | -.037                        | -.324  | .746        |
| DkatO           | .571                                | .873           | .078                         | .654   | .514        |
| DprotO          | -1.171                              | .873           | -.160                        | -1.341 | .181        |

| Konfession     | Reg. | Vorhersagewert                  |
|----------------|------|---------------------------------|
| Katholisch     | West | $b_0 + b_1 = -.428$             |
| Protestantisch | West | $b_0 + b_2 = 2.600$             |
| Konfessionslos | West | $b_0 = .629$                    |
| Katholisch     | Ost  | $b_0 + b_1 + b_3 + b_4 = -.057$ |
| Protestantisch | Ost  | $b_0 + b_2 + b_3 + b_5 = 1.229$ |
| Konfessionslos | Ost  | $b_0 + b_3 = .429$              |

Die beobachteten Gruppenmittelwerte werden exakt geschätzt.

<sup>a</sup>Abhängige Variable: EvalSPD

<sup>b</sup>Ausschließliche Auswahl von Fällen, bei denen group > .

## Varianzanalyse mit mehreren Faktoren: Interaktionseffekte bei Effekt-Kodierung

Ganz analog ist das Vorgehen bei Effektkodierung.

| Konst. | Ekath | Eprot | Ereg | Ekath×Reg. | Eprot×Reg. |                                |
|--------|-------|-------|------|------------|------------|--------------------------------|
| 1      | 1     | 0     | -1   | -1         | 0          | bei Katholiken im Westen       |
| ...    | ...   | ...   | ...  |            |            |                                |
| 1      | 0     | 1     | -1   | 0          | -1         | bei Protestanten im Westen     |
| ...    | ...   | ...   | ...  |            |            |                                |
| 1      | -1    | -1    | -1   | 1          | 1          | bei Konfessionslosen im Westen |
| ...    | ...   | ...   | ...  |            |            |                                |
| 1      | 1     | 0     | 1    | 1          | 0          | bei Katholiken im Osten        |
| ...    | ...   | ...   | ...  |            |            |                                |
| 1      | 0     | 1     | 1    | 0          | 1          | bei Protestanten im Osten      |
| ...    | ...   | ...   | ...  |            |            |                                |
| 1      | -1    | -1    | 1    | -1         | -1         | bei Konfessionslosen im Osten  |
| ...    | ...   | ...   | ...  |            |            |                                |

# Varianzanalyse mit mehreren Faktoren: Interaktionseffekte bei Effekt-Kodierung

## Bericht

Mittelwert

| region Ost/West | EvalSPD         |                     |                     |           |
|-----------------|-----------------|---------------------|---------------------|-----------|
|                 | Konfess         |                     |                     |           |
|                 | 1.00 katholisch | 2.00 protestantisch | 3.00 konfessionslos | Insgesamt |
| 0 West          | -.4286          | 2.6000              | .6286               | .9333     |
| 1 Ost           | -.0571          | 1.2286              | .4286               | .5333     |
| Insgesamt       | -.2429          | 1.9143              | .5286               | .7333     |

## Koeffizienten

| Mode          | nicht standardisierte Koeffizienten |                | standardisierte Koeffizienten | T      | Signifikanz |
|---------------|-------------------------------------|----------------|-------------------------------|--------|-------------|
|               | B                                   | Standardfehler | Beta                          |        |             |
| 1 (Konstante) | .733                                | .178           |                               | 4.113  | .000        |
| Ekat          | -.976                               | .252           | -.292                         | -3.872 | .000        |
| Eprot         | 1.181                               | .252           | .353                          | 4.684  | .000        |
| Ereg          | -.200                               | .178           | -.073                         | -1.122 | .263        |
| EkatO         | .386                                | .252           | .115                          | 1.530  | .128        |
| EprotO        | -.486                               | .252           | -.145                         | -1.926 | .055        |

| Konfession     | Reg. | Vorhersagewert                             |
|----------------|------|--|
| Katholisch     | West | $b_0 + b_1 - b_3 - b_4 = -.428$            |
| Protestantisch | West | $b_0 + b_2 - b_3 - b_5 = 2.600$            |
| Konfessionslos | West | $b_0 - b_1 - b_2 - b_3 + b_4 + b_5 = .629$ |
| Katholisch     | Ost  | $b_0 + b_1 + b_3 + b_4 = -.057$            |
| Protestantisch | Ost  | $b_0 + b_2 + b_3 + b_5 = 1.229$            |
| Konfessionslos | Ost  | $b_0 - b_1 - b_2 + b_3 - b_4 - b_5 = .429$ |

Die beobachteten Gruppenmittelwerte werden wiederum exakt geschätzt.

a. Abhängige Variable: EvalSPD

b. Ausschließliche Auswahl von Fällen, bei denen group > .

## Varianzanalyse mit mehreren Faktoren: Korrelationen der Prädiktoren

Im Unterschied zur Dummy-Kodierung korrelieren bei Effekt-Kodierung und gleichen Fallzahlen in den Subgruppen die Designvariablen verschiedener Faktoren weder untereinander, noch gibt es Korrelationen zwischen Designvariablen für Haupteffekte und Interaktionseffekte.

**Korrelationen<sup>a</sup>**

|                          |         | EvalSPD | Ekat  | Eprot | Ereg  | EkatO | EprotO |
|--------------------------|---------|---------|-------|-------|-------|-------|--------|
| Korrelation nach Pearson | EvalSPD | 1.000   | -.115 | .207  | -.073 | .043  | -.088  |
|                          | Ekat    | -.115   | 1.000 | .500  | .000  | .000  | .000   |
|                          | Eprot   | .207    | .500  | 1.000 | .000  | .000  | .000   |
|                          | Ereg    | -.073   | .000  | .000  | 1.000 | .000  | .000   |
|                          | EkatO   | .043    | .000  | .000  | .000  | 1.000 | .500   |
|                          | EprotO  | -.088   | .000  | .000  | .000  | .500  | 1.000  |

a. Ausschließliche Auswahl von Fällen, bei denen group > .00

**Korrelationen<sup>a</sup>**

|                          |                 | EvalSPD | Dkat  | Dprot | region<br>Ost/West | DkatO | DprotO |
|--------------------------|-----------------|---------|-------|-------|--------------------|-------|--------|
| Korrelation nach Pearson | EvalSPD         | 1.000   | -.253 | .306  | -.073              | -.129 | .081   |
|                          | Dkat            | -.253   | 1.000 | -.500 | .000               | .632  | -.316  |
|                          | Dprot           | .306    | -.500 | 1.000 | .000               | -.316 | .632   |
|                          | region Ost/West | -.073   | .000  | .000  | 1.000              | .447  | .447   |
|                          | DkatO           | -.129   | .632  | -.316 | .447               | 1.000 | -.200  |
|                          | DprotO          | .081    | -.316 | .632  | .447               | -.200 | 1.000  |

a. Ausschließliche Auswahl von Fällen, bei denen group > .00

## ML-Schätzung der Regressionskoeffizienten

In der Regel werden die Regressionskoeffizienten nach der OLS-methode geschätzt.

Anstelle der OLS-Regression können die Regressionskoeffizienten unter der Normalverteilungsannahme auch nach der Maximum-Likelihood-Methode geschätzt werden (ML-Schätzung). Bei dieser Methode werden die Regressionskoeffizienten so festgelegt, dass die tatsächlich in der Stichprobe beobachteten Realisationen die größte Auftretenschance hätten, wenn die zu schätzenden Regressionskoeffizienten mit ihren Schätzungen übereinstimmen würden.

Wenn für die Residuen die Normalverteilungsannahme gilt, berechnet sich die Wahrscheinlichkeitsdichte jedes Residuums nach:

$$f(Y|x) = \frac{1}{\sqrt{2\pi \cdot \sigma_{Y|x}^2}} \cdot e^{-\frac{1}{2} \cdot \frac{(Y - \mu_{Y|x})^2}{\sigma_{Y|x}^2}}$$

## ML-Schätzung der Regressionskoeffizienten

In der für die Schätzung verwendeten Likelihood-Funktion wird die Dichtefunktion als mathematische Funktion der Regressionskoeffizienten aufgefasst.

Da bei voneinander unabhängigen Realisierungen die gemeinsame Dichte aller Realisierungen das Produkt der Dichtefunktion der einzelnen Fälle ist, ergibt sich als zu maximierende Likelihood-Funktion der gesamten Stichprobe:

$$L(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_\zeta^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi} \cdot \hat{\sigma}_\zeta} \cdot \exp\left(-\frac{(y_i - \hat{\beta}_0 - \hat{\beta}_1 \cdot x_i^2)^2}{2 \cdot \hat{\sigma}_\zeta^2}\right) = \max$$

Für die Lösung wird anstelle der Maximierung der Likelihood-Funktion äquivalent die negative Log-Likelihood-Funktion minimiert:

$$-\ln(L(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_\zeta^2)) = -0.5 \cdot n \cdot \ln(2\pi \cdot \hat{\sigma}_\zeta^2) - \sum_{i=1}^n \frac{(y_i - \hat{\beta}_0 - \hat{\beta}_1 \cdot x_i^2)^2}{2 \cdot \hat{\sigma}_\zeta^2} = \min$$

Berechnet man diese Lösung, ergeben sich für die Regressionskoeffizienten die gleichen Schätzfunktionen wie bei der OLS-Lösung.

Beide Methoden führen hier also zum gleichen Ergebnis.

## Eigenschaften der ML-Schätzung

Ein Vorteil der Methode liegt darin, dass ML-Schätzer unter sehr allgemeinen Bedingungen statistisch erwünschte Eigenschaften aufweisen:

- die Schätzer sind konsistent, asymptotisch erwartungstreu und asymptotisch normalverteilt
- die Standardfehler sind verglichen mit den Standardfehlern vieler anderer Schätzmethoden minimal
- die Varianzen und Kovarianzen der Schätzer lassen sich konsistent aus den Stichprobendaten schätzen.

Im Falle des linearen Regressionsmodells kommt hinzu,

- + dass die Schätzer der Regressionskoeffizienten bei beliebiger Stichprobengröße erwartungstreu und exakt normalverteilt sind
- + dass durch Anwendung eines einfachen Korrekturfaktor auch des ML-Schätzer der Residualvarianz bei beliebiger Stichprobengröße erwartungstreu geschätzt werden kann.