

# Lineare Regression

## Einfache Regression

Beispieldatensatz: trinkgeld.sav

Ziel: Vorhersage eines Kriteriums/Regressand Y durch einen Prädiktor/Regressor X.

H0: Y lässt sich nicht durch X erklären, das heißt es besteht kein Zusammenhang.

n = 3

x1 = 10    y1 = 2

x2 = 30    y2 = 3

x3 = 50    y3 = 7

| <b>Aufgenommene/Entfernte Variablen<sup>b</sup></b> |                         |                     |           |
|---|-------------------------|---------------------|-----------|
| Modell  | Aufgenommene Variablen  | Entfernte Variablen | Methode   |
| 1   | Rechnungsbetrag in Euro | .                   | Einschluß |

a. Alle gewünschten Variablen wurden eingegeben.  
b. Abhängige Variable: gezahltes Trinkgeld in Euro

Gibt die Variablen im Modell wieder. Voreingestellte Methode ist Einschluß. Im vorliegenden Modell ist nur ein Modell angegeben ( es gibt ja auch nur einen potentiellen Regressor - sprich abhängige Variable x)

| <b>Modellzusammenfassung</b> |                   |           |                        |                              |
|------------------------------|-------------------|-----------|------------------------|------------------------------|
| Modell                       | R                 | R-Quadrat | Korrigiertes R-Quadrat | Standardfehler des Schätzers |
| 1                            | ,945 <sup>a</sup> | ,893      | ,786                   | 1,22474                      |

a. Einflußvariablen : (Konstante), Rechnungsbetrag in Euro

Bestimmtheitsmaß (Anpassungsgüte, Vorhersageleistung) R-Quadrat von Bedeutung. Hier führt es durch den hohen Wert zur Annahme, dass y durch x beeinflusst wird.

R=r Korrelationskoeffizient nach Bravais-Pearson. Misst Zusammenhang zwischen X und Y.

| ANOVA <sup>b</sup> |                                |              |    |                     |       |                   |
|--------------------|--------------------------------|--------------|----|---------------------|-------|-------------------|
| Modell             |                                | Quadratsumme | df | Mittel der Quadrate | F     | Sig.              |
| 1                  | Regression                     | 12,500       | 1  | 12,500              | 8,333 | ,212 <sup>a</sup> |
|                    | Nicht standardisierte Residuen | 1,500        | 1  | 1,500               |       |                   |
|                    | Gesamt                         | 14,000       | 2  |                     |       |                   |

a. Einflußvariablen : (Konstante), Rechnungsbetrag in Euro  
b. Abhängige Variable: gezahltes Trinkgeld in Euro

Wichtig: F-Test, testet H0, dass alle Koeffizienten der Regressoren den Wert 0 besitzen. Ist eher im multiplen linearen Regressionsmodell von Interesse. Hoher F-Wert: Ablehnung von H0 -> es besteht ein Zusammenhang zwischen X und Y

| Koeffizienten <sup>a</sup> |                         |                                     |                |                               |       |      |
|----------------------------|-------------------------|-------------------------------------|----------------|-------------------------------|-------|------|
| Modell                     |                         | Nicht standardisierte Koeffizienten |                | Standardisierte Koeffizienten | T     | Sig. |
|                            |                         | Regressionskoeffizient B            | Standardfehler | Beta                          |       |      |
| 1                          | (Konstante)             | ,250                                | 1,479          |                               | ,169  | ,893 |
|                            | Rechnungsbetrag in Euro | ,125                                | ,043           | ,945                          | 2,887 | ,212 |

a. Abhängige Variable: gezahltes Trinkgeld in Euro

Man beachte hier die Werte für die Schätzer alpha (0,25) und beta (0,125)  
 alpha = Koeffizient der Konstante  
 beta = Koeffizient von x (hier Rechnungsbetrag)  
 standardisierter Beta-koeffizient (hier 0,945): erklärte Varianz, entspricht r.

## Multiple Regression (Beispiel 1)

Beispieldatensatz: regression.sav

Ziel: Vorhersage von Y durch mehrere X.

H0: Y lässt sich nicht durch X erklären, es gibt keinen Zusammenhang zwischen Y und X<sub>n</sub>

n = 25

Es gibt eine Abhängige Variable Y und x1-x4, also 4 unabhängige Variablen (Regressoren).

Es wurde eine *schrittweise* lineare Regression gewählt.

Schrittweise Selektion derjenigen Regressoren, die einen signifikanten Einfluss haben.

| Modellzusammenfassung                     |                   |           |                        |                              |
|---|-------------------|-----------|------------------------|------------------------------|
| Modell                                    | R                 | R-Quadrat | Korrigiertes R-Quadrat | Standardfehler des Schätzers |
| 1   | ,717 <sup>a</sup> | ,513      | ,492                   | 13,048                       |
| 2   | ,862 <sup>b</sup> | ,743      | ,719                   | 9,704                        |
| a. Einflußvariablen : (Konstante), x2     |                   |           |                        |                              |
| b. Einflußvariablen : (Konstante), x2, x1 |                   |           |                        |                              |

Im ersten Modell wurde nur die UV x2 aufgenommen, im zweiten Modell wurde x1 dann hinzugenommen, was zu einer Erhöhung von R-Quadrat führt

| ANOVA <sup>c</sup>   |                                |              |    |                     |        |                   |
|--|--------------------------------|--------------|----|---------------------|--------|-------------------|
| Modell   |                                | Quadratsumme | df | Mittel der Quadrate | F      | Sig.              |
| 1  | Regression                     | 4131,439     | 1  | 4131,439            | 24,268 | ,000 <sup>a</sup> |
|  | Nicht standardisierte Residuen | 3915,601     | 23 | 170,244             |        |                   |
|  | Gesamt                         | 8047,040     | 24 |                     |        |                   |
| 2  | Regression                     | 5975,415     | 2  | 2987,707            | 31,728 | ,000 <sup>b</sup> |
|  | Nicht standardisierte Residuen | 2071,625     | 22 | 94,165              |        |                   |
|  | Gesamt                         | 8047,040     | 24 |                     |        |                   |
| a. Einflußvariablen : (Konstante), x2<br>b. Einflußvariablen : (Konstante), x2, x1<br>c. Abhängige Variable: y |                                |              |    |                     |        |                   |

F-Wert ist hoch: Ablehnung der Hypothese "Untauglichkeit des Modells" -> Es besteht ein Zusammenhang zwischen X und Y.

Vergleich der Signifikanzwerte der Modelle -> welches Modell ist besser?

P-Wert ist bei SPSS in der Regel voreingestellt auf  $\leq 0,05$ . Das heißt: ist die Signifikanz  $\leq 0,05$  dann ist  $H_0$  zu verwerfen.

| Koeffizienten <sup>a</sup> |             |                                     |                |                               |       |      |
|----------------------------|-------------|-------------------------------------|----------------|-------------------------------|-------|------|
| Modell                     |             | Nicht standardisierte Koeffizienten |                | Standardisierte Koeffizienten | T     | Sig. |
|                            |             | Regressionkoeffizient B             | Standardfehler | Beta                          |       |      |
| 1                          | (Konstante) | 21,383                              | 7,170          |                               | 2,982 | ,007 |
|                            | x2          | 2,170                               | ,441           | ,717                          | 4,926 | ,000 |
| 2                          | (Konstante) | -1,709                              | 7,461          |                               | -,229 | ,821 |
|                            | x2          | 2,747                               | ,353           | ,907                          | 7,791 | ,000 |
|                            | x1          | 3,774                               | ,853           | ,515                          | 4,425 | ,000 |
| a. Abhängige Variable: y   |             |                                     |                |                               |       |      |

Schätzungen der Modellparameter wurde für beide Modelle durchgeführt.

Standardisierter Beta: je höher desto bedeutsamer die Wirkung von X auf Y.

Hoher T-Wert: Ablehnung H0 -> Es besteht Zusammenhang

| Ausgeschlossene Variablen <sup>c</sup> |    |                    |        |      |                       |                         |
|--|----|--------------------|--------|------|-----------------------|-------------------------|
| Modell                                 |    | Beta In            | T      | Sig. | Partielle Korrelation | Kollinearitätsstatistik |
|  |    |                    |        |      |                       | Toleranz                |
| 1                                      | x1 | ,515 <sup>a</sup>  | 4,425  | ,000 | ,686                  | ,863                    |
|  | x3 | -,140 <sup>a</sup> | -,916  | ,369 | -,192                 | ,912                    |
|  | x4 | -,058 <sup>a</sup> | -,391  | ,700 | -,083                 | ,981                    |
| 2                                      | x3 | -,155 <sup>b</sup> | -1,396 | ,177 | -,291                 | ,911                    |
|  | x4 | -,019 <sup>b</sup> | -,173  | ,864 | -,038                 | ,975                    |

a. Einflußvariablen im Modell: (Konstante), x2  
b. Einflußvariablen im Modell: (Konstante), x2, x1  
c. Abhängige Variable: y

Diese Tabelle gibt Aufschluss über die im Modell nicht verwendeten Variablen.

## Multiple Regression (Beispiel 2)

Beispieldatensatz [klinik.sav](#)

Ziel: Es soll überprüft werden, ob ein Regressionsmodell zur Erklärung der Gewichtsabnahme mit den potentiellen erklärenden Variablen bmi, dauer des Aufenthalts und Geschlecht sinnvoll ist.

H0 könnte wie folgt beschrieben werden: Keine der Variablen trägt sinnvoll dazu bei die Gewichtsabnahme zu erklären.

Diesmal wurde die Methode "Einschluß" gewählt.

| Modellzusammenfassung |                   |           |                        |                              |
|-----------------------|-------------------|-----------|------------------------|------------------------------|
| Modell                | R                 | R-Quadrat | Korrigiertes R-Quadrat | Standardfehler des Schätzers |
| 1                     | ,685 <sup>a</sup> | ,469      | ,370                   | 4,41445                      |

a. Einflußvariablen : (Konstante), Geschlecht, bmi, Dauer des Aufenthalts (in Tagen)

R-Quadrat ist im mittleren Bereich, was bedeutet dass ein Teil der Gewichtsabnahme durch die Regressoren erklärt werden könnte.

| ANOVA <sup>b</sup> |                                |              |    |                     |       |                   |
|--------------------|--------------------------------|--------------|----|---------------------|-------|-------------------|
| Modell             |                                | Quadratsumme | df | Mittel der Quadrate | F     | Sig.              |
| 1                  | Regression                     | 275,819      | 3  | 91,940              | 4,718 | ,015 <sup>a</sup> |
|                    | Nicht standardisierte Residuen | 311,798      | 16 | 19,487              |       |                   |
|                    | Gesamt                         | 587,617      | 19 |                     |       |                   |

a. Einflußvariablen : (Konstante), Geschlecht, bmi, Dauer des Auefnthalts (in Tagen)  
b. Abhängige Variable: gewichtsabnahme

Der F-Wert ist niedrig die Signifikanz  $\leq 0,05$ . Das bedeutet dass das vorliegende Modell durchaus zur Erklärung der Gewichtsabnahme sinnvoll sein kann.

| Koeffizienten <sup>a</sup> |                                  |                                     |                |                               |        |      |
|----------------------------|----------------------------------|-------------------------------------|----------------|-------------------------------|--------|------|
| Modell                     |                                  | Nicht standardisierte Koeffizienten |                | Standardisierte Koeffizienten | T      | Sig. |
|                            |                                  | Regressionskoeffizient B            | Standardfehler | Beta                          |        |      |
| 1                          | (Konstante)                      | -18,408                             | 8,584          |                               | -2,145 | ,048 |
|                            | bmi                              | ,388                                | ,181           | ,391                          | 2,142  | ,048 |
|                            | Dauer des Auefnthalts (in Tagen) | ,550                                | ,183           | ,565                          | 3,004  | ,008 |
|                            | Geschlecht                       | ,837                                | 2,036          | ,077                          | ,411   | ,687 |

a. Abhängige Variable: gewichtsabnahme

Betrachtet werden die Signifikanzwerte. Der Wert von Geschlecht ist  $\geq 0,05$ , das heißt, diese Variable könnte eventuell 'aussortiert' werden. Die anderen beiden scheinen signifikanten Einfluss zu haben.