

# Regressionsanalyse (1)

Wilfried Mann,  
Mettmann

Mit Hilfe der Regressionsanalyse wird anhand eines Regressionsmodells die Beziehung zwischen einer abhängigen und einer oder mehrerer unabhängigen Variablen analysiert.

## 1. Allgemeine Definitionen

Die Regressionsanalyse ist eine Kombination aus deskriptiver Statistik und varianzanalytischer Hypothesentests zur Beurteilung der Schätzergebnisse, als Schluss auf die Grundgesamtheit.

### Regressionsmodell

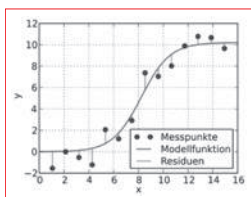
Es dient zur Untersuchung der Art der Beziehungen zwischen einer **abhängigen Variable** (*endogene, erklärte Variable oder Zielgröße bzw. Regressor*) und einer oder mehrerer **unabhängigen Variablen** (*exogene, vorherbestimmte, erklärende Variable oder Einflussgröße bzw. Regressand*), wobei zusätzlich eine zufällige Komponente, der Störterm (*auch Störgröße, Störvariable, Residuum, Rauschen*), in die Modellgleichung eingeht.

### Modellarten

Einfache Regression, lineare Regression, *multiple lineare Regression* und nicht lineare Regression.

### Schätzmethode

Meist wird mit der Kleinstquadrat-Methode (nach Carl Friedrich Gauß) gearbeitet, bei der die Parameter so geschätzt werden, dass die Summe der quadrierten Abweichungen der Regressionskurve von den Datenpunkten minimiert wird. Zur Schätzung der Parameter von Regressions-Modellen werden i.d.R. bestimmte Annahmen über die stochastischen Eigenschaften des additiven Störterms (Residuum) getroffen.



Im Rahmen der deskriptiven Statistik ist dieses Verfahren eine Optimierungsmethode für Messdaten, wobei die Lösung Überbestimmungen erfordert.

Quelle: Wikipedia

### Modellansatz

Einfache lineare Regression (Polynom 1. Ordnung)

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

hierbei ist

$y_i$  = Zielgröße

$\beta_0$  = Achsenabschnitt

$\beta_1$  = Regressionskoeffizient

$x_i$  = unabhängige Variable

$\varepsilon_i$  = Residuum

$i$  = Anzahl der Beobachtungen (1 bis  $n$ )

Multiple lineare Regression (Polynom  $k$ . Ordnung)

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i$$

hierbei ist ergänzend

$k$  = Anzahl der unabhängigen Variablen (1 bis  $k$ )

### Beurteilung der Regressionsergebnisse

Die wichtigsten Kennzahlen sind

- das multiple Bestimmtheitsmaß ( $B$  oder  $R^2$ ) als Anteil der durch die Regression erklärten Varianz für das untersuchte Modell, dazu der
- F-Wert zur Beurteilung der Signifikanz der Regression und
- t-Werte zur Beurteilung der Signifikanz des Einflusses einzelner erklärender Variablen auf die erklärte Variable.

## 2. Bedingungen/Annahmen

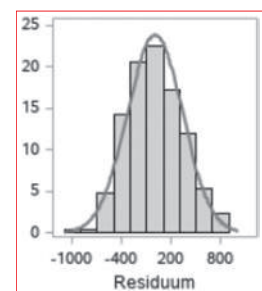
Bei der Durchführung einer Regressionsanalyse sind zunächst folgende Bedingungen einzuhalten:

- Es müssen gleichmäßige, zufällig verteilte, stetige Variablen (verhältnisskaliert, mindestens intervallskaliert) in einem Aussagegebiet vorliegen.
- Es ist ein lineares Modell zu definieren.
- Multikollinearität (Korrelation zwischen unabhängigen Variablen) ist auszuschließen. Hier kann ggf. ein vorgelagerter t-Test auf Korrelation Klarheit schaffen.
- Normalverteilung der Ziel- und Einflussgrößen sollte zur Beurteilung von Hypothesentests vorliegen.

Weiter wird im Analyseverfahren von Annahmen ausgegangen. Es wird zum Beispiel unterstellt, dass die Zielgröße (Regressor) abhängig ist. Ob die Regressanden tatsächlich einen signifikanten Einfluss auf die Zielgröße haben, wird erst im Laufe der Analyse mit Hilfe eines t-Test festgestellt.

Abschließend sind die Residuen während der Analyse zu prüfen auf:

- Exogenität (der Erwartungswert ist »Null«)
- gleichmäßige Streuung (Homoskedastizität, keine Heteroskedastizität)
- zufällig und voneinander unabhängig (keine Autokorrelation)
- Normalverteilung, ggf. sind Ausreißer zu eliminieren. Dieser wichtige Test wird in der Regel optisch durchgeführt.



## 3. Zusammenfassung

Die Durchführung einer Regressionsanalyse ist ein anspruchsvoller iterativer Prozess, der nicht nur statistisches Knowhow erfordert, sondern auch auf fachspezifische Sachkenntnis aufbaut. Zur praktischen Umsetzung wird heute eine Vielzahl von Software-Produkten auf dem Markt angeboten, wobei die »multiple lineare Regressionsanalyse« das etablierte Statistikverfahren auch im Rahmen der Verkehrswertermittlung geworden ist. Eine Vertiefung des Verfahrens in der Anwendung an Kaufpreise, mit Hinweisen zu erreichbaren Zielen, der Datenstruktur sowie zu typischen Kennzahlen, wird im nächsten Glossar, Regressionsanalyse (2), vorgestellt.