# Multiple lineare Regression

## Beispiel 1

In einer Umfrage wurden Körpergrösse, Gewicht und Schuhgrösse von erwachsenen Männern ermittelt:

| Grösse (cm) | Gewicht (kg) | Schuhgrösse (EU) |
|---|---|---|
| 183 | 68 | 45 |
| 171 | 76 | 42 |
| 196 | 93 | 45 |
| 175 | 58 | 38 |

*Ziel:* Ein Modell, das es erlaubt, bei erwachsenen Männern die Schuhgrösse aus ihrer Grösse und ihrem Gewicht zu schätzen.

## Begriffe

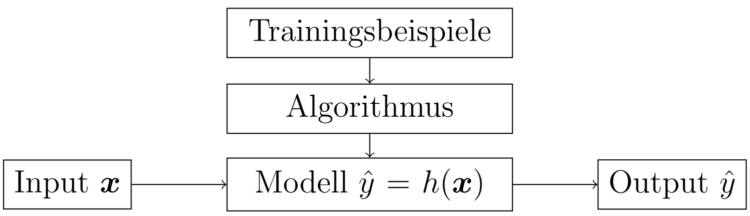
Jede Tabellenzeile stellt ein *Trainingsbeispiel* dar.

Ein Trainingsbeispiel besteht aus den *Inputvariablen Körpergrösse* und *Gewicht* sowie der *Output-* oder *Zielvariable Schuhgrösse.*

Das $i$-te Trainingsbeispiel wird durch das Paar $(\boldsymbol{x}^{(i)}, y^{(i)})$ dargestellt, wobei $\boldsymbol{x}^{(i)} = (x_1^{(i)}, x_2^{(i)})$ der Zeilenvektor mit der Körpergrösse $x_1^{(i)}$, und dem Gewicht $x_2^{(i)}$ des $i$-ten Mannes ist.

Beachte:

- Variablen in fetter Schrift bezeichnen Vektoren.
- Der Index $^{(i)}$ ist kein Exponent sondern eine Variable, mit der man die Trainingsbeispiele durchnummeriert.
- $\left(x^{(2)}, y^{(2)}\right) =$
- $y^{(4)} =$
- $x_1^{(1)} =$
- $x_2^{(3)} =$




Ein Dach über einer Variable bedeutet, dass der jeweilige Wert eine *Schätzung* des „wahren“ Werts darstellt.

Der Input $\boldsymbol{x}$ erhält hier keinen Index, da es sich in der Regel um neue Inputdaten handelt, für die man den jeweiligen Outputwert $\hat{y}$ schätzen möchte.

Bei der multiplen linearen Regression besteht das Modell aus einer Linearkombination der Inputvariablen. Wenn eine Problemstellung $d$ Inputvariablen hat, so gilt

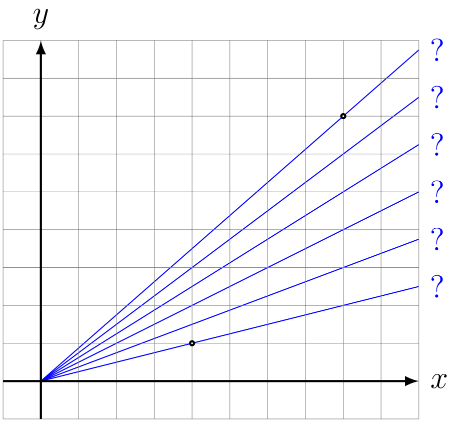
$$\hat{y} = h(\boldsymbol{x}) = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_d x_d = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x} \quad \text{mit} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \ldots \\ \beta_d \end{pmatrix}$$

Die Zahlen $\beta_1$, $\beta_2$, …, $\beta_d$ sind die gesuchten *Modellparameter*.

Um anschaulich zeigen zu können, wie der Vektor $\boldsymbol{\beta}$ bestimmt wird, verwenden wir ein sehr einfaches Beispiel und kehren danach wieder zum Beispiel 1 zurück.

### Beispiel 2

Welche Ursprungsgerade $y = \beta x$ passt am besten durch die Punkte $(8, 7)$ und $(4, 1)$?

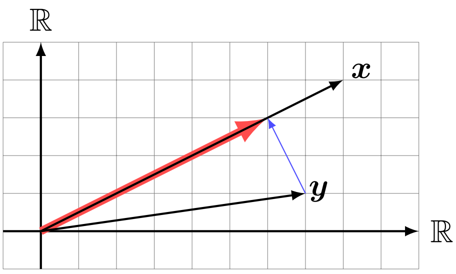



Gesucht wird eine Zahl $\beta$ mit der Eigenschaft

$$\begin{aligned} \beta \cdot 8 &= 7 \\ \beta \cdot 4 &= 1 \end{aligned} \quad \Leftrightarrow \quad \beta \cdot \underbrace{\begin{pmatrix} 8 \\ 4 \end{pmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{pmatrix} 7 \\ 1 \end{pmatrix}}_{\boldsymbol{y}}$$

Dieses „Gleichungssystem“ ist nicht erfüllbar. Wir können aber eine Zahl $\hat{\beta}$ suchen, die eine möglichst gute Näherung für die Lösung darstellt; d. h. für die $\hat{\beta} \cdot \boldsymbol{x}$ möglichst nahe bei $\boldsymbol{y}$ liegt.

Zur Veranschaulichung stellen wir den Vektor $\boldsymbol{x} = \begin{pmatrix} 8 & 4 \end{pmatrix}^{\mathrm{T}}$ mit den $x$-Koordinaten und den Vektor $\boldsymbol{y} = \begin{pmatrix} 7 & 1 \end{pmatrix}^{\mathrm{T}}$ mit den $y$-Koordinaten in einem gemeinsamen Koordinatensystem dar.



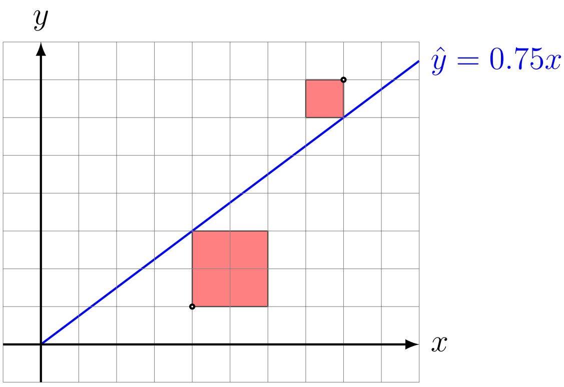

Wir suchen jetzt ein $\hat{\beta}$, für das $\hat{\beta}\boldsymbol{x}$ möglichst nahe bei $\boldsymbol{y}$ liegt.

$\hat{\beta}\boldsymbol{x}$ ist die orthogonale Projektion von $\boldsymbol{y}$ auf $\boldsymbol{x}$.
Den Faktor $\hat{\beta}$ berechnet man bekanntlich mit folgender Formel:

$$\hat{\beta} = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} = \left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\right)^{-1}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$$

Konkret: $\beta = \left[\begin{pmatrix} 8 & 4 \end{pmatrix}\begin{pmatrix} 8 \\ 4 \end{pmatrix}\right]^{-1}\begin{pmatrix} 8 & 4 \end{pmatrix}\begin{pmatrix} 7 \\ 1 \end{pmatrix} = \frac{1}{80} \cdot 60 = \frac{3}{4}$

Im Streudiagramm erhält man folgendes Bild:




**Beispiel 1 (Fortsetzung)**

Modell: $\hat{y} = \beta_1 x_1 + \beta_2 x_2$

Die Werte der Inputvariablen $x_1$ und $x_2$ werden in einer Matrix zusammengefasst:

$$\boldsymbol{X} = \begin{pmatrix} 183 & 68 \\ 171 & 76 \\ 196 & 93 \\ 175 & 58 \end{pmatrix}$$

Ebenso die Werte der Outputvariablen:

$$\boldsymbol{y} = \begin{pmatrix} 45 \\ 42 \\ 45 \\ 38 \end{pmatrix}$$

**Die Matrixform der Projektionsformel**

Um die folgende Formel herleiten zu können, müssten weitere mathematische Voraussetzungen erarbeitet werden, für die uns im Rahmen dieses Kapitels jedoch die Zeit fehlt.

$$\boldsymbol{\beta} = \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}$$

Wie man aber sehen kann, handelt es sich um eine Verallgemeinerung der Formel, die wir zur Lösung von Beispiel 2 verwendet haben.

Damit erhält man für die Daten von Beispiel 1:

$$\boldsymbol{\beta} = \begin{pmatrix} 0.2097 \\ 0.06036 \end{pmatrix} \quad \Rightarrow \quad \hat{y} = 0.2097x_1 + 0.06036x_2$$

**Prognosen**

**Das Bestimmtheitsmass**

Ohne Herleitung:

$$R^2 = \frac{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{X}\boldsymbol{\beta}}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{y}}$$

mit $G = I_n - \frac{1}{n}(J_n J_n^{\mathrm{T}})$

wobei $I_n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$ und $J = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$

Die Grösse $R^2$ drückt aus, wie gut Daten und Modell übereinstimmen.

Für $R^2$ gilt: $0 \leq R^2 \leq 1$

Im Gegensatz zur einfachen linearen Regression gibt es bei der multiplen linearen Regression kein Vorzeichen, das angibt, ob der lineare Zusammenhang fallender oder wachsender Natur ist.

Werte von $R^2$, die nahe bei 1 liegen, bedeuten einen guten Zusammenhang; Werte, die Nahe bei 0 liegen, bedeuten einen schlechten Zusammenhang.

Im Beispiel 1: $R^2 = 0.8267$

Das folgende Programm nimmt uns die mühsame Eingabe der obigen Formeln ab.

Da uns der TI-84+ für Matrizen nur die Variablen [A]–[J] zur Verfügung stellt, können wir nicht die in der Literatur üblichen Bezeichnungen gebrauchen und verwenden statt dessen:

$\boldsymbol{X} \to$ [A], $\boldsymbol{y} \to$ [B], $\boldsymbol{\beta} \to$ [C]

Das Programm setzt also voraus, dass die Werte der Inputvariablen in der Matrix [A] und die Werte der Outputvariablen in der Matrix [B] gespeichert sind. Der Vektor $\boldsymbol{\beta}$ mit den berechnete Modellparametern heisst dann [C].

### Ein Programm für den TI-84+

```
PROGRAM:LSQM
:([A]ᵀ*[A])⁻¹*[A]ᵀ*[B]→[C]
:dim([A])→L₃
:L₃(1)→N
:{N,1}→dim([J])
:Fill(1,[J])
:identity(N)-1/N*[J]*[J]ᵀ→[G]
:[C]ᵀ*[A]ᵀ*[G]*[A]*[C]→[D]
:[B]ᵀ*[G]*[B]→[E]
:[D](1,1)/[E](1,1)→R
:Disp [C]
:Disp "R²:",R
```

### Beispiel 1 (Revisited)

Das Modell $\hat{y} = \beta_1 x_1 + \beta_2 x_2$ kann mit einer Konstanten $\beta_0$ zu

$$\hat{y} = \beta_0 \cdot 1 + \beta_1 x_1 + \beta_2 x_2$$

erweitert werden. In diesem Fall ist

$$\boldsymbol{X} = \begin{pmatrix} 1 & 183 & 68 \\ 1 & 171 & 76 \\ 1 & 196 & 93 \\ 1 & 175 & 58 \end{pmatrix}$$

und damit: $\hat{y} = 15.44x_0 + 0.1087x_1 + 0.09979x_2$

Wegen $R^2 = 0.5583$ bringt die Aufnahme eines zusätzlichen Parameters keine Verbesserung.

**Beispiel 3**

Sophie möchte untersuchen, ob und wie einige ihrer Lebensgewohnheiten ihren Notenerfolg beeinflussen. Deshalb sammelt sie Daten über sich.

| Kaffee (#) | Frühstück (0/1) | Lernzeit (h) | Note |
|---|---|---|---|
| 2 | 0 | 1 | 4.5 |
| 1 | 1 | 2 | 5 |
| 0 | 1 | 3 | 5 |
| 1 | 0 | 2 | 5.5 |
| 2 | 0 | 0 | 3 |

Lineares Modell: $\hat{y} = \beta_0 \cdot 1 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3$

$$\boldsymbol{X} = \begin{pmatrix} 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 2 \\ 1 & 0 & 1 & 3 \\ 1 & 1 & 0 & 2 \\ 1 & 2 & 0 & 0 \end{pmatrix} \to \text{[A]}; \; \boldsymbol{y} = \begin{pmatrix} 4.5 \\ 5 \\ 5 \\ 5.5 \\ 3 \end{pmatrix} \to \text{[B]}$$

Welche Modellparameter erhalten wir jetzt mit der Methode der kleinsten Quadrate?

**Prognosen erstellen**

Welche Note $y$ könnte man bei $x = (2, 0, 2)$ erwarten?

Welche Note $y$ könnte man bei $x = (4, 0, 0)$ erwarten?

**Polynomielle Regression**

Das oben beschriebene lineare Modell kann zusätzlich um Potenzen der unabhängigen Variablen erweitert werden. Die Gleichung

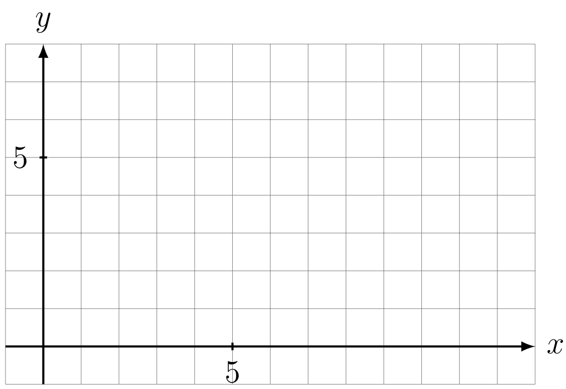
$$\hat{y} = \beta_0 \cdot 1 + \beta_1 \cdot x_1 + \beta_2 \cdot x_1^2$$

versucht, im Sinne der Methode der kleinsten Quadrate, eine Parabel durch die gegeben Punkte $(x^{(i)}, y^{(i)})$ zu legen.

*Achtung:* Auch wenn jetzt Quadrate in der Gleichung auftreten, ist das Modell immer noch *linear*, da es sich um eine *Linearkombination* der unabhängigen Variablen (und allfälliger Potenzen davon) handelt.

**Beispiel 4**

Stelle die Punkte $(2,2)$, $(4,5)$, $(7,6)$, $(10,4)$ in einem Streudigramm dar:




Modell: $y = \beta_0 \cdot 1 + \beta_1 \cdot x + \beta_2 \cdot x^2$

$$\begin{pmatrix} \quad & \quad & \quad \\ \\ \\ \end{pmatrix} \to [A] \quad \text{und} \quad \begin{pmatrix} \quad \\ \\ \\ \end{pmatrix} \to [B]$$

`LSQM`:

**Bemerkung**

Je mehr Terme man in das Modell aufnimmt, desto besser kann sich die Kurve den gegebenen Werten anpassen.

Modell: $y = \beta_0 \cdot 1 + \beta_1 \cdot x + \beta_2 \cdot x^2 + \beta_3 \cdot x^3$

$$\begin{pmatrix} \quad & \quad & \quad & \quad \\ \\ \\ \end{pmatrix} \to [A] \quad \text{und} \quad \begin{pmatrix} \quad \\ \\ \\ \end{pmatrix} \to [B]$$

`LSQM`: ;

## Was schief gehen kann (Teil 1)

| $x_0$ | $x_1$ | $x_2$ | y |
|---|---|---|---|
| 1 | 2 | 4 | 1 |
| 1 | 3 | 6 | 3 |
| 1 | 4 | 8 | 4 |
| 1 | 5 | 10 | 6 |

`LSQM`:

## Was schief gehen kann (Teil 2)

| $x_0$ | $x_1$ | $x_2$ | y |
|---|---|---|---|
| 1 | 2 | 4 | 1 |
| 1 | 3 | 7 | 3 |

`LSQM`: