Allgemeines Lineares Modell

BSc Psychologie, SoSe 2024

Joram Soch

(1) Regression

Methode der kleinsten Quadrate

Einfache lineare Regression

Selbstkontrollfragen

Methode der kleinsten Quadrate

Einfache lineare Regression

Selbstkontrollfragen

Anwendungsszenario

Psychotherapie

Mehr Therapiestunden

⇒ Höhere Wirksamkeit?

Unabhängige Variable

- Anzahl Therapiestunden

Abhängige Variable

- Symptomreduktion

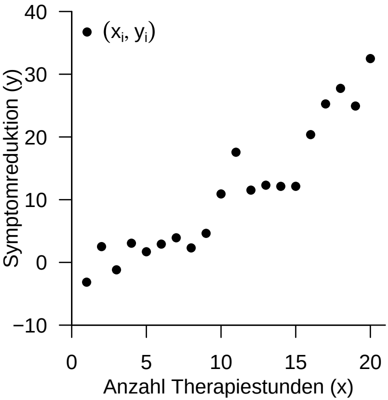
Methode der kleinsten Quadrate

Beispieldatensatz

$i = 1, ..., 20$ Patient:innen, y_i Symptomreduktion bei Patient:in i, x_i Anzahl Therapiestunden von Patient:in i

y_i	x_i
-3.15	1
2.52	2
-1.18	3
3.06	4
1.70	5
2.91	6
3.92	7
2.31	8
4.63	9
10.91	10
17.56	11
11.52	12
12.31	13
12.12	14
12.13	15
20.37	16
25.26	17
27.75	18
24.93	19
32.49	20

Beispieldatensatz



Welcher funktionale Zusammenhang zwischen x und y liegt den Daten zugrunde?

Definition (Gerade)

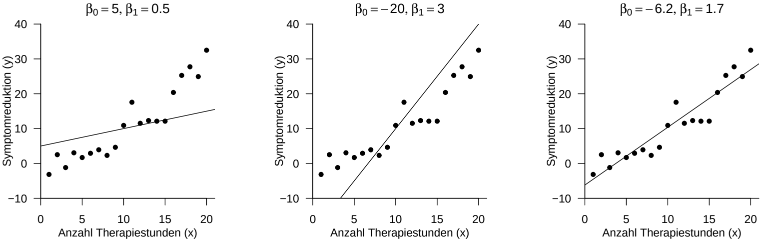
Sei $\beta := (\beta_0, \beta_1)^{\mathrm{T}} \in \mathbb{R}^2$ ein zweidimensionaler Vektor reeller Zahlen. Dann ist eine Gerade durch folgende linear-affine Funktion gegeben:

$$f_\beta : \mathbb{R} \to \mathbb{R}, x \mapsto f_\beta(x) := \beta_0 + \beta_1 x\ . \tag{1}$$

Bemerkungen

- β_0 bestimmt den Schnittpunkt der Gerade mit der y-Achse.
- β_0 wird auch als *Offset-Parameter* oder *intercept* bezeichnet.
- β_1 gibt das Ausmaß der y-Einheitsdifferenz pro x-Einheitsdifferenz wieder.
- β_1 wird auch als *Steigungsparameter* oder *slope* bezeichnet.

Linear-affine Funktionen: $f_\beta(x) := \beta_0 + \beta_1 x$



Definition (Ausgleichsgerade)

$\{(x_1, y_1), ..., (x_n, y_n)\} \subset \mathbb{R}^2$ sei ein Datensatz. Weiterhin sei $q(\beta)$ die Summe der quadrierten vertikalen Abweichungen der y_i von den Funktionswerten $f_\beta(x_i)$:

$$q : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}, \beta \mapsto q(\beta) := \sum_{i=1}^{n}(y_i - f_\beta(x_i))^2 = \sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i))^2 \ . \tag{2}$$

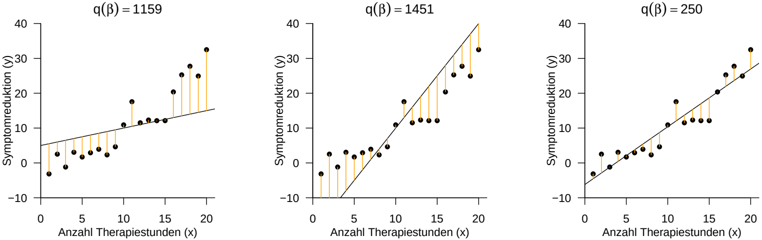
Dann heißt die Gerade f_β, für die die Funktion q ihr Minimum annimt, die *Ausgleichsgerade für den Datensatz* $\{(x_1, y_1), ..., (x_n, y_n)\}$.

Bemerkungen

- Wir nehmen hier ohne Beweis an, dass das Minimum von q eindeutig ist.

Funktion der quadrierten vertikalen Abweichungen:

$$q(\beta) := \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2 \tag{3}$$



$y_i - (\beta_0 + \beta_1 x_i)$ für $i = 1, ..., n$

Animation zur Ausgleichsgerade

Methode der kleinsten Quadrate

Theorem (Ausgleichsgeradenparameter)

Für einen Datensatz $\{(x_1, y_1), ..., (x_n, y_n)\} \subset \mathbb{R}^2$ hat die Ausgleichsgerade die Form

$$f_\beta : \mathbb{R} \to \mathbb{R}, x \mapsto f_\beta(x) := \hat{\beta}_0 + \hat{\beta}_1 x \,, \tag{4}$$

mit den Ausgleichsgeradenparametern

$$\hat{\beta}_1 = \frac{c_{xy}}{s_x^2} \text{ und } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \,, \tag{5}$$

wobei c_{xy} die Stichprobenkovarianz der (x_i, y_i)-Werte, s_x^2 die Stichprobenvarianz der x_i-Werte und $\bar{x}$ und $\bar{y}$ die Stichprobenmitteln der x_i- bzw. y_i-Werte sind.

Bemerkungen

- Mit den Definitionen von c_{xy} und s_x^2 gilt also:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \,. \tag{6}$$

- Man spricht hier von der Stichprobenkovarianz c_{xy}, auch wenn die Werte $x_1, ..., x_n$ oft nicht als Realisierungen einer Stichprobe $\xi_1, ..., \xi_n$ verstanden werden, sondern als gegebene Zahlen.

Methode der kleinsten Quadrate

Beweis

Wir betrachten die Summe der quadrierten vertikalen Abweichungen der y_i von den Funktionswerten $f(x_i)$ als Funktion von β_0 und β_1 und bestimmen Werte $\hat{\beta}_0$ und $\hat{\beta}_1$, für die diese Funktion ihr Minimum annimmt, die Summe der quadrierten vertikalen Abweichungen der y_i von den Funktionswerten $f(x_i)$ also minimal ist. Wir betrachten also die Funktion

$$q : \mathbb{R}^2 \to \mathbb{R}, (\beta_0, \beta_1) \mapsto q(\beta_0, \beta_1) := \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 . \tag{7}$$

Um das Minimum dieser Funktion zu bestimmen, berechnen wir zunächst die partiellen Ableitungen hinsichtlich β_0 und β_1 und setzen diese gleich 0. Es ergibt sich zunächst

$$\begin{aligned}
\frac{\partial}{\partial \beta_0} q(\beta_0, \beta_1) &= \frac{\partial}{\partial \beta_0} \left(\sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 \right) \\
&= \sum_{i=1}^{n} \frac{\partial}{\partial \beta_0} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 \\
&= \sum_{i=1}^{n} 2 \left(y_i - (\beta_0 + \beta_1 x_i)\right) \frac{\partial}{\partial \beta_0} \left(y_i - \beta_0 - \beta_1 x_i\right) \\
&= -2 \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i\right)
\end{aligned} \tag{8}$$

Methode der kleinsten Quadrate

Beweis (fortgeführt)

Weiterhin ergibt sich

$$
\begin{aligned}
\frac{\partial}{\partial \beta_1} q(\beta_0, \beta_1) &= \frac{\partial}{\partial \beta_1} \left(\sum_{i=1}^n \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 \right) \\
&= \sum_{i=1}^n \frac{\partial}{\partial \beta_1} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 \\
&= \sum_{i=1}^n 2\left(y_i - (\beta_0 + \beta_1 x_i)\right) \frac{\partial}{\partial \beta_1} \left(y_i - \beta_0 - \beta_1 x_i\right) \\
&= -2 \sum_{i=1}^n \left(y_i - \beta_0 - \beta_1 x_i\right) x_i
\end{aligned}
\tag{9}
$$

Nullsetzen beider partieller Ableitungen ergibt dann

$$
\begin{aligned}
& \frac{\partial}{\partial \beta_0} q(\beta_0, \beta_1) = 0 \text{ und } \frac{\partial}{\partial \beta_1} q(\beta_0, \beta_1) = 0 \\
\Leftrightarrow & -2 \sum_{i=1}^n \left(y_i - \beta_0 - \beta_1 x_i\right) = 0 \text{ und } -2 \sum_{i=1}^n \left(y_i - \beta_0 - \beta_1 x_i\right) x_i = 0 \\
\Leftrightarrow & \sum_{i=1}^n \left(y_i - \beta_0 - \beta_1 x_i\right) = 0 \text{ und } \sum_{i=1}^n \left(y_i - \beta_0 - \beta_1 x_i\right) x_i = 0
\end{aligned}
\tag{10}
$$

Beweis (fortgeführt)

und weiter

$$
\begin{aligned}
&\sum_{i=1}^n y_i - \sum_{i=1}^n \beta_0 - \beta_1 \sum_{i=1}^n x_i = 0 \text{ und } \sum_{i=1}^n y_i x_i - \sum_{i=1}^n \beta_0 x_i - \beta_1 \sum_{i=1}^n x_i^2 = 0 \\
&\Leftrightarrow \beta_0 n + \beta_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \text{ und } \beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i
\end{aligned}
\tag{11}
$$

Das sich hier ergebende Gleichungssystem

$$
\begin{aligned}
\beta_0 n + \beta_1 \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i \\
\beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n y_i x_i
\end{aligned}
\tag{12}
$$

wird *System der Normalengleichungen* genannt und beschreibt die notwendige Bedingung für ein Minimum von q. Auflösen dieses Gleichungssystems nach β_0 und β_1 liefert dann die Werte $\hat{\beta}_0$ und $\hat{\beta}_1$ des Theorems.

Beweis (fortgeführt)

Um dies zu sehen, halten wir zunächst fest, dass mit der ersten Gleichung des Systems der Normalengleichungen gilt

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \Leftrightarrow \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y} \Leftrightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{13}$$

Einsetzen der Form von $\hat{\beta}_0$ in die zweite Gleichung des Systems der Normalengleichungen ergibt dann zunächst

$$\begin{aligned}
\hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n y_i x_i \\
\Leftrightarrow \left(\bar{y} - \hat{\beta}_1 \bar{x}\right) \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n y_i x_i \\
\Leftrightarrow \bar{y} \sum_{i=1}^n x_i - \hat{\beta}_1 \bar{x} \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n y_i x_i \\
\Leftrightarrow -\hat{\beta}_1 \bar{x} \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n x_i \\
\Leftrightarrow \hat{\beta}_1 \left(\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i\right) &= \sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n x_i
\end{aligned} \tag{14}$$

Beweis (fortgeführt)

Wir halten nun zunächst fest, dass gilt

$$
\begin{aligned}
\sum_{i=1}^n x_i^2 - \bar{x}\sum_{i=1}^n x_i &= \sum_{i=1}^n x_i^2 - 2\bar{x}\sum_{i=1}^n x_i + \bar{x}\sum_{i=1}^n x_i \\
&= \sum_{i=1}^n x_i^2 - 2\bar{x}\sum_{i=1}^n x_i + n\left(\frac{1}{n}\sum_{i=1}^n x_i\right)\bar{x} \\
&= \sum_{i=1}^n x_i^2 - 2\bar{x}\sum_{i=1}^n x_i + n\bar{x}^2 \\
&= \sum_{i=1}^n \left(x_i^2 - 2\bar{x}x_i + \bar{x}^2\right) \\
&= \sum_{i=1}^n \left(x_i - \bar{x}\right)^2
\end{aligned}
\tag{15}
$$

Beweis (fortgeführt)

Weiterhin halten wir zunächst fest, dass gilt

$$
\begin{aligned}
\sum_{i=1}^n y_i x_i - \bar{y}\sum_{i=1}^n x_i & = \sum_{i=1}^n y_i x_i - \bar{y}\sum_{i=1}^n x_i - n\bar{y}\bar{x} + n\bar{y}\bar{x} \\
& = \sum_{i=1}^n y_i x_i - \bar{y}\sum_{i=1}^n x_i - \sum_{i=1}^n y_i\bar{x} + \sum_{i=1}^n \bar{y}\bar{x} \\
& = \sum_{i=1}^n y_i x_i - \sum_{i=1}^n y_i\bar{x} - \sum_{i=1}^n \bar{y}x_i + \sum_{i=1}^n \bar{y}\bar{x} \\
& = \sum_{i=1}^n \left(y_i x_i - y_i\bar{x} - \bar{y}x_i + \bar{y}\bar{x}\right) \\
& = \sum_{i=1}^n \left(y_i - \bar{y}\right)\left(x_i - \bar{x}\right).
\end{aligned} \tag{16}
$$

Methode der kleinsten Quadrate

Beweis (fortgeführt)

In der Fortsetzung von (14) ergibt sich dann

$$
\begin{aligned}
\hat{\beta}_1 \left(\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i\right) &= \sum_{i=1}^{n} y_i x_i - \bar{y}\sum_{i=1}^{n} x_i \\
\Leftrightarrow \hat{\beta}_1 \left(\sum_{i=1}^{n} (x_i - \bar{x})^2\right) &= \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) \\
\Leftrightarrow \hat{\beta}_1 &= \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \\
\Leftrightarrow \hat{\beta}_1 &= \frac{c_{xy}}{s_x^2}
\end{aligned}
\tag{17}
$$

$\square$

Beispieldatensatz: Analyse

```
# Einlesen des Beispieldatensatzes
fname       = file.path(getwd(), "Daten", "Regression_Simulation.csv")
D           = read.table(fname, sep = ",", header = TRUE)

# Stichprobenstatistiken
x_bar       = mean(D$x_i)               # Stichprobenmittel der x_i-Werte
y_bar       = mean(D$y_i)               # Stichprobenmittel der y_i-Werte
s2x         = var(D$x_i)                # Stichprobenvarianz der x_i-Werte
cxy         = cov(D$x_i, D$y_i)         # Stichprobenkovarianz der (x_i,y_i)-Werte

# Ausgleichsgeradenparameter
beta_1_hat  = cxy/s2x                   # \hat{\beta}_1, Steigungsparameter
beta_0_hat  = y_bar - beta_1_hat*x_bar  # \hat{\beta}_0, Offset Parameter

# Ausgabe
cat(  "beta_0_hat:", beta_0_hat,
    "\nbeta_1_hat:", beta_1_hat)
```

```
> beta_0_hat: -6.19
> beta_1_hat: 1.66
```

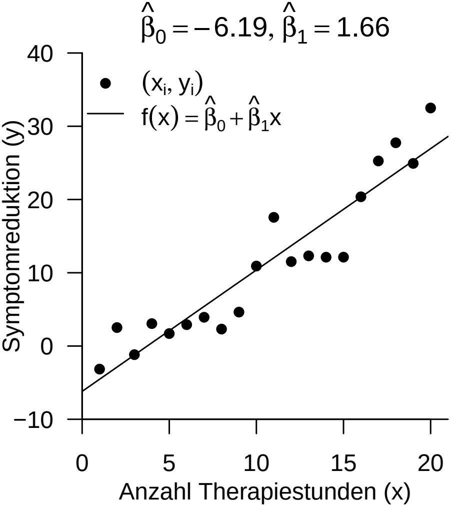
Beispieldatensatz: Visualisierung

```
# Datenwerte
plot(D$x_i, D$y_i,
    pch         = 16,
    xlab        = "Anzahl Therapiestunden (x)",
    ylab        = "Symptomreduktion (y)",
    xlim        = c(0,21),
    ylim        = c(-10, 40),
    main        = TeX("$\\hat{\\beta}_0 = -6.19, \\hat{\\beta}_1 = 1.66$"))

# Ausgleichsgerade
abline(
    coef        = c(beta_0_hat, beta_1_hat),
    lty         = 1,
    col         = "black")

# Legende
legend("topleft", c(TeX("$(x_i,y_i)$"), TeX("$f(x) = \\hat{\\beta}_0 + \\hat{\\beta}_1x$")),
    lty        = c(0,1),
    pch        = c(16, NA),
    bty        = "n")
```

Beispieldatensatz: Visualisierung



Definition (Polynom)

Sei $\beta := (\beta_0, ..., \beta_k)^{\mathrm{T}} \in \mathbb{R}^{k+1}$ ein $(k+1)$-dimensionaler Vektor reeller Zahlen. Dann ist ein Polynom k-ten Grades durch folgende Funktion gegeben:

$$f_\beta : \mathbb{R} \to \mathbb{R}, x \mapsto f_\beta(x) := \sum_{i=0}^{k} \beta_i \, x^i \; . \tag{18}$$

Definition (Ausgleichspolynom)

$\{(x_1, y_1), ..., (x_n, y_n)\} \subset \mathbb{R}^2$ sei ein Datensatz. Weiterhin sei $q(\beta)$ die Summe der quadrierten vertikalen Abweichungen der y_i von den Funktionswerten $f_\beta(x_i)$:

$$q : \mathbb{R}^{k+1} \to \mathbb{R}_{\geq 0}, \beta \mapsto q(\beta) := \sum_{i=1}^{n} \left(y_i - f_\beta(x_i)\right)^2 = \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{k} \beta_j \, x_i^j\right)^2 \tag{19}$$

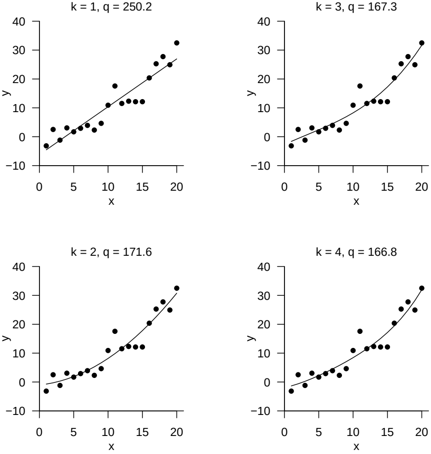
Dann heißt das Polynom k-ten Grades f_β, für das die Funktion q ihr Minimum annimt, das *Ausgleichspolynom k-ten Grades* für den Datensatz $\{(x_1, y_1), ..., (x_n, y_n)\}$.

Bemerkungen

- Wir nehmen hier ohne Beweis an, dass das Minimum von q eindeutig ist.
- Die Ausgleichsgerade ist das Ausgleichspolynom ersten Grades.
- Die Paramterwerte $\hat{\beta}_0, ..., \hat{\beta}_k$, für die q bei gegebenem Datensatz ihr Minimum annimmt, werden an späterer Stelle im Rahmen der Theorie des Allgemeinen Linearen Modells bestimmt werden.

Methode der kleinsten Quadrate

Beispieldatensatz: Ausgleichspolynome ersten bis vierten Grades ($k = 1, ..., 4$)



$\bullet\ (x_i, y_i) \qquad — \ f_{\hat{\beta}}(x) = \sum_{i=0}^{k} \hat{\beta}_i \, x^i$

Methode der kleinsten Quadrate

Einfache lineare Regression

Selbstkontrollfragen

Motivation

Eine Ausgleichsgerade erlaubt Aussagen über unbeobachtete y-Werte für gegebene x-Werte. Der Wert von $q(\hat{\beta})$ quantifiziert zudem die Güte der Ausgleichsgeradenpassung. Eine Ausgleichsgerade erlaubt allerdings nur implizite Aussagen über die mit der Anpassung verbundene Unsicherheit.

In der einfachen linearen Regression wird die Idee einer Ausgleichsgerade um eine probabilistische Komponente erweitert (normalverteilte Fehlervariable), um quantitative Aussagen über die mit einer Ausgleichsgeradenanpassung verbundene Unsicherheit machen zu können. Weiterhin erlaubt die einfache lineare Regression, einen Hypothesentest-basierten Zugang zur Einschätzung der angepassten Parameterwerte $\hat{\beta}_0$ und $\hat{\beta}_1$.

Wir betrachten hier zunächst nur das probabilistische Modell der einfachen linearen Regression sowie die auf ihm basierende Maximum-Likelihood-Schätzung der Parameter β_0 und β_1. Die Bewertung von Parameterschätzerunsicherheit sowie parameterzentrierte Hypothesentests behandeln wir an späterer Stelle in verallgemeinerter Form (siehe Einheiten (5)-(8) in *Allgemeines Lineares Modell*).

Einfache lineare Regression

Definition (Modell der einfachen linearen Regression)

Für $i = 1, ..., n$ sei

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \tag{20}$$

wobei

- y_i beobachtbare Zufallsvariablen sind, die Werte einer abhängigen Variable modellieren,
- $x_i \in \mathbb{R}$ fest vorgegebene sogenannte *Prädiktorwerte* oder *Regressorwerte* sind, die Werte einer unabhängigen Variable modellieren
- $\beta_0, \beta_1 \in \mathbb{R}$ wahre, aber unbekannte *Offset- und Steigungsparameterwerte* sind und
- $\varepsilon_i \sim N(0, \sigma^2)$ unabhängig und identisch normalverteilte nicht-beobachtbare Zufallsvariablen mit wahrem, aber unbekanntem *Varianzparameter* $\sigma^2 > 0$ sind, die Fehler- oder Störvariablen modellieren.

Dann heißt (20) *Modell der einfachen linearen Regression*.

Bemerkung

- Wir bezeichnen hier mit y_i sowohl die Zufallsvariable, über deren Wahrscheinlichkeitsverteilung wir Aussagen machen können (vormals υ_i), als auch eine konkrete Realisierung dieser Zufallsvariable, die Teil eines Datensatzes darstellt. Es wird also hier nicht zwischen der Zufallsvariable υ_i und ihrer Realisierung y_i unterschieden.
- Das Modell der einfachen linearen Regression hat drei Parameter: $\beta_0, \beta_1 \in \mathbb{R}$ und $\sigma^2 \in \mathbb{R}_{>0}$.

Theorem (Datenverteilung der einfachen linearen Regression)

Das Modell der einfachen linearen Regression

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, ..., n \tag{21}$$

lässt sich äquivalent in der Datenverteilungsform

$$y_i \sim N\left(\mu_i, \sigma^2\right) \text{ u.v. für } \text{ mit } \mu_i := \beta_0 + \beta_1 x_i \text{ und } i = 1, ..., n \tag{22}$$

schreiben.

Bemerkung

- Die Werte der abhängigen Variable werden im Modell der einfachen linearen Regression also durch unabhängige normalverteilte Zufallsvariablen mit im Allgemeinen unterschiedlichen Erwartungswertparametern modelliert.

Beweis

Wir zeigen die Äquivalenz für ein i. Die Unabhängigkeit der y_i zeigen wir an späterer Stelle im Rahmen des Allgemeinen Linearen Modells. Die Äquivalenz beider Modellformen für ein i folgt direkt aus der Transformation normalverteilter Zufallsvariablen durch linear-affine Funktionen. Speziell gilt im vorliegenden Fall für $\varepsilon_i \sim N(0, \sigma^2)$, dass

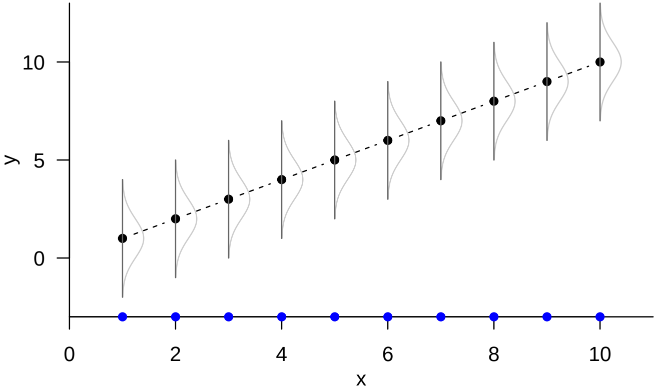
$$y_i = f(\varepsilon_i) \text{ mit } f : \mathbb{R} \to \mathbb{R}, \varepsilon_i \mapsto f(\varepsilon_i) := \varepsilon_i + (\beta_0 + \beta_1 x_i) \tag{23}$$

Mit dem Transformationstheorem für Wahrscheinlichkeitsdichtefunktionen bei linear-affinen Abbildungen (siehe Einheit (7) in *Wahrscheinlichkeitstheorie und Frequentistische Inferenz*) folgt dann

$$\begin{aligned} p(y_i) &= \frac{1}{|1|} p_{\varepsilon_i} \left(\frac{y_i - \beta_0 - \beta_1 x_i}{1} \right) \\ &= N\left(y_i - \beta_0 - \beta_1 x_i; 0, \sigma^2\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (y_i - \beta_0 - \beta_1 x_i - 0)^2 \right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 \right) \\ &= N\left(y_i; \beta_0 + \beta_1 x_i, \sigma^2\right) \end{aligned} \tag{24}$$

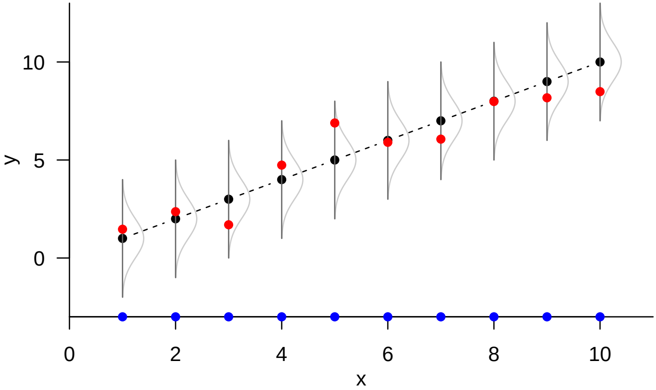
Definiert man $\mu_i := \beta_0 + \beta_1 x_i$ für $i = 1, ..., n$, ergibt sich die Aussage des Theorems. $\square$

Modell der einfachen linearen Regression



• x_i • $\beta_0 + \beta_1 x_i$ für $\beta_0 := 0,\ \beta_1 := 1$ — $N(y_i; \beta_0 + \beta_1 x_i, \sigma^2)$ für $\sigma^2 := 1$.

Realisierung des Modells der einfachen linearen Regression



• x_i • $\beta_0 + \beta_1 x_i$ für $\beta_0 := 0,\ \beta_1 := 1$ — $N(y_i; \beta_0 + \beta_1 x_i, \sigma^2)$ für $\sigma^2 := 1$ • (x_i, y_i)

Theorem (Maximum-Likelihood-Schätzung)

Es sei

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, ..., n \tag{25}$$

das Modell der einfachen linearen Regression. Dann sind Maximum-Likelihood-Schätzer der Modellparameter β_0, β_1 und σ^2 gegeben durch

$$\hat{\beta}_1 := \frac{c_{xy}}{s_x^2}, \quad \hat{\beta}_0 := \bar{y} - \hat{\beta}_1 \bar{x} \quad \text{und} \quad \hat{\sigma}^2 := \frac{1}{n} \sum_{i=1}^n \Big(y_i - \big(\hat{\beta}_0 + \hat{\beta}_1 x_i \big) \Big)^2 . \tag{26}$$

Bemerkungen

- In der Vergangenheit (siehe Einheit (9) in *Wahrscheinlichkeitstheorie und Frequentistische Inferenz*) haben wir Maximum-Likelihood-Schätzer mit ML-Superskripten ($^{\text{ML}}$) gekennzeichnet. Aus Gründen der Übersichtlichkeit verzichten wir hier auf die ML-Superskripte und kennzeichnen Maximum-Likelihood-Schätzer durch das "Dach"-Zeichen ($\hat{\ }$).
- Die ML-Schätzer für β_0 und β_1 sind offenbar mit den Ausgleichsgeradenparametern identisch.

Einfache lineare Regression

Beweis

Wir zeigen zunächst, dass die Ausgleichsgerardenparameter $\hat{\beta}_0$ und $\hat{\beta}_1$ den entsprechenden ML-Schätzern gleichen. Dazu halten wir zunächst fest, dass aufgrund der Unabhängigkeit der $y_1, ..., y_n$ die Likelihood-Funktion des Modells der einfachen linearen Regression bezüglich β_0 und β_1 die folgende Form hat:

$$
\begin{aligned}
L : \mathbb{R}^2 \to \mathbb{R}_{>0}, (\beta_0, \beta_1) \mapsto L(\beta_0, \beta_1) &:= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - (\beta_0 + \beta_1 x_i))^2\right) \\
&= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2\right) .
\end{aligned}
\tag{27}
$$

Für die Exponentialfunktion gilt: Wenn $a < b \leq 0$ gilt, dann gilt $\exp(a) < \exp(b)$. Somit wird der Exponentialterm dieser Likelihood-Funktion maximal, wenn der Term

$$
q := \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \geq 0 \tag{28}
$$

minimal und damit $-q$ maximal wird. Im Rahmen des Beweises der Ausgleichsgeradenform haben wir aber schon gezeigt, dass der Term (28) für

$$
\hat{\beta}_1 := \frac{c_{xy}}{s_x^2} \text{ und } \hat{\beta}_0 := \bar{y} - \hat{\beta}_1 \bar{x} \tag{29}
$$

minimal wird, und damit $\hat{\beta}_1$ und $\hat{\beta}_0$ die Likelihood-Funktion maximieren.

Beweis (fortgeführt)

In einem zweiten Schritt betrachten wir nun die Likelihood-Funktion des Modells der einfachen linearen Regression bezüglich σ^2 an der Stelle von $\hat{\beta}_0$ und $\hat{\beta}_1$. Wir erhalten die Likelihood-Funktion

$$L : \mathbb{R}_{>0} \to \mathbb{R}_{>0}, \sigma^2 \mapsto L(\sigma^2) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2\right) \tag{30}$$

und die entsprechende Log-Likelihood-Funktion

$$\ell : \mathbb{R}_{>0} \to \mathbb{R}, \sigma^2 \mapsto \ell(\sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 \tag{31}$$

In Analogie zu der Herleitung des ML-Schätzers für σ^2 im Normalverteilungsmodell (vgl. Einheit (9) in *Wahrscheinlichkeitstheorie und Frequentistische Inferenz*) ergibt sich unter Beachtung von

$$\hat{\mu} = \hat{\beta}_0 + \hat{\beta}_1 x_i \tag{32}$$

dann hier

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 \,. \tag{33}$$

□

Beispieldatensatz: Parameterschätzung

```
# Einlesen des Beispieldatensatzes
fname       = file.path(getwd(), "Daten", "Regression_Simulation.csv")
D           = read.table(fname, sep = ",", header = TRUE)

# Stichprobenstatistiken
n           = length(D$y_i)                                 # Anzahl Datenpunkte
x_bar       = mean(D$x_i)                                   # Stichprobenmittel der x_i-Werte
y_bar       = mean(D$y_i)                                   # Stichprobenmittel der y_i-Werte
s2x         = var(D$x_i)                                    # Stichprobenvarianz der x_i-Werte
cxy         = cov(D$x_i, D$y_i)                             # Stichprobenkovarianz der (x_i,y_i)-Werte

# Parameteterschätzer
beta_1_hat  = cxy/s2x                                       # \hat{\beta}_1, Steigungsparameter
beta_0_hat  = y_bar - beta_1_hat*x_bar                      # \hat{\beta}_0, Offset-Parameter
sigsqr_hat  = (1/n)*sum((D$y_i-(beta_0_hat+beta_1_hat*D$x_i))^2) # \hat{sigma}^2, Varianzparameter

# Ausgabe
cat(  "beta_0_hat:", beta_0_hat,
    "\nbeta_1_hat:", beta_1_hat,
    "\nsigsqr_hat:", sqrt(sigsqr_hat))
```

```
> beta_0_hat: -6.19
> beta_1_hat: 1.66
> sigsqr_hat: 3.54
```

Beispieldatensatz: Analyse mit `lm()`

```
# Einlesen des Beispieldatensatzes
library(car)
fname       = file.path(getwd(), "Daten", "Regression_Simulation.csv")
D           = read.table(fname, sep = ",", header = TRUE)

# Analyse mit lm()
model       = lm(formula = D$y_i ~ D$x_i, data = D)
print(model)
```

```
>
> Call:
> lm(formula = D$y_i ~ D$x_i, data = D)
>
> Coefficients:
> (Intercept)        D$x_i
>       -6.19         1.66
```

Methode der kleinsten Quadrate

Einfache lineare Regression

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die funktionale Form eine linear-affinen Funktion an.
2. Erläutern Sie die Bedeutung der Parameter einer linear-affinen Funktion.
3. Definieren Sie den Begriff der Ausgleichsgerade.
4. Erläutern Sie die intuitive Bedeutung der Funktion der quadrierten vertikalen Abweichungen.
5. Geben Sie das Theorem zur Ausgleichsgerade wieder.
6. Skizzieren Sie den Beweis des Theorems zur Ausgleichsgeraden.
7. Definieren Sie den Begriff des Ausgleichspolynoms.
8. Erläutern Sie die Motivation des einfachen linearen Regressionsmodells in Bezug auf die Ausgleichsgerade.
9. Definieren Sie das Modell der einfachen linearen Regression.
10. Geben Sie das Theorem zur Datenverteilung der einfachen linearen Regression wieder.
11. Skizzieren das Modell der einfachen linearen Regression per Hand.
12. Skizzieren Sie eine Realisierung des Modells der einfachen linearen Regression per Hand.
13. Geben Sie das Theorem zur ML-Schätzung der Parameter der einfachen linearen Regression an.
14. Skizzieren Sie den Beweis des Theorems zur ML-Schätzung der Parameter der einfachen linearen Regression.