Tutorium

Allgemeines Lineares Modell

BSc Psychologie SoSe 2023

2. Termin: (1) Regression

Sean Mulready

Selbstkontrollfragen - Regression

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.

1. Geben Sie die funktionale Form eine linear-affinen Funktion an.

Für $\beta := (\beta_0, \beta_1)^T \in \mathbb{R}^2$ hat die linear-affine Funktion $f_\beta : \mathbb{R} \to \mathbb{R}, x \mapsto f_\beta(x)$ die funktionale Form $f_\beta(x) := \beta_0 + \beta_1 x$.

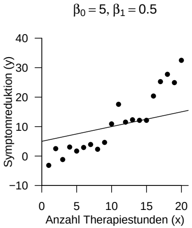
linear-*affin*, da die Funktion nicht unbedingt durch den Nullpunkt der y-Achse geht

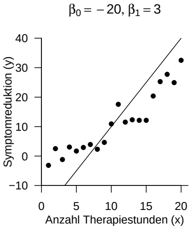
2. Erläutern Sie die Bedeutung der Parameter einer linear-affinen Funktion.

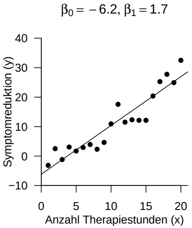
Für linear-affine Funktionen $f_\beta(x) := \beta_0 + \beta_1 x$ ist

- β_0 der Schnittpunkt von Gerade und y-Achse ("Offset Parameter") und
- β_1 die y-Differenz pro x-Einheitsdifferenz ("Steigungsparameter").

Beispiele:







3. Definieren Sie den Begriff der Ausgleichsgerade.

Definition (Ausgleichsgerade)

Für $\beta := (\beta_0, \beta_1)^T \in \mathbb{R}^2$ heißt die linear-affine Funktion

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

für die für eine Wertemenge $\{(x_1, y_1), \ldots, (x_n, y_n)\} \subset \mathbb{R}^2$ die Funktion

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

der quadrierten vertikalen Abweichungen der y_i von den Funktionswerten $f_\beta(x_i)$ ihr Minimum annimt, die *Ausgleichsgerade* für die Wertemenge $\{(x_1, y_1), \ldots, (x_n, y_n)\}$.

4. Erläutern Sie die intuitive Bedeutung der Funktion der quadrierten vertikalen Abweichungen.

Die Funktion der quadrierten vertikalen Abweichungen $q(\beta)$ quantifiziert die Abweichungen der Funktionswerte $f_\beta(x_i)$ der Ausgleichsgeraden f_β von den (beobachteten) Werten y_i.

Die Abweichungen werden *quadriert*, damit sich positive und negative Abweichungen nicht zu null ausgleichen.

Intuitiv misst die Funktion $q(\beta)$ "wie gut" eine Ausgleichsgerade mit bestimmten Werten für β_0 und β_1 in das Streudiagramm der Datenpunkte "rein passt".

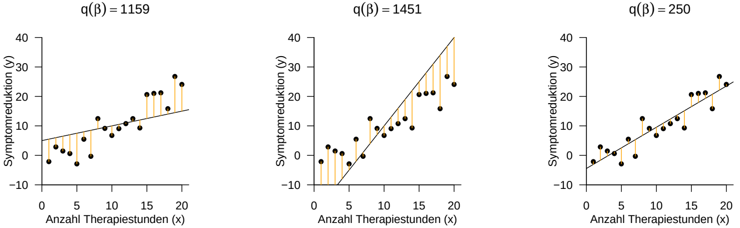
Je größer ein Funktionswert $q(\beta)$, desto größer ist die aufsummierte Abweichung und desto "schlechter" passt die Ausgleichsgerade f_β in das Streudiagramm der Datenpunkte.

Je geringer ein Funktionswert $q(\beta)$, desto geringer ist die aufsummierte Abweichung und desto "besser" passt die Ausgleichsgerade f_β in das Streudiagramm der Datenpunkte.

Zur Veranschaulichung

Funktion der quadrierten vertikalen Abweichungen

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



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

5. Geben Sie das Theorem zur Ausgleichsgerade wieder.

Theorem (Ausgleichsgerade)

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

$$f_\beta : \mathbb{R} \to \mathbb{R}, x \mapsto f_\beta(x) := \hat{\beta}_0 + \hat{\beta}_1 x, \tag{3}$$

wobei mit der Stichprobenkovarianz c_{xy} der (x_i, y_i)-Werte, der Stichprobenvarianz s_x^2 der x_i-Werte und den Stichprobenmitteln $\bar{x}$ und $\bar{y}$ der x_i- und y_i-Werte, respektive, gilt, dass

$$\hat{\beta}_1 = \frac{c_{xy}}{s_x^2} \text{ und } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{4}$$

6. Skizzieren Sie den Beweis des Theorems zur Ausgleichsgeraden.

Wir wollen zeigen, dass die Summe der quadrierten Abweichungen $q(\beta_0, \beta_1)$ für die Werte $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$ und $\hat{\beta}_1 = \frac{c_{xy}}{s_x^2}$ ihr Minimum annimmt, wobei $\bar{x}$ und $\bar{y}$ die Stichprobenmittel der x_i- und y_i-Werte, repektive, sind, c_{xy} die Stichprobenkovarianz der (x_i, y_i)-Werte und s_x^2 die Stichprobenvarianz der x_i-Werte entspricht.

Dafür bestimmen wir zunächst das Minimum der Funktion $q(\beta_0, \beta_1)$, indem wir die partiellen Ableitungen hinsichtlich β_0 und β_1 berechnen und diese gleich 0 setzen. Formal,

$$\frac{\partial}{\partial\beta_0} q(\beta_0, \beta_1) = 0 \text{ und } \frac{\partial}{\partial\beta_1} q(\beta_0, \beta_1) = 0$$

Durch die partielle Ableitungen und das anschließende Umstellen von Termen ergibt sich ein Gleichungssystem, das das *System der Normalengleichungen* genannt wird und die notwendige Bedingung für ein Minimum von q beschreibt. Auflösen dieses Gleichungssystems nach β_0 und β_1 liefert dann die Werte $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$ und $\hat{\beta}_1 = \frac{c_{xy}}{s_x^2}$ des Theorems.

7. Definieren Sie den Begriff des Ausgleichspolynoms.

Definition (Ausgleichspolynom)

Für $\beta := (\beta_0, ..., \beta_k)^T \in \mathbb{R}^{k+1}$ heißt die Polynomfunktion $k$ten Grades

$$f_\beta : \mathbb{R} \to \mathbb{R}, x \mapsto f_\beta(x) := \sum_{i=0}^{k} \beta_i x^i, \tag{5}$$

für die für eine Wertemenge $\{(x_1, y_1), ..., (x_n, y_n)\} \subset \mathbb{R}^2$ die Funktion

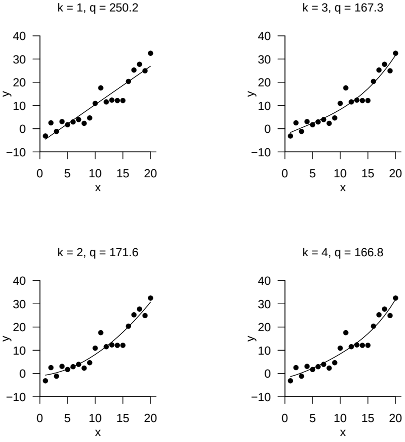
$$q : \mathbb{R}^{k+1} \to \mathbb{R}_{\geq 0}, \beta \mapsto q(\beta) := \sum_{i=1}^{n} \Big(y_i - f_\beta(x_i)\Big)^2 = \sum_{i=1}^{n} \left(y_i - \sum_{i=0}^{k} \beta_i x^i\right)^2 \tag{6}$$

der quadrierten vertikalen Abweichungen der y_i von den Funktionswerten $f_\beta(x_i)$ ihr Minimum annimt, das *Ausgleichspolynom $k$ten Grades* für die Wertemenge $\{(x_1, y_1), ..., (x_n, y_n)\}$.

Methode der kleinsten Quadrate - Selbstkontrollfragen

Zur Veranschaulichung

Beispieldatensatz Ausgleichspolynome 1ten bis 4ten Grades



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

8. Erläutern Sie die Motivation des einfachen linearen Regressionsmodells in Bezug auf die Ausgleichsgerade.

Eine Ausgleichsgerade erlaubt Aussagen über unbeobachtete y Werte für x Werte. Der Wert von $q(\hat{\beta})$ quantifiziert 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 (normalverteilte Fehlervariable) erweitert, 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 $\hat{\beta}_1$ sowie das Bestimmen von Konfidenzintervallen, die eine quantitative Aussage über die mit dem Schätzwert assoziierte Unsicherheit ermöglichen.

9. Definieren Sie das Modell der einfachen linearen Regression.

Definition (Modell der einfachen linearen Regression)

Für $i = 1, ..., n$ sei

$$\upsilon_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{7}$$

wobei

- $x_i \in \mathbb{R}$ fest vorgegebene sogenannte *Prädiktorwerte* oder *Regressorwerte* sind,
- $\beta_0, \beta_1 \in \mathbb{R}$ wahre, aber unbekannte, Parameterwerte sind und
- $\varepsilon_i \sim N(0, \sigma^2)$ unabhängige und identisch normalverteilte nicht-beobachtbare Zufallsvariablen mit wahrem, aber unbekanntem, Parameter $\sigma^2 > 0$ sind.

Dann heißt (7) *Modell der einfachen linearen Regression*.

10. Geben Sie das Theorem zur Datenverteilung der einfachen linearen Regression wieder.

Theorem (Datenverteilung der einfachen linearen Regression)

Das Modell der einfachen linearen Regression

$$\upsilon_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, ..., n \tag{8}$$

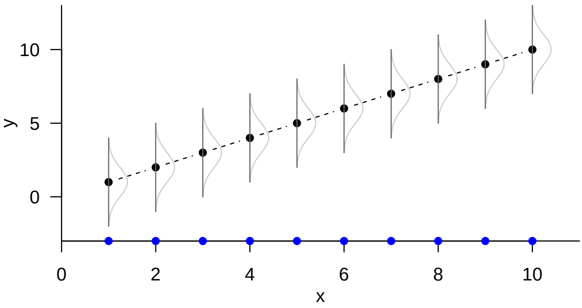
lässt sich äquivalent in der Form

$$\upsilon_i \sim N\left(\beta_0 + \beta_1 x_i, \sigma^2\right) \text{ u.v. für } i = 1, ..., n \tag{9}$$

schreiben.

11. Skizzieren das Modell der einfachen linearen Regression per Hand.

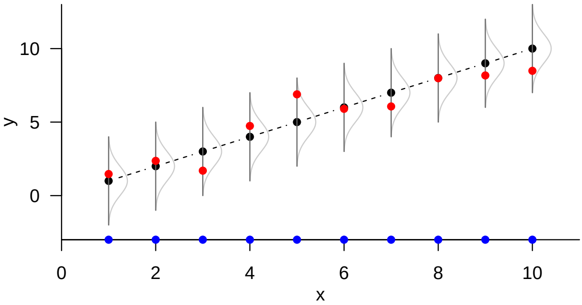
Modell der einfachen linearen Regression



• 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 $\sigma^2 := 1.$

12. Skizzieren Sie eine Realisierung des Modells der einfachen linearen Regression per Hand.

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)

13. Geben Sie das Theorem zur ML-Schätzung der Parameter der einfachen linearen Regression an.

Theorem (Maximum Likelihood Schätzung)

Es sei

$$\upsilon_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \dots, n \tag{10}$$

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

$$\hat{\beta}_1 := \frac{c_{xy}}{s_x^2}, \quad \hat{\beta}_0 := \bar{y} - \hat{\beta}_1 \bar{x} \quad \text{und } \hat{\sigma}^2 := \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i \right) \right)^2 . \tag{11}$$

14. Skizzieren Sie den Beweis des Theorems zur ML-Schätzung der Parameter der einfachen linearen Regression.

Teil 1/2: $\hat{\beta}_0$ und $\hat{\beta}_1$

Wir wollen zunächst zeigen, dass die Ausgleichsgeradenparameter $\hat{\beta}_0$ und $\hat{\beta}_1$ den entsprechenden ML Schätzern gleichen.

Um die ML Schätzer zu bestimmen, formulieren wir zunächst die Likelihood-Funktion des Modells der einfachen linearen Regression in Abhängigkeit von β_0 und β_1.

Die Likelihood-Funktion ist definiert als der Wert der gemeinsamen Verteilung der $\upsilon_1, ..., \upsilon_n$ in Abhängigkeit von den Parametern β_0 und β_1.

Aufgrund der Unabhängigkeit der $\upsilon_1, ..., \upsilon_n$ können wir die gemeinsame Verteilung als Produkt der einzelnen Wahrscheinlichkeitsdichtefunktionen, also als Produkt von Dichtefunktionen der univariaten Normalverteilung aufschreiben.

Die funktionale Form der Dichtefunktionen der univariaten Normalverteilung enthält eine Exponentialfunktion. Mit den Eigenschaften einer Exponentialfunktion können wir dieses Produkt umschreiben zu einer Exponentialfunktion von einem Term, der im Wesentlichen aus der negativen Summe der quadrierten Abweichungen (i.e. der Funktion q) besteht.

Weil für eine Exponentialfunktion gilt, dass für $a < b \leq 0$ gilt, dass $\exp(a) < \exp(b)$, wird der Exponentialterm der Likelihood-Funktion maximal, wenn q minimal und entsprechend $-q$ maximal wird.

Wie im Beweis der Ausgleichsgeradenform gezeigt, wissen wir, dass q für $\hat{\beta}_0$ und $\hat{\beta}_1$, wie sie auch im Theorem zur ML-Schätzung der Parameter der einfachen linearen Regression angegeben sind, minimal wird, und damit $\hat{\beta}_1$ und $\hat{\beta}_0$ die Likelihood-Funktion maximieren.

14. Skizzieren Sie den Beweis des Theorems zur ML-Schätzung der Parameter der einfachen linearen Regression.

Teil 2/2: $\hat{\sigma^2}$

Als nächstes wollen wir zeigen, dass $\hat{\sigma}^2$ dem ML-Schätzer entspricht.

Dazu betrachten wir analog zu oben die Likelihood-Funktion des Modells der einfachen linearen Regression, jedoch als Funktion von σ^2 und formulieren die entsprechende log-Likelihood-Funktion.

Wir wollen das $\hat{\sigma}^2$ bestimmen, für das die (log-)Likelihood-Funktion maximal wird.

Um die log-Likelihood-Funktion zu maximieren, bilden wir die 1. Ableitung, setzen diese gleich 0 und lösen nach σ^2 auf. Durch umstellen erhalten wir dann die Formel zur Schätzung von σ^2, also $\hat{\sigma}^2$, wie sie im Theorem angeben ist.