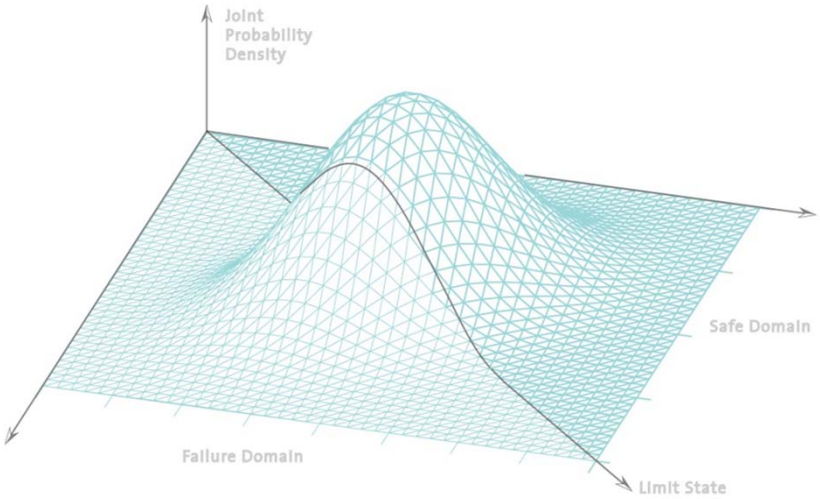


Statistik und Wahrscheinlichkeitsrechnung

Übung 9

Inhalt der heutigen Übung

- Informationen zur Testatprüfung
- Besprechung der der Hausübung E.6
- Gemeinsames Lösen der Übungsaufgaben
  - E.13: Bayes‘sches Updating
  - E.14: Regressionsanalyse
- Keine Hausübung

Testatprüfung am Donnerstag 5.Mai

Wann?

Donnerstag, 5. Mai, 8:00 Uhr

Dauer der Prüfung: 60 min

Wo?

Die Raumaufteilung wird noch auf unserer Homepage veröffentlicht:

www.ibk.ethz.ch/fa/education/ss_statistics

Testatprüfung am Donnerstag 5.Mai

Inhalt

60 min. Multiple Choice

Gesamter Stoff bis einschliesslich Vorlesung 9 und Übung 9.

Hilfsmittel

Eine DIN A4 Seite doppelseitig Formelsammlung erlaubt.

Keine weiteren Hilfsmittel erlaubt!

Aufgabe E.13

Die Betondruckfestigkeit X von Probekörpern einer bestimmten Produktion wird als normalverteilt angenommen:

$$X \sim N\left(\mu_X, \sigma_X\right)$$

Aus früheren Testergebnissen ist der Mittelwert und die Standardabweichung der Druckfestigkeit bekannt. Nach Berücksichtigung der statistischen Unsicherheit ist der Mittelwert der Betondruckfestigkeit ebenfalls normalverteilt (Einheit in *MPa*):

$$\mu_X \sim N\left(\mu'_{\mu_X} = 35, \sigma'_{\mu_X} = 3\right)$$

Die Standardabweichung wird als bekannt (deterministisch) angenommen:

$$\sigma_X = 10MPa$$

Aufgabe E.13

Um die Verteilung des Parameters μ_X zu aktualisieren, wurden 20 Versuchskörper auf ihre Druckfestigkeit geprüft. Die Ergebnisse sind in der Tabelle gegeben.

Bestimme die *a posteriori* Verteilung des Mittelwertes sowie die prädiktive Verteilung der Beton-Druckfestigkeit

$$\mu_X \sim N\left(\mu'_{\mu_X} = 35, \sigma'_{\mu_X} = 3\right)$$

$$\sigma_X = 10MPa$$

Es wird nur die Verteilung des Mittelwertes aktualisiert.

Nr.	Druck-festigkeit [MPa]	Nr.	Druck-festigkeit [MPa]
1	24.4	11	33.3
2	27.6	12	33.5
3	27.8	13	34.1
4	27.9	14	34.6
5	28.5	15	35.8
6	30.1	16	35.9
7	30.3	17	36.8
8	31.7	18	37.1
9	32.2	19	39.2
10	32.8	20	39.7

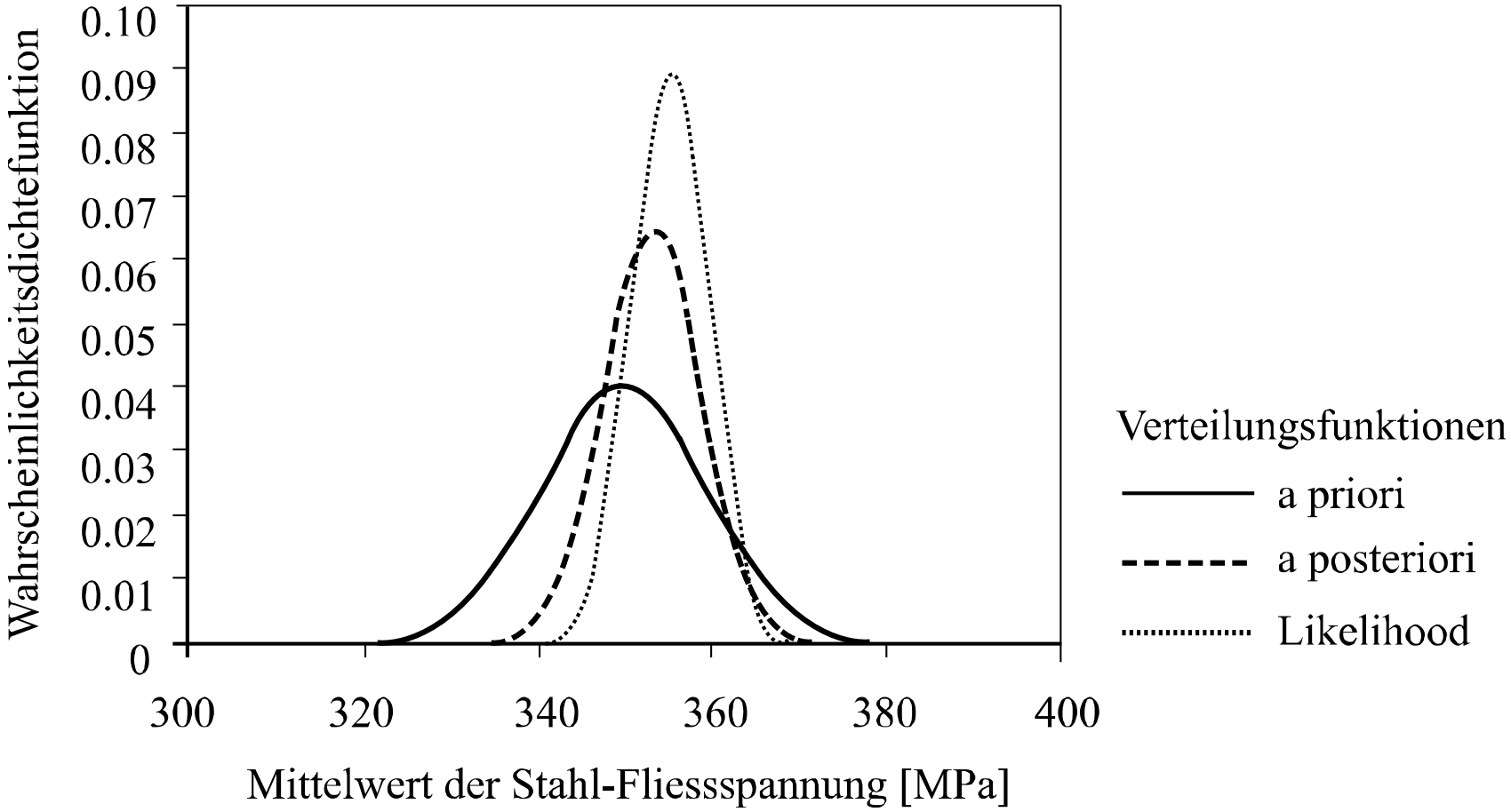
Aktualisierung von Wahrscheinlichkeitsverteilungen

Durch neue Messungen $\hat{x}$ können wir unsere *A Priori* Wahrscheinlichkeitsdichte für den Mittelwert μ_X aktualisieren – Dazu verwenden wir den Satz von Bayes.

A Posteriori *Likelihood* *A Priori*

$$f_{\mu_X}''(\mu_X \mid \hat{\mathbf{x}}) = \frac{f_X(\hat{\mathbf{x}} \mid \mu_X) \cdot f_{\mu_X}'(\mu_X)}{\int f_X(\hat{\mathbf{x}} \mid \mu_X) \cdot f_{\mu_X}'(\mu_X) d\mu_X}$$

Aktualisierung von Wahrscheinlichkeitsverteilungen



Aufgabe E.13

Die konjugierte a priori Verteilung für das Aktualisieren des Mittelwertes μ_X einer Normalverteilung bei bekannter Standardabweichung σ_X ist die Normalverteilung.

$$\mu_X \sim N\left(\mu'_{\mu_X} = 35, \sigma_{\mu_X}' = 3\right)$$

$$\sigma_X = 10MPa$$

Für diesen Fall gibt es eine analytische Lösung!

Konjugierte Verteilungen:

Sowohl die a priori als auch die a posteriori Verteilung gehören zur selben Verteilungsfamilie.

Aktualisierung von Wahrscheinlichkeitsverteilungen

Wollen wir den Mittelwert μ_X einer Normalverteilung aktualisieren, ergibt sich folgender Zusammenhang:

$$f_{\mu_X}''(\mu_X \mid \hat{\mathbf{x}}) = \frac{f_X(\hat{\mathbf{x}} \mid \mu_X) \cdot f_{\mu_X}'(\mu_X)}{\int f_X(\hat{\mathbf{x}} \mid \mu_X) \cdot f_{\mu_X}'(\mu_X) d\mu_X}$$

Die a priori Verteilung $f_{\mu_X}'(\mu_X \mid \hat{\mathbf{x}})$ und die Likelihood $f_{\mu_X}(\hat{\mathbf{x}} \mid \mu_X)$ sind Normalverteilungen.

⇨ Dann ist auch die a posteriori Verteilung $f_{\mu_X}''(\mu_X \mid \hat{\mathbf{x}})$ eine Normalverteilung!

Aufgabe E.13

Die konjugierte a priori Verteilung für das Aktualisieren des Mittelwertes μ_X einer Normalverteilung bei bekannter Standardabweichung σ_X ist die Normalverteilung.

$$\mu_X \sim N\left(\mu'_{\mu_X} = 35, \sigma_{\mu_X}' = 3\right)$$

$$\sigma_X = 10MPa$$

Für die analytische Lösung benötigen wir den Stichprobenmittelwert $\overline{x}$ und die Stichprobengrösse n

$$\overline{x} = 32.665$$
$$n = 20$$

Nr.	Druck-festigkeit [MPa]	Nr.	Druck-festigkeit [MPa]
1	24.4	11	33.3
2	27.6	12	33.5
3	27.8	13	34.1
4	27.9	14	34.6
5	28.5	15	35.8
6	30.1	16	35.9
7	30.3	17	36.8
8	31.7	18	37.1
9	32.2	19	39.2
10	32.8	20	39.7

Aufgabe E.13

$\sigma_X = 10$ $\quad \bar{x} = 32.665$ $\quad n = 20$

A priori Verteilung für μ_X :

$$\mu_X \sim N\left(\mu'_{\mu_X} = 35, \sigma_{\mu_X}' = 3\right)$$

Parameter der *a posteriori* Verteilung für μ_X :

$$\mu_{\mu_X}'' = \frac{\frac{\mu_{\mu_X}'}{n} + \frac{\bar{x}}{n'}}{\frac{1}{n'} + \frac{1}{n}}$$

$n = 20$

Stichprobengrösse
(Anzahl Daten)

$$n' = \frac{\sigma_X^2}{\sigma'^2_{\mu_X}}$$

«Stichprobengewichtung»
der a priori Verteilung

⇨ Je mehr Daten zum Aktualisieren der A priori Verteilung verwendet werden, desto grösser ist der Einfluss der Likelihood (der Daten) auf die a posteriori Verteilung

⇨ Je kleiner die Unsicherheit in der A priori Verteilung ist, desto grösser ist ihr Einfluss auf die a posteriori Verteilung

Aufgabe E.13

$\sigma_X = 10$ $\quad \bar{x} = 32.665$

$n = 20$

A priori Verteilung für μ_X :

$$\mu_X \sim N\left(\mu'_{\mu_X} = 35, \sigma_{\mu_X}' = 3\right)$$

Parameter der *a posteriori* Verteilung für μ_X :

$$\mu_{\mu_X}'' = \frac{\frac{\mu_{\mu_X}'}{n} + \frac{\bar{x}}{n'}}{\frac{1}{n'} + \frac{1}{n}} = \frac{\frac{35}{20} + \frac{32.665}{11.11}}{\frac{1}{11.11} + \frac{1}{20}} = \underline{\underline{33.499}}$$

$$n' = \frac{\sigma_X^2}{\sigma'^2_{\mu_X}} = \frac{10^2}{3^2} = \frac{100}{9} = 11.11$$

$$\sigma_{\mu_X}'' = \sqrt{\frac{\frac{\sigma_X^2}{n'} \cdot \frac{\sigma'^2_{\mu_X}}{n}}{\frac{\sigma'^2_{\mu_X}}{n'} + \frac{\sigma'^2_{\mu_X}}{n}}} = \sqrt{\frac{\frac{10^2}{11.11} \cdot \frac{3^2}{20}}{\frac{3^2}{11.11} + \frac{3^2}{20}}} = \underline{\underline{1.793}}$$

Aufgabe E.13

$\sigma_X = 10$ $\quad \bar{x} = 32.665$

$n = 20$

A priori Verteilung für μ_X :

$$\mu_X \sim N\left(\mu'_{\mu_X} = 35, \sigma_{\mu_X}' = 3\right)$$

Parameter der *a posteriori* Verteilung für μ_X :

$$\mu_{\mu_X}'' = \frac{\frac{\mu_{\mu_X}'}{n} + \frac{\bar{x}}{n'}}{\frac{1}{n'} + \frac{1}{n}} = \frac{\frac{35}{20} + \frac{32.665}{11.11}}{\frac{1}{11.11} + \frac{1}{20}} = \underline{\underline{33.499}}$$

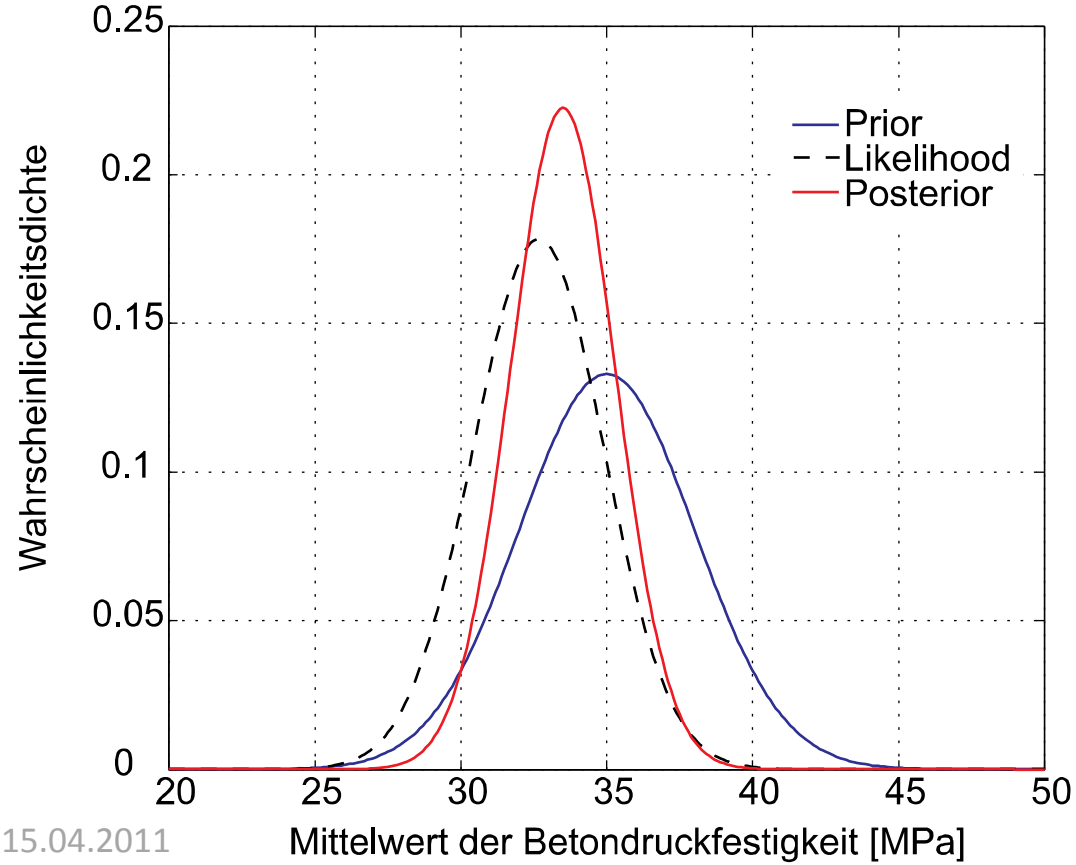
$$\sigma_{\mu_X}'' = \sqrt{\frac{\frac{\sigma_X^2}{n'} \cdot \frac{\sigma'^{\,2}_{\mu_X}}{n}}{\frac{\sigma'^{\,2}_{\mu_X}}{n'} + \frac{\sigma'^{\,2}_{\mu_X}}{n}}} = \sqrt{\frac{\frac{10^2}{11.11} \cdot \frac{3^2}{20}}{\frac{3^2}{11.11} + \frac{3^2}{20}}} = \underline{\underline{1.793}}$$

A posteriori Verteilung :

$$\mu_X \sim N\left(\mu''_{\mu_X} = 33.5, \sigma_{\mu_X}'' = 1.8\right)$$

Aufgabe E.13

In der *A Posteriori* Wahrscheinlichkeitsdichtefunktion für den Mittelwert μ_X sind die Informationen aus der *A priori* Verteilung und der *Likelihood* (den Daten) enthalten.



A priori Verteilung für μ_X :

$$\mu_X \sim N\left(\mu'_{\mu_X} = 35, \sigma_{\mu_X}' = 3\right)$$

A posteriori Verteilung :

$$\mu_X \sim N\left(\mu''_{\mu_X} = 33.5, \sigma_{\mu_X}'' = 1.8\right)$$

Aufgabe E.13

Die (a posteriori) *prädiktive* Verteilung ist die neue Verteilung unserer Zufallsvariablen X nach dem Aktualisieren der Verteilungsparameter:

$$\underbrace{f_X\left(x|\hat{\mathbf{x}}\right)}_{\text{Prädiktive Dichte}} = \int \underbrace{f_X\left(x\,|\,\mu_X,\sigma_X\right)}_{\text{Dichte bei gegebenen Parametern}} \cdot \underbrace{f''\left(\mu_X\,|\,\hat{\mathbf{x}}\right)}_{\text{A posteriori Verteilung der Paramter}} d\,\mu_X$$

Aufgabe E.13

In unserem Beispiel ist die (a posteriori) *prädiktive* Verteilung wiederum eine Normalverteilung:

$$f_X\left(x\middle|\hat{\mathbf{x}}\right) = \frac{1}{\sqrt{2\pi}\sigma'''_{\mu_X}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu'''_{\mu_X}}{\sigma'''_{\mu_X}}\right)^2\right)$$

$$X \sim N(\mu''', \sigma''') \qquad \mu''' = \mu''$$

A posteriori Mittelwert

Aufgabe E.13

In unserem Beispiel ist die (a posteriori) *prädiktive* Verteilung wiederum eine Normalverteilung:

$$f_X\left(x\middle|\hat{\mathbf{x}}\right) = \frac{1}{\sqrt{2\pi}\sigma'''_{\mu_X}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu'''_{\mu_X}}{\sigma'''_{\mu_X}}\right)^2\right)$$

$$X \sim N(\mu''', \sigma''')$$

$$\mu''' = \mu''$$

$$\sigma'''^2 = \underbrace{\sigma''^2_{\mu_X}}_{\text{Streuung des Mittelwerts}} + \underbrace{\sigma^2_X}_{\text{Streuung um den Mittelwert}}$$

Aufgabe E.13

In unserem Beispiel ist die (a posteriori) *prädiktive* Verteilung wiederum eine Normalverteilung:

$$f_X\left(x\middle|\hat{\mathbf{x}}\right)=\frac{1}{\sqrt{2\pi}\sigma'''_{\mu_X}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu'''_{\mu_X}}{\sigma'''_{\mu_X}}\right)^2\right)$$

$$X\sim N(\mu''',\sigma''')$$

$$\mu'''=\mu''=\underline{\underline{33.499MPa}}$$

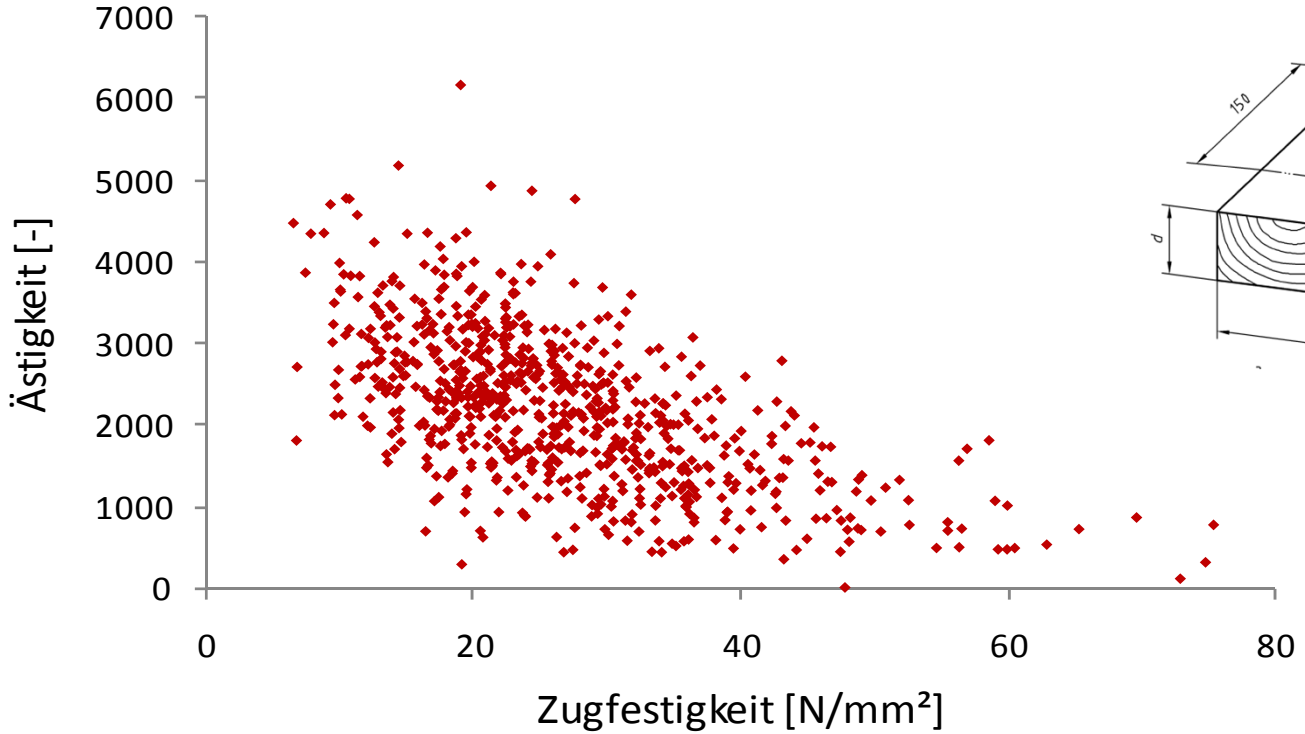
$$\sigma'''^2=\sigma''^2_{\mu_X}+\sigma^2_X$$

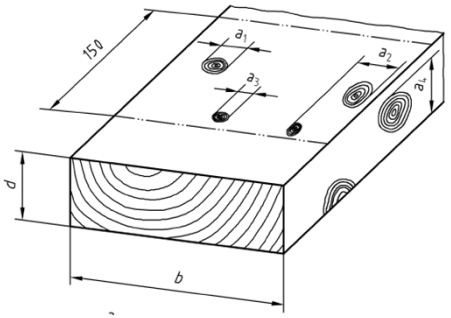
$$=1.793^2+10^2=103.214$$

$$\sigma'''=\underline{\underline{10.159MPa}}$$

(Lineare) Regression

Problemstellung:
Der funktionale Zusammenhang zwischen zwei Zufallsvariablen soll bestimmt werden. → Im Falle der linearen Regression eine Gerade.

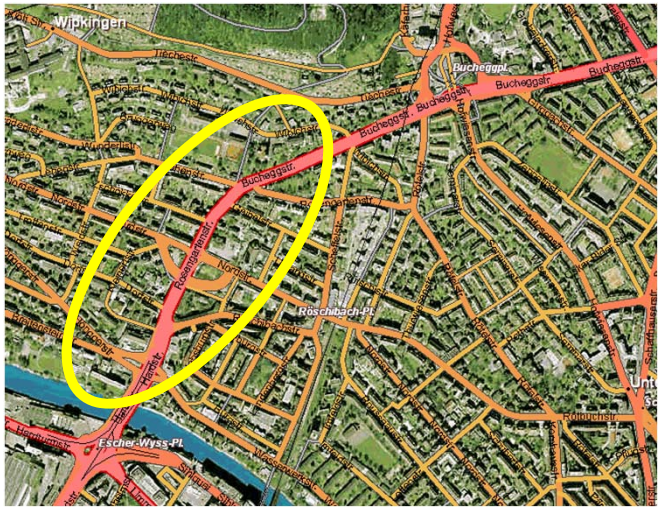




Aufgabe E.14

Datum	Richtung 1	Richtung 2
01.04.2001	32618	24609
02.04.2001	33380	29965
03.04.2001	34007	30629
04.04.2001	33888	30263
05.04.2001	35237	31405
06.04.2001	35843	31994
07.04.2001	33197	26846
08.04.2001	30035	22762
09.04.2001	32158	30366
10.04.2001	33406	29994
11.04.2001	34576	30958
12.04.2001	34013	30680
13.04.2001	24846	19735
14.04.2001	28252	21145
15.04.2001	25365	17805
16.04.2001	24862	18123
17.04.2001	32472	28117
18.04.2001	33245	28858
19.04.2001	33788	29080
20.04.2001	34076	30313

Uns wurden Verkehrsdaten der Rosengartenstrasse in Zürich zur Auswertung übergeben. Richtung 1 gibt die Verkehrsbelastung zum Bucheggplatz, Richtung 2 die Belastung zum Escher-Wyss-Platz an.



Aufgabe E.14

Der Zusammenhang der Verkehrsdaten für die beiden Richtungen soll mit Hilfe einer linearen Regression bestimmt werden.

a) Erstelle ein Regressionmodell für die Abhängigkeit zwischen den Daten der beiden Richtungen: Bestimme die Regressions-koeffizienten sowie die Varianz des Residualwertes und der geschätzten Modellparameter. Verwende hierzu nur die ersten 10 Datenpaare.

b) Das Regressionsmodell soll nun mit neuen Daten aktualisiert werden. Bestimme die a posteriori Regressionskoeffizienten und quantifiziere die Unsicherheit des aktualisierten Regressionsmodells. Verwende hierzu die zweite Hälfte des Datensatzes.

Aufgabe E.14

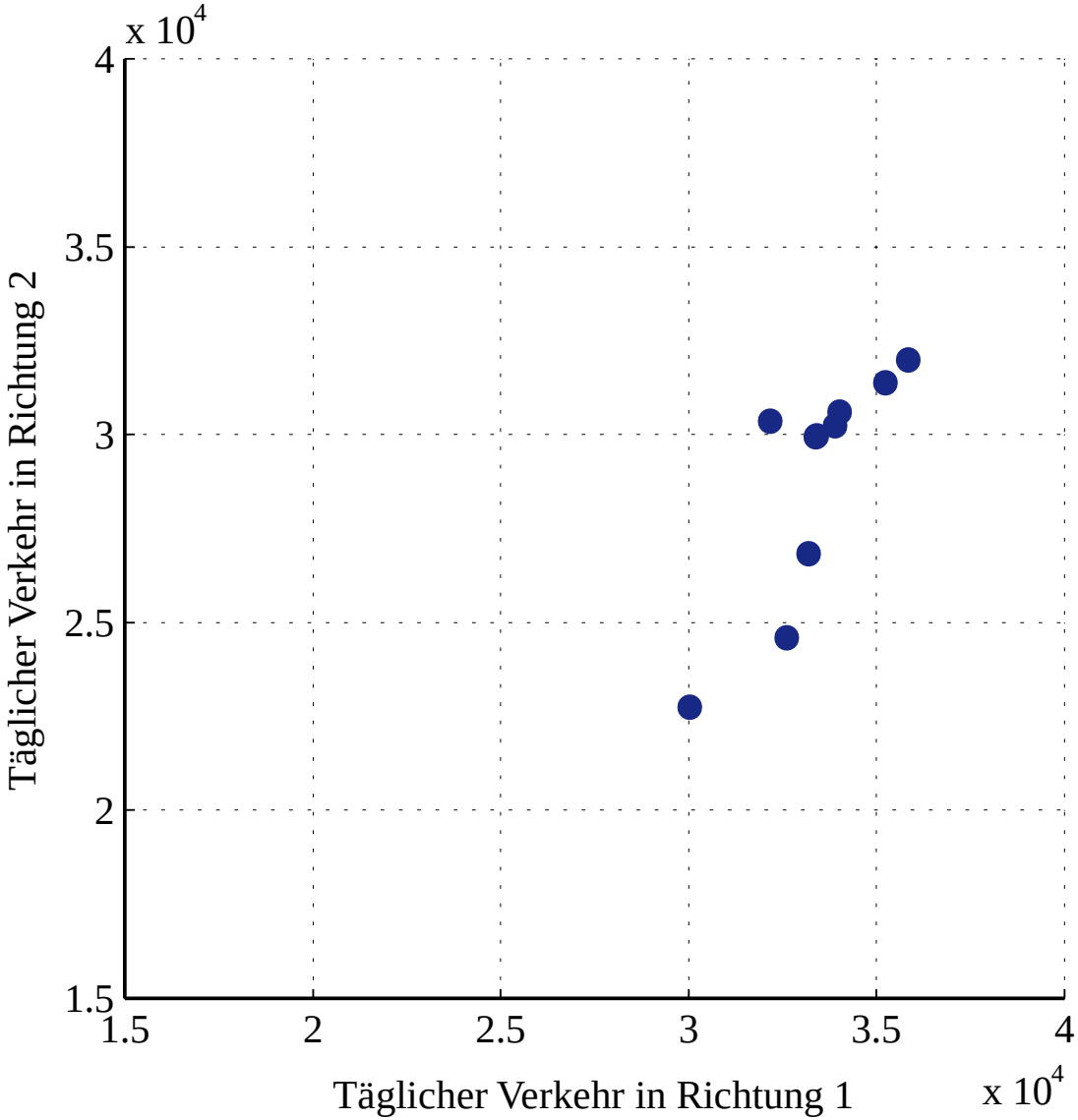
Der Zusammenhang der Verkehrsdaten für die beiden Richtungen soll mit Hilfe einer linearen Regression bestimmt werden.

Regressionsgerade:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

X Verkehr Richtung 1

Y Verkehr Richtung 2



Aufgabe E.14

a) Erstelle ein Regressionmodell für die Abhängigkeit zwischen den Daten der beiden Richtungen: Bestimme die Regressions-koeffizienten sowie die Varianz des Residualwertes und der geschätzten Modellparameter.

Formeln für die Regressionskoeffizienten:

$$\beta_0 = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i - \beta_1 \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i = \bar{y} - \beta_1 \bar{x}$$

$$\beta_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i\hat{x}_i - \bar{y}\frac{1}{n}\sum_{i=1}^{n}\hat{x}_i}{\frac{1}{n}\sum_{i=1}^{n}\hat{x}_i^2 - \bar{x}\frac{1}{n}\sum_{i=1}^{n}\hat{x}_i} = \frac{\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i\hat{x}_i - \bar{y}\bar{x}}{\frac{1}{n}\sum_{i=1}^{n}\hat{x}_i^2 - \bar{x}^2} = \frac{s_{XY}}{s_X^2}$$

Aufgabe E.14

Mit den Verkehrsdaten vom 1. bis zum 10. April 2001 ergeben sich die folgenden Werte:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} \hat{x}_i = 33'377 \qquad \frac{1}{n}\sum_{i=1}^{n} \hat{x}_i^2 = 1'116'363'757$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i = 28'883 \qquad \frac{1}{n}\sum_{i=1}^{n} \hat{x}_i \hat{y}_i = 967'681'266$$

$$\Rightarrow \beta_1 = \frac{\frac{1}{n}\sum_{i=1}^{n} \hat{y}_i \hat{x}_i - \bar{y}\bar{x}}{\frac{1}{n}\sum_{i=1}^{n} \hat{x}_i^2 - \bar{x}^2} = \frac{967'681'266 - 28'883 \cdot 33'377}{1'116'363'757 - (33'377)^2} = 1.554$$

$$\Rightarrow \beta_0 = \bar{y} - \beta_1 \bar{x} = 28'883 - 1.554 \cdot 33'377 = -22'986$$

Aufgabe E.14

Die Regressionsgerade hat die folgende Form:

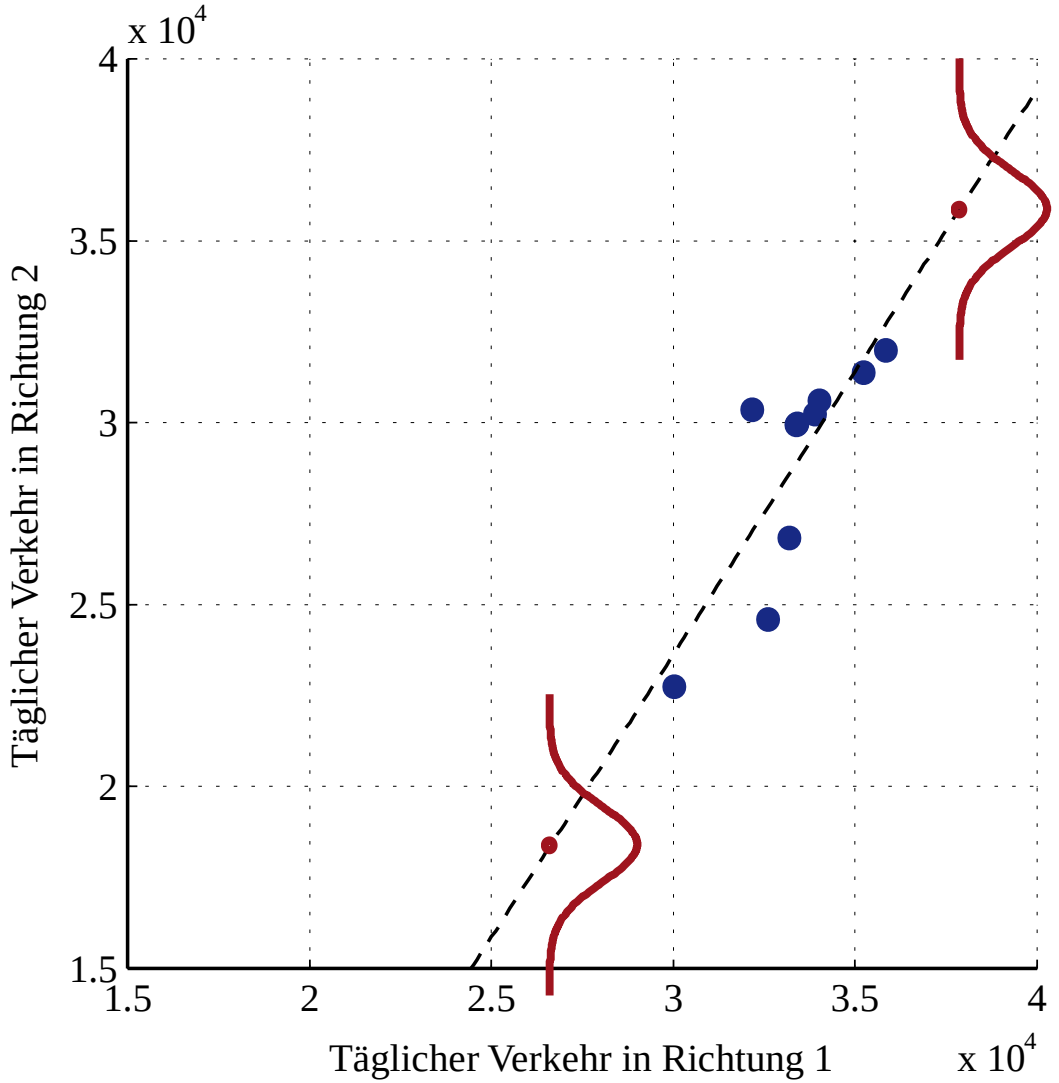
$$Y = -22'986 + 1.554X + \varepsilon$$

Der Zusammenhang ist jedoch nicht deterministisch:

$$Y|X \sim N(\beta_0 + \beta_1 X, \sigma_\varepsilon)$$

$$\varepsilon \sim N(0, \sigma_\varepsilon)$$

Residualwert, Fehler



Aufgabe E.14

Anhand der Summe der quadratischen Fehler kann man beurteilen, wie gut das Regressionsmodell die Daten abbilden kann.

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(\hat{y}_i - \left(\beta_0 + \beta_1 \hat{x}_i \right) \right)^2 = 28'814'936$$

Berechnung der Standardabweichung des Fehlers:

$$\sigma_\varepsilon = \sqrt{\frac{\sum_{i=1}^{n} \varepsilon_i^2}{n-k}} = \sqrt{\frac{28'814'936}{10-2}} = \sqrt{3\acute{}601\acute{}867} = 1897.9$$

Anzahl Daten

Anzahl Parameter

Aufgabe E.14

Auf das gleiche Ergebnis kommt man mit der Matrix-Schreibweise:

Schätzung der Regressionskoeffizienten:

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \left(\hat{\mathbf{X}}^T \hat{\mathbf{X}}\right)^{-1} \hat{\mathbf{X}}^T \hat{\mathbf{y}}$$

Quantifizierung der statistischen Unsicherheit:

$$\sigma_\varepsilon^2 = \left(\hat{\mathbf{y}} - \hat{\mathbf{X}}\boldsymbol{\beta}\right)^T \left(\hat{\mathbf{y}} - \hat{\mathbf{X}}\boldsymbol{\beta}\right) / \left(n - k\right)$$

$$Cov\left(\boldsymbol{\beta}\right) = \sigma_\varepsilon^2 \mathbf{V}_\beta \qquad \mathbf{V}_\beta = \left(\hat{\mathbf{X}}^T \hat{\mathbf{X}}\right)^{-1}$$

$$\hat{\mathbf{X}} = \begin{pmatrix} 1 & 32618 \\ 1 & 33380 \\ 1 & 34007 \\ 1 & 33888 \\ 1 & 35237 \\ 1 & 35843 \\ 1 & 33197 \\ 1 & 30035 \\ 1 & 32158 \\ 1 & 33406 \end{pmatrix} \qquad \hat{\mathbf{y}} = \begin{pmatrix} 24609 \\ 29965 \\ 30629 \\ 30263 \\ 31405 \\ 31994 \\ 26846 \\ 22762 \\ 30366 \\ 29994 \end{pmatrix}$$

Aufgabe E.14

$$\hat{\mathbf{X}}^T\hat{\mathbf{X}} = \begin{pmatrix} 1 & \cdots & 1 \\ \hat{x}_1 & \cdots & \hat{x}_n \end{pmatrix} \begin{pmatrix} 1 & \hat{x}_1 \\ \vdots & \vdots \\ 1 & \hat{x}_i \\ \vdots & \vdots \\ 1 & \hat{x}_n \end{pmatrix} = \begin{pmatrix} n \cdot 1^2 & \sum_{i=1}^{n} \hat{x}_i \\ \sum_{i=1}^{n} \hat{x}_i & \sum_{i=1}^{n} \hat{x}_i^2 \end{pmatrix} = \begin{pmatrix} 10 & 333'769 \\ 333'769 & 11'163'637'569 \end{pmatrix}$$

$$\mathbf{V}_\beta = \left(\hat{\mathbf{X}}^T\hat{\mathbf{X}}\right)^{-1} = \begin{pmatrix} 10 & 333'769 \\ 333'769 & 11'163'637'569 \end{pmatrix}^{-1} = \begin{pmatrix} 47.58 & -1.42 \cdot 10^{-3} \\ -1.42 \cdot 10^{-3} & 4.26 \cdot 10^{-8} \end{pmatrix}$$

Aufgabe E.14

$$\hat{\mathbf{X}}^T\hat{\mathbf{y}} = \begin{pmatrix} 1 & \cdots & 1 \\ \hat{x}_1 & \cdots & \hat{x}_n \end{pmatrix} \begin{pmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} \hat{y}_i \\ \sum_{i=1}^{n} \hat{x}_i \hat{y}_i \end{pmatrix} = \begin{pmatrix} 288'833 \\ 9'676'812'660 \end{pmatrix}$$

Nun lassen sich die Regressionskoeffizienten bestimmen:

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \left(\hat{\mathbf{X}}^T\hat{\mathbf{X}}\right)^{-1}\hat{\mathbf{X}}^T\hat{\mathbf{y}}$$

$$= \begin{pmatrix} 47.58 & -1.42\cdot 10^{-3} \\ -1.42\cdot 10^{-3} & 4.26\cdot 10^{-8} \end{pmatrix} \begin{pmatrix} 288'833 \\ 9'676'812'660 \end{pmatrix} = \begin{pmatrix} -22'986 \\ 1.554 \end{pmatrix}$$

Aufgabe E.14

Die Fehlervarianz berechnet sich wie zuvor:

$$\sigma_\varepsilon^2 = \frac{1}{n-k}\left(\hat{\mathbf{y}} - \hat{\mathbf{X}}\boldsymbol{\beta}\right)^T\left(\hat{\mathbf{y}} - \hat{\mathbf{X}}\boldsymbol{\beta}\right) = \frac{1}{n-k}\sum_{i=1}^{n}\left(\hat{y}_i - \left(\beta_0 + \beta_1\hat{x}_i\right)\right)^2$$
$$= 3'601'867$$

Hieraus ergibt sich die Kovarianz-Matrix der Modellparameter:

$$Cov(\boldsymbol{\beta}) = \sigma_\varepsilon^2\mathbf{V}_\beta = 3'601'867\begin{pmatrix} 10 & 333'769 \\ 333'769 & 11'163'637'569 \end{pmatrix}$$
$$= \begin{pmatrix} 1.71\cdot 10^8 & -5.12\cdot 10^3 \\ -5.12\cdot 10^3 & 0.154 \end{pmatrix} = \begin{pmatrix} Var[\beta_0] & Cov[\beta_0,\beta_1] \\ Cov[\beta_0,\beta_1] & Var[\beta_1] \end{pmatrix}$$

Aufgabe E.14

Regressionsmodell:

$$Y \mid X \sim N\left(\beta_0 + \beta_1 X, \sigma_\varepsilon\right)$$

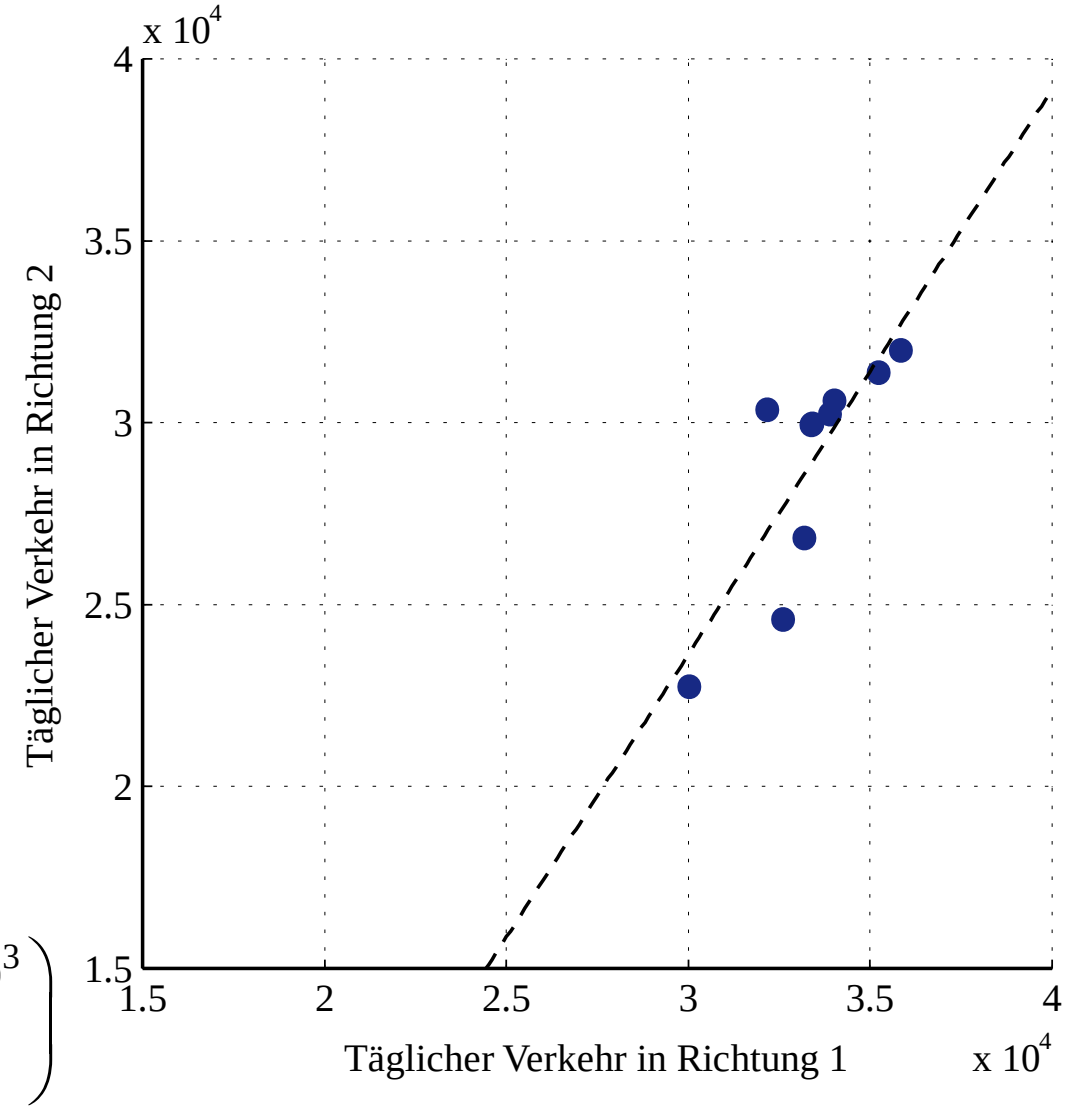
Regressionskoeffizienten:

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} -22'986 \\ 1.554 \end{pmatrix}$$

Unsicherheit des Modells:

$$\sigma_\varepsilon = 1897.9$$

$$Cov(\boldsymbol{\beta}) = \begin{pmatrix} 1.71 \cdot 10^8 & -5.12 \cdot 10^3 \\ -5.12 \cdot 10^3 & 0.154 \end{pmatrix}$$



Aufgabe E.14

b) Das Regressionsmodell soll nun mit neuen Daten aktualisiert werden. Bestimme die a posteriori Regressionskoeffizienten und quantifiziere die Unsicherheit des aktualisierten Regressionsmodells. Verwende hierzu die zweite Hälfte des Datensatzes.

$$\boldsymbol{\beta}'' = \mathbf{V}_\beta'' \left(\left(\mathbf{V}_\beta' \right)^{-1} \boldsymbol{\beta}' + \hat{\mathbf{X}}_n{}^T \hat{\mathbf{y}}_n \right) \qquad \left(\mathbf{V}_\beta'' \right)^{-1} = \left(\mathbf{V}_\beta' \right)^{-1} + \hat{\mathbf{X}}_n^T \hat{\mathbf{X}}_n$$

A posteriori Modell — A priori Modell — neue Daten

Die Informationen aus dem bestehenden a priori Regressionsmodell und den neuen Daten werden für das a posteriori Modell kombiniert.

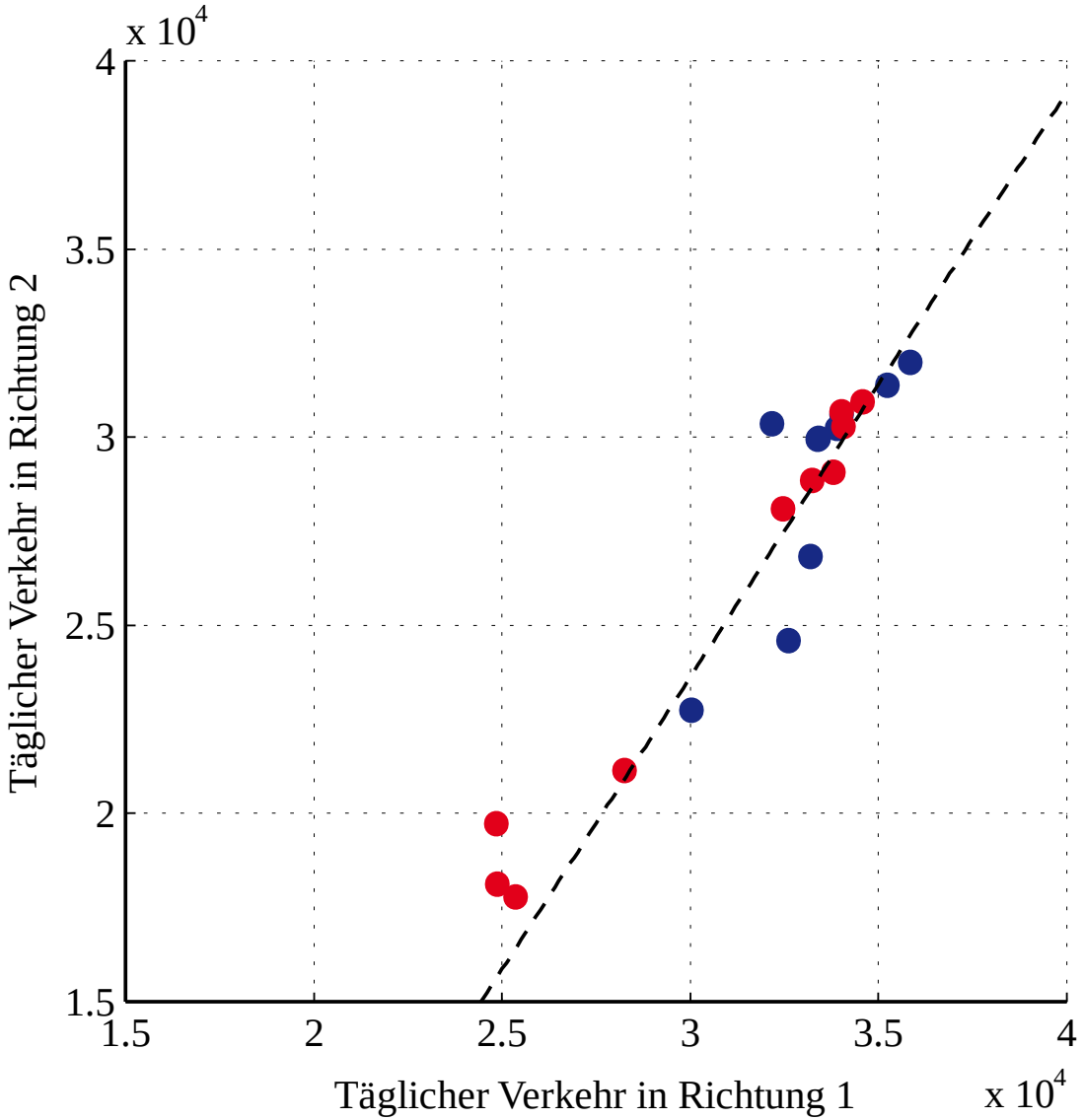
Aufgabe E.14

Das a priori Modell aus Teilaufgabe a) soll mit neuen Daten aktualisiert werden.

Das a posteriori Modell wird folgendermassen bestimmt:

$$\boldsymbol{\beta}'' = \mathbf{V}_\beta''\left(\left(\mathbf{V}_\beta'\right)^{-1}\boldsymbol{\beta}' + \hat{\mathbf{X}}_n^T\hat{\mathbf{y}}_n\right)$$

$$\left(\mathbf{V}_\beta''\right)^{-1} = \left(\mathbf{V}_\beta'\right)^{-1} + \hat{\mathbf{X}}_n^T\hat{\mathbf{X}}_n$$



Aufgabe E.14

Zunächst berechnen wir $\hat{\mathbf{X}}_n^T \hat{\mathbf{X}}_n$ und $\hat{\mathbf{X}}_n^T \hat{\mathbf{y}}_n$ für die neuen Daten:

$$\hat{\mathbf{X}}_n = \begin{pmatrix} 1 & 34576 \\ 1 & 34013 \\ 1 & 24846 \\ 1 & 28252 \\ 1 & 25365 \\ 1 & 24862 \\ 1 & 32472 \\ 1 & 33245 \\ 1 & 33788 \\ 1 & 34076 \end{pmatrix} \qquad \hat{\mathbf{y}}_n = \begin{pmatrix} 30958 \\ 30680 \\ 19735 \\ 21145 \\ 17805 \\ 18123 \\ 28117 \\ 28858 \\ 29080 \\ 30313 \end{pmatrix}$$

$$\hat{\mathbf{X}}_n^T \hat{\mathbf{X}}_n = \begin{pmatrix} n \cdot 1^2 & \sum_{i=1}^{n} \hat{x}_i \\ \sum_{i=1}^{n} \hat{x}_i & \sum_{i=1}^{n} \hat{x}_i^2 \end{pmatrix} = \begin{pmatrix} 10 & 305'495 \\ 305'495 & 9'491'848'963 \end{pmatrix}$$

$$\hat{\mathbf{X}}_n^T \hat{\mathbf{y}}_n = \begin{pmatrix} \sum_{i=1}^{n} \hat{y}_i \\ \sum_{i=1}^{n} \hat{x}_i \hat{y}_i \end{pmatrix} = \begin{pmatrix} 254'814 \\ 7'991'745'111 \end{pmatrix}$$

Aufgabe E.14

Die neuen Daten werden mit dem a priori Modell kombiniert:

$$\left(\mathbf{V}_\beta''\right)^{-1} = \overbrace{\left(\mathbf{V}_\beta'\right)^{-1}}^{\text{A priori}} + \overbrace{\hat{\mathbf{X}}_n^T\hat{\mathbf{X}}_n}^{\text{neu}} = \begin{pmatrix} 10 & 333'769 \\ 333'769 & 11'163'637'569 \end{pmatrix} + \begin{pmatrix} 10 & 305'495 \\ 305'495 & 9'491'848'963 \end{pmatrix}$$

$$= \begin{pmatrix} 20 & 639'264 \\ 639'264 & 20'655'486'532 \end{pmatrix}$$

$$\mathbf{V}_\beta'' = \begin{pmatrix} 20 & 639'264 \\ 639'264 & 20'655'486'532 \end{pmatrix}^{-1} = \begin{pmatrix} 4.64 & -1.44\cdot10^{-4} \\ -1.44\cdot10^{-4} & 4.49\cdot10^{-9} \end{pmatrix}$$

Nun können die a posteriori Regressionskoeffizienten bestimmt werden:

$$\boldsymbol{\beta}'' = \mathbf{V}_\beta''\left(\left(\mathbf{V}_\beta'\right)^{-1}\boldsymbol{\beta}' + \hat{\mathbf{X}}_n^T\hat{\mathbf{y}}\right) = \begin{pmatrix} -14'733 \\ 1.311 \end{pmatrix} = \begin{pmatrix} \beta_0'' \\ \beta_1'' \end{pmatrix}$$

Aufgabe E.14

Das a priori Modell wurde mit den neuen Daten aktualisiert.

Die a posteriori Regressionsgerade hat die folgende Form:

$$Y = -14'733 + 1.311X + \varepsilon$$