

Modul E – Lösung

Aufgabe E.1

1. Formuliere die Null- und die Alternativhypothese

Unsere Nullhypothese H_0 lautet, dass der Mittelwert μ der Grundgesamtheit gleich 30 MPa ist. Die Alternativhypothese H_1 lautet dann einfach: $\mu \neq 30$ MPa.

$$H_0: \mu = 30 \text{ MPa} \qquad H_1: \mu \neq 30 \text{ MPa}$$

2. Wähle ein Signifikanzniveau α

$$\alpha = 10\% .$$

3. Definiere die Art der Stichprobenerhebung (welche Daten, wie viele Daten?)

15 Stichproben der Druckfestigkeit der Betonproduktion des betreffenden Tages.

4. Formuliere eine Entscheidungsregel

$$P(30 - \Delta \leq \bar{X} \leq 30 + \Delta) = 1 - \alpha$$

5. Berechne

$$P(30 - \Delta \leq \bar{X} \leq 30 + \Delta) = 1 - \alpha \Rightarrow$$

$$P(30 - \Delta \leq \bar{X} \leq 30 + \Delta) = 0.9 \Rightarrow 2\Phi\left(\frac{30 + \Delta - 30}{\sigma_x / \sqrt{n}}\right) - 1 = 0.9 \Rightarrow \Phi\left(\frac{\Delta}{\sqrt{16.36/15}}\right) = 0.95$$

\bar{X} ist die Stichprobenstatistik. In der Tabelle der Standardnormalverteilung kann der z-Wert nachgeschaut werden (Anhang T, Tabelle T.1):

$$\frac{\Delta}{\sqrt{16.36/15}} = 1.645 \Rightarrow \Delta = 1.72 .$$

Folglich: Wenn den Stichprobenmittelwert \bar{x} der 15 Proben im Intervall liegt:

$[28.28 \text{ MPa} \leq \bar{x} \leq 31.72 \text{ MPa}]$, dann kann die Nullhypothese nicht zurückgewiesen werden auf einem Signifikanzniveau von 10%.

Stichprobenmittelwert: 32.25 MPa.

6. Beurteile die Nullhypothese H_0

Da 32.25 MPa ausserhalb des berechneten Intervalls liegt, wird die Nullhypothese verworfen.

Das gleiche Vorgehen wird für den Hypothesentest mit dem Signifikanzniveau von 1% durchgeführt. Das erhaltene Intervall ist $[27.31MPa \leq \bar{X} \leq 32.69MPa]$. Da der Stichprobenmittelwert sich innerhalb dieses Intervalls befindet, kann die Nullhypothese nicht verworfen werden.

Aufgabe E.3

In einem Prüflabor werden jeden Tag 30 Messungen durchgeführt, um die Qualität des Wassers zu kontrollieren. Jedes Messergebnis folgt einer Normalverteilung mit einem Mittelwert $\mu = 23 \text{ ng/ml}$ und einer Standardabweichung $\sigma = 4.3 \text{ ng/ml}$.

- a) Mit welcher Wahrscheinlichkeit wird eine Messung einen Wert kleiner als 23 ng/ml aufweisen?

$$P[X \leq 23] = P\left[\frac{X - 23}{4.3} \leq \frac{23 - 23}{4.3}\right] = \Phi(0) = 0.5$$

- b) Mit welcher Wahrscheinlichkeit liegt eine Messung im Intervall $[19.5 \text{ ng/ml}; 20.5 \text{ ng/ml}]$?

$$P[19.5 < X \leq 20.5] = P\left[\frac{19.5 - 23.0}{4.3} < \frac{X - 23.0}{4.3} \leq \frac{20.5 - 23.0}{4.3}\right] = \Phi(-0.58) - \Phi(-0.81) = 0.073$$

- c) Mit welcher Wahrscheinlichkeit ergibt sich ein Tagesmittelwert von kleiner als 20 ng/ml ?

$$P(\bar{X} < 20) = P\left(\frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} < \frac{20 - \mu_X}{\sigma_X / \sqrt{n}}\right) = \Phi\left(\frac{20 - 23}{0.79}\right) = 7.3 \cdot 10^{-5}$$

- d) Das Labor hat ein neues Instrument für die Messungen der Wasserqualität erworben und führt Eichmessungen durch, d.h. es vergleicht die Messwerte des neuen Instruments mit dem alten Instrument. Es werden 15 Proben mit dem neuen Instrument gemessen, und deren Mittelwert liegt bei $\bar{x} = 19 \text{ ng/ml}$, mit einer Standardabweichung von $s = 5 \text{ ng/ml}$. Teste auf einem 5% Signifikanzniveau, ob die neuen Messwerte zur gleichen Grundgesamtheit gehören wie die alten Messwerte.

Testen des Mittelwertes mit unbekannter Varianz \rightarrow t-Verteilung

1. Festlegen der Nullhypothese

$$H_0: \mu_N = 23 \text{ ng/ml}$$

$$H_1: \mu_N \neq 23 \text{ ng/ml}$$

2. Festlegen der operativen Regel:

$$\mu_A - t_{\alpha/2} \frac{s_N}{\sqrt{n}} \leq \bar{x}_N \leq \mu_A + t_{\alpha/2} \frac{s_N}{\sqrt{n}} \quad \rightarrow \text{zweiseitiger Hypothesentest}$$

3. Auswahl des Signifikanzniveaus:

gegeben, $\alpha = 0.05$.

4. Berechnung der Akzeptanzkriterien

Berechnen der oberen und unteren Intervallgrenzen:

$$\mu_A - t_{\alpha/2} \frac{s_N}{\sqrt{n}} = 23 - 2.145 \cdot \frac{5}{\sqrt{15}} \qquad \mu_A + t_{\alpha/2} \frac{s_N}{\sqrt{n}} = 23 + 2.145 \cdot \frac{5}{\sqrt{15}}$$

Das Intervall der t-Statistik entspricht demzufolge [20.23 ng/ml; 25.77 ng/ml].

5. Durchführen des Tests:

Der Stichprobenmittelwert von $\bar{x}_N = 19 \text{ ng/ml}$ liegt ausserhalb dieses Intervalls.

6. Rückschluss auf dem gewählten Signifikanzniveau:

Die Nullhypothese, dass es sich um die gleichen Populationen handelt, wird auf einem Signifikanzniveau von 5% verworfen.

Aufgabe E.4

1. Festlegen der Nullhypothese

$$H_0 : \mu = 23.7 \quad H_1 : \mu \neq 23.7$$

2. Festlegen der operativen Regel

$$\mu_X - \Delta \leq \bar{X} \leq \mu_X + \Delta$$

$$\mu_X - k_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} \leq \bar{X} \leq \mu_X + k_{\alpha/2} \frac{\sigma_X}{\sqrt{n}}$$

wobei σ die Standardabweichung der Fahrzeit ist, und n die Stichprobenanzahl.

3. Wähle ein Signifikanzniveau α

$$\alpha = 5\% .$$

4. Berechnung der Akzeptanzkriterien

$$\mu - 1.96 \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 1.96 \frac{\sigma}{\sqrt{n}}$$

$$23.7 - 1.96 \frac{3}{\sqrt{13}} \leq \bar{X} \leq 23.7 + 1.96 \frac{3}{\sqrt{13}}$$

$$22.07 \leq \bar{X} \leq 25.33$$

5. Durchführung des Tests und Überprüfung der Akzeptanz

$$\bar{x} = 22.3 \text{ Minuten}$$

Der Stichprobenmittelwert befindet sich im Intervall $[22.07 \leq \bar{x} \leq 25.33]$.

6. Rückschluss auf gewähltem Signifikanzniveau

Die Nullhypothese kann auf einem Signifikanzniveau von 5% nicht verworfen werden.

Aufgabe E.5

Aus Verkehrszählungen liegt eine Datenserie vor, die den täglichen Verkehrsfluss in der Rosengartenstrasse in Zürich beschreibt (Tabelle E.5.1).

Tabelle E.5.1: Anzahl der beobachteten Fahrzeuge pro Tag, sortiert.

Tag (i)	Anzahl Fahrzeuge
1	3600
2	4500
3	5400
4	6500
5	7000
6	7500
7	8700
8	9000
9	9500

- a) Erstelle ein Wahrscheinlichkeitspapier für die folgende Dichtefunktion:

$$f_x(x) = \begin{cases} \frac{2}{10000^2} x & 0 \leq x \leq 10000 \\ 0 & \text{sonst} \end{cases}$$

Durch Integration erhält man die Verteilungsfunktion:

$$F_x(x) = \begin{cases} 0 & x \leq 0 \\ \left(\frac{x}{10000}\right)^2 & 0 < x \leq 10000 \\ 1 & x > 10000 \end{cases}$$

Eine lineare Funktion von x erhält man, indem man auf beiden Seiten die Wurzel nimmt:

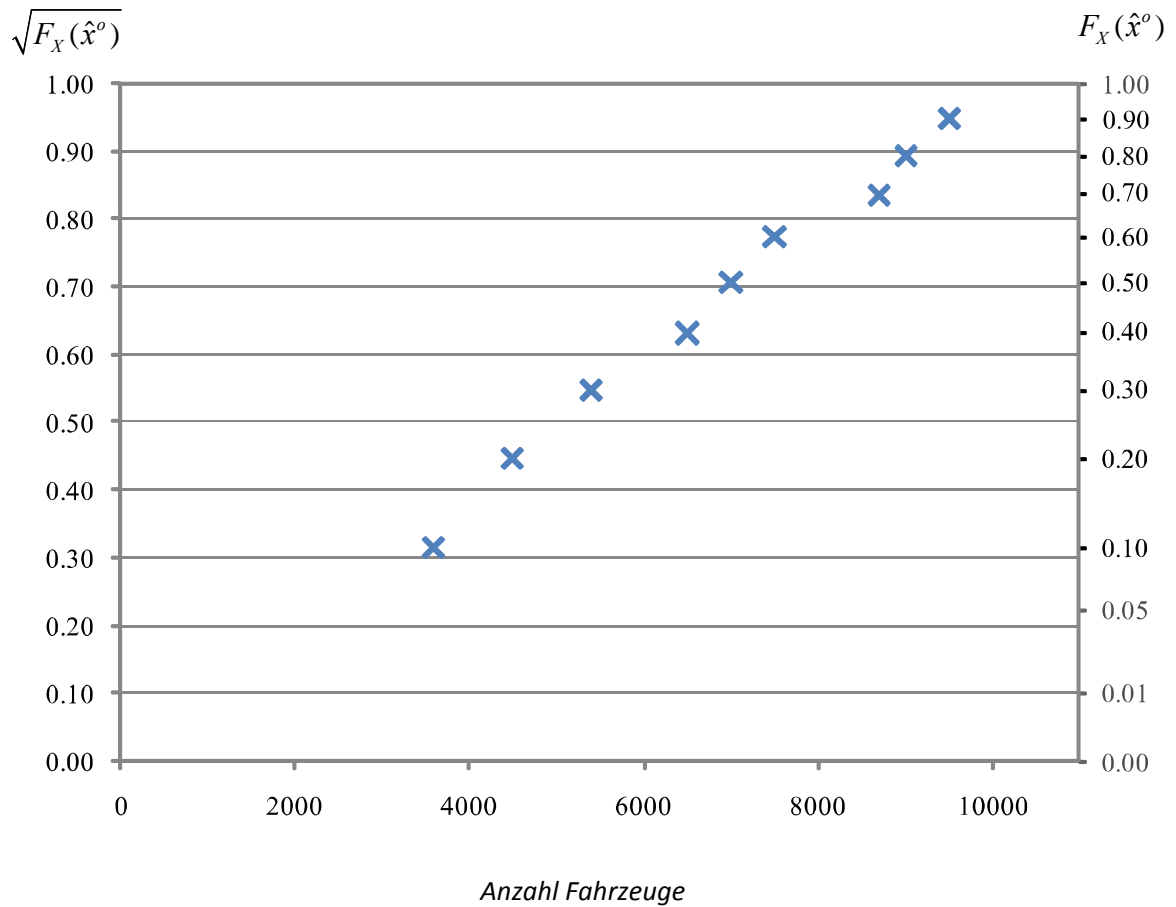
$$F_x(x) = \left(\frac{x}{10000}\right)^2 \Leftrightarrow \sqrt{F_x(x)} = \frac{x}{10000}$$

Für die Werte der beobachteten kumulativen Wahrscheinlichkeitsverteilungsfunktion $F_X(\hat{x}^o) = \frac{i}{N+1}$

im Intervall $[0;1]$ erhält man die folgende Tabelle:

Beobachtung i	Anz. Autos (x-Achse)	$F_X(\hat{x}^o)$	$\sqrt{F_X(\hat{x}^o)}$
		0	0
1	3600	0.1	0.32
2	4500	0.2	0.45
3	5400	0.3	0.55
4	6500	0.4	0.63
5	7000	0.5	0.71
6	7500	0.6	0.77
7	8700	0.7	0.84
8	9000	0.8	0.89
9	9500	0.9	0.95
		1.0	

Nun kann damit ein Wahrscheinlichkeitspapier erstellt werden, indem die y-Achse neu skaliert wird. Die Werte werden dann in das Papier eingetragen. Wenn die Daten auf einer geraden Linie liegen, dann kann angenommen werden, dass sie der gegebenen Verteilung folgen.



- b) Überprüfe mit Hilfe des erstellten Wahrscheinlichkeitspapiers, ob der tägliche Verkehrsfluss mit dieser Verteilung angenommen werden kann.

Die Daten liegen etwa auf einer Geraden, somit kann der Verteilungstyp akzeptiert werden.

Aufgabe E.6

Zur Dimensionierung eines Parkhauses wurden die Ankunftszeiten von Fahrzeugen aufgenommen. Die Zeitdifferenzen zwischen ankommenden Fahrzeugen sind in Tabelle E.6.1 gegeben.

- a) Zeichne das Wahrscheinlichkeitspapier für die Exponentialverteilung und trage die Werte der Zeitdifferenzen ein. Beurteile, ob die Intervalle einer Exponentialverteilung folgen.
- b) Bestimme den Erwartungswert der Zeitdifferenz grafisch aus der in a) erzeugten Grafik unter der Annahme, dass die Daten exponentialverteilt sind. Berechne den Erwartungswert und vergleiche die beiden Werte.

Verwende für die beobachtete kumulative Wahrscheinlichkeitsverteilungsfunktion: $F_x(\hat{x}_i^o) = \frac{i}{N+1}$

Tabelle E.6.1: Zeitdifferenz zwischen den Ankunftszeiten der Fahrzeuge.

i	Zeitdifferenz \hat{x}_i^o (Sekunden)
1	1.52
2	6.84
3	9.12
4	10.64
5	15.20
6	21.28
7	30.40
8	30.40
9	34.20
10	60.80
11	78.28
12	95.76

a) Die Wahrscheinlichkeitsverteilungsfunktion der Exponentialfunktion ist:

$$F(t) = 1 - \exp(-\lambda t)$$

Wir versuchen nun, eine lineare Beziehung zwischen x und y zu finden. Mit dem natürlichen Logarithmus erhält man eine lineare Beziehung zwischen x und y :

$$\ln(1 - F(t)) = -\lambda t$$

$(t, \ln(1 - F(t)))$ stellt eine Gerade dar mit der Steigung $-\lambda$. Die Annahme, dass die Daten exponentialverteilt sind, wird unter Verwendung eines Wahrscheinlichkeitspapiers beurteilt, siehe Abbildung E.6.1.

No.	Interval [s]	$F(t_i)$	$1-F(t_i)$	$\ln(1-F(t))$
1	1.52	0.077	0.923	-0.080
2	6.84	0.154	0.846	-0.167
3	9.12	0.231	0.769	-0.262
4	10.64	0.308	0.692	-0.368
5	15.2	0.385	0.615	-0.486
6	21.28	0.462	0.538	-0.619
7	30.4	0.538	0.462	-0.773
8	30.4	0.615	0.385	-0.956
9	34.2	0.692	0.308	-1.179
10	60.8	0.769	0.231	-1.466
11	78.28	0.846	0.154	-1.872
12	95.76	0.923	0.077	-2.565

Tabelle E.6.1: Rechentabelle für die Erstellung des Wahrscheinlichkeitspapiers.

Aufgrund der Abbildung E.6.1 darf behauptet werden, dass die Daten einer Exponentialverteilung folgen.

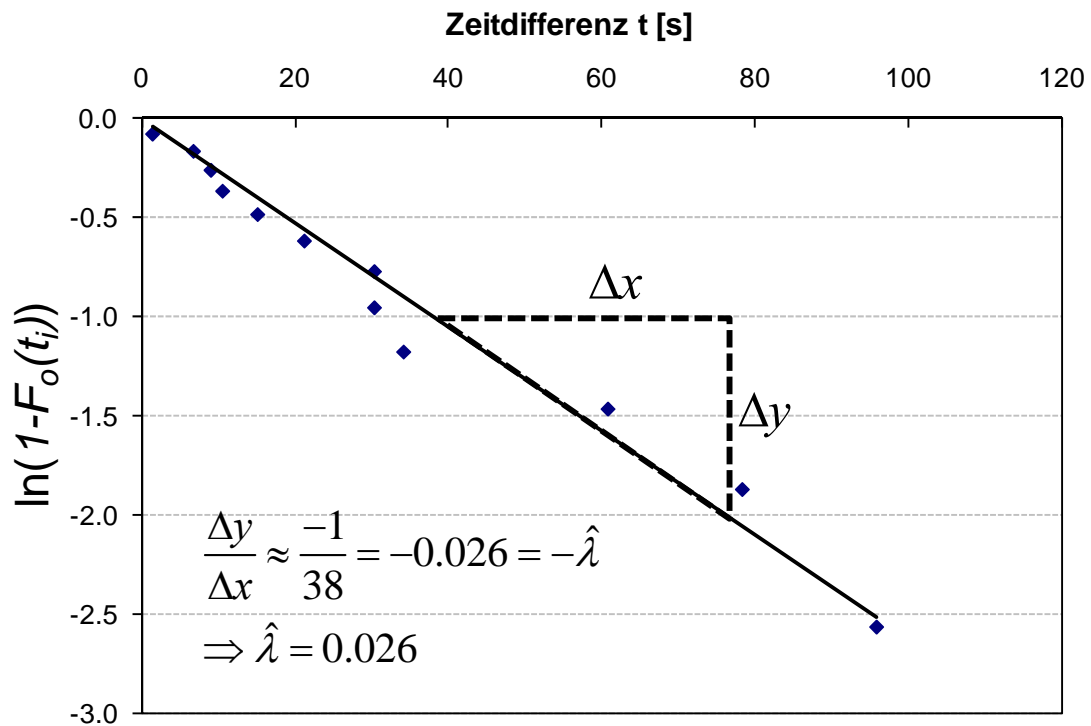
b) Der Stichprobenmittelwert wird berechnet als:

$$\bar{x} = \frac{1}{12}(1.52 + 6.84 + 9.12 + 10.64 + 15.2 + 21.28 + 30.4 + 30.4 + 34.2 + 60.8 + 78.28 + 95.76) = 32.87$$

Da die negative Steigung der Geraden in der Abbildung dem Parameter λ entspricht, kann der Parameter als ca. 0.026 geschätzt werden. Der Mittelwert wird dann gemäss der folgenden Beziehung hergeleitet:

$$\frac{1}{\hat{\lambda}} = \frac{1}{0.026} = 38.46 \text{ [s]}$$

Abb. E.6.1: Wahrscheinlichkeitspapier für die Exponentialfunktion, und eingetragene Beobachtungen der Zeitdifferenzen zwischen den Ankunftszeiten.



Aufgabe E.7

Um die Druckfestigkeit von Beton einer bestimmten Produktionsmenge zu modellieren, wurden 20 Stichproben gemessen. Die Ergebnisse sind in der Tabelle dargestellt. Es wird angenommen, dass die Grundgesamtheit der Stichproben einer Normalverteilung $N(\mu, \sigma)$ folgt.

Tabelle 9.4: Messwerte der Druckfestigkeit von den Betonproben.

Nr. der Messung	Druckfestigkeit [N/mm ²]	Nr. der Messung	Druckfestigkeit [N/mm ²]
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

Schätze die unbekannt Parameter (μ, σ) anhand der Momentenmethode.

$$m_1 = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = \mu \quad m_2 = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = \sigma^2 + \mu^2$$

$$\mu = m_1 = \frac{1}{n} \sum_{i=1}^n \hat{x}_i \quad \sigma^2 = m_2 - m_1^2 = \frac{1}{n} \sum_{i=1}^n \hat{x}_i^2 - \left(\frac{1}{n} \sum_{i=1}^n \hat{x}_i\right)^2$$

$$\mu = 32.67$$

$$\sigma = 4.04$$

Aufgabe E.8

Gleiche Aufgabenstellung wie E.7.

a) Beschreibe die Likelihood-Funktion.

Die Likelihood-Funktion lautet allgemein: $L(\boldsymbol{\theta} | \hat{x}_1, \hat{x}_2, \dots, \hat{x}_n) = \prod_{i=1}^n f_X(\hat{x}_i | \boldsymbol{\theta})$

und für die Normalverteilung: $L(\mu, \sigma | \hat{\mathbf{x}}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\hat{x}_i - \mu)^2}{2\sigma^2}\right]$

Anstelle der Likelihood-Funktion empfiehlt es sich, die Log-Likelihood-Funktion zu verwenden:

$$\begin{aligned} l = \ln L &= \sum_{i=1}^n \ln \left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\hat{x}_i - \mu)^2}{2\sigma^2}\right] \right] \\ &= -n \ln(\sqrt{2\pi}) - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (\hat{x}_i - \mu)^2 \end{aligned}$$

b) Schätze die unbekannt Parameter (μ, σ) anhand der MLM.

$$\mu = \frac{1}{n} \sum_{i=1}^n \hat{x}_i = 32.67 \quad \sigma^2 = \frac{1}{n} \sum_{i=1}^n (\hat{x}_i - \mu)^2 = 4.04$$

c) Schätze die Parameter der Exponentialverteilung mit der MLM und vergleiche die kumulative Verteilungsfunktion mit den beobachteten Werten.

Die Likelihood-Funktion ist:

$$L = \prod_{i=1}^n \lambda e^{-\lambda \hat{x}_i}$$

Die Log-Likelihood-Funktion ist:

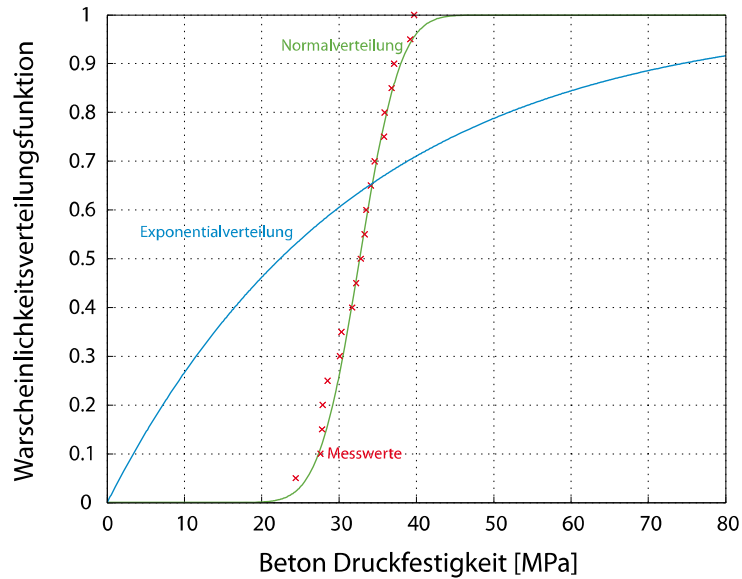
$$l = \ln L = \sum_{i=1}^n \ln(\lambda e^{-\lambda \hat{x}_i}) = n \ln \lambda - \lambda \sum_{i=1}^n \hat{x}_i$$

Den MLM Schätzwert für den Parameter der Exponentialverteilung erhält man aus:

$$\frac{\partial l}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^n \hat{x}_i = 0$$

und somit einen Wert von: $\lambda = \frac{n}{\sum_{i=1}^n \hat{x}_i} = 0.031$

Kumulative Verteilungsfunktionen:



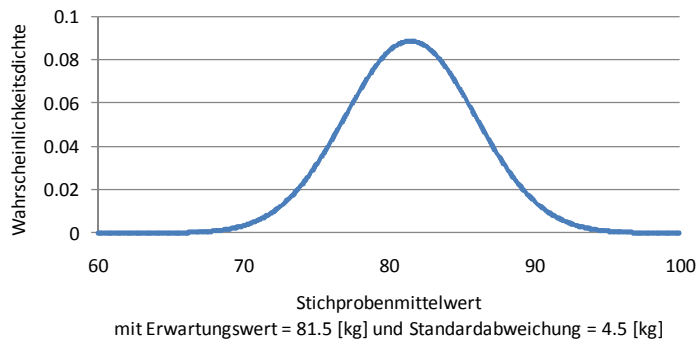
Aufgabe E.9

- a) Schätze anhand der Stichprobe den Erwartungswert $E(\bar{X})$ und die Varianz $Var(\bar{X})$ des Stichprobenmittelwertes \bar{X} .

$$E[\bar{X}] = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{4} 326 = 81.5$$

$$Var[\bar{X}] = \frac{1}{n} \sigma_X^2 = \frac{1}{4} 81 = 20.25$$

- b) Stelle die Streuung des Mittelwertes grafisch dar.



- c) Ermittle den Bereich, in dem der wahre Mittelwert μ_X mit 95% Konfidenz zu erwarten ist.

$$P \left[-k_{\alpha/2} < \frac{\bar{X} - \mu_X}{\sigma_X \frac{1}{\sqrt{n}}} < k_{\alpha/2} \right] = 1 - \alpha$$

$$P \left[\bar{X} - k_{\alpha/2} \sigma_X \frac{1}{\sqrt{n}} < \mu_X < \bar{X} + k_{\alpha/2} \sigma_X \frac{1}{\sqrt{n}} \right] = 1 - \alpha$$

$$P \left[81.5 - k_{\alpha/2} 9 \frac{1}{\sqrt{4}} < \mu_X < 81.5 + k_{\alpha/2} 9 \frac{1}{\sqrt{4}} \right] = 0.95$$

$$P \left[81.5 - 1.96 \cdot 9 \frac{1}{\sqrt{4}} < \mu_X < 81.5 + 1.96 \cdot 9 \frac{1}{\sqrt{4}} \right] = 0.95$$

$$\underline{\underline{P[72.68 < \mu_X < 90.32] = 0.95}}$$

Aufgabe E.10

- a) Berechne das Konfidenzintervall für die Schätzung des Stichprobenmittelwertes \bar{X} nach 15 Produktionstagen bei gegebenem $\alpha = 0.05$.

$$P \left[-k_{\alpha/2} < \frac{\bar{X} - \mu_X}{\sigma_X \frac{1}{\sqrt{n}}} < k_{\alpha/2} \right] = 1 - \alpha$$

$$P \left[-k_{\alpha/2} \sigma_X \frac{1}{\sqrt{n}} < \bar{X} - \mu_X < k_{\alpha/2} \sigma_X \frac{1}{\sqrt{n}} \right] = 1 - \alpha$$

$$P \left[-1.96 < \frac{\bar{X} - \mu_X}{1430 \frac{1}{\sqrt{150}}} < 1.96 \right] = 1 - 0.05$$

$$P \left[-228.85 < \bar{X} - \mu_X < 228.85 \right] = 0.95$$

$$P \left[\bar{X} - 228.85 < \mu_X < \bar{X} + 228.85 \right] = 0.95$$

- b) Berechne aus a) das Konfidenzintervall für einen gegebenen Stichprobenmittelwert von $\bar{x} = 11'000$ [MPa].

$$P[\bar{X} - 228.85 < \mu_X < \bar{X} + 228.85] = 0.95$$

$$P[11'000 - 228.85 < \mu_X < 11'000 + 228.85] = 0.95$$

$$P[10'771.15 < \mu_X < 11'228.85] = 0.95$$

Das Konfidenzintervall für den wahren Mittelwert liegt zwischen 10'771.15 [MPa] und 11'228.85 [MPa].

- c) Wie viele Stichprobenentnahmen wären notwendig, um zu zeigen, dass der Mittelwert aller Stichproben mit mindestens $\alpha = 0.01$ im gleichen Konfidenzintervall wie unter b) berechnet liegt?

$$P\left[\bar{X} - k_{\alpha/2} \sigma_X \frac{1}{\sqrt{n}} < \mu_X < \bar{X} + k_{\alpha/2} \sigma_X \frac{1}{\sqrt{n}}\right] = 1 - \alpha$$

$$\rightarrow \bar{X} - k_{\alpha/2} \sigma_X \frac{1}{\sqrt{n}} = 10'771.15$$

$$\rightarrow 11'000 - 2.60 \cdot 1430 \frac{1}{\sqrt{n}} = 10'771.15$$

$$\rightarrow n = \left(\frac{1}{0.0616}\right)^2$$

$$\rightarrow \underline{\underline{n = 263.53 \rightarrow n \geq 264}}$$

Aufgabe E.11

Die Druckfestigkeit von 30 Holzproben parallel zur Faser wurde bestimmt. Die Ergebnisse sind in Tabelle 10.1 gegeben.

Tabelle 11.1: Druckfestigkeit von Holzproben.

Nr.	Druckfestigkeit [MPa]	Nr.	Druckfestigkeit [MPa]	Nr.	Druckfestigkeit [MPa]
1	12.8	11	23.4	21	29.3
2	16.3	12	26.8	22	29.5
3	16.6	13	26.9	23	30.3
4	16.9	14	27	24	32.1
5	17.2	15	27.1	25	32.3
6	17.9	16	27.2	26	33.5
7	19.5	17	27.2	27	33.9
8	21.9	18	27.5	28	35.6
9	22.3	19	27.9	29	39.2
10	22.5	20	28.3	30	43.5

- a) Passe die Exponentialverteilung und die Weibullverteilung den Daten an. Bestimme dazu die Parameter dieser Verteilungen mit der Methode der Momente.

Das erste und zweite Stichprobenmoment sind:

$$m_1 = 26.41$$

$$m_2 = 747.55$$

Die Exponentialverteilung hat die folgende kumulative Verteilungsfunktion:

$$F_X(x) = 1 - \exp(-\lambda x), \quad x > 0$$

Mit dem Mittelwert $\mu = \frac{1}{\lambda}$

So kann der Parameter λ geschätzt werden zu: $\hat{\lambda} = \frac{1}{m_1} = 0.038$

Die (zweiparametrische) Weibullverteilung hat die folgende kumulative Verteilungsfunktion:

$$F_X(x) = 1 - \exp\left(-\left(\frac{x}{u}\right)^k\right), \quad x > 0$$

mit dem Mittelwert und der Standardabweichung:

$$\mu = u\Gamma\left(1 + \frac{1}{k}\right); \quad \sigma = u\sqrt{\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)} \quad (\text{Siehe Skript Tabelle D.2})$$

Die unbekannt Parameter u und k können wie folgt bestimmt werden:

Parameter k :

$$\frac{\sigma}{\mu} = \frac{\sqrt{m_2 - m_1^2}}{m_1} = \frac{u\sqrt{\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)}}{u\Gamma\left(1 + \frac{1}{k}\right)} = \frac{\sqrt{\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)}}{\Gamma\left(1 + \frac{1}{k}\right)} = \sqrt{\frac{\Gamma(1+2/k) - \Gamma^2(1+1/k)}{\Gamma^2(1+1/k)}} - 1$$

$$\text{Mit } \Gamma(p) = \int_0^{\infty} x^{p-1} \cdot e^{-x} dx.$$

Da $\Gamma^2(1+1/k)$ nicht so einfach zu lösen ist, empfiehlt sich hier ein numerisches/approximatives Vorgehen (z.B. Excel Solver). So erhält man:

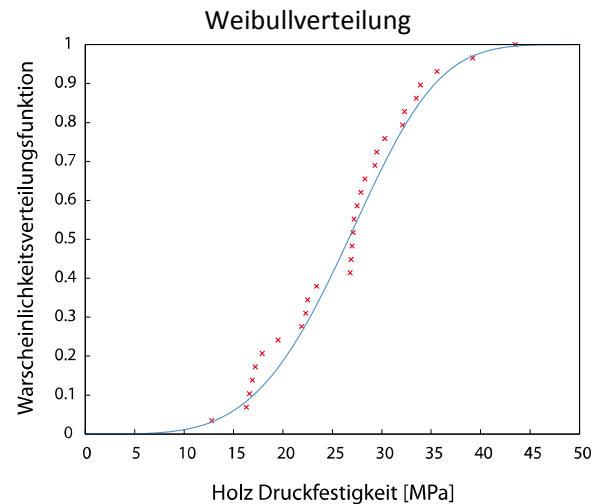
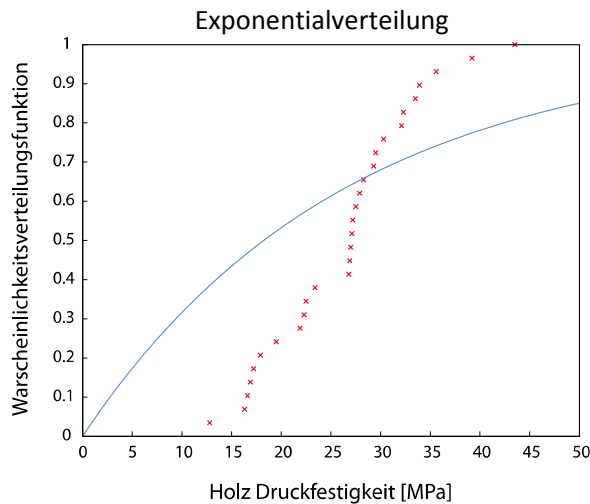
$$k \approx \left(\frac{\sigma}{\mu}\right)^{-1.09} = 4.21 \quad \text{siehe Skript, Tabelle D2}$$

Parameter u :

$$\mu = u\Gamma\left(1 + \frac{1}{k}\right)$$

$$u = \frac{m_1}{\Gamma(1+1/k)} = \frac{m_1}{\Gamma(1.24)} = 29.05$$

b) Zeichne die kumulative Verteilungsfunktion der beiden Verteilungen und zeichne jeweils die kumulative Verteilung der Stichprobe ein.



c) **Teste die Güte der Anpassung für die Verteilungen mit dem χ^2 Test auf einem Signifikanzniveau von 10%. Benutze dazu die Intervalle in Tabelle 2.**

Die Statistik für den χ^2 – Test ist: $\epsilon_m^2 = \sum_{i=1}^k \frac{(N_i - n \cdot p_i)^2}{n \cdot p_i}$

wobei N_i die Anzahl der Beobachtungen im Intervall i , n die gesamte Anzahl Beobachtungen und p_i die erwartete Wahrscheinlichkeit (berechnet aus der gewählten Verteilungsfunktion) sind.

Die Entscheidungsregel lautet dann: Wenn $P[\epsilon^2 \leq \chi] = 1 - \alpha$ erfüllt ist kann die Hypothese angenommen werden.

für die Exponentialverteilung:

ϵ^2 folgt der Chi-Quadrat Verteilung mit 2 FHG.

Wir haben 4 Intervalle. Das letzte Intervall ist abhängig von den 3 anderen – Reduktion um 1 FHG.

Bestimmung des Parameters λ aus den beobachteten Daten – Reduktion um 1 FHG.

$4 - 1 - 1 = 2$ FHG und $\alpha=0.1 \sim \chi = 4.6052$

$\epsilon_m^2 = \sum_{i=1}^k \frac{(N_i - n \cdot p_i)^2}{n \cdot p_i} = 46.52$ last sich mit Hilfe Tabelle 10.2 berechnen:

Tabelle 10.2 Intervalle für den χ^2 Test

Intervall	Häufigkeit N_i	Wahrscheinlichkeit P [Stichprobe in diesem Intervall]	Erwartete Häufigkeit	Normalisierte Quadrate der Differenzen
0-20	7	0.532	15.96	5.03
20-25	4	0.081	2.43	1.01
25-30	11	0.067	2.01	40.21
30-	8	0.32	9.60	0.27
Summe	30	1	30	46.52

Da $\varepsilon^2 = 46.52$ grösser als $\chi = 4.61$ ist, müssen wir die Hypothese, dass die Daten einer Exponentialverteilung folgen, auf einem Signifikanzniveau von 10% verwerfen.

für die Weibullverteilung:

ε^2 folgt der Chi-Quadrat Verteilung mit 1 FHG.

$\nu = m - 1 - j = 4 - 1 - 2 = 1$ FHG und $\alpha = 0.1 \sim \chi = 2.7055$

$\varepsilon_m^2 = \sum_{i=1}^k \frac{(N_i - n \cdot p_i)^2}{n \cdot p_i} = 2.36$ lässt sich mit Hilfe Tabelle 11.3 berechnen:

Tabelle 11.3 Intervalle für den χ^2 Test

Intervall	Häufigkeit	Wahrscheinlichkeit P [Stichprobe in diesem Intervall]	Erwartete Häufigkeit	Normalisierte Quadrate der Differenzen
0-20	7	0.19	5.7	0.296
20-25	4	0.22	6.6	1.024
25-30	11	0.27	8.1	1.038
30-	8	0.32	9.6	0.267
Summe	30	1	30	2.63

Da $\varepsilon^2 = 2.36$ kleiner als $\chi^2 = 2.71$ ist, können wir die Hypothese, dass die Daten einer Weibullverteilung folgen, auf einem Signifikanzniveau von 10% nicht verwerfen.

Zu Beachten:

Der χ^2 Test hängt stark von der Wahl der Anzahl und der Grössen der Klassen ab.
Jede Klasse sollte mindestens 5 Beobachtungen enthalten, um ein plausibles Resultat zu erlangen.

- d) Wir nehmen nun an, dass wir die Parameter für die Verteilungen aus a) aus der Literatur entnommen haben.
Teste die Güte der Anpassung für die Verteilungen mit dem Kolmogorov-Smirnov Test auf einem Signifikanzniveau von 10%.

für die Exponentialverteilung:

Die Stichprobenstatistik für den Kolmogorov-Smirnov Test berechnet sich mit Hilfe des Berechnungsblattes in Tabelle 10.4 und ist:

$$\varepsilon_{\max} = \max_{i=1}^n \left[\left| F_o(x_i^o) - F_p(x_i^o) \right| \right] = 0.4 .$$

Da $\varepsilon_{\max} = 0.4$ grösser als der Tabellenwert $c=0.218$ ist, müssen wir die Hypothese, dass die Daten einer Exponentialverteilung folgen, auf einem Signifikanzniveau von 10% verwerfen.

Table 11.4: Berechnungsblatt für den Kolmogorov-Smirnov Test.

Nr.	[Mpa]	$F_X(x)$	i/n	$ i/n - F_X(x) $
1	12.80	0.39	0.03	0.35
2	16.30	0.46	0.07	0.40
3	16.60	0.47	0.10	0.37
4	16.90	0.47	0.13	0.34
5	17.20	0.48	0.17	0.31
6	17.90	0.49	0.20	0.29
7	19.50	0.52	0.23	0.29
8	21.90	0.56	0.27	0.30
9	22.30	0.57	0.30	0.27
10	22.50	0.57	0.33	0.24
11	23.40	0.59	0.37	0.22
12	26.80	0.64	0.40	0.24
13	26.90	0.64	0.43	0.21
14	27.00	0.64	0.47	0.17
15	27.10	0.64	0.50	0.14
16	27.20	0.64	0.53	0.11
17	27.20	0.64	0.57	0.08
18	27.50	0.65	0.60	0.05
19	27.90	0.65	0.63	0.02
20	28.30	0.66	0.67	0.01
21	29.30	0.67	0.70	0.03
22	29.50	0.67	0.73	0.06
23	30.30	0.68	0.77	0.08
24	32.10	0.70	0.80	0.10
25	32.30	0.71	0.83	0.13
26	33.50	0.72	0.87	0.15
27	33.90	0.72	0.90	0.18
28	35.60	0.74	0.93	0.19
29	39.20	0.77	0.97	0.19
30	43.50	0.81	1.00	0.19

Weibullverteilung:

Die Stichprobenstatistik für den Kolmogorov-Smirnov Test berechnet sich mit Hilfe des Berechnungsblattes in Tabelle 10.5 und ist:

$$\varepsilon_{\max} = \max_{i=1}^n \left[\left| F_o(x_i^o) - F_p(x_i^o) \right| \right] = 0.11.$$

Da $\varepsilon_{\max} = 0.11$ grösser als der Tabellenwert $c=0.218$ ist, müssen wir die Hypothese, dass die Daten einer Exponentialverteilung folgen, auf einem Signifikanzniveau von 10% annehmen.

Table 11.5: Berechnungsblatt für den Kolmogorov-Smirnov Test.

Nr.	[Mpa]	$F_X(x)$	i/n	$ i/n - F_X(x) $
1	12.8	0.03	0.03	0.00
2	16.3	0.08	0.07	0.02
3	16.6	0.09	0.10	0.01
4	16.9	0.10	0.13	0.04
5	17.2	0.10	0.17	0.06
6	17.9	0.12	0.20	0.08
7	19.5	0.17	0.23	0.06
8	21.9	0.26	0.27	0.00
9	22.3	0.28	0.30	0.02
10	22.5	0.29	0.33	0.04
11	23.4	0.33	0.37	0.04
12	26.8	0.51	0.40	0.11
13	26.9	0.51	0.43	0.08
14	27	0.52	0.47	0.05
15	27.1	0.53	0.50	0.03
16	27.2	0.53	0.53	0.00
17	27.2	0.53	0.57	0.04
18	27.5	0.55	0.60	0.05
19	27.9	0.57	0.63	0.06
20	28.3	0.59	0.67	0.07
21	29.3	0.65	0.70	0.05
22	29.5	0.66	0.73	0.08
23	30.3	0.70	0.77	0.07
24	32.1	0.78	0.80	0.02
25	32.3	0.79	0.83	0.04
26	33.5	0.84	0.87	0.03
27	33.9	0.85	0.90	0.05
28	35.6	0.90	0.93	0.03
29	39.2	0.97	0.97	0.00
30	43.5	1.00	1.00	0.00

Aufgabe E.12

a) Passe eine Normalverteilung den gegebenen Beobachtungen an. Verwende dazu den Mittelwert aus der Literatur und bestimme den Parameter σ mit der Maximum Likelihood Methode.

$$\text{Normalverteilung: } f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$$\text{Likelihood: } L(\theta|\hat{\mathbf{x}}) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\hat{x}_i-\mu}{\sigma}\right)^2\right)$$

$$\theta_1 = \hat{\mu}_X = 168.921$$

Somit muss nur der zweite Parameter der Normalverteilung (Standardabweichung) anhand der Beobachtungen abgeschätzt werden:

$$\theta_2 = \hat{\sigma}_X = \sqrt{\frac{\sum_{i=1}^n (\hat{x}_i - \theta_1)^2}{n}} = 6.4048$$

b) Teste die Güte der Anpassung, für die Normalverteilung mit dem χ^2 Test auf einem Signifikanzniveau von 10%. Benutze dazu die Intervalle in Tabelle 4

ε^2 folgt der Chi-Quadrat Verteilung mit 3 FHG.

Das letzte Intervall ist abhängig von den 4 anderen – Reduktion um 1 FHG.

Bestimmung der Parameter μ und σ aus den beobachteten Daten – Reduktion um 2 FHG.

5 – 1 – 2 = 2 FHG und $\alpha=0.1 \sim \chi=4.6052$

$$\varepsilon_m^2 = \sum_{i=1}^k \frac{(N_i - n \cdot p_i)^2}{n \cdot p_i} = 2.6090 \text{ last sich mit Hilfe Tabelle 12.1 berechnen:}$$

Tabelle 12.1 Intervalle für den χ^2 Test

Intervall	Häufigkeit	Wahrscheinlichkeit P [Stichprobe in diesem Intervall]	Erwartete Häufigkeit	Normalisierte Quadrate der Differenzen
[0-160]	5	0.0819	3.1121	1.1453
]160-165]	9	0.1884	7.1580	0.4740
]165-170]	8	0.2966	11.2709	0.9492
]170-175]	10	0.2618	9.9475	0.0003
]175- ∞ [6	0.1714	6.5115	0.0402
Summe	38	1		2.6090

Mit dem aus der Tabelle abgelesenen Wert für $\chi = 4.6052$ folgt:

Da $\varepsilon^2 = 2.6090$ kleiner ist als $\chi^2 = 4.6052$, können wir die Hypothese, dass die Daten einer Normalverteilung folgen, auf einem Signifikanzniveau von 10% annehmen.

Aufgabe E.13

Druckfestigkeit von Beton: Normalverteilte Zufallsvariable X mit unsicherem Mittelwert μ_X und bekannter Standardabweichung $\sigma_X = 10 \text{MPa}$.

A priori Verteilung:

Mittelwert μ_X normalverteilt

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

Daten:

Stichprobe von 20 Versuchskörpern: $n = 20$

Stichprobenmittelwert der Druckfestigkeit: $\bar{x} = 32.665 \text{MPa}$

Auf Basis der Daten soll die Verteilung des Mittelwertes aktualisiert werden.

Bestimmung der a posteriori Verteilung des Mittelwertes:

Da die gewählten Verteilungsmodelle konjugierte Verteilungen sind, ist die a posteriori Verteilung ebenfalls eine Normalverteilung.

Gewichtung („Stichprobengrösse“) der a priori Information:

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

Berechnung der a posteriori Verteilungsparameter für die Verteilung von μ_X :

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

$$\sigma_{\mu_X}'' = \sqrt{\frac{\frac{\sigma_X^2 \cdot \sigma_{\mu_X}^{\prime 2}}{n'} + \frac{\sigma_{\mu_X}^{\prime 2}}{n}}{\frac{1}{n'} + \frac{1}{n}}} = \sqrt{\frac{\frac{10^2 \cdot 3^2}{11.11} + \frac{3^2}{20}}{\frac{1}{11.11} + \frac{1}{20}}} = \underline{\underline{1.793MPa}}$$

Die a posteriori Verteilung des Mittelwertes μ_X ergibt sich somit zu:

$$\mu_X \sim N(\mu_{\mu_X}'' = 33.499MPa, \sigma_{\mu_X}'' = 1.793MPa)$$

Bestimmung der prädiktiven Verteilung der Betondruckfestigkeit:

Die a posteriori prädiktive Verteilung ist die neue Verteilung der Druckfestigkeit X nach dem Aktualisieren des Verteilungsparameters μ_X

$$f_X(x|\hat{\mathbf{x}}) = \int f_X(x|\mu_X, \sigma_X) \cdot f''(\mu_X|\hat{\mathbf{x}}) d\mu_X$$

Durch die Wahl von konjugierten Verteilungsmodellen ist die prädiktive Verteilung ebenfalls eine Normalverteilung mit den Parametern μ''' und σ''' . Diese berechnen sich folgendermassen:

$$\mu''' = \mu_{\mu_X}'' = 33.499MPa$$

$$\sigma'''^2 = \sigma_{\mu_X}''^2 + \sigma_X^2 = 1.793^2 + 10^2 = 103.214MPa^2$$

$$\sigma''' = 10.159MPa$$

Die a posteriori prädiktive Verteilung der Druckfestigkeit ergibt sich also zu:

$$\mu_X \sim N(\mu''' = 33.499MPa, \sigma''' = 10.159MPa)$$

Aufgabe E.14

a) Erstellung eines Regressionsmodells (auf Basis der ersten 10 Datenpaare):

Lineare Regression:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad \varepsilon \sim N(0, \sigma_\varepsilon)$$

X : Verkehrsbelastung in Richtung 1

Y : Verkehrsbelastung in Richtung 2

Zunächst werden aus den Daten vom 1. bis zum 10. April die folgenden Werte berechnet:

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^n \hat{x}_i = 33'377 & \frac{1}{n} \sum_{i=1}^n \hat{x}_i^2 &= 1'116'363'757 \\ \bar{y} &= \frac{1}{n} \sum_{i=1}^n \hat{y}_i = 28'883 & \frac{1}{n} \sum_{i=1}^n \hat{x}_i \hat{y}_i &= 967'681'266 \end{aligned}$$

Hieraus lassen sich die Regressionskoeffizienten β_0 und β_1 bestimmen:

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

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

Die Regressionsgerade hat also die folgende Form:

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

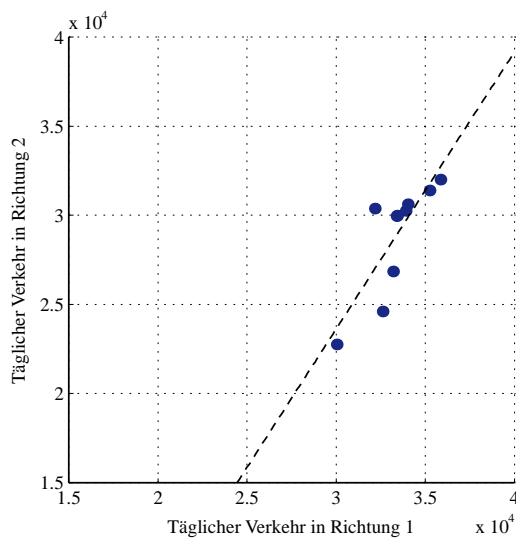


Abbildung E.14.1: A priori Regressionsmodell.

Gegeben einen Wert für X folgt die Zufallsvariable Y einer Normalverteilung:

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

Die Standardabweichung des Fehlers lässt sich folgendermassen bestimmen:

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

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

n ist die Anzahl der zur Schätzung verwendeten Daten

k ist die Anzahl der geschätzten Parameter (Regressionskoeffizienten)

Zur Bestimmung der Kovarianzmatrix zwischen den beide Regressionskoeffizienten muss zunächst die Matrix \mathbf{V}_β berechnet werden:

$$\hat{\mathbf{X}}^T \hat{\mathbf{X}} = \begin{pmatrix} 1 & \cdots & 1 \\ \hat{x}_1 & \cdots & \hat{x}_n \end{pmatrix} \begin{pmatrix} 1 & \hat{x}_1 \\ \vdots & \vdots \\ 1 & \hat{x}_i \\ \vdots & \vdots \\ 1 & \hat{x}_n \end{pmatrix} = \begin{pmatrix} n \cdot 1^2 & \sum_{i=1}^n \hat{x}_i \\ \sum_{i=1}^n \hat{x}_i & \sum_{i=1}^n \hat{x}_i^2 \end{pmatrix} = \begin{pmatrix} 10 & 333'769 \\ 333'769 & 11163'637'569 \end{pmatrix}$$

$$\mathbf{V}_\beta = (\hat{\mathbf{X}}^T \hat{\mathbf{X}})^{-1} = \begin{pmatrix} 10 & 333'769 \\ 333'769 & 11163'637'569 \end{pmatrix}^{-1} = \begin{pmatrix} 47.58 & -1.42 \cdot 10^{-3} \\ -1.42 \cdot 10^{-3} & 4.26 \cdot 10^{-8} \end{pmatrix}$$

Ausserdem fliesst die Standardabweichung des Fehlers in die Kovarianzmatrix ein:

$$\text{Cov}(\boldsymbol{\beta}) = \sigma_\varepsilon^2 \mathbf{V}_\beta = 3'601'867 \begin{pmatrix} 10 & 333'769 \\ 333'769 & 11163'637'569 \end{pmatrix} = \begin{pmatrix} 1.71 \cdot 10^8 & -5.12 \cdot 10^3 \\ -5.12 \cdot 10^3 & 0.154 \end{pmatrix}$$

b) Aktualisieren des Regressionsmodells (auf Basis der letzten 10 Datenpaare):

Das a posteriori Regressionsmodell lässt sich folgendermassen bestimmen:

A posteriori Regressionskoeffizienten:

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

Hierin ist

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

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

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

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

Diese Ergebnisse kombinieren wir mit den Angaben aus dem a priori Regressionsmodell (siehe Teilaufgabe a)):

$$\begin{aligned} (\mathbf{V}_\beta'')^{-1} &= (\mathbf{V}_\beta')^{-1} + \hat{\mathbf{X}}_n^T \hat{\mathbf{X}}_n = \begin{pmatrix} 10 & 333'769 \\ 333'769 & 11'163'637'569 \end{pmatrix} + \begin{pmatrix} 10 & 305'495 \\ 305'495 & 9'491'848'963 \end{pmatrix} \\ &= \begin{pmatrix} 20 & 639'264 \\ 639'264 & 20'655'486'532 \end{pmatrix} \end{aligned}$$

$$\mathbf{V}_\beta'' = \begin{pmatrix} 20 & 639'264 \\ 639'264 & 20'655'486'532 \end{pmatrix}^{-1} = \begin{pmatrix} 4.64 & -1.44 \cdot 10^{-4} \\ -1.44 \cdot 10^{-4} & 4.49 \cdot 10^{-9} \end{pmatrix}$$

Nun können die a posteriori Regressionskoeffizienten berechnet werden:

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

Die neue Regressionsgerade hat also die folgende Form:

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

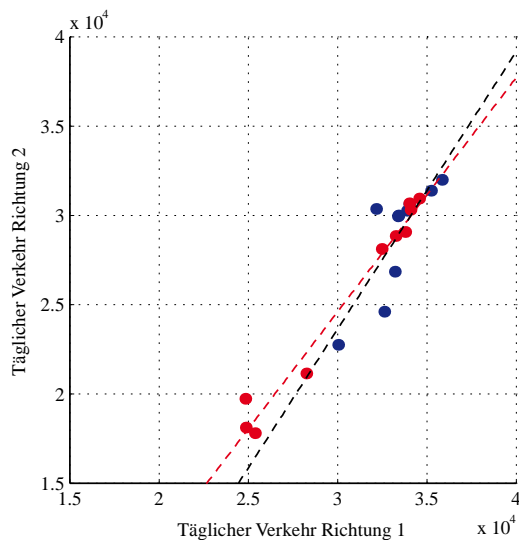


Abbildung E.14.1: A posteriori Regressionsmodell.