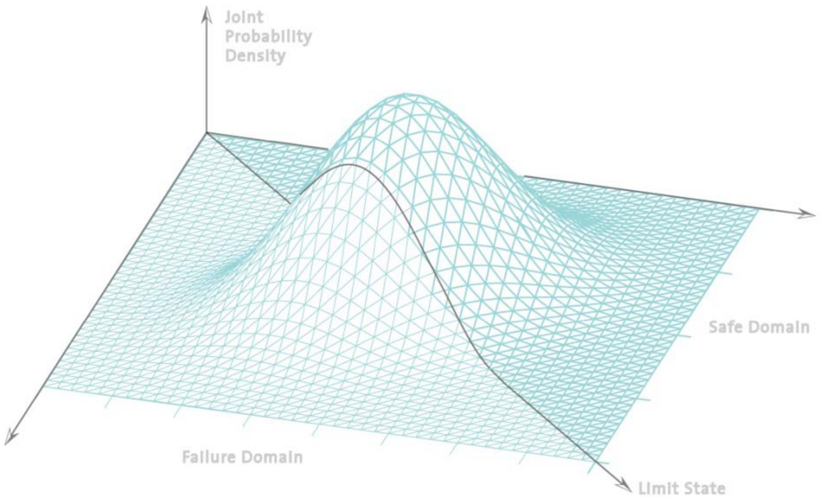


Statistik und Wahrscheinlichkeitsrechnung

Übung 10

Übung 11
am Dienstag, 25.05.2010

Start 8:00 Uhr

Markus HIL E 1
Mathias HIL E 6
Gerhard HCI D8
Katharina HCI D2

2. Teilprüfung am Donnerstag 27.Mai

Wann?

Dienstag, 27. Mai, 8:00 Uhr – 9:30 Uhr

Bitte um 7:45 Uhr am Hörsaal sein!

Wo?

Studierende mit Nachnamen A – J : HIL E 4

Studierende mit Nachnamen K – Z : HCI G 3

2. Teilprüfung am Donnerstag 27.Mai

Inhalt

Multiple Choice + 1 Übung zum Rechnen

Gesamter Stoff bis einschliesslich Vorlesung 11 und Übung 10.

Erlaubte Hilfsmittel

Alle Unterlagen erlaubt.

Taschenrechner (ohne Kommunikationsmittel) erlaubt, auch programmierbare.

Keine Kommunikationsmittel (z.B. Telefon) erlaubt.

Korrektur 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 :

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

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

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

Korrektur Aufgabe E.13

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

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

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

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

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

$$\longrightarrow \sigma''' = 10.159\,MPa$$

$$\mu''' = 33.499\,MPa$$

Inhalt der heutigen Übung

- Gruppenaufgabe E.1
- Aufgaben E.10: Stichprobenstatistiken / Konfidenzintervalle
- Aufgabe E.4: Hypothesentests
- Aufgabe E.11 – Fortsetzung (χ^2- und K.-S.-Test)
- Vorstellen der Gruppenaufgabe E.12 (Parameterschätzung)

Gruppenaufgabe E.1

Um die Qualität des Betons auf einer Baustelle zu prüfen, wird die Druckfestigkeit des hergestellten Betons getestet.

Erfahrungsgemäss ist bekannt, dass die Druckfestigkeit einer Normalverteilung folgt und die Varianz der Druckfestigkeit für diese Betonsorte 16.36 [MPa^2] ist.

Akzeptanzkriterium für die Qualität des Betons auf der Baustelle ist, dass der Mittelwert des Betons gleich $30 \pm \Delta \; [MPa]$ ist. Dies wird täglich am jeweils hergestellten Beton gemessen. Um die Homogenität der Verarbeitung zu gewährleisten, sind sowohl kleinere wie auch grössere Werte nicht akzeptabel.

Gruppenaufgabe E.1

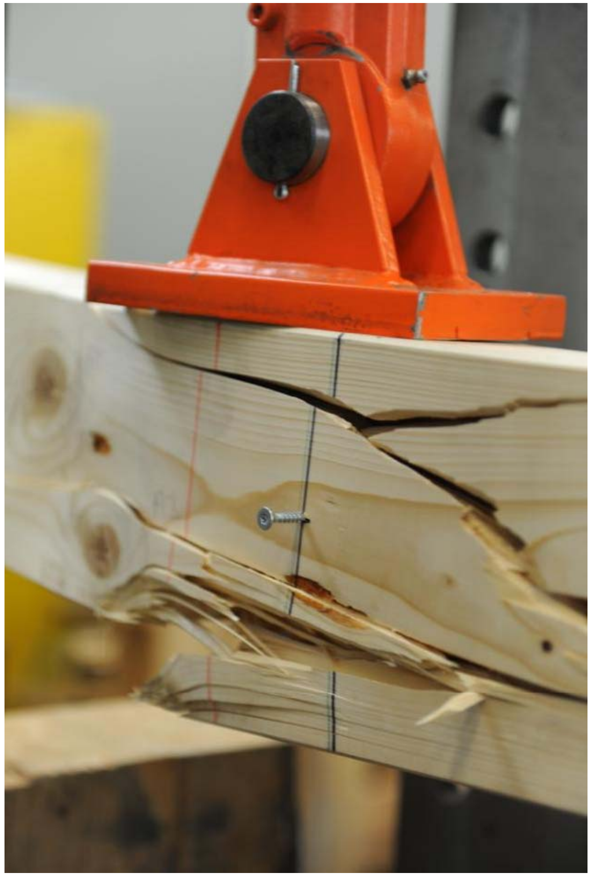
Nummer der Probe (i)	Druckfestigkeit [MPa]
1	24.4
2	26.5
3	27.8
4	29.2
5	39.2
6	37.8
7	35.1
8	30.8
9	30.3
10	39.7
11	38.4
12	33.3
13	33.5
14	28.1
15	34.6

Aus einer Tagesproduktion werden 15 Proben entnommen und ihre Druckfestigkeit getestet (siehe Tabelle)

Kann die Qualität des Betons akzeptiert werden?

Teste die Hypothese jeweils für ein Signifikanzniveau von 10 % und 1 %.

Aufgabe E.10



In einem Sägewerk wird das Schnittholz einer bestimmten Sortierklasse einer Qualitätskontrolle unterzogen. Pro Produktionstag wird eine Stichprobe mit zehn Brettern entnommen und jedes Brett auf seine Steifigkeit getestet. Der Mittelwert der Steifigkeit für jede Stichprobe ist normalverteilt und wird in Kontrolltabellen festgehalten. Basierend auf Erfahrung kann angenommen werden, dass die Standardabweichung der Steifigkeit unabhängig von der Qualität $\sigma = 1430 \text{ MPa}$ beträgt. Qualitätsschwankungen äussern sich nur in Form von Schwankungen des Mittelwertes.

Aufgabe E.10

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

b) Berechne aus a) das Konfidenzintervall für einen gegebenen Stichprobenmittelwert von $\overline{X} = 11'000$ [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?

Aufgabe E.10

a) Berechne das Konfidenzintervall für die Schätzung des Stichprobenmittelwertes nach 15 Produktionstagen.

Was ist gegeben?

$\sigma = 1430 \text{ MPa}$

15 Produktionstage → $n = 10 \cdot 15 = 150$

Stichprobenmittelwert normalverteilt

$\alpha = 0.05$

Lösungsansatz?

Für den Fall, dass der Mittelwert unsicher und die Varianz bekannt ist, ist das so genannte zweiseitige und symmetrische Konfidenzintervall des Mittelwertes gegeben durch:

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

Aufgabe E.10

a) Berechne das Konfidenzintervall für die Schätzung des Stichprobenmittelwertes nach 15 Produktionstagen.

Was ist gegeben?

$\sigma = 1430 \text{ MPa}$

15 Produktionstage → $n = 10 \cdot 15 = 150$

Stichprobenmittelwert normalverteilt

$\alpha = 0.05$

Lösungsansatz?

$$k_{\alpha/2} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = \Phi^{-1}\left(1 - \frac{0.05}{2}\right) = \Phi^{-1}(0.975) = \underline{\underline{1.96}}$$

Aufgabe E.10

a) Berechne das Konfidenzintervall für die Schätzung des Stichprobenmittelwertes nach 15 Produktionstagen.

Was ist gegeben?

$\sigma = 1430 \text{ MPa}$

15 Produktionstage → $n = 10 \cdot 15 = 150$

Stichprobenmittelwert normalverteilt

$\alpha = 0.05$

Lösungsansatz?

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

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

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

Aufgabe E.10

b) Berechne aus a) das Konfidenzintervall für einen beobachteten Stichprobenmittelwert von $\bar{x} = 11'000 MPa$

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

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

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

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

Aufgabe E.10

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?

Was ist gegeben?

$$\alpha = 0.01$$

$$\sigma = 1430 \text{ [MPa]}$$

$$\bar{x} = 11'000 \text{[MPa]}$$

$$\text{Konfidenzintervall } \left[10'771.15 < \mu_X < 11'228.85\right]$$

$$\text{Stichprobenmittelwert normalverteilt}$$

Aufgabe E.10

$$\alpha = 0.01$$
$$\sigma = 1430 \text{ [MPa]}$$
$$\bar{x} = 11'000 \text{[MPa]}$$
$$\text{Konfidenzintervall } \left[10'771.15 < \mu_X < 11'228.85\right]$$

Stichprobenmittelwert normalverteilt

Wie gross ist n?

Symmetrie

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

$$\text{Konfidenzintervall}\left[10'771.15 < \mu_X < 11'228.85\right]$$

ist gleich

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

?

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

Aufgabe E.10

$\alpha = 0.01$

$\sigma = 1430$ [MPa]

$\bar{x} = 11'000$[MPa]

Konfidenzintervall $[10'771.15 < \mu_X < 11'228.85]$

Stichprobenmittelwert normalverteilt

Wie gross ist n?

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

$$k_{\alpha/2} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = \Phi^{-1}\left(1 - \frac{0.01}{2}\right) = 2.60$$

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

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

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

Konfidenzintervall

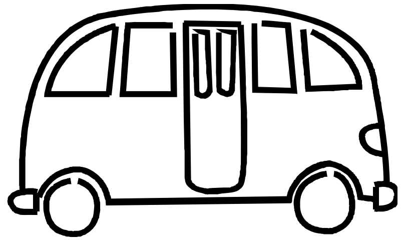
- Stichprobenstatistiken beinhalten Unsicherheiten – die Angabe von Intervallen ist notwendig, in denen ein Parameter in $(1-\alpha)\%$ der Fälle liegt.

Hypothesentest

- Dient der Aussage, ob aufgrund vollzogener Beobachtungen ein Wert signifikant unterschiedlich vom angenommenen Wert der Grundgesamtheit abweicht.

- Konfidenzintervalle können auch als Grundlage für Hypothesentests dienen (siehe E.4 als Beispiel dafür): Testen wir z.B. die Nullhypothese $\mu_X = 23.7$, dann ist das $(1-\alpha)$ Kriterium für das Verwerfen der Nullhypothese, dass das Konfidenzintervall den Wert 23.7 nicht enthält.

Aufgabe E.4



Ein Student liest in einem Bericht über Verkehrsanalysen, dass die mittlere Fahrzeit mit einem Personenwagen von seinem Wohnort Baden bis zur ETH Hönggerberg während des Berufsverkehrs 23.7 Minuten, mit einer Standardabweichung von 3 Minuten, beträgt.

Bei seinen nächsten 13 Fahrten während des Berufsverkehrs notiert er sich seine eigenen Fahrzeiten und kommt auf einen Durchschnitt von 22.3 Minuten.

Berechne auf einem Signifikanzniveau von 5 %, ob der Bericht mit diesen Messungen übereinstimmen kann, unter der Annahme einer normalverteilten Fahrzeit und unter der Annahme, dass auch bei seinen Messungen die Standardabweichung 3 Minuten beträgt.

Aufgabe E.4

Gegeben: Fahrzeit Baden-Hönggerberg

Normalverteilte Grundgesamtheit: $\mu = 23.7$ [min], $\sigma = 3$ [min]

Stichprobe: $\bar{x} = 22.3$ [min]

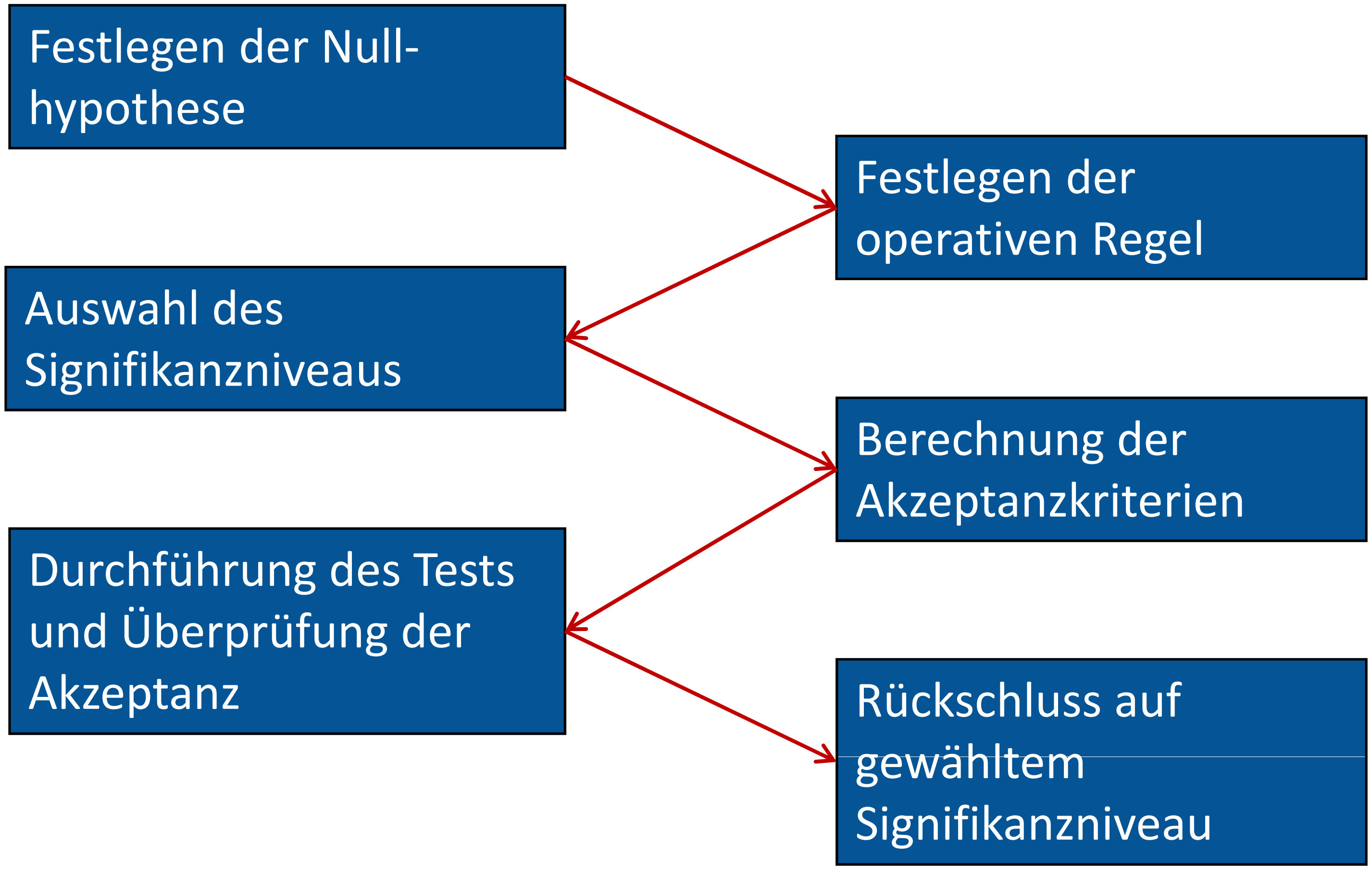
$n = 13$

Annahme: gleiche Standardabweichung wie Grundgesamtheit

Gesucht: Aussage darüber, ob auf einem Signifikanzniveau von 5 % behauptet werden kann, dass der Bericht mit diesen Messungen übereinstimmt.

Testen von Hypothesen

Generelles Vorgehen beim Hypothesentest:



Aufgabe E.4

$$\mu_X = 23.7[\text{min}],\ \sigma_X = 3[\text{min}]$$

$$\bar{x} = 22.3[\text{min}],\ n = 13,\ \sigma_X = 3[\text{min}]$$

Hypothesentest:

Kann auf einem einem Signifikanzniveau von 5 % behauptet werden, dass der Bericht mit diesen Messungen übereinstimmt?

1. **Hypothese:**
$$H_0 : \mu_X = 23.7[\text{min}]$$
$$H1 : \mu_X \neq 23.7[\text{min}]$$

2. **Regel für Ablehnung oder Akzeptanz (operative Regel):**

   Der Stichprobenmittelwert muss sich in einem bestimmten Intervall rund um den wahren Mittelwert befinden, um die Nullhypothese akzeptieren zu können.

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

3. **Signifikanzniveau:** $\alpha = 0.05$

Aufgabe E.4

$$\mu_X = 23.7[\min],\ \sigma_X = 3[\min]$$

$$\bar{x} = 22.3[\min],\ n = 13,\ \sigma_X = 3[\min]$$

4. Berechnen:

$$\mu_X - \Delta \le \overline{X} \le \mu_X + \Delta$$

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

$$\rightarrow k_{\alpha/2}:\ \Phi^{-1}\left(1-\alpha/2\right) = \Phi^{-1}(0.975) \rightarrow k_{\alpha/2} = 1.96$$

$$\mu_X - 1.96\frac{\sigma_X}{\sqrt{n}} \le \overline{X} \le \mu_X + 1.96\frac{\sigma_X}{\sqrt{n}}$$

$$23.7 - 1.96\frac{3}{\sqrt{13}} \le \overline{X} \le 23.7 + 1.96\frac{3}{\sqrt{13}}$$

$$22.07 \le \overline{X} \le 25.33$$

Aufgabe E.4

$$\mu_X = 23.7[\text{min}],\ \sigma_X = 3[\text{min}]$$
$$\bar{x} = 22.3[\text{min}],\ n = 13,\ \sigma_X = 3[\text{min}]$$

5. Test durchführen

$$\mu_X - \Delta \leq \overline{X} \leq \mu_X + \Delta \rightarrow 22.07 \leq \overline{X} \leq 25.33$$

Der Stichprobenmittelwert beträgt $\bar{x} = 22.3[\text{min}]$, liegt also im Intervall.

6. Rückschluss

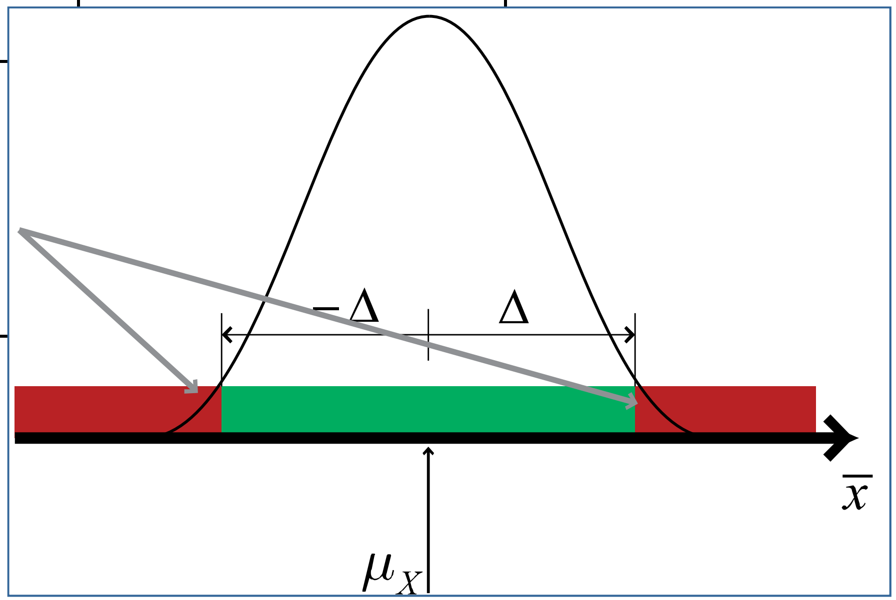
Der Stichprobenmittelwert befindet sich im Intervall $\left[22.07 \leq \bar{x} \leq 25.33\right]$. Die Nullhypothese, dass der wahre Mittelwert 23.7 [min] beträgt, kann anhand dieser Messungen nicht verworfen werden.

Testen von Hypothesen

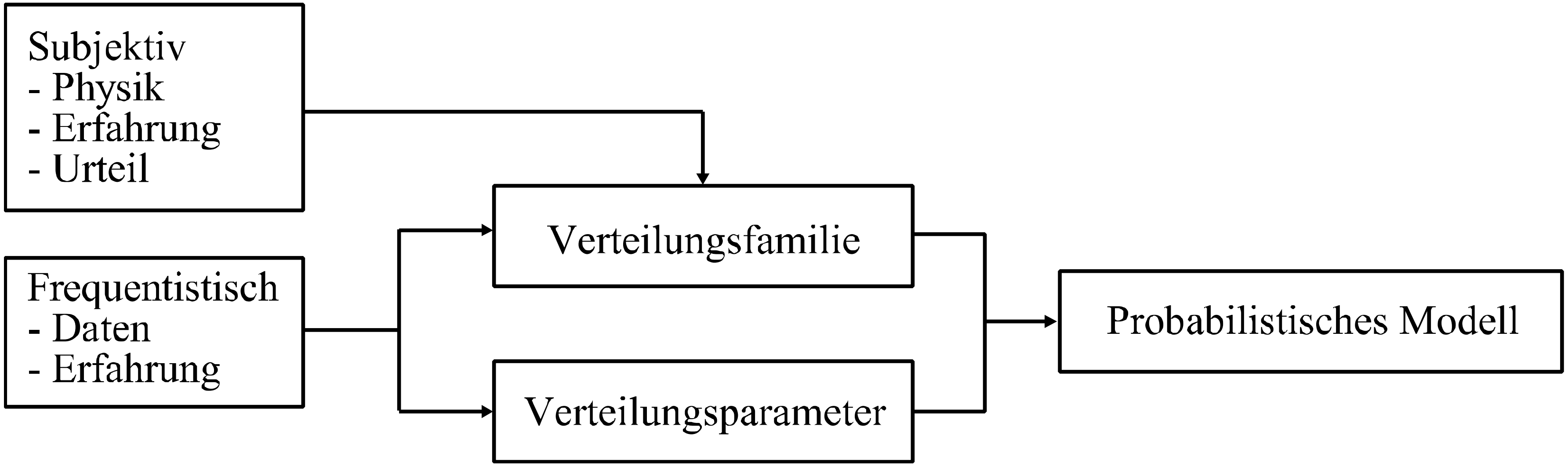
Wahrheit / Urteil	H_0 trifft zu.	H_0 trifft nicht zu.
Akzeptanz von H_0.	Richtiges Urteil.	Fehler 2. Art.
Ausschluss von H_0.	Fehler 1. Art.	Richtiges Urteil.

Testen von Hypothesen

Wahrheit / Urteil	H_0 trifft zu.	H_0 trifft nicht zu.
Akzeptanz von H_0.	Richtiges Urteil.	Fehler 2. Art.
Ausschluss von H_0.	Fehler 1. Art.	



Güte der Anpassung



χ^2-Test

- Für diskrete und kontinuierliche Verteilungen anwendbar (Kolmogorov-Smirnov-Test nur für kontinuierliche Verteilungen).

- 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.

Kolmogorov-Smirnov-Test

Die Idee:

Bei diesem Test werden die Abweichungen der angenommenen kumulativen Verteilungsfunktion und der beobachteten kumulativen Verteilungsfunktion bestimmt. Die maximale Abweichung sollte dabei möglichst klein sein.

- Ist nur für kontinuierliche Verteilungen anwendbar.

Aufgabe E.11

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

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

Intervall	Häufigkeit	Wahrscheinlichkeit P [Stichprobe in diesem Intervall]	Erwartete Häufigkeit
0-20			
20-25			
25-30			
30-∞			

Aufgabe E.11

Exponentialverteilung

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

Die Statistik für den Chi-Quadrat Test ist:

$$\varepsilon_m^2 = \sum_{i=1}^{k} \frac{\left(N_i - n \cdot p_i\right)^2}{n \cdot p_i}$$

Und die operative Regel lautet dann:

$$P\left[\varepsilon^2 > \chi\right] = \alpha \qquad P\left[\varepsilon^2 \leq \chi\right] = 1 - \alpha$$

ε^2 folgt der χ^2-Quadrat Verteilung mit 2 Freiheitsgraden. Warum?

Aufgabe E.11

Exponentialverteilung

ε^2 folgt der χ^2 -Verteilung mit 2 Freiheitsgraden.

Wir haben 4 Intervalle.
Das letzte Intervall ist abhängig von den 3 anderen – Reduktion um 1 Freiheitsgrad.
Bestimmung des Parameters λ aus den beobachteten Daten – Reduktion um 1 Freiheitsgrad.

$v = m - 1 - j$ 4 – 1 – 1 = 2 FHG und α=0.1 ~ χ = 4.6052

v \ $F(q)$	0.75	0.90	0.95	0.98	0.99	0.995	0.999
1	1.3233	2.7055	3.8415	5.4119	6.6349	7.8794	10.8276
2	2.7726	$\boxed{4.6052}$	5.9915	7.8240	9.2103	10.5966	13.8155
3	4.1083	6.2514	7.8147	9.8374	11.3449	12.8382	16.2662
4	5.3853	7.7794	9.4877	11.6678	13.2767	14.8603	18.4668
5	6.6257	9.2364	11.0705	13.3882	15.0863	16.7496	20.5150

Skript, Anhang, Tab. 3

Aufgabe E.11

Exponentialverteilung

Intervall	Häufigkeit N_i	Wahrscheinlichkeit P [Stichprobe in diesem Intervall]	Erwartete Häufigkeit	Normalisierte Quadrate der Differenzen
0-20	7	0.532		
20-25	4	0.081		
25-30	11	0.067		
30- ∞	8	0.32		
Summe	30	1		

$$\int_{20}^{25} \lambda \cdot e^{-\lambda x} dx$$

mit den zu $\lambda = 0.038$ geschätzten Parameter

beobachtete Werte N_1 im Intervall 1

Faustregel:

N_i sollte in jedem Intervall ≥ 5 sein.

Aufgabe E.11

Exponentialverteilung

Intervall	Häufigkeit N_i	Wahrscheinlichkeit P [Stichprobe in diesem Intervall]	Erwartete Häufigkeit	Normalisierte Quadrate der Differenzen
0-20	7	0.532		
20-25	4	0.081		
25-30	11	0.067		
30- ∞	8	0.32		
Summe	30	1		

beobachtete Werte N_1 im Intervall 1

$$\int_{20}^{25} \lambda \cdot e^{-\lambda x} dx$$

mit den zu $\lambda = 0.038$ geschätzten Parameter

alternative Berechnungsmethode z. B. für 2. Intervall:

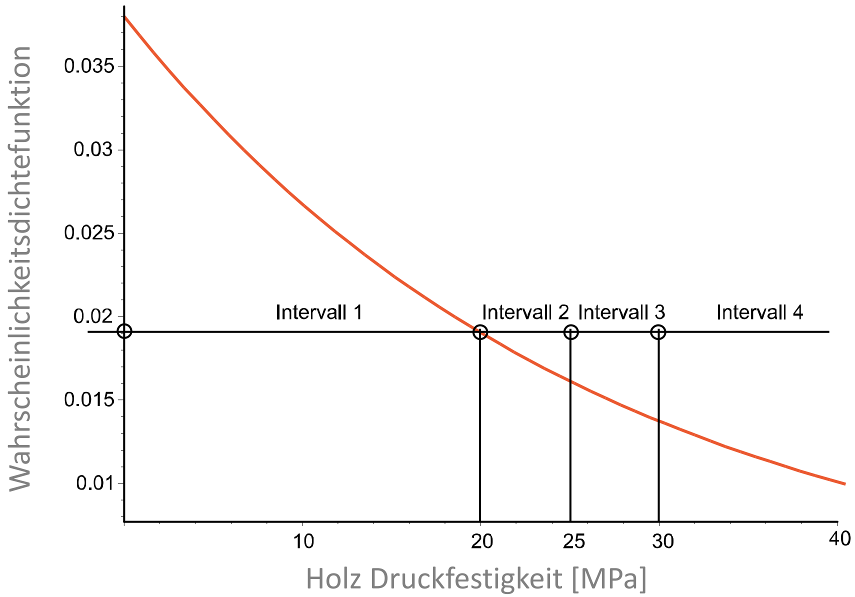
$$F_X(25) - F_X(20)$$

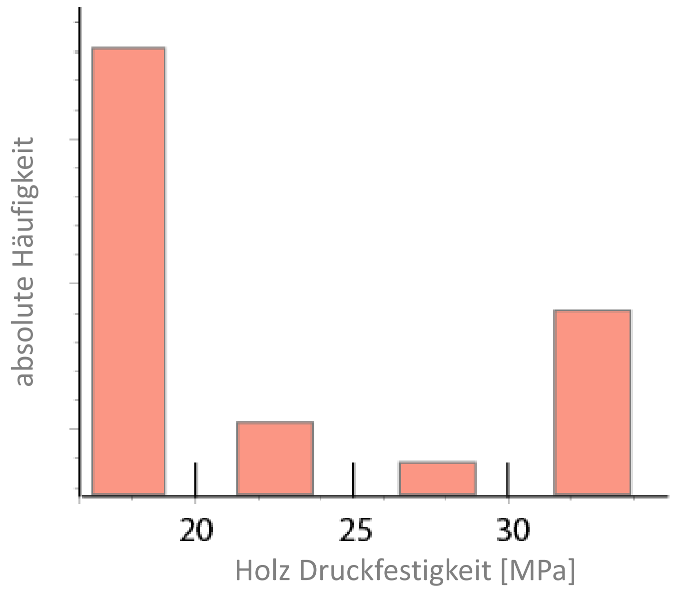
Aufgabe E.11

Exponentialverteilung

Hier haben wir eine stetige Wahrscheinlichkeitsdichtefunktion gegeben und wollen den Chi-Quadrat Test anwenden.

Für den Test müssen wir deshalb die Wahrscheinlichkeitsdichtefunktion diskretisieren. – Hier in 4 Intervalle !





Aufgabe E.11

Exponentialverteilung

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**

beobachtete Werte N_1 im Intervall 1

$$\int_{20}^{25} \lambda \cdot e^{-\lambda x} dx$$

mit den zu $\lambda = 0.038$ geschätzten Parameter

$$n \cdot p_1 = 30 \cdot 0.532$$

$$\varepsilon_m^2 = \frac{(N_2 - n \cdot p_2)^2}{n \cdot p_2}$$

$$\varepsilon_m^2 = \sum_{i=1}^{k} \frac{(N_i - n \cdot p_i)^2}{n \cdot p_i}$$

Aufgabe E.11

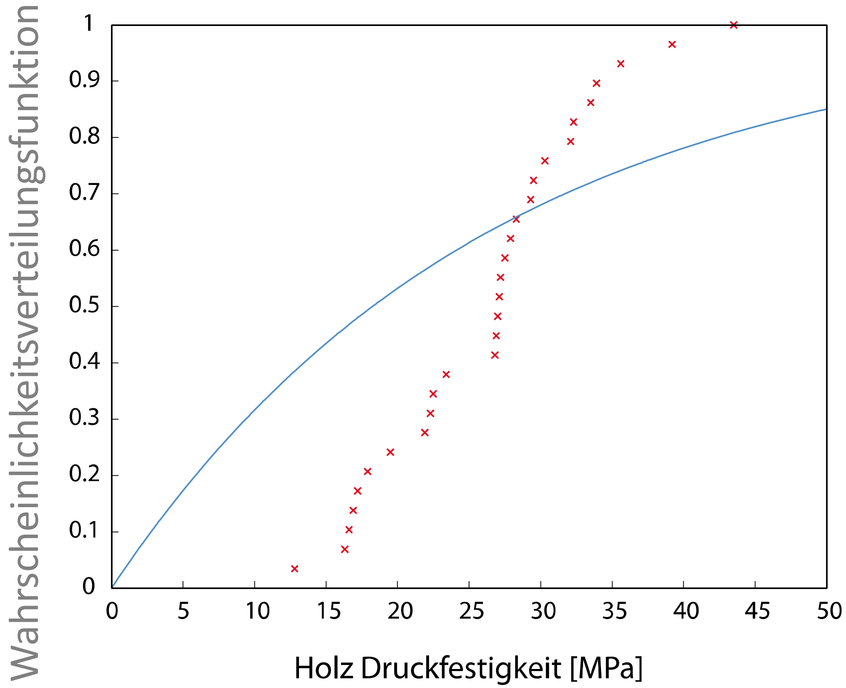
Exponentialverteilung

c) Teste die Güte der Anpassung für beide Verteilungen mit dem χ^2 Test auf einem Signifikanzniveau von 10%.

$$\varepsilon_m^2 = \sum_{i=1}^{k} \frac{(N_i - n \cdot p_i)^2}{n \cdot p_i} = 46.52$$

Mit der Entscheidungsregel :

$$P\left[\varepsilon^2 \leq \chi\right] = 1 - \alpha$$



Und dem aus der Tabelle ermittelten Wert für $\chi = 4.6052$ folgt:

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

Aufgabe E.11

Weibullverteilung

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

Die Statistik für den Chi-Quadrat Test ist:

$$\varepsilon_m^2 = \sum_{i=1}^{k} \frac{\left(N_i - n \cdot p_i\right)^2}{n \cdot p_i}$$

Und die operative Regel lautet dann:

$$P\left[\varepsilon^2 > \chi\right] = \alpha \qquad P\left[\varepsilon^2 \leq \chi\right] = 1 - \alpha$$

ε^2 folgt der χ^2-Quadrat Verteilung mit **1** Freiheitsgrad. Warum?

Aufgabe E.11

Weibullverteilung

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

Wir haben m= 4 Intervalle.
Das letzte Intervall ist abhängig von den 3 anderen – Reduktion um 1 Freiheitsgrad.
Bestimmung der Parameter k und u aus den beobachteten Daten – Reduktion um j=2 Freiheitsgrade.

$v = m - 1 - j$ = 4 – 1 – 2 = 1 FHG und a=0.1 ~ χ = 2.7055

v \ $F(q)$	0.75	0.90	0.95	0.98	0.99	0.995	0.999
1	1.3233	$\boxed{2.7055}$	3.8415	5.4119	6.6349	7.8794	10.8276
2	2.7726	4.6052	5.9915	7.8240	9.2103	10.5966	13.8155
3	4.1083	6.2514	7.8147	9.8374	11.3449	12.8382	16.2662
4	5.3853	7.7794	9.4877	11.6678	13.2767	14.8603	18.4668
5	6.6257	9.2364	11.0705	13.3882	15.0863	16.7496	20.5150

Skript, Anhang, Tab. 3

Aufgabe E.11

Weibullverteilung

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

beobachtete Werte N_1 im Intervall 1

$$\int_{20}^{25} \frac{k}{u} \cdot \left(\frac{x}{u}\right)^{k-1} \cdot e^{-\left(\frac{x}{u}\right)^{k}} dx$$

mit den zu
k = 4.21 und u = 29.05
geschätzten Parameter

$$n \cdot p_1 = 30 \cdot 0.19$$

$$\varepsilon_m^2 = \frac{(N_2 - n \cdot p_2)^2}{n \cdot p_2}$$

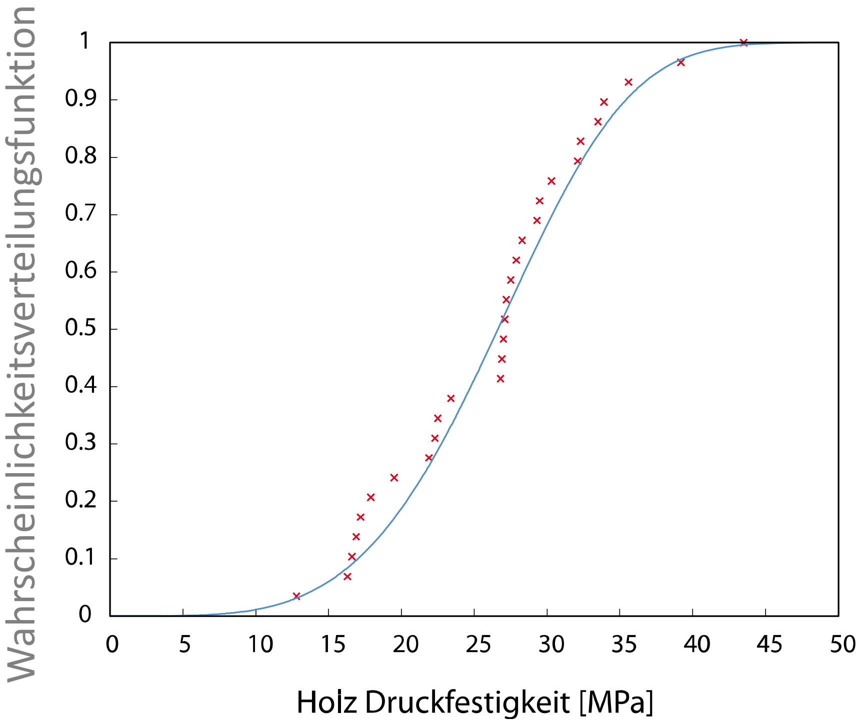
$$\varepsilon_m^2 = \sum_{i=1}^{k} \frac{(N_i - n \cdot p_i)^2}{n \cdot p_i}$$

Aufgabe E.11

Weibullverteilung

c) Teste die Güte der Anpassung für beide Verteilungen mit dem χ^2 Test auf einem Signifikanzniveau von 10%.

$$\varepsilon_m^2 = \sum_{i=1}^{k} \frac{\left(N_i - n \cdot p_i\right)^2}{n \cdot p_i} = 2.36$$



Mit der Entscheidungsregel :

$$P\left[\varepsilon^2 \leq c\right] = 1 - \alpha$$

Und den aus der Tabelle ermittelten Wert für $\chi = 2.71$ folgt:
Da $\varepsilon^2 = 2.36$ kleiner als $\chi = 2.71$ ist, können wir die Hypothese, dass die Daten einer Weibullverteilung folgen, auf einem Signifikanzniveau von 10% nicht verwerfen.

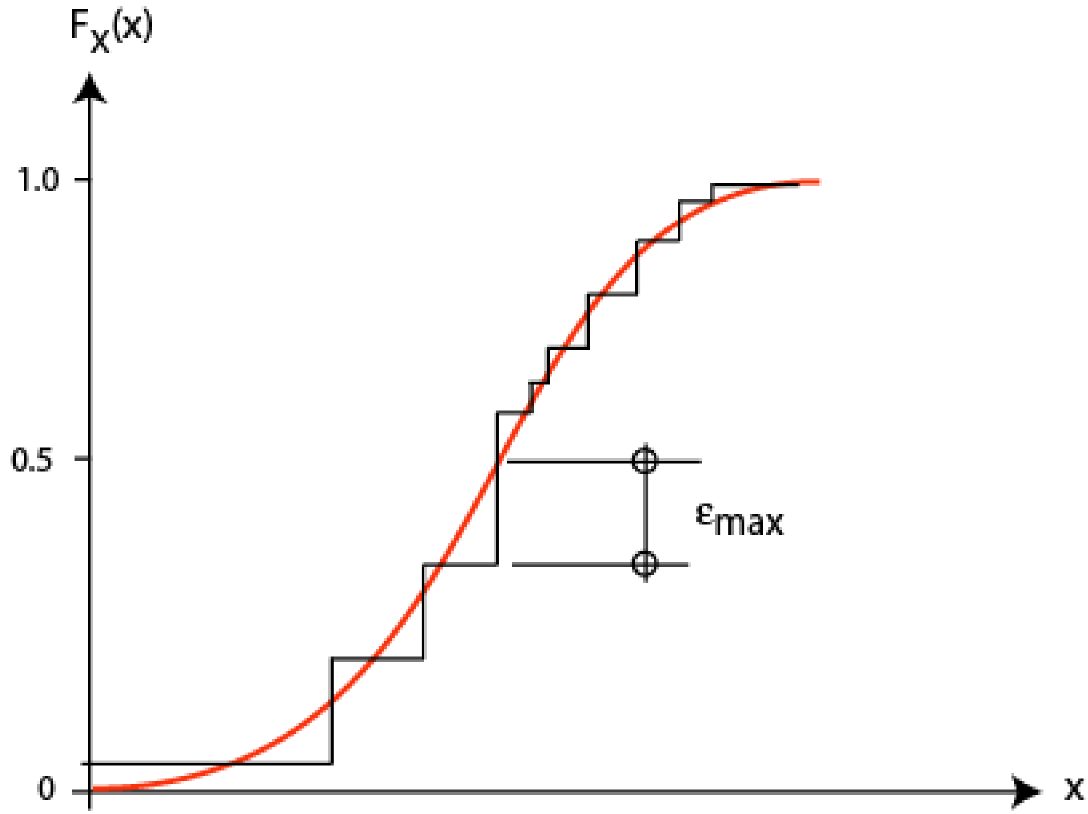
Aufgabe E.11

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

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

Aufgabe E.11

d) Teste die Güte der Anpassung beider Verteilungen mit dem Kolmogorov-Smirnov Test auf einem Signifikanzniveau von 10%.



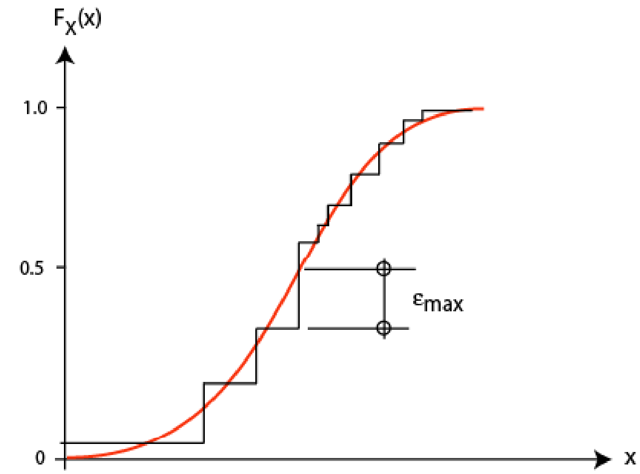
Aufgabe E.11

Die beobachtete kumulative Verteilungsfunktion wird berechnet:

$$F_{o,X}(x) = \frac{i}{n}$$

i ← Rang in der geordneten Tabelle

n ← Gesamtzahl der Beobachtungen



Es ergibt sich die folgende Statistik:

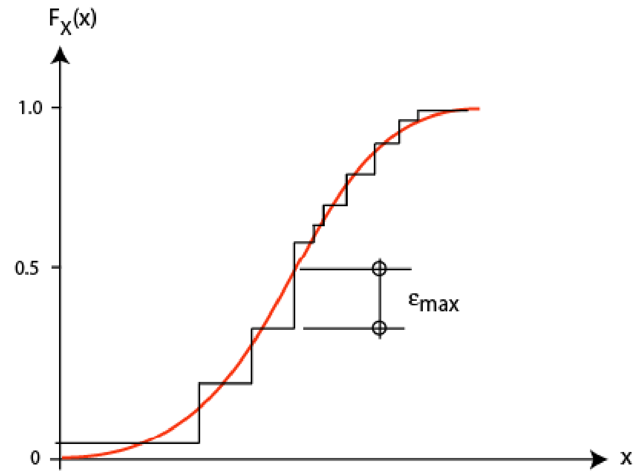
$$\varepsilon_{\max} = \max_{i=1}^{n}\left[\left|F_{o,X}(x_i) - F_p(x_i)\right|\right] = \max_{i=1}^{n}\left[\left|\frac{i}{n} - F_p(x_i)\right|\right]$$

Aufgabe E.11

Die Entscheidungsregel lautet:

$$P[\varepsilon_{\max} > c] = \alpha$$

$$P[\varepsilon_{\max} \le c] = 1 - \alpha$$



c kann für ein bestimmtes Signifikanzniveau der Tabellen entnommen werden:

n	$\alpha = 0.01$	$\alpha = 0.02$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.2$
28	0.300	0.279	0.250	0.225	0.197
29	0.295	0.275	0.246	0.221	0.193
30	0.290	0.270	0.242	0.218	0.190
31	0.285	0.266	0.238	0.214	0.187

Skript, Anhang, Tab. 4

Aufgabe E.11

Exponentialverteilung

d) Teste die Güte der Anpassung beider Verteilungen mit dem Kolmogorov-Smirnov Test auf einem Signifikanzniveau von 10%.

Die Stichprobenstatistik für den Kolmogorov-Smirnov-Test ist also

$$\varepsilon_{\max} = \max_{i=1}^{n}\left[\left|\frac{i}{n} - F_X(x_i)\right|\right]$$

Wobei $F_X(x)$ die angenommene Verteilungsfunktion ist.

Schauen wir uns die Exponentialverteilung an:

$$F_X(x) = 1 - e^{-\lambda x}$$

Mit λ= 0.038 können wir die Werte berechnen

Aufgabe E.11

Exponentialverteilung

Nr.	[Mpa]	FX(x)	i/n	| i/n - FX(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

$$F_X(x) = 1 - e^{-0.038x} = 1 - e^{-0.038 \cdot 17.2} = 0.48$$

$$\frac{i}{n} = \frac{10}{30} = 0.333$$

$$\left| \frac{i}{n} - F_p(x_i) \right| =$$

$$|0.57 - 0.64| = 0.08$$

Aufgabe E.11

Exponentialverteilung

Nr.	[Mpa]	FX(x)	i/n	| i/n - FX(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

$$\max\left|\frac{i}{n} - FX(x)\right| = 0.40$$

$$F_X(x) = 1 - e^{-0.038x} = 1 - e^{-0.038\cdot 17.2} = 0.48$$

$$\frac{i}{n} = \frac{10}{30} = 0.333$$

$$\left|\frac{i}{n} - F_p(x_i)\right| =$$

$$|0.57 - 0.64| = 0.08$$

Aufgabe E.11

Exponentialverteilung

Mit der Entscheidungsregel können wir die Hypothese nun bewerten:

$$P\left[\varepsilon_{\max} \leq c\right] = 1 - \alpha$$

α = 0.1 können wir den Wert *c* aus der Tabelle ablesen

n	$\alpha = 0.01$	$\alpha = 0.02$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.2$
28	0.300	0.279	0.250	0.225	0.197
29	0.295	0.275	0.246	0.221	0.193
30	0.290	0.270	0.242	$\boxed{0.218}$	0.190
31	0.285	0.266	0.238	0.214	0.187

Skript, Anhang, Tab. 4

c hat demnach den Wert 0.218

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

Aufgabe E.11

Weibullverteilung

d) Teste die Güte der Anpassung beider Verteilungen mit dem Kolmogorov-Smirnov Test auf einem Signifikanzniveau von 10%.

Da der Tabellenwert c unabhängig von der gewählten Verteilung und auch unabhängig vom Freiheitsgrad ist, gilt dieser auch für die Teststatistik der Weibullverteilung.

Die Stichprobenstatistik für den Kolmogorov-Smirnov-Test ist also

$$\varepsilon_{\max} = \max_{i=1}^{n}\left[\left|\frac{i}{n} - F_X(x_i)\right|\right]$$

Wobei $F_X(x)$ die angenommene Verteilungsfunktion ist.

$$F_X(x) = 1 - e^{-\left(\frac{x}{u}\right)^k}$$

Mit u= 29.05 und k=4.21 können wir die Werte berechnen.

Aufgabe E.11

Weibullverteilung

Nr.	[Mpa]	FX(x)	i/n	| i/n - FX(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

$$F_X(x) = 1 - e^{-\left(\frac{16.9}{29.05}\right)^{4.21}} = 0.10$$

$$\frac{i}{n} = \frac{10}{30} = 0.333$$

$$\max\left|\frac{i}{n} - FX(x)\right| = 0.11$$

$$\left|\frac{i}{n} - F_p(x_i)\right| =$$

$$|0.7 - 0.77| = 0.07$$

Aufgabe E.11

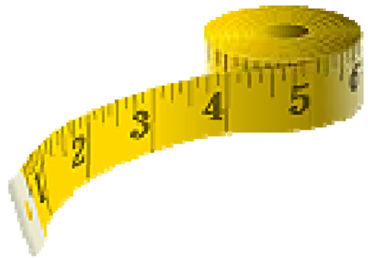
Weibullverteilung

Mit der Entscheidungsregel können wir die Hypothese nun beurteilen:

$$P\left[\varepsilon_{\max} \leq c\right] = 1 - \alpha$$

Da ε_{max} = 0.11 kleiner ist als c=0.218, können wir die Hypothese, dass die Daten einer Weibullverteilung folgen, auf einem Signifikanzniveau von 10% annehmen.

Aufgabe E.12 - *Gruppenaufgabe*



Anhand eines Teiles des in der ersten Vorlesung erhobenen Datensatzes, welcher die Körpergrösse aller Frauen beinhaltet, soll Folgendes durchgeführt werden:

a) Passe die Normalverteilung den Beobachtungen an. Bestimme dazu die Parameter μ und σ mit der Maximum-Likelihood-Methode.

b) Teste die Güte der Anpassung für die Normalverteilung mit dem χ^2 Test auf einem Signifikanzniveau von 10%.

Aufgabe E.12 - *Gruppenaufgabe*

Datensatz der erhobenen Körpergrössen aller Frauen im FS08:

Nr.	Grösse [cm]	Nr.	Grösse [cm]	Nr.	Grösse [cm]	Nr.	Grösse [cm]
1	158	11	164	21	170	31	175
2	158	12	165	22	170	32	175
3	158	13	165	23	172	33	176
4	160	14	165	24	172	34	176
5	160	15	166	25	172	35	176
6	162	16	166	26	173	36	177
7	162	17	168	27	174	37	178
8	164	18	168	28	174	38	183
9	164	19	169	29	175		
10	164	20	170	30	175		

$$x_M = 168.92cm$$
$$S_M = 6.41cm$$