# ÜBUNG 10

## Lösungen Aufgabe 10.1

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

*Tabelle 10.1: Druckfestigkeit von Holzproben.*

| Nr. | Druck-festigkeit [MPa] | Nr. | Druck-festigkeit [MPa] | Nr. | Druck-festigkeit [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:

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

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

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

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:

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

Die unbekannten Parameter $u$ und $k$ können wie folgt bestimmt werden:

Parameter $k$ :

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

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 nummerisches/approximatives Vorgehen(z.B. Excel Solver). So erhält man:

$$k \approx \left(\frac{\sigma}{\mu}\right)^{-1.09} = 4.21$$

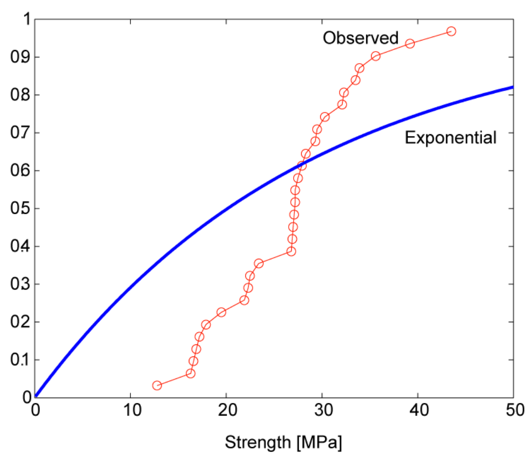
Parameter $u$:

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

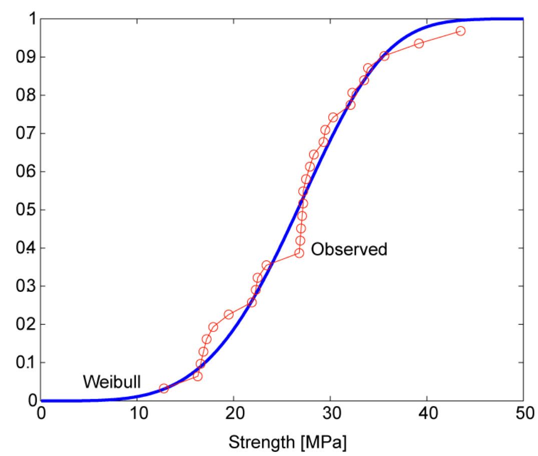
$$u = \frac{m_1}{\Gamma(1+1/k)} = \frac{m_1}{\Gamma(1.24)} = 29.05$$

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

Exponentialverteilung




Weibullverteilung




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

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

, 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 Entscheidungsregelautet dann: Wenn $P\left[\varepsilon^2 \le \chi\right] = 1 - \alpha$ erfüllt ist kann die Hypothese angenommen werden.

<u>Exponentialverteilung:</u>

$\varepsilon^2$ 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 $\lambda$ aus den beobachteten Daten – Reduktion um 1 FHG.
4 – 1 – 1 = 2 FHG und $\alpha$=0.1 ~ χ = 4.6052

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

*Tabelle 10.2 Intervalle für den $\chi^2$ 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 χ=4.61 ist, müssen wir die Hypothese, dass die Daten einer Exponentialverteilung folgen, auf einem Signifikanzniveau von 10% verwerfen.

<u>Weibullverteilung:</u>

$\varepsilon^2$ folgt der Chi-Quadrat Verteilung mit 1 FHG.

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

$\varepsilon_m^2 = \sum_{i=1}^{k} \frac{\left(N_i - n \cdot p_i\right)^2}{n \cdot p_i} = 2.36$ last sich mit Hilfe Tabelle 10.3 berechnen:

*Tabelle 10.3 Intervalle für den $\chi^2$ 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 $\chi$=2.71 ist, können wir die Hypothese, dass die Daten einer Weibullverteilung folgen, auf einem Signifikanzniveau von 10% nicht verwerfen.

Zu Beachten:
Der $\chi^2$ 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) Teste die Güte der Anpassung beider Verteilungen mit dem Kolmogorov-Smirnov Test auf einem Signifikanzniveau von 10%**

<u>Exponentialverteilung:</u>

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

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

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 10.4: Berechnungsblatt für den Kolmogorov-Smirnov Test.*

| Nr. | [Mpa] | $F_X(x)$ | i/n | $\lvert i/n - F_X(x) \rvert$ |
|---|---|---|---|---|
| 1 | 12.80 | 0.39 | 0.03 | 0.35 |
| 2 | 16.30 | 0.46 | 0.07 | $\boxed{0.40}$ |
| 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 |

<u>Weibullverteilung:</u>

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

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

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 10.5: Berechnungsblatt für den Kolmogorov-Smirnov Test.*

| Nr. | [Mpa] | $F_X(x)$ | i/n | $\lvert i/n - F_X(x) \rvert$ |
|---|---|---|---|---|
| 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 |

## Lösungen Aufgabe 10.2

Anhand eines Teiles des in der ersten Vorlesung erhobenen Datensatzes (Tabelle 10.6), welcher die Körpergrösse aller Frauen beinhaltet, soll folgendes durchgeführt werden:

*Tabelle 10.6: Körpergrössen*

| 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 | | |

**a) Überprüfe mittels eines T-Tests ob der Mittelwert einem gegebenen Literaturwert von μ = 168 cm entspricht, auf einem Signifikanzniveau von 20%.**

Intervall für den Stichprobenmittelwert berechnen:

$$\mu_L - t_{\alpha/2}\frac{s_M}{\sqrt{n}} \le \overline{x}_M \le \mu_L + t_{\alpha/2}\frac{s_M}{\sqrt{n}}$$

Untere Intervallgrenze: $\mu_L - t_{\alpha/2}\frac{s_M}{\sqrt{n}} = 168 - 1.303\cdot\frac{6.41}{\sqrt{38}} = 166.65$

Obere Intervallgrenze: $\mu_L + t_{\alpha/2}\frac{s_M}{\sqrt{n}} = 168 + 1.303\cdot\frac{6.41}{\sqrt{38}} = 169.35$

Der Stichprobenmittelwert wurde berechnet zu: $\overline{x}_M = 168.92cm$

Das Intervall der t-Statistik wurde berechnet als [166.65cm; 169.35cm].
Der Stichprobenmittelwert liegt im Intervall.

Die Nullhypothese, dass es sich um die gleichen Populationen handelt, wird auf einem Signifikanzniveau von 20% angenommen.

**b) Passe die Normalverteilung den Beobachtungen an. Verwende dazu den Literaturmittelwert und bestimme den Parameter σ mit der Maximum Likelihood Methode.**

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

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

Um diese Aufgabe zu lösen wollen wir den Literaturwert verwenden und als ersten Parameter einsetzen:

$$\theta_1 = \mu_L = 168$$

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

$$\theta_2 = \widehat{\sigma}_X = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{x}_i - \theta_1\right)^2}{n}} = 10.934$$

**c) 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**

$\varepsilon^2$ folgt der Chi-Quadrat Verteilung mit 3 FHG.
Das letzte Intervall ist abhängig von den 4 anderen – Reduktion um 1 FHG.

Bestimmung nur des Parameters $\sigma$ aus den beobachteten Daten – Reduktion um 1 FHG.

5– 1 – 1 = 3 FHG und $\alpha$=0.1 ~ χ = 6.2514

$\varepsilon_m^2 = \sum_{i=1}^{k} \frac{\left(N_i - n \cdot p_i\right)^2}{n \cdot p_i} = 6.948$ last sich mit Hilfe Tabelle 10.7 berechnen:

Mit dem aus der Tabelle abgelesenen Wert für χ = 6.2514 folgt:
Da $\varepsilon^2$ =6.948 grösser als χ= 6.2514 ist, können wir die Hypothese, dass die Daten dieser Normalverteilung folgen, auf einem Signifikanzniveau von 10% nicht annehmen.

*Tabelle 10.7 Intervalle für den $\chi^2$ Test*

| **Intervall** | **Häufigkeit** | **Wahrscheinlichkeit P [ Stichprobe in diesem Intervall ]** | **Erwartete Häufigkeit** | **Normalisierte Quadrate der Differenzen** |
|---|---|---|---|---|
| 0-160 | 5 | 0.2322 | 8.824 | 1.657 |
| 161-165 | 9 | 0.1597 | 6.069 | 1.418 |
| 166-170 | 8 | 0.1807 | 6.867 | 0.1869 |
| 171-175 | 10 | 0.1664 | 6.323 | 2.1383 |
| 176- | 6 | 0.2610 | 9.918 | 1.548 |
| Summe | 38 | 1 | 38 | 6.948 |