Statistische Methoden der Datenanalyse WS 2017/18

Prof. Dr. Ulrich Landgraf

Lösung zum Aufgabenblatt 4 vom 22.11.2017

Aufgabe 1 Für die männlichen Teilnehmer lautet die Verteilung:

$$f_m(x) = \frac{1}{\sqrt{2\pi} \cdot 10}\, e^{-\frac{(x-180)^2}{200}}$$

und für die Studentinnen

$$f_w(x) = \frac{1}{\sqrt{2\pi} \cdot 8}\, e^{-\frac{(x-164)^2}{128}} .$$

Die Verteilung für alle Teilnehmer zusammen ist die gewichtete Summe, wobei die männlichen Teilnehmer mit dem Gewicht $75/(50+75)$ und die weiblichen mit dem Gewicht $50/(50+75)$ eingehen:

$$f_{ges}(x) = \frac{0.6}{\sqrt{2\pi} \cdot 10}\, e^{-\frac{(x-180)^2}{200}} + \frac{0.4}{\sqrt{2\pi} \cdot 8}\, e^{-\frac{(x-164)^2}{128}} .$$

a) Hier muss mit $f_{ges}(x)$ gearbeitet werden. Die Zahl der Studierenden im gesuchten Intervall erhält man, indem man die Wahrscheinlichkeit mit der Gesamtzahl multipliziert:

$$N = 125 \cdot \left(\frac{0.6}{\sqrt{2\pi} \cdot 10} \int_{164}^{180} e^{-\frac{(x-180)^2}{200}}\, dx + \frac{0.4}{\sqrt{2\pi} \cdot 8} \int_{164}^{180} e^{-\frac{(x-164)^2}{128}}\, dx \right)$$

Mit der in der Vorlesung verwendeten Transformation

$$z = \frac{x-\mu}{\sqrt{2}\,\sigma}$$

schreibt man das um in

$$\begin{aligned} N &= 125 \cdot \left(\frac{0.6}{\sqrt{\pi}} \int_{-\frac{1,6}{\sqrt{2}}}^{0} e^{-z^2}\, dz + \frac{0.4}{\sqrt{\pi}} \int_{0}^{\frac{2}{\sqrt{2}}} e^{-z^2}\, dz \right) \\ &= \frac{125}{2} \cdot \left(0.6 \cdot \mathrm{Erf}\left(\frac{1,6}{\sqrt{2}}\right) + 0.4 \cdot \mathrm{Erf}\left(\frac{2}{\sqrt{2}}\right) \right) \\ &= 62.5 \cdot (0.6 \cdot 0.890 + 0.4 \cdot 0.955) \\ &= 57.25 \end{aligned}$$

Man wird also 57 Personen erwarten, die zwischen 164 und 180 cm groß sind.

b) In diesem Fall sucht man die Integralgrenze:

$$75 \cdot \frac{1}{\sqrt{2\pi} \cdot 10} \int_{-\infty}^{x} e^{-\frac{(x-180)^2}{200}}\, dx = 15$$

Wir transformieren wieder:

$$75 \cdot \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\frac{x-180}{\sqrt{2}\cdot 10}} e^{-z^2}\, dz = 15$$

und schreiben das Integral um:

$$
\begin{aligned}
75 \cdot \frac{1}{2} \cdot \left(1 + \frac{2}{\sqrt{\pi}} \int_0^{\frac{x-180}{\sqrt{2}\cdot 10}} e^{-z^2}\, dz\right) &= 15 \\
75 \cdot \frac{1}{2} \cdot \left(1 + \mathrm{Erf}\left(\frac{x-180}{\sqrt{2}\cdot 10}\right)\right) &= 15 \\
\mathrm{Erf}\left(\frac{x-180}{\sqrt{2}\cdot 10}\right) &= 0.4 - 1 \\
\frac{x-180}{\sqrt{2}\cdot 10} &= \mathrm{ErfInverse}(\,-0.6) \\
x &= 180 + 10 \cdot \sqrt{2} \cdot \mathrm{ErfInverse}\,(-0.6) \\
x &= 180 + 10 \cdot \sqrt{2} \cdot (-0.595116) \\
x &= 171.6 \text{ cm}
\end{aligned}
$$

c) Hier ist der Ansatz:

$$50 \cdot \frac{1}{\sqrt{2\pi}\cdot 8} \int_x^{\infty} e^{-\frac{(x-164)^2}{128}}\, dx = 10$$

Transformiert:

$$50 \cdot \frac{1}{\sqrt{\pi}} \int_{\frac{x-164}{\sqrt{2}\cdot 8}}^{\infty} e^{-z^2}\, dz = 10$$

Umgeschrieben:

$$
\begin{aligned}
50 \cdot \frac{1}{2} \cdot \left(1 - \frac{2}{\sqrt{\pi}} \int_0^{\frac{x-164}{\sqrt{2}\cdot 8}} e^{-z^2}\, dz\right) &= 10 \\
50 \cdot \frac{1}{2} \cdot \left(1 - \mathrm{Erf}\left(\frac{x-164}{\sqrt{2}\cdot 8}\right)\right) &= 10 \\
\mathrm{Erf}\left(\frac{x-164}{\sqrt{2}\cdot 8}\right) &= 1 - 0.4 \\
\frac{x-164}{\sqrt{2}\cdot 8} &= \mathrm{ErfInverse}\,(0.6) \\
x &= 164 + 8 \cdot \sqrt{2} \cdot 0.595116 \\
x &= 170.7 \text{ cm}
\end{aligned}
$$

Aufgabe 2

Wenn man alle Maße in mm angibt, lautet die Normalverteilung:

$$f(x) = \frac{1}{\sqrt{2\pi}\cdot 0.2}\, e^{-\frac{(x-250)^2}{0{,}08}} \;.$$

a) Die gesuchte Wahrscheinlichkeit lautet:

$$P(249.7 < x < 250.3) = \frac{1}{\sqrt{2\pi}\cdot 0.2} \int_{249.7}^{250.3} e^{-\frac{(x-250)^2}{0.08}}\, dx$$

Transformiert ergibt das

$$
\begin{aligned}
P(249.7 < x < 250.3) &= \frac{1}{\sqrt{\pi}} \int_{\frac{249.7-250}{0.2\cdot\sqrt{2}}}^{\frac{250.3-250}{0.2\cdot\sqrt{2}}} e^{-z^2}\, dz \\
&= \frac{2}{\sqrt{\pi}} \int_0^{\frac{1.5}{\sqrt{2}}} e^{-z^2}\, dz \\
&= \mathrm{Erf}\left(\frac{1.5}{\sqrt{2}}\right) \\
&= 0.866
\end{aligned}
$$

Die gesuchte Wahrscheinlichkeit beträgt 86,6%.

b) Die gesuchte Wahrscheinlichkeit lautet:

$$
P(x < 250.1) = \frac{1}{\sqrt{2\pi}\cdot 0.2} \int_{-\infty}^{250.1} e^{-\frac{(x-250)^2}{0.08}}\, dx
$$

Transformiert ergibt das

$$
\begin{aligned}
P(x < 250.1) &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\frac{250.1-250}{0.2\cdot\sqrt{2}}} e^{-z^2}\, dz \\
&= \frac{1}{2}\left(1 + \frac{2}{\sqrt{\pi}} \int_0^{\frac{1}{2\cdot\sqrt{2}}} e^{-z^2}\, dz\right) \\
&= \frac{1}{2}\left(1 + \mathrm{Erf}\left(\frac{\sqrt{2}}{4}\right)\right) \\
&= \frac{1}{2}(1 + 0.3829) \\
&= 0.691
\end{aligned}
$$

Die gesuchte Wahrscheinlichkeit beträgt 69,1%.

c) Hier muss man die zwei (gleichen) Intervalle [-0.6,-0.3] und [0.3,0.6] für die Abweichung $x - 250.0$ berücksichtigen (Einheiten in mm!). Daher hat man

$$
P(0.3 < |x - 250.0| < 0.6) = 2 \cdot \frac{1}{\sqrt{2\pi}\cdot 0.2} \int_{0.3}^{0.6} e^{-\frac{u^2}{0.08}}\, du
$$

Zweckmäßigerweise wurde hier die Interationsvariable gleich als $u = x - 250$ gewählt. Wieder transformiert man mit $z = u/(\sqrt{2}\,\sigma)$:

$$
\begin{aligned}
P(0.3 < |x - 250.0| < 0.6) &= 2 \cdot \frac{1}{\sqrt{\pi}} \int_{\frac{0.3}{\sqrt{2}\cdot 0.2}}^{\frac{0.6}{\sqrt{2}\cdot 0.2}} e^{-z^2}\, dz \\
&= \frac{2}{\sqrt{\pi}} \left(\int_0^{\frac{0.6}{\sqrt{2}\cdot 0.2}} e^{-z^2}\, dz - \int_0^{\frac{0.3}{\sqrt{2}\cdot 0.2}} e^{-z^2}\, dz \right)
\end{aligned}
$$

$$
\begin{aligned}
&= \operatorname{Erf}\left(\frac{0.6}{\sqrt{2}\cdot 0.2}\right) - \operatorname{Erf}\left(\frac{0.3}{\sqrt{2}\cdot 0.2}\right) \\
&= \operatorname{Erf}\left(\frac{3}{2}\sqrt{2}\right) - \operatorname{Erf}\left(\frac{3}{4}\sqrt{2}\right) \\
&= 0.9973 - 0.8664 \\
&= 0.1309
\end{aligned}
$$

Die gesuchte Wahrscheinlichkeit beträgt hier 13,1%.

d) Jetzt ist die Grenze des Integrals gesucht, so dass

$$
\frac{1}{\sqrt{2\pi}\cdot 0.2}\int_{-x}^{x} e^{-\frac{(x-250)^2}{0.08}}\,dx = 0.95
$$

Auch hier transformiert man die Integrationsvariable:

$$
\begin{aligned}
\frac{1}{\sqrt{\pi}}\int_{-\frac{x-250}{\sqrt{2}\cdot 0.2}}^{\frac{x-250}{\sqrt{2}\cdot 0.2}} e^{-z^2}\,dz &= 0.95 \\
\frac{2}{\sqrt{\pi}}\int_{0}^{\frac{x-250}{\sqrt{2}\cdot 0.2}} e^{-z^2}\,dz &= 0.95 \\
\operatorname{Erf}\left(\frac{x-250}{\sqrt{2}\cdot 0.2}\right) &= 0.95 \\
x &= 250 + 0.2\cdot\sqrt{2}\cdot \operatorname{ErfInverse}(0.95) \\
x &= 250 + 0.2\cdot\sqrt{2}\cdot 1.3859 \\
x &= 250.39\ \text{mm}
\end{aligned}
$$

Aufgabe 3

Die Simulation (siehe Programmcode) ergibt, dass die gesuchte Wahrscheinlichkeitsverteilung eine Dreiecksform hat. Mit dem Ansatz:

$$
f(x) = \begin{cases} a\cdot x & 0 \le x \le 1 \\ -a\cdot x + 2\,a & 1 < x \le 2 \\ 0 & \text{sonst} \end{cases}
$$

muss man nur noch die Konstante a aus der Normierung bestimmen:

$$
\begin{aligned}
\int_{-\infty}^{\infty} f(x)\,dx &= \int_0^1 ax\,dx + \int_1^2 (-ax+2a)\,dx \\
&= \left[\frac{1}{2}\,ax^2\right]_0^1 + \left[-\frac{1}{2}\,ax^2 + 2ax\right]_1^2 \\
&= \frac{1}{2}\,a + \left(-2a + 4a + \frac{1}{2}\,a - 2a\right) \\
&= \frac{1}{2}\,a + \frac{1}{2}\,a = a \overset{!}{=} 1
\end{aligned}
$$

Damit heißt die gesuchte Wahrscheinlichkeitsverteilung

$$f(x) = \begin{cases} x & 0 \leq x \leq 1 \\ 2-x & 1 < x \leq 2 \\ 0 & \text{sonst} \end{cases}$$

Der Erwartungswert beträgt:

$$\begin{aligned} E(x) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx &= \int_0^1 x \cdot x\, dx + \int_1^2 x \cdot (2-x)\, dx \\ &= \int_0^1 x^2\, dx + \int_1^2 (2x - x^2)\, dx \\ &= \left[\frac{1}{3}x^3\right]_0^1 + \left[x^2 - \frac{1}{3}x^3\right]_1^2 \\ &= \frac{1}{3} + \left(4 - \frac{8}{3} - 1 + \frac{1}{3}\right) \\ &= \frac{1}{3} + \left(3 - \frac{7}{3}\right) \\ &= 1 \end{aligned}$$

Die Standardabweichung berechnet man so:

$$\begin{aligned} V(x) &= \int_{-\infty}^{\infty} x^2 \cdot f(x)\, dx - \left(E(x)\right)^2 \\ &= \int_0^1 x^2 \cdot x\, dx + \int_1^2 x^2 \cdot (2-x)\, dx - 1 \\ &= \int_0^1 x^3\, dx + \int_1^2 (2x^2 - x^3)\, dx - 1 \\ &= \left[\frac{1}{4}x^4\right]_0^1 + \left[\frac{2}{3}x^3 - \frac{1}{4}x^4\right]_1^2 - 1 \\ &= \frac{1}{4} + \left(\frac{16}{3} - 4 - \frac{2}{3} + \frac{1}{4}\right) - 1 \\ &= \frac{1}{2} + \frac{14}{3} - 5 \\ &= \frac{14}{3} - \frac{9}{2} = \frac{28}{6} - \frac{27}{6} = \frac{1}{6} \\ \Rightarrow \qquad \sigma &= \frac{1}{\sqrt{6}} \approx 0.4082 \end{aligned}$$

Bei der Gleichverteilung hingegen ist der Mittelwert

$$E(x) = \int_0^1 x\, dx = \left[\frac{1}{2}x^2\right]_0^1 = \frac{1}{2}$$

und die Standardabweichung

$$V(x) = \int_0^1 x^2\, dx - \left(E(x)\right)^2$$

$$
\begin{aligned}
&= \left[\frac{1}{3}x^3\right]_0^1 - \left(\frac{1}{2}\right)^2 \\
&= \frac{1}{3} - \frac{1}{4} = \frac{1}{12} \\
\Rightarrow \qquad \sigma &= \frac{1}{\sqrt{12}} \approx 0.2887
\end{aligned}
$$

Aufgabe 4 Siehe auch Programmcode

(ii) Aus einer Gleichverteilung ($f(x) = 1$ im Intervall [0,1]) soll eine Exponentialverteilung

$$g(y; \tau) = \frac{1}{\tau} e^{-\frac{y}{\tau}}$$

entstehen. Wir schreiben wie im Vorlesungsskript:

$$g(y) = f(x) \cdot \left|\frac{dx}{dy}\right|$$

Einsetzen von $f(x)$ und $g(y)$ liefert

$$\frac{1}{\tau} e^{-\frac{y}{\tau}} = 1 \cdot \left|\frac{dx}{dy}\right| \qquad (*)$$

Durch Integrieren können wir $x(y)$ bis auf ein Vorzeichen die Transformation $x(y)$ erhalten:

$$x(y) = \int \frac{1}{\tau} e^{-\frac{y}{\tau}}\, dy = \left[-e^{-\frac{y}{\tau}}\right]$$

Allerdings war x eine positive Zufallsvariable, während das Ergebnis hier stets negativ ist. Da die Funktion nur bis auf ein Vorzeichen festgelegt war, dürfen wir auch das umgekehrte Vorzeichen wählen, denn dann ist Relation (*) auch erfüllt. Wir legen also fest:

$$x(y) = e^{-\frac{y}{\tau}}$$

und bilden die Umkehrfunktion:

$$y(x) = -\tau \cdot \ln(x)$$

Da x ja nur zwischen 0 und 1 variiert, ist stets $\ln(x) < 0$ und daher y (mit der Bedeutung einer Zerfallszeit) stets größer 0, wie es sein soll.