ETH Zürich HS 2019
RW, D-MATL, D-MAVT
Dr. Christoph Czichowsky

Koordinator
Florian Krach

Stochastik
Musterlösung 5

1. Aufgrund langjähriger Untersuchungen ist bekannt, dass der Bleigehalt X in einer Bodenprobe annähernd normalverteilt ist. Ausserdem weiss man, dass der Erwartungswert 32 ppb beträgt und dass die Standardabweichung 6 ppb beträgt.

a) Mache eine Skizze der Dichte von X und zeichne die Wahrscheinlichkeit, dass eine Bodenprobe zwischen 26 und 38 ppb Blei enthält, in die Skizze ein.

b) Wie gross ist die Wahrscheinlichkeit, dass eine Bodenprobe höchstens 40 ppb Schwermetall enthält? Berechne *ohne* Taschenrechner.
Hinweis: Gehe zur standardisierten Zufallsvariablen Z über und benutze die Tabelle der Standardnormalverteilung.

c) Wie gross ist die Wahrscheinlichkeit, dass eine Bodenprobe höchstens 27 ppb Schwermetall enthält?

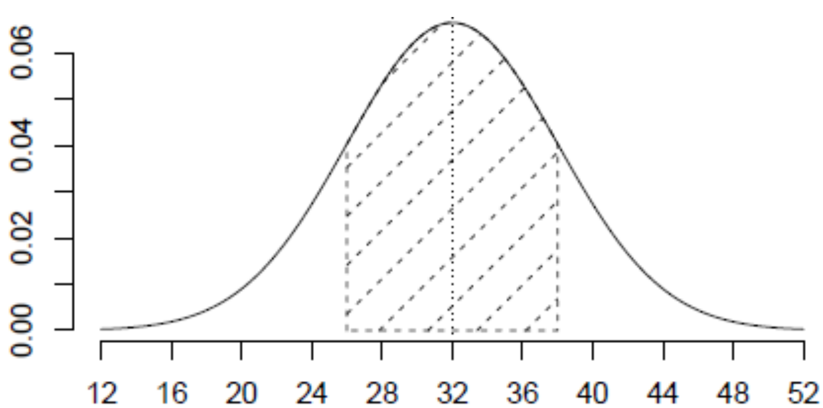
d) Welcher Bleigehalt wird mit einer Wahrscheinlichkeit von 97.5% unterschritten? Das heisst, bestimme dasjenige c, so dass die Wahrscheinlichkeit, dass der Bleigehalt kleiner oder gleich c ist, genau 97.5% beträgt.

e) Welcher Bleigehalt wird mit einer Wahrscheinlichkeit von 10% unterschritten?

f) Wie gross ist die Wahrscheinlichkeit, die in Aufgabe a) eingezeichnet wurde?

Lösung:

a) Siehe Bild.



Bitte wenden!

b) X bezeichne den Bleigehalt. Es gilt:

$$X \sim \mathcal{N}(\mu, \sigma^2) \qquad \text{mit } \mu = 32 \text{ und } \sigma^2 = 6^2.$$

Ohne Computer geht man aus praktischen Gründen (Tabelle!) normalerweise zur standardisierten Zufallsvariablen $Z = (X - \mu)/\sigma$ über. Es gilt: $Z \sim \mathcal{N}(0, 1)$.

$$P(X \leq 40) = P(Z \leq \frac{40 - 32}{6}) = P(Z \leq 1.33) = \Phi(1.33) = 0.9082$$

c) $P(X \leq 27) = P(Z \leq -0.83) = \Phi(-0.83) = 1 - \Phi(0.83) = 0.2033$

d) $P(X \leq c) = 0.975 = P(Z \leq \frac{c-32}{6}) = \Phi(\frac{c-32}{6})$
Mit Hilfe der Tabelle findet man $\Phi(1.96) = 0.975$ (Bemerkung: 1.96 ist das 97.5%-Quantil der Standardnormalverteilung). Also muss gelten:

$$\frac{c - 32}{6} = 1.96 \text{ und deshalb } c = 32 + 1.96 \cdot 6 = 43.76$$

e) Aus der Tabelle: $\Phi(1.28) = 0.9$ und $\Phi(-1.28) = 1 - 0.9 = 0.1$. Somit $c = 32 - 1.28 \cdot 6 = 24.31$.

f)

$$\begin{aligned} P(26 \leq X \leq 38) &= P(\frac{26 - 32}{6} \leq Z \leq \frac{38 - 32}{6}) = P(-1 \leq Z \leq 1) = \Phi(1) - \Phi(-1) \\ &= 2 \cdot \Phi(1) - 1 = 2 \cdot 0.8413 - 1 = 0.6826 \end{aligned}$$

2. Die Geschwindigkeit X eines Teilchens der Masse m sei durch die sogenannte *Rayleigh Verteilung* modelliert, d.h. X hat die Dichte

$$f(x) = \begin{cases} xe^{-x^2/2}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

a) Zeigen Sie, dass f eine Dichte ist.

b) Berechnen Sie die Wahrscheinlichkeit, dass das Teilchen eine Geschwindigkeit zwischen 2 und 5 besitzt.

c) Berechnen Sie die kumulative Verteilungsfunktion der kinetischen Energie $Y = \frac{1}{2}mX^2$. Bestimmen Sie den Erwartungswert von Y. Wie heisst die Verteilung von Y?

Lösung:

a) Nach Definition ist $f(x) \geq 0$ für alle x. Weiter gilt

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_0^{\infty} xe^{-x^2/2}\, dx = -e^{-x^2/2}\Big|_0^{\infty} = 1.$$

Siehe nächstes Blatt!

b) Es gilt

$$\begin{aligned} P[2 \leq X \leq 5] &= \int_2^5 f(x)dx = -e^{x^2/2}\Big|_2^5 \\ &= e^{-2} - e^{-12.5} = 13.5332\%. \end{aligned}$$

c) $F_Y(y)$ bezeichne die Verteilungsfunktion von Y. Für alle $y < 0$ ist $F_Y(y) = 0$. Für $y \geq 0$ hat man

$$\begin{aligned} F_Y(y) &= P[Y \leq y] = P[\frac{1}{2}mX^2 \leq y] = P\left[X \leq \sqrt{2y/m}\right] \\ &= \int_0^{\sqrt{2y/m}} xe^{-x^2/2}\,dx = -e^{-x^2/2}\Big|_0^{\sqrt{2y/m}} = 1 - e^{-y/m}. \end{aligned}$$

Dies bedeutet, dass Y eine Exponentialverteilung hat mit Parameter $1/m$. Deshalb folgt $E[Y] = m$.

3. Ein System bestehe aus 2 Maschinen, welche voneinander unabhängige Lebensdauern T_1 und T_2 besitzen mit Dichten (Die Maßeinheit von t sei Stunden)

$$f_{T_1}(t) = \begin{cases} \frac{1}{1000}\exp(-\frac{1}{1000}t) & t \geq 0 \\ 0 & \text{sonst,} \end{cases} \qquad f_{T_2}(t) = \begin{cases} \frac{1}{1500}\exp(-\frac{1}{1500}t) & t \geq 0 \\ 0 & \text{sonst.} \end{cases}$$

a) Berechne die Wahrscheinlichkeit, dass nach 200 Stunden beide Maschinen noch funktionieren.

b) Berechne die Wahrscheinlichkeit, dass nach 200 Stunden noch mindestens eine Maschine funktioniert.

c) Angenommen Maschine 1 funktioniert noch nach 200 Stunden, was ist dann die Wahrscheinlichkeit, dass sie noch weitere 200 Stunden funktioniert. Können wir daraus eine allgemeine Aussage für die Exponentialverteilung ableiten?

Lösung:

a) Die gesuchte Wahrscheinlichkeit ist gegeben durch

$$\begin{aligned} &P(T_1 \geq 200, T_2 \geq 200) \overset{\text{unabhängig}}{=} P(T_1 \geq 200) \cdot P(T_2 \geq 200) \\ &= (1 - P(T_1 \leq 200))(1 - P(T_2 \leq 200)) = (1 - F_{T_1}(200))(1 - F_{T_2}(200)), \end{aligned}$$

wobei F_{T_i} die kumulative Verteilungsfunktion von T_i ist, d.h. für $t > 0$

$$F_{T_1}(t) = \int_0^t f_{T_1}(s)ds = 1 - e^{-\frac{1}{1000}t}.$$

$$F_{T_2}(t) = \int_0^t f_{T_2}(s)ds = 1 - e^{-\frac{1}{1500}t}.$$

Bitte wenden!

Somit ist die Wahrscheinlichkeit:

$$P(T_1 \geq 200, T_2 \geq 200) = e^{-\frac{200}{1000}} \cdot e^{-\frac{200}{1500}} = 0.7165.$$

b) Die gesuchte Wahrscheinlichkeit ist gegeben durch

$$\begin{aligned} &P(T_1 \geq 200 \text{ oder } T_2 \geq 200) = P\left((T_1 \leq 200 \text{ und } T_2 \leq 200)^c\right) \\ &= 1 - P(T_1 \leq 200, T_2 \leq 200) \overset{\text{unabhängig}}{=} 1 - F_{T_1}(200)F_{T_2}(200) \\ &= 1 - (1 - e^{-\frac{200}{1000}})(1 - e^{-\frac{200}{1500}}) = 0.9774. \end{aligned}$$

c) Wir zeigen dass die Exponentialverteilung gedächtnislos ist, also dass für eine Zufallsvariable $X \sim \text{Exp}(\lambda), \lambda > 0$ und $s, t \geq 0$ gilt

$$P[X > s + t \mid X > s] = P[X > t].$$

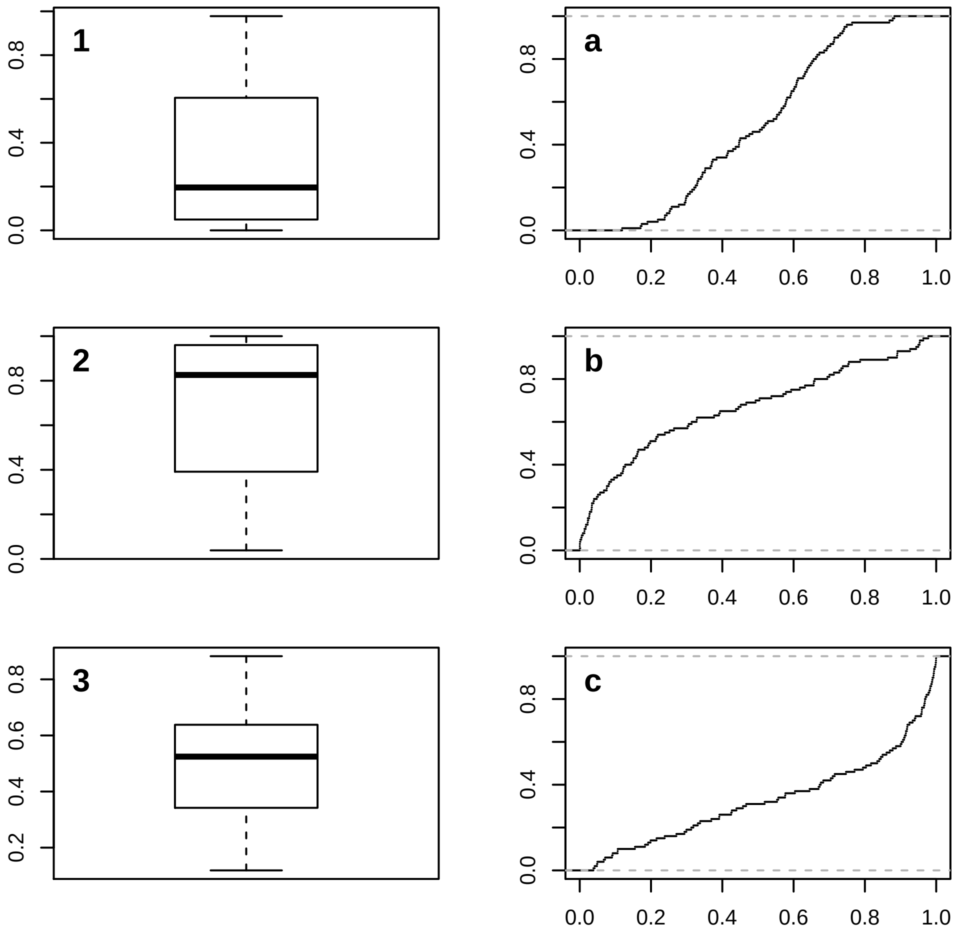
Unter Verwendung von $P[X > x] = 1 - F_X(x) = e^{-\lambda x}$ für alle $x \geq 0$, folgt

$$P[X > s + t \mid X > s] = \frac{P[X > s + t, X > s]}{P[X > s]} = \frac{P[X > s + t]}{P[X > s]} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P[X > t].$$

Damit berechnen wir die gesuchte Wahrscheinlichkeit

$$P[T_1 > 400 \mid T_1 > 200] = P[T_1 > 200] = e^{-\frac{200}{1000}} = 0.8187.$$

4. Für drei Stichproben vom Umfang n = 100 wurden je ein Boxplot und die empirische Verteilungsfunktion gezeichnet. Ordne die Boxplots den entsprechenden empirischen Verteilungsfunktionen zu:



Lösung: 1b, 2c, 3a

5. Gegeben seien die folgenden Daten:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	2.75	7.47	1.72	-0.45	6.12	0.16	-1.13	0.55	0.54	0.00	-0.47	0.83	2.30	0.77	16.85

a) Zeichne von Hand ein Histogramm der Daten. Bilde dazu Klassen $(c_{k-1}, c_k]$, $k = 1, \ldots, 10$ mit $c_0 = -2, c_1 = 0, c_2 = 2, c_3 = 4, \ldots, c_9 = 16, c_{10} = 18$. Was kann man über die Verteilung der Daten aussagen (Symmetrie, extreme Werte)?

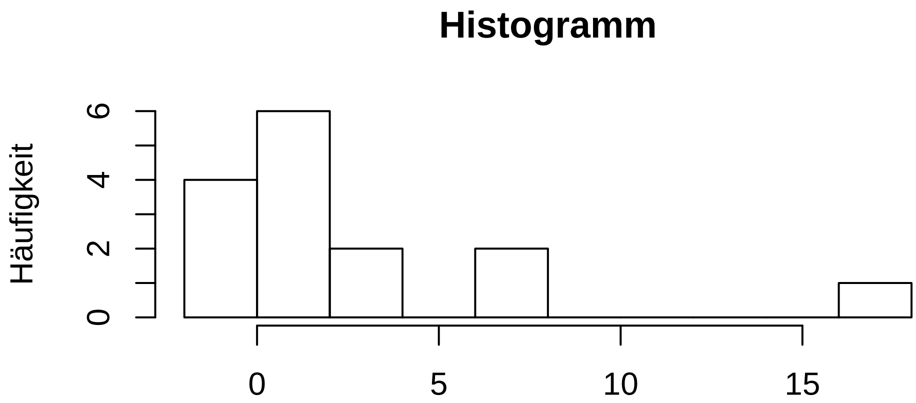
b) Bestimme den Mittelwert, den Median und die Standardabweichung der Daten.

c) Beim Wert $x_{15} = 16.85$ könnte es sich um einen Schreibfehler handeln. Ersetze x_{15} durch den Wert 6.85 und berechne erneut den Mittelwert, den Median und die Standardabweichung. Was stellst du fest?

Lösung:

a) Histogramm ist wie folgt. Die Verteilung ist positiv schief. Ein Wert liegt auffallend weit weg vom Rest (Ausreisser).

Bitte wenden!



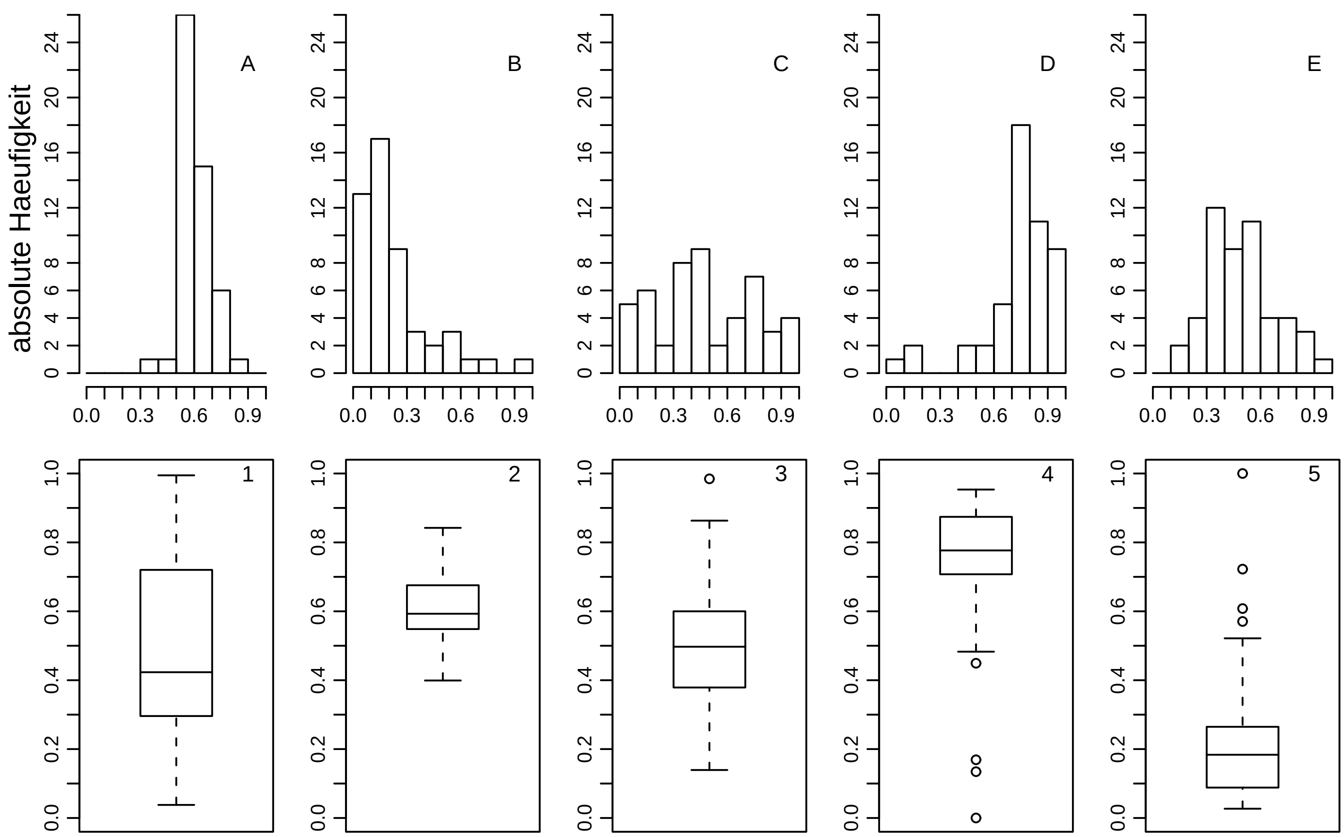
b) Der Mittelwert ist $\overline{x} = 2.53$. Der Median ist $m = 0.77$. Die Standardabweichung ist $\hat{\sigma} = 4.63$.

c) Der Mittelwert ist neu $\overline{x}_{\text{neu}} = 1.87$. Der Median ist neu $m_{\text{neu}} = 0.77$. Die Standardabweichung ist neu $\hat{\sigma}_{\text{neu}} = 2.77$. Mittelwert und Standardabweichung ändern sich stark. Der Median bleibt gleich. Der Median ist ein sogenannt robustes Lokationsmass. Es gibt auch robuste Streuungsmasse, z.B. den MAD ("median absolute deviation"), welcher definiert ist als der Median der absoluten Abweichungen vom Median.

6. Für fünf Stichproben vom Umfang $n = 100$ wurden je ein Histogramm und ein Boxplot gezeichnet. Ordne die fünf Boxplots den entsprechenden Histogrammen zu. Gib für jede Zuordnung eine kurze Begründung!

Lösung:

A2: Sehr kleine Streuung, Median ungefähr bei $0.5 - 0.6$.
B5: Lage gegen kleinere Werte hin verschoben, extreme Werte gegen oben, positive Schiefe.
C1: Grösste Streuung, Median gegen kleine Werte hin verschoben, keine extremen Werte.
D4: Lage gegen grössere Werte hin verschoben, extreme Werte gegen unten, negative Schiefe.
E3: Fast symmetrisch, Median ungefähr bei 0.5, grössere Streuung als A2, aber weniger extreme Werte als B5 und D4.


absolute Haeufigkeit
A
B
C
D
E
1
2
3
4
5
0.0 0.3 0.6 0.9
0 2 4 6 8 12 16 20 24
0.0 0.2 0.4 0.6 0.8 1.0