

## Musterlösung zu Serie 8

1. a) Damit  $f(x)$  eine Dichte ist, muss die Fläche des Dreiecks gleich 1 sein. Es muss also gelten

$$\frac{c \cdot 20}{2} = 1.$$

Daraus folgt  $c = \frac{1}{10}$ . Die Dichte lässt sich somit durch die Funktion

$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{10} \left(1 - \frac{x}{20}\right) & 0 \leq x \leq 20 \\ 0 & x > 20 \end{cases}$$

beschreiben.

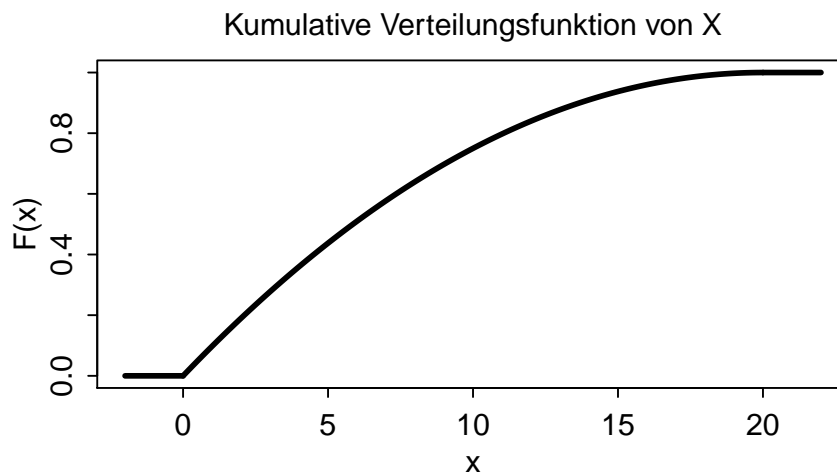
- b) Die kumulative Verteilungsfunktion von  $X$  lässt sich durch Integration der Dichtefunktion berechnen: Für  $0 \leq x \leq 20$  gilt:

$$F(x) = P[X \leq x] = \int_0^x f(t) dt = \int_0^x \left(\frac{1}{10} - \frac{t}{200}\right) dt = \frac{x}{10} - \frac{x^2}{400}$$

Für  $x \leq 0$  ist  $F(x) = 0$  und für  $x \geq 20$  gilt  $F(x) = 1$ .

Insbesondere gilt:  $P[X < 5] = F(5) = 0.4375$  und  $P[X < 10] = F(10) = 0.75$ .

- c) Die kumulative Verteilungsfunktion wurde bereits in b) berechnet.  
Skizze von  $F(x)$ :



- d)

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{20} x \left[ \frac{1}{10} \left(1 - \frac{x}{20}\right) \right] dx = \frac{1}{10} \left( \frac{x^2}{2} - \frac{x^3}{60} \right) \Bigg|_0^{20} = \frac{20}{3}$$

$$\text{Var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$$

$$\mathbf{E}[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^{20} x^2 \frac{1}{10} \left(1 - \frac{x}{20}\right) dx = \frac{1}{10} \left( \frac{x^3}{3} - \frac{x^4}{80} \right) \Bigg|_0^{20} = \frac{200}{3}$$

$$\text{Var}(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \frac{200}{3} - \left(\frac{20}{3}\right)^2 = \frac{200}{9}$$

Also ist die Standardabweichung  $sd(X) = \sqrt{\text{Var}(X)} = \sqrt{2} \cdot 10/3 \approx 4.71$ .

Für den Median  $\tilde{m}$  muss gelten:  $F(\tilde{m}) \stackrel{!}{=} 0.5$ . Der Median liegt sicher im Intervall  $[0,20]$  und somit haben wir

$$\frac{\tilde{m}}{10} - \frac{\tilde{m}^2}{400} \stackrel{!}{=} 0.5 \quad \implies \quad \tilde{m} = 20 - 10\sqrt{2} \approx 5.858.$$

e)

$$\begin{aligned} P[K \leq 120'000] &= P[40'000 \cdot \sqrt{X} \leq 120'000] = P[\sqrt{X} \leq 3] \\ &= P[X \leq 9] = F(9) = \frac{9}{10} - \frac{9^2}{400} = 0.6975 \end{aligned}$$

f) Die Exponential-verteilung hat die Dichte  $g(x)$ :

$$g(x) = \begin{cases} 0 & x \leq 0 \\ \lambda \exp(-\lambda x) & x > 0 \end{cases}.$$

Wenn  $X$  exponentialverteilt ist, dann ist der Erwartungswert

$$E[X] = \frac{1}{\lambda}.$$

Für  $\lambda = \frac{3}{20}$  erhalten wir somit denselben Erwartungswert wie in der bisherigen Verteilung.g) Die kumulative Verteilungsfunktion  $G(x)$  ist für  $x > 0$ 

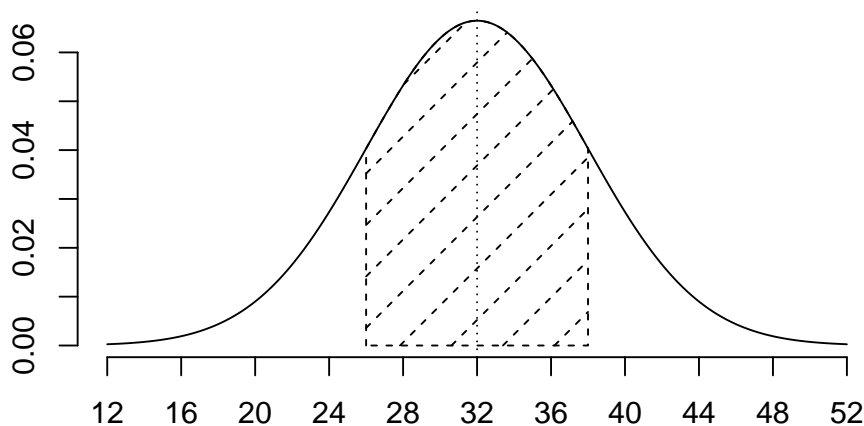
$$G(x) = P[X \leq x] = 1 - \exp(-\lambda x).$$

Daher

$$\begin{aligned} P[K \leq 120'000] &= P[40'000 \cdot \sqrt{X} \leq 120'000] = P[\sqrt{X} \leq 3] \\ &= P[X \leq 9] = G(9) = 1 - \exp\left(-\frac{3}{20} \cdot 9\right) = 1 - 0.259 = 0.741 \end{aligned}$$

Wenn die Dauer der Baustellen als exponentialverteilt angenommen wird, ist die Wahrscheinlichkeit also grösser, dass die Kosten einer Baustelle unter 120'000 Fr. liegen, verglichen mit der ursprünglich angenommenen Verteilung, obwohl der Erwartungswert für die Dauer der Baustellen identisch ist für beide Verteilungen.

2. a) Skizze:

b)  $X$  bezeichne den Bleigehalt. Es gilt:

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

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

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

Mit R kann die Wahrscheinlichkeit direkt (ohne Transformation) berechnet werden:

```
> pnorm(40, mean=32, sd=6)
```

```
[1] 0.9087888
```

Die kleine Differenz zur Zahl welche “von Hand” berechnet wurde beruht auf einem Rundungsfehler:

```
> pnorm(1.33, mean=0, sd=1)      # == pnorm(1.33)
```

```
[1] 0.9082409
```

```
> pnorm((40-32)/6, mean=0, sd=1) # == pnorm((40-32)/6)
```

```
[1] 0.9087888
```

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  $\Phi(1.96) = 0.975$ . Also muss gelten:

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

Mit R kann man die Zahl wie folgt berechnen:

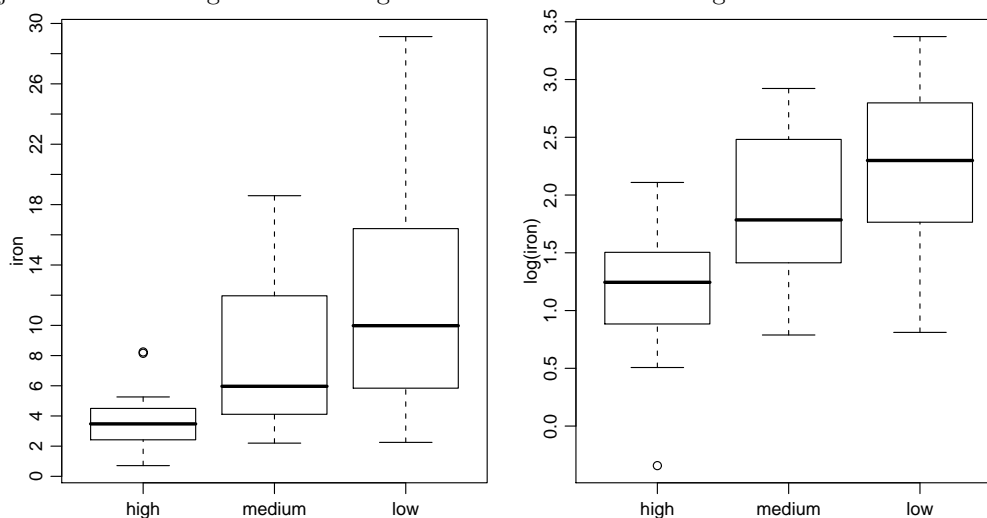
```
> qnorm(0.975, mean=32, sd=6)
```

```
[1] 43.75978
```

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

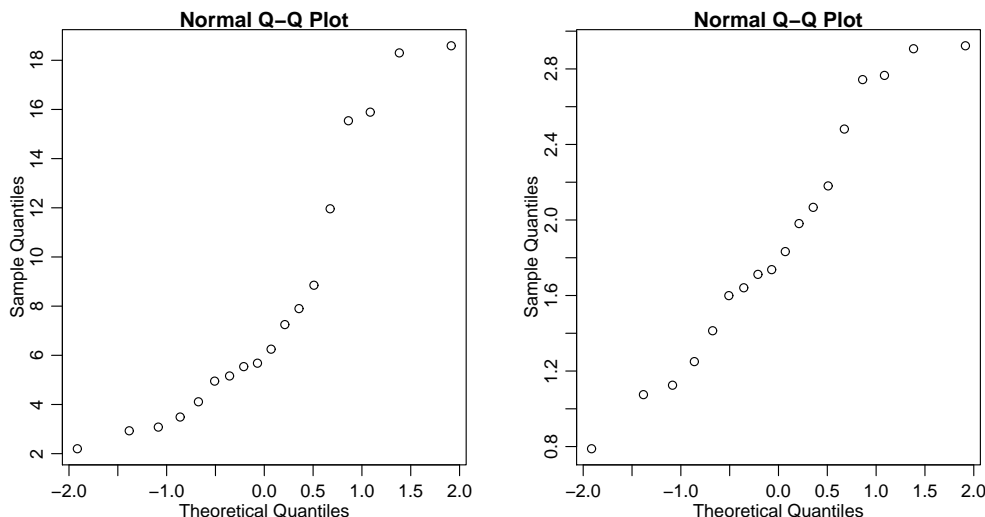
f)  $\Phi(1) - \Phi(-1) = 2 * \Phi(1) - 1 = 2 * 0.8413 - 1 = 0.6826$

3. a) Die drei Gruppen unterscheiden sich sowohl in der Lage wie auch in der Streuung. Bei tiefer Dosis ist der Anteil des zurückgehaltenen Eisens höher als bei hoher Dosis. Je kleiner die Dosis, desto grösser wird die Streuung.
- b) Wenn man die Daten logarithmiert, so wird die Varianz “stabilisiert”, d.h. alle Gruppen zeigen jetzt eine ähnlich grosse Streuung. Der Unterschied in der Lage ist immer noch ersichtlich.



- c) Bei normalverteilten Daten sollte man im Normalplot ungefähr eine Gerade erhalten. Aus den Plots sieht man, dass dies bei den ursprünglichen Daten (links) nicht der Fall ist, während man in der Abbildung mit den logarithmierten Daten (rechts) eher eine Gerade erkennt. Die Lognormalverteilung scheint also besser zu passen.

**R:** `qqnorm(iron[, "medium"]); qqnorm(log(iron[, "medium"]))`



4. a) Sei  $X$  die Anzahl Kunden, welche die neue Speisekarte bevorzugen.  $X$  ist binomialverteilt mit Parametern  $n = 356$  und  $\pi = 0.8$ . Daher gilt  $\mathbf{E}[X] = n \cdot \pi = 356 \cdot 0.8 = 284.8$
- b) Die Wahrscheinlichkeit, dass keiner der ersten vier befragten Kunden die neue Karte bevorzugt, der fünfte Kunde jedoch die neue Karte bevorzugt ist gegeben durch:  $0.2 \cdot 0.2 \cdot 0.2 \cdot 0.2 \cdot 0.8 = 0.2^4 \cdot 0.8 = 0.0013$ .

Die Wahrscheinlichkeit, dass drei der ersten vier befragten Kunden die neue Karte bevorzugen ist gegeben durch:  $P[X = 3] = \binom{4}{3} \cdot 0.8^3 \cdot 0.2^1 = 0.4096$ .

- c)  $Z = \frac{X - \mathbf{E}[X]}{\sqrt{\text{Var}(X)}} = \frac{X - n\pi}{\sqrt{n\pi(1-\pi)}}$  ist standardnormalverteilt. Also gilt

$$\begin{aligned} P(X \leq 261) &= P\left(\frac{X - n\pi}{\sqrt{n\pi(1-\pi)}} \leq \frac{261 - 356 \cdot 0.8}{\sqrt{356 \cdot 0.8 \cdot 0.2}}\right) = P(Z \leq -3.153) \\ &= \Phi(-3.153) = 1 - \Phi(3.153) \approx 8 \cdot 10^{-4} \end{aligned}$$

- d) 1. **Modell:**  $X$ : Anzahl Kunden, die die neue Karte bevorzugen.  $X \sim \text{Binomial}(n, \pi)$ , mit  $n = 356$ .
2. **Nullhypothese:**  $H_0 : \pi = \pi_0 = 0.8$   
**Alternative:**  $H_A : \pi \neq 0.8$
3. **Teststatistik:**  $\bar{X}_n$ : Anzahl Kunden, die die neue Karte bevorzugen geteilt durch Gesamtanzahl Kunden.  
**Verteilung der Teststatistik unter  $H_0$ :**  $\bar{X}_n \approx \mathcal{N}(\mu, \sigma_X^2/n)$  mit  $\mu = \pi_0 = 0.8$  und  $\sigma_X^2 = \pi_0(1 - \pi_0) = 0.8 \cdot 0.2 = 0.16$
4. **Signifikanzniveau:**  $\alpha = 0.05$
5. **Verwerfungsbereich für die Teststatistik:** Analog zum Beispiel im Skript ist der Verwerfungsbereich gegeben durch

$$\begin{aligned} K &= \left(-\infty, \pi_0 - \frac{\sqrt{\pi_0(1-\pi_0)}}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right] \cup \left[\pi_0 + \frac{\sqrt{\pi_0(1-\pi_0)}}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \infty\right) \\ &= \left(-\infty, 0.8 - \frac{\sqrt{0.16}}{\sqrt{356}} \cdot 1.96\right] \cup \left[0.8 + \frac{\sqrt{0.16}}{\sqrt{356}} \cdot 1.96, \infty\right) \\ &= (-\infty, 0.758] \cup [0.842, \infty). \end{aligned}$$

6. **Testentscheid:** Der beobachtete Wert der Teststatistik ist  $\bar{x}_n = \frac{261}{356} = 0.733$ . Da dieser Wert im Verwerfungsbereich  $K$  liegt, wird die Nullhypothese auf dem 5%-Signifikanzniveau verworfen. Die Zweifel von Lecker und Co., dass wirklich 80% der Kunden die neue Karte bevorzugen sind berechtigt.

Alternativ kann die Aufgabe (analog zur Aufgabe 3d) in Serie 6) wie folgt gelöst werden:

1. **Modell:**  $X$ : Anzahl Kunden, die die neue Karte bevorzugen.  $X \sim \text{Binomial}(n, \pi)$ , mit  $n = 356$ .
2. **Nullhypothese:**  $H_0 : \pi = \pi_0 = 0.8$   
**Alternative:**  $H_A : \pi \neq 0.8$
3. **Teststatistik:**  $X$ : Anzahl Kunden, die die neue Karte bevorzugen.  
**Verteilung der Teststatistik unter  $H_0$ :**  $X \sim \text{Binomial}(356, 0.8)$ .
4. **Signifikanzniveau:**  $\alpha = 0.05$
5. **Verwerfungsbereich für die Teststatistik:** Da eine zweiseitige Alternativhypothese getestet wird ist der Verwerfungsbereich von der Form  $K = [0, c_u] \cup [c_o, n]$ , wobei

$$c_u = n\pi_0 - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{n\pi_0(1 - \pi_0)} = 270.007 \text{ abgerundet,}$$

$$c_o = n\pi_0 + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{n\pi_0(1 - \pi_0)} = 299.592 \text{ aufgerundet.}$$

Also ist der Verwerfungsbereich gegeben durch:  $K = [0, 270] \cup [300, 356]$

6. **Testentscheid:** Der beobachtete Wert der Teststatistik ist  $x = 261$ . Da dieser Wert im Verwerfungsbereich  $K$  liegt, wird die Nullhypothese auf dem 5%-Signifikanzniveau verworfen. Die Zweifel von Lecker und Co., dass 80% der Kunden die neue Karte bevorzugen sind berechtigt.

Die zwei Lösungsansätze sind äquivalent, der Unterschied basiert lediglich auf unterschiedlichen Definitionen der Teststatistik.