Bachelorprüfung: Statistik Musterlösung

1. a) **(1 Punkt)** Gepaart (0.5 P), da jeweils dieselbe Person einmal mit dem einen Balltyp und einmal mit dem anderen Balltyp aufschlägt. (0.5 P)

b) **(1 Punkt)** $H_0 : \mu_X = \mu_Y$ (0.5 P), $H_A : \mu_X > \mu_Y$ (0.5 P)

c) **(2 Punkte)**

1. **Modell**: D_i i.i.d. $\sim \mathcal{N}(\mu_D, \sigma_D^2)$, $i = 1, \ldots, n$.
2. **Nullhypothese** $H_0 : \mu_X = \mu_Y$;
   Alternativhypothese $H_A : \mu_X > \mu_Y$
3. **Teststatistik:** $T = \dfrac{\sqrt{n}\hat{\mu}_D}{\hat{\sigma}_D}$
   Verteilung der Teststatistik unter H_0**:** $T \sim t_{11}$
4. **Signifikanzniveau:** $\alpha = 0.05$
5. **Verwerfungsbereich:**
   $K = (t_{11,1-\alpha}, \infty) = (1.796, \infty)$ (1 P)
6. **Testentscheid:** $t = \frac{\sqrt{12} \cdot 1.27}{3.386} = 1.30 \notin K$ (0.5 P), also wird H_0 beibehalten (0.5 P)

d) **(2 Punkte)** Das einseitige 95%-Vertrauensintervall ist gegeben durch

$$\left(\hat{\mu}_D - t_{11,1-\alpha}\frac{\hat{\sigma}_D}{\sqrt{n}}, \infty\right) = \left(1.27 - 1.796 \cdot \frac{3.386}{\sqrt{12}}, \infty\right) = (-0.485, \infty).$$

e) **(1 Punkt)** Wenn σ_D bekannt ist, müssen wir mit der Standardnormalverteilung und nicht mehr mit der t-Verteilung arbeiten. Der Verwerfungsbereich sieht dann wie folgt aus.

$$K = \left(\Phi^{-1}(1-\alpha), \infty\right)$$

2. a)
- R_n: Anzahl roter Ameisen aus n ausgewählten Ameisen.
- $p = 0.1$: Wahrscheinlichkeit, dass eine Ameise rot ist.

$$\begin{aligned} \mathrm{P}\left[R_5 = 3\right] &= \binom{5}{3} p^3(1-p)^2 \qquad \textbf{(0.5 P)} \\ &= 10 * 0.00081 = 0.0081 \qquad \textbf{(0.5 P)} \end{aligned}$$

b) **(2 Punkte)**

$$\begin{aligned} R_{150} &\sim \mathrm{Bin}\left(150, 0.1\right) \qquad \textbf{(0.5 P)} \\ \mathbf{E}\left[R_{150}\right] &= 150 \cdot 0.1 = 15 \qquad \textbf{(0.5 P)} \\ \mathrm{Var}\left(R_{150}\right) &= 150 \cdot 0.1 \cdot 0.9 = 13.5 \qquad \textbf{(1 P)} \end{aligned}$$

c) **(2 Punkte)**
Normalapproximation (i.o. da $150 * 0.1 * 0.9 = 13.5 > 9$):

$$\begin{aligned} R_{150} &\sim \mathcal{N}\left(15, 13.5\right) \\ \mathrm{P}\left[15 \leq R_{150} \leq 20\right] &= \mathrm{P}\left[0 \leq Z \leq \frac{20-15}{\sqrt{13.5}}\right] \qquad \textbf{(1 P)} \\ &= \mathrm{P}\left[Z \leq 1.36\right] - 0.5 \qquad \textbf{(0.5 P)} \\ &= 0.9131 - 0.5 = 0.4131 \qquad \textbf{(0.5 P)} \end{aligned}$$

d)
1. **Modell**: R_n: Anzahl rote Ameisen; $R_n \sim \mathrm{Binomial}(n, p)$ mit $n = 150$
2. **Nullhypothese** $H_0 : p = p_0 = 0.1$;
   Alternativhypothese $H_A : p \neq 0.1$. **(0.5 P)**
3. **Teststatistik:** $Z = \dfrac{R_n - np}{\sqrt{np(1-p)}}$
   Verteilung der Teststatistik unter H_0**:** $Z \approx \mathcal{N}(0, 1)$
4. **Signifikanzniveau:** $\alpha = 0.05$
5. **Verwerfungsbereich:**
   $K = (-\infty, -\Phi^{-1}(1-\frac{\alpha}{2})] \cap [\Phi^{-1}(1-\frac{\alpha}{2}), \infty)(-\infty, -1.96] \cap [1.96, \infty)$ **(0.5 P)**
6. **Testentscheid:** $z = 1.361 \notin K$ **(1 P)**, daher wird die Nullhypothese *nicht* verworfen. **(0.5 P)**

3. 1) d: QQ-Plot und Tukey-Anscombe Plot sehe beide gut aus.

2) c: Die Geradengleichung aus dem R-Output ist $y = 62.366 + 3.570 \cdot x$. Für $x = 130$ ergibt sich (c).

3) a: Ja, weil der p-Wert kleiner 5% ist.

4) d: Schätzwert (3.570) minus hypothetischer Wert (0) gemessen in Vielfachen des Std. Error: $t = \frac{3.570-0}{0.347}$ ergibt (d).

5) f: "degrees of freedom"(23) plus Anzahl geschätzter betas (2) gibt 25.

6) d: Wir brauchen das 99.5%-Quantil der t-Verteilung mit 23 Freiheitsgraden (2.807) und den Std. Error aus dem R-Output (0.347).

7) b: Abgekürzt: Nein, weil der t-Wert (2.382) grösser als das 97.5%-Quantil der t-Verteilung mit 23 Freiheitsgraden ist (2.069). Ausführlich: Aus Estimate und t-Value kann man Std. Error berechnen. Damit kann man mit den üblichen Formeln das exakte 95% Vertrauensintervall ausrechnen.

8) e: Die gesuchte Zahl steht bei "Residual standard error" im R-Output.

4.

1) d: Verwende $E[aX + b] = aE[X] + b$; $Var(aX + b) = a^2 Var(X)$
2) c
3) d: Eigentlich müsste man $f(x)$ integrieren (Stammfunktion ausrechnen) und kommt auf (d). Einfacher ist es wohl, wenn man die möglichen Lösungen für $F(x)$ ableitet und schaut, bei welcher Lösung $f(x)$ raus kommt.
4) d: Kontinuierliche Messgrösse, bei der viele verschiedene Einflussfaktoren zusammenspielen. Daher ist die Normalverteilung ein plausibler Kandidat.
5) b: Die Summe von exponentialverteilten Zufallsvariablen ist wieder exponentialverteilt (das wurde nicht jedes Jahr behandelt); wenn das nicht gelten würde, würde der Zentrale Grenzwertsatz als gute Approximation die Normalverteilung liefern.
6) c: Siehe Definition des p-Werts.