Bachelorprüfung: Statistik Musterlösung

1. **a)** Gepaart. **(0.5 P)** Jeder Stern wird jeweils mit dem neuen und dem traditionellen Verfahren vermessen. **(0.5 P)**

b) $H_0 : \mu_X = \mu_Y$, $H_A : \mu_X \neq \mu_Y$ **(1P)**

c) Der Verwerfungsbereich ist

$$K = \left(-\infty, -\frac{t_{9,97.5\%}0.46}{\sqrt{10}}\right) \cup \left(\frac{t_{9,97.5\%}0.46}{\sqrt{10}}, \infty\right) = (-\infty, -0.3290) \cup (0.3290, \infty).$$

(2 P) *wenn K komplett richtig, für jeden falschen Eintrag im Endresultat einen Punkt Abzug (minimal 0 Punkte).*

Da $-0.21 \notin K$, wird H_0 beibehalten. **(1 P)**

d) $\left[-0.21 - \frac{t_{9,97.5\%} \cdot 0.46}{\sqrt{10}}, -0.21 + \frac{t_{9,97.5\%} \cdot 0.46}{\sqrt{10}}\right] = [-0.5390, 0.1190]$ **(2P)** *Für jeden falschen Eintrag im Endresultat einen Punkt Abzug.*

e) Nein, denn die Standartabweichung ist nicht bekannt. **(1P)**

2. **a) (1 Punkt)** Wir benutzen folgende Notation: $R = \text{Reise}$; $A = \text{Absage}$.

$$P[3R\ 1A] = \binom{4}{3} \cdot 0.9^3 \cdot 0.1^1 = 0.2916 = 29.16\%$$

b) (1.5 Punkte) S_n sei der Anzahl Personen, die den Flug nehmen möchten. S_n ist binomialverteilt. Mit 28 Passagiere haben wir:

$$\begin{aligned} S_{28} &\sim Bin(28, 0.9) && (0.5Pt) \\ \mathbf{E}\left[S_{28}\right] &= 28 \cdot 0.9 = 25.2 && (0.5Pt) \\ \operatorname{Var}\left(S_{28}\right) &= 28 \cdot 0.9 \cdot 0.1 = 2.52 && (0.5Pt) \end{aligned}$$

c) (1 Punkt)

$$\begin{aligned} P[\text{Zu viele Leute}] &= P[k = 27] + P[k = 28] && (0.5Pt) \\ &= \binom{28}{27} \cdot 0.9^{27} \cdot 0.1^1 + \binom{28}{28} \cdot 0.9^{28} \cdot 0.1^0 \\ &= 0.1628 + 0.05233 = 0.215154 = 21.52\% && (0.5Pt) \end{aligned}$$

d) (1.5 Punkte) Mit einer Normalverteilung (0.5Pt),

$$\begin{aligned} & S_n \sim \mathcal{N}(np, npq) \\ \Rightarrow\ & S_{890} \sim \mathcal{N}(801, 80.1) && (0.5Pt\ ,\ 0.5Pt) \end{aligned}$$

e) (3 Punkte)

1. **Modell:** Anzahl erscheinende Personen: $S_n \sim \text{Binomial}(n, p)$, $n = 890$
2. **Nullhypothese:** $H_0 : p = p_0 = 0.9$
   Alternative: $H_A : p \neq 0.9$ (0.5 Pt)
3. **Teststatistik:** $Z = \dfrac{S_n - np_0}{\sqrt{np_0 q_0}}$. (1 Pt)
   Verteilung der Teststatistik unter H_0: $Z \approx \mathcal{N}(0, 1)$.
4. **Signifikanzniveau:** $\alpha = 0.05$
5. **Verwerfungsbereich für die Teststatistik:**

$$K = \Big(-\infty, -\Phi^{-1}(1 - \frac{\alpha}{2})\Big] \cup \Big[\Phi^{-1}(1 - \frac{\alpha}{2}), \infty\Big) = (-\infty, -1.96] \cup [1.96, \infty)\ \text{(0.5 Pt)}$$

6. **Testentscheid:** $z = 8.268 \in K$, daher wird die Nullhypothese deutlich verworfen. (0.5 Pt)

3. Je einen Punkt für:

1) d: `degrees of freedom` (90) plus Anzahl gefitteter betas (4) gibt 94.
2) a: β_1 gehört zu der Variable `log(cavol)`. Also müssen wir in der Tabelle in der Zeile `logcavol` und der Spalte `Std. Error` nachschauen. Der Wert wurde aber gelöscht; also muss man ihn aus `t value` und `Estimate` rekonstruieren: $\frac{0.642}{8.551} = 0.075$.
3) a: `Residual standard error` gibt die geschätzte Standardabweichung. Das Quadrat davon liefert die gesuchte Varianz.
4) b: Nein, weil der p-Wert grösser als 1% ist.
5) c: Dazu brauchen wir das 99.5%-Quantil der t-Verteilung mit 90 Freiheitsgraden (1.987), den Std. Error (0.181) und den Estimate (0.468).
6) c: In der Zeile `age` wurden zu viele Informationen gelöscht; man kann den p-Wert nicht mehr bestimmen und daher die Aufgabe nicht lösen.
7) b: Exakt sind die Modellannahmen nie erfüllt (jedes Modell ist falsch, aber manche sind nützlich). Gemäss der plots scheinen die Annahmen aber immerhin plausibel. An den beiden Enden des QQ-Plots kann man Abweichungen von der Gerade erkennen, was auf Ausreisser hinweist.

4. Je einen Punkt für:

1) a: Verwende $E[aX+b] = aE[X] + b$ und $Var(aX+b) = a^2 Var(X)$.
2) c: Die Ableitung von $F(x)$ ergibt $f(x)$.
3) a: $F(x) = 0.5$ nach x auflösen.
4) a: Poisson-Approximation (wurde nicht in jedem Vorlesungsjahr besprochen)
5) f: Wartezeit, nicht-negativ, nach oben offen: Exponential-Verteilung
6) a: Formel für einseitiges Vertrauensintervall im Skript nachschauen.