Bachelorprüfung: Statistik Musterlösung

1. a)

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2 \ (= 235.84)^2 \ \textbf{(1 Punkt)}$$

wobei $\bar{d} = -20$.

b) Gepaart. Jeder Fahrer fährt einmal mit und einmal ohne System **(1 Punkt)**.

c) $H_0 : \mu \geq 0,\ H_A : \mu < 0$ **(1 Punkt)**.

d)
1. **Modell:** $d_1, \ldots, d_n$ i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$
2. **Nullhypothese:** $\mu \geq \mu_0 = 0$
   Alternative: $\mu < 0$
3. **Teststatistik:** $T = \dfrac{\sqrt{n}(\bar{d} - \mu_0)}{\hat{\sigma}}$
   Verteilung der Teststatistik unter H_0: $T \sim t_{n-1}$.
4. **Signifikanzniveau:** $\alpha = 0.05$
5. **Verwerfungsbereich für die Teststatistik:**

$$K = (-\infty, -t_{n-1,1-\alpha}] = (-\infty, -2.132]$$

6. **Testentscheid:** $t = \frac{\sqrt{5}\cdot(-20)}{235.84} = -0.190 \notin K$, daher wird die Nullhypothese beibehalten. **(2 Punkte)**

e) $I = \left(-\infty, \bar{d} + \dfrac{\hat{\sigma}}{\sqrt{n}} t_{n-1,1-\alpha}\right] = (-\infty, 204.85]$ **(1 Punkt)**

f) In diesem Fall ist die Teststatistik $T = -0.71 \notin K$; also wird die Nullhypothese beibehalten. **(1 Punkt)**
Eine grosse Varianz bedeutet, dass (für festes Signifikanzlevel) die Nullhypothese nur schwer verworfen werden kann. Denn wenn die Werte stark streuen, kann nur mit kleiner Sicherheit angenommen werden, dass die Nullhypothese nicht doch zutrifft. Somit wird hier die Nullhypothese – die Varianz ist zwar deutlich kleiner als oben, aber immer noch sehr gross – wieder nicht verworfen. **(1 Punkt)**

2. **a)** $0.8 \cdot 356 = 284.8$ **(1 Punkt)**

b) $0.2^4 \cdot 0.8 = 0.0013$, **(1 Punkt)** bzw.
$4 \cdot 0.2 \cdot 0.8^3 = 0.4096$ **(1 Punkt)**

c) $Z = \dfrac{S_n - n\pi}{\sqrt{n\pi(1-\pi)}}$ ist normalverteilt. Es gilt hier $Z = -3.153$. Also

$$P(Z \leq -3.153) = \Phi(-3.153) = 1 - \Phi(3.153) \approx 8 \cdot 10^{-4}$$ **(2 Punkt)**

d)
1. **Modell:** Anzahl zufriedene Kunden: $S_n \sim \text{Binomial}(n, \pi)$, $n = 356$.
2. **Nullhypothese:** $H_0 : \pi = \pi_0 = 0.8$
   Alternative: $H_A : \pi \neq 0.8$ **(1 Punkt)**
3. **Teststatistik:** $Z = \dfrac{S_n - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}}$.
   Verteilung der Teststatistik unter H_0: $Z \approx \mathcal{N}(0,1)$.
4. **Signifikanzniveau:** $\alpha = 0.05$
5. **Verwerfungsbereich für die Teststatistik:**

$$K = \left(-\infty, -\Phi^{-1}(1-\frac{\alpha}{2})\right] \cup \left[\Phi^{-1}(1-\frac{\alpha}{2}), \infty\right) = (-\infty, -1.960] \cup [1.960, \infty)$$

6. **Testentscheid:** $z = -3.153 \in K$, daher kann die Nullhypothese verworfen werden.

3. Je einen Punkt für:

1) a: Anzahl Beobachtungen minus Anzahl geschätzter betas: 16-2 = 14
2) b: `Residual standard error` im Quadrat.
3) e: Der t-Wert ist der geschätzte Parameter (-80.667) minus der hypothetische Parameter (0) gemessen in Vielfachen von `Std. Error`: $t = \frac{-80.667-0}{1.420} = -56.81$.
4) c: Wir brauchen den Estimate (-80.667), den zugehörigen Std. Error (1.420) und das 97.5%-Quantil der t-Verteilung mit 14 Freiheitsgraden (2.145).
5) b: Wir brauchen das 99.5%-Quantil der t-Verteilung mit 14 Freiheitsgraden (2.977), den Estimate (0.521) und den zugehörigen Std. Error (0.007). Daraus ergibt sich das 99%-Vertrauensintervall. Es enthält die 0 nicht. (Man hätte das Quantil gar nicht nachschauen müssen. Der Std. Error ist so viel kleiner als der Esitmate, dass man sofort sehen kann, dass das Vertrauensintervall den Wert 0 nicht enthält.)
6) a: Berechne 99%-Vertrauensintervall von β_0. Es enthält die 0 nicht, also wird H_0 verworfen. (Wie vorhin kann man das auch sofort und ohne Tabelle sehen.)
7) a: In Geradengleichung $y = -80.667 + 0.521 \cdot x$ einsetzen.
8) d: Im Tukey-Anscombe Plot (links) sieht man eine v-Form. Das spricht für einen systematischen Fehler.

4. Je einen Punkt für:

1) e: Für unabhängige Zufallsvariablen gilt $Var(X+Y) = Var(X) + Var(Y)$.
2) f: Im allgemeinen hat die Funktion einer Zufallsvariable eine ganz andere Verteiung als die Zufallsvariable selbst.
3) d: $P(R = ungerade|S = 4) = \frac{P(R=ungerade \cap S=4)}{P(S=4)}$. Die möglichen Elementarereignisse im Zähler sind (r,b) = (1,3) oder (r,b) = (2,2). Es gibt also zwei solche Elementarereignisse. Die möglichen Elementarereignisse im Nenner sind (r,b) = (1,3), (r,b) = (2,2), (r,b) = (3,1). Es gibt also drei solche Elementarereignisse. Der Quotient ist also $\frac{2}{3}$.
4) a: Immer, weil bei einer stetigen Zufallsvariable gilt: $P[X = x] = 0$ für alle Werte von x.
5) b: Da 0.05 im 95%-Vertrauensintervall liegt, kann H_0 auf dem 5% Niveau nicht verworfen werden. Dann kann sie erst recht nicht bei dem strengeren Niveau von 1% verworfen werden.
6) f: $E[X] = -1000 \cdot \frac{36}{37} + 35000 \cdot \frac{1}{37} \approx -27$. Diese Zahl steht nicht in der Auswahl, also ist (f) richtig.
7) e (1b,2c,3a)