Bachelorprüfung: Mathematik IV – Statistik
Musterlösung

1. **a)** **(Total 1P)** Der Test muss gepaart sein, da zu jedem Läufer genau ein Wert mit jedem Schuhtyp vorliegt und die beiden Werte eines Läufers von dessen Lauffähigkeiten abhängen.

b) **(Total 1P)** Nullhypothese: $H_0 : \mu_x = \mu_y$ **(1/2P)**
Alternativhypothese: $H_A : \mu_x < \mu_y$ **(1/2P)**

c) **(Total 3 1/2P)** Für den gepaarten t-Test müssen wir die Kennzahlen der Differenz, $\bar{x} - \bar{y} = -0.22$ und $\hat{\sigma}_{x-y} = 0.26$, verwenden **(1/2P)**. Der Wert der Teststatistik ist

$$T = \frac{\sqrt{n}(\bar{x} - \bar{y})}{\hat{\sigma}_{x-y}} = \frac{\sqrt{10}(46.02 - 46.24)}{0.26} = -2.68 \ . \ (\mathbf{1P})$$

Verworfen wird die Nullhypothese, wenn die Teststatistik extremer ist (in negativer Richtung) als das 95%-Quantil der t-Verteilung mit 9 Freiheitsgraden. Der Verwerfungsbereich ist also $(-\infty, -t_{9;0.95}) = (-\infty, -1.833)$ **(1P)**. Da die Teststatistik in diesem Bereich liegt, wird beim t-Test hier die Nullhypothese verworfen **(1P)**.

d) **(Total 2 1/2P)** Das Vertrauensintervall erstreckt bis $t_{9;0.95}/\sqrt{10}$-mal die Schätzung der Standardabweichung oberhalb der Schätzung des Durchschnitts $\mu_x - \mu_y$. D.h., es ist

$$\begin{aligned} & (-\infty, \bar{x} - \bar{y} + \hat{\sigma}_{x-y} \cdot t_{9;0.95}/\sqrt{10}) \ (\mathbf{1P}) \\ = \ & (-\infty, -0.22 + 0.26 \cdot 1.833/\sqrt{10}) = (-\infty, -0.069) \ . \ (\mathbf{1P}) \end{aligned}$$

2. **a)** **(1 Punkt)** Anzahl gewonnener Pokerrunden durch Markus X ist $\mathrm{Bin}\,(n, \pi)$.

b) **(1 Punkt)** Daher ist

$$\mathbf{E}\,[X] = n\pi = 20 * 0.2 = 4 \; (\mathbf{1/2P})$$

und

$$\mathrm{Var}\,(X) = n\pi(1-\pi) = 20 * 0.2 * 0.8 = 3.2, \; (\mathbf{1/2P})$$

d.h. es werden 4 Siege von Markus erwartet.

c) **(2 Punkte)** Verwende Zentralen Grenzwertsatz:

$$Y = \sum_{k=1}^{100} X_i \; (\mathbf{1/2P})$$

$$Y \sim \mathcal{N}\left(20, 4^2\right) \; (\mathbf{1/2P})$$

Daher ist

$$\begin{aligned} \mathrm{P}\,[Y >= 28] &= \mathrm{P}\left[\frac{Y-20}{4} >= 2\right] = 1 - \mathrm{P}\,[Z < 2] \; (\mathbf{1/2P}) \\ &= 1 - \mathrm{P}\,[Z \leq 2] = 1 - 0.977 = 0.022 = 2.2\% \; (\mathbf{1/2P}) \end{aligned} \tag{1}$$

d) **(2 Punkte)**
H_0: Markus spielt gleich gut wie alle anderen Spieler; $\pi_0 = 0.2$
H_A: Markus spielt besser als die anderen Spieler; $\pi > \pi_0 = 0.2$
$T : Y = \sum_{k=1}^{100} X_i \sim \mathcal{N}\,(20, 4^2)$
(1/2P)
Die Wahrscheinlichkeit von 28 oder mehr durch Markus gewonnener Runden unter der Nullhypothese ist $2.2\% > 1\%$. Daher kann die Nullhypothese auf dem 1% nicht verworfen werden. **(1/2P)**
Das 99%-Quantil der Standard-Normalverteilung ist 2.326. Daher ist das 99%-Quantil der Verteilung von Y unter der Nullhypothese gleich $20 + 2.326 * 4 = 29.3$ und der Verwerfungsbereich des Test $[30, 100]$. **(1P)**

e) **(2 Punkte)** Die geschätzte Gewinnwahrscheinlichkeit von Markus ist

$$\hat{\pi} = \frac{X}{n} = 0.28 \; (\mathbf{1P})$$

Das zweiseitige 95% Vertrauensintervall von π ist

$$[\hat{\pi} \pm 1.96\sqrt{\hat{\pi}(1-\hat{\pi})/n}] = [0.28 \pm 1.96\sqrt{0.28 * (1-0.28)/100}] = [0.192, 0.368] \; (\mathbf{1P})$$

3. Im Folgenden bezeichne f die Trefferquote von Fritz und h die Trefferquote von Hansi, d.h. $f = 0.6$ und $h = 0.95$. Wir nehmen an, dass sämtliche Schüsse unabhängig voneinander abgefeuert werden; wir wenden deshalb jeweils den Multiplikationssatz für unabhängige Ereignisse an.

a) **(1 Punkt)** $f^5 \approx 0.077$, d.h. mit ca. 8%-iger Wahrscheinlichkeit trifft Fritz 5-mal hintereinander.

b) **(2 Punkte)** Es bezeichne X die benötigte Anzahl Schüsse. Das Ereignis $\{X = n\}$ tritt ein, wenn Fritz zunächst $(n-1)$-mal verfehlt und dann trifft, d.h.

$$\mathrm{P}\left[X = n\right] = (1-f)^{n-1} f.$$

Insbesondere gilt

$$\mathrm{P}\left[X > 3\right] = 1 - \mathrm{P}\left[X \leq 3\right] = 1 - (f + (1-f)f + (1-f)^2 f) = 0.064.$$

Mit 6%-iger Wahrscheinlichkeit braucht Fritz also mehr als 3 Schüsse.

c) **(2 Punkte)**

$$\begin{aligned}\mathrm{P}\left[\text{Ziel wird getroffen}\right] &= 1 - \mathrm{P}\left[\text{Fritz und Hansi verfehlen beide das Ziel}\right] \\ &= 1 - (1-f)(1-h) = 0.98,\end{aligned}$$

das Ziel wird mit also mit 98%-iger Wahrscheinlichkeit getroffen.

d) **(2 Punkte)** Die Wahrscheinlichkeit, dass Hansi verfehlt und Fritz mindestens einmal trifft ist $p_1 = (1-h)(1-(1-f)^3) = 0.0468$. Die Wahrscheinlichkeit, dass Hansi trifft und Fritz mindestens zweimal trifft ist $p_2 = h\left(\binom{3}{2} f^2(1-f) + f^3\right) = 0.6156$. Da sich diese beiden Ereignisse gegenseitig ausschliessen, dürfen wir die Einzelwahrscheinlichkeiten addieren; wir erhalten $p_1 + p_2 \approx 0.66$, d.h. Fritz gewinnt mit 66%-iger Wahrscheinlichkeit.

e) **(3 Punkte)** Gesucht ist die Wahrscheinlichkeit des Ereignisses $A = \bigcup_{j=1}^{\infty} A_j$. Weil die A_j sich gegenseitig ausschliessen, gilt:

$$\begin{aligned}\mathrm{P}\left[A\right] &= \sum_{j=1}^{\infty} \mathrm{P}\left[A_j\right] = \sum_{j=1}^{\infty} (1-f)^{j-1}(1-h)^{j-1} f = f \sum_{j=1}^{\infty} ((1-f)(1-h))^{j-1} \\ &= f \sum_{j=0}^{\infty} ((1-f)(1-h))^{j} = \frac{f}{1-(1-f)(1-h)} \approx 0.61.\end{aligned}$$

Fritz trifft mit 61%-iger Wahrscheinlichkeit als erster.

4. 1) **(1 Punkt)** Die Anzahl Beobachtungen n ist gegeben durch die Summe der Anzahl Freiheitsgrade und der Anzahl Parameter $n = n - p + p = 9 + 2 = 11$. Somit ist Antwort c) korrekt.

2) **(1 Punkt)** Der Intercept lässt sich schnell aus den Werten im angegebenen R-Output berechen. Man multipliziert die t-Teststatistik mit der Standardabweichung und erhält so den Intercept. Dass heisst, $51.25 \cdot -1.831 \approx -93.8$. Demnach ist Antwort b) richtig.

3) **(1 Punkt)** Um diese Frage zu beantworten betrachtet man den entsprechenden P-Wert. Ist der P-Wert kleiner als 0.05 wird die Nullhypothese auf dem 5% Niveau verworfen. Der P-Wert für $H_0 : \beta = 0$ und $H_A : \beta \neq 0$ ist, wie man aus dem R-Output erkennt, $0.00249 < 0.05$. Somit ist Antwort a) richtig.

4) **(1 Punkt)** Die t-Teststatistik lässt sich schnell aus den Werten im angegebenen R-Output berechen. Die t-Teststatistik erhält man nämlich, indem man den geschätzten Wert durch seine Standardabweichung teilt. Somit ergiebt sich $\frac{59.05}{14.24} \approx 4.15$. Richtig ist also Antwort d).

5) **(1 Punkt)** Das zweiseitige Vertrauensintervall für β ist gegeben durch $\hat{\beta} \pm t_{9;0.975} \cdot \hat{\sigma}_\beta$. Das Quantil $t_{9;0.975}$ lässt sich in der Tabelle der t-Verteilung nachschauen, es gilt $t_{9;0.975} = 2.262$. Somit ist Antwort d) korrekt.

6) **(1 Punkt)** Der Tukey-Anscombe Plot zeigt, dass die Annahme der konstanten Varianz der Fehler ϵ_i verletzt ist. Somit ist ist Aussage c) zutreffend.

7) **(1 Punkt)** Durch das hinzufügen einer Beobachtung an der Stelle $(2.5, 150)$ wird die Regressionsgerade flacher, dass bedeutet $\hat{\beta}$ wird kleiner. Weiter wird $\hat{\sigma}$ grösser, da durch das Hinzufügen der Beobachtung die Residuen mehrheitlich dem Betrage nach grösser werden und somit auch $\hat{\sigma}$. Das heisst Antwort a) ist korrekt.

8) **(1 Punkt)** Bei der einfachen linearen Regression erhält man die empirische Korrelation zwischen erklärender Variable und Zielvariable aus der Wurzel von R^2. Somit ist $\hat{\rho} = \sqrt{R^2} = \sqrt{0.6566} \approx 0.81$, also ist Antwort e) korrekt.

9) **(1 Punkt)** Die Nullhypotesen $H_{0,1} : \beta_1 = 0$ und $H_{0,2} : \beta_2 = 0$ werden nicht verworfen. Aber der F-Test hat einen kleinen P-Wert. Somit ist Antwort a) richtig.

10) **(1 Punkt)** In der multiplen Regression lässt sich R^2 nicht mehr als quadratische Stichprobenkorrelation zwischen erklärenden Variablen und Zielvariablen interpretieren. Somit ist Antwort b) richtig.

Übersicht

1) c)
2) b)
3) a)
4) d)
5) d)
6) c)

7) a)

8) e)

9) a)

10) b)

5. 1) b)

2) f)
$X + Y = -2 + \frac{5}{4}X$. Also gilt $\mathbf{E}[X + Y] = -2 + \frac{5}{4}2 = 0.5$.

3) b)
Siehe 2: $X + Y = -2 + \frac{5}{4}X$. Also ist $\text{Var}(X + Y) = \frac{25}{16}\,\text{Var}(X) = \frac{25}{16}$. Bem.: X und Y sind nicht unabhängig, daher darf man die Varianzen nicht addieren!

4) c)
Zentraler Grenzwertsatz: Summe von vielen Zufallsvariablen ist (approx.) normalverteilt.

5) a)
Da die Teststatistik grösser 0 ist (d.h. sie "liegt auf der richtigen Seite") ist der p-Wert 0.01.

6) b)
$\sqrt{2}$ ist das 0.8416 Quantil der $\mathcal{N}(0, 2)$ Verteilung.

7) b)