Bachelorprüfung: Statistik Musterlösung

1. a) **(1 Punkt)** Es muss ein ungepaarter Test durchgeführt werden, da die Test- und Kontrollmessungen auf unterschiedlichen Pisten durchgeführt werden (und nicht je 1 Test- und 1 Kontrollmessung pro Piste). Auf einen ungepaarten Test weist zudem die unterschiedliche Anzahl Messungen in den beiden Gruppen klar hin.

   b) **(2 Punkte)** Die gepoolte Varianz berechnet sich aus den Varianzen der einzelnen Messgruppen durch $S_{pool}^2 = \frac{1}{26+36-2} \cdot ((26-1) \cdot \hat{\sigma}_X^2 + (36-1) \cdot \hat{\sigma}_Y^2)$, also ergibt sich umgekehrt:

$$\hat{\sigma}_Y^2 = \frac{1}{36-1} \cdot ((26+36-2) \cdot S_{pool}^2 - (26-1) \cdot \hat{\sigma}_X^2) = 19.6 \ .$$

   c) **(2 Punkte)** Die Grösse, über die man etwas zeigen will, ist der mittlere Unterschied $\mu_X - \mu_Y$ der Schneeschmelzraten zwischen Behandlungs- und Kontrollgruppe. Der Effekt, den man nachweisen möchte, ist ein einseitiger. Also haben wir die Nullhypothese

$$H_0 : \ \mu_X - \mu_Y = 0$$

   und die Alternativhypothese

$$H_1 : \ \mu_X - \mu_Y < 0$$

   .

   d) **(4 Punkte)** Ein (einseitiges, da wir einseitig testen) Konfidenzintervall für die erwartete Differenz $\mu_X - \mu_Y$ konstruiert man aus dem Konfidenzintervall $(-\infty, t_{60,95\%})$, das man noch mit der geschätzten Standardabweichung $S_{pool} * \sqrt{1/26 + 1/36}$ von $\hat{\mu}_X - \hat{\mu}_Y$ skalieren und um den Schätzer $\hat{\mu}_X - \hat{\mu}_Y$ verschieben muss. Insgesamt ergibt sich also das folgende Intervall als Konfidenzintervall:

$$(-\infty, t_{60,95\%} * S_{pool} * \sqrt{1/26 + 1/36} + (\hat{\mu}_X - \hat{\mu}_Y)) = (-\infty, -2.52) \ .$$

   Da 0, der Wert von $\mu_X - \mu_Y$ unter der Nullhypothese, ausserhalb des ermittelten Konfidenzintervalls liegt, wird der entsprechende t-Test (ungepaart, beim Niveau 5%) die Nullhypothese verwerfen.

   e) **(1 Punkt)** Ein Vorzeichentest setzt gepaarte Daten voraus – die hier nicht vorliegen.

2. **a)** **(1 Punkt)** Man hat ein seltenes Ereignis (Unfall eines einzelnen Verkehrsteilnehmers) bei vielen unabhängigen Versuchen (insgesamt viele Verkehrsteilnehmer).

b) **(3 Punkte)** Da der Parameter λ gleich dem Erwartungswert ist, verwenden wir den Stichprobenmittelwert als Schätzung für λ:

$$\hat{\lambda} = \frac{1}{52}(0 \cdot 4 + 1 \cdot 8 + \ldots + 6 \cdot 3 + 7 \cdot 1) = \frac{150}{52} \approx 2.88.$$

Die Wahrscheinlichkeit, dass eine $\mathrm{Pois}\left(\hat{\lambda}\right)$-verteilte Zufallsvariable einen Wert grösser oder gleich 3 annimmt, ist:

$$1 - \sum_{k=0}^{2} e^{-\hat{\lambda}} \frac{\hat{\lambda}^k}{k!} \approx 0.55.$$

c) **(1 Punkt)** Die jährliche Unfallzahl ist Poisson-verteilt mit Parameter $52 \cdot 3 = 156$.

d) **(2 Punkte)** Nullhypothese: Die Anzahl Unfälle im Jahr 2007 ist Poisson-verteilt mit Parameter $\lambda_0 = 156$. Alternative: Die Anzahl Unfälle im Jahr 2007 ist Poisson-verteilt mit Parameter $\lambda_1 < \lambda_0$.
Der Test ist einseitig durchzuführen, da wir wissen wollen, ob die Anzahl Unfälle abgenommen hat.

e) **(3 Punkte)** Wir verwenden $N \sim \mathrm{Pois}\,(\lambda_0)$ als Teststatistik. Gesucht ist $c \in \mathbb{N}$, so dass $\mathrm{P}\,[N \leq c] \leq 0.05$. Die Zufallsvariable N ist in guter Näherung normalverteilt mit Erwartungswert λ_0 und Varianz λ_0; wir erhalten:

$$c \approx \lambda_0 + \sqrt{\lambda_0} \cdot \Phi^{-1}(0.05) \approx 135.$$

Der Verwerfungsbereich zum 5%-Niveau ist daher gegeben durch $\{N \leq 135\}$; wir verwerfen die Nullhypothese, es besteht ein signifikanter Unterschied in den Unfallzahlen zwischen 2006 und 2007. Der P-Wert beträgt $\Phi\left(\frac{131-\lambda_0}{\sqrt{\lambda_0}}\right) \approx 0.023$.

3. a) (2 Punkte)

- S_n: Anzahl Männer die Medizin studieren.
- $p = 0.5$: Wahrscheinlichkeit, dass ein Medizinstudent ein Mann ist (Männer und Frauen gleich verteilt)

$$\begin{aligned} S_{22} &\sim \operatorname{Bin}(22, 0.5) \qquad \textbf{(1 P)} \\ \mathbf{E}[S_{22}] &= 22 \cdot 0.5 = 11 \qquad \textbf{(0.5 P)} \\ \operatorname{Var}(S_{22}) &= 22 \cdot 0.5 \cdot 0.5 = 5.5 \qquad \textbf{(0.5 P)} \end{aligned}$$

b) (1 Punkt)

$$\begin{aligned} \mathrm{P}[S_{22} = 10] &= \binom{22}{10} \cdot 0.5^{10} \cdot 0.5^{22-10} \qquad \textbf{(0.5 P)} \\ &= 0.154 \qquad \textbf{(0.5 P)} \end{aligned}$$

c) (2.5 Punkte)

Normalapproximation:

$$\begin{aligned} S_{22} &\sim \mathcal{N}(11, 5.5) \\ \mathrm{P}[S_{22} > 10] &= 1 - \mathrm{P}[S_{22} < 10] \qquad \textbf{(0.5 P)} \\ &= 1 - \mathrm{P}\left[\frac{S_{22} - 11}{\sqrt{5.5}} < -0.426\right] \qquad \textbf{(0.5 P)} \\ &= \mathrm{P}[X < 0.426 = 0.43] = 0.6664 \end{aligned}$$

Poissonapproximation:

$$\begin{aligned} S_{22} &\sim \operatorname{Pois}(n \cdot p) \\ &= \operatorname{Pois}(11) \qquad \textbf{(0.5 P)} \\ \mathrm{P}[S_{22} = 10] &= \frac{11^{10} \cdot \exp(-11)}{10!} \\ &= 0.119 \qquad \textbf{(0.5 P)} \end{aligned}$$

Die Poissonapproximation ist in diesem Fall nicht angebracht, da n zu klein (und p zu gross) ist. **(0.5 P)**

d) (3 Punkte)

- Normalapproximation: $S_{22} \sim \mathcal{N}(11, 5.5)$ **(0.5 P)**
- Nullhypothese vs. Alterntivhypothese

$$\begin{aligned} H_0 &: p = 0.5 \\ H_1 &: p \neq 0.5 \qquad \textbf{(0.5 P)} \end{aligned}$$

- 2 seitiger Z-Test auf 5% Niveau: Verwerfe Nullhypothese falls

$$|u| > u_{1-\frac{\alpha}{2}} = 1.96 \qquad \textbf{(0.5 P)}$$

wobei

$$u = \frac{16 - 11}{\sqrt{5.5}} \qquad \textbf{(1 P)}$$

- Testentscheid: H_0 wird verworfen, da $u = 2.132 > 1.96$. **(0.5 P)**

e) (1.5 Punkte)

$$u = u_{1-\frac{p}{2}} = 2.13 \Rightarrow \frac{p}{2} = \mathrm{P}\left[X > 2.13\right] \Rightarrow p = 0.0332 \qquad \textbf{(1 P)}$$

für Signifikanzniveaus $> 3.3\%$ wird die Nullhypothese verworfen **(0.5 P)**

4. 1) c.

2) b, logarithmieren der Zielvariable ist eine Standardlösung wenn ein kegelförmiges Anwachsen der Streuung mit $\hat{y}_i$ auftritt.

3) a, $n = 14$ und $p = 2$.

4) b, $\frac{\hat{\beta}_1}{\widehat{s.e.}(\hat{\beta}_1)} = 0.95465/0.06349 = 15.036$.

5) a, $t_{12;0.975} = 2.179 < 15.036$.

6) d, $\hat{\beta}_1 \pm \widehat{s.e.}(\hat{\beta}_1) \cdot t_{12;0.975}$.

7) c, es wird einen zweiseitigen Test ausgeführt.

8) e, $\log(\text{p1980}) = 1.13 + 0.95 \log(42.3) = 4.71$ und $\text{p1980} = \exp(4.71) = 110.75$.

9) d, nach Definition von R^2.

10) a, der F-Test hat einen sehr kleinen p-Wert.

5. 1) $\mathrm{P}[T \geq 1000] = 1 - \mathrm{P}[T < 1000] = 1 - (1 - \exp(-0.0004916 \cdot 1000)) = \exp(-0.4916) = 0.6116$, also b).

2) d).

3) $\mathrm{P}[T \leq t_H] = 0.5 \Leftrightarrow 1 - \exp(\lambda \cdot t_H) = 0.5 \Leftrightarrow \exp(\lambda \cdot t_H) = 0.5 \Leftrightarrow \lambda \cdot t_H = \ln(2)$ $\Leftrightarrow t_H = \ln(2)/\lambda \approx 1410$, also c)

4) Da man eine stetige Verteilungsfunktion haben will, muss $F(16) = \alpha \cdot 16^{1/4} = \alpha \cdot 2 = 1$ gelten, d.h. $\alpha = 1/2$, also b).

5) $F(m) = 1/2 \Leftrightarrow 1/4 \cdot m^{1/2} = 0.5 \Leftrightarrow m^{1/2} = 2 \Rightarrow m = 4$, also c).

6) $f(x) = 1/8 \cdot x^{-1/2}$ auf $[0, 16]$. $\mathbf{E}[X] = \int_0^{16} 1/8 \cdot x^{1/2} dx = 16/3$, also f).

7) $\mathrm{Var}(Y) = 1/9 \cdot \mathrm{Var}(X) = 1/9 \cdot 9 = 1$, also e).

8) $Y \sim \mathcal{N}(\mathbf{E}[Y] = 3, \mathrm{Var}(Y) = 1) \Rightarrow \mathrm{P}[Y \geq 3] = \Phi(0) = 0.5$, also c),

9) Y hängt linear von X ab mit einer positiven Steigung, daher ist die Korrelation 1, also e).