Bachelorprüfung: Statistik Musterlösung

1. **a)** **(1P)** ungepaart (0.5 P); Bücher werden in jeder Buchhandlung zufällig ausgewählt. Die beiden Stichproben sind also unabhängig von einander (0.5 P).

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

c) **(2.5P)**

1. **Modell**: X_i i.i.d. $\sim \mathcal{N}(\mu_X, \sigma_X^2)$, $i = 1, \ldots, n_1$, und Y_i i.i.d. $\sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, $i = 1, \ldots, n_2$.
2. **Nullhypothese** $H_0 : \mu_X = \mu_Y$;
   Alternativhypothese $H_A : \mu_X > \mu_Y$
3. **Teststatistik:**
   $$T \overbrace{=}^{0.5P} \frac{\bar{X} - \bar{Y}}{s_{\text{pool}}} \sqrt{\frac{1}{1/n_1 + 1/n_2}}$$
   Verteilung der Teststatistik unter H_0**:** $T \sim t_{28}$
4. **Signifikanzniveau:** $\alpha = 0.05$
5. **Verwerfungsbereich:**
   $K = (t_{28,1-\alpha}, \infty)$ (0.5 P) $= (1.701, \infty)$ (0.5 P)
6. **Testentscheid:**
   $$t = \frac{18 - 15.2}{3.45} \sqrt{\frac{1}{1/15 + 1/15}} \overbrace{=}^{0.5P} 2.223 \in K,$$
   also wird H_0 verworfen (0.5 P)

d) **(2P)** $[\bar{X} - \bar{Y} - t_{28,1-\alpha} \cdot s_{\text{pool}} \cdot \sqrt{1/n_1 + 1/n_2}, +\infty) = [0.621, +\infty)$ (1P+1P)

2. a) **(1 Punkt)**

- S_n: Anzahl Flaschen mit Defekt.
- $p = 0.1$: Wahrscheinlichkeit, dass eine Flasche einen Mangel aufweist.

$$\begin{aligned}\mathrm{P}[S_5 = 3] &= \binom{5}{3} p^3(1-p)^2 \quad \textbf{(0.5 P)}\\ &= 10 * 0.00081 = 0.0081 \quad \textbf{(0.5 P)}\end{aligned}$$

b) **(1 Punkt)**

$$\begin{aligned}\mathrm{P}[\text{die ersten drei Flaschen haben einen Mangel}] &= p^3(1-p)^2 \quad \textbf{(0.5 P)}\\ &= 0.00081 \quad \textbf{(0.5 P)}\end{aligned}$$

c) **(2 Punkte)**

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

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

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

e)
1. **Modell**: S_n: Anzahl Flaschen mit Mangel; $S_n \sim \mathrm{Binomial}(n, p)$ mit $n = 150$
2. **Nullhypothese** $H_0 : p = p_0 = 0.1$;
   Alternativhypothese $H_A : p \neq 0.1$. **(0.5 P)**
3. **Teststatistik:** $Z = \dfrac{S_n - np}{\sqrt{np(1-p)}}$ **(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:** $z = 1.361 \notin K$, daher wird die Nullhypothese *nicht* verworfen. **(0.5 P)**

3. 1) d: “degrees of freedom” plus “Anzahl geschätzter betas” = 28 + 2 = 30

2) a: Muss man aus dem Streudiagramm schätzen; zwischen 25 und 50 steigt die Gerade um etwa 600. D.h., bei groesse = 0 hat sie etwa die Höhe 200. Die einzige Lösung in dieser Grössenordnung ist (a).

3) c: Der t-Wert misst die Abweichung zwischen geschätztem Parameterwert (22.81) und hypothetischem Parameterwert (0) in Einheiten des Std. Error (1.93). Also: $t = \frac{22.81-0}{1.93} = 11.819$.

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

5) d: Man braucht das 97.5% Quantil der t-Verteilung mit 28 Freiheitsgraden (2.048) und den Std. Error (1.93).

6) c: Der gesuchte Wert ist $\hat{\beta}_1 \cdot 70 + \hat{\beta}_0$. Weil $\beta_1 = 22.81$ ist (c) richtig.

7) a: Der QQ-Plot sieht gut aus, aber beim Tukey-Anscombe Plot sieht man eine trichterförmige Öffnung. Also ist die Annahme der konstanten Fehlervarianz verletzt.

8) d: Die geschätzte Gerade minimiert die Summer der quadratischen, vertikalen Abweichungen von den Punkten. Wenn bei (80,3000) noch ein Punkt hinzugefügt wird (der liegt dann deutlich über der bisherigen Geraden), wird die Gerade steiler, um auch diesen Punkt einigermassen gut approximieren zu können. Ausserdem liegen die Punkte tendenziell weiter von der Geraden weg. Also ist (d) richtig.

9) b: Hier wird wieder ein Ausreisser hinzugefügt, diesmal allerdings unterhalb der Geraden. Dadurch wird die Gerade flacher, d.h., $\hat{\beta}_1$ wird kleiner. Ausserdem streuen die Punkte nun tendenziell mehr um die Gerade. Also wird $\hat{\sigma}$ und deshalb auch $\sigma\beta_1$ grösser. Der t-Wert ist der Quotient $\frac{\hat{\beta}_1}{\sigma\beta_1}$; der Zähler wird kleiner, der Nenner wird grösser; also wird der Quotient kleiner.

4. 1) b) The margin of error ME equals 1/2 the length of the confidence interval. $ME = z^*\sigma/\sqrt{n} = 1.96 * 5/\sqrt{n}$. Solve for n.

2) d) Properties of Poisson Distribution.

3) b) An increase in the level of significance is associated with an increase in the Power.

4) d) Set $F(x) = 0.5$ and solve for x.

5) a) The data is not normally distributed, so we should use the Wilcoxon Test. The P-value is less than 0.05, and so we reject the null hypothesis.