Statistik 2, Wiederholungsübung

Henry Haustein

Aufgabe 1

(a) Es gibt genau $\binom{49}{6}$ verschiedene Möglichkeiten beim 6-aus-49-Lotto. Wir wollen jetzt die Anzahl der für uns *günstigen* Möglichkeiten (alle Zahlenkombinationen, wo es genau 5 Richtige gibt) bestimmen. Bei 5 Richtigen müssen von den 6 gezogenen Zahlen 5 richtig sein, es gibt also $\binom{6}{5}$ Möglichkeiten dafür. Da auf einem Tippschein 6 Zahlen angekreuzt werden müssen, muss eine dieser Zahlen falsch sein, also aus der Menge der nicht gezogenen Zahlen kommen. Dafür gibt es $\binom{43}{1}$ Möglichkeiten. Die Wahrscheinlichkeit für das Ereignis *5 Richtige* liegt also bei

$$\frac{\binom{6}{5}\cdot\binom{43}{1}}{\binom{49}{6}} = \frac{258}{13983816} = \frac{43}{2330636}$$

(b) Selbiges Vorgehen wie oben, nur hier können wir das Ereignis in zwei Ereignisse *kein Richtiger* und *1 Richtiger* aufteilen. Die gesuchte Wahrscheinlichkeit ist also

$$\frac{\binom{6}{0}\cdot\binom{43}{6}+\binom{6}{1}\cdot\binom{43}{5}}{\binom{49}{6}} = \frac{11872042}{13983816} = \frac{848003}{998844}$$

Aufgabe 2

(a) Die Ungleichung von Tschebyscheff gibt die untere Schranke für die Wahrscheinlichkeit an, dass $X \in [\mu-\varepsilon, \mu+\varepsilon]$. Hier ist $\mu = 40$ und damit $\varepsilon = 10$. Es folgt dann

$$\begin{aligned}\mathbb{P}(|X-\mu| \le \varepsilon) &\ge 1\frac{\mathrm{Var}(X)}{\varepsilon^2} \\ &\ge 1-\frac{5^2}{10^2} \\ &\ge 0.75\end{aligned}$$

(b) Wir führen eine neue Zufallsvariable $\bar{X}_n$ ein, die dem Mittelwert von n Messungen entspricht. Für diese Zufallsvariable gilt nun bezüglich Erwartungswert und Varianz:

$$\begin{aligned}\mathbb{E}(\bar{X}_n) &= \mathbb{E}(X_i) = 40 \\ \mathrm{Var}(\bar{X}_n) &= \frac{1}{n}\cdot\mathrm{Var}(X_i) = \frac{25}{n}\end{aligned}$$

Mit der Ungleichung von Tschebyscheff (und $\varepsilon = 10$) folgt damit

$$\begin{aligned}\mathbb{P}(|\bar{X}_n - \mu| \leq \varepsilon) &\geq 1 - \frac{\operatorname{Var}(\bar{X}_n)}{\varepsilon^2} \\ &\geq 1 - \frac{\frac{25}{n}}{10^2} \\ &\geq 1 - \frac{\frac{25}{9}}{10^2} \\ &\geq \frac{35}{36}\end{aligned}$$

(c) Standardisieren wir die Zufallsvariable zuerst: $\frac{X-40}{5} \sim \Phi$. Die Grenzen von 50 bzw. 30 müssen wir derselben Transformation unterziehen, sodass sich die die benötigte Wahrscheinlichkeit ergibt:

$$\begin{aligned}\mathbb{P}(30 \leq X \leq 50) &= \Phi\left(\frac{50-40}{5}\right) - \Phi\left(\frac{30-40}{5}\right) \\ &= \Phi(2) - \Phi(-2) \\ &= 0.9545\end{aligned}$$

Interessant ist, dass die konkreten Zahlen in dieser Aufgabe gar nicht so wichtig sind, denn es gilt allgemein:

- Im Intervall der Abweichung $\pm\sigma$ vom Erwartungswert sind 68,27 % aller Messwerte zu finden
- Im Intervall der Abweichung $\pm 2\sigma$ vom Erwartungswert sind 95,45 % aller Messwerte zu finden
- Im Intervall der Abweichung $\pm 3\sigma$ vom Erwartungswert sind 99,73 % aller Messwerte zu finden

siehe dazu auch `https://de.wikipedia.org/wiki/Normalverteilung`

(d) Bringen wir die Ideen aus (b) und (c) zusammen und standardisieren wieder zuerst: $\frac{\bar{X}_n - \mathbb{E}(\bar{X}_n)}{\sqrt{\operatorname{Var}(\bar{X}_n)}} = \frac{\bar{X}_n - 40}{\frac{5}{n}}$ und dann ergibt sich:

$$\begin{aligned}\mathbb{P}(30 \leq \bar{X}_n \leq 50) &= \Phi\left(\frac{50-40}{\frac{5}{n}}\right) - \Phi\left(\frac{30-40}{\frac{5}{n}}\right) \\ &= \Phi(2n) - \Phi(-2n) \\ &= \Phi(18) - \Phi(-18) \\ &\approx 1\end{aligned}$$

(e) Die Breite eines Konfidenzintervalls ist

$$2 \cdot z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Damit gilt

$$\begin{aligned}
2 \cdot z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} &= 10 \\
\sqrt{n} &= z_{1-\frac{\alpha}{2}} \frac{\sigma}{5} \\
n &= \left(z_{1-\frac{\alpha}{2}} \frac{\sigma}{5}\right)^2 \\
&= \left(1.95996 \frac{5}{5}\right)^2 \\
&= 3.8414 \\
&\approx 4
\end{aligned}$$

(f) Das Konfidenzintervall für unbekannte Varianz ist gegeben durch

$$\begin{aligned}
KI &= \left[\bar{X} - t_{n-1;1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}; \bar{X} + t_{n-1;1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right] \\
&= \left[40 - 2.306 \frac{\sqrt{60.5}}{\sqrt{9}}; 40 + 2.306 \frac{\sqrt{60.5}}{\sqrt{9}}\right] \\
&= [34.0212; 45.9788]
\end{aligned}$$

Aufgabe 3

(a) Ableiten der Verteilungsfunktion liefert die Dichtefunktion:

$$f(x) = \begin{cases} 2a \cdot \exp(-2ax) & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

(b) Aufstellen der Likelihood-Funktion:

$$\begin{aligned}
L &= \prod_{i=1}^{n} f(x_i) \\
&= \prod_{i=1}^{n} 2a \cdot \exp(-2ax) \\
&= (2a)^n \cdot \prod_{i=1}^{n} \exp(-2ax)
\end{aligned}$$

Berechnen der log-Likelihood-Funktion, um das Produkt in eine Summe umzuwandeln:

$$\begin{aligned}
l &= n \cdot \log(2a) + \sum_{i=1}^{n} \log(\exp(-2ax)) \\
&= n \cdot \log(2a) - \sum_{i=1}^{n} 2ax \\
&= n \cdot \log(2a) - 2a \sum_{i=1}^{n} x_i
\end{aligned}$$

Ableiten nach a und Nullsetzen liefert den Likelihood-Schätzer für a:

$$\begin{aligned}
\frac{\partial l}{\partial a} &= 2n\frac{1}{2a} - 2\sum_{i=1}^{n} x_i = 0 \\
\frac{n}{a} &= 2\sum_{i=1}^{n} x_i \\
a &= \frac{n}{2\sum_{i=1}^{n} x_i} \\
&= \frac{1}{2\bar{x}}
\end{aligned}$$

(c) Der Mittelwert $\bar{x}$ ist 5, also ist $a = \frac{1}{2\bar{x}} = \frac{1}{2\cdot 5} = \frac{1}{10}$

Aufgabe 4

(a) Wir führen einen rechtsseitigen Test durch:
$H_0 : \mu \leq 7.6$
$H_1 : \mu > 7.6$
Die Teststatistik ergibt sich zu

$$\begin{aligned}
T &= \frac{\bar{X} - 7.6}{\sigma}\sqrt{n} \\
&= \frac{7.8 - 7.6}{0.5}\sqrt{12} \\
&= 1.3856
\end{aligned}$$

Der kritische Wert ist $z_{1-\alpha} = z_{0.95} = 1.6449$ und damit kann die Nullhypothese nicht abgelehnt werden.

(b) Das Konfidenzintervall für unbekannte Varianz ist gegeben durch

$$\begin{aligned}
KI &= \left[\bar{X} - t_{n-1;1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}; \bar{X} + t_{n-1;1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right] \\
&= \left[7.8 - 2.1788\frac{0.6}{\sqrt{12}}; 7.8 + 2.1788\frac{0.6}{\sqrt{12}}\right] \\
&= [7.4226; 8.1774]
\end{aligned}$$

(c) Beschäftigen wir uns zuerst mit der Operationscharakteristik. Das ist die Wahrscheinlichkeit, dass wir H_0 nicht ablehnen. Wir lehnen H_0 genau dann nicht ab, wenn $T < z_{1-\frac{\alpha}{2}}$ ist, also

$$\begin{aligned}
L(\mu) &= \mathbb{P}(T \leq z_{1-\frac{\alpha}{2}}) \\
&= \mathbb{P}\left(\frac{\bar{X} - \mu_0}{\sigma}\sqrt{n} \leq z_{1-\frac{\alpha}{2}}\right) \\
&= \mathbb{P}\left(\frac{\bar{X} - \mu_0}{\sigma}\sqrt{n} \leq z_{1-\frac{\alpha}{2}} + \frac{\mu_0 - \mu}{\sigma}\sqrt{n}\right) \\
&= \Phi\left(z_{1-\frac{\alpha}{2}} + \frac{\mu_0 - \mu}{\sigma}\sqrt{n}\right) \\
&= \Phi\left(1.6449 + \frac{7.6 - \mu}{0.5}\sqrt{12}\right)
\end{aligned}$$

Die Gütefunktion ist dann $1 - L(\mu)$.

(d) Dafür müssen wir die Gütefunktion an der Stelle $\mu = 7.5$ berechnen:

$$\begin{aligned} G(7.5) &= 1 - L(7.5) \\ &= 1 - \Phi\left(1.6449 + \frac{7.6 - 7.5}{0.5}\sqrt{12}\right) \\ &= 1 - \Phi(2.3328) \\ &= 0.0098 \end{aligned}$$

Aufgabe 5

(a) Wir führen hier einen χ^2-Anpassungstest durch:

Ereignis	0	1	2	Σ
S_i	320	48	32	400
p_i	0.64	0.32	0.04	1
np_i	256	128	16	400

Die Teststatistik ergibt sich zu $T = 82$, der kritische Wert ist $\chi^2_{3-1;1-\alpha} = \chi^2_{2,0.9} = 4.6052$, also wird die Nullhypothese abgelehnt.

(b) Wir führen hier einen χ^2-Anpassungstest durch:

Ereignis	0	1	2	Σ
S_i	320	48	32	400
p_i	0.7408	0.2222	0.0333	≈ 1
np_i	296.32	88.90	13.33	≈ 400

Die Teststatistik ergibt sich zu $T = 46.8333$, der kritische Wert ist $\chi^2_{3-1;1-\alpha} = \chi^2_{2,0.9} = 4.6052$, also wird die Nullhypothese abgelehnt.

Aufgabe 6

Die Variable *Arbeitsgeschwindigkeit* hat metrisches Skalenniveau, die Variable *Qualität* hat nur ordinales Skalenniveau. Wir können hier also einen Unabhängigkeitstest mittels Rangkorrelationskoeffizient machen. Die Ränge sind

Arbeiterin	A	B	C	D	E	F	F	H
R(Geschwindigkeit)	1	2	3	4	5	6	7	8
R(Qualität)	4	5	3	2	8	6	1	7

Die Teststatistik ergibt sich zu $T = 0.1819$ und der kritische Wert für den zweiseitigen Test $H_0 : \rho = 0$ vs. $H_1 : \rho \neq 0$ ist $z_{1-\frac{\alpha}{2}} = 1.9599$. Die Nullhypothese kann also nicht abgelehnt werden (hier ist $\rho = 0.1905$).

Aufgabe 7

(a) Wir führen hier einen Zweistichproben-t-Test durch. Der Mittelwert der ersten Waage ist $\bar{X}_1 = 1000$, der Mittelwert der zweiten Waage ist $\bar{X}_2 = 1000.01$. Die Varianz der ersten Waage ist $S_1^2 = 5 \cdot 10^{-5}$, die der zweiten Waage $S_2 = 0.00035$. Da die Varianzen bei beiden Waagen gleich sein sollen, muss diese erst geschätzt werden:

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \\
&= \frac{(5-1) \cdot 5 \cdot 10^{-5} + (5-1) \cdot 0.00035}{5 + 5 - 2} \\
&= 2 \cdot 10^{-4}
\end{aligned}$$

Die Teststatistik ergibt sich dann zu

$$\begin{aligned}
T &= \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \\
&= \frac{1000 - 1000.01}{\sqrt{2 \cdot 10^{-4} \left(\frac{1}{5} + \frac{1}{5}\right)}} \\
&= -1.1180
\end{aligned}$$

Der kritische Wert ist $t_{n_1+n_2-2;1-\frac{\alpha}{2}} = t_{8;0.95} = 1.8595$, die Nullhypothese kann damit nicht abgelehnt werden.

(b) Die Nullhypothese wird dann abgelehnt, wenn $|T| > t_{krit}$, also dann, wenn t_{krit} 1.180 unterschreitet.