Lineare und Logistische Regression
Dr. Lucia Kleint
HS 2020

Fachhochschule Nordwestschweiz

Übungsblatt LE3

Lösungen

Lösung 1.

Hypothesen

$$H_0 : p = 0.174$$

$$H_1 : p \neq 0.174$$

Da auf Ungleichheit getestet wird, ist dies ein zweiseitiger Hypothesentest. Stichprobe $n = 2000, p_1 = 0.191$. $\alpha = 5\% = 0.05$.

Diese Aufgabe ist analog zu einem Zufallsexperiment mit n Wiederholungen. Dabei ist die Testgrösse

$$T = \frac{p_1 - p_0}{\sqrt{p_0 q_0 / n}}, \text{ wobei } q_0 = 1 - p_0$$

also in diesem Fall

$$T = \frac{0.191 - 0.174}{\sqrt{0.174 \cdot 0.826/2000}} = 2.01$$

Da der Test beidseitig ist, verwendet man $z_{0.025}$, und gemäss Tabelle (oder Skript) ist dies 1.96. Da $T > z_{0.025}$, muss H_0 verworfen werden, d.h. die Wirkung scheint sich verändert zu haben.

Lösung 2.

Hypothesen

$$H_0 : p \leq 0.5$$

$$H_1 : p > 0.5$$

Dies ist ein einseitiger Test mit $\alpha = 5\%$. Stichprobe: $n = 10000$. Schätzung für die Wahrscheinlichkeit $p_1 = \frac{5090}{10000} = 0.509$.

Berechnung Testgrösse

$$T = \frac{0.509 - 0.5}{\sqrt{0.5 \cdot 0.5/10000}} = 1.8$$

$z_{0.05} = 1.64$ (Skript). Da $1.8 > 1.64$, also $T > z_{0.05}$, muss H_0 verworfen werden.

Lösung 3.

Hypothesen

$$H_0 : \mu = 60 \text{ d.h. Maschine stimmt}$$
$$H_1 : \mu \neq 60$$

Da auf Ungleichheit getestet wird, ist dies ein zweiseitiger Hypothesentest. Stichprobe $n = 40$, was genügend gross ist, um eine Normalverteilung anzunehmen (und somit keinen t Test). $\alpha = 5\% = 0.05$.

Berechnung Testgrösse

$$T = \frac{\bar{x} - \mu_0}{\sigma}\sqrt{n} = \frac{59.92 - 60}{0.2}\sqrt{40} = -2.53$$

Da der Test beidseitig ist, verwendet man $z_{0.025}$, und gemäss Tabelle (oder Skript) ist dies 1.96. Da $|T| > z_{0.025}$, muss H_0 verworfen werden, d.h. man muss die Maschine neu einstellen.

Lösung 4.

Hypothesen

$$H_0 : \mu \geq 1.45 \text{ d.h. keine Verbesserung}$$
$$H_1 : \mu < 1.45 \text{ d.h. es gab eine Verbesserung}$$

Dies ist ein einseitiger Test mit $\alpha = 5\%$. Stichprobe: $n = 200$.

Berechnung Testgrösse

$$T = \frac{\bar{x} - \mu_0}{\sigma}\sqrt{n} = \frac{1.41 - 1.45}{0.6}\sqrt{200} = -0.94$$

$z_{0.05} = 1.64$ (Skript). Da $0.94 < 1.64$, also $|T| < z_{0.05}$, kann H_0 nicht verworfen werden. In diesem Fall kann man H_0 nicht verwerfen, d.h. man kann nicht sagen, ob es zu einer Verbesserung gekommen ist. Beide Möglichkeiten sind noch offen, da man H_0 nie beweisen kann.

Lösung 5.

Hypothesen

$$H_0 : \mu = 60 \text{ d.h. Ausstoss stimmt}$$
$$H_1 : \mu \neq 60$$

(a) Da sowohl zu tiefe, wie auch zu hohe Ausstosse nicht gewünscht sind, betrachten wir dies als zweiseitigen Hypothesentest. Stichprobe $n = 20$, was eher tief ist, deswegen verwendet man einen t Test, da auch die Varianz der Grundgesamtheit unbekannt ist. $\alpha = 10\% = 0.1$.

Berechnung Testgrösse

$$t = \frac{\bar{x} - \mu_0}{\sigma}\sqrt{n} = \frac{60.2 - 60}{\sqrt{0.5}}\sqrt{20} = 1.26$$

Aus einer Tabelle (oder Python `stats.t.ppf(1-0.1/2., 19)`) findet man den t-Wert für das 0.95 Quantil und 19 Freiheitsgraden: $t_{n-1;1-\alpha/2} = t_{19;0.95} = 1.73$

Da $t < t_{19;0.95}$ kann man H_0 nicht ablehnen.

(b) Im Vergleich zu a) ändert sich die Testgrösse zu

$$t = \frac{\bar{x} - \mu_0}{\sigma}\sqrt{n} = \frac{60.2 - 60}{\sqrt{0.5}}\sqrt{60} = 2.19$$

Eine Stichprobe von 60 ist genügend hoch, um eine Normalverteilung zu verwenden, aber generell ergeben sowohl t-Verteilung wie auch die Normalverteilung bei dieser Stichprobengrösse ähnliche Werte. t-Verteilung: `stats.t.ppf(1-0.1/2., 59) = 1.67`. Normalverteilung: $z_{0.05} = 1.64$. Da $t > z_{0.05}$ kann man H_0 ablehnen. Hier sieht man deutlich, dass es sein kann, dass man bei einer zu kleinen Stichprobe keine Rückschlüsse ziehen kann.

Lösung 6.

Hypothesen

$$H_0 : \mu = 0 \text{ d.h. kein Unterschied}$$

$$H_1 : \mu \neq 0$$

Berechnung Testgrösse

$$t = \frac{\bar{x} - \mu_0}{\sigma}\sqrt{n} = \frac{0.2 - 0}{0.5}\sqrt{30} = 2.19$$

t-Verteilung: `stats.t.ppf(1-0.05/2., 29) = 2.045`

Da t > 2.045 kann man die Nullhypothese ablehnen. Somit gibt es statistisch signifikante Unterschiede zwischen den beiden Geräten!

Lösung 7.

(a) Dauer = $b_0 + b_1$ * Distanz = 0.12 + 0.0036 * Distanz

(b) Hypothesen

$$H_0 : b_1 = 0 \text{ (also kein Zusammenhang, nur konstante Funktion)}$$

$$H_1 : b_1 \neq 0$$

Das Ziel ist die Berechnung der Testgrösse

$$t = \frac{b_1 - 0}{SE(b_1)}$$

Das "0" stammt aus H_0 und SE steht für standard error, also die Standardabweichung von b_1. Die Testgrösse misst die Anzahl Standardabweichungen, die b_1 von 0 entfernt ist.

Die Standardabweichung von b_1 ist definiert als

$$SE(b_1) = \sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

hier ist σ die Standardabweichung, x_i sind die einzelnen x-Werte und $\bar{x}$ ist ihr Mittelwert.

Nun berechnet man die Standardabweichung bzw. die Varianz mittels Residuen (s. Fahrmeir, S. 467).

$$\sigma^2 = \frac{1}{10-2}\sum_{i=1}^{n} r_i^2 = \frac{1}{8}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 0.23$$

$\hat{y}_i$ bezeichnet die gefitteten Werte. $\Rightarrow \sigma = 0.48$

Die einzelnen Werte berechnet man am besten in einem Programm und erhält für die Testgrösse

$$t = \frac{b_1 - 0}{SE(b_1)} = \frac{0.0036}{\sqrt{\frac{0.48^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}} = \frac{0.0036}{(0.48/1139.24)} = 8.54$$

Die t-Verteilung (Python oder Tabelle) ist $t_{8;0.975} = 2.306$. Da $t > t_{8;0.975}$ wird H_0 abgelehnt und somit besteht ein signifikanter Zusammenhang zwischen der Distanz und der Fahrdauer.