ETH Zürich HS 2019
RW, D-MATL, D-MAVT
Dr. Christoph Czichowsky

Koordinator
Florian Krach

# Stochastik

## Interaktive Aufgabe für die Übung 11

**1. Limitation von $p$-Werten und statistischen Tests.**

In dieser Aufgabe wollen wir zeigen, dass ein signifikanter $p$-Wert bzw. ein signifikantes Testergebnis unter Umständen nur wenig darüber aussagen kann, wie wahrscheinlich es ist, dass die Alternativhypothese in Wirklichkeit stimmt. Sei dafür $X \sim N(\mu, 1)$ mit unbekanntem $\mu$. Unsere Nullhypothese ist $H_0 : \mu = 0$ und unsere Alternative $H_A : \mu = m > 0$.

**a)** Geben sie einen Test und einen Verwerfungsbereich für diesen Test mit Niveau $\alpha \in (0, 1)$ an.

Wir führen nun eine weitere Zufallsvariable ein, die kontrolliert, wie wahrscheinlich es ist, dass in Wirklichkeit die Nullhypothese stimmt. Seien $M \sim \text{Bernoulli}(p)$, mit $p \in (0, 1)$ und $Z \sim N(0, 1)$ unabhängig. Wir treffen nun die Annahme, dass die Nullhypothese mit einer Wahrscheinlichkeit von $1 - p$ wahr ist (das kann als Vorwissen (*a priori* Wahrscheinlichkeit) interpretiert werden [1]). Diese Annahme können wir umsetzen, indem wir $X := m \cdot M + Z$ setzen.

**b)** Überprüfe, dass die Null- und Alternativhypothese mit den Wahrscheinlichkeiten $1 - p$ und $p$ eintreffen.

**c)** Wir führen den Test durch und erhalten ein signifikantes Ergebnis auf Niveau $\alpha$ (also würden wir $H_0$ verwerfen). Gegeben diesem Testergebnis, wie Wahrscheinlich ist es, dass die Alternativhypothese stimmt (*a posterior* Wahrscheinlichkeit)?

**d)** Berechne diese Wahrscheinlichkeit für $\alpha = 0.01$, $m = 1$ und $p = 0.05$.

**Lösung:**

**a)** Wir verwenden einen einseitigen $Z$-Test. Der Verwerfungsbereich ist somit $(z_{1-\alpha}, \infty)$, wobei $z_{1-\alpha} = \Phi^{-1}(1 - \alpha)$.

**b)** $X$ ist verteilt wie in der Nullhypothese genau dann wenn $M = 0$ gilt. Das ist mit Wahrscheinlichkeit $1 - p$ der Fall. Entsprechend gilt, dass $X$ verteilt ist wie in der Alternativhypothese, genau dann wenn $M = 1$.

[1] Bemerkung: Natürlich kann in Wirklichkeit etwas nur entweder wahr oder falsch sein. Aber es kann sein, dass wir bereits Vorwissen über die Wahrscheinlichkeiten der Hypothesen haben. Ein Beispiel: Wir warten auf die Tram, die Verspätung hat und wir fragen uns, ob ein aussergewöhnliches Ereignis (ein Unfall, eine Geburt in der Tram etc.) ($H_A$) oder ein erhöhtes Verkehrs- und Personenaufkommen ($H_0$) die Ursache ist. Dann wissen wir zwar nicht ob in diesem speziellen Fall $H_0$ oder $H_A$ stimmt, aber wir können z.B. die historischen Daten der VBZ, verwenden um die Wahrscheinlichkeit für ein aussergewöhnliches Ereignis zu schätzen (z.B. $p = 0.0001\%$). Somit haben wir also das Vorwissen, das $H_A$ mit Wahrscheinlichkeit $p$ eintritt. Dieser Ansatz fällt in das Gebiet der *Bayes'schen-Statistik*.

**c)** Wir verwenden den Satz von Bayes und bekommen

$$\begin{aligned} P(H_A \text{ stimmt} \mid \text{Testergebnis}) &= P(M=1|X > z_{1-\alpha}) \\ &= \frac{P(X > z_{1-\alpha}|M=1)P(M=1)}{P(X > z_{1-\alpha}|M=1)P(M=1) + P(X > z_{1-\alpha}|M=0)P(M=0)} \\ &= \frac{(1-\Phi(z_{1-\alpha}-m)) \cdot p}{(1-\Phi(z_{1-\alpha}-m)) \cdot p + \alpha \cdot (1-p)} \end{aligned}$$

**d)** Siehe R-Code. Für die gegebenen Werte erhalten wir die Wahrscheinlichkeit $P(M = 1|X > z_{1-\alpha}) = 0.3271$.

**2.** Sei $X$ eine Zufallsvariable. Wir haben 10 Beobachtungen $(x_1, \ldots, x_{10})$. Wir nehmen an, dass $X$ $\mathcal{N}(\mu, \sigma^2)$-verteilt ist. Es wurde gesagt, dass $\mu = 7$ gilt, aber wir vermuten, dass $\mu$ grösser ist. Das arithmetische Mittel ist $\overline{x} = 8.13$ und die empirische Standardabweichung ist $s = 2.97$.

**a)** Führen Sie einen geeigneten Test auf dem 5%-Niveau durch.

**b)** Führen Sie den gleichen Test auf dem 20%-Niveau durch.

**c)** Welches ist das kleinste Niveau, auf dem der Test die Nullhypothese noch verwirft (d.h. der $P$-Wert)?

**d)** Jetzt haben wir 90 Beobachtungen, wobei $\overline{x} = 8.13$ und $s = 2.98$. Führen Sie den gleichen Test auf dem 5%-Niveau durch.

**e)** Wie gross ist der $P$-Wert?

**Lösung:**

**a)** Da die Daten normalverteilt sind und $\sigma$ unbekannt ist, führen wir einen $t$-Test zum Niveau $\alpha = 5\%$ durch. Wir wollen testen, ob $\mu > 7$. Daraus ergibt sich $\mu_0 := 7$ und ein einseitiger Test mit den Hypothesen

$$H_0 : \mu = \mu_0 = 7 \qquad\qquad H_A : \mu > \mu_0.$$

Mit $t_{n-1,1-\alpha} = t_{9,0.95} = 1.833$ ist der Verwerfungsbereich gegeben durch $K = [1.833; \infty)$. Die Realisierung $t$ der Teststatistik $T$ errechnet sich zu

$$t = \sqrt{n}\frac{\overline{x}_n - \mu_0}{s} = \sqrt{10}\frac{8.13-7}{2.97} = 1.203 \notin K,$$

weshalb die Nullhypothese nicht verworfen wird.

**b)** Mit $t_{n-1,1-\alpha} = t_{9,0.80} = 0.883$ ist der Verwerfungsbereich $K = [0.883; \infty)$. Die Realisierung der Teststatistik $T$ ist immer $t = 1.203$. Die Nullhypothese wird diesmal verworfen.

**c)** Der $P$-Wert für die gegebene Realisierung der Teststatistik errechnet sich zu $p = P(T \geq 1.203) = 1 - P(T \leq 1.203) = 1 - 0.864 = 0.136$ (lineare Interpolation zwischen $t_{9,0.80} = 0.883$ und $t_{9,0.90} = 1.383$). Wenn wir einen Test zum Niveau $\alpha \geq 13.6\%$ durchführen, wird die Nullhypothese verworfen. Wenn $\alpha < 13.6\%$, wird die Nullhypothese nicht verworfen.

**d)** Der Verwerfungsbereich ist $K = [t_{n-1,1-\alpha}; \infty) = [t_{89,0.95}; \infty) \approx [t_{90,0.95}; \infty) = [1.662; \infty)$. Die Realisierung $t$ der Teststatistik $T$ errechnet sich zu

$$t = \sqrt{n}\frac{\overline{x}_n - \mu_0}{s} = \sqrt{90}\frac{8.13 - 7}{2.98} = 3.597 \in K,$$

weshalb die Nullhypothese (ziemlich klar) verworfen wird.

**e)** Der $P$-Wert für die gegebene Realisierung der Teststatistik errechnet sich zu $p = P(T \geq 3.597) = 1 - P(T \leq 3.597)$. In der Tabelle ist $t_{90,0.995} = 2.632$, also $p < 0.005 = 0.5\%$. Man bekommt $p = 0.03\%$ mit Excel. Wir könnten also einen Test zum Niveau $\alpha = 0.03\%$ durchführen und die Nullhypothese würde immer noch verworfen.