# Severity

Roland Schäfer

4. April 2024

## 1 Nomenklatur

Im Folgenden wird $\mu$ für den *wahren Mittelwert* verwendet. Der Mittelwert unter der Nullhypothese sei $\mu_0$. Gemessene Werte werden als $\mu_n$ mit $n$ als Index angegeben (bei Mayo auch $\bar{x}$). Die unter der Auswertung von Severity betrachteten Partikularhypothesen bezeichnen wir hier als $\mu'$. Bei Mayo heißen diese Hypothesen $\mu_1$.

## 2 Test

Ein einseitiger Test liefert die frequentistische Wahrscheinlichkeit, das konkrete Ergebnis oder ein extremeres zu finden, wenn die $H_0$ korrekt ist. Analog zum Artikel anhand eines z-Tests über Mittelwerte: Wir betrachten die $H_0$ in (1).

$$H_0 : \mu_0 = 0 \tag{1}$$

In Worten: *Der Mittelwert unter der Nullhypothese ist 0.* Für die Illustration nehmen wir einen einseitigen -- hier rechtsseitigen -- Test gemäß (2). In Worten: *Der Mittelwert unter der Nullhypothese ist größer als 0.*

$$H_1 : \mu > 0 \tag{2}$$

Die Varianz sei bekannt ($\sigma = 2$), und wir gehen von einer Stichprobengröße von $n = 100$ aus. Damit ist der Standardfehler gegeben gemäß (3).

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{2}{10} = 0.2 \tag{3}$$

Wir betrachten die drei möglichen Ausgänge des Experiments: $\mu_1 = 0.4$, $\mu_2 = 0.6$ und $\mu_3 = 1.0$. In einem Fisherschen Rahmen können diese Beobachtungen die $H_0$ zurückweisen, hier zu $sig = 0.025$ (bzw. $2\sigma$). Am Beispiel $\mu_1$ gezeigt in (4) mit $\mathcal{N}$ als kumulativer Verteilungsfunktion der Standardnormalverteilung.

$$P(\mu_1 \geq 0.4; \mu = \mu_0 = 0) = 1 - \mathcal{N}(\frac{\mu_1}{SE}) = 1 - \mathcal{N}(2) = 0.023 \tag{4}$$

Anders formuliert erreichen wir $2\sigma$, denn (5).

$$\frac{\mu_1}{SE} = \frac{0.4}{0.2} = 2 \tag{5}$$

# 3 Severity

## 3.1 Grundidee

Der Test aus Abschnitt 2 verläuft bei einer binären Entscheidung für oder gegen eine Zurückweisung der $H_0$ in allen drei betrachteten Fällen gleich. Die $H_0$ wird zurückgewiesen. Der p-Wert gibt zusätzlich darüber Auskunft, wie gut die Evidenz für die Zurückweisung war, denn (6).

$$P(\mu_1 \geq 0.4; \mu = 0) < P(\mu_2 \geq 0.6; \mu = 0) < P(\mu_3 \geq 1.0; \mu = 0) \tag{6}$$

Der Ausgang $\mu_3$ liefert also stärkere Evidenz gegen die $H_0$ als der Ausgang $\mu_2$ usw. Severity quantifiziert darüberhinaus, **wie gut die Evidenz für konkrete Abweichungen von der $\mathbf{H_0}$** ist. Sie beantwortet also Fragen wie: *Wie gut ist die Evidenz $\mu_1 = 0.4$ für eine Abweichung von $\gamma = 0.2$ von $H_0$?* Die Abweichung $\gamma$ ist hier eine *Effektstärke* im Sinn von Power-Berechnungen.
Dazu betrachten wir zusätzlich zur $H_1 : \mu > 0$ auf Basis eines konkreten signifikanten Ausgangs eines Experiments weitere Partikularhypothesen $H'$ über den wahren Wert $\mu$ wie in (7).

$$\begin{aligned} H' &: \mu > \mu' \\ \mu' &= \mu_0 + \gamma \end{aligned} \tag{7}$$

Der Unterschied zwischen $\mu$ und $\mu_0$ ist hier eventuell relevant. Der Test weist die $H_0$ über einen arbiträr gesetzten Wert $\mu_0$ zurück und sagt im Prinzip damit wenig über $\mu$. Severity quantifiziert die Evidenz für Schätzwerte $\mu'$ des wahren Werts $\mu$ als Abweichung von $\mu_0$ um die Differenz $\gamma$. Diese Betrachtung ist zulässig, sofern der Test bereits gezeigt hat, dass es gute Evidenz dafür gibt, dass die $H_0$ (in die erwartete Richtung) inkorrekt ist.

## 3.2 Wann ist Severity niedrig?

Wir setzen als Beispiel $\mu' = 0.2$. Die Severity für $\mu'$ soll **niedrig** sein, wenn bei $\mu' = \mu = 0.2$ der konkrete Messwert $\mu_1$ trotzdem sehr häufig (= frequentistisch wahrscheinlich) ist. Dies ist generell der Fall, wenn die Stichprobe klein oder die Varianz groß ist. Es ist unabhängig davon auch der Fall, wenn die Differenz zwischen $\mu'$ und $\mu_1$ größer bzw. positiver wird, wenn wir also eine stärkere Inferenz bezüglich der Punktschätzung des wahren Werts tätigen wollen. Im betrachteten Beispiel ($\mu' = 0.2$) ist $\mu' - \mu_1 = 0.2 - 0.4 = -0.2$. Würden wir hingegen eine Partikularhypothese $\mu' = 0.6$ betrachten, wäre $\mu' - \mu_1 = 0.6 - 0.4 = 0.2$. Bei gleichbleibender Varianz und Stichprobengröße sollte dies auch (wenn $\mu' = \mu = 0.2$) intuitiv unwahrscheinlicher sein, denn eine Beobachtung von $0.4$ liefert schlechtere Evidenz für eine Abweichung um $0.6$ von $0$ als für eine Abweichung von $0.2$ von $0$. Es wird deutlich, dass die ursprüngliche $H_0$ und die Richtung der Ausgangshypothese mit Severity zusammenhängen. Bei einem linksseitigen Test sollte Severity hingegen kleiner werden, je kleiner (bzw. je negativer) die Differenz zwischen $\mu'$ und $\mu_1$ wird.

## 3.3 Wann ist Severity hoch?

Die Severity für $\mu' = 0.2$ soll nun **hoch** sein, wenn der Messwert $\mu_1 = 0.4$ selten zu erwarten ist, falls $\mu' = \mu$.
(Wird fortgesetzt.)

## 3.4 Veranschaulichung und Berechnung

Abbildung 1 zeigt die Situation für $\mu' = 0.2$. Die schwarze Kurve zeigt die Dichte der Standardnormalverteilung für den ursprünglichen Test, der mit $p = 0.023$ die $H_0$ zurückweisen konnte. Die blaue Schleppe für den Beobachtungswert $\mu_1 = 0.4$ entspricht $2.3\%$ der frequentistisch erwartbaren Werte. Unter der Annahme, dass $\mu = \mu' = 0.2$, zeigt bei den gleichen Parametern $\sigma$ und $n$ die rote Kurve die erwartete Verteilung der Messwerte um $\mu' = 0.2$. Die grüne Schlepppe (ebenfalls für den Beobachtungswert $\mu_1 = 0.4$) entspricht dem Anteil der erwarteten Messwerte, die dann größer oder gleich $0.4$ sind. Da $SE = 0.2$ und $\mu_1 - \mu' = 0.2$, entspricht dies $1 - \mathcal{N}(1) = 0.16$. In diesem Fall wären also $\mathcal{N}(1) = 0.84$ ($84\%$) der Werte kleiner als $0.4$, also (8). Die Klausel *is true* nach dem Semikolon wurde als redundant ausgelassen.

$$SEV(\mu > 0.2) = P(\bar{X} < 0.4; \mu \leq 0.2) \tag{8}$$

Berechnet wird hier die *minimale* Severity. Abbildung 2 zeigt dasselbe für $\mu' = 0.1$. Trivialerweise $\mu_1 - \mu' = 0.4 - 0.1 = 0.3$. Bei $SE = 0.2$ entspricht die Fläche unter der Kurve minus der grünen Schleppe einem Anteil von $\mathcal{N}(1.5) = 0.93$. Kleinere $\delta$ entsprechen größeren Wahrscheinlichkeiten, also einer größeren Severity.

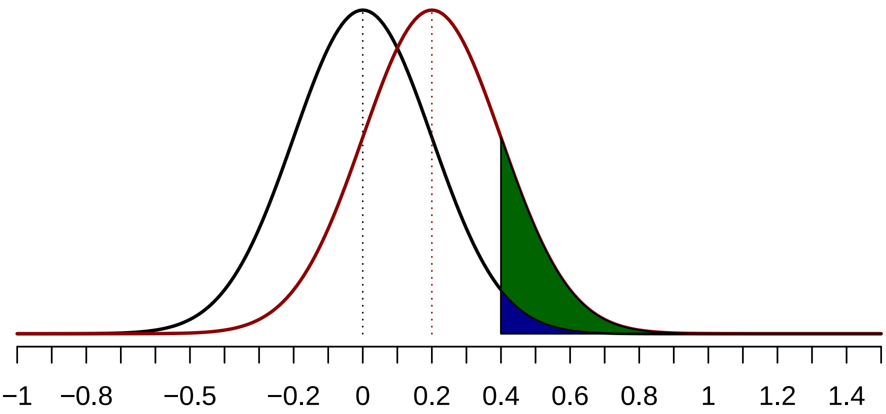


Abbildung 1: Severity for H': mu'>0.2 bei der Beobachtung mu1=0.4

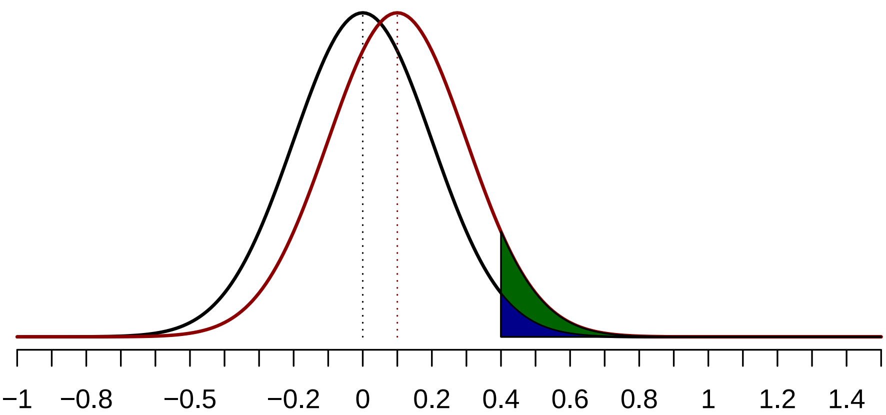


Abbildung 2: Severity for H': mu'>0.1 bei der Beobachtung mu1=0.4

Paralleles gilt für größere beobachtete Abweichungen von $0$ wie in Abbildung 3 mit $\mu_1 = 0.6$ und $\mu' = 0.2$. Hier gilt $\mathcal{N}(2) = 0.98$ wegen $\mu_1 - \mu' = 0.6 - 0.2 = 0.4$ bei $SE = 0.2$. Steigt der Beobachtungswert $\mu_1$ oder sinkt der Schätzwert $\mu'$, dessen Severity zu bewerten ist, wird die Severity größer. Daher stellt die Berechnung mit (9) allgemein eine Untergrenze für SEV dar.

$$SEV(\mu > \mu') = P(\bar{X} < \mu_1; \mu < \mu') = \mathcal{N}(\frac{\mu_1 - \mu'}{SE}) \tag{9}$$

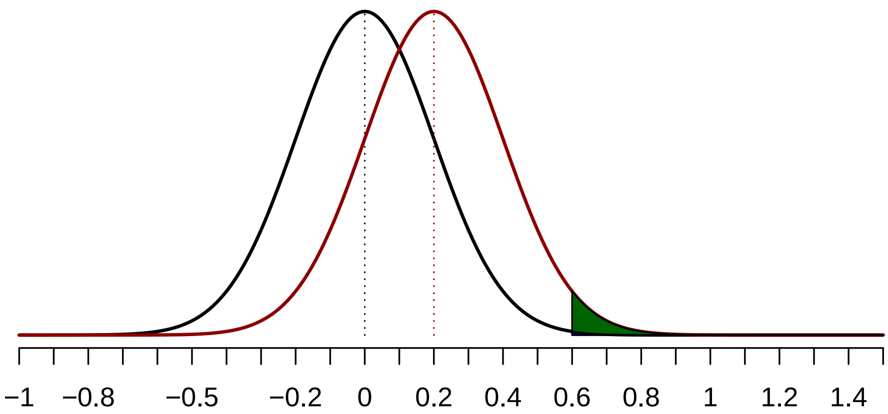


Abbildung 3: Severity for H': mu'>0.2 bei der Beobachtung mu1=0.6

Für drei Beobachtungen ($\mu_1 = 0.4$, $\mu_2 = 0.6$, $\mu_3 = 1.0$) zeigt Abbildung 4 die Severity-Kurven für $\delta \in [0, 1]$.

# 4 Zweiseitige Tests

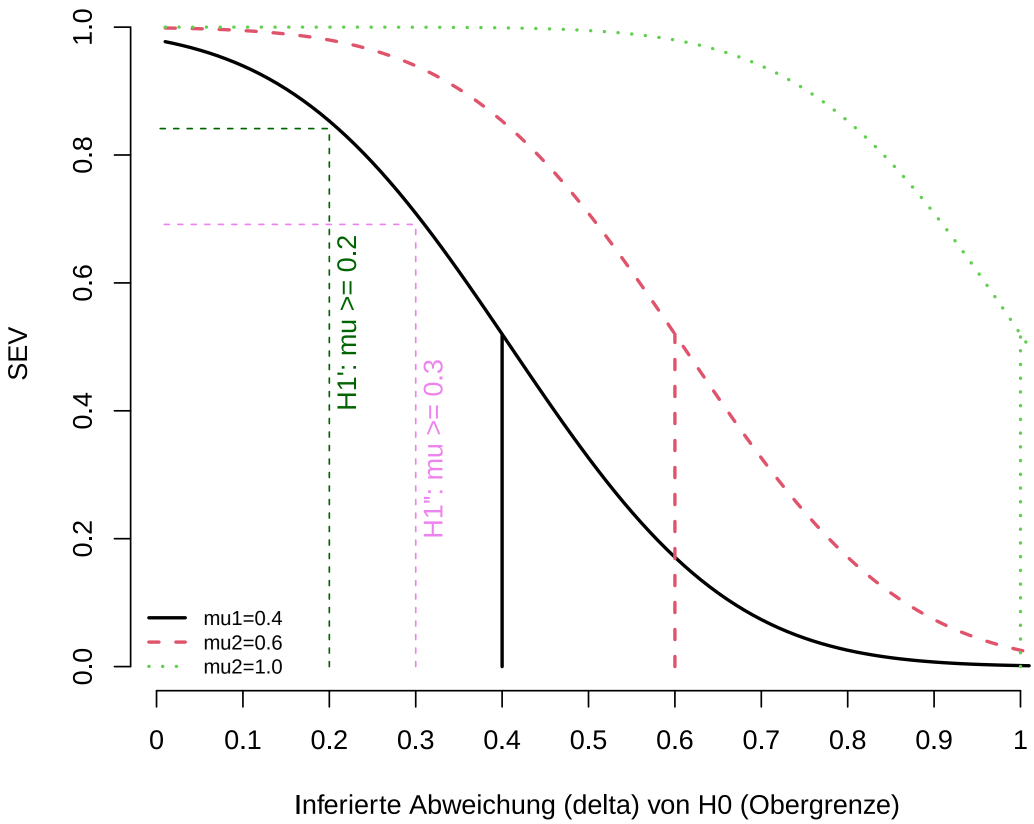


Abbildung 4: Severity-Kurven für verschiedene Beobachtungen