Souvenir W: The Severity Interpretation of Negative Results (SIN) for Test T+

Applying our general abbreviation: SEV(test T+, outcome \(x\), inference \(H\)), we get “the severity with which \(\mu \leq \mu_1\) passes test T+, with data \(x_0\):"

\[
\text{SEV}(T+, d(x_0), \mu \leq \mu_1),
\]
where \(\mu_1 = (\mu_0 + \gamma)\), for some \(\gamma \geq 0\). If it’s clear which test we’re discussing, we use our abbreviation: \(\text{SEV}(\mu \leq \mu_1)\). We obtain a companion to the severity interpretation of rejection (SIR), Section 4.4, Souvenir R:

**SIN (Severity Interpretation for Negative Results)**

(a) If there is a very *low* probability that \(d(x_0)\) would have been larger than it is, even if \(\mu > \mu_1\), then \(\mu \leq \mu_1\) passes with *low* severity: \(\text{SEV}(\mu \leq \mu_1)\) is low.

(b) If there is a very *high* probability that \(d(x_0)\) would have been larger than it is, were \(\mu > \mu_1\), then \(\mu \leq \mu_1\) passes the test with *high* severity: \(\text{SEV}(\mu \leq \mu_1)\) is high.

To break it down, in the case of a statistically insignificant result:

\[
\text{SEV}(\mu \leq \mu_1) = \Pr(d(X) > d(x_0); \mu \leq \mu_1 \text{ false}).
\]
We look at \{d(X) > d(x_0)\} because severity directs us to consider a “worse fit” with the claim of interest. That \(\mu \leq \mu_1\) is false within our model means that \(\mu > \mu_1\). Thus:

\[
\text{SEV}(\mu \leq \mu_1) = \Pr(d(X) > d(x_0); \mu > \mu_1).
\]

Now \(\mu > \mu_1\) is a composite hypothesis, containing all the values in excess of \(\mu_1\). How can we compute it? As with power calculations, we evaluate severity at a point \(\mu_1 = (\mu_0 + \gamma)\), for some \(\gamma \geq 0\), because for values \(\mu \geq \mu_1\) the severity increases. So we need only to compute

\[
\text{SEV}(\mu \leq \mu_1) > \Pr(d(X) > d(x_0); \mu = \mu_1).
\]

To compute SEV we compute \(\Pr(d(X) > d(x_0); \mu = \mu_1)\) for any \(\mu_1\) of interest. Swapping out the claims of interest (in significant and insignificant results), gives us a single criterion of a good test, severity.