Souvenir X: Power and Severity Analysis

Let’s record some highlights from Tour I:

First, ordinary power analysis versus severity analysis for Test $T^+$:

*Ordinary Power Analysis:* If $\Pr(d(X) \geq c_\alpha; \mu_1) = \text{high}$ and the result is not significant, then it’s an indication or evidence that $\mu \leq \mu_1$.

*Severity Analysis:* If $\Pr(d(X) \geq d(x_0); \mu_1) = \text{high}$ and the result is not significant, then it’s an indication or evidence that $\mu \leq \mu_1$.

It can happen that claim $\mu \leq \mu_1$ is warranted by severity analysis but not by power analysis.

If you add $k\sigma_X$ to $d(x_0)$, $k > 0$, the result being $\mu_1$, then $\text{SEV}(\mu \leq \mu_1) = \text{area to the right of } -k$ under the standard Normal ($\text{SEV} > 0.5$).

If you subtract $k\sigma_X$ from $d(x_0)$, the result being $\mu_1$, then $\text{SEV}(\mu \leq \mu_1) = \text{area to the right of } k$ under the standard Normal ($\text{SEV} \leq 0.5$).

For the general case of Test $T^+$, you’d be adding or subtracting $k\sigma_\bar{x}$ to $(\mu_0 + d(x_0)\sigma_\bar{x})$. We know that adding $0.85\sigma_\bar{x}$, $1\sigma_\bar{x}$, and $1.28\sigma_\bar{x}$ to the cut-off for rejection in a test $T^+$ results in $\mu$ values against which the test has 0.8, 0.84, and 0.9 power. If you treat the observed $\bar{x}$ as if it were being contemplated as the cut-off, and add $0.85\sigma_\bar{x}$, $1\sigma_\bar{x}$, and $1.28\sigma_\bar{x}$, you will arrive at $\mu_1$ values such that $\text{SEV}(\mu \leq \mu_1) = 0.8, 0.84, \text{and } 0.9$, respectively. That’s because severity goes in the same direction as power for non-rejection in $T^+$.

For familiar numbers of $\sigma_\bar{x}$’s added/subtracted to $\bar{x} = \mu_0 + d_0\sigma_\bar{x}$:

<table>
<thead>
<tr>
<th>Claim</th>
<th>$\mu \leq \bar{x} - 1\sigma_\bar{x}$</th>
<th>$\mu \leq \bar{x}$</th>
<th>$\mu \leq \bar{x} + 1\sigma_\bar{x}$</th>
<th>$\mu \leq \bar{x} + 1.65\sigma_\bar{x}$</th>
<th>$\mu \leq \bar{x} + 1.98\sigma_\bar{x}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEV</td>
<td>0.16</td>
<td>0.5</td>
<td>0.84</td>
<td>0.95</td>
<td>0.975</td>
</tr>
</tbody>
</table>
Now an overview of severity for test T+: Normal testing: $H_0: \mu \leq \mu_0$ vs. $H_1: \mu > \mu_0$ with $\sigma$ known. The severity reinterpretation is set out using discrepancy parameter $\gamma$. We often use $\mu_1$ where $\mu_1 = \mu_0 + \gamma$.

Reject $H_0$ (with $x_0$) licenses inferences of the form $\mu > [\mu_0 + \gamma]$, for some $\gamma \geq 0$, but with a warning as to $\mu \leq [\mu_0 + \kappa]$, for some $\kappa \geq 0$.

Non-reject $H_0$ (with $x_0$) licenses inferences of the form $\mu \leq [\mu_0 + \gamma]$, for some $\gamma \geq 0$, but with a warning as to values fairly well indicated $\mu > [\mu_0 + \kappa]$, for some $\kappa \geq 0$.

The severe tester reports the attained significance levels and at least two other benchmarks: claims warranted with severity, and ones that are poorly warranted.

Talking through SIN and SIR. Let $d_0 = d(x_0)$.

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

(a) *low*: If there is a very low probability that $d_0$ 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 (i.e., your test wasn’t very capable of detecting discrepancy $\mu_1$ even if it existed, so when it’s not detected, it’s poor evidence of its absence).

(b) *high*: If there is a very high probability that $d_0$ 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 (i.e., your test was highly capable of detecting discrepancy $\mu_1$ if it existed, so when it’s not detected, it’s a good indication of its absence).

**SIR (Severity Interpretation for Significant Results)**

If the significance level is small, it’s indicative of some discrepancy from $H_0$, we’re concerned about the magnitude:

(a) *low*: If there is a fairly high probability that $d_0$ would have been larger than it is, even if $\mu = \mu_1$, then $d_0$ is not a good indication $\mu > \mu_1$: $\text{SEV}(\mu > \mu_1)$ is low.\(^9\)

(b) *high*: Here are two ways, choose your preferred:

- (b-1) If there is a very high probability that $d_0$ would have been smaller than it is, if $\mu \leq \mu_1$, then when you observe so large a $d_0$, it indicates $\mu > \mu_1$: $\text{SEV}(\mu > \mu_1)$ is high.

---

\(^9\) A good rule of thumb to ascertain if a claim $C$ is warranted is to think of a statistical *modus tollens* argument, and find what would occur with high probability, were claim $C$ false.
• (b-2) If there’s a very low probability that so large a \( d_0 \) would have resulted, if \( \mu \) were no greater than \( \mu_1 \), then \( d_0 \) indicates \( \mu > \mu_1 \): SEV(\( \mu > \mu_1 \)) is high.\(^{10}\)

\(^{10}\) For a shorthand that covers both severity and FEV for Test T+ with small significance level (Section 3.1):

(FEV/SEV): If \( d(x_0) \) is not statistically significant, then \( \mu \leq x + k_\sigma/\sqrt{n} \) passes the test T+ with severity \( (1 - \epsilon) \)

(FEV/SEV): If \( d(x_0) \) is statistically significant, then \( \mu > x - k_\sigma/\sqrt{n} \) passes test T+ with severity \( (1 - \epsilon) \),

where \( \Pr(d(X) > k) = \epsilon \) (Mayo and Spanos (2006), Mayo and Cox (2006).)