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 $\Pr(d(\mathbf{X}) \geq c_\alpha; \mu_1)$ = high and the result is not significant, then it's an indication or evidence that $\mu \leq \mu_1$.

Severity Analysis: If $\Pr(d(\mathbf{X}) \geq d(\boldsymbol{x}_0); \mu_1)$ = 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.

8
- If you add $k\sigma_{\overline{X}}$ to $d(\boldsymbol{x}_0)$, $k > 0$, the result being μ_1, then $\text{SEV}(\mu \leq \mu_1)$ = area to the right of $-k$ under the standard Normal (SEV > 0.5).
- If you subtract $k\sigma_{\overline{X}}$ from $d(\boldsymbol{x}_0)$, the result being μ_1, then $\text{SEV}(\mu \leq \mu_1)$ = area to the right of k under the standard Normal (SEV ≤ 0.5).

For the general case of Test T+, you'd be adding or subtracting $k\sigma_{\overline{X}}$ to $(\mu_0 + d(\boldsymbol{x}_0)\sigma_{\overline{X}})$. We know that adding $0.85\sigma_{\overline{X}}$, $1\ \sigma_{\overline{X}}$, and $1.28\sigma_{\overline{X}}$ to the cut-off for rejection in a test T+ results in μ values against which the test has 0.8, 0.84, and 0.9 power. If you treat the observed $\overline{x}$ as if it were being contemplated as the cut-off, and add $0.85\sigma_{\overline{X}}$, $1\sigma_{\overline{X}}$, and $1.28\sigma_{\overline{X}}$, you will arrive at μ_1 values such that $\text{SEV}(\mu \leq \mu_1)$ = 0.8, 0.84, and 0.9, respectively. That's because severity goes in the same direction as power for non-rejection in T+. For familiar numbers of $\sigma_{\overline{X}}$'s added/subtracted to $\overline{x} = \mu_0 + d_0\sigma_{\overline{X}}$:

Claim	$(\mu \leq \overline{x} - 1\sigma_{\overline{X}})$	$(\mu \leq \overline{x})$	$(\mu \leq \overline{x} + 1\sigma_{\overline{X}})$	$(\mu \leq \overline{x} + 1.65\sigma_{\overline{X}})$	$(\mu \leq \overline{x} + 1.98\sigma_{\overline{X}})$
SEV	0.16	0.5	0.84	0.95	0.975

Now an overview of severity for test T+: Normal testing: H_0: $\mu \leq \mu_0$ vs. H_1: $\mu > \mu_0$ with σ known. The severity reinterpretation is set out using discrepancy parameter γ. We often use μ_1 where $\mu_1 = \mu_0 + \gamma$.

> Reject H_0 (with $\boldsymbol{x}_0$) 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 $\boldsymbol{x}_0$) 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 $d_0 = d(\boldsymbol{x}_0)$.

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: SEV($\mu \leq \mu_1$) is low (i.e., your test wasn't very capable of detecting discrepancy μ_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: SEV($\mu \leq \mu_1$) is high (i.e., your test was highly capable of detecting discrepancy μ_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 $\mu > \mu_1$: 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 $\mu > \mu_1$: 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 μ were no greater than μ_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($\boldsymbol{x}_0$) is not statistically significant, then $\mu \leq \overline{x} + k_\varepsilon \sigma/\sqrt{n}$ passes the test T+ with severity $(1 - \varepsilon)$

(FEV/SEV): If d($\boldsymbol{x}_0$) is statistically significant, then $\mu > \overline{x} - k_\varepsilon \sigma/\sqrt{n}$ passes test T+ with severity $(1 - \varepsilon)$,

where $\Pr(d(\boldsymbol{X}) > k_\varepsilon) = \varepsilon$ (Mayo and Spanos (2006), Mayo and Cox (2006).)