Sample Size (Rule of Thumb) | | 23 Aug 22

Sample Size Equation:

$$n \simeq \frac{16\sigma^2}{\delta^2} = \frac{16(\sigma^2)}{(\mu_0-\mu_1)^2} = \frac{16}{\Delta^2} \quad \text{where } \Delta = \frac{\mu_0-\mu_1}{\sigma}$$

Assumtions here are that variances are homogeneous ($\sigma_0^2 = \sigma_1^2 = \sigma^2$) and samples are of an equal size.

This formula is based on two-sample t-test which is aimed to determine whether difference between two populations is statistically significant. In other words, we want to test whether the metric μ_c is the same as μ_t where μ_c = mean of control group variable, μ_t = mean of treatment group variable:

$H_0: \mu_c = \mu_t$

$H_A: \mu_c \neq \mu_t$

Central limit theorem:

$$\bar{x} \sim N\left(\mu_c - \mu_t, \frac{2\sigma^2}{n}\right)$$ ← difference of two populations is normal distribution

We can transform it to standard normal variable Z:

$$Z = \frac{\bar{x} - (\mu_c - \mu_t)}{\sqrt{2}\sigma/\sqrt{n}} \sim N(0,1)$$

By the definition: α = significance level

Type II error rate (β) = P(accepting H_0 when difference exists) = β

$$\beta = P\left(\left|\frac{\bar{x}}{\sqrt{2}\sigma/\sqrt{n}}\right| \leq Z_{\alpha/2}\right) = P\left(-Z_{\alpha/2} \leq \frac{\bar{x}}{\sqrt{2}\sigma/\sqrt{n}} \leq Z_{\alpha/2}\right) =$$

$$= P\left(-Z_{\alpha/2} - \frac{\mu_c-\mu_t}{\sqrt{2}\sigma/\sqrt{n}} \leq \frac{\bar{x}-(\mu_c-\mu_t)}{\sqrt{2}\sigma/\sqrt{n}} \leq Z_{\alpha/2} - \frac{\mu_c-\mu_t}{\sqrt{2}\sigma/\sqrt{n}}\right) =$$

$$= \Phi\left(Z_{\alpha/2} - \frac{\mu_c-\mu_t}{\sqrt{2}\sigma/\sqrt{n}}\right) - \Phi\left(-Z_{\alpha/2} - \frac{\mu_c-\mu_t}{\sqrt{2}\sigma/\sqrt{n}}\right)$$

(Φ cummulative distribution function of normal distribution)

There are two possibilities: 1) $\mu_c > \mu_t$; 2) $\mu_c < \mu_t$.
Results are the same for both scenarios, so let's look @ one of them.
Assume $\mu_c > \mu_t$

$$\beta = \Phi\left(Z_{\alpha/2} - \frac{\mu_c - \mu_t}{\sqrt{2}\,\sigma/\sqrt{n}}\right) - \underbrace{\Phi\left(-Z_{\alpha/2} - \frac{\mu_c - \mu_t}{\sqrt{2}\,\sigma/\sqrt{n}}\right)}_{< \Phi(-Z_{\alpha/2}) \approx 0}$$

because $\frac{\mu_c - \mu_t}{\sqrt{2}\,\sigma/\sqrt{n}} > 0$

$\Rightarrow \boxed{\beta = \Phi(-Z_\beta)}$ Math trick: $x = \Phi(-Z_x)$ when $0 < x < 1$

$$\Rightarrow -Z_\beta = Z_{\alpha/2} - \frac{\mu_c - \mu_t}{\sqrt{2}\,\sigma/\sqrt{n}} \Rightarrow n \approx \frac{2(Z_{\alpha/2} + Z_\beta)^2 \sigma^2}{(\mu_c - \mu_t)^2}$$

Choose $\underbrace{\alpha = 0.05}_{\text{standards}}$, $\frac{\alpha}{2} = 0.025$, $\beta = 0.2$ $\quad \underbrace{\text{Power} = 1 - \beta = 0.8}_{\text{standards}}$.

$$\Rightarrow n \approx \frac{2(Z_{\alpha/2} + Z_\beta)^2 \sigma^2}{(\mu_c - \mu_t)^2}$$

for $\mu_c < \mu_t$ $\Rightarrow$ $n \approx \dfrac{2(Z_{\frac{\alpha}{2}} - Z_\beta)^2 \sigma^2}{(\mu_t - \mu_c)^2}$