Sample Size & Power Calculations

\[ X_1, \ldots, X_n \text{ iid Normal}(\mu_A, \sigma_A) \quad Y_1, \ldots, Y_m \text{ iid Normal}(\mu_B, \sigma_B) \]

Test \( H_0 : \mu_A = \mu_B \) vs \( H_a : \mu_A \neq \mu_B \) at \( \alpha = 0.05 \).

Test statistic: \( T = \frac{\bar{X} - \bar{Y}}{SD(\bar{X} - \bar{Y})} \).

\[ \rightarrow \] Critical value: \( C \) such that \( \Pr(|T| > C \mid \mu_A = \mu_B) = \alpha \).

Power: \( \Pr(|T| > C \mid \mu_A \neq \mu_B) \)
Power depends on...

- The design of your experiment
- What test you’re doing
- Chosen significance level, $\alpha$
- Sample size
- True difference, $\mu_A - \mu_B$
- Population SD’s, $\sigma_A$ and $\sigma_B$.

The case of known population SDs

Suppose $\sigma_A$ and $\sigma_B$ are known.

Then $\bar{X} - \bar{Y} \sim \text{Normal}(\mu_A - \mu_B, \sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}})$

Test statistic: $\tilde{Z} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}}$

If $H_0$ is true (i.e. $\mu_A = \mu_B$), we have $\tilde{Z} \sim \text{Normal}(0,1)$.

$\implies C = z_{\alpha/2}$ so that $\Pr(|\tilde{Z}| > C \mid \mu_A = \mu_B) = \alpha$.

For example, for $\alpha = 0.05$, $C = qnorm(0.975) = 1.96$. 
Power when the population SDs are known

If \( \mu_A - \mu_B = \Delta \), then
\[
Z = \frac{(X - Y) - \Delta}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}} \sim \text{Normal}(0,1)
\]

Pr \left( \left| \frac{X - Y}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}} \right| > 1.96 \right) = \Pr \left( \frac{X - Y}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}} > 1.96 \right) + \Pr \left( \frac{X - Y}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}} < -1.96 \right)

= \Pr \left( \frac{X - Y - \Delta}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}} > 1.96 - \frac{\Delta}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}} \right) + \Pr \left( \frac{X - Y - \Delta}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}} < -1.96 - \frac{\Delta}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}} \right)

= \Pr \left( Z > 1.96 - \frac{\Delta}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}} \right) + \Pr \left( Z < -1.96 - \frac{\Delta}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}} \right)

Calculations in R

Power = \Pr \left( Z > 1.96 - \frac{\Delta}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}} \right) + \Pr \left( Z < -1.96 - \frac{\Delta}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}} \right)

C <- qnorm(0.975)
se <- sqrt( sigmaA^2/n + sigmaB^2/m )
power <- 1-pnorm(C-delta/se) + pnorm(-C-delta/se)
Power curves

Special case: equal standard deviations and sample sizes.

\[
\text{Power} = \text{Pr} \left( Z > 1.96 - \frac{\Delta}{\sqrt{\frac{2 \sigma^2}{n}}} \right) + \text{Pr} \left( Z < -1.96 - \frac{\Delta}{\sqrt{\frac{2 \sigma^2}{n}}} \right)
\]

Power depends on . . .

\[
\text{Power} = \text{Pr} \left( Z > C - \frac{\Delta}{\sqrt{\frac{\sigma_A^2}{n} + \sigma_B^2/m}} \right) + \text{Pr} \left( Z < -C - \frac{\Delta}{\sqrt{\frac{\sigma_A^2}{n} + \sigma_B^2/m}} \right)
\]

- Choice of \( \alpha \) (which affects \( C \))
  Larger \( \alpha \) \( \rightarrow \) less stringent \( \rightarrow \) greater power.

- \( \Delta = \mu_A - \mu_B \) = the true “effect.”
  Larger \( \Delta \) \( \rightarrow \) greater power.

- Population SDs, \( \sigma_A \) and \( \sigma_B \)
  Smaller \( \sigma \)'s \( \rightarrow \) greater power.

- Sample sizes, \( n \) and \( m \)
  Larger \( n, m \) \( \rightarrow \) greater power.
Choice of sample size

We mostly influence power via \( n \) and \( m \).

Power is greatest when \( \frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m} \) is as small as possible.

Suppose the total sample size \( N = n + m \) is fixed.

\[
\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m} \text{ is minimized when } n = \frac{\sigma_A}{\sigma_A + \sigma_B} \times N \text{ and } m = \frac{\sigma_B}{\sigma_A + \sigma_B} \times N
\]

For example:

- If \( \sigma_A = \sigma_B \), we should choose \( n = m \).
- If \( \sigma_A = 2 \sigma_B \), we should choose \( n = 2m \).

That means, if \( \sigma_A = 4 \) and \( \sigma_B = 2 \), we might use \( n = 20 \) and \( m = 10 \).

Calculating the sample size

Suppose we seek 80\% power to detect a particular value of \( \mu_A - \mu_B = \Delta \), in the case that \( \sigma_A \) and \( \sigma_B \) are known.

(For convenience here, let’s pretend that \( \sigma_A = \sigma_B \) and that we plan to have equal sample sizes for the two groups.)

Power \( \approx \Pr \left( Z > C - \frac{\Delta}{\sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}} \right) = \Pr \left( Z > 1.96 - \frac{\Delta \sqrt{n}}{\sigma \sqrt{2}} \right) \)

\[\rightarrow \text{ Find } n \text{ such that } \Pr \left( Z > 1.96 - \frac{\Delta \sqrt{n}}{\sigma \sqrt{2}} \right) = 80\% .\]

Thus \( 1.96 - \frac{\Delta \sqrt{n}}{\sigma \sqrt{2}} = \text{qnorm}(0.2) = -0.842 \).

\[\rightarrow \sqrt{n} = \frac{\sigma}{\Delta} \{1.96 - (-0.842)\} \sqrt{2} \quad \rightarrow \quad n = 15.7 \times (\frac{\sigma}{\Delta})^2 \]
Equal but unknown population SDs

\( X_1, \ldots, X_n \) iid Normal(\( \mu_A, \sigma \)) \hspace{1cm} \( Y_1, \ldots, Y_m \) iid Normal(\( \mu_B, \sigma \))

Test \( H_0 : \mu_A = \mu_B \) vs \( H_a : \mu_A \neq \mu_B \) at \( \alpha = 0.05 \).

\[
\hat{\sigma}_p = \sqrt{\frac{s_A^2(n-1) + s_B^2(m-1)}{n+m-2}} \hspace{1cm} \text{SD}(\bar{X} - \bar{Y}) = \hat{\sigma}_p \sqrt{\frac{1}{n} + \frac{1}{m}}
\]

Test statistic: \( T = \frac{\bar{X} - \bar{Y}}{\text{SD}(\bar{X} - \bar{Y})} \).

In the case \( \mu_A = \mu_B \), \( T \) follows a t distribution with \( n + m - 2 \) d.f.

\[ \rightarrow \text{Critical value: } C = \text{qt}(0.975, n+m-2) \]

Power: equal but unknown pop’n SDs

\[
\text{Power} = \text{Pr}\left( \frac{|\bar{X} - \bar{Y}|}{\hat{\sigma}_p \sqrt{\frac{1}{n} + \frac{1}{m}}} > C \right)
\]

\[ \rightarrow \text{In the case } \mu_A - \mu_B = \Delta, \text{ the statistic } \frac{\bar{X} - \bar{Y}}{\hat{\sigma}_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \text{ follows a non-central t distribution.} \]

This distribution has two parameters:

\[ \rightarrow \text{The degrees of freedom (as before)} \]

\[ \rightarrow \text{The non-centrality parameter, } \frac{\Delta}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \]

\[
C \leftarrow \text{qt}(0.975, n + m - 2) \\
\text{se} \leftarrow \text{sigma} \times \text{sqrt} \left( \frac{1}{n} + \frac{1}{m} \right) \\
\text{power} \leftarrow 1 - \text{pt}(C, n+m-2, \text{ncp=delta/se}) + \text{pt}(-C, n+m-2, \text{ncp=delta/se})
\]
**Power: equal population SDs**

Power curves

![Power curves](image)

**A built-in function: power.t.test()**

Calculate power (or determine the sample size) for the t-test when:
- Sample sizes equal
- Population SDs equal

Arguments:
- `n = sample size`
- `delta = \Delta = \mu_2 - \mu_1`
- `sd = \sigma = population SD`
- `sig.level = \alpha = significance level`
- `power = the power`
- `type = type of data (two-sample, one-sample, paired)`
- `alternative = two-sided or one-sided test`
Examples

A. n = 10 for each group; effect = \( \Delta = 5 \); pop’n SD = \( \sigma = 10 \)

\[
\text{power.t.test(n=10, delta=5, sd=10)}
\]

\[\rightarrow 18\%\]

B. power = 80%; effect = \( \Delta = 5 \); pop’n SD = \( \sigma = 10 \)

\[
\text{power.t.test(delta=5, sd=10, power=0.8)}
\]

\[\rightarrow n = 63.8 \rightarrow 64 \text{ for each group}\]

C. power = 80%; effect = \( \Delta = 5 \); pop’n SD = \( \sigma = 10 \); one-sided

\[
\text{power.t.test(delta=5, sd=10, power=0.8, alternative="one.sided")}
\]

\[\rightarrow n = 50.2 \rightarrow 51 \text{ for each group}\]

Unknown and different pop’n SDs

\(X_1, \ldots, X_n\) iid Normal\((\mu_A, \sigma_A)\) \quad \(Y_1, \ldots, Y_m\) iid Normal\((\mu_B, \sigma_B)\)

Test \(H_0: \mu_A = \mu_B\) vs \(H_a: \mu_A \neq \mu_B\) at \(\alpha = 0.05\).

Test statistic: \(T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_A^2}{n} + \frac{s_B^2}{m}}}\)

To calculate the critical value for the test, we need the null distribution of \(T\) (that is, the distribution of \(T\) if \(\mu_A = \mu_B\)).

To calculate the power, we need the distribution of \(T\) given the value of \(\Delta = \mu_A - \mu_B\).

We don’t really know either of these.
Power by computer simulation

- Specify n, m, \(\sigma_A\), \(\sigma_B\), \(\Delta = \mu_A - \mu_B\), and the significance level, \(\alpha\).
- Simulate data under the model.
- Perform the proposed test and calculate the P-value.
- Repeat many times.

\[\text{Example:}\]
\[n = 5, m = 10, \sigma_A = 1, \sigma_B = 2, \quad \Delta = 0.0, 0.5, 1.0, 1.5, 2.0 \text{ or } 2.5.\]
Determining sample size

The things you need to know:

- Structure of the experiment
- Method for analysis
- Chosen significance level, $\alpha$ (usually 5%)
- Desired power (usually 80%)
- Variability in the measurements
  → If necessary, perform a pilot study, or use data from prior experiments or publications.
- The smallest meaningful effect
Reducing sample size

- Reduce the number of treatment groups being compared.

- Find a more precise measurement (e.g., average survival time rather than proportion dead).

- Decrease the variability in the measurements.
  - Make subjects more homogenous.
  - Use stratification.
  - Control for other variables (e.g., weight).
  - Average multiple measurements on each subject.