STAT 500

Sample Size Determination
for Two-Sample Inference

Sample Size Determination

- Frequently asked design questions:
  How many experimental units should I use?

- Four possible considerations:

  1. Include as many as you can afford or find

  2. Desired precision (standard error) of some estimator

  3. Width of a confidence interval

  4. Power of a hypothesis test

Based on Standard Error

- Choose a quantity that you want to estimate
- For example, population mean ($\boldsymbol{\mu}$)
  - Estimate the mean from one population using $\bar{\boldsymbol{Y}}$
  - The standard error of the estimator is $\boldsymbol{\sigma}/\sqrt{\boldsymbol{n}}$
  - The estimate for standard deviation $\hat{\boldsymbol{\sigma}} = \boldsymbol{S}$, sample standard deviation

Based on Standard Error

- Difference in population means ($\mu_1 - \mu_2$):

  Standard error of $\bar{Y}_1 - \bar{Y}_2 = S_p\sqrt{1/n_1 + 1/n_2}$

- Assuming $n_1 = n_2 = n$, we have:

  Standard error of $\bar{Y}_1 - \bar{Y}_2 = S_p\sqrt{2/n}$

- Specify an acceptable value for the s.e. and solve for n

Based on Standard Error

Estimation of the difference in two population means:

$$s.e. = \frac{\sqrt{2}S_p}{\sqrt{n}} \quad \Rightarrow \quad n = \frac{2S_p^2}{(s.e.)^2}$$

- Requires a value for S_p from:
  - a previous study
  - a pilot study
  - a guess

Based on Confidence Interval

- Width of CI (assuming $n_1 = n_2 = n$) is

$$\text{width} = 2\, t_{2(n-1),1-\alpha/2}\, S_p \sqrt{2/n}$$

- Find n to achieve specified width

$$n = 8\left(\frac{t_{2(n-1),1-\alpha/2} S_p}{\text{width}}\right)^2$$

- One difficulty is that n enters twice (sample size and d.f. for the t-distribution)

Based on Confidence Interval

- Compute initial value using $z_{1-\alpha/2}$ in place of the t-value

$$n_0 = 8\left(\frac{z_{1-\alpha/2}S_p}{\text{width}}\right)^2$$

- Then improve using

$$n = 8\left(\frac{t_{2(n_0-1),1-\alpha/2}S_p}{\text{width}}\right)^2$$

Based on Hypothesis Test

Four Possible Outcomes for Hypothesis Test

Decision	H_0 is true	H_0 is false
Reject H_0	**Type I Error**	Good Decision
Fail to reject H_0	Good Decision	**Type II Error**

Based on Hypothesis Test

Hypothesis Testing Errors

- Type I error $\Rightarrow$ reject H_o when it is true
  - Probability of Type I error
  - α = P(reject H_o | H_o is true)
  - Threshold for p-value, significance level
  - Value of α is typically set to be small, such as 0.05.
- Type II error $\Rightarrow$ fail to reject H_o when it is false
  - Probability of Type II error
  - β = P(fail to reject H_o | H_o is false)

Based on Hypothesis Test

- Power of a statistical test $= \mathbf{1} - \boldsymbol{\beta}$
- Function of a particular alternative to the null hypothesis:
  Power $= 1 - \boldsymbol{\beta} =$ P(reject $\boldsymbol{H_o}|$ specific alternative)
- For fixed $\boldsymbol{\alpha}$, power is determined by
  - How much the alternative deviates from the null hypothesis (true effect size, e.g., $\boldsymbol{\delta} = \boldsymbol{\mu_1} - \boldsymbol{\mu_2}$)
  - population variance ($\boldsymbol{\sigma^2}$)
  - sample sizes ($\boldsymbol{n_1}$, $\boldsymbol{n_2}$)

Based on Hypothesis Test

We can use power to determine the sample size because, for any given test, type I error rate α, power $1-\beta$, standard deviation σ or S, effect size $\delta = \mu_1 - \mu_2$, and sample size n are all related. Specifying four enables us to calculate the fifth.

Based on Hypothesis Test

- For a t-test of $H_o : \mu_1 = \mu_2$ against a two-sided alternative with
  - Equal sample sizes,
  - Type I error $= \alpha$,
  - Power $=$1-β for detecting $\delta = \mu_1 - \mu_2$,
  - Pooled estimate of the population variance denoted by S_p^2

  the required sample size for each group is

$$n = \frac{\left(t_{2(n-1),1-\alpha/2} + t_{2(n-1),1-\beta}\right)^2 (2S_p^2)}{\delta^2}$$

Based on Hypothesis Test

- As before, n enters twice. Use the same two-step approach. First compute

$$n_0 = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 \left(2S_p^2\right)}{\delta^2}$$

- Then update

$$n = \frac{\left(t_{2(n_0-1),1-\alpha/2} + t_{2(n_0-1),1-\beta}\right)^2 \left(2S_p^2\right)}{\delta^2}$$

- Common to use power values of 80%, 90% or 95%. Just as arbitrary as using α=5%.

- Can adapt to one-sided alternative by replacing $\alpha/2$ with α in the previous formulas

Derivation of Sample Size Formula

- Based on the model assumptions, we have the key result

$$\frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

- While controlling the type I error rate at α, we reject H_0 if

$$\left|\frac{(\bar{Y}_1 - \bar{Y}_2) - 0}{\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}}\right| > t_{(n_1+n_2-2),1-\alpha/2}$$

Derivation of Sample Size Formula

$$
\begin{aligned}
1-\beta &= Pr(\text{reject } H_0: \mu_1-\mu_2=0 \mid \mu_1-\mu_2=\delta) \\
&= Pr\left(\left|\frac{(\bar{Y}_1-\bar{Y}_2)-0}{\sqrt{S_p^2(\frac{1}{n_1}+\frac{1}{n_2})}}\right| > t_{(n_1+n_2-2),1-\alpha/2} \mid \mu_1-\mu_2=\delta\right) \\
&= Pr\left(|(\bar{Y}_1-\bar{Y}_2)-0| > t_{(n_1+n_2-2),1-\alpha/2}\sqrt{S_p^2(\frac{1}{n_1}+\frac{1}{n_2})} \mid \mu_1-\mu_2=\delta\right) \\
&= Pr\left((\bar{Y}_1-\bar{Y}_2) > t_{(n_1+n_2-2),1-\alpha/2}\sqrt{S_p^2(\frac{1}{n_1}+\frac{1}{n_2})} \mid \mu_1-\mu_2=\delta\right) \\
&\quad + Pr\left((\bar{Y}_1-\bar{Y}_2) < -t_{(n_1+n_2-2),1-\alpha/2}\sqrt{S_p^2(\frac{1}{n_1}+\frac{1}{n_2})} \mid \mu_1-\mu_2=\delta\right)
\end{aligned}
$$

Derivation of Sample Size Formula

$$1 - \beta \;=\; Pr\left(\frac{(\bar{Y}_1 - \bar{Y}_2) - \delta}{\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} > \frac{t_{(n_1+n_2-2),1-\alpha/2}\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})} - \delta}{\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} \mid \mu_1 - \mu_2 = \delta \right)$$

$$+ \; Pr\left(\frac{(\bar{Y}_1 - \bar{Y}_2) - \delta}{\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} < \frac{-t_{(n_1+n_2-2),1-\alpha/2}\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})} - \delta}{\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} \mid \mu_1 - \mu_2 = \delta \right)$$

This implies that

$$-t_{(n_1+n_2-2),1-\beta} \approx \frac{t_{(n_1+n_2-2),1-\alpha/2}\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})} - \delta}{\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

Derivation of Sample Size Formula

With $n_1 = n$ and $n_2 = Cn$ and $\delta > 0$ we have

$$-t_{((1+C)n-2),1-\beta} \approx \frac{t_{((1+C)n-2),1-\alpha/2}\sqrt{S_p^2(\frac{1}{n}(1+\frac{1}{C}))} - \delta}{\sqrt{S_p^2(\frac{1}{n}(1+\frac{1}{C}))}}$$

Solve for n:

$$n = \frac{(t_{((1+C)n-2),1-\alpha/2} + t_{((1+C)n-2),1-\beta})^2 S_p^2(1+\frac{1}{C})}{\delta^2}$$

Setting $n_1 = n_2 = n$, we get

$$n = \frac{(t_{2(n-1),1-\alpha/2} + t_{2(n-1),1-\beta})^2 2S_p^2}{\delta^2}$$

Example - Bone Density Study

Client's request:

I want to do a randomized experiment to compare two treatments for preventing bone loss in elderly women. I want to use a .05 level t-test that has at least an 80% chance of detecting of a difference of 4 units between mean loss in bone density for the two treatments. How many subjects do I need?

- After some discussion the client explains that she would like to use treatment groups of the same size

Example - Bone Density Study

- Further discussion about previous studies of bone density loss in elderly women suggests that the measured responses should be approximately normally distributed for each treatment $\Rightarrow$ a t-test could be used
- The null hypothesis is $H_o : \mu_1 = \mu_2$
- The alternative is two-sided and the deviation of interest is $\delta = |\mu_1 - \mu_2| = 4$
- Type I error level is $\alpha = 0.05$
- Power $=$ 1-β $=$ 0.80, and β=0.20
- Review of previous studies suggests $S_p^2 \approx 25$

Example - Bone Density Study

- Compute initial sample size value

$$
\begin{aligned}
n_0 &= \frac{\left(z_{0.975} + z_{0.80}\right)^2 \left(2S_p^2\right)}{\delta^2} \\
&= \frac{(1.96 + 0.841)^2 \, (2 \times 25)}{(4)^2} \\
&= 24.52 \quad \Rightarrow \quad 25
\end{aligned}
$$

Example - Bone Density Study

- Then compute

$$
\begin{aligned}
n &= \frac{\left(t_{48,0.975} + t_{48,0.80}\right)^2 \left(2S_p^2\right)}{\delta^2} \\
&= \frac{(2.01 + 0.849)^2 \,(2 \times 25)}{(4)^2} \\
&= 25.54 \quad \Rightarrow \quad 26
\end{aligned}
$$

- Use 26 subjects in each treatment group

Based on Confidence Interval

- Width of CI (assuming $n_1 = n_2 = n$) is

$$\text{width} = 2\, t_{2(n-1),1-\alpha/2}\, S_p \sqrt{2/n}$$

- Find n to achieve specified width

$$n = 8 \left(\frac{t_{2(n-1),1-\alpha/2} S_p}{\text{width}} \right)^2$$

Example - Bone Density Study

Instead of performing a test, suppose the client in the previous example wanted to construct a 95% confidence interval for the difference in the mean bone density loss for the two treatments of width 4 units, e.g.

$$(\bar{Y}_1 - \bar{Y}_2) \pm 2$$

- $\alpha = 0.05$ and $S_p^2 \approx 25$
- First compute

$$\begin{aligned} n_0 &= 8\left(\frac{z_{0.975} \times S_p}{\text{width}}\right)^2 \\ &= 8\left(\frac{1.96 \times 5}{4}\right)^2 = 48.02 \quad \Rightarrow \quad 49 \end{aligned}$$

Example - Bone Density Study

- Then improve using

$$
\begin{aligned}
\boldsymbol{n} &= \mathbf{8}\left(\frac{\boldsymbol{t}_{\mathbf{96,0.975}} \times \boldsymbol{S}_{\boldsymbol{p}}}{\text{width}}\right)^{\mathbf{2}} \\
&= \mathbf{8}\left(\frac{\mathbf{1.99} \times \mathbf{5}}{\mathbf{4}}\right)^{\mathbf{2}} \\
&= \mathbf{49.50} \quad \Rightarrow \quad \mathbf{50}
\end{aligned}
$$

- Use 50 subjects in each group

Example - Bone Density Study

Suppose the client simply wanted to use enough subjects to make the standard error for the estimated difference of the means no larger than 1.0, e.g.

$$\text{s.e. of } (\bar{Y}_1 - \bar{Y}_2) = S_p\sqrt{\frac{2}{n}} = 1$$

- $S_p^2 \approx 25$
- Compute

$$n = \frac{2 \times S_p^2}{(1)^2} = \frac{2(25)}{1} = 50$$

- Use 50 subjects in each group

Design of Comparative Studies

- Previously only considered $n_1 = n_2 = n$, why?
  - Could consider $n_1 = 2n_2$, same concept, modified formula
  - For fixed value of $n_1 + n_2$, $n_1 = n_2$ gives estimator of difference in means with smallest s.e. (assuming homogeneous variances)

Design of Comparative Studies

- Consider unequal group sizes in experiments when
  - One treatment is much more expensive than another
  - One treatment is limited or not readily available
  - Variation in responses differs between treatments
- Unequal sample sizes in observational studies are more common

Design of Comparative Studies

- Distribution of the t-statistic is not distorted by unequal variances when $n_1 = n_2$

- Size for one group may be limited by availability or higher cost, but

  $n_1 = 10,\ n_2 = 40$ is better than $n_1 = 10,\ n_2 = 10$

  $n_1 = 10,\ n_2 = 40$ is not as good as $n_1 = 25,\ n_2 = 25$