STAT 512 su 2021 Lec 09 slides

Sample size calculations

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Let $X_1, \dots, X_n$ be iid rvs with mean μ and variance σ^2.

Large-sample $(1-\alpha)100\%$ CI for μ is

$$\bar{X}_n \quad \pm \quad \underbrace{z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\text{margin of error}}$$

Strategy: Choose n to make the *margin of error* (ME) sufficiently small.

To achieve $\text{ME} \leq M^*$, set $n = \left\lceil \left(\frac{z_{\alpha/2}}{M^*}\right)^2 \sigma^2 \right\rceil$

Exercise: Derive above formula.

Exercise: Researchers want a 95% CI for μ with $\text{ME} \leq 1/2$. Believed that $\sigma \approx 2$. Recommend a sample size.

Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Bernoulli}(p)$. Then $\mu = p$ and $\sigma^2 = p(1-p)$.

A large-sample $(1-\alpha)100\%$ CI for p is

$$\bar{X}_n \quad \underbrace{\pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}}_{\text{margin of error}}$$

To achieve $\text{ME} \leq M^*$, set $n = \left\lceil \left(\frac{z_{\alpha/2}}{M^*}\right)^2 p(1-p) \right\rceil$

Replace p with

- an estimate from a previous study
- the value $1/2$, at which $p(1-p)$ is maximized (err on the large side).

Example: Want 99% CI for prop. of voters who will vote for a candidate with a ME not exceeding two percentage points. What sample size do we need?

For large n_1, n_2, a large-sample $(1-\alpha)100\%$ CI for $\mu_1 - \mu_2$ is

$$\bar{X} - \bar{Y} \pm \underbrace{z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}_{\text{margin of error}}$$

How should we choose n_1 and n_2?

To achieve $\text{ME} \leq M^*$, set

$$n_1 = \left\lceil \left(\frac{\sigma_1}{\sigma_1 + \sigma_2}\right) n^* \right\rceil \quad \text{and} \quad n_2 = \left\lceil \left(\frac{\sigma_2}{\sigma_1 + \sigma_2}\right) n^* \right\rceil,$$

where $n^* = \left(\frac{z_{\alpha/2}}{M^*}\right)^2 (\sigma_1 + \sigma_2)^2$.

Exercise: Derive the above.

Exercise: Suppose we have $\hat{\sigma}_1 = 2$ and $\hat{\sigma}_2 = 3$ from a previous study. We want a 99% CI for $\mu_1 - \mu_2$ with $\text{ME} \leq 1/2$. Recommend sample sizes n_1 and n_2.

For large n_1, n_2, a large-sample $(1-\alpha)100\%$ CI for $p_1 - p_2$ is

$$\hat{p}_1 - \hat{p}_2 \pm \underbrace{z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}_{\text{margin of error}}$$

How should we choose n_1 and n_2?

Just as before with $\sigma_1^2 = p_1(1-p_1)$ and $\sigma_2^2 = p_2(1-p_2)$.

Can use $p_1 = 1/2$ and $p_2 = 1/2$ to err on large side.