STAT 515 Lec 13 slides

Sample size calculations

Karl 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.

How to choose the sample size n?

Two questions for determining sample size

1. How narrow do we want confidence intervals to be?
2. What level of confidence do we want to have?

μ p

Recall: Our CIs for a mean and a proportion were of the form

$$\bar{X}_n \pm \text{ME} \quad \text{and} \quad \hat{p}_n \pm \text{ME},$$

where ME was the *margin of error*.

Strategy: Choose n large enough to make ME as small as desired.

For large n, an approximate $(1-\alpha)\times 100\%$ CI for μ is

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le M$$

$$\bar{X}_n \pm \underbrace{z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\text{ME}}.$$

$$\iff \frac{z_{\alpha/2}\,\sigma}{M} \le \sqrt{n}$$

$$\left(\frac{z_{\alpha/2}\,\sigma}{M}\right)^2 \le n$$

Sample size for estimating μ

For a maximum desired margin of error M and confidence level $1-\alpha$, take

$$n = \left\lceil \left(\frac{z_{\alpha/2}\cdot\sigma}{M}\right)^2 \right\rceil.$$

In the above, $\lceil x \rceil$ is the smallest integer greater than or equal to x ("round up").

The unknown σ can be replaced by an estimate from a previous or pilot study.

Exercise: Derive the sample size formula. ✓

Exercise: Researchers wish to estimate the mean gestation period of tortoises with the margin of error at the 95% confidence level no greater than 0.5 months.

Previous studies have estimated the standard deviation of gestation periods to be roughly 2 months. What sample size should be used?

$\alpha = 0.05$

$\sigma = 2$

$M = \frac{1}{2}$

$$n \geq \left(\frac{z_{\alpha/2}\, \sigma}{M} \right)^2 = \left(\frac{z_{0.025} \cdot 2}{1/2} \right)^2 = \left(\frac{(1.96) \cdot 2}{1/2} \right)^2$$

$$= 16 \cdot (1.96)^2 = 61.47$$

Take $n = 62$

For large n, an approximate $(1-\alpha)\times 100\%$ CI for p is Wald type interval.

$$\hat{p}_n \pm \underbrace{z_{\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}}_{\text{ME}}.$$

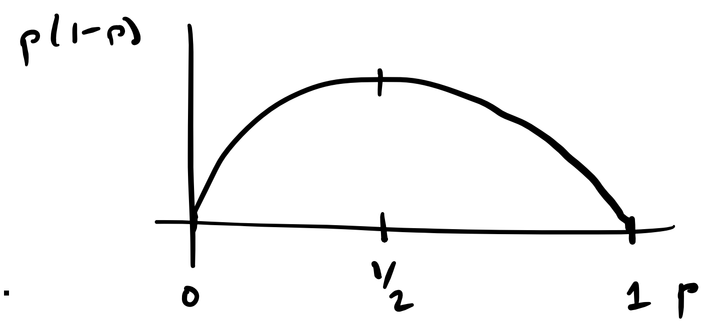
Sample size for estimating p

For a maximum desired margin of error M and confidence level $1-\alpha$, take

$$n = \left\lceil \left(z_{\alpha/2} \cdot \frac{\sqrt{p(1-p)}}{M} \right)^2 \right\rceil.$$

The unknown p can be replaced by

1. an estimate from a previous or pilot study
2. or simply by $p = 1/2$ (results in largest n).



Exercise: Derive the sample size formula.

$$\underbrace{z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}}_{ME} \leq M \iff \frac{z_{\alpha/2}\sqrt{p(1-p)}}{M} \leq \sqrt{n}$$

$$\iff \left(\frac{z_{\alpha/2}\sqrt{p(1-p)}}{M}\right)^2 \leq n$$

Exercise: We wish to estimate with 99% confidence the proportion of voters who will vote for a particular candidate with a ME no greater than 2 percentage points. What sample size do we need?

$\alpha = 0.01$

use $p = 1/2$

$M = 0.02$

$$n \geq \left(\frac{z_{\alpha/2} \sqrt{p(1-p)}}{M} \right)^2$$

$$= \left(\frac{z_{\frac{0.01}{2}} \sqrt{\frac{1}{2}(1-\frac{1}{2})}}{0.02} \right)^2$$

$z_{0.005}$

$$= \left(\frac{2.576 \cdot \frac{1}{2}}{.02} \right)^2 = 4147.36$$

take $\boxed{n = 4{,}148}$

Use $\alpha = 0.05$, I get

$$n \geq \left(\frac{\overset{1.96}{Z_{\frac{0.05}{2}}} \sqrt{\frac{1}{2}\left(1 - \frac{1}{2}\right)}}{0.02} \right)^2 = 2{,}401$$