Review

One Sample CI on pop. mean (μ)

1/ σ known

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

σ known $\rightarrow$ z

2/ σ unknown

$$\bar{x} \pm t_{(n-1,\, \alpha/2)} \cdot \frac{s}{\sqrt{n}}$$

σ unknown $\rightarrow$ t

One Sample CI on proportion (p)

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

proportions $\rightarrow$ z

STA258H5
University of Toronto Mississauga

Al Nosedal and Omid Jazi

Winter 2023

$n = ?$

SAMPLE SIZE SELECTION USING CONFIDENCE INTERVALS

↳ Calculating min sample sizes

Performed to satisfy certain criteria

- Interval is within a specified margin of error
- A specified confidence level

$\Big\}$ Both

1) when σ known

$$\bar{x} \pm \boxed{z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$$

equate this to desired margin of error with specified confidence level and solve for n.

When σ is Known

Parameter : μ.

Confidence interval :

$$\bar{Y} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right).$$

Empirical Rule

For any sample from a population that is close to Normally distributed:

- about 68% of all observations will lie in the interval $\mu \pm \sigma$
- about 95% of all observations will lie in the interval $\mu \pm 2\sigma$
- about 99.7% of all observations will lie in the interval $\mu \pm 3\sigma$

This suggests that for a sample, from a population that is close to Normally distributed,

$$\sigma \approx \frac{\text{Sample Range}}{4}$$

Crude estimate of pop. st dev

Analyzing pharmaceuticals

A manufacturer of pharmaceutical products analyzes a specimen from each batch of a product to verify the concentration of the active ingredient. The chemical analysis is not perfectly precise. Repeated measurements on the same specimen give slightly different results. Suppose we know that the results of repeated measurements follow a Normal distribution with mean μ equal to the true concentration and standard deviation $\sigma = 0.0068$ grams per liter. (That the mean of the population of all measurements is the true concentration says that the measurements process has no bias. The standard deviation describes the precision of the measurement.) The laboratory analyzes each specimen *n* times and reports the mean result.

σ known

Analyzing pharmaceuticals

margin of error

Confidence level

Management asks the laboratory to produce results accurate to within ± 0.005 with 95% confidence. How many measurements must be averaged to comply with this request?

$n = ?$

Example (Slide 5-6)

Since σ known $(\sigma = 0.0068)$ margin of error $= 0.005$. Conf level $= 0.95$

$$\bar{x} \pm \boxed{z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$$

margin of error $= 0.005$

Since σ known,
$z_{\alpha/2} = 1.96$ for a 95% conf level
(see previous examples)

$$z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 0.005$$

$$(1.96)\frac{(0.0068)}{\sqrt{n}} = 0.005$$

$$(\sqrt{n})^2 = (2.6656)^2$$

Sample size $\rightarrow$ $n = 7.105$

Sample size has to be a whole number. always round up

$$n = 8$$

How to calculate n when σ is not known?

Suppose we have raw data calculated with a number of data points

$x_1, x_2, \ldots, x_n$ ⟶ calculate S with data

$$\bar{X} \pm \underbrace{t_{(n-1, \alpha/2)}}_{} \cdot \frac{S}{\sqrt{n}}$$

use the sample size used to calculate S to determine $t_{(n-1, \alpha/2)}$

Solution

The desired margin of error is $m = 0.005$. For 95% confidence, our table gives $z_{0.025} = 1.96$. We know that $\sigma = 0.0068$. Therefore,

$$n = \left(\frac{z_{0.025}\sigma}{m}\right)^2 = \left(\frac{(1.96)(0.0068)}{0.005}\right)^2 = 7.1$$

Solution

Because 7 measurements will give a slightly larger margin of error than desired, and 8 measurements a slightly smaller margin of error, the lab must take 8 measurments on each specimen to meet management's demand. **Always round up to the next higher whole number when finding n.**

Example

σ known

Smith Travel Research provides information on the one-night cost of hotel rooms through-out the United States. Use \$ 2 as the desired margin of error and \$ 22.50 as the planning value for the population standard deviation to find the sample size recommended in a), b), and c).

a. A 90% confidence interval estimate of the population mean cost of hotel rooms.

b. A 95% confidence interval estimate of the population mean cost of hotel rooms.

c. A 99% confidence interval estimate of the population mean cost of hotel rooms.

Solution

a) Confidence level $= 0.90$ $(1-\alpha = 0.90)$, $B = 2$, $\sigma = 22.50$. From Table 3, $Z_{0.05} = 1.65$

$$n = \left(\frac{Z_{\frac{\alpha}{2}}\sigma}{B}\right)^2 = \left(\frac{(1.65)(22.50)}{2}\right)^2 \approx 344.5664$$

Final answer: **345**.

Solution

b) Confidence level $= 0.95$ $(1 - \alpha = 0.95)$, $B = 2$, $\sigma = 22.50$. From Table 3, $Z_{0.025} = 1.96$

$$n = \left(\frac{Z_{\frac{\alpha}{2}} \sigma}{B} \right)^2 = \left(\frac{(1.96)(22.50)}{2} \right)^2 \approx 486.2025$$

Final answer: **487**.

Solution

c) Confidence level $= 0.99$ $(1 - \alpha = 0.99)$, $B = 2$, $\sigma = 22.50$. From Table 3, $Z_{0.005} = 2.58$

$$n = \left(\frac{Z_{\frac{\alpha}{2}}\sigma}{B}\right)^2 = \left(\frac{(2.58)(22.50)}{2}\right)^2 \approx 842.4506$$

Final answer: **843**.

HW?

We want to construct a 95% CI for the length of iron rods produced by a certain factory. We know that these rods range in length from about 0.96m to about 1.04m. If we want the entire width of the confidence interval to be equal to 0.05m, what is the required sample size?

$$\sigma \approx \frac{\text{range}}{4}$$

$$n \approx 3$$

HW?

Suppose that you want to estimate the mean pH of rainfalls in an area that suffers from heavy pollution due to the discharge of smoke from a power plant. Assume that σ is in the neighborhood of 0.5 pH and that you want your estimate to lie within 0.1 of μ with probability near 0.99. Approximately how many rainfalls must be included in your sample (one pH reading per rainfall)?

$n = 166$

Interval Estimate of p

Draw a simple random sample of size n from a population with unknown proportion p of successes. An (approximate) confidence interval for p is:

$$\hat{p} \pm z_* \left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

where z_* is a number coming from the Standard Normal that depends on the confidence level required.

Calculating min sample size for proportions

$$\hat{p} \pm \boxed{z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

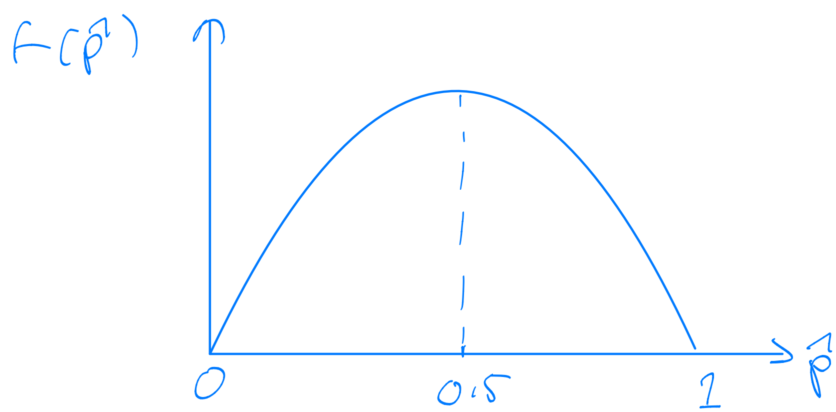
margin of error

If information regarding a good estimate of p is provided, then use it as p-hat (not common)

If information is not available use the conservative value for $\hat{p}$, $\hat{p} = 0.5$

$f(\hat{p}) = \hat{p}(1-\hat{p}) = \hat{p} - \hat{p}^2$

graphically



$0 < \hat{p} < 1$

Calculus

$$f(\hat{p}) = \hat{p} - \hat{p}^2$$

$$f'(\hat{p}) = \frac{d}{d\hat{p}} \hat{p} - \hat{p}^2 = 1 - 2\hat{p}$$

exercise

Set $f'(p) = 0$, solve for $\hat{p}$ $\longrightarrow$ $\hat{p} = 0.5$

Problem

Aisha Shariff and Yvette Ye are the candidates for mayor in a large city. You are planning a sample survey to determine what percent of the voters plan to vote for Shariff. This is a population proportion p. You will contact an SRS of registered voters in the city. You want to estimate p with 95% confidence and a margin of error no greater than 3%, or 0.03. How large a sample do you need?

$n = ?$

$\leq$

CI for a CI on proportions

$$\hat{p} \pm \boxed{z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

margin of error $\le 3\%$

For a 95% CI on p

$$z_{\alpha/2} = 1.96$$

No information provided on a good estimate of p.

use $\hat{p} = 0.5$

$$z_{\alpha/2} \cdot \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \le 3\%$$

$$1.96\,\frac{\sqrt{0.5(1-0.5)}}{\sqrt{n}} \le 0.03$$

$$\left(\frac{98}{3}\right)^2 \le \left(\sqrt{n}\right)^2$$

round up $\uparrow$ $1067.11 \le n$

min sample size $n = 1068$

Solution

We use the guess $p^* = 0.5$. The sample size you need is

$$n = \left(\frac{1.96}{0.03}\right)^2 (0.5)(1 - 0.5) = 1067.1$$

You should round up the result up to $n = 1068$. (Rounding down would give a margin of error slightly greater than 0.03).

Determining the Sample Size

Sample Size for an Interval Estimate of a Population Proportion.

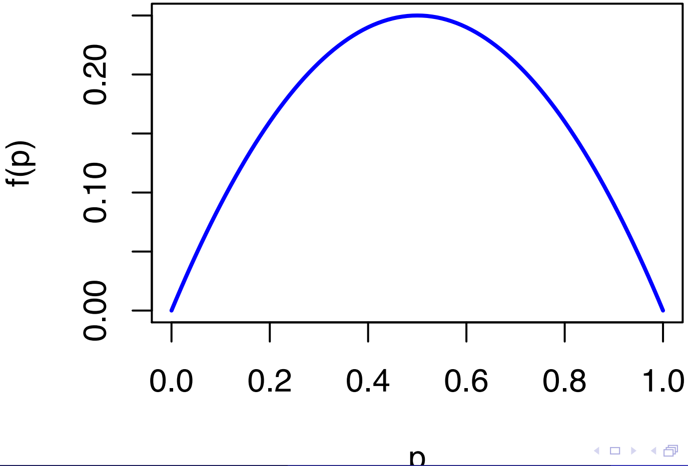
$$n = \left(\frac{z_*}{E}\right)^2 p^*(1 - p^*)$$

In practice, the planning value p^* can be chosen by one of the following procedures.

1. Use the sample proportion from a previous sample of the same or similar units.

2. Use a planning value of $p^* = 0.5$.

$f(p) = p(1-p)$



Problem

The percentage of people not covered by health care insurance in 2007 in the USA was 15.6%. A congressional committee has been charged with conducting a sample survey to obtain more current information.
a. What sample size would you recommend if the committee's goal is to estimate the current proportion of individuals without health care insurance with a margin of error of 0.03? Use a 95% confidence level.
b. Repeat part a) using a 99% confidence level.

Solution

a. $n = \left(\frac{z_*}{E}\right)^2 p^*(1-p^*) = \left(\frac{1.96}{0.03}\right)^2 (0.156)(1-0.156) = 563$

b. $n = \left(\frac{z_*}{E}\right)^2 p^*(1-p^*) = \left(\frac{2.58}{0.03}\right)^2 (0.156)(1-0.156) = 974$

HW?

A consumer advocacy group would like to find the proportion of consumers who bought the newest generation of iPhone and were happy with their purchase. How large a sample should they take to estimate p with 2% margin of error and 90% confidence?

$n = 1692$