Confidence Intervals - 1 Mean 1 Proportion

David Armstrong

UCI

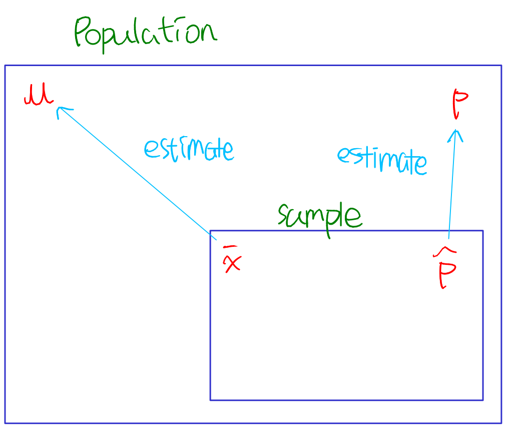
Confidence Intervals

- Provides the **interval of likely values** for the population parameter
- We are ___% confident that the **population parameter** is captured within the interval.
- If we took m random samples from a population and created ___% confidence intervals for each of those m samples, then if m is large, ___% of those confidence intervals would contain the **population parameter**.
- In order to create a confidence interval for a population parameter, we use two things

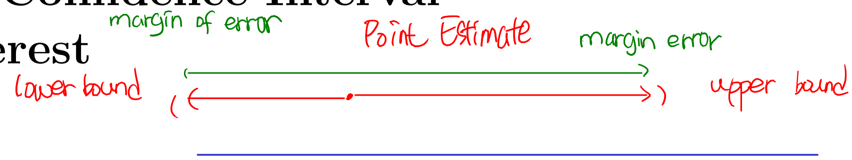
  We use the sample statistic as the best estimate of the population parameter.

  We also need a measure of spread.

- The **sample statistic** **ESTIMATES** the **population parameter**.



General Formula for a Confidence Interval for a Parameter of Interest



Point Estimate ± Margin of Error

$(PE - ME, PE + ME)$

- Point Estimate

  a single point that estimates the parameter

  sample statistic

  the midpoint of your confidence interval

  determines the location of the interval

- Margin of Error

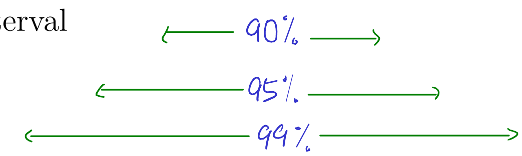
  an estimate of the variability of the population parameter

  distance from the midpoint of the interval to the edge z^* t^*

  critical value * spread

  determines the precision of the interval

  increasing the ME decreases precision of the interval



- critical value

  accounts for the level of confidence

  increasing confidence increases the width of the interval (decreases precision)

- spread

  accounts for the variability

  increasing the sample size decreases the width of the interval (increases precision)

Specific Formulas for Confidence Intervals

Parameter Description	Confidence Interval	Parameter Estimated
1 population proportion	$\hat{p} \pm Z\sqrt{\frac{\hat{p}\hat{q}}{n}}$	p
1 population mean (σ unknown)	$\bar{X} \pm t_{n-1}\frac{s}{\sqrt{n}}$	μ

poin estimate $\pm$ margin of error

$\hat{p} \pm z^* SE$

$\bar{x} \pm t^*_{n-1}(SE)$

Degree of Freedom = $n-1$

Assumptions for Confidence Intervals

1 population proportion

- $n\hat{p} \geq 10$ and $n\hat{q} \geq 10$
- Sample size is less than 10% of the population size; if we are sampling is w/out replacement
- The sample can be regarded as a simple random sample from the population of interest.
- The data values are assumed to be independent of each other.

1 population mean

CLT Bell shaped

- We need to have a large enough sample size ($n \geq 30$). For $n < 30$ with extreme skewness or outliers, you cannot use this method.

$x \sim N$ or AN

- Sample size is less than 10% of the population size; if we are sampling is w/out replacement
- The sample can be regarded as a simple random sample from the population of interest.
- The data values are assumed to be independent of each other.

$p = ?$ $\quad n = 875 \quad x = 33 \quad \hat{p} = \frac{x}{n} = 0.0606$

Ex: The drug Lipitor is meant to lower cholesterol levels. In a clinical trial of 875 randomly selected patients who received 12 mg doses of Lipitor daily, 53 reported a headache as a side effect.

a. What is the point estimate for the true population proportion of Lipitor users who will experience a headache as a side effect?

$PE = \hat{p} = 0.0606$ $\qquad \hat{q} = 1 - \hat{p} = 0.9394$

b. Verify that the requirements for constructing a confidence interval are satisfied.

Random ✓

Independent ✓

< 10% ✓

$n\hat{p} = 875(0.0606) = 53$ $\qquad n\hat{q} = 875(0.9394) = 822$

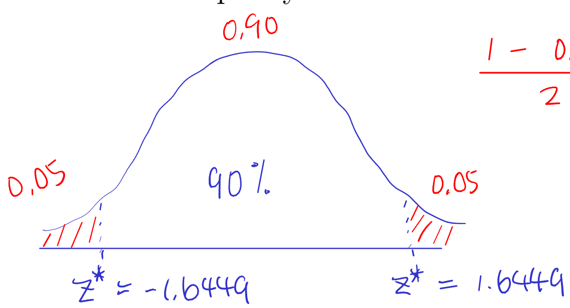
$53 > 10$ ✓ $\qquad 822 > 10$ ✓

c. Construct a 90% confidence interval for the population proportion of Lipitor users who will report a headache as a side effect.

$$PE \pm ME = \hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = 0.0606 \pm (1.6449)\sqrt{\frac{0.0606(0.9394)}{875}}$$

$$= (0.0473, 0.0739)$$

d. Interpret your confidence interval.



$$\frac{1 - 0.90}{2} = \frac{0.10}{2} = 0.05$$

qnorm(0.05, 0, 1) = ± 1.6449

We are 90% confidence the true population proportion of Lipitor users who will report a headache as side effect is captured within the interval (0.0473, 0.0739).

$\theta =$ $\quad$ $n = 10$

$\mu = ?$ $\quad$ $\bar{x} = 3.2$ $\quad$ $s = 0.78 \to t$

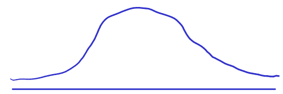
Ex: Suppose that a random survey of 10 teenagers found that the average amount of time they spend on the Internet each day is 3.2 hours with a sample standard deviation of 0.78 hours.

a. What assumptions must be made in order for a confidence interval to be valid?

< 10%

Random

Independent



Assume Bell shaped

$n \geq 30$? No

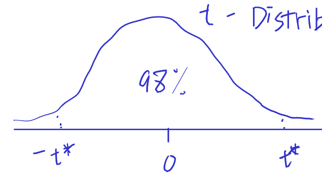
b. What are the point estimate and the standard error of the population average amount of time teenagers spend on the Internet each day?

$PE_{\bar{x}} = \bar{x} = 3.2$ $\quad$ $SE_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{0.78}{\sqrt{10}} = 0.2467$

c. Assuming the necessary conditions are met, calculate a 98% confidence interval for the average amount of time a teenager spends on the Internet each day.

$t^* = qt(0.98 + \frac{1-0.98}{2} = 0.99,\ df = n-1 = 9) = 2.8214$

t - Distribution



$PE \pm ME = \bar{x} \pm t^* (\frac{s}{\sqrt{n}}) = 3.2 \pm (2.8214)(\frac{0.78}{\sqrt{10}})$

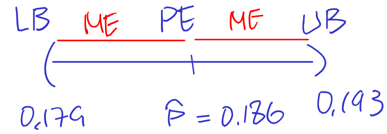
$= (2.5041, 3.8959)$

d. Interpret your confidence interval in the context of the problem using a complete sentence.

We are 98% confident the true population mean of the average amount of time a teenager spends on the internet each day is captured within the interval.

EX: A study of 10,485 randomly selected 30-39 year old Americans conducted by the Center for Disease Control in 2000 found with 95% confidence that the true proportion → p of 30-39 Americans that are overweight is between 0.179 and 0.193.

a. Find the point estimate for the true proportion.



$$PE = \frac{UB + LB}{2} = \frac{0.193 + 0.179}{2} = 0.186$$

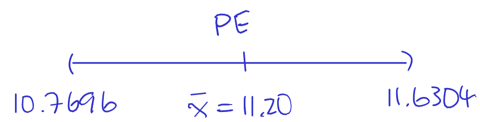
b. Find the margin of error.

$$ME = \begin{cases} UB - PE = 0.193 - 0.186 \\ PE - LB = 0.186 - 0.179 \\ \frac{UB - LB}{2} = \frac{0.193 - 0.179}{2} \end{cases} \rightarrow 0.07$$

Ex: The IRS is investigating the income obtained from tips in various types of services. In a small pilot study at a gourmet restaurant the agents randomly selected 30 charge-card receipts and computed the average tip to be \$11.20 with a standard deviation of \$1.798. A confidence interval for the average tip given by patrons of this restaurant is (\$10.7696 , \$11.6304) $\theta = ?$

$\mu =$ $\bar{x} = 11.20$ $s = 1.798$ $n = 30$

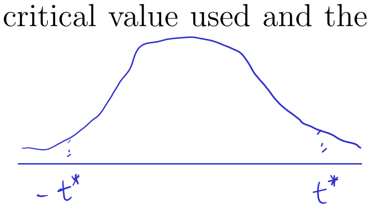
a. The margin of error for the confidence interval is:



$$ME = t^* \frac{s}{\sqrt{n}} = UB - PE = 11.6304 - 11.20$$

$$= 0.4304$$

b. Based on the above confidence interval, determine the critical value used and the confidence level.



$$ME = t^* \frac{s}{\sqrt{n}}$$

$$t^* = \frac{ME}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{0.4304}{\left(\frac{1.798}{\sqrt{30}}\right)} = 1.3111$$

pt (1.3111 , 29) − pt (−1.3111 , 29)

= 0.80

↑

1 − 2 pt(−1.3111 , 29)