Confidence Intervals

Dr. Ab Mosca (they/them)

Slides based off slides courtesy of OpenIntro and John McGreedy of Johns Hopkins University

Warm Up

4.4 Triathlon times, Part I. In triathlons, it is common for racers to be placed into age and gender groups. Friends Leo and Mary both completed the Hermosa Beach Triathlon, where Leo competed in the *Men, Ages 30 - 34* group while Mary competed in the *Women, Ages 25 - 29* group. Leo completed the race in 1:22:28 (4948 seconds), while Mary completed the race in 1:31:53 (5513 seconds). Obviously Leo finished faster, but they are curious about how they did within their respective groups. Can you help them? Here is some information on the performance of their groups:

- The finishing times of the *Men, Ages 30 - 34* group has a mean of 4313 seconds with a standard deviation of 583 seconds.
- The finishing times of the *Women, Ages 25 - 29* group has a mean of 5261 seconds with a standard deviation of 807 seconds.
- The distributions of finishing times for both groups are approximately Normal.

Remember: a better performance corresponds to a faster finish.

(a) Write down the short-hand for these two normal distributions.
(b) What are the Z-scores for Leo's and Mary's finishing times? What do these Z-scores tell you?
(c) Did Leo or Mary rank better in their respective groups? Explain your reasoning.
(d) What percent of the triathletes did Leo finish faster than in his group?
(e) What percent of the triathletes did Mary finish faster than in her group?
(f) If the distributions of finishing times are not nearly normal, would your answers to parts (b) - (e) change? Explain your reasoning.

Warm Up

(a) Let X denote the finishing times of *Men, Ages 30 - 34* and Y denote the finishing times of emphWomen, Ages 25 - 29. Then,

$$X \sim N(\mu = 4313, \sigma = 583)$$
$$Y \sim N(\mu = 5261, \sigma = 807)$$

(b) The Z scores can be calculated as follows:

$$Z_{Leo} = \frac{x - \mu}{\sigma} = \frac{4948 - 4313}{583} = 1.09$$
$$Z_{Mary} = \frac{y - \mu}{\sigma} = \frac{5513 - 5261}{807} = 0.31$$

Leo finished 1.09 standard deviations above the mean of his group's finishing time and Mary finished 0.31 standard deviations above the mean of her group's finishing time.

(c) Mary ranked better since she she has a lower Z score indicating that her finishing time is relatively shorter.

(d) Leo:

$$\begin{aligned} P(Z > 1.09) &= 1 - P(Z < 1.09) \\ &= 1 - 0.8621 \\ &= 0.1379 \rightarrow 13.79\% \end{aligned}$$

(e) Mary:

$$\begin{aligned} P(Z > 0.31) &= 1 - P(Z < 0.31) \\ &= 1 - 0.6217 \\ &= 0.3783 \rightarrow 37.83\% \end{aligned}$$

(f) Answer to part (b) would not change as Z scores can be calculated for distributions that are not normal. However, we could not answer parts (c)-(e) since we cannot use the Z table to calculate probabilities and percentiles without a normal model.

Warm Up

4.18 Chickenpox, Part I. Boston Children's Hospital estimates that 90% of Americans have had chickenpox by the time they reach adulthood.[32]

(a) Suppose we take a random sample of 100 American adults. Is the use of the binomial distribution appropriate for calculating the probability that exactly 97 out of 100 randomly sampled American adults had chickenpox during childhood? Explain.

(b) Calculate the probability that exactly 97 out of 100 randomly sampled American adults had chickenpox during childhood.

(c) What is the probability that exactly 3 out of a new sample of 100 American adults have *not* had chickenpox in their childhood?

(d) What is the probability that at least 1 out of 10 randomly sampled American adults have had chickenpox?

(e) What is the probability that at most 3 out of 10 randomly sampled American adults have *not* had chickenpox?

Warm Up

(a) In order to determine if we can we use the binomial distribution to calculate the probability of finding exactly 97 people out of a random sample of 100 American adults had chickenpox in childhood, we need to check if the binomial conditions are met:

1. Independent trials: In a random sample, whether or not one adult has had chickenpox does not depend on whether or not another one has.
2. Fixed number of trials: $n = 100$.
3. Only two outcomes at each trial: Have or have not had chickenpox.
4. Probability of a success is the same for each trial: $p = 0.90$.

(b) Let X be number of people who have had chickenpox in childhood, using a binomial distribution with $n = 100$ and $p = 0.90$:

$$P(X = 97) = \binom{100}{97} \times 0.90^{97} \times 0.10^{3} = 0.0059$$

(c) P(97 out of 100 did have chickenpox) = P(3 out of 100 did not have chickenpox in childhood) = 0.0059

(d) P(at least 1) = P(greater than or equal to 1):

$$\begin{aligned} P(X \geq 1) &= P(X = 1) + P(X = 2) + \cdots + P(X = 10) \\ &= 1 - P(X = 0) \\ &= 1 - 0.10^{10} \\ &\approx 1 \end{aligned}$$

(e) P(at most 3 did not have chickenpox) = P(less than or equal to 3 where p = 0.10)

$$\begin{aligned} P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= \binom{10}{0} \times 0.10^{0} \times 0.90^{10} + \binom{10}{1} \times 0.10^{1} \times 0.90^{9} + \binom{10}{2} \times 0.10^{2} \times 0.90^{8} + \binom{10}{3} \times 0.10^{3} \times 0.90^{7} \\ &= 0.3487 + 0.3874 + 0.1937 + 0.0574 \\ &= 0.9872 \end{aligned}$$

Plan for Today

- Confidence Intervals (CI’s) for means
- Confidence Intervals for proportions

Confidence intervals

- A plausible range of values for the population parameter is called a *confidence interval*.
- Using only a sample statistic to estimate a parameter is like fishing in a murky lake with a spear, and using a confidence interval is like fishing with a net.

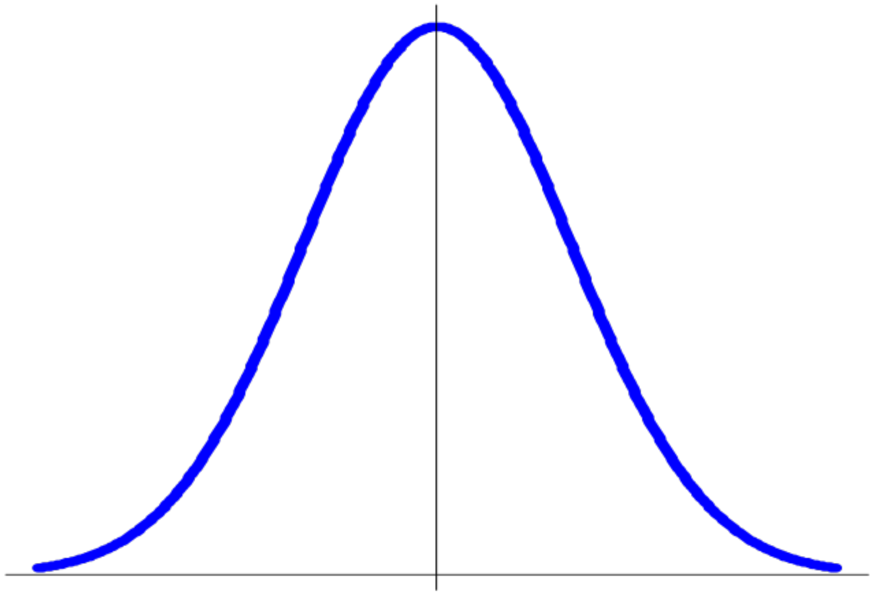
We can throw a spear where we saw a fish but we will probably miss. If we toss a net in that area, we have a good chance of catching the fish.

- If we report a point estimate, we probably won't hit the exact population parameter. If we report a range of plausible values we have a good shot at capturing the parameter.

Photos by Mark Fischer (http://www.flickr.com/photos/fischerfotos/7439791462) and Chris Penny (http://www.flickr.com/photos/clearlydived/7029109617) on Flickr.

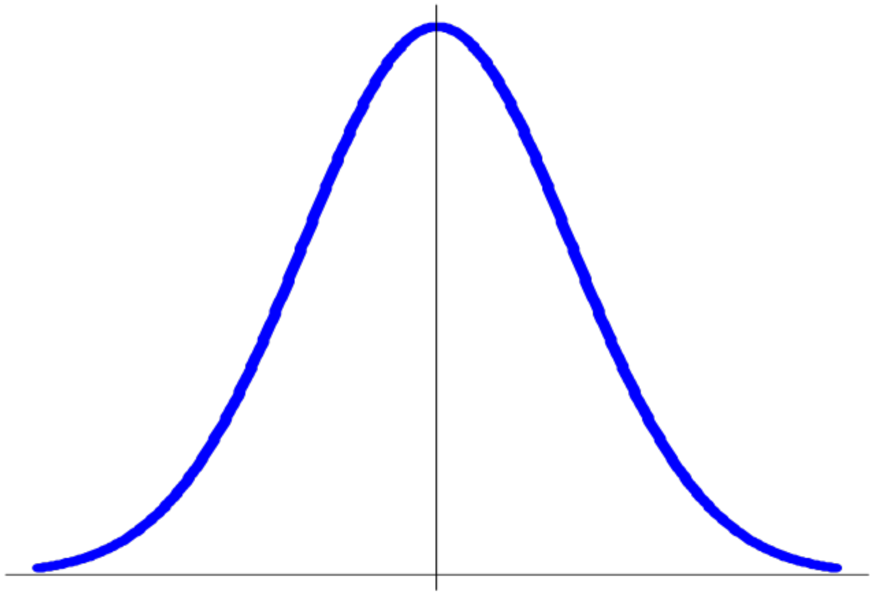
The Central Limit Theorem, Revisited—1

- The CLT states that if all possible random samples of the same size, *n*, were taken from the same population, and a summary statistic were computed (mean, proportion) for each sample, then the distribution of the summary statistic values across these samples is:



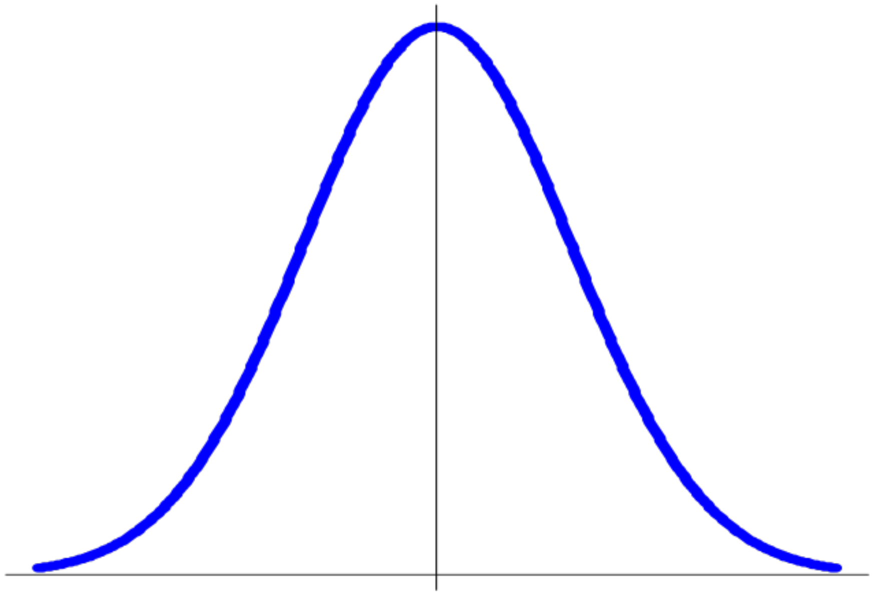
The Central Limit Theorem, Revisited—2

- Most (95%) of the summary statistic values fall within two standard deviations of the truth they are estimating (even more [99%] fall within three standard deviations)



The Central Limit Theorem, Revisited—3

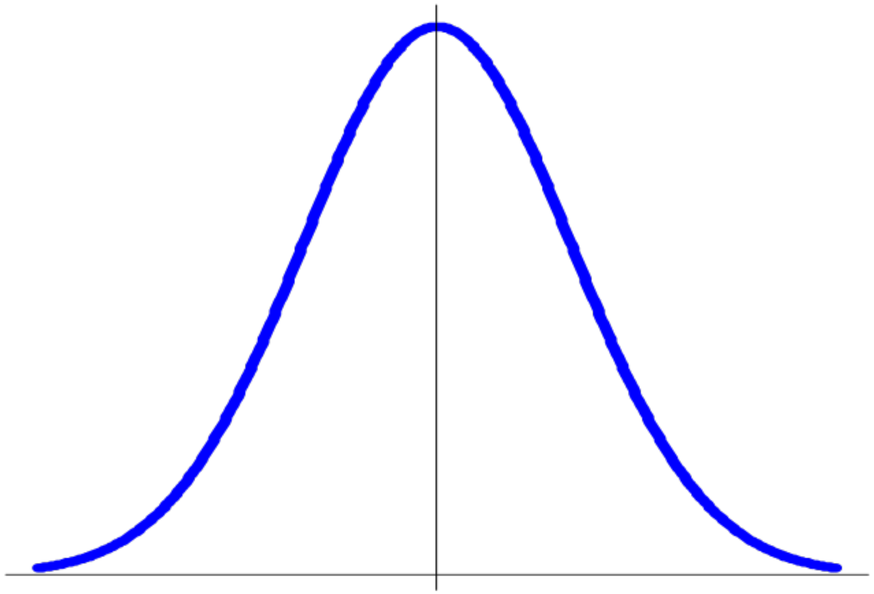
- In research, only one sample will be taken from each population under study
  - So how will the CLT help research?



The Central Limit Theorem, Revisited—3

- In research, only one sample will be taken from each population under study
  - So how will the CLT help research?

By allowing us to take a point estimate and extrapolate a range of very likely true population values.

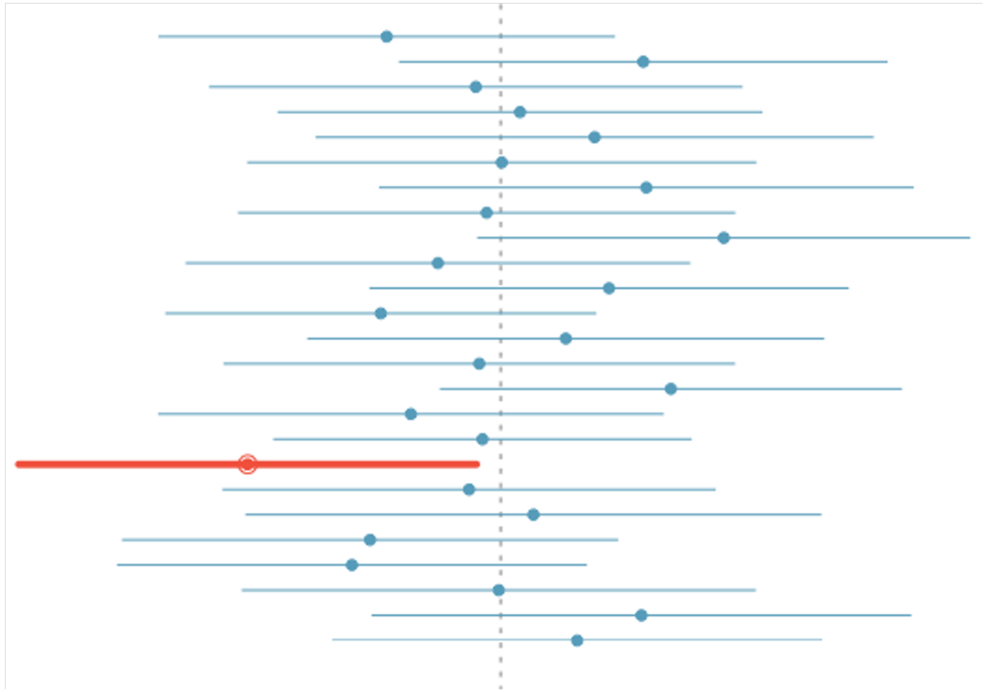


Examples with Means

What does 95% confident mean?

- Suppose we took many samples and built a confidence interval from each sample using the equation *point estimate* ± *2 x SE*.
- Then about 95% of those intervals would contain the true population mean (μ).

- The figure shows this process with 25 samples, where 24 of the resulting confidence intervals contain the true average number of exclusive relationships, and one does not.



Example: 95% CI for SBP Mean, Bases on Sample of 113 Men—1

- Random sample of 113 men taken from a clinical population
  - Sample summary statistics include:

$$(\text{Estimate of } \mu)\text{: } \bar{x} = 123.6 \text{ mmHg}$$

$$(\text{Estimate of } \sigma)\text{: } s = 12.9 \text{ mmHg}$$

- We can estimate the standard error of sample means based on random samples of 113 men from this population by:

$$\text{SE}(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{12.9 \text{ mmHg}}{\sqrt{113}} \approx 1.2 \text{ mmHg}$$

Example: 95% CI for SBP Mean, Bases on Sample of 113 Men—1

- Rand… …nical popu…
  - Sa…

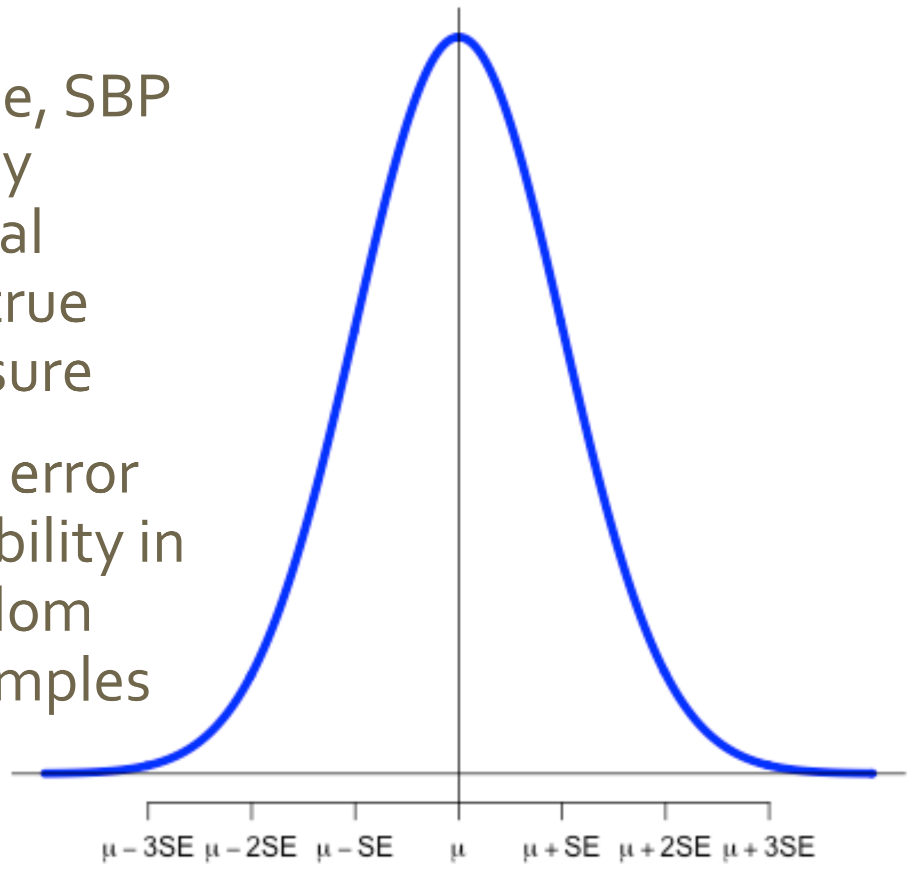
SE of a statistic is the standard deviation of the sampling distribution

- We can estimate the standard error of sample means based on random samples of 113 men from this population by:

$$\mathrm{SE}(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{12.9 \text{ mmHg}}{\sqrt{113}} \approx 1.2 \text{ mmHg}$$

Example: 95% CI for SBP Mean, Bases on Sample of 113 Men—2

- The standard error estimate quantifies how far, on average, SBP means based on 113 randomly sampled men from this clinical population will fall from the true population mean blood pressure
- In other words, this standard error estimate quantifies the variability in sample means based on random samples of 113, across the samples



Example: 95% CI for SBP Mean, Bases on Sample of 113 Men—3

- Since the CLT tells us the theoretical distribution of all possible sample means based on random samples of n=113 is approximately normal, we can estimate a 95% CI for the true population mean by

$$\bar{x} \pm 2\widehat{SE}(\bar{x}) = 123.6 \text{ mmHg } \pm 2(1.2 \text{ mmHg})$$

Example: Heritage Health Length of Stay—1

- All patients with at least one inpatient stay in 2011 (n=12,928)

$$(\text{Estimate of } \mu)\text{: } \bar{x} = 4.3 \text{ days}$$

$$(\text{Estimate of } \sigma)\text{: } s = 4.9 \text{ days}$$

- We can estimate the standard error of sample means based on random samples of 12,928 persons from this population by

$$\widehat{SE}(\bar{x}) = \frac{s}{\sqrt{n}} =$$

Example: Heritage Health Length of Stay—1

- All patients with at least one inpatient stay in 2011 (n=12,928)

$$\text{(Estimate of } \mu\text{)}: \bar{x} = 4.3 \text{ days}$$

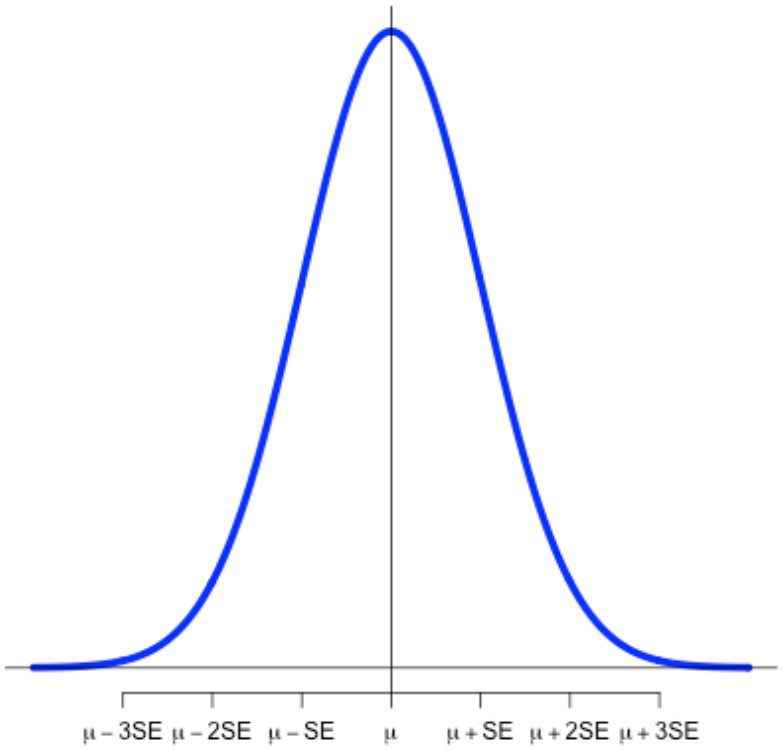
$$\text{(Estimate of } \sigma\text{)}: s = 4.9 \text{ days}$$

- We can estimate the standard error of sample means based on random samples of 12,928 persons from this population by

$$\widehat{SE}(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{4.9 \text{ days}}{\sqrt{12,928}} \approx 0.04 \text{ days}$$

Example: Heritage Health Length of Stay—2

- The standard error estimate quantifies how far length of stay means based on 12,928 patients from the insurance population will fall from the true population mean length of stay



Example: Heritage Health Length of Stay—3

- Since the CLT tells us the theoretical distribution of all possible sample means based on samples of n=12,928 is approximately normal, we can estimate a 95% CI for the true population mean by

$$\bar{x} \pm 2\widehat{SE}(\bar{x}) =$$

Example: Heritage Health Length of Stay—3

- Since the CLT tells us the theoretical distribution of all possible sample means based on samples of n=12,928 is approximately normal, we can estimate a 95% CI for the true population mean by

$$\bar{x} \pm 2\widehat{SE}(\bar{x}) = 4.3 \text{ days} \pm 2(0.04 \text{ days})$$

Ex. Average number of exclusive relationships

A random sample of 50 college students were asked how many exclusive relationships they have been in so far. This sample yielded a mean of 3.2 and a standard deviation of 1.74. **Estimate the true average number of exclusive relationships using this sample.**

Average number of exclusive relationships

A random sample of 50 college students were asked how many exclusive relationships they have been in so far. This sample yielded a mean of 3.2 and a standard deviation of 1.74. **Estimate the true average number of exclusive relationships using this sample.**

The approximate 95% confidence interval is defined as

point estimate ± 2 x SE

$$SE = \frac{s}{\sqrt{n}}$$

Average number of exclusive relationships

A random sample of 50 college students were asked how many exclusive relationships they have been in so far. This sample yielded a mean of 3.2 and a standard deviation of 1.74. Estimate the true average number of exclusive relationships using this sample.

The approximate 95% confidence interval is defined as : point estimate ± 2 x SE

$\bar{x} = 3.2$ $\quad\quad$ $s = 1.74$

$$SE = \frac{s}{\sqrt{n}} = \frac{1.74}{\sqrt{50}} \approx 0.25$$

$\bar{x} \pm 2 \times SE$ $\quad\quad$ $\rightarrow 3.2 \pm 2 \times 0.25.$

$\rightarrow (3.2 - 0.5, 3.2 + 0.5)$

$\rightarrow (2.7, 3.7)$

Practice

Which of the following is the correct interpretation of this confidence interval?

(2.7, 3.7)

We are 95% confident that

(a) the average number of exclusive relationships college students in this sample have been in is between 2.7 and 3.7.

(b) college students on average have been in between 2.7 and 3.7 exclusive relationships.

(c) a randomly chosen college student has been in 2.7 to 3.7 exclusive relationships.

(d) 95% of college students have been in 2.7 to 3.7 exclusive relationships.

Practice

Which of the following is the correct interpretation of this confidence interval?

(2.7, 3.7)

We are 95% confident that

(a) the average number of exclusive relationships college students in this sample have been in is between 2.7 and 3.7.

(b) college students on average have been in between 2.7 and 3.7 exclusive relationships.

(c) a randomly chosen college student has been in 2.7 to 3.7 exclusive relationships.

(d) 95% of college students have been in 2.7 to 3.7 exclusive relationships.

A Note on Level of Confidence

- 95% confidence intervals are the "industry standard" in research
- It is certainly possible, however, to estimate intervals with different levels of confidence.
  - For example:
  - 90% confidence interval for a population mean

$$\bar{x} \pm 1.65 \times \left(\frac{s}{\sqrt{n}}\right)$$

  - 99% confidence interval for a population mean

$$\bar{x} \pm 2.58 \times \left(\frac{s}{\sqrt{n}}\right)$$

Summary

- 95% confidence intervals for a population mean μ, based on a data in random sample taken from the population, can be constructed by:

$$\bar{x} \pm 2\widehat{SE}(\bar{x})$$

$$\text{Where } \widehat{SE}(\bar{x}) = \frac{s}{\sqrt{n}}$$

- The standard error of the sample mean quantifies the variation in sample means across random samples of the same size (from the same population)
- The level of confidence can be changed by adjusting the number of standard errors subtracted from and added to the sample mean

Examples with Proportions

Ex. Facebook's categorization of user interests

Most commercial websites (e.g. social media platforms, news out- lets, online retailers) collect a data about their users' behaviors and use these data to deliver targeted content, recommendations, and ads. To understand whether Americans think their lives line up with how the algorithm-driven classification systems categorizes them, Pew Research asked a representative sample of 850 American Facebook users how accurately they feel the list of categories Facebook has listed for them on the page of their supposed interests actually represents them and their interests. 67% of the respondents said that the listed categories were accurate. Estimate the true proportion of American Facebook users who think the Facebook categorizes their interests accurately.

https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/

Ex. Facebook’s categorization of user interests

Most commercial websites (e.g. social media platforms, news out- lets, online retailers) collect a data about their users’ behaviors and use these data to deliver targeted content, recommendations, and ads. To understand whether Americans think their lives line up with how the algorithm-driven classification systems categorizes them, Pew Research asked a representative sample of 850 American Facebook users how accurately they feel the list of categories Facebook has listed for them on the page of their supposed interests actually represents them and their interests. 67% of the respondents said that the listed categories were accurate. **Estimate the true proportion of American Facebook users who think the Facebook categorizes their interests accurately.**

https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/

Calculate the 95% confidence interval for the true proportion

Ex. Facebook's categorization of user interests

Most commercial websites (e.g. social media platforms, news out- lets, online retailers) collect a data about their users' behaviors and use these data to deliver targeted content, recommendations, and ads. To understand whether Americans think their lives line up with how the algorithm-driven classification systems categorizes them, Pew Research asked a representative sample of 850 American Facebook users how accurately they feel the list of categories Facebook has listed for them on the page of their supposed interests actually represents them and their interests. 67% of the respondents said that the listed categories were accurate. **Estimate the true proportion of American Facebook users who think the Facebook categorizes their interests accurately.**

https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/

Calculate the 95% confidence interval for the true proportion

The approximate 95% confidence interval is defined as

$$\textit{point estimate} \pm 2 \times SE$$

Ex. Facebook's categorization of user interests

Most commercial websites (e.g. social media platforms, news out- lets, online retailers) collect a data about their users' behaviors and use these data to deliver targeted content, recommendations, and ads. To understand whether Americans think their lives line up with how the algorithm-driven classification systems categorizes them, Pew Research asked a representative sample of 850 American Facebook users how accurately they feel the list of categories Facebook has listed for them on the page of their supposed interests actually represents them and their interests. 67% of the respondents said that the listed categories were accurate. **Estimate the true proportion of American Facebook users who think the Facebook categorizes their interests accurately.**

https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/

Calculate the 95% confidence interval for the true proportion

The approximate 95% confidence interval is defined as

$$point\ estimate \pm \ 2 \ \times SE$$

Ex. Facebook's categorization of user interests

Most commercial websites (e.g. social media platforms, news out- lets, online retailers) collect a data about their users' behaviors and use these data to deliver targeted content, recommendations, and ads. To understand whether Americans think their lives line up with how the algorithm-driven classification systems categorizes them, Pew Research asked a representative sample of 850 American Facebook users how accurately they feel the list of categories Facebook has listed for them on the page of their supposed interests actually represents them and their interests. 67% of the respondents said that the listed categories were accurate. **Estimate the true proportion of American Facebook users who think the Facebook categorizes their interests accurately.**

https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/

Calculate the 95% confidence interval for the true proportion

The approximate 95% confidence interval is defined as

$$point\ estimate \pm 2 \times SE$$

$$SE = \sqrt{\frac{p\,(1-p)}{n}}$$

population proportion

sample size

Ex. Facebook's categorization of user interests

Most commercial websites (e.g. social media platforms, news out- lets, online retailers) collect a data about their users' behaviors and use these data to deliver targeted content, recommendations, and ads. To understand whether Americans think their lives line up with how the algorithm-driven classification systems categorizes them, Pew Research asked a representative sample of 850 American Facebook users how accurately they feel the list of categories Facebook has listed for them on the page of their supposed interests actually represents them and their interests. 67% of the respondents said that the listed categories were accurate. **Estimate the true proportion of American Facebook users who think the Facebook categorizes their interests accurately.**

https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/

Calculate the 95% confidence interval for the true proportion

The approximate 95% confidence interval is defined as

$$point\ estimate \pm 2 \times SE$$

$$SE = \sqrt{\frac{\boxed{p}(1-p)}{n}}$$

population proportion

sample size

Ex. Facebook's categorization of user interests

Most commercial websites (e.g. social media platforms, news out- lets, online retailers) collect a data about their users' behaviors and use these data to deliver targeted content, recommendations, and ads. To understand whether Americans think their lives line up with how the algorithm-driven classification systems categorizes them, Pew Research asked a representative sample of 850 American Facebook users how accurately they feel the list of categories Facebook has listed for them on the page of their supposed interests actually represents them and their interests. 67% of the respondents said that the listed categories were accurate. **Estimate the true proportion of American Facebook users who think the Facebook categorizes their interests accurately.**

https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/

population proportion

Calculate the 95% confidence interval for the true proportion

The approximate 95% confidence interval is defined as

$$point\ estimate \pm 2 \times SE$$

$$SE = \sqrt{\frac{\boxed{p}(1-p)}{n}}$$

sample size

Ex. Facebook's categorization of user interests

$$\hat{p} = 0.67 \qquad n = 850$$

The approximate 95% confidence interval is defined as

$$point\ estimate \pm 2 \times SE$$

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.67 \times 0.33}{850}} \approx 0.016$$

$$\begin{aligned} \hat{p} \pm 2 \times SE &= 0.67 \pm 2 \times 0.016 \\ &= (0.67 - 0.03, 0.67 + 0.03) \\ &= (0.64, 0.70) \end{aligned}$$

Ex. Facebook's categorization of user interests

95% confidence interval: $(0.64, 0.70)$

Which of the following is the correct interpretation of this confidence interval? We are 95% confident that...

(a) 64% to 70% of American Facebook users in this sample think Facebook categorizes their interests accurately.

(b) 64% to 70% of all American Facebook users think Facebook categorizes their interests accurately

(c) there is a 64% to 70% chance that a randomly chosen American Facebook user's interests are categorized accurately.

(d) there is a 64% to 70% chance that 95% of American Facebook users' interests are categorized accurately.

Ex. Facebook's categorization of user interests

Which of the following is the correct interpretation of this confidence interval? We are 95% confident that...

(a) 64% to 70% of American Facebook users in this sample think Facebook categorizes their interests accurately.

(b) 64% to 70% of all American Facebook users think Facebook categorizes their interests accurately

(c) there is a 64% to 70% chance that a randomly chosen American Facebook user's interests are categorized accurately.

(d) there is a 64% to 70% chance that 95% of American Facebook users' interests are categorized accurately.

Width of an interval

If we want to be more certain that we capture the population parameter, i.e. increase our confidence level, should we use a wider interval or a smaller interval?

Width of an interval

If we want to be more certain that we capture the population parameter, i.e. increase our confidence level, should we use a wider interval or a smaller interval?

A wider interval.

Width of an interval

If we want to be more certain that we capture the population parameter, i.e. increase our confidence level, should we use a wider interval or a smaller interval?

A wider interval.

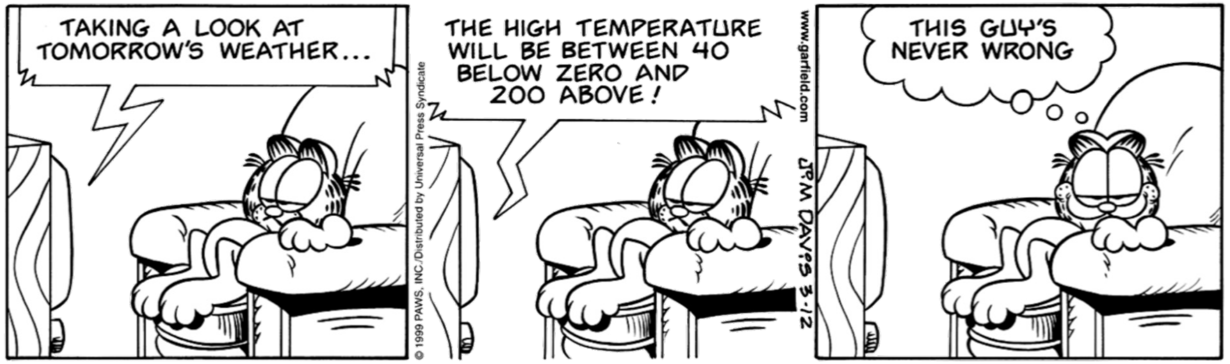
Can you see any drawbacks to using a wider interval?

Width of an interval

If we want to be more certain that we capture the population parameter, i.e. increase our confidence level, should we use a wider interval or a smaller interval?

A wider interval.

Can you see any drawbacks to using a wider interval?



If the interval is too wide it may not be very informative.

Changing the confidence level

point estimate $\pm$ z^* *x SE*

- In a confidence interval, z^* *x SE* is called the *margin of error*, and for a given sample, the margin of error changes as the confidence level changes.
- In order to change the confidence level we need to adjust z^* in the above formula.
- Commonly used confidence levels in practice are 90%, 95%, 98%, and 99%.
- For a 95% confidence interval, z^* is about 2.
- However, using the standard normal (z) distribution, it is possible to find the appropriate z^* for any confidence level.

Interpreting confidence intervals

Confidence intervals are ...

- always about the population
- are not probability statements
- only about population parameters, not individual observations
- only reliable if the sample statistic they're based on is an unbiased estimator of the population parameter