Confidence Intervals

- Need to estimate population parameter
- Only have sample, so use sample percentage to estimate SE ("bootstrap" method)
- Once we get our estimate of SE, we know that the sample percentage will be within 2 SEs of the population percentage 95% of the time.
- So the sample percentage ± 2 SE is called a 95% confidence interval.
- 95% is NOT the probability that the pop. percentage will be in the interval.

Confidence Interval for Averages

- **Central Limit Theorem**

  The probability histogram for the sum (or average) of a large number of independent draws can be approximated by the normal curve.

- So, once we know the box average and box SD, we can compute probabilities for the sum of a large number of draws. But what if we *don't* know these values?

What do we know about the sum?

- EV of sum = (number of draws) x (box average)
- SE of sum = (√number of draws) x (box SD)
- So, SE of sum goes up as the square root of the number of draws
- What about the average?
- SE of average = box SD/(√number of draws)

  So, the SE of the average decreases as √n.

Example

- A box of tickets averages out to 75, and the SD is 10. 100 draws at random are made with replacement from this box.

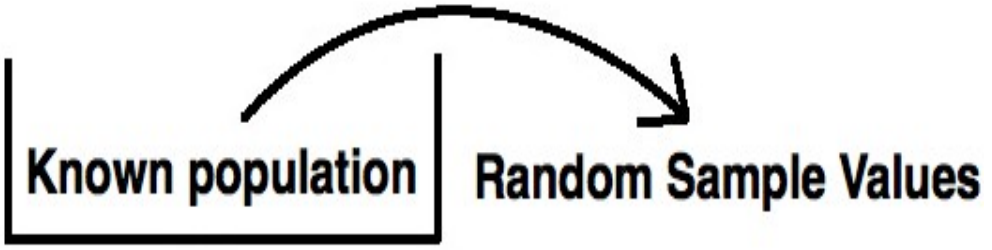
  The sum of the draws will be around _____, give or take _____.

  The average of the draws will be around _____, give or take _____.

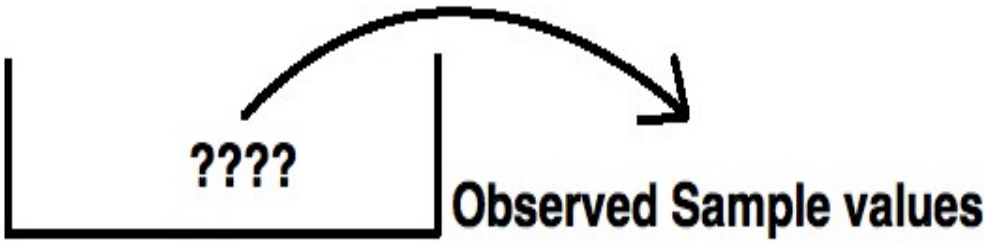
  The (approx.) chance that the average will be in the range 65 to 85 is _____.

  The (approx.) chance that the average will be in the range 74 to 76 is _____.

So we know how to solve this problem:



How do we solve *this* problem?

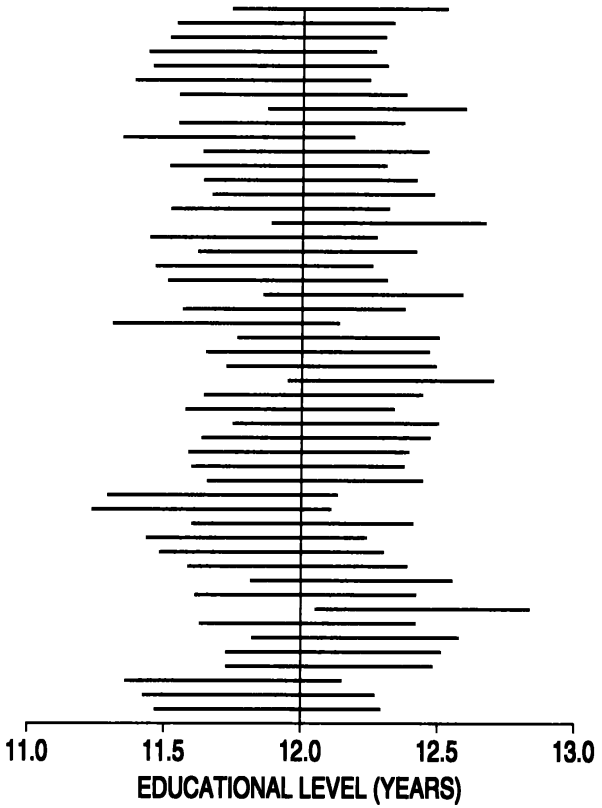


Example

As part of an opinion survey, a simple random sample of 400 people age 25 and over is taken in a certain town in Appalachia. Average educational level from the sample is about 11.6 years. The SD of the sample is 4.1 years. Find a 95% confidence interval for the average educational level of all persons age 25 and over in this town.

A 95% CI would be _____±_____.

Example



- Why do these intervals have different centers?
- Why do they have different lengths?
- How many of them would you expect to contain 12?
- How many do?

Example

100 teenagers are interviewed as they leave a shopping mall and are asked how much money they spent that day. Can a confidence interval for the average expenditure of a teenager in a mall be derived from the information that is collected?

A sample of size 1000 is taken from a population of size 100,000 adults and you are given the following information:
(1) The percent of the sample who own cars
(2) The average weekly earnings of the people in the sample.
Can you find the SE for the population parameter for (1)? For (2)?

Things to avoid

- Don't suggest that the parameter is a variable.

  (“There is a 95% chance that the population percentage is between 42% and 67%.”)
- Don't claim that other samples will agree with yours.
- Don't be *certain* about the parameter (just quite sure).