Confidence Intervals

- what is a confidence interval?
- finding and interpreting confidence intervals

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

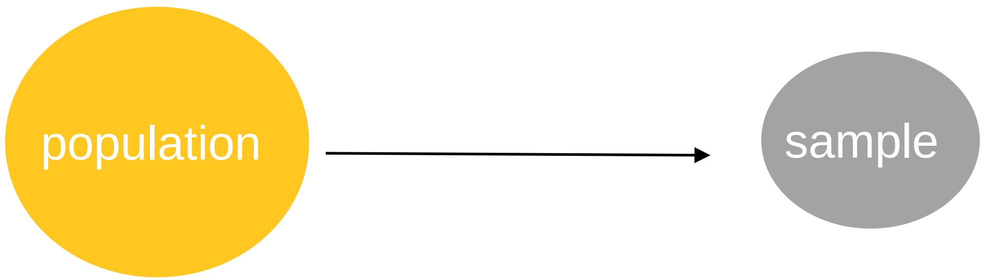
Using a confidence interval is like fishing with a net.

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.

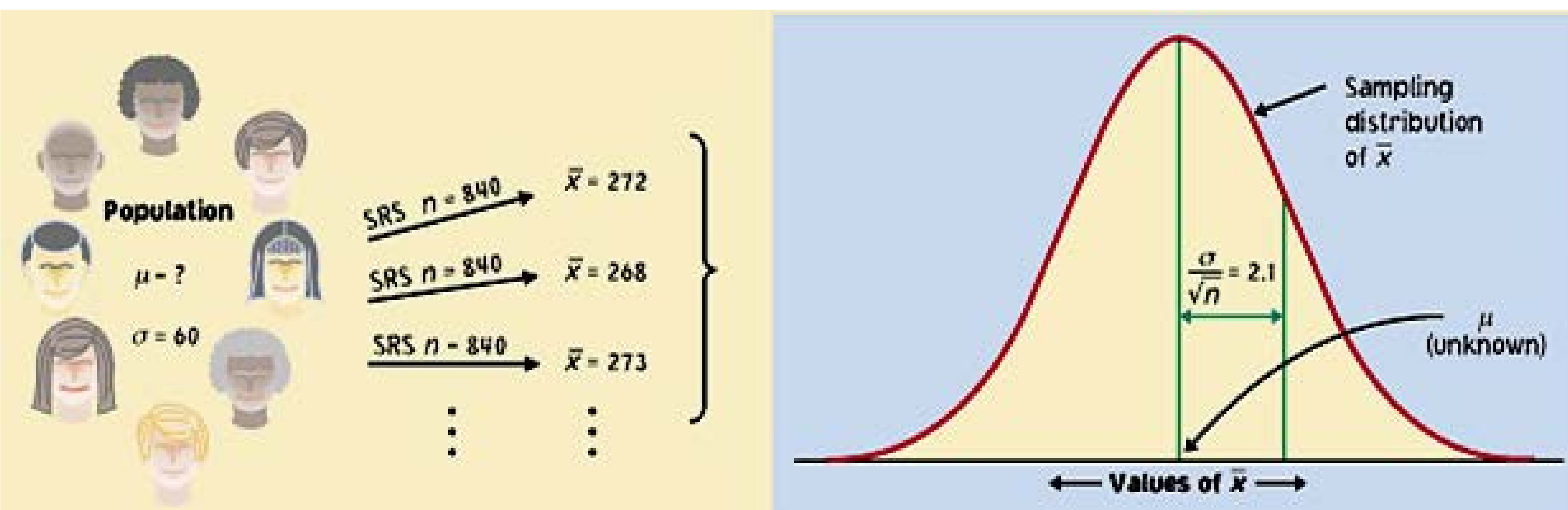
Sampling distribution and CLT



μ : population mean ← $\bar{x}$: sample mean

Although the sample mean $\bar{x}$ is a unique number for any particular sample, *if you pick a different sample you will probably get a different sample mean.*

In fact, you could get many different values for the sample mean, and virtually none of them would actually equal the true population mean, μ.



confidence interval for a population mean

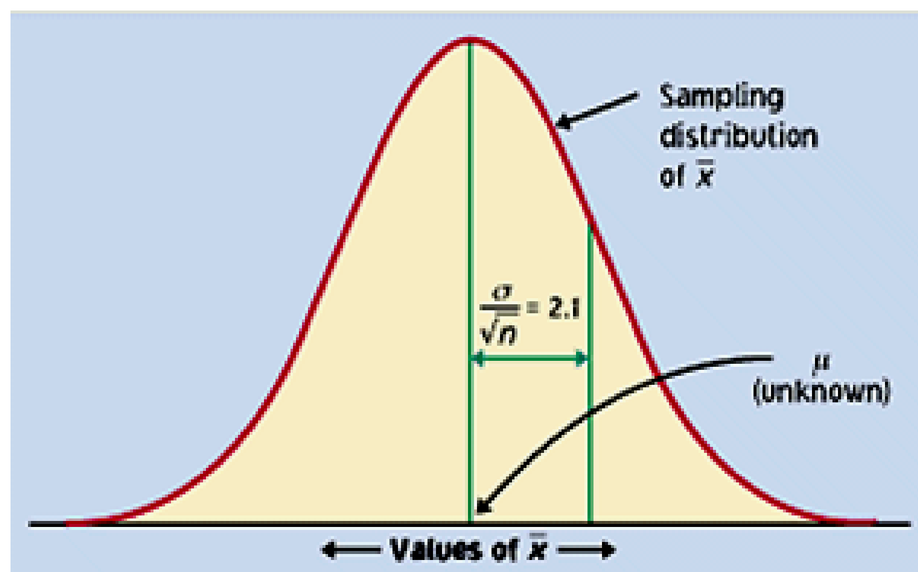
population → sample

μ : population mean ← $\bar{x}$: sample mean

point estimate : $\bar{x}$

Central Limit Theorem (CLT):

$$\bar{x} \sim N\left(mean = \mu,\ SE = \frac{\sigma}{\sqrt{n}}\right)$$

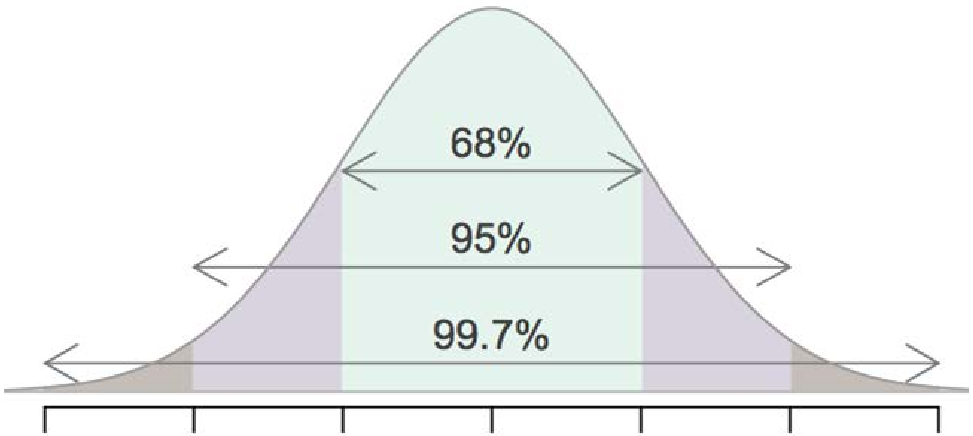


finding the confidence interval

Central Limit Theorem (CLT):

$\bar{x} \sim N(mean= \mu, SE= \frac{\sigma}{\sqrt{n}})$

For 95% of random samples, the unknown true population mean is going to be within 2 standard errors of that sample's mean.



95% CI $\approx \bar{x} \pm 2$ SE

CI stands for confidence interval

SE stands for standard error

In above formula, the 2 standard errors, is actually called the **margin of error (ME)**.

The **margin of error** for a 95% confidence interval is roughly 2 standard error.

If population distribution is normal, 95% CI for a population mean : $\bar{x} \pm 1.96$SE

If σ is unknown, use *s* (sample standard deviation) for SE,

conditions for the confidence interval :

Independence: sampled observations must be independent

- random sample/assignment
- if sampling without replacement, $n<10\%$ of population

Sample size/skew:

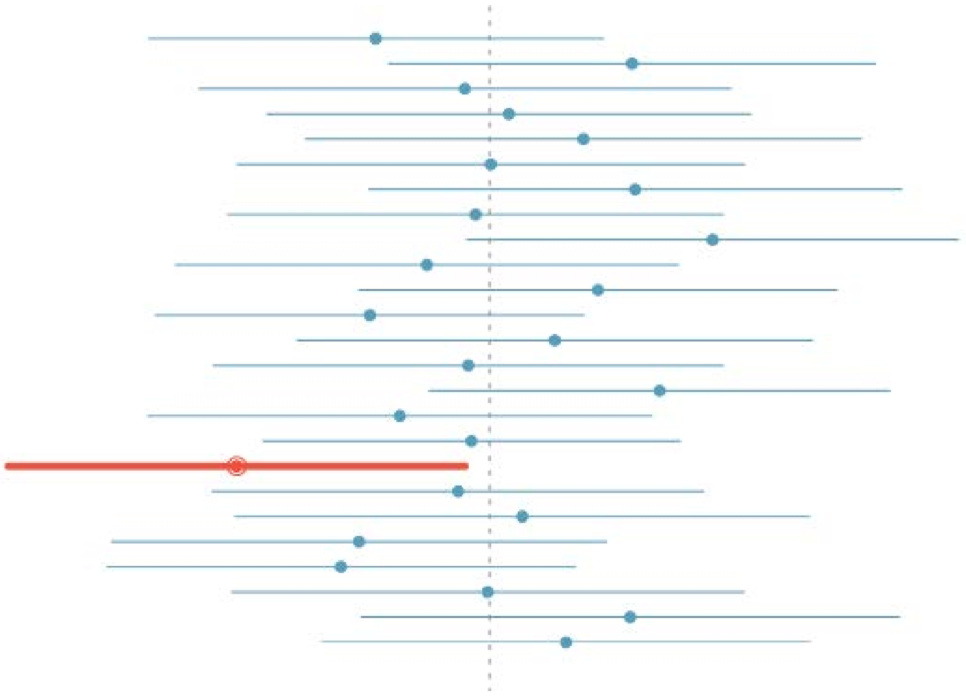
Either the population distribution is normal, or

if the population distribution is skewed, the sample size is large

(*rule of thumb:* $n>30$)

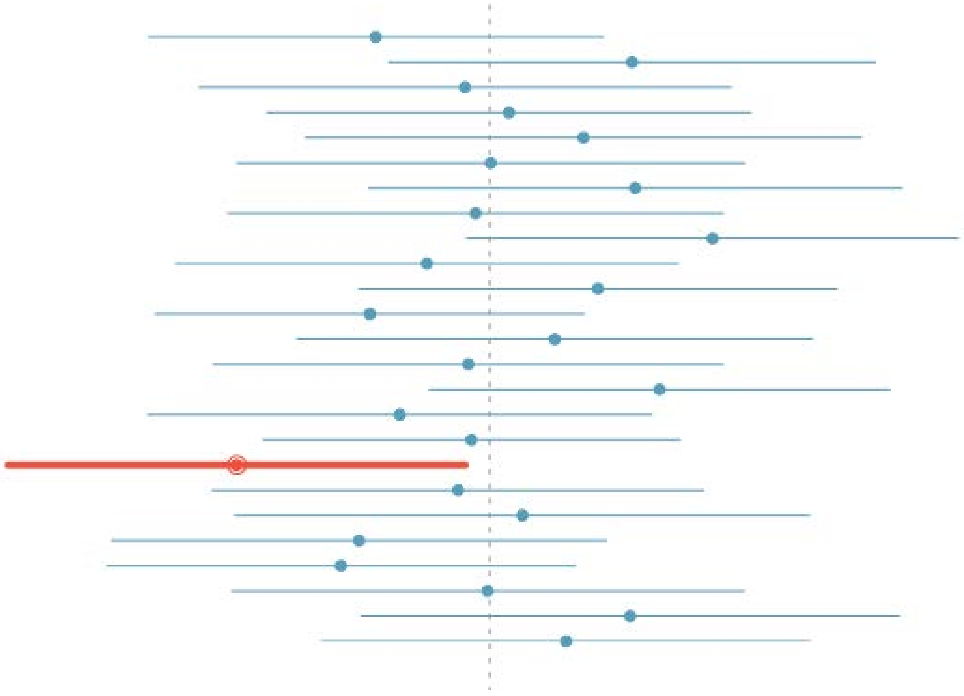
What does 95% confident mean?

- 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 population mean, and one does not.



Interpreting confidence intervals

- A confidence level describes what would happen if we took many samples and built a confidence interval from each sample: the confidence level describes the fraction of those intervals that captures the population parameter



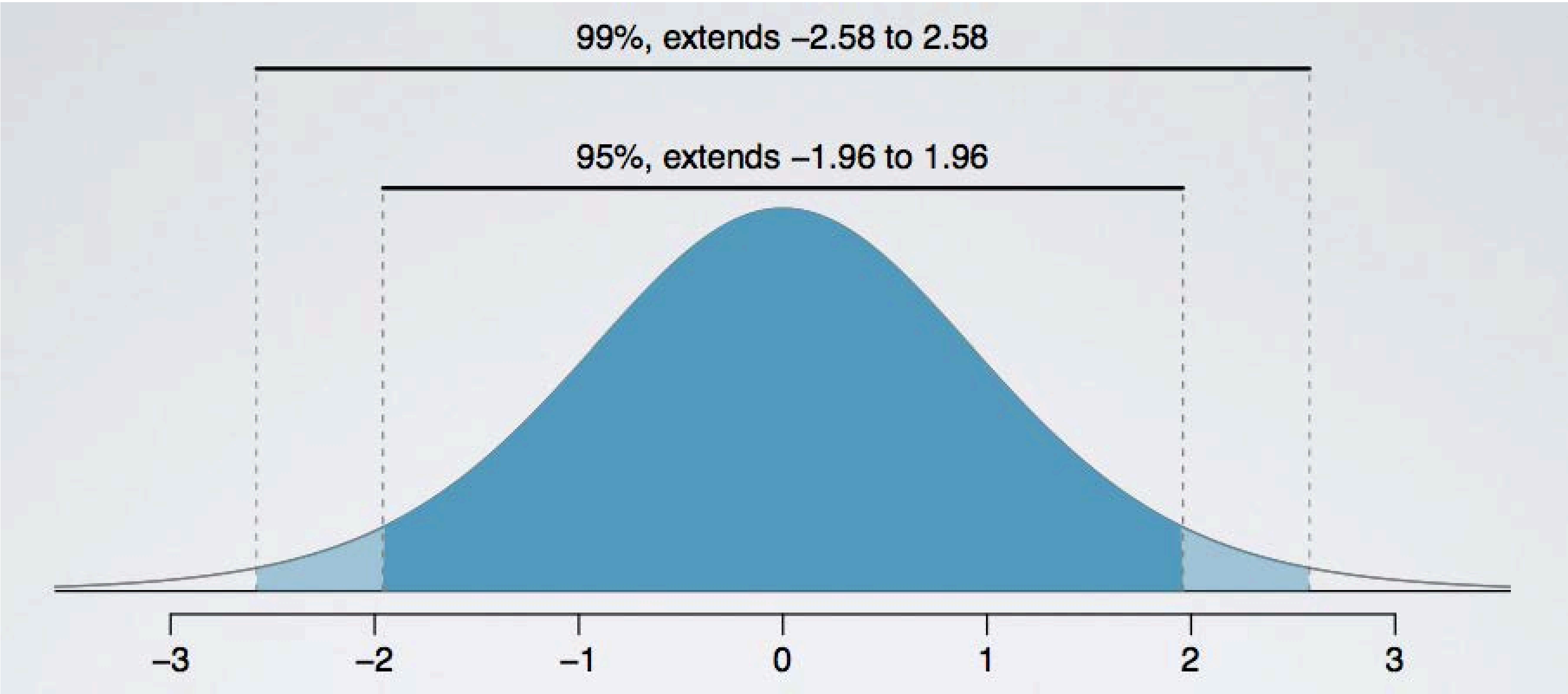
- While it is useful to think of confidence level as a probability there are technical reasons why it is not a probability (So never describe it as such)

changing the confidence level

if a point estimate closely follows a normal distribution with standard error SE, then a confidence interval for the population parameter is

$$\boxed{\text{point estimate} \pm z^\star \times SE}$$

- In a confidence interval, $z^\star \times SE$ 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 $z^\star$ in the above formula. Commonly used confidence levels in practice are 90%, 95% and 99%.
- For a 95% confidence interval, $z^\star = 1.96$.
- However, using the standard normal (z) distribution, it is possible to find the appropriate $z^\star$ for any confidence level.



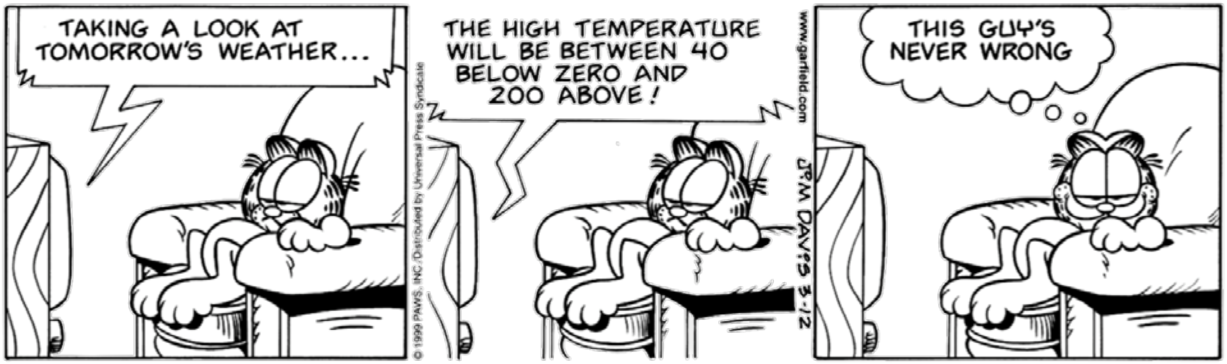
99%, extends −2.58 to 2.58
95%, extends −1.96 to 1.96
−3
−2
−1
0
1
2
3

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.

Practice

Which of the below Z scores is the appropriate $z\star$ when calculating a 98% confidence interval?

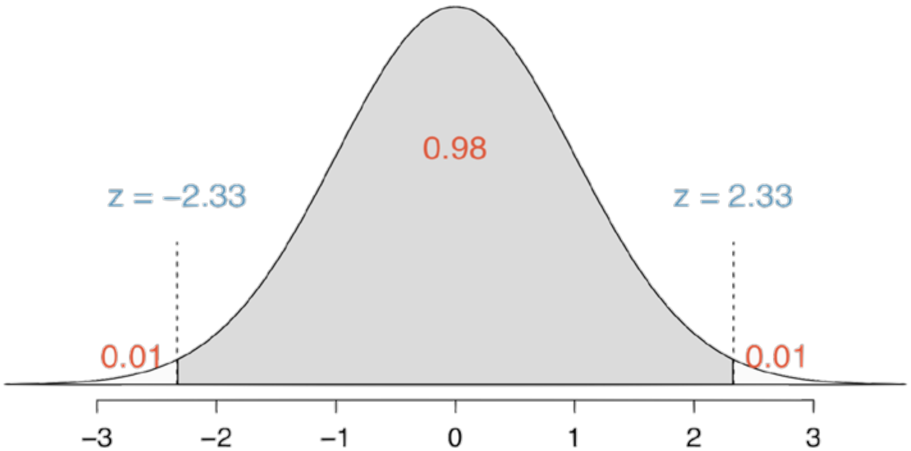
(a) $Z = 2.05$

(b) $Z = 1.96$

(c) $Z = 2.33$

(d) $Z = -2.33$

(e) $Z = -1.65$




```
> qnorm(.01)
[1] -2.326348
```

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831

Practice

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.

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

The approximate 95% confidence interval is defined as

point estimate ± 2 x SE

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

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

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

$\rightarrow \quad (2.7, 3.7)$

Practice

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

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.

Summary

- The concept of a confidence interval.

  Analogy: fishing with a net instead of a spear

- Constructing a 95% confidence interval. :

  point estimate $\pm$ 1.96$\times$SE or point estimate $\pm$ 2$\times$SE

- Changing the confidence level.

  Adjust “1.96” in the confidence interval formula

- Interpreting confidence intervals.

  “We are 95% confident that…”

  Never use the word “probability”.

  Remember that the interval only tries to capture the population parameter

Assignment

In 2013, the Pew Research Foundation reported that 45% of U.S. adults report that “they live with one or more chronic conditions”. However, this value was based on a sample, so it may not be a perfect estimate for the population parameter of interest on its own. The study reported a standard error of about 1.2%, and a normal model may reasonably be used in this setting.

Problem 1. Create a 95% confidence interval for the proportion of U.S. adults who live with one or more chronic conditions. Also interpret the confidence interval in the context of the study

Assignment

Problem 2. Identify each of the following statements as true or false. Provide an explanation to justify each of your answers.

a. We can say with certainty that this confidence interval contains the true percentage of U.S. adults who suffer from a chronic illness.
b. If we repeated this study 1,000 times and constructed a 95% confidence interval for each study, then approximately 950 of those confidence intervals would contain the true fraction of U.S. adults who suffer from chronic illnesses.
c. The poll provides statistically significant evidence (at the = 0:05 level) that the percentage of U.S. adults who suffer from chronic illnesses is below 50%.
d. Since the standard error is 1.2%, only 1.2% of people in the study communicated uncertainty about their answer.