IMS 11 (ish): Introduction to Hypothesis Testing

Libraries For Code Today

```r
library(openintro)
library(tidyverse)
library(infer)
```

Recall the First day of Class

I recently heard a claim that:

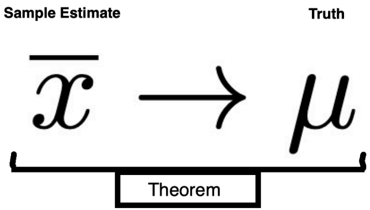
The average height of a student at Smith is 62.5 inches.

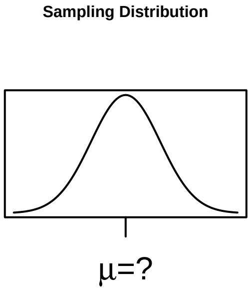
Recall the First day of Class

- We then collected a sample.
- We calculated a estimate for the mean height of students at Smith.
- We discussed as a class if the value under then the claim *seems reasonable* based on our estimate.
- How can we determine whether a claim *seems reasonable* in any situation?

Fundamental Theorem of Statistics

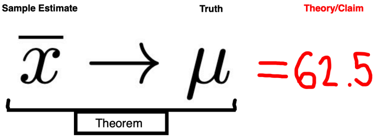
- Later we learned about the FTS
  - Estimates for the mean (and proportion) will eventually converge to the truth.
  - The sampling distribution of the mean (and proportion) are normally distributed.
- We know what the the shape/spread of the distribution is, but the theorem does not explicitly tell us the center.



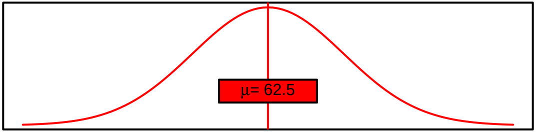


If the claim was true...

Our logic on the first day was correct: if the claim was true, then the sample estimate, the truth, and the claim should be the same.

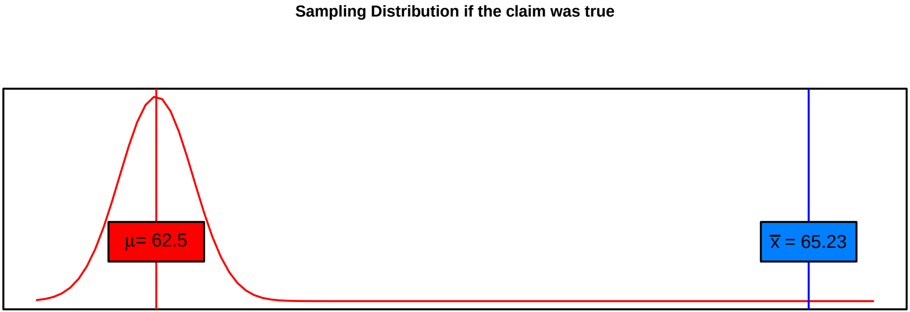


What the sampling distribution looks like if the claim was true:



If the claim was true...

The data we observed: $\overline{x} = 65.23, n = 529, \sigma_x = 3.55$.



- Do the claim and the estimate seem reasonably similar?
- What are all the reasons that might explain why the sample estimate and the claim are *not* the same?

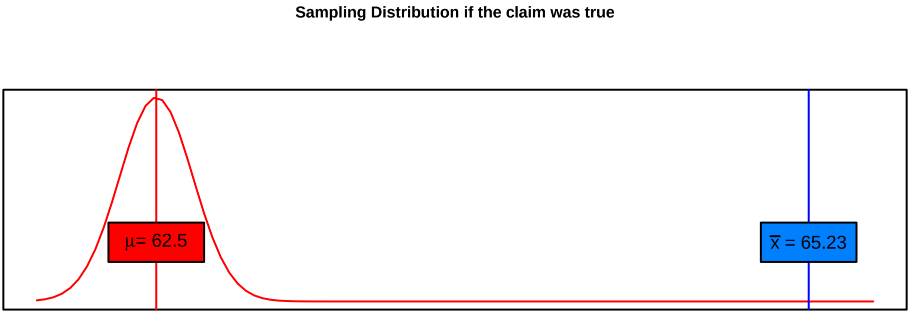
Why the estimate and claim might differ...

- The sample was not representative.
- Random variation for $\overline{x}$.
- The claim was false.

What steps can we take to minimize these occurances?

When should we reject the claim?

Suppose we took all the precautions that we brainstormed, and still observed the following data.



What is the most plausible explanation for why the claim and the estimated mean are so dissimilar?

Introduction to Hypothesis Testing

- We will formalize the process we just did using a **significance test** to analyze evidence in favor of the claim.
- We already learned some of the tools used in significance testing: confidence intervals. We will add some more tools though.
- The previous example was based on real data that we collected.
- A new situation to consider:

Example

Since the Fair Labor Standards Act was passed in 1938, the standard work week in the United States has been 40 hours. In recent years, the standard work week has fallen to less than 40 hours in most of Western Europe and in Australia. But many believe that the work-oriented culture in the United States has resulted in pressure among workers to put in longer hours than the 40-hour standard. We will focus only on investigating men for simplicity.

Introduction to Hypothesis Testing

Definition

In statistics, a **hypothesis** is a statement about a population, usually claiming that a population parameter takes a particular numerical value or falls in a certain range of values.

- The main goal of many research studies is to check whether the data support certain statements or predictions, i.e. a **hypotheses**.
- A **significance test** is a method of using data to summarize the evidence about a hypothesis.

Introduction to Hypothesis Testing

A significance test about a hypothesis has five steps.

1) Frame the research question in terms of hypotheses.
2) Collect data (check conditions)
3) Model the randomness that would occur if the null hypothesis was true.
4) Analyze the data.
5) Form a conclusion.

Step 1: Frame the research question in terms of hypotheses

Each significance test has two hypotheses:

- The null hypothesis (H_0) is a statement that the parameter takes a particular value. Usually represents a skeptical perspective or a perspective of no relationship between the variables.
- The alternative hypothesis (H_A) states that the parameter falls in some alternative range of values. Usually represents what the researcher hopes to show.

Example

- $\mu=$ mean hours worked per week in the USA by men
- H_0: $\mu = 40$
- H_A: $\mu \neq 40$

Step 1: Frame the research question in terms of hypotheses

- We can form hypothesis tests in one of three ways.
- In each forms H_0 must have one of the following signs: $=$, $\geq$, $\leq$
- The format we pick depends on the question asked: Is the average hours worked per week by men in the USA 40 hrs? Is it under 40 hrs? Is it over 40 hrs?

Two-Sided	*One-sided*	
$H_0 : \mu = 40$	$H_0 : \mu \leq 40$	$H_0 : \mu \geq 40$
$H_A : \mu \neq 40$	$H_A : \mu > 40$	$H_A : \mu < 40$

Step 2: Collect data.

If a research question can be formed into two hypotheses, we can collect data to run a hypothesis test.

- Make sure the data are *independent*.
- Make sure to collect *enough* data.
- Make sure data *represents* the population being studied.
- Consider an observational study vs experiment.

Example

For those who were working in 2012, the General Social Survey asked, 'How many hours did you work last week?'. For the 100 men included in the survey, the mean was 43.5 hours with a standard deviation of 15.3 hours.

Step 3: Model the randomness that would occur if the null hypothesis was true.

To summarize the evidence against the null hypothesis, we need to assess whether the $\overline{x}$ is unusual under H_0.

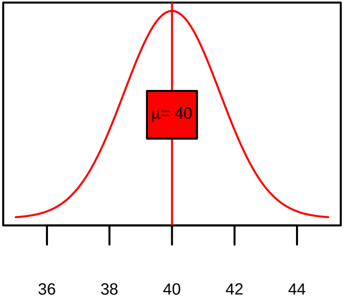
- We presume that H_0 is true, since the burden of proof is on the alternative, H_A.
- We consider the sorts of values we'd expect to get for the $\overline{x}$, according to its sampling distribution under H_0.
- If $\overline{x}$ falls well out in a tail of the sampling distribution under H_0, it is far from what H_0 predicts. If H_0 were true, such a value would be unusual.

Step 3: Model the randomness that would occur if the null hypothesis was true.

Example:

$$\overline{x} \sim N\left(\mu = 40, \sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} \approx 1.53\right)$$

Sampling Distribution Expected under $H_0 : \mu = 40$



Step 4: Analyze the data.

We choose one of the tools below to help us determine when $H_0 : \mu = 40$ and $\overline{x}$ are dissimilar.

- Confidence intervals
- P-values (coming soon)
- Test statistics (coming soon)

The tools above summarize the evidence. They describes how unusual the data would be if H_0 were true.

Step 4: Analyze the data.

Calculate a $(1-\alpha)100\%$ confidence interval.

Example:

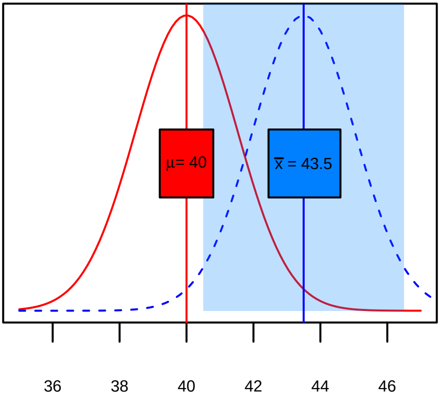
```
lower <- 43.5 - 1.96 *1.53
upper <- 43.5 + 1.96 *1.53
c(lower, upper)
```

```
[1] 40.5012 46.4988
```

The 95% confidence interval for hours worked per week for men in the USA is (40.5012, 46.4988).

Step 4: Analyze the data.

Example:



Step 4: Analyze the data.

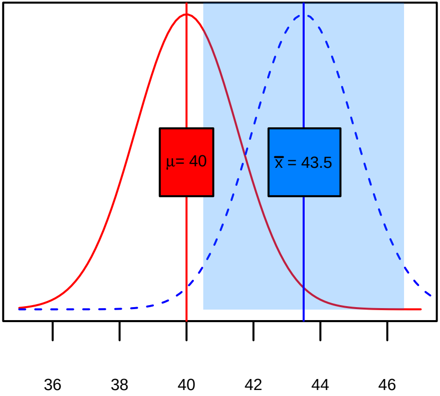
Example:

- Red solid line: sampling distribution for $\overline{x}$ under H_0. The range of values (and associated probabilities) that we would expect if $\mu = 40$.
- Blue region: the bounds for our confidence interval. This is the range of values we would expect μ to be based on our data. (i.e. where we pinned the tail)

Step 5: Form a conclusion.

- We want to check if the claim (μ under H_0) is similar to the sample mean we observed $\overline{x}$.
- Confidence intervals, p-values, and test statistics, give us an impartial way of determining what is "similar enough".
- Determine whether the data provide evidence against the null hypothesis.
- Be sure to write the conclusion in plain language so casual readers can understand the results.
- Do a gut check!

Step 5: Form a conclusion.



- Does it seem likely that we would observe a sample mean of $\overline{x} = 43.5$ (or something near it) if true mean was $\mu = 40$?
- What do you conclude about the average hours worked per week for men in the USA?

Step 5: Form a conclusion.

To make a conclusion based on confidence intervals we check to see if the hypothesized mean is a value our interval ‘caught’.

- If the hypothesized μ is in our interval, then we have evidence to support the claim.
- If the hypothesized μ is not in the interval, then we have evidence against the claim.

Typical values for α are 0.05, 0.10, 0.01.

Step 5: Form a conclusion

Hypothesis test conclusions should have three components:

- What did you check?
- Are you in favor or against H_0? (always stay in terms of H_0!)
- Describe your conclusion in context, avoiding technical jargon and definitive language.

Example

Our 95% confidence interval (40.5012, 46.4988) does not contain 40. Thus, we evidence against H_0. We conclude that the average hours worked per week for adult males in the USA is likely not 40 hrs.

Step 5: Form a conclusion.

Note: we have not **proved** anything. Recall our three reasons why $H_0 : \mu = 40$ and $\overline{x}$ could be dissimilar:

- Sample is not representative.
- Random chance.
- H_0 is false.

The first option is unlikely because we are responsible statisticians. We have found that random chance (second option) is very unlikely (our confidence level is usually high). So it seems that the last option is the most plausible.

Introduction to Hypothesis Testing

- A significance test analyzes the strength of the evidence against the null hypothesis H_0.
- We start by presuming that H_0 is true.
- The approach used in hypothesis testing is called a *proof by contradiction*.
- To convince ourselves that H_A is true, we must show that data contradict H_0.
- The approach can also be thought of as a *process of elimination*; however, we never fully eliminated any option. We just conclude that the other options were very unlikely.

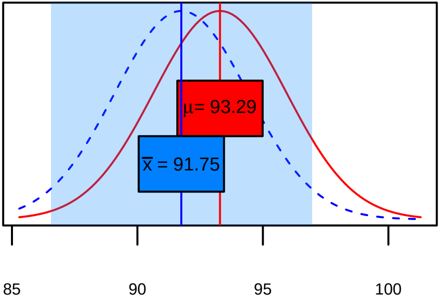
Practice

The average time for all runners who finished the Cherry Blossom Race in 2006 was 93.29 minutes (93 minutes and about 17 seconds). The sample mean and sample standard deviation of the sample of 100 runners from the 2017 Cherry Blossom Race are 91.75 and 26.55 minutes, respectively. We want to determine whether the pace for runners in 2017 in this race has changed since 2006.

Follow the steps we just learned:

1. Frame the research question in terms of hypotheses.
2. ~~Collect data~~
   - Already done :D
3. Model the randomness that would occur if the null hypothesis was true.
   - What is the distribution under H_0?
4. Analyze the data.
   - Calculate the 95% confidence interval.
5. Form a conclusion.
   - State in context.

Practice



Hypotheses Testing with the Bootstrap

- We can follow all the same steps before to do hypotheses testing.
- This time we use a bootstrap confidence interval instead of a traditional confidence interval.
- The conclusion should be the same (almost always).

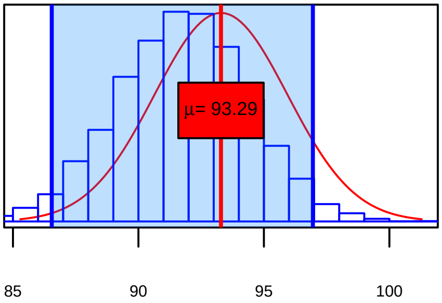
```r
# Generating the data
set.seed(62)
index <- sample(1:nrow(cherryblossom::run17), 100, F)
cherry17 <- cherryblossom::run17[index, ]
cherry17$net_sec <- cherry17$net_sec/60

# Bootstrap means
bootstrap_means <- cherry17 %>%
                  rep_sample_n(size=100, reps=10000, replace =TRUE) %>%
                  summarize(boot_mean = mean(net_sec))

# Bound CI Bounds
lower <- quantile(bootstrap_means$boot_mean, 0.025)
upper <- quantile(bootstrap_means$boot_mean, 0.975)
c(lower, upper)
```

```
    2.5%    97.5%
86.58407 96.81121
```

Hypotheses Testing with the Bootstrap



Practice

[IMS 11.6]: Write the null and alternative hypotheses in words and using symbols for each of the following situations.

a) Since 2008, chain restaurants in California have been required to display calorie counts of each menu item. Prior to menus displaying calorie counts, the average calorie intake of diners at a restaurant was 1100 calories. After calorie counts started to be displayed on menus, a nutritionist collected data on the number of calories consumed at this restaurant from a random sample of diners. Do these data provide convincing evidence of a difference in the average calorie intake of a diners at this restaurant?

b) Based on the performance of those who took the GRE exam between July 1, 2004 and June 30, 2007, the average Verbal Reasoning score was calculated to be 462. In 2021 the average verbal score was slightly higher. Do these data provide convincing evidence that the average GRE Verbal Reasoning score has changed since 2021?

Practice

Many students brag that they have more than 100 friends on Facebook. Suppose a sample was taken of 50 students, which had a mean of 131.25 friends with a standard deviation of 67.66. We want to determine if average number of friends students have on Facebook is 100. Follow the steps below.

1. Frame the research question in terms of hypotheses.
2. ~~Collect data.~~
3. Model the randomness that would occur if the null hypothesis was true.
4. Analyze the data.
    - Calculate a 99% CI
5. Form a conclusion.