

# DATA SCIENCE FOR ECONOMISTS

ECON 220 LAB

Jafet Baca-Obando

Week 11, Hypothesis Testing – 11/07/2025

# Outline

- 01 One-sample  $t$ -test
- 02 Two-sample  $t$ -test
- 03 Two-sample proportion test
- 04 Chi-squared test

# One-sample $t$ -test

- This test checks if the **mean (average) of a single group** is equal to a specific, known value.
- It's used to test a claim about a population average.

- Example from the code:
- Claim: The average work week is 40 hours
- Null hypothesis: The average is 40,  $H_0: \mu = 40$
- Alternative hypothesis: The average is NOT 40,  $H_a: \mu \neq 40$

# One-sample $t$ -test

- What is our decision rule?
  - If the  $p$ -value is below a certain **significance** threshold (for example, 0.05), then we reject the null hypothesis  $H_0$ .
  - In other words, we reject the null hypothesis when the  $p$ -value is very low (or the corresponding statistic is very high).

# Two-sample $t$ -test

- This test compares the **means (averages)** of two independent groups to see if they are different from each other.
- For example, are the average hours worked for men and women different?

- **Claim:** Average hours are the same for males and females.
- Null:  $H_0: \mu_{male} = \mu_{female}$
- Alternative:  $H_a: \mu_{male} \neq \mu_{female}$

# Two-sample proportions test

- This test compares the **proportions** of two independent groups to see if they are different.
- For example, is the proportion of overworked people the same for males and females?





- **Claim:** Proportions of overworked people are the same for males and females.
- Null:  $H_0: p_{male} = p_{female}$
- Alternative:  $H_a: p_{male} \neq p_{female}$

# Chi-squared test

- This test is used to compare **proportions across more than two groups**. It checks if there is a significant association between two categorical variables.
- Question: Is the proportion of "overworked" people different across the 4 U.S. regions?

- **Claim:** Proportions of overworked people are the same across regions.
- Null:  $H_0: p_i = p_j$
- Alternative:  $H_a: p_i \neq p_j$  for at least one region  $j$ .

# Summary of tests

One-sample $t$ -test		<ul style="list-style-type: none"><li>• Compares one group's mean to a known value. (e.g., Avg. hours vs. 40)</li></ul>
Two-sample $t$ -test		<ul style="list-style-type: none"><li>• Compares two groups' means to each other. (e.g., Male vs. Female avg.)</li></ul>
Two-sample proportions test		<ul style="list-style-type: none"><li>• Compares two groups' proportions. (e.g., Male vs. Female "overworked" %)</li></ul>
Chi-squared		<ul style="list-style-type: none"><li>• Compares proportions across *more than two* groups. (e.g., "Overworked" % by Region)</li></ul>



# To-do list

- **Complete Data Exercise 7**
  - Upload Jupyter notebook (.ipynb file) and HTML file on **November 9**
- **Complete Data Exercise 8**
  - Upload Jupyter notebook (.ipynb file) and HTML file on **November 16**