# Hypothesis Testing

Jeffrey Woo

School of Data Science, University of Virginia

## Motivation

The general approach to hypothesis testing is the following: we perform probability calculations to distinguish patterns seen in data between those that are due to **chance** and those that **reflect a real feature** of the phenomenon under study.

1 Hypotheses

2 Evaluating Evidence
- P-Value Approach
- Critical Value Approach

## Example

You are in charge of quality control in your food company. You randomly sample 40 packs of cherry tomatoes, each labeled 1/2 lb. (227 g), and find their average weight is 226.5 g. Obviously, we cannot expect boxes filled with whole tomatoes to all weigh exactly half a pound. Thus,

- is the weight in our sample due to chance?
- is the weight in our sample evidence the machine that sorts the tomatoes needs revision?

# Stating Hypotheses

Hypothesis testing uses sample data to decide on the validity of a hypothesis. A **hypothesis** is an assumption or a theory about the characteristics of one or more variables in one or more populations.

- What you want to know: Does the calibrating machine that sorts cherry tomatoes into packs need revision?
- The same question reframed statistically: Is the population mean $\mu$ for the distribution of weights of cherry tomato packages different from 227 g (i.e., half a pound)?

# Stating Hypotheses

The statement being tested in a test of significance is called the **null hypothesis,** $H_0$. The test of significance is designed to assess the strength of the evidence against the null hypothesis. The null hypothesis is usually a statement of "no effect" or "no difference."

The **alternative hypothesis,** $H_a$**,** is the statement we suspect is true instead of the null hypothesis.

- $H_0 : \mu = 227g$
- $H_a : \mu \neq 227g$

## One-sided and Two-sided Tests

- A two-sided test of the population mean has the following hypotheses:
    - $H_0 : \mu =$ specific number ($\mu_0$)
    - $H_a : \mu \neq$ specific number ($\mu_0$)
- A one-sided test of the population mean has the following hypotheses:
    - $H_0 : \mu =$ specific number ($\mu_0$)
    - $H_a : \mu <$ specific number ($\mu_0$)

  OR

    - $H_0 : \mu =$ specific number ($\mu_0$)
    - $H_a : \mu >$ specific number ($\mu_0$)

# One-sided and Two-sided Tests

What determines the choice of a one-sided versus a two-sided test is what we know about the problem **before** we perform a test of statistical significance.

**Question:** You are in charge of quality control in your food company. You randomly sample 40 packs of cherry tomatoes, each labeled 1/2 lb. (227 g), and find their average weight is 226.5 g. A consumer advocacy group is trying to claim that consumers are being cheated by the food company. What should be the null and alternative hypothesis in this scenario?

1. Hypotheses

2. Evaluating Evidence
   - P-Value Approach
   - Critical Value Approach

# Evaluating Evidence

Next, we evaluate the evidence our data provides **against** the null hypothesis. This evaluation is done by

- assuming the null hypothesis is true
- computing a test statistic to measure how dissimilar our sample is with the null hypothesis
- comparing our test statistic with a benchmark to decide if we have enough evidence against the null hypothesis

We then end by making a relevant conclusion

# Test Statistics

- The test statistic measures how dissimilar our sampled data is with the null hypothesis.
- In a hypothesis test for a mean, the test statistic is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}. \tag{1}$$

- The larger (in magnitude) the test statistic, the more evidence we have against the null hypothesis.

# Significance Level

Before deciding if our test statistic provides enough evidence against the null hypothesis, we first decide on an appropriate **significance level**, $\alpha$. The scientific standard is 0.05, although this value should change based on the context of your problem.

- The significance level, $\alpha$, is the probability of wrongly rejecting the null hypothesis (when the null hypothesis is true).
- The benchmark with which we decide if we have enough evidence against the null hypothesis is based on $\alpha$. There are actually two, equivalent, approaches.

1. Hypotheses
2. Evaluating Evidence
   - P-Value Approach
   - Critical Value Approach

# The p-value Approach

**P-value:** the probability of obtaining your particular random sample result (or more extreme) if the null hypothesis were true.

- A high p-value implies that the random sample result is consistent with $H_0$.
- A small p-value implies that random variation alone is unlikely to account for the difference between $H_0$ and the observation from our random sample. Our sample is inconsistent with $H_0$.
- The smaller the p-value, the stronger the evidence against $H_0$.
- With a small p-value we reject $H_0$, and say that our data support $H_a$.

We reject $H_0$ when the p-value is **less than** the significance level, $\alpha$.

# Distribution of t Statistic

Since the test statistic for testing a mean is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}},$$

our test statistic is compared with a $t_k$ distribution.

# Finding the P-Value

The p-value is represented by the area under the sampling distribution for values at least as extreme, in the direction of $H_a$, as that of our random sample.

# Decision

We compare the p-value with the **significance level**, $\alpha$.

- If the p-value is equal to or less than $\alpha$, then we reject $H_0$. Our data support $H_a$.
- If the p-value is greater than $\alpha$, then we fail to reject $H_0$. Our data do not support $H_a$.

Rejecting $H_0$ is said to be a "statistically significant result". Failing to reject $H_0$ is said be a "statistically insignificant result".

1. Hypotheses
2. Evaluating Evidence
   - P-Value Approach
   - Critical Value Approach

# The Critical Value Approach

**Critical value:** the critical value is the value of the test statistic that results in a p-value equal to the significance level.

- The larger the test statistic, the more evidence we have against $H_0$.
- If the magnitude of the test statistic is larger than the critical value, we reject $H_0$.

# Finding the Critical Value

# Finding the Critical Value

- Find the critical value of 2-sided $t$ test at $\alpha = 0.1$ with 10 df.
- Find the critical value of 1-sided $t$ test at $\alpha = 0.02$ with 55 df.
- Find the critical value of 1-sided $t$ test at $\alpha = 0.01$ with 70 df.

## Summary

- Based on your question of interest, write $H_0$ and $H_a$.
- Evaluate how dissimilar your sample is from $H_0$ by calculating the test statistic.
- Compare your sample with a benchmark. Two equivalent approaches:
    1. Compare p-value with significance level $\alpha$.
    2. Compare your test statistic with the critical value.
- Write relevant conclusion for your analysis.

# Worked Examples

**Question 1:** You are in charge of quality control in your food company. You randomly sample 40 packs of cherry tomatoes, each labeled 1/2 lb. (227 g), and find their average weight is 226.5 g. Is the weight in our sample evidence the machine that sorts the tomatoes needs revision? Suppose the sample standard deviation is 1.5g.

**Question 2:** A consumer advocacy group is trying to claim that consumers are being cheated by the food company. Carry out an appropriate hypothesis test.