STA 103 Lecture 11: Hypothesis Testing (One-sample, Two-sample Inference)

Instructor: Wookyeong Song

Department of Statistics, University of California, Davis

Sep 2nd, 2025

Hypothesis Testing

- When estimating an unknown parameter θ from data, there are generally two types of statements you can make
  - 1) θ is likely within $[a, b]$, (where a and b comes from the data $X_1, X_2, \ldots, X_n$.
  - 2) θ is likely not bigger than b or smaller than a.
- 1) is confidence intervals approach and 2) is hypothesis testing approach.
- The two are effectively equivalent but often 2) can be done more convincingly.

Why Do We Need Hypothesis Testing

- In business and economics, decision-makers often need to evaluate claims or test ideas based on sample data.
  - Has a new marketing campaign **increased** average sales?
  - Is the defect rate of a new supplier **lower** than the current one?
  - Do customers spend **more** time on the redesigned app?
- In such cases, we want to use data from a sample to make a judgment about a population parameter. But because of natural sampling variability, we cannot be 100% certain.
- Hypothesis testing provides a structured statistical framework for making such decisions under uncertainty.

Basic Logic of Hypothesis Testing

- The null hypothesis H_0 represents the status quo (assumed true unless strong evidence suggests otherwise).
- The alternative hypothesis H_a represents what we want to find evidence for.
- We collect sample data and assess how likely it is to observe such data if H_0 were true.
- If the observed data is very unlikely under H_0, we reject H_0 in favor of H_a.
- **Key Idea**: We do **not** prove anything with certainty. Instead, we evaluate whether the sample data provides *statistically significant evidence* against the null hypothesis H_0.

State the Hypothesis Testing

- Two-sided Hypothesis Testing (Checking both increase and decrease):

$$H_0 : \theta = \theta_0 \quad H_a : \theta \neq \theta_0.$$

- One-sided Hypothesis Testing "$>$" (i.e., higher satisfaction, more effective drug):

$$H_0 : \theta \leq \theta_0 \text{ (equivalently, } \theta = \theta_0) \quad H_a : \theta > \theta_0.$$

- One-sided Hypothesis Testing "$<$" (i.e., lower average cost, fewer defects):

$$H_0 : \theta \geq \theta_0 \text{ (equivalently, } \theta = \theta_0) \quad H_a : \theta < \theta_0.$$

- **Remark**: The null H_0 **must** have the equality $=$. If the statement we want to prove is inequality $>$ or $<$, then we put it on **alternative** H_a.

Examples

- Two-sided Hypothesis Testing: A factory claims that its energy drink bottles contain exactly **500 ml** on average.

$$H_0 : \mu = 500 \quad H_a : \mu \neq 500.$$

- One-sided Hypothesis Testing "$>$": Historically, **40%** of users respond to promotional emails. A new email template is introduced to check for an increase in response rate.

$$H_0 : p \leq 0.4 \text{ (equivalently, } p = 0.4) \quad H_a : p > 0.4.$$

- One-sided Hypothesis Testing "$<$": Treatment group has a new drug for cholesterol, and control group has a sugar pill. μ_X is the mean of cholesterol taking the new drug for 6 months, and μ_Y is the cholesterol mean taking the sugar pill for 6 months. We want to check the new drug **reduce** the cholesterol.

$$H_0 : \mu_X - \mu_Y \geq 0 \text{ (equivalently, } \mu_X - \mu_Y = 0) \quad H_a : \mu_X - \mu_Y < 0.$$

Case 1: Hypothesis Testing for One-sample Mean μ

- **Example**: A factory claims that its energy drink bottles contain exactly $\mu = 500$ (ml) on average. They sample $n = 100$ energy drink bottles and it shows: $\bar{X} = 503.2$ (ml) and $s = \hat{\sigma} = 10$ (ml).

- **Question 1**: Find 95% confidence interval (CI).
- **Answer**: Notice that
  - The parameter $\theta = \mu$,
  - The estimate $\hat{\theta} = \bar{X} = 503.2$,
  - The standard error $\mathrm{SE}(\hat{\theta}) = \mathrm{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}} = \frac{10}{\sqrt{100}} = 1.$
  - Z-score $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$, when $\alpha = 0.05$.

- The 95% approximate CI is

$$\mu \in \bar{X} \pm z_{1-\frac{\alpha}{2}} \times \mathrm{SE}(\bar{X})$$

$$\mu \in \bar{X} \pm z_{0.975} \times \frac{s}{\sqrt{n}}$$

$$= 503.2 \pm 1.96 \times 1 = [501.24, 505.16].$$

- $\mu = 500$ is not included in the CI, so the claim that its energy drink bottles contain exactly $\mu = 500$ could be wrong!

Case 1: Hypothesis Testing for One-sample Mean μ

- For these approximate CI, we estimate population σ by sample estimate s:

$$\mathrm{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}} = \frac{10}{\sqrt{100}} = 1.$$

- A somewhat more rigorous argument can be made by a Hypothesis Testing (HT).

- **Question 2**: Make a statistical decision

$$H_0 : \mu = 500 \quad H_a : \mu \neq 500$$

  with significance level $\alpha = 0.05$

- Note: Common choice of significance level $\alpha = 0.05$ or 0.01, which corresponds to 95%, 99% confidence interval. The significance level α is the probability of allowing error rejecting H_0 when the null H_0 is true.

Case 1: Hypothesis Testing for One-sample Mean μ

- Note the Z-score of the point estimator $\bar{X}$ is

$$Z = \frac{\bar{X} - E(\bar{X})}{\mathrm{sd}(\bar{X})} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

- If we are assuming the null hypothesis H_0, i.e., $\mu = 500$, then

$$\begin{aligned} Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - 500}{\sigma/\sqrt{n}} = \frac{503.2 - 500}{\sigma/\sqrt{n}} \\ \approx \frac{503.2 - 500}{10/\sqrt{100}} = 3.2. \end{aligned}$$

- Note the probability of observing more extreme value than 3.2, $P(|Z| \geq 3.2) = 2P(Z \geq 3.2) = 2 \times 0.0007 = 0.0014$, this is quite rare and likely due to an incorrect null assumption $H_0 : \mu = 500$.

Case 1: Hypothesis Testing for One-sample Mean μ

- The direction of observing extreme values follows the alternative hypothesis, here $H_a : \mu \neq 500$, then the extreme direction should be two sided. $P(|Z| \geq 3.2) = 0.0014$ is called p-value for testing the null hypothesis H_0. Formal decision-making should be based on comparing p-value 0.0014 and significance level $\alpha = 0.05$.

- If p-value is **less** than significance level, we reject the null H_0, meaning that there is significant evidence that H_0 is not true. Then, we take the alternative H_a, instead.

- If p-value is **greater** than significance level, we do not reject the null H_0, meaning that there is no significance evidence that H_0 is not true. Then we just keep the null hypothesis H_0.

Relationship between CI and HT

- Results from the CI and HT are closely related.
- A confidence interval asks: "What range of values are plausible for the parameter?"
- A hypothesis test asks: "Is there enough evidence to reject a specific value?"
- For a **two-tailed** test at $100(1-\alpha)\%$ confidence (α significance), you can use a CI to perform the test:
  - If the null H_0 is **not inside** the CI, then **reject** H_0.
  - If the null H_0 is **inside** the CI, then **fail to reject** H_0, take H_a instead.
- However, CI is about estimation, HT is about decision-making.

Case 1: Hypothesis Testing for One-sample Mean μ

- **Question 3 (Add'l)**: If we are assuming

$$H_0 : \mu \leq 500, \quad H_a : \mu > 500,$$

make a statistical decision with the same scenario, significance level $\alpha = 0.01$.

- **Answer**: if we are assuming $\mu \leq 500$ then the least rare this could be is

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \geq \frac{\bar{X} - 500}{\sigma/\sqrt{n}} = \frac{503.2 - 500}{\sigma/\sqrt{n}}$$
$$\approx \frac{503.2 - 500}{10/\sqrt{100}} = 3.2.$$

- Then the p-value is

$$P(Z \geq 3.2) = 0.0007,$$

this is quite rare and likely due to an incorrect H_0. Formal decision making is that p-value 0.007 is lower than significance level $\alpha = 0.01$. Thus, we reject the null H_0 and take the alternative H_a which is energy drink bottles contain more than $H_a : \mu > 500$.

Case 2: Hypothesis Testing for One-sample Proportion p

- Historically, 40% of users respond to promotional emails. A new email template is introduced to check for an increase in response rate. After releasing a new template, a survey of $n = 200$ customers finds that 90 report being satisfied.

- **Question**: Does this sample provide evidence that the satisfaction rate has increased with significance level $\alpha = 0.05$?

- State the hypothesis:

$$H_0 : p \leq 0.4 \text{ (equivalently, } p = 0.4) \qquad H_a : p > 0.4.$$

- The null H_0 **must** have the equality $=$. If what we want to observe is $>$, then we usually put it on alternative H_a. Here, if we want to check the satisfaction rate p is increased than benchmark 0.4, our $H_a : p > 0.4$.

Case 2: Hypothesis Testing for One-sample Proportion p

- Compute the Test Statistic:

$$\hat{p} = \frac{90}{200} = 0.45.$$
$$Z = \frac{\hat{p} - E(\hat{p})}{\mathrm{sd}(\hat{p})} = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} = \frac{0.45 - 0.4}{\sqrt{0.4(1-0.4)/200}} = 1.44.$$

- Find the p-value:

$$\text{p-value} = P(Z > 1.44) = 0.0749.$$

- **Caveat**: There is **no** such plug-in estimator or conservative estimator for the standard error SE, because we know what the true population proportion p is under the null hypothesis H_0.

- Since p-value 0.0749 is higher than significance level $\alpha = 0.05$, we **do not reject the null** hypothesis so there is **no** significant evidence that the new template increase the satisfaction rate.

Case 3: Hypothesis Testing for Two-sample Mean Difference

- Person X daily expenditures over the last 10 years:

$$\$72.15, \$50.10, \ldots$$

- Person Y daily expenditures over the last 10 years

$$\$26.25, \$42.15, \ldots$$

- Randomly sample daily receipts:

Person X	Person Y
$\bar{X} = \$53.92$	$\bar{Y} = \$49.03$
$s_X = 5.03$	$s_Y = 4.1$
$n_1 = 41$	$n_2 = 39$

- **Question**: Does this data suggest Person X has had larger average daily expenditure over the past 10 years?

Case 3: Hypothesis Testing for Two-sample Mean Difference

- **Question 1**: More specifically, state the hypothesis testing with significance level $\alpha = 0.05$.

- **Answer**: Since we want to know whether $\mu_X > \mu_Y$ or not, and the inequality $\mu_X > \mu_Y$ do not have "$=$", we just put this on the alternative H_a.

- Then,

$$H_0 : \mu_X = \mu_Y , \quad H_a : \mu_X > \mu_Y .$$

Case 3: Hypothesis Testing for Two-sample Mean Difference

- **Question 2**: Conduct the hypothesis testing with significance level $\alpha = 0.05$

$$H_0 : \mu_X = \mu_Y, \quad H_a : \mu_X > \mu_Y.$$

- **Answer**: Under the null H_0, $\mu_X - \mu_Y = 0$. The Z-score is

$$\begin{aligned} Z &= \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\mathrm{sd}(\bar{X} - \bar{Y})} = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}}} \\ &= \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}}} \approx \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{n_1} + \frac{s_Y^2}{n_2}}} \\ &= \frac{53.92 - 49.03}{\sqrt{\frac{5.03^2}{41} + \frac{4.1^2}{39}}} = 4.776. \end{aligned}$$

Case 3: Hypothesis Testing for Two-sample Mean Difference

- No way! Indeed, the p-value is

$$\text{p-value} = P(Z > 4.776) < 10^{-6}.$$

- So H_0 must be false and $\mu_X > \mu_Y$. Formally, the p-value is less than significance level $\alpha = 0.05$. It is significant under $\alpha = 0.05$ to say that the expenditure of Person X is higher than Person Y.