FUNDAMENTALS OF STATISTICS (4ECON013C)

TUTORIAL 7

Hypothesis testing (z test) about μ and π

Learning outcomes:

- Define the hypothesis testing.
- Conduct hypothesis testing for μ (population mean) and π (population proportion).
- Understand Type I and Type II errors.
- Compute and interpret the p-value in hypothesis testing.

What is a hypothesis?

A **hypothesis** is a statement or assumption about a population parameter (like the mean, proportion, or difference between groups).

There are **two types** of hypotheses in hypothesis testing:

- **Null Hypothesis** (H_0, read as "H-naught" or "H zero") — this is the default or status quo assumption. It usually states that there is no effect or no difference. Example: "The new drug is not effective in treating the disease."

The null hypothesis will contain one of the following relational operators: "=", "≥", "≤".

- **Alternative Hypothesis** (H_1 or H_a) — this is what you want to test or prove. It states that there is an effect or a difference. Example: "The new drug is effective in treating the disease."

The alternative hypothesis will contain one of the following relational operators: "≠", "<", ">".

A **hypothesis testing** is a statistical method to decide whether the data you collected provide enough evidence to reject the **null hypothesis** in favor of the **alternative hypothesis**.

While conducting a hypothesis testing different sources or textbooks provide different number of steps. But in general, the following steps need to be undertaken in hypothesis testing:

1. **Setting up the hypotheses:**

- *Case 1: Two-sided (Two-tailed):* $H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$

  Example: Is the average age in Uzbekistan different than 30 years?

  $H_0: \mu = 30$ $H_a: \mu \neq 30$

- *Case 2: Left-sided (Left-tailed):* $H_0: \mu \geq \mu_0$ $H_a: \mu < \mu_0$

  Example: Is the mean range of Tesla Model S less than 575 kilometers?

H_0: $\mu \geq 575$ (or $\mu = 575$) H_a: $\mu < 575$

- *Case 3: Right-sided (Right-tailed):* H_0: $\mu \leq \mu_0$ H_a: $\mu > \mu_0$

  Example: Is average income in Uzbekistan more than $400?

  H_0: $\mu \leq 400$ (or $\mu = 400$) H_a: $\mu > 400$

2. Calculating the test statistic value:

In this lesson we are conducting a z-test (test based on z distribution), therefore our test statistic value is called z_{stat}.

σ known:	σ unknown:
$z_{stat} = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$z_{stat} = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ (if $n \geq 30$)

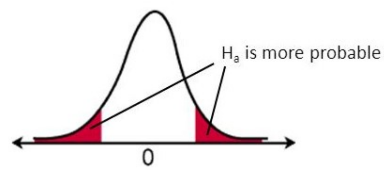
3. Formulating decision rule (rejection region):

We need a clear rule to decide whether to reject or not to reject the null hypothesis. Therefore, we set the boundaries in the standard normal distribution curve:

Case 1: Two-sided: *Reject H_0 if:* $|z_{stat}| > z_{\alpha/2}$ (z critical)

H_0: $\mu = \mu_0$

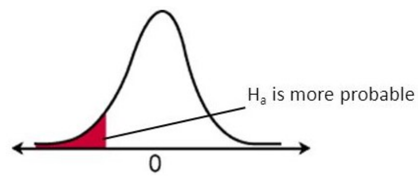
H_a: $\mu \neq \mu_0$



Case 2: Left-sided: *Reject H_0 if:* $z_{stat} < -z_\alpha$ (z critical)

H_0: $\mu \geq \mu_0$

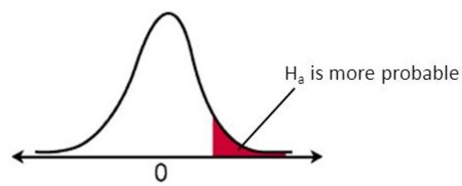
H_a: $\mu < \mu_0$



Case 3: Right-sided: *Reject H_0 if:* $z_{stat} > z_\alpha$ (z critical)

H_0: $\mu \leq \mu_0$

H_a: $\mu > \mu_0$



4. Making a decision:

Decision can be one of the following two:

- Reject the null hypothesis.
- Fail to reject (do not reject) the null hypothesis.

Once you make a decision, then you need to provide a practical interpretation of the decision in the context of research question. Your statement can start as follows: "there is sufficient evidence to support the claim …" or "there is no sufficient evidence to support the claim …".

Type I and Type II errors in hypothesis testing

After making a decision, sometimes the researcher might commit a mistake. These mistakes are called Type I and Type II errors.

Null hypothesis is …	***True***	***False***
Not rejected	Correct decision True negative Probability = 1 - α	Type II error False negative Probability = β
Rejected	Type I error False positive Probability = α	Correct decision True positive Probability = 1-β

Type I error (*false-positive*): rejecting a null hypothesis that is actually true in the population.

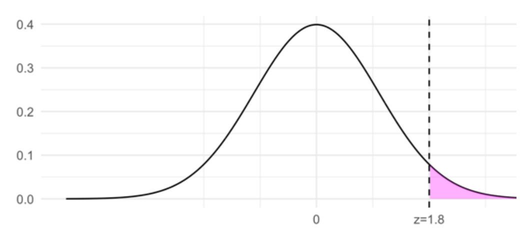
Type II error (*false-negative*): failing to reject a null hypothesis that is actually false in the population.

The **p-value** is the probability of observing a test statistic as extreme as (or more extreme than) the one you got if the null hypothesis is true. If p-value is less than alpha (significance level of the test), then we reject the null hypothesis. This is the second way of making a decision that is called "p-value approach" as an alternative to "critical value approach" in which we compare the test statistic value with a critical value. But in both approaches you end up with the same decision.

Depending on the case of hypothesis (left, right or two-tail), the calculation of the p-value differs. The area of the p-value is shaded in pink in the following curves:

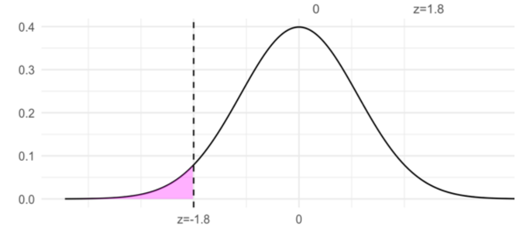
- *For right-tailed hypothesis:*

  p-value = $P(z > z_{stat})$



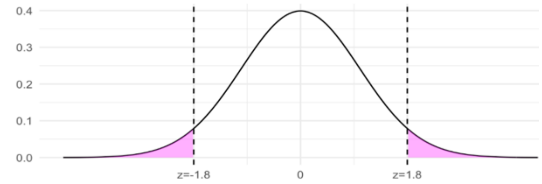
- *For left-tailed hypothesis:*

  p-value = $P(z < z_{stat})$



- *For two-tailed hypothesis:*

  p-value = 2*P(z > z_{stat}) = 2* P(z < z_{stat})



When one is conducting a hypothesis testing about a population proportion (π), the hypothesis and the calculation of z stat are different than what we saw for a population mean.

- *Case 1: Two-sided (Two-tailed):* H_0: $\pi = \pi_0$ H_a: $\pi \neq \pi_0$
- *Case 2: Left-sided (Left-tailed):* H_0: $\pi = \pi_0$ H_a: $\pi < \pi_0$
- *Case 3: Right-sided (Right-tailed):* H_0: $\pi = \pi_0$ H_a: $\pi > \pi_0$

The z statistic is found as follows:

$$\mathbf{z_{stat}} = \frac{\boldsymbol{p - \pi_0}}{\sqrt{\frac{\boldsymbol{\pi_0(1-\pi_0)}}{\boldsymbol{n}}}}$$

here, p = sample proportion

π_0 = hypothesized population proportion under the null hypothesis

n = total sample size

Example 7.1 A fund manager claims the portfolio's average monthly return is 1.5%**.** You want to test whether the manager's strategy actually produces higher returns. Assume the monthly return has known standard deviation σ = 2% and you take a sample of n = 36 months. The sample average is 1.7%. Use $\alpha = 0.05$.

- *Hypotheses:*

H_0: $\mu \leq 1.5$

H_a: $\mu > 1.5$ (Right tail)

- Test statistic:

$z_{stat} = \frac{1.7-1.5}{\frac{2}{\sqrt{36}}} = 0.6$

- *Rejection region:*

We reject H_0 if $z_{stat} > z_\alpha = 1.645$

- *Decision:* Since $z_{stat} = 0.6$ is smaller than z critical, we fail to reject the null hypothesis.

There is no evidence to claim manager's strategy actually produces higher returns than 1.5% monthly.

Example 7.2 (optional, non-examinable) Refer to the example 7.1. If the true mean is 2.5% monthly, then what is the value of β, the probability of Type II error?

Solution. To find β, first we need to find the threshold critical value of $\bar{X}$ that rejects H_0:

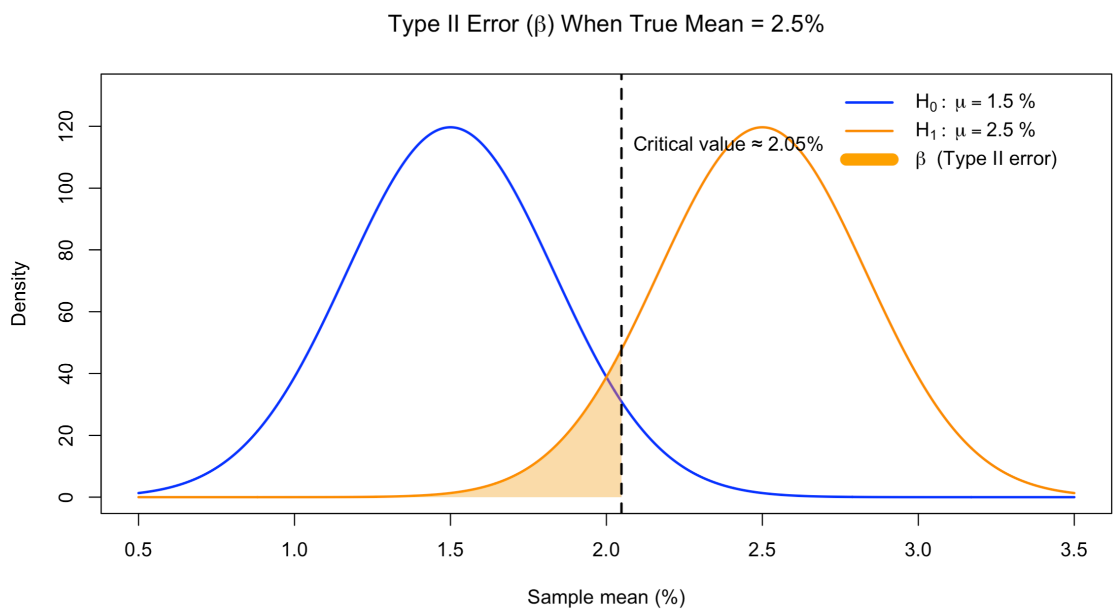
Critical $\bar{X} = \mu_0 + z_\alpha * SE$ = 1.5+1.645*0.33(3) ≈ 2.048

So, this test rejects the null hypothesis if $\bar{X} > 2.048$

If the true mean (μ_a) is 2.5%, we need to find $P(\bar{X} < 2.048 \mid \mu = 2.5)$.

$\beta = P(\bar{X} < 2.048) = P(z < \frac{2.048-2.5}{0.333}) = P(z < -1.36) = P(z > 1.36) = 1 - 0.9131 = 0.0869$

The visual representation is given below:



Example 7.3 A fund manager claims that at least half (50%) of the fund's monthly returns beat the market benchmark (S&P 500)**.** You suspect the fund performs worse than claimed, and you want to test this using past performance data. You review 100 months of data and find that the fund outperformed the market benchmark in 40 months. You want to test at 10% significance level.

Solution.

We are given p = 40/100 = 0.40, n = 100, α = 0.10.

Hypotheses:

H_0: $\pi \geq 0.5$ H_a: $\pi < 0.5$

Test statistic:

$z_{stat} = \frac{0.40-0.5}{\sqrt{\frac{0.5*0.5}{100}}} = -2.0$

Rejection region:

We reject H_0 if p-value < 0.10.

Decision: p-value = P(z < -2) = 1 – P(z < 2) = 1 – 0.9772 = 0.0228 < 0.10, so we reject H_0.

There is statistical evidence that the fund outperforms the market benchmark less than 50% of the times.

TASKS

1. **True/False questions.** Explain your choice.

- **a.** If the p-value is less than α (significance level), we fail to reject the null hypothesis.
- **b.** In a one-mean z-test, the population standard deviation (σ) is assumed to be known or sample size needs to be at least 30 to use sample standard deviation (s).
- **c.** A one-proportion z-test is appropriate to test if more than 60% of customers use online banking.
- **d.** The null hypothesis should always contain the inequality ($\neq$) operator.

2. Think about each of the following statements. For each, state the null and alternative hypotheses and say whether they will need one-tailed (left or right) or two-tailed tests.

- **a.** As per the Ministry of Labor of the Republic of Uzbekistan, the unemployment rate decreased by one percentage point, reaching 5.8% as of November 1, 2024. If you are testing whether the actual unemployment rate is higher than the reported figure, what would the null and alternative hypotheses look like?
- **b.** A politician's election campaign is conducting a survey in a region. They found 65 out of 120 randomly selected voters prefer this candidate. Do they have enough evidence to claim that majority of the voters prefer this candidate?
- **c.** You read a research article which stated that average income per household in Tashkent is $500. But you suspect that this value is different from true population parameter, μ, average income in Tashkent.
- **d.** According to Statistical Committee of Uzbekistan, average age of Uzbekistan was 28.5 years in 2017. A student researcher believes the average age has decreased over the past eight years in the country. She took random sample of 1000 people and calculated the sample average age.

3. In these exercises, a null hypothesis is given. Without using the terms "null hypothesis" and "alternative hypothesis", identify the type I and type II errors. Discuss which type of error has the greater consequence in each situation.

- **a.** Frank wants to go rock-climbing using his own equipment.
  H_0: Frank's rock-climbing equipment is safe.
- **b.** There was a car accident on the road. Onlookers take the victim to the hospital.
  H_0: The victim of the car accident is alive when he arrives at the hospital.

c. You are at work and you receive a notification that a fire alarm goes off in your house. H_0: There is no fire in the house.

4. Belissimo Pizza claims that their delivery time in Tashkent is within 35 minutes and offers the pizza for free if the delivery exceeds this time. To test whether their average delivery is slower than their claimed time, a researcher looked at randomly selected pizza delivery records of 50 and found the sample average of 39 minutes with a standard deviation of 0.25 hours. Use the significance level of 5%. What type of error is possible with your decision?

5. The manufacturer of a patient medicine claimed that it was 90% effective in relieving an allergy for a period of 8 hours. In a random sample of 200 people suffering from the allergy, the medicine provided relief for 160 people. Determine whether there is evidence to challenge the manufacturer's claim at 10% significance level?

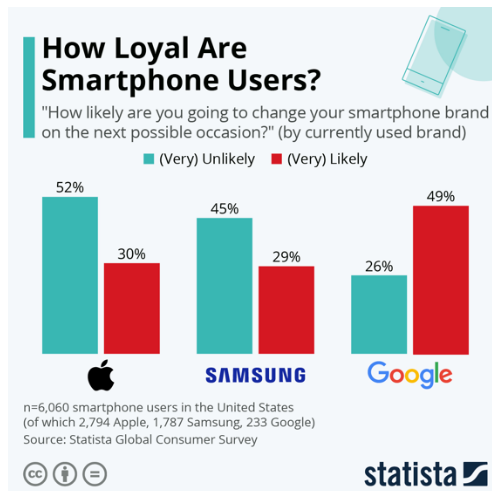
6. You are given the following survey conducted by Statista. According to survey data, can you conclude that majority of iPhone users are loyal to the Apple products? Use alpha = 0.01.

a. State the hypotheses.

b. Compute the test statistic value.

c. What is the p-value of this test?

d. What is the conclusion and interpretation?



7. Most air travelers now use e-tickets. Electronic ticketing allows passengers to not worry about a paper ticket, and it costs the airline companies less than paper ticketing. However, recently the airlines have received complaints from passengers regarding their e-tickets, particularly when connecting flights and a change of airlines were involved. To investigate the problem, an independent watchdog agency contacted a random sample of 30 airports and collected information on the number of complaints the airport had with e-tickets for the month of March. The information is reported in the Table.

14	14	16	15	16	14	13	13	15	18
12	15	15	14	13	13	15	12	10	16
8	10	14	15	20	17	15	15	14	19

At the 0.05 significance level, can the watchdog agency conclude the mean number of complaints per airport is less than 15 per month? You can use MS Excel to calculate sample mean and standard deviation.

a. State the null and alternative hypotheses.

b. Provide the test statistic (z_{stat}).

c. State the rejection region.

d. Provide your decision and interpretations.

8. You come across an article that reports the average income in Uzbekistan to be $600, based on a cluster sampling method. However, you are skeptical about the result because recent evidence from the World Bank shows a growing income disparity in the country. According to the World Bank:

> "For the most recent period (2022–2023), the poorest 10 percent saw their incomes grow by 6 percent, while the richest 10 percent experienced more than 30 percent income growth. The Gini Coefficient**,** a measure of income inequality, increased from 0.31 in 2022 to 0.35 in 2023."

This growing inequality suggests that the population is becoming more diverse in terms of income levels. In such a context, cluster sampling may not provide accurate estimates of the population mean because clusters (e.g., regions, neighborhoods, or institutions) can differ significantly in their income distributions. If some clusters contain mostly high-income or low-income individuals, the sample mean could be biased, especially when clusters are not perfectly representative of the entire population.

To obtain more reliable results, you decide to conduct your own study using a stratified sampling method, ensuring that each income group or region is properly represented in proportion to its size in the population.

Your study yields the following results:

- Sample mean income: $540
- Sample standard deviation: $160
- Sample size: n = 350

You want to test whether the true mean income in Uzbekistan is different from the $600 reported in the article, at a 2% significance level ($\alpha = 0.02$).

HOMEWORK

9. Suppose you wanted to conduct a blind tasting test for your Marketing class to find out whether there is a significant difference between the proportion of consumers who prefer Pepsi over Coca Cola. You randomly invited 210 tasters to try two cups of colas while their eyes were closed and each individual indicated his/her preferred drink. 90 of them preferred Pepsi over Coke, 10 of them were undecided, and the remaining consumers preferred Coca Cola. Can you conclude with 95% confidence that the percentage of those who prefer Pepsi is different than 50%?

- **a.** State your hypotheses.
- **b.** Calculate the test statistic.
- **c.** Provide your rejection region.
- **d.** What is your decision and interpretations?
- **e.** What is the p-value for this test?

10. Statistics can help to decide the authorship of literary works. Suppose poems by Alisher Navoi are known to contain an average of $\mu = 8.9$ new words (words not used in the poet's other works). The standard deviation of the number of new words is $\sigma = 2.5$. Now a manuscript with 30 new poems have come to light, and scholars are debating whether it is Navoi's work. The new poems contain an average of 10.2 words not used in the poet's known works. We expect poems by another author to contain more new words, so we test whether average new words are more than known $\mu = 8.9$.

- **a.** State the null and alternative hypotheses.
- **b.** Provide the test statistic (zstat).
- **c.** State the rejection region ($\alpha = 0.05$).
- **d.** Provide your decision and interpretations.

11. Unoccupied seats on flights cause airlines to lose revenue. Suppose a large airline wants to estimate its mean number of unoccupied seats per flight over the past year. To accomplish this, the records of 225 flights are randomly selected and the number of unoccupied seats is noted for each of the sampled flights. The sample mean is 11.6 seats and the sample standard deviation is 4.1 seats. Test whether average number of unoccupied seats significantly exceeds 11 seats per flight.

- **a.** State the null and alternative hypotheses.
- **b.** Provide the test statistic (zstat).
- **c.** State the rejection region ($\alpha = 0.10$).
- **d.** Provide your decision and interpretations.

12. One of the largest problems on college campuses is alcohol abuse by underage students. Although all 50 states have mandated by law that no one under the age of 21 may possess or purchase alcohol, many college students report that alcohol is readily available. More problematic is that these same students report that they drink with one goal in mind—to get drunk. Universities are acutely aware of the problem of binge drinking, defined as consuming five or more drinks in a row three or more times in a two-week period. An extensive survey of college students reported that 44% of U.S. college students engaged in binge drinking during the two weeks before the survey. The president of a large Midwestern university stated publicly that binge drinking was not a problem on her campus of 25,000 undergraduate students. A service fraternity conducted a survey of 2,500 undergraduates attending the university and found that 1,200 of the 2,500 students had engaged in binge drinking.

- **a.** Is there sufficient evidence to indicate that the percentage of students engaging in binge drinking at the university is greater than the percentage found in the national survey? Use α of 0.05.
- **b.** Place a 99% confidence interval on the percentage of binge drinkers at the university.