8 Hypothesis testing

In this chapter we will learn how researchers study the effectiveness of certain drugs, or determine whether the characteristics of a certain part of a population differ from those of the entire population. The general methodology relies on sampling the population under study. We need to make here a point that it is very important, namely to make sure that we have an unbiased sample. A sample is biased if it is not representative for the entire population.

We consider the following problem. A population has mean and standard deviation μ and σ respectively. We apply a treatment to the population and then we take a sample of n individuals and calculate the sample mean $\overline{x}$. We want to decide whether the treatment was effective or not. Using the CLT we can find a confidence interval for the mean after the treatment was applied and based on the confidence level we can decide weather the the new mean is different significantly from the original one. If the treatment has not been effective, then the mean stays the same, and the sample mean is too close to the original mean and the effect of the treatment cannot be shown. This is the situation when the null hypothesis H_0 cannot be rejected.

$H_0 =$ Null Hypothesis and it can be formulated in three ways, depending on the application

- H_0 : There is no change in the mean
- H_0: There is no significant increase in the mean
- H_0: There is no significant decrease in the mean

If we find that the new mean is significantly different, larger or smaller than the original mean, then we accept the alternate hypothesis H_A.

- H_A : The new mean is different from the original one
- H_A : The new mean is larger that the original one
- H_A : The new mean is smaller from the original one

8.1 Z-test for the Population Mean

For hypothesis testing we can follow the following steps.

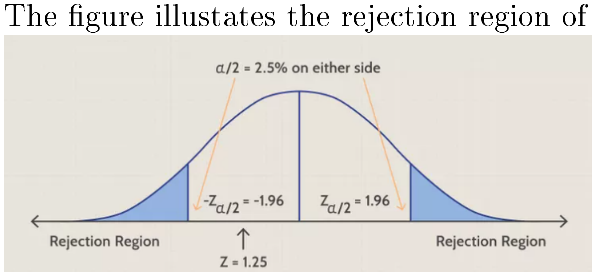
- **Formulate H_0 and H_A.**
- **Set criteria for decision making.** Using the significance level we calculate the critical z-score.
- **Collect data.** Calculate the standard deviation of the mean and the actual z-score. If needed we need to we can also calculate the p-value.
- **Draw conclusions.** Using the z-score or the p-value, we can reject or accept the Null Hypothesis.

Based on the problem we consider we can have a two-tailed or one-tailed hypothesis test.

8.1.1 Two-tailed test

If we are interested in showing that a treatment results in a significan difference for the population mean we need to use a two-tailed test. Let μ_0 denote the population mean before the treatment and μ denote the population mean after the treatment. We also have the significance level α, standard deviation σ, sample size n, and sample mean $\overline{x}$. In this case the above steps can be formulated in detail:

- **Formulate H_0 and H_A.**
  - $H_0 : \mu = \mu_0$
  - $H_A : \mu \neq \mu_0$
- **Set criteria for decision making.** The null hypothesis can be rejected if the probability that the new mean is the same as the old one is very low, i.e., less than the significance level. We can calculate the critical z-score
$$z_{\frac{\alpha}{2}} = \left|\Phi^{-1}\left(\frac{\alpha}{2}\right)\right|.$$
  The figure illustates the rejection region of a two-tailed test.



- **Collect data.** From CLT we know that the standard deviation of the mean is $\sigma_M = \frac{\sigma}{\sqrt{n}}$. The actual z-score (also called test statistic) can be calculated
$$z = \frac{\overline{x} - \mu_0}{\sigma_M}.$$
We can make a decision based solely on the z-score, however in some situations the p-value gives more information. The p-value can be calculated as
$$p = P(Z < -z \text{ or } Z > z).$$
The p-value is the probability that the z-score that we obtained can be observed as test statistic when the null hypothesis is true. p-values that are small (less than the significance level) warrant the null hypothesis to be rejected, while a large p-value (larger than the significance level) suggests that we cannot reject the null hypothesis. There is a mnemonic to remember this rule: "If p is high, the null will fly, if p is low the null must go".

- **Draw conclusions.** Using the z-score or the p-value, we can reject or accept the null hypothesis. If the z-score is in the rejection region, we can reject the null hypothesis, otherwise we cannot reject it.

Example 1. Mean statistics exam scores are at 85% with standard deviation of 8%. A Study group is formed with 9 students. After their participation in the study group their average score was 90%. Is the score significantly different from the population mean? Use significance level of 5%.

We introduce the notations $\mu_0 = 85$, $\sigma = 8$, $n = 9$,$\overline{x} = 90$, $\alpha = 0.05$.
The null and alternate hypotheses are
H_0 : Mean score did not change ($\mu = 85$)
H_A : Mean score is significanly different for the study group $\mu \neq 85$.
As the significance level is $\alpha = 0.05$, we have
$$z_{\frac{\alpha}{2}} = z_{0.025} = 1.96.$$
We can calculate the standard deviation of the mean
$$\sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{8}{3} = 2.67$$
and the z-score, also called the test statistic
$$z = \frac{x - \mu}{\sigma_M} = 1.87.$$

As a conclusion, since the z-score falls within the acceptance region of H_0, we cannot reject it. So, there is no significant difference between the means at significance level of 5%.

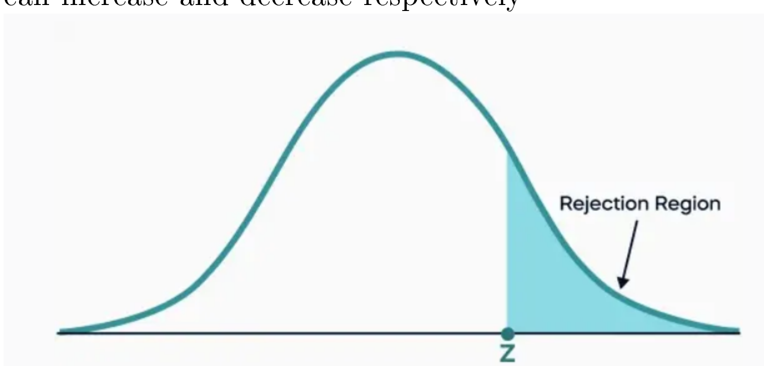
8.1.2 One-tailed test

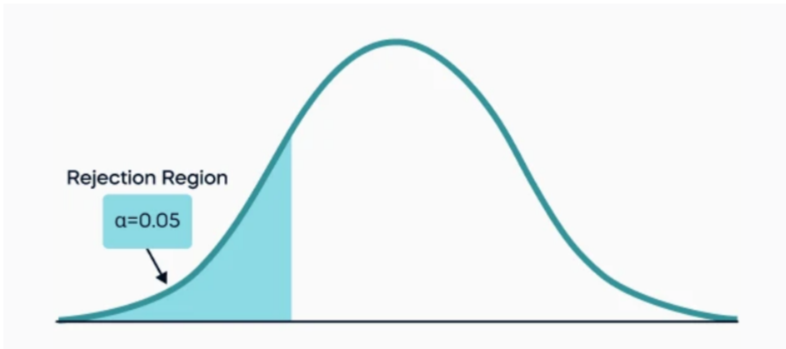
If we are interested in showing that a treatment results in a significan increase or decrease for the population mean we need to use a one-tailed test. In this case the hypothesis testing steps can be formulated as follows:

- **Formulate H_0 and H_A.**
  - H_0 : No significant increase (decrease) $\mu \leq \mu_0$ ($\mu \geq \mu_0$)
  - H_A : Significant increase (decrease) $\mu > \mu_0$ ($\mu < \mu_0$)
- **Set criteria for decision making.** The null hypothesis can be rejected if the probability that the new meanhas not increased (decreased) is very low, i.e., less than the significance level. We can calculate the critical z-score

  $$z_{\frac{\alpha}{2}} = \left|\Phi^{-1}\left(\alpha\right)\right|.$$

  The figures illustate the rejection region of a one-tailed test for significan increase and decrease respectively





- **Collect data.** We have $\sigma_M = \frac{\sigma}{\sqrt{n}}$ and

$$z = \frac{\overline{x} - \mu_0}{\sigma_M}.$$

  If the problem at hand requires us to compute the p-value, then we compute it as $p = P(Z > z)$ for increase and $p = P(Z < z)$ for decrease. The p-value is interpreted as the probability that our test statistic is obtained while the null hypothesis is true.

- **Draw conclusions.** Using the z-score or the p-value, we can reject or accept the null hypothesis. If the z-score is in the rejection region, we can reject the null hypothesis, otherwise we cannot reject it.

Example 2. We solve the question in the previous example but we weant to know whether there is a significant increase in test scores for the study gorup.

The null and alternate hypotheses are
H_0 : Mean score did not increase: $\mu \leq 85$
H_A : Mean score did significanly increase: $\mu > 85$.
As the significance level is $\alpha = 0.05$, we have

$$z_\alpha = z_{0.05} = 1.645.$$

We have the standard deviation of the mean

$$\sigma_M = \frac{\sigma}{\sqrt{n}} = 2.67$$

and the z-score

$$z = \frac{x - \mu}{\sigma_M} = 1.87.$$

As a conclusion, since the z-score falls within the rejection region of H_0, we reject the null hypothesis. So, there is a significant increase of the mean, at significance level of 5%. Let us calculate the p-value as well

$$p = P(Z > 1.87) = 0.0307.$$

The interpretation of the p-value is the probability that the mean has not increased. Since the p-value is less than the significance level, then the null hypothesis can be rejected.

8.2 Z-test for population proportion

We want to see how a certain treatment changes a population proportion. Let p denote the population proportion after the treatment is applied, while p_0 denotes the original population proportion. To perform Hypothesis testing for population proportion we need to follow the same process as for a population mean. The only difference being that the standard deviation for the population proportion is given by

$$\sigma_M = \sqrt{\frac{p_0 q_0}{n}},$$

where p_0 is the hypothesized population proportion and $q_0 = 1 - p_0$. Further we take $p' = \frac{x}{n}$ is the measured population proportion.

Example 3. 60% of gamers like our game. A new feature is added and we want to assess whether it will make our game more popular, so we ask 50 people if they like the game with the added feature and 35 responded yes. Does that show that our game is more popular with the new feature added? Use 95% confidence level.

To test the hypothesis considered we need to do a hypothesis test. Let p denote the population proportion of people who like our game, p_0 being the proportion before we added the feature.

Our hypotheses is

$H_0 : p \leq 0.6$

$H_A : p > 0.6$

The significance level is $\alpha = 0.05$. The test is a one-tailed (right-tailed) test with the critical z-score being $z_{0.05} = 1.645$. We have

$$p' = \frac{x}{n} = \frac{35}{50} = 0.7$$

The standard deviation of the mean is

$$\sigma_M = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{0.6 \cdot 0.4}{50}} = 0.069.$$

The z-score is

$$z = \frac{0.7 - 0.6}{0.069} = 1.56.$$

Since the z-score is less than the critical z-score, we cannot reject the null hypothesis.

8.3 T-test

If the population standard deviation is not known, instead of the z-test we can use Student's t-test for the population mean. Student's t-test can be performed using the sample mean and standard deviation as follows

- Formulate H_0 and H_A.
- Set criteria for decision making. Using the significance level and $\nu = n-1$ the number of degrees of freedom, we calculate the critical t-score t_α for one-tailed test, and $t_{\frac{\alpha}{2}}$ for two tailed test.
- **Collect data.** We calculate $s_M = \frac{s}{\sqrt{n}}$ and the actual t-score (test statistic) $t = \frac{\overline{x}-\mu_0}{s_M}$.
- **Draw conclusions.** Using the t-score or the p-value, we can reject or accept the Null Hypothesis.

Example 4. We collected 31 energy bars from different stores. The label states that they contain $20g$ of protein. Is the label correct given that $\overline{x} = 21.4$, $s = 2.54$, at significance level $\alpha = 0.05$?

$H_0 : \mu = 20$

$H_A : \mu \neq 20$

The number of degrees of freedom is $\nu = 30$. The critical t-score is $t_{\frac{\alpha}{2}} = t_{0.025} = 2.042$.

We calculate $s_M = \frac{s}{\sqrt{n}} = 0.456$ and the test statistic

$$t = \frac{\overline{x} - \mu_0}{s_M} = \frac{21.4 - 20}{0.456} = 3.07.$$

As a conclusion we can reject H_0.

8.4 Comparing two population means

In this section we use the tools for hypothesis testing using two samples from two populations, to compare their means. These methods are often used in testing efficiency of certain drugs. For this case one would consider a subgroup of the population to receive the drug, while another subgroup from the population to receive a placebo. Let $\overline{x}_i$ denote the mean of a sample of size n_i from a population with mean μ_i and standard deviation σ_i, $i = 1, 2$.

To perform the hypothesis test we use the result that the difference is normally distributed with distribution

$$\overline{x}_1 - \overline{x}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right).$$

Next we observe that comparing μ_1 and μ_2 is equivalent to comparing their difference to 0.

Example 5. We want to compare two brands of bateries. One of them has population standard deviation of $\sigma_1 = 0.33$ while the other has $\sigma_2 = 0.36$. We collect a sample of 20 batteries of each type and we find that type 1 has sample mean of 3 years, while type 2 batteris have sample mean of 2.9 years. Can we conclude that type 1 batteries last longer than type 2 batteries? We will use a 98% confidence level.

We first formulate our hypotheses:
$H_0 : \mu_1 \leq \mu_2$ (i.e., $\mu_1 - \mu_2 \leq 0$)
$H_A : \mu_1 > \mu_2$ (i.e., $\mu_1 - \mu_2 > 0$)
We will perform a one-tailed test at significance level $\alpha = 2\%$. The critical z-score is $z_{0.02} = 2.05$.

The standard deviation of the mean difference can be calculated as

$$\sigma_{MD} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 0.109$$

z-score can be calculated as

$$z = \frac{\overline{x}_1 - \overline{x}_2}{\sigma_{MD}} = \frac{3 - 2.9}{0.109} = 0.917$$

as the z-score is less than the critical z-score we cannot reject the null hypothesis.

8.5 Exercises

1. A particular brand of tires claims that its deluxe tire averages at least 50,000 miles before it needs to be replaced. From past studies of this tire, the standard deviation is known to be 8,000. A survey of owners of that tire design is conducted. From the 28 tires surveyed, the mean lifespan was 46,500 miles with a standard deviation of 9,800 miles. Using alpha = 0.05, is the data highly inconsistent with the claim?

2. From generation to generation, the mean age when smokers first start to smoke varies. However, the standard deviation of that age remains constant of around 2.1 years. A survey of 40 smokers of this generation was done to see if the mean starting age is at least 19. The sample mean was 18.1 with a sample standard deviation of 1.3. Do the data support the claim at the 5% level?

3. The cost of a daily newspaper varies from city to city. However, the variation among prices remains steady with a standard deviation of 20¢. A study was done to test the claim that the mean cost of a daily newspaper is $1.00. Twelve costs yield a mean cost of 95¢ with a standard deviation of 18¢. Do the data support the claim at the 1% level?

4. An article in the San Jose Mercury News stated that students in the California state university system take 4.5 years, on average, to finish their undergraduate degrees. Suppose you believe that the mean time is longer. You conduct a survey of 49 students and obtain a sample mean of 5.1 with a sample standard deviation of 1.2. Do the data support your claim at the 1% level?

5. The mean number of sick days an employee takes per year is believed to be about ten. Members of a personnel department do not believe this figure. They randomly survey eight employees. The number of sick days they took for the past year are as follows: 12; 4; 15; 3; 11; 8; 6; 8. Let x = the number of sick days they took for the past year. Should the personnel team believe that the mean number is ten? Use 5% significance level.

6. Your statistics instructor claims that 60 percent of the students who take her Elementary Statistics class go through life feeling more en-

riched. For some reason that she can't quite figure out, most people don't believe her. You decide to check this out on your own. You randomly survey 64 of her past Elementary Statistics students and find that 34 feel more enriched as a result of her class. Now, what do you think? Use 5% significance level.

7. According to an article in Bloomberg Businessweek, New York City's most recent adult smoking rate is 14%. Suppose that a survey is conducted to determine this year's rate. Nine out of 70 randomly chosen N.Y. City residents reply that they smoke. Conduct a hypothesis test to determine if the rate is still 14% or if it has decreased. Use 5% significance level.

8. A student at a four-year college claims that mean enrollment at four–year colleges is higher than at two–year colleges in the United States. Two surveys are conducted. Of the 35 two–year colleges surveyed, the mean enrollment was 5,068 with a standard deviation of 4,777. Of the 35 four-year colleges surveyed, the mean enrollment was 5,466 with a standard deviation of 8,191. Conduct a hypothesis test at 1% significance level to prove or disprove the claim.

9. Mean entry-level salaries for college graduates with mechanical engineering degrees and electrical engineering degrees are believed to be approximately the same. A recruiting office thinks that the mean mechanical engineering salary is actually lower than the mean electrical engineering salary. The recruiting office randomly surveys 50 entry level mechanical engineers and 60 entry level electrical engineers. Their mean salaries were $46,100 and $46,700, respectively. Their standard deviations were $3,450 and $4,210, respectively. Conduct a hypothesis test to determine if you agree that the mean entry-level mechanical engineering salary is lower than the mean entry-level electrical engineering salary. Use 5% significance level.