Hypothesis Testing

By Dr. Bidisha Bhabani

What is hypothesis?

- A hypothesis is a proposition formulated for empirical testing. By empirical testing we mean relying on practical experience rather than theories.
- Types of Hypothesis:
  - Research Hypothesis
    - It is any hypothesis framed by a researcher in order to study a relationship between variables, effectiveness of a process etc.
  - Statistical Hypothesis
    - It is regarding a population. A level of precision is attached for the test of a statistical hypothesis
  - Substantial Hypothesis
    - A statistically significant difference does not imply or mean a material, substantive difference

Objective of Hypothesis Testing

- The objective of hypothesis testing is to either reject or retain a null hypothesis.
- In many cases, for example, in regression models, one would like to reject the null hypothesis to establish statistically significant relationship between the dependent and the independent variable.
- However, in goodness of fit tests, that are used for checking whether the data follows a specific distribution or not, we would like to retain the null hypothesis.

INTRODUCTION TO HYPOTHEIS TESTING

- Hypothesis testing is a statistical process of either rejecting or retaining a claim or belief or association related to a business context, product, service, processes, etc
- Hypothesis test consists of two complementary statements called null hypothesis and alternative hypothesis, and only one of them is true
- Hypothesis is an integral part of many predictive analytics techniques such as multiple linear regression and logistic regression

IMPORTANCE OF HYPOTHESIS TESTING

- Hypothesis is an integral part of many predictive analytics techniques such as multiple linear regression and logistic regression.
- It plays an important role in providing evidence of an association relationship between an outcome variable and predictor variables.

Few examples of such claims are listed below:

- Children who drink the health drink Complan (a health drink owned by the company Heinz in India) are likely to grow taller
- If you drink Horlicks, you can grow taller, stronger, and sharper (3 in 1).

Exploratory Data Analysis and Confirmatory Data Analysis

- Data analysis in general can be classified as exploratory data analysis or confirmatory data analysis

- In exploratory data analysis, the idea is to look for new or previously unknown hypothesis or suggest hypotheses

- In the case of confirmatory data analysis, the objective is to test the validity of a hypothesis (confirm whether the hypothesis is true or not) using techniques such as hypothesis testing and regression.

HYPOTHESIS TESTING STEPS

1. Describe the hypothesis in words. Hypothesis is described using a population parameter (such as mean, standard deviation, proportion, etc.) about which a claim (hypothesis) is made. Few sample claims (hypothesis) are:

   - Average time spent by teenagers using social media is more than middle aged people.
   - On average women upload more photos in social media than men.

HYPOTHESIS TESTING STEPS

2) Based on the claim made in step 1, define null and alternative hypotheses. Initially we believe that the null hypothesis is true. In general, null hypothesis means that there is no relationship between the two variables under consideration (for example, null hypothesis for the claim 'teenagers use social media more than middle aged people' will be 'there is no relationship between age and the average time spent in social media'). Null and alternative hypotheses are defined using a population parameter.

3) Identify the test statistic to be used for testing the validity of the null hypothesis. Test statistic will enable us to calculate the evidence in support of null hypothesis. The test statistic will depend on the probability distribution of the sampling distribution; for example, if the test is for mean value and the mean is calculated from a large sample and if the population standard deviation is known, then the sampling distribution will be a normal distribution and the test statistic will be a *Z*-statistic (standard normal statistic).

HYPOTHESIS TESTING STEPS

4. Decide the criteria for rejection and retention of null hypothesis. This is called **significance value** traditionally denoted by symbol α. The value of α will depend on the context and usually 0.1, 0.05, and 0.01 are used. Significance value α is the Type I error.
5. Calculate the *p*-value (probability value), which is nothing but the conditional probability of observing the test statistic value when the null hypothesis is true. In simple terms p-value is the evidence in support of the null hypothesis.
6. Take the decision to reject or retain the null hypothesis based on the *p*-value and significance value α. The null hypothesis is rejected when *p*-value is less than α and the null hypothesis is retained when *p*-value is greater than or equal to α.

Description of Hypothesis

Hypotheses are claims that are usually stated in simple words as listed below:

- Average annual salary of machine learning experts is different for males and females.
- On average people with Ph.D. in analytics earn more than people with Ph.D. in engineering.
- The average box-office collection of comedy genre movies is more than that of action movies.
- Average life of vegetarians is more than meat eaters.
- Proportion of married people defaulting on loan repayment is less than proportion of singles defaulting on loan repayment.

Null and Alternative Hypothesis

- **Null hypothesis**, usually denoted as H_0 (H zero and H naught), refers to the statement that there is no relationship or no difference between different groups with respect to the value of a population parameter.
- The null hypothesis refers to a specified value of the population parameter, not a sample statistic

- **Alternative hypothesis**, usually denoted as H_A (or H_1), is the complement of null hypothesis.

Hypothesis statement to definition of null and alternative hypothesis

S. No.	Hypothesis Description	Null and Alternative Hypothesis
1	Average annual salary of machine learning experts is different for males and females. (In this case, the null hypothesis is that there is no difference in male and female salary of machine learning experts)	H0: $\mu m = \mu f$ HA: $\mu m \neq \mu f$ μm and μf are average annual salary of male and female machine learning experts, respectively.
2	On average people with Ph.D. in analytics earn more than people with Ph.D. in engineering.	H0: $\mu a \leq \mu e$ HA: $\mu a > \mu e$ μa = Average annual salary of people with Ph.D. in analytics. μe = Average annual salary of people with Ph.D. in engineering. It is essential to have the equal sign in null hypothesis statement.

Hypothesis Testing

Hypothesis test checks the validity of the null hypothesis based on the evidence from the sample. At the beginning of the test, we assume that the null hypothesis is true. Since the researcher may believe in alternative hypothesis, she/he may like to reject the null hypothesis. However, in many cases (such as goodness of fit tests), we would like to retain or fail to reject the null hypothesis.

Test Statistic

- Test statistic is the standardized difference between the estimated value of the parameter being tested calculated from the sample(s) and the hypothesis value (that is standardized difference between $\overline{x}$ and μ) in order to establish the evidence in support of the null hypothesis.

- It measures the standardized distance (measured in terms of number of standard deviations) between the value of the parameter estimated from the sample(s) and the value of the null hypothesis.

P - Value

- The *p*-value is the conditional probability of observing the statistic value when the null hypothesis is true.

- For example, consider the following hypothesis:

Average annual salary of machine learning experts is at least 100,000. The corresponding null hypothesis is H_0: $\mu_m \leq 100{,}000$. Assume that estimated value of the salary from a sample is 1,10,000 (that is $\bar{X} = 1{,}10{,}000$ and assume that the standard deviation of population is known and standard error of the sampling distribution is 5000 (that is, $\sigma/\sqrt{n} = 5000$ where *n* is the sample size using which $\bar{X} = 1{,}10{,}000$ was calculated).

Hypothesis Testing

- The standardized distance between estimated salary from hypothesis salary is (1,10,000 – 1,00,000)/5000 = 2.

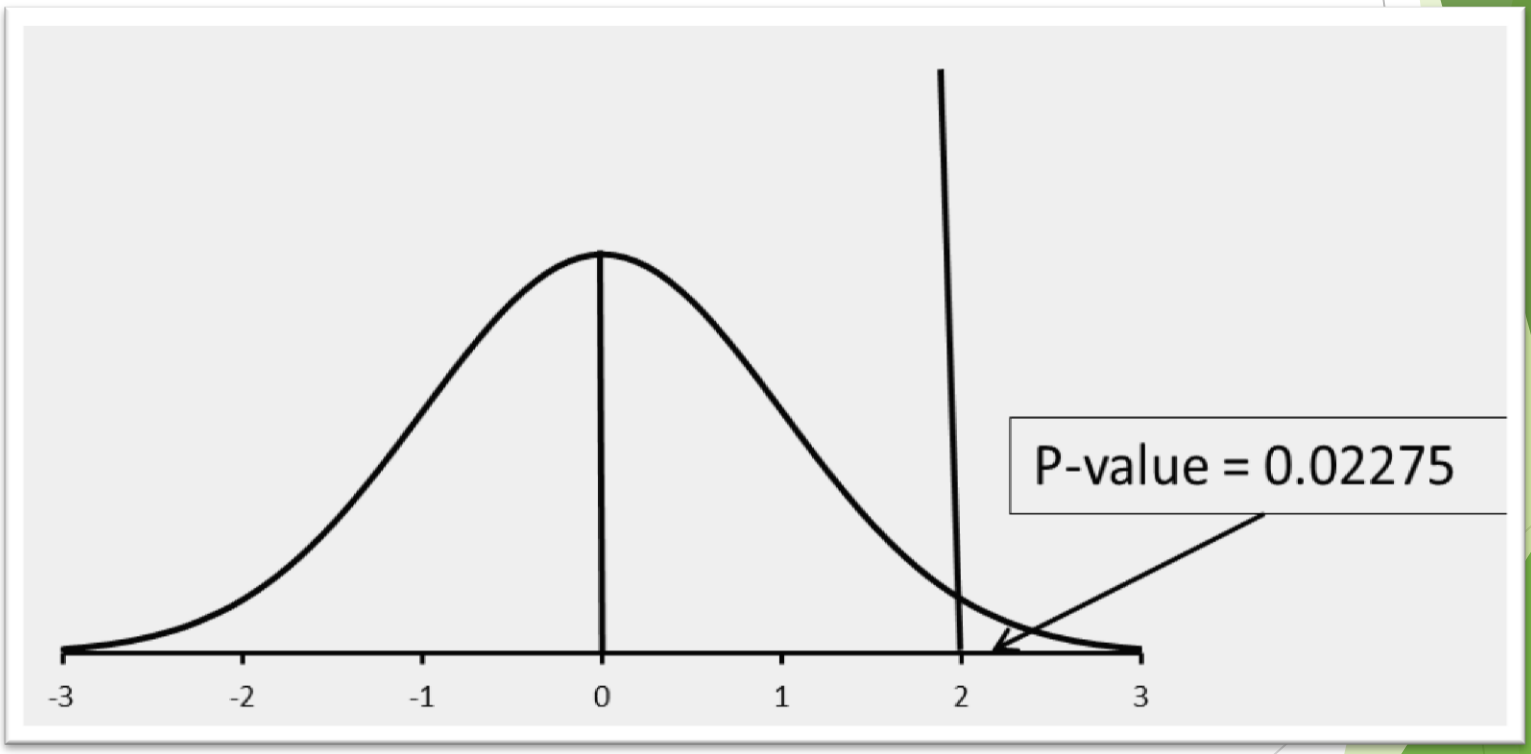
- That is, the standardized distance between estimated value and the hypothesis value is 2 and we can now find the probability of observing this statistic value from the sample if the null hypothesis is true (that is if $\mu_m \leq 100{,}000$).

- A large standardized distance between the estimated value and the hypothesis value will result in a low *p*-value.

- Note that the value 2 is actually the value under a standard normal distribution since it is calculated from

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

Standard normal distribution and the *p*-value corresponding to Z = 2 are shown below:



Hypothesis Testing

- Probability of observing a value of 2 and higher from a standard normal distribution is 0.02275.
- That is, if the population mean is 1,00,000 and standard error of the sampling distribution is 5000 then probability of observing a sample mean greater than or equal to 1,10,000 is 0.02275.
- The value 0.02275 is the *p*-value, which is the evidence in support of the statement in the null hypothesis.

p-value = P(Observing test statistics value | null hypothesis is true)

P-value

Note that the p-value is a conditional probability. It is the conditional probability of observing the statistic value given that the null hypothesis is true. P-value is the evidence in support of null hypothesis.

Decision Criteria – Significance Value

- Significance level, usually denoted by α, is the criteria used for taking the decision regarding the null hypothesis (reject or retain) based on the calculated *p*-value.
- The significance value α is the maximum threshold for *p*-value.
- The decision to reject or retain will depend on whether the calculated *p*-value crosses the threshold value α or not

Decision making under hypothesis testing

Criteria	Decision
p-value < α	Reject the null hypothesis
p-value $\geq \alpha$	Retain (or fail to reject) the null hypothesis

- Usually α = 0.05 is used by researchers (recommended by Fisher, 1956); however, values such as 0.1, 0.02, and 0.01 are also frequently used.

- The value of α chosen is very low (0.05) for reason that we start the process of hypothesis testing with an assumption that null hypothesis is true

- The value of statistic for which the probability is α is called the **critical value.**

- The areas beyond the critical values are known as **rejection region**.

Significance Value (α)

The significance value α is the threshold conditional probability of rejecting a null hypothesis when it is true. It is the value of Type I error.

	Level of significance		
	1% (0.01)	*5% (0.05)*	*10% (0.1)*
Two tailed test	$\lvert Z_\alpha \rvert = 2.58$	$\lvert Z_\alpha \rvert = 1.96$	$\lvert Z_\alpha \rvert = 0.645$
Right tailed test	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$	$Z_\alpha = 1.28$
Left tailed test	$Z_\alpha = -2.33$	$Z_\alpha = -1.645$	$Z_\alpha = -1.28$

One-Tailed and Two-Tailed Test

Consider the following three hypotheses

1. Salary of machine learning experts on average is at least US $100,000.
2. Average waiting time at the London Heathrow airport security check is less than 30 minutes.
3. Average annual salaries of male and female MBA students are different at the time of graduation.

Example 1

Statement 1 – Salary of machine learning experts on average is at least US $100,000:

The null and alternative hypotheses in this case are given by

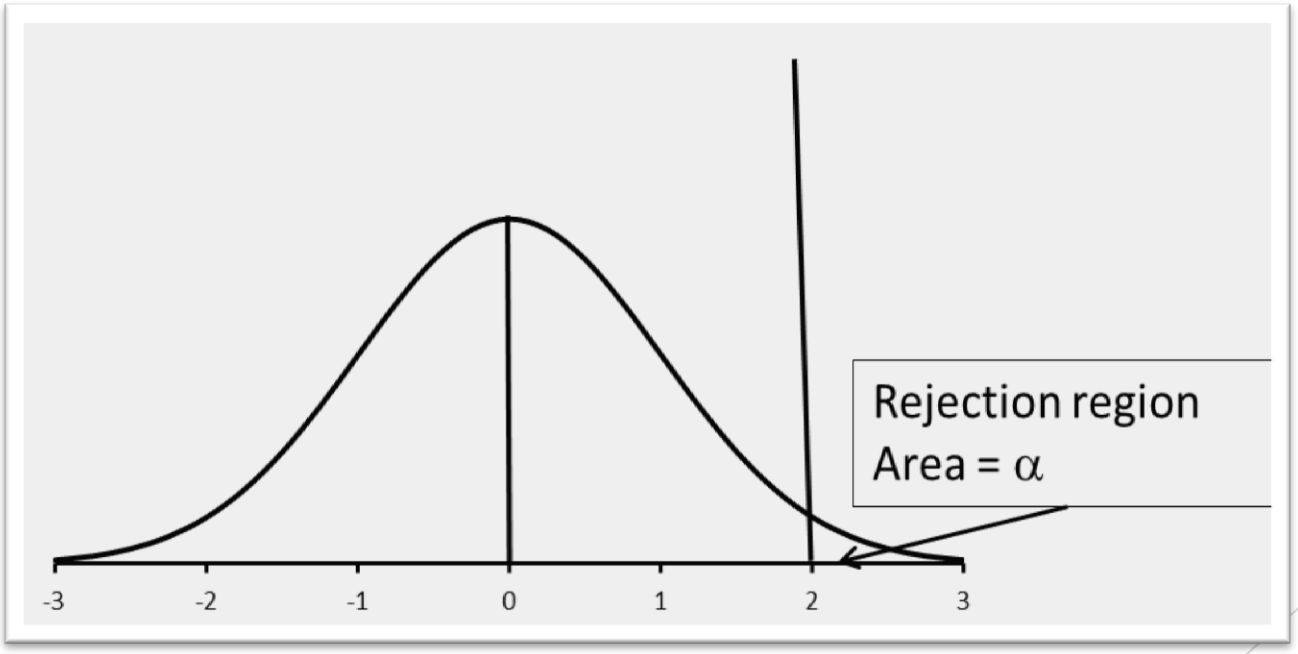
H_0: $\mu_m \leq 100{,}000$

H_A: $\mu_m > 100{,}000$

where μ_m is the average annual salary of machine learning experts. Note that the equality symbol is always part of the null hypothesis since we have to measure the difference between estimated value from the sample and the hypothesis value. In this case, reject or retain decision will depend on the direction of deviation of the estimated parameter from the sample from hypothesis value.

Solution

Below figure shows the rejection region on the right side of the distribution. Since the rejection region is only on one side this is a one-tailed test (right tailed test). Specifically, since the alternative hypothesis in this case is $\mu_m > 100{,}000$, this is called right-tailed test.



Example 2

- **Statement 2 – Average waiting time at the London Heathrow airport security check is less than 30 minutes:**

The null and alternative hypotheses in this case are given by

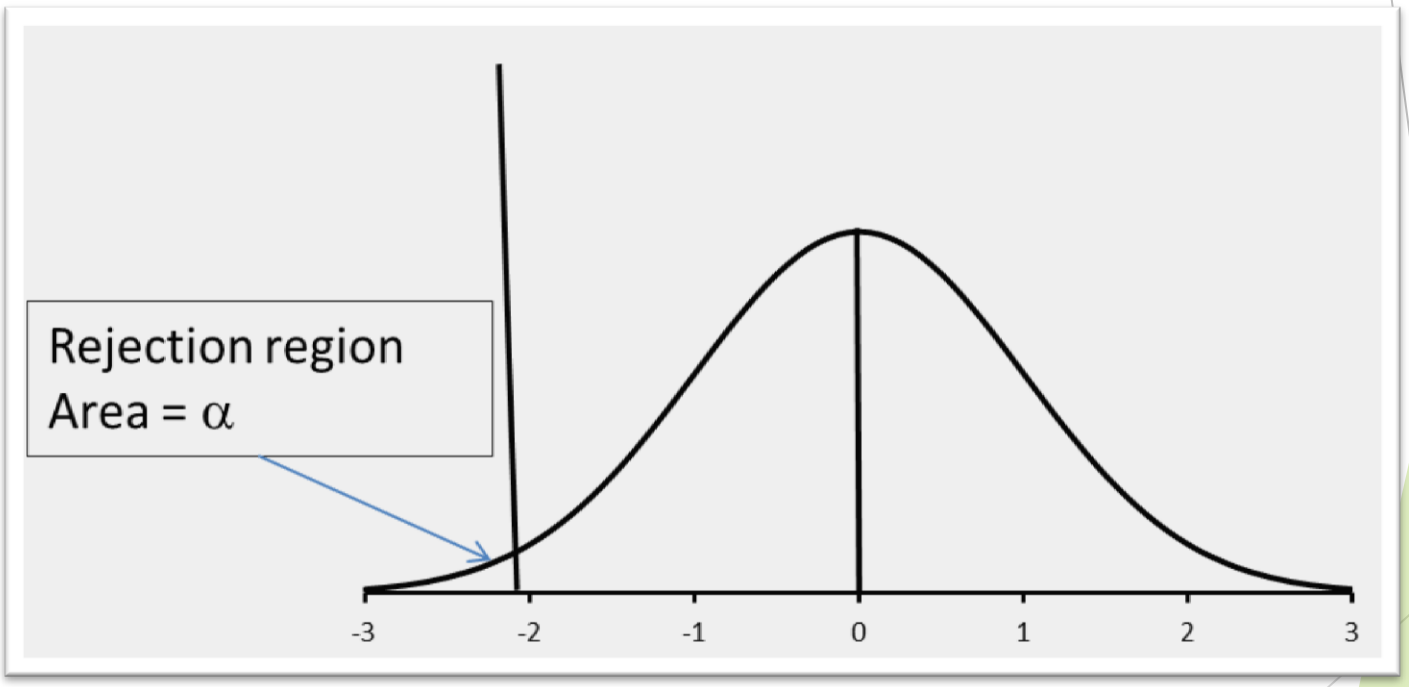
H_0: $\mu_w \geq 30$

H_A: $\mu_w < 30$

where μ_w is the average waiting time at London Heathrow security check. In this case, reject region will on the left side (known as left tailed test) of the distribution as shown in Figure

Solution

Rejection region in case of left-sided test



Example 3

Statement 3 – Average salary of male and female MBA students at graduation is different:

The null and alternative hypotheses in this case are given by

H_0: $\mu_m = \mu_f$

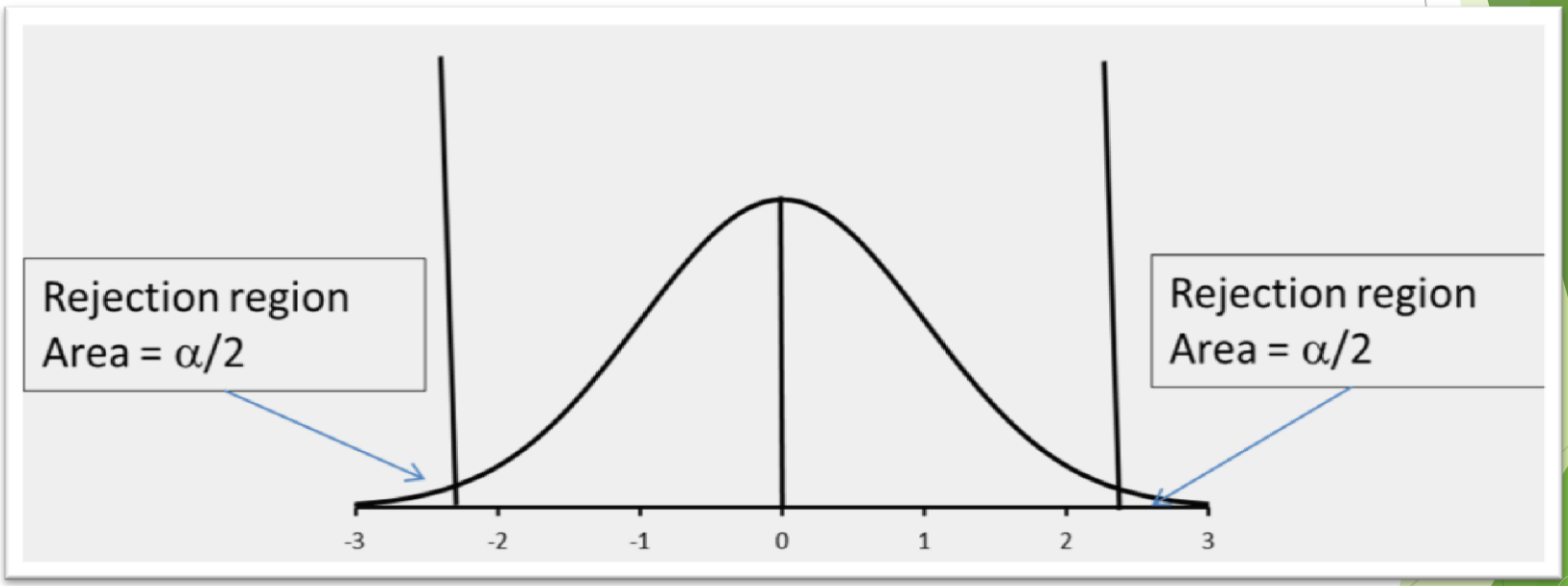
H_A: $\mu_m \neq \mu_f$

Where μ_m and μ_f are the average salaries of male and female MBA students, respectively, at the time of graduation.

In this case, the rejection region will be on either side of the distribution and if the significance level is α then the rejection region will be $\alpha/2$ on either side of the distribution. Since the rejection region is on either side of the distribution, it will be a two-tailed test.

Solution

Rejection region in case of two-tailed test



Decisions of Hypothesis Test

In hypothesis test we end up with the following two decisions:

- Reject null hypothesis.
- Fail to reject (or retain) null hypothesis.

Type I Error, Type II Error, and Power of the Hypothesis Test

TYPE I ERROR

Type I Error: Conditional probability of rejecting a null hypothesis when it is true is called Type I Error or False Positive (falsely believing claim made in alternative hypothesis is true). The significance value α is the value of Type I error.

Type I Error = α= P(Rejecting null hypothesis | H_0 is true)

- Probability value (p-value) is the evidence against the null hypothesis whereas significance value α is the error based on repetitive sampling.

- Whereas the significance level α refers to incorrect rejection of null hypothesis when it is true under **repeated trials**.

TYPE II ERROR

- **Type II Error:** Conditional probability of failing to reject a null hypothesis (or retaining a null hypothesis) when the alternative hypothesis is true is called Type II Error or False Negative (falsely believing there is no relationship). Usually Type II error is denoted by the symbol β. Mathematically, Type II error can be defined as follows:

Type II Error = β = P(Retain null hypothesis | H_0 is false)

Power of the Hypothesis Test

The value $(1 - \beta)$ is known as the power of hypothesis test. That is, the power of the test is given by

power of test = $1 - \beta$ = P(Reject null hypothesis | H_0 is false)

Description of Type I error, Type II error, and the Power of Test

	Decision made about null hypothesis based on the hypothesis test	
Actual value of H_0	Reject H_0	Retain H_0
H_0 is true	Type I error $P(\text{Reject } H_0 \mid H_0 = \text{true}) = \alpha$	Correct Decision $P(\text{Retain } H_0 \mid H_0 = \text{true}) = (1 - \alpha)$
H_0 is false	Correct Decision (Power of test) $P(\text{Reject } H_0 \mid H_0 = \text{false}) = 1 - \beta$	Type II Error $P(\text{Retain } H_0 \mid H_0 = \text{false}) = \beta$

Hypothesis Testing for Population Mean with known Variance: *Z*-Test

- *Z*-test (also known as one-sample *Z*-test) is used when a claim (hypothesis) is made about the population parameter such as population mean or proportion when population variance is known.
- Since the hypothesis test is carried out with just one sample, this test is also known as **one-sample *Z*-test**.
- *Z*-test uses CLT to conduct a hypothesis test for population mean when the population variance is known; the test statistics for *Z*-test is given by

$$Z\text{-statistic} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

- The critical value in this case will depend on the significance value α and whether it is a one-tailed or two-tailed test

Critical value for different values of α

	Approximate Critical Values		
α	Left-tailed test	Right-tailed test	Two-tailed test
0.1	−1.28	1.28	−1.64 and 1.64
0.05	−1.64	1.64	−1.96 and 1.96
0.02	-2.05	2.05	-2.33 and 2.33
0.01	−2.33	2.33	−2.58 and 2.58

Condition for rejection of null hypothesis H_0

Type of Test	Condition	Decision
Left-tailed test	Z-statistic < Critical value	Reject H_0
	Z-statistic ≥ Critical value	Retain H_0
Right-tailed test	Z-statistic > Critical value	Reject H_0
	Z-statistic ≤ Critical value	Retain H_0
Two-tailed test	|Z-statistic| > |Critical Value|	Reject H_0
	|Z-statistic| ≤ |Critical Value|	Retain H_0

Example 1: One-Sample Z-Test

- **Problem:** A coffee shop claims that the average waiting time for customers is less than 5 minutes. To test this claim, a random sample of 25 customers is taken, and the average waiting time is found to be 4.6 minutes with a standard deviation of 1.2 minutes. Test the hypothesis at a 5% significance level.

Solution:

- **Formulate Hypotheses:**
  - Null Hypothesis (H0): $\mu \geq 5$ (The average waiting time is 5 minutes or more)
  - Alternative Hypothesis (H1): $\mu < 5$ (The average waiting time is less than 5 minutes)
- **Set Significance Level (Alpha):**
  - $\alpha = 0.05$
- **Calculate the Test Statistic (Z):** $z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$
- **Compare with Critical Value or P-value:**
  - For a one-tailed test, compare the Z-value with the critical Z-value for $\alpha = 0.05$.
  - If Z < Z_critical, reject the null hypothesis.
- **Conclusion:**
  - If the Z-value falls in the rejection region, reject the null hypothesis.
  - Conclude that there is enough evidence to support the claim that the average waiting time is less than 5 minutes.