Hypothesis Testing – Summary

Concepts in Hypothesis Testing I

The statistical analyses learnt in Inferential Statistics enable you try to make inferences about population mean and other population data from the sample data. However, you could not confirm the confirm the conclusions you made about the population about the data. It is here that **hypothesis testing** comes into the picture.

Understanding Hypothesis Testing

What is a Hypothesis?

When we perform an analysis on a population sample — the analysis could be descriptive, inferential, or exploratory in nature — we get certain information from which we can make **claims about the entire population**. These are just the claims; we can't be sure if they're actually true. This kind of claim or assumption is called a **hypothesis**.

Example: The average commute time of employees of a company to and fro office is 35 minutes

What is Hypothesis Testing?

There are ways to check if your hypothesis has any truth to it, and if the hypothesis is true then apply it to the population parameters. This is called hypothesis testing. The goal is to determine whether there is enough evidence to infer that the hypothesis about the population parameter is true. In hypothesis testing, we confirm our assumptions about the population based on sample data.

Difference between Inferential Statistics & Hypothesis Testing

Inferential statistics is used to find the mean of a population parameter when you have no initial number to start with. So, you start with the sampling activity and find out the sample mean. Then, you estimate the population mean from the sample mean using the confidence interval.

Hypothesis testing is used to confirm your conclusion (or hypothesis) about the population mean (which you know from EDA or your intuition). Through hypothesis testing, you can determine whether there is enough evidence to conclude if the hypothesis about a population parameter is true or not.

- **Null hypothesis** (H_0): The status quo
- **Alternate hypothesis** (H_1): The challenge to the status quo

Example: Suppose a man has been charged with murder. In the criminal trial for this case, the jury has to decide whether the defendant is innocent or guilty. Now, this can be turned into two hypotheses. You can claim that the defendant is innocent, and you can claim that the defendant is not innocent, i.e. guilty.

Therefore, you have two opposing hypotheses about the defendant. These two opposing hypotheses are called the null hypothesis and the alternate hypothesis.

- The **null hypothesis** is the prevailing belief about a population; it states that there is no change or no difference in the situation. In our criminal trial example, the defendant was considered innocent. So, the null hypothesis claims that he is innocent, just like he was before the murder charge. Null Hypothesis is denoted by H_0

- The **alternate hypothesis**, or **research hypothesis** as it is also called, is the claim that opposes the null hypothesis. If you were the prosecutor in the trial, your claim would be that the defendant is guilty, and you would try to prove this. So, the alternate hypothesis is an assumption that competes with the null hypothesis. Alternate Hypothesis is denoted by H_1

Outcome of Hypothesis Testing

Suppose, at the end of the trial, the jury reaches a decision. What would be the outcome in terms of hypothesis testing?

If the defendant is found guilty, it means that the jury rejects the null hypothesis in favour of the alternate hypothesis. The jury decides that there is enough evidence to support the alternate hypothesis, and to conclude that the defendant is guilty.

On the other hand, if the jury acquits the defendant, it means that there is not enough evidence to support the alternate hypothesis. Keep in mind that this does not mean that the defendant is innocent, it just means that there is not enough evidence to conclude that he is guilty. In other words, we cannot accept the null hypothesis; we can only fail to reject it.

Therefore, in hypothesis testing, if there is sufficient evidence to support the alternate hypothesis, you reject the null hypothesis; and if there is not sufficient evidence to support the alternate hypothesis, you fail to reject the null hypothesis. So, you should never say that you “accept” the null hypothesis.

You should never say that you “accept” the null hypothesis.

Formulating Null & Alternate Hypotheses

If your claim statement has words like “at least”, “at most”, “less than”, or “greater than”, **you cannot formulate the null hypothesis just from the claim statement** (because it’s not necessary that the **claim is always about the status quo**).

You can use the following rule to formulate the null and alternate hypotheses:

- **The null hypothesis** always has the following signs: $=$ OR $\leq$ OR $\geq$
- **The alternate hypothesis** always has the following signs: $\neq$ OR $>$ OR $<$

For example:

Situation 1: Flipkart claimed that its total valuation in December 2016 was at least $14 billion. Here, the claim contains ≥ sign (i.e. the at least sign), so **the null hypothesis is the original claim**.

The hypothesis in this case can be formulated as:

Figure 1 - Hypotheses for Situation 1

Situation 2: Flipkart claimed that its total valuation in December 2016 was greater than $14 billion. Here, the claim contains > sign (i.e. the ‘more than’ sign), so **the null hypothesis is the complement of the original claim**.

The hypothesis in this case can be formulated as:

Figure 2 - Hypotheses for Situation 2

To summarize this, you cannot decide the status quo or formulate the null hypotheses from the claim statement, you need to take care of signs in writing the null hypothesis. Null Hypothesis never contains ≠ or > or < signs. It always has to be formulated using ≠ or > or < signs.

Once you have formulated the null and alternate hypotheses, the next most important step of hypothesis testing is — **making the decision to either reject or fail to reject the null hypothesis**

Situation 1: If sample mean is greater than UCV or less than LCV, i.e. sample mean lies in the criticals region.

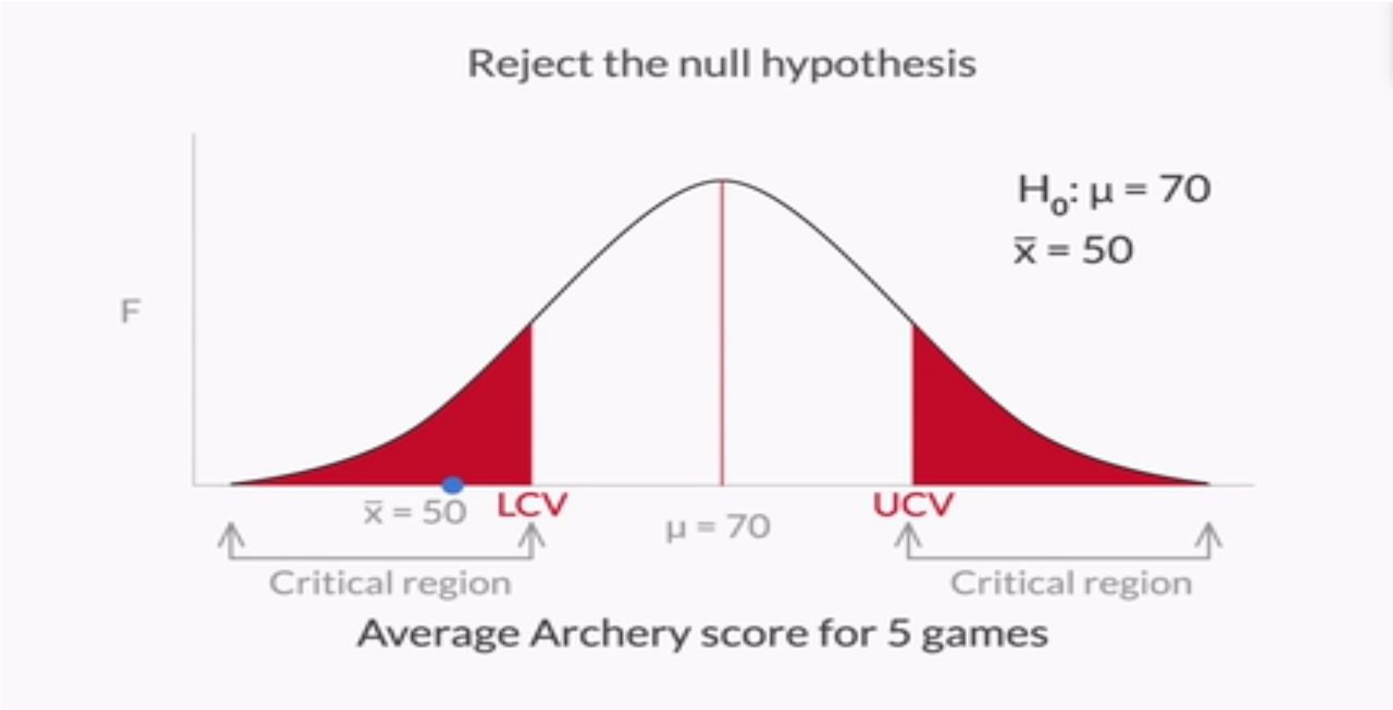


Figure 3 – Rejecting a Null Hypothesis

Situation 2: If sample mean is less than UCV or greater than LCV, i.e. sample mean lies in the acceptance region.

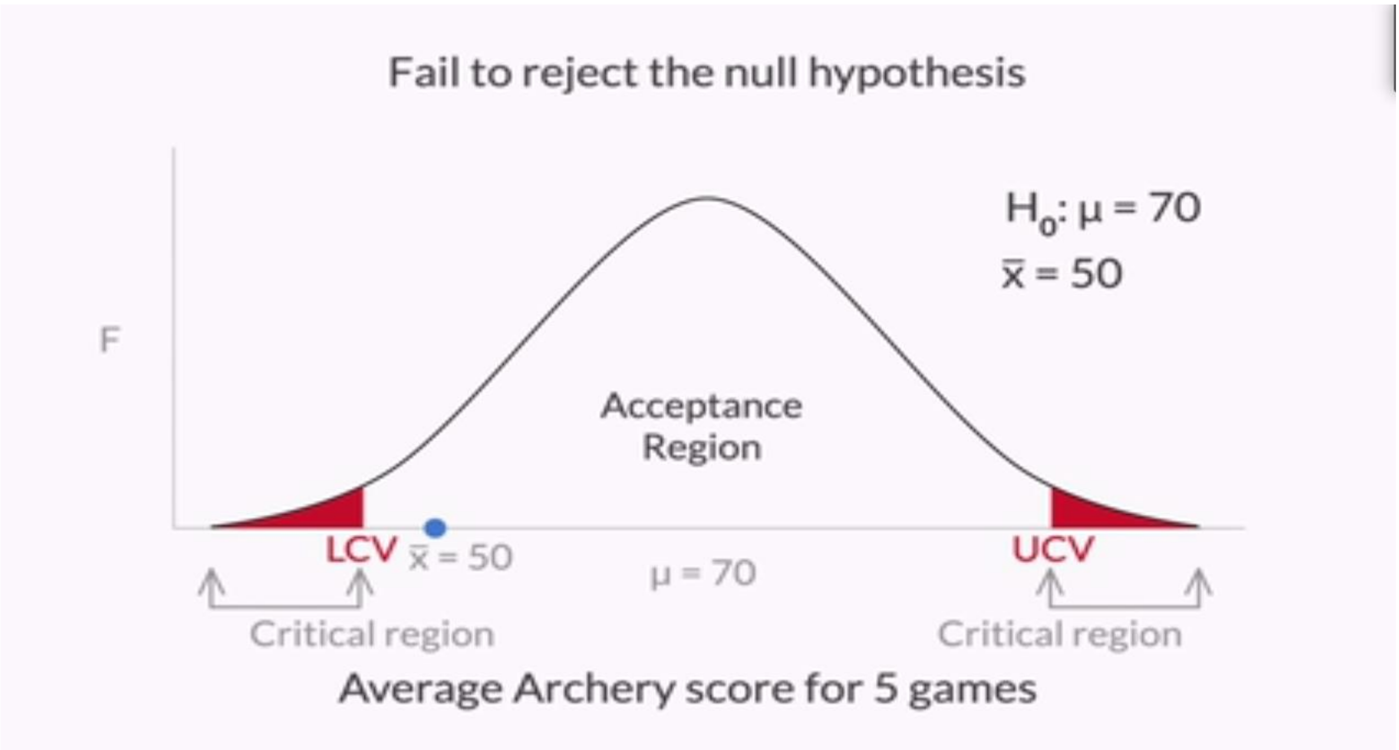


Figure 4 – Failing to reject a Null Hypothesis

The formulation of the null and alternate hypotheses determines the type of them test and the position of the critical regions in the normal distribution.

You can tell the type of the test and the position of the critical region on the basis of the '**sign' in the alternate hypothesis.**

≠ in H_1 → Two-tailed test → Rejection region on **both sides** of distribution

< in H_1 → Lower-tailed test → Rejection region on **left side** of distribution

> in H_1 → Upper-tailed test → Rejection region on **right side** of distribution

Critical Value Method

After formulating the hypothesis, the steps you have to follow to **make a decision** using **the critical value method** are as follows:

1. Calculate the value of Zċ from the given value of α (significance level). Take it a 5% if not specified in the problem.
2. Calculate the critical values (UCV and LCV) from the value of Zċ.
3. Make the decision on the basis of the value of the sample mean x with respect to the critical values (UCV AND LCV).

You can download the z-table from this link.

Critical Value Method – Complete Example

Assume that you are the owner of multiple AC stores. You want to know about the mean demand of AC units per month per store during summer. Till now you have been ordering 350 AC units per store per month based on the historic demand. But this time because of intense heat waves, you anticipate that the demand might go up. So you want to check your assumption that the average units required in one month will be different from 350 units per store.

In this case you are assuming that 350 units is the average number of units that are sold every month. When you try visualising it, you use histogram and the mean comes out to be 350 approximately. This becomes the mean / average of population. Following figure 5 shows the histogram for this.

Next you will define the null and the alternate hypothesis. You start with the null hypothesis, i.e. the assumption about the status quo. So you assume that H_0 is true, and this implies that your population mean is still equal to 350. In this AC hypothesis problem, the assumption is that the average demand for AC units per store in one month is 350 units. So your null hypothesis H_0 states that the mean demand of ACs is 350 units per store every month.

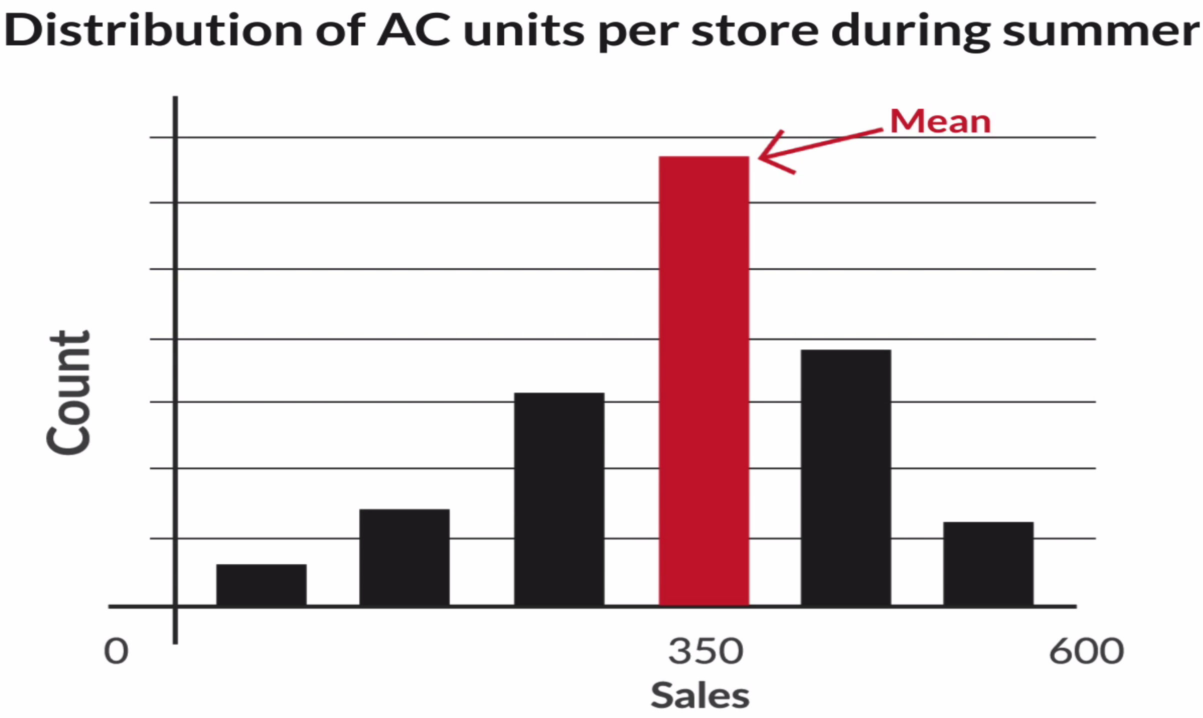


Figure 5 - Histogram showing the average number of ACs

Now that your null hypothesis is clear, the next step is to state the alternative or research hypothesis. H_1 is the opposite of H_0 — it challenges the status quo. This is the assumption that you try to prove, and you should examine all evidence with respect to H_1. In your AC sales problem, H_1 is that the population mean $\mu \neq 350$.

> You should always examine the evidences with respect to alternative hypothesis NOT with respect to null hypothesis.

You know that the population standard deviation sigma (σ) is 90, i.e. the distribution obtained every year, containing the sales numbers of every store, has a standard deviation of 90. This year after the sales are over, you take a random sample of 25 stores and plot them. The mean sales turns out to be 370.16. This is your evidence. You can clearly see that it differs from the assumed population mean of 350 units per store. Following figure 6 shows the same.

You can see that it differs from the assumed population mean of 350 units per store. As you are working on samples, you will compute the standard error. The standard error can be calculated as standard deviation /

sqrt(number of samples). You calculate the standard error, because you want to know is the value 370.16 has significant distance from mean 350, so that the null hypothesis can be rejected.

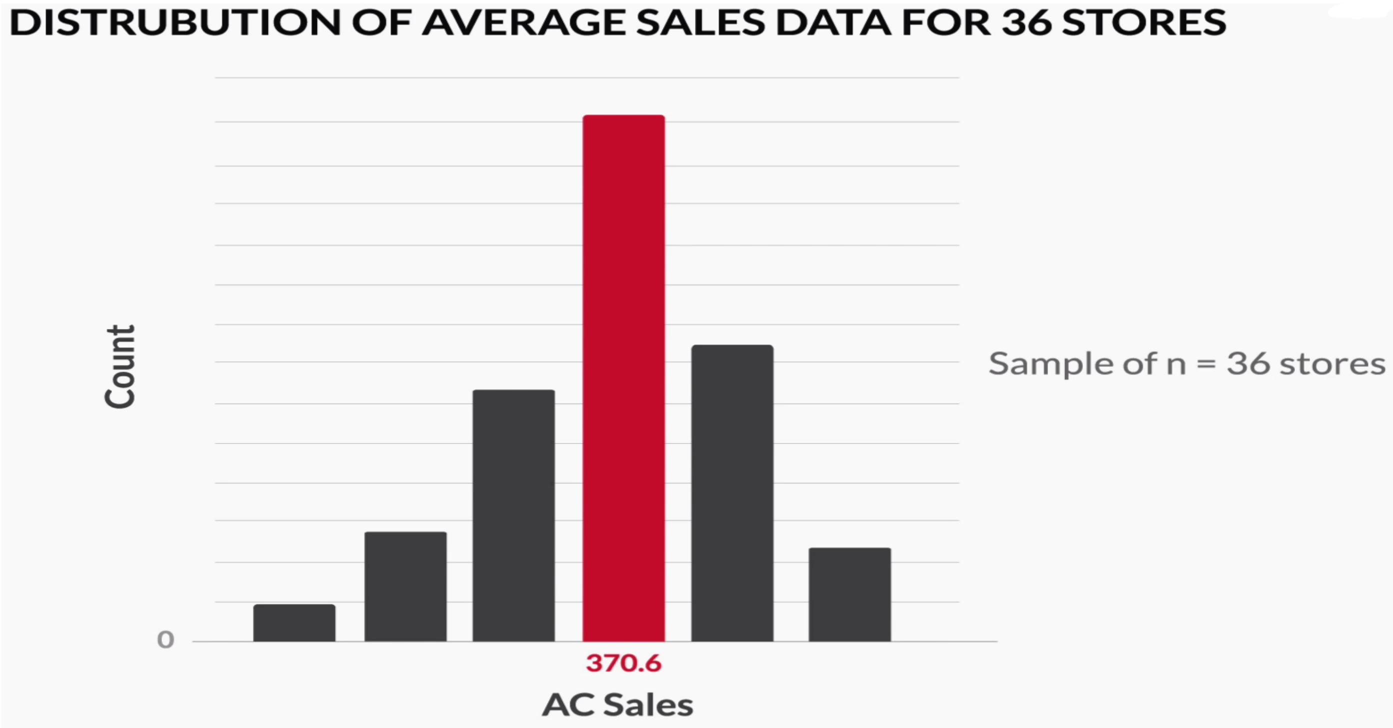


Figure 6 - The mean of sample is 370.16

So sampling distribution of sample means can be drawn in the graph with the given information as:

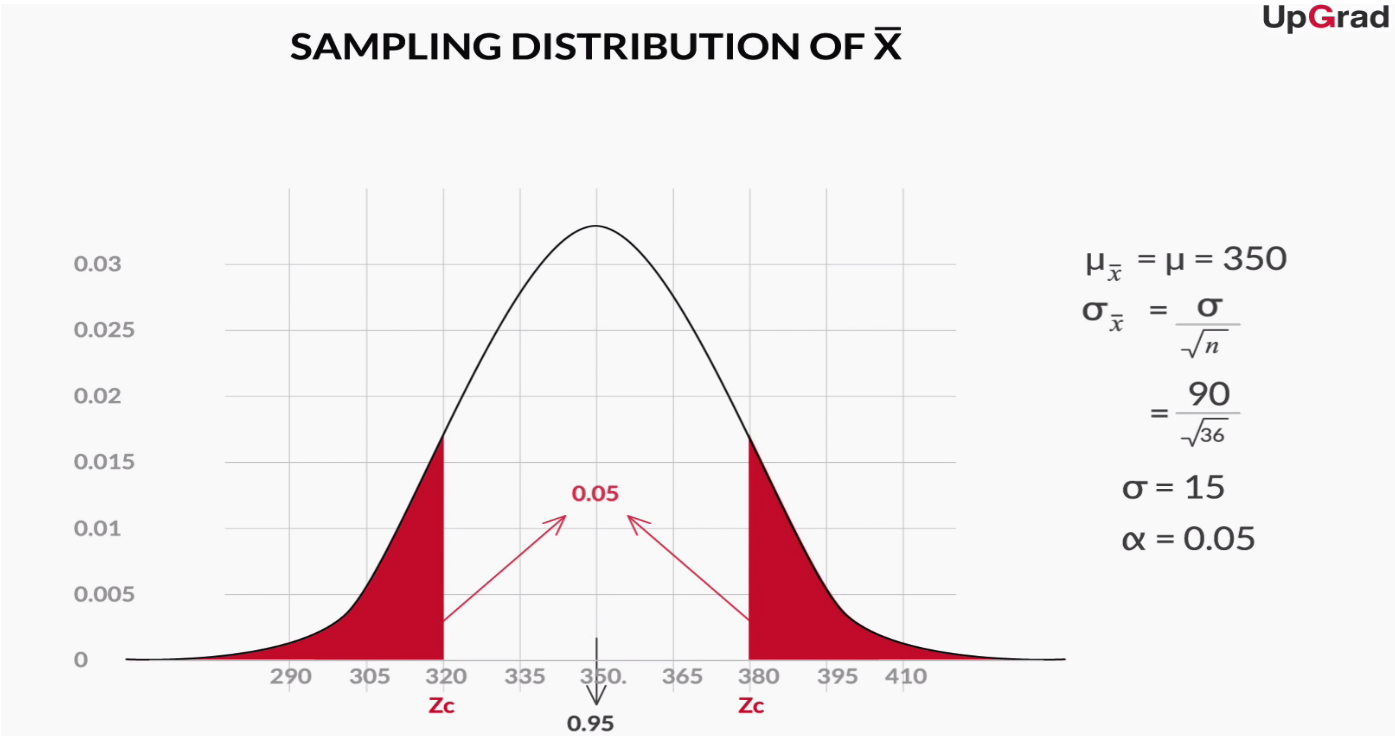


Figure 7 – Sampling distribution of sample means

Formulating Hypotheses:

The first step would be to formulate the hypotheses:

$$H_0: \mu = 350$$

There is no change in status quo

$$H_1: \mu \neq 350$$

The status quo has changed

Figure 8 – Formulating Hypotheses for AC sales problem

Making a Decision – Critical Value Method:

The first step of critical value method is to find Zc. To do that, you calculate the cumulative probability of UCV from the value of α, which is further used to find the z-critical value (Zc) for UCV.

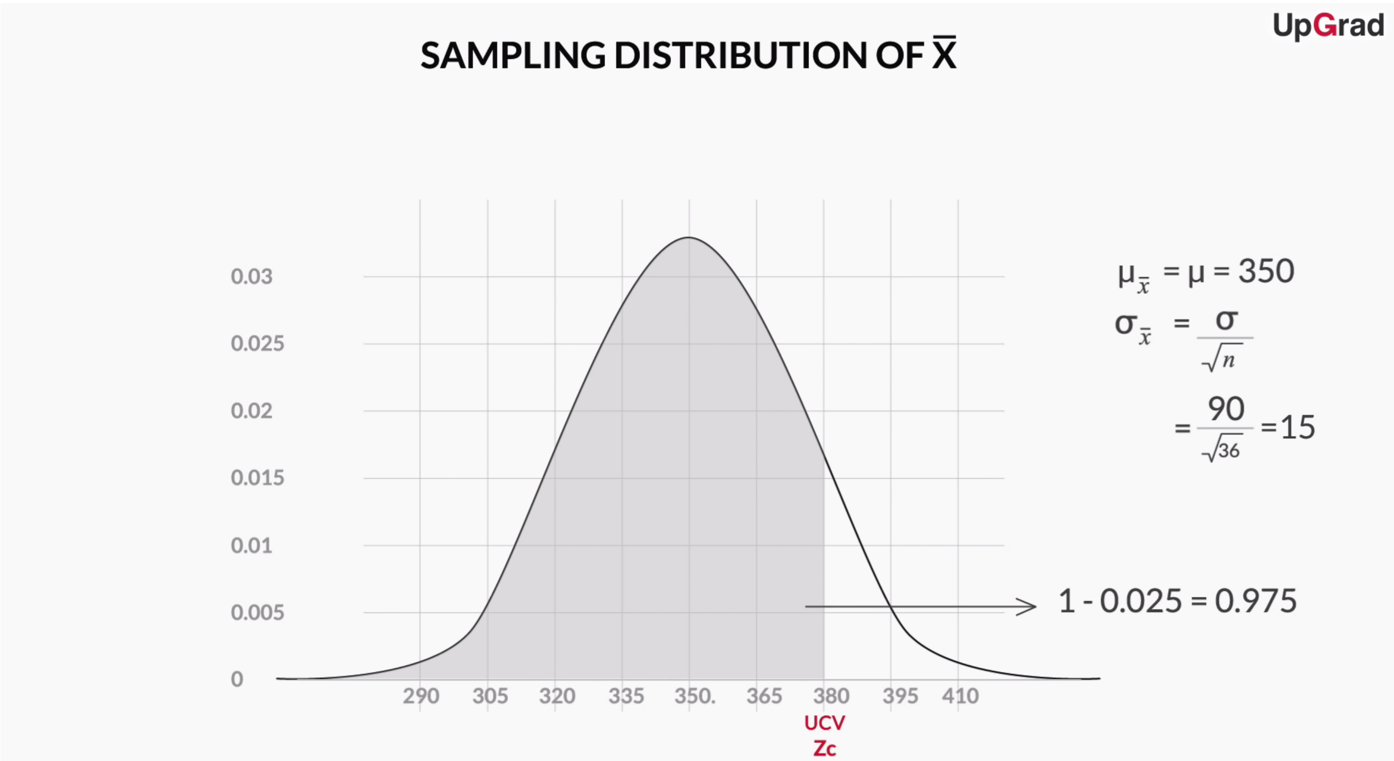
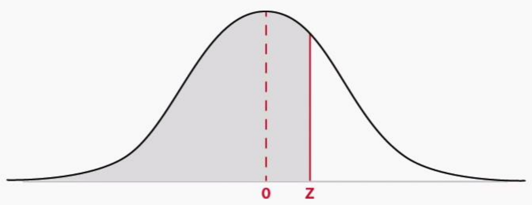


Figure 9 – Formulating Hypotheses for AC sales problem

Then you find the z-score of cumulative probability of UCV (Zc in this case).

TABLE OF STANDARD NORMAL PROBABILITY FOR POSITIVE Z-SCORE

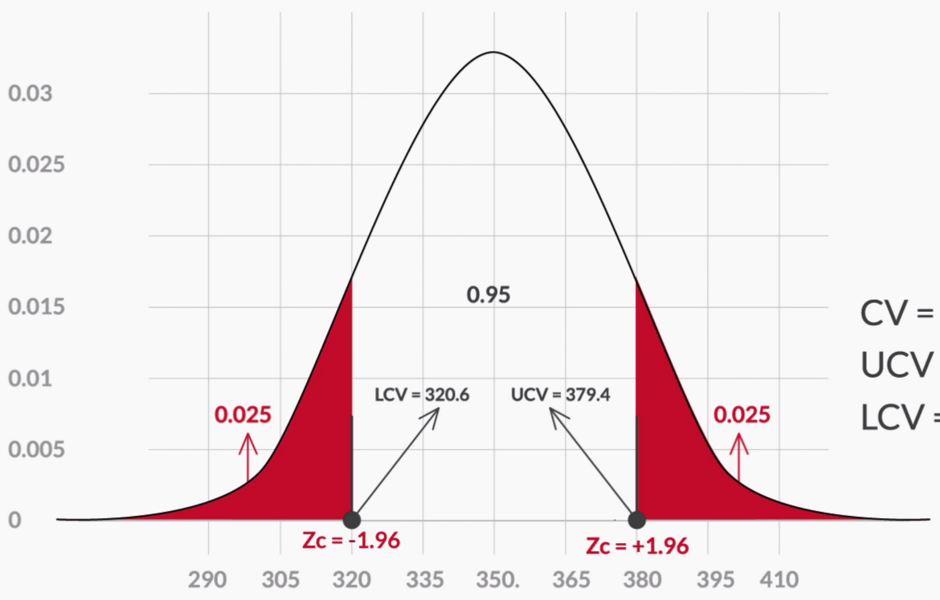
1.96



z	.00	.01	.02	.03	.04	.05	.06
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686
$\boxed{1.9}$	.9713	.9719	.9726	.9732	.9738	.9744	$\boxed{.9750}$

Figure 10 – Find z-value for cumulative probability of UCV

Then you calculate the critical values (UCV and LCV) from the value of Zc.



$$\mu_{\bar{x}} = \mu = 350$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{90}{\sqrt{36}} = 15$$

$$CV = \mu + (Zc \times \sigma_{\bar{x}})$$

$$UCV = 350 + 1.96 \times 15 = 379.4$$

$$LCV = 350 - 1.96 \times 15 = 320.6$$

Figure 11 – Finding the UCV and LCV

As sample mean lies is less than UCV and greater than LCV, i.e. it lies in the acceptance region,

Decision: Fail to reject the null hypothesis

Concepts in Hypothesis Testing - II

There are various methods similar to the critical value method to statistically make your decision about the hypothesis. One such method is the p-value method. This is an important method and is used more frequently in the industry.

p-value Method

What is p-value?

A P-value measures the strength of evidence in support of a null hypothesis. Suppose the test statistic in a hypothesis test is equal to *K*. The P-value is the probability of observing a test statistic as extreme as *K*, assuming the null hypothesis is true. If the P-value is less than the significance level, we reject the null hypothesis.

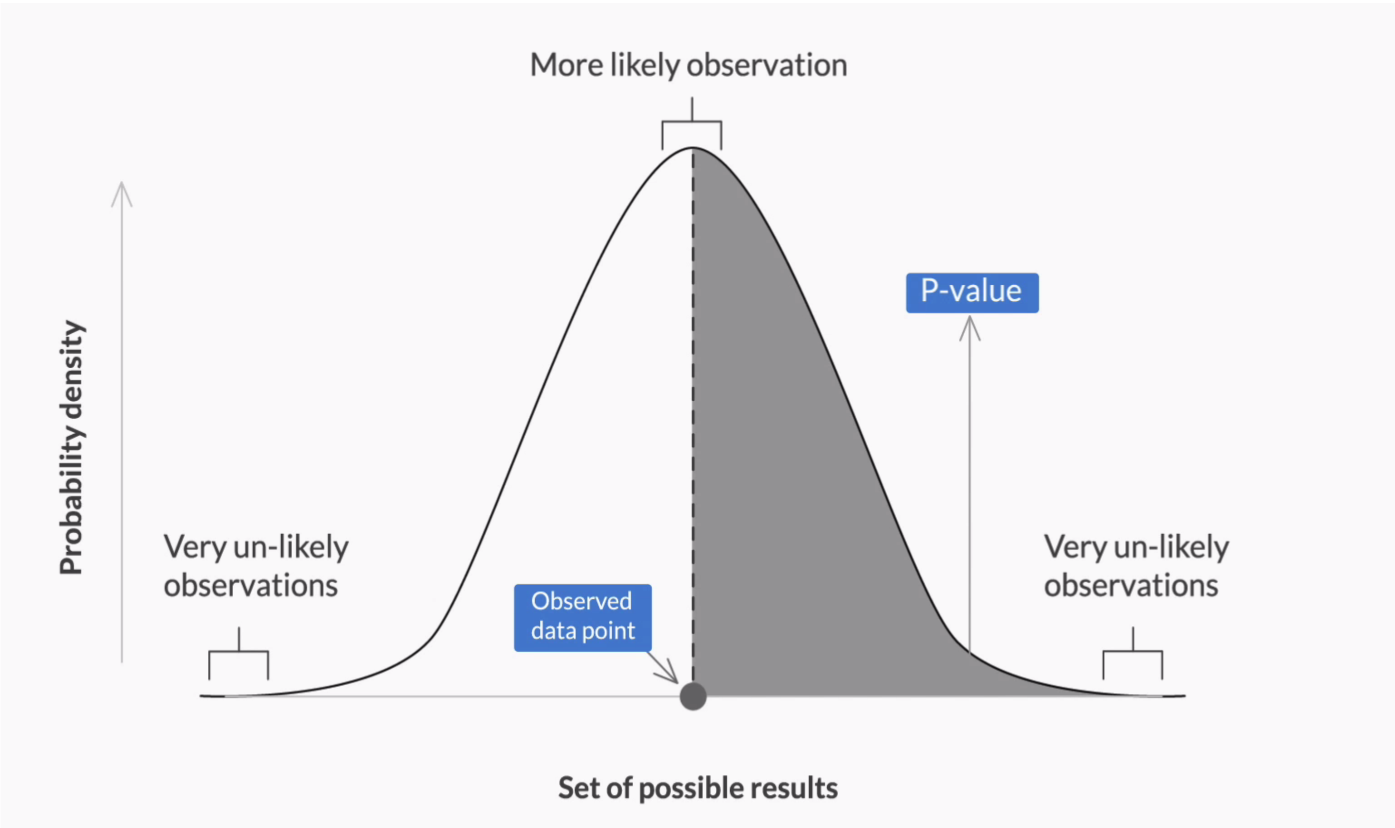


Figure 12 – Interpretation of p-value

After formulating the hypothesis, the steps you have to follow to **make a decision** using **the p-value method** are as follows:

1. Calculate the value of z-score for the sample mean point on the distribution
2. Calculate the p-value from the cumulative probability for the given z-score using the z-table
3. Make a decision on the basis of the p-value (multiply it by 2 for a two-tailed test) with respect to the given value of α (significance value).

To find the correct p-value from the z-score, first find the **cumulative probability** by simply looking at the z-table, which gives you the area under the curve till that point.

Situation 1: The sample mean is on the right side of the distribution mean (the z-score is positive)

Example: z-score for sample point = + 3.02

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026

Figure 13- z-table for positive z-scores

Cumulative probability of sample point = 0.9987

For one-tailed test → p = 1 - 0.9987 = 0.0013
For two-tailed test → p = 2 (1 - 0.9987) = 2 * 0.0013 = 0.0026

Situation 2: The sample mean is on the left side of the distribution mean (the z-score is negative)

Example: z-score for sample point = -3.02

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026

Figure 14 - z-table for negative z-scores

Cumulative probability of sample point = 0.0013

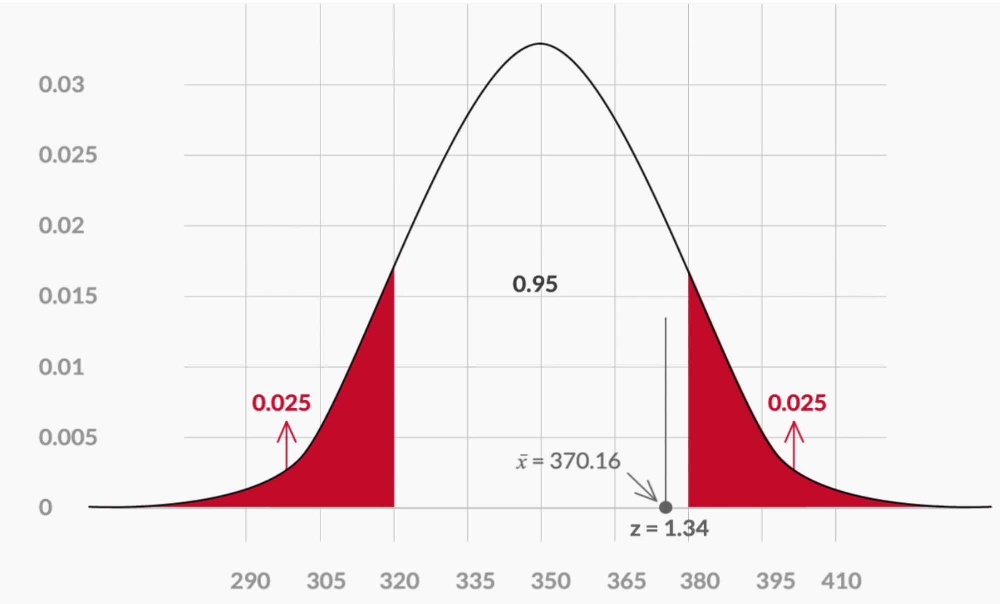
For one-tailed test $\rightarrow$ p = 0.0013
For two-tailed test $\rightarrow$ p = 2 * 0.0013 = 0.0026

p-value Method – Complete Example

Taking the same AC sales problem, hypotheses for the situation would remain the same.

Making a Decision

So you start by finding out the z-value for given sample mean.



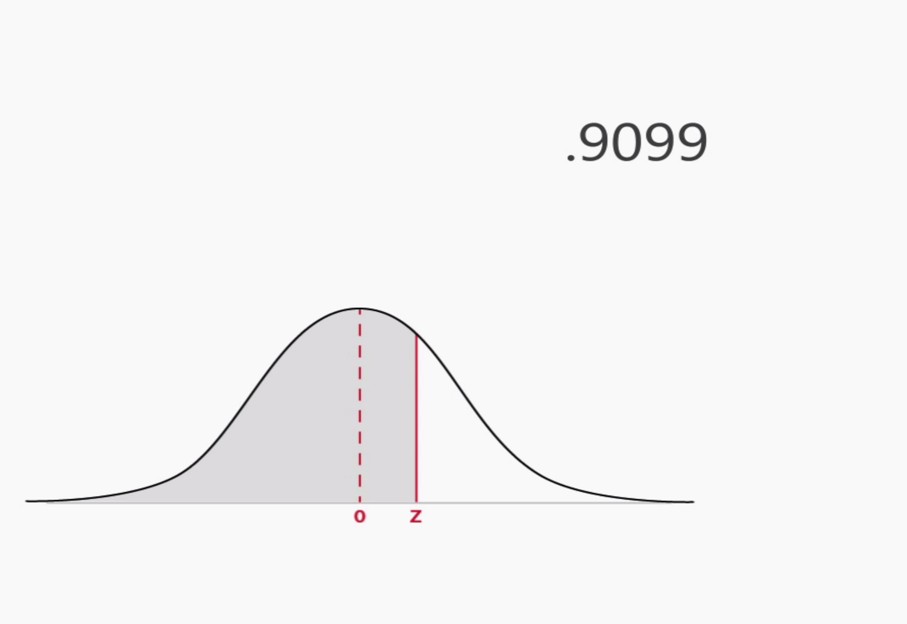
$$\mu_{\bar{x}} = \mu = 350$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{90}{\sqrt{36}} = 15$$

$$Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{370.16 - 350}{15} = 1.34$$

Figure 15 – Finding z-value for sample mean

Then you find the z-value for the calculated z-value for sample mean.



z	.00	.01	.02	.03	.04
0.0	.5000	.5040	.5080	.5120	.5160
0.1	.5398	.5438	.5478	.5517	.5557
0.2	.5793	.5832	.5871	.5910	.5948
0.3	.6179	.6217	.6255	.6293	.6331
0.4	.6554	.6591	.6628	.6664	.6700
0.5	.6915	.6950	.6985	.7019	.7054
0.6	.7257	.7291	.7324	.7357	.7389
0.7	.7580	.7611	.7642	.7673	.7704
0.8	.7881	.7910	.7939	.7967	.7995
0.9	.8159	.8186	.8212	.8238	.8264
1.0	.8413	.8438	.8461	.8485	.8508
1.1	.8643	.8665	.8686	.8708	.8729
1.2	.8849	.8869	.8888	.8907	.8925
1.3	.9032	.9049	.9066	.9082	.9099
1.4	.9192	.9207	.9222	.9236	.9251
1.5	.9332	.9345	.9357	.9370	.9382
1.6	.9452	.9463	.9474	.9484	.9495
1.7	.9554	.9564	.9573	.9582	.9591
1.8	.9641	.9649	.9656	.9664	.9671
1.9	.9713	.9719	.9726	.9732	.9738
2.0	.9772	.9778	.9783	.9788	.9793
2.1	.9821	.9826	.9830	.9834	.9838
2.2	.9861	.9864	.9868	.9871	.9875
2.3	.9893	.9896	.9898	.9901	.9904
2.4	.9918	.9920	.9922	.9925	.9927
2.5	.9938	.9940	.9941	.9943	.9945
2.6	.9953	.9955	.9956	.9957	.9959

Figure 16 – Finding cumulative probability for z-value

Then you find the p-value by the approach mentioned above.

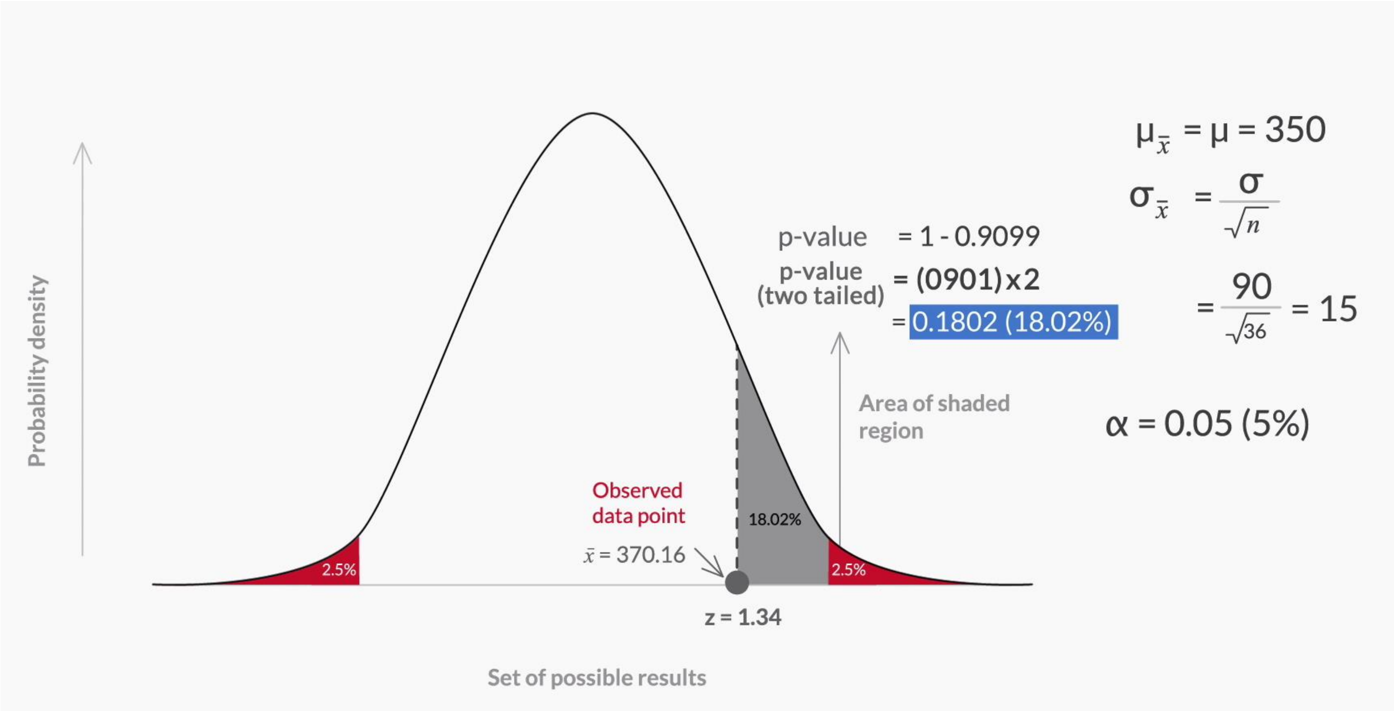


Figure 17 – Finding p-value from cumulative probability

As p-value (0.1802) is greater than the value of α (0.05),

Decision: Fail to reject the null hypothesis

Types of errors

There are two possible errors we can commit during hypothesis testing —

- type I error
- type II error.

The type I error occurs when the null hypothesis is true but we reject it, i.e. reject H_0 when it is true.

Example:

Just imagine, if the defendant is innocent of the murder, but is still convicted and given the death penalty, it would be a gross miscarriage of justice. For a case like this, the type I error should have a 0.001 probability, i.e. the jury should be convinced beyond reasonable doubt that the defendant is guilty, or an innocent man might go to the gallows. On the other hand, for a civil trial, say, for damages in a car accident, the type I error can have a larger margin like 0.49, i.e. upon a preponderance of the evidence.

The probability of type I error is denoted by alpha (α) and is usually 0.05 or 0.01, i.e. only a 5% or 1% chance. The type I error is also called the level of significance of the hypothesis test.

The type II error occurs when the null hypothesis is false but we fail to reject it, i.e. fail to reject H_0 when it is false.

Example:

If the defendant is guilty, but the jury acquits him, it would be a type II error. In practical terms, this is the most serious error you can make. If you let a murderer walk away, he might end up killing more people.

The probability of type II error is denoted by beta (β).

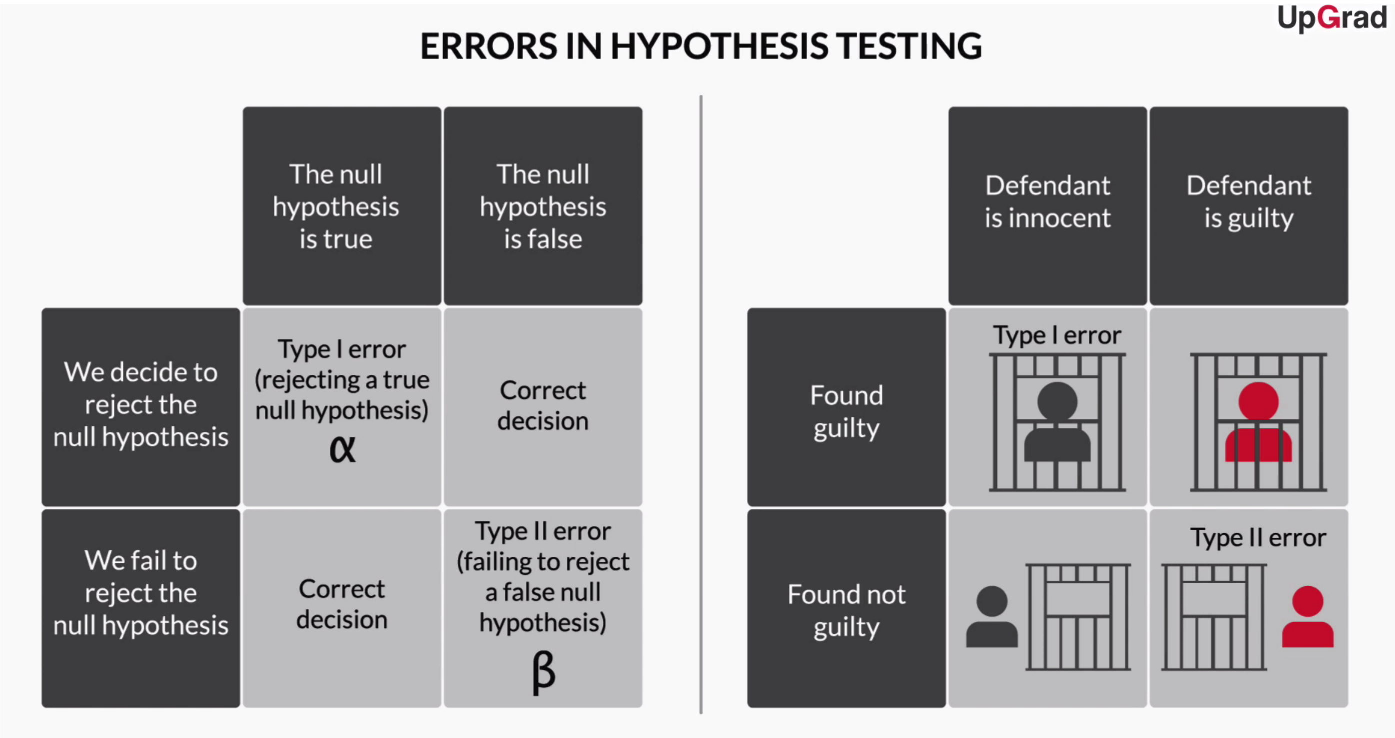


Figure 18 – Types of errors in Hypothesis Testing

T - Distribution

What is a T-distribution?

A T-distribution (or Student T distribution) is similar to the normal distribution in many cases; for example, it is symmetrical about its central tendency. However, it is shorter than the normal distribution and has a flatter tail, which would eventually mean that it has a larger standard deviation.

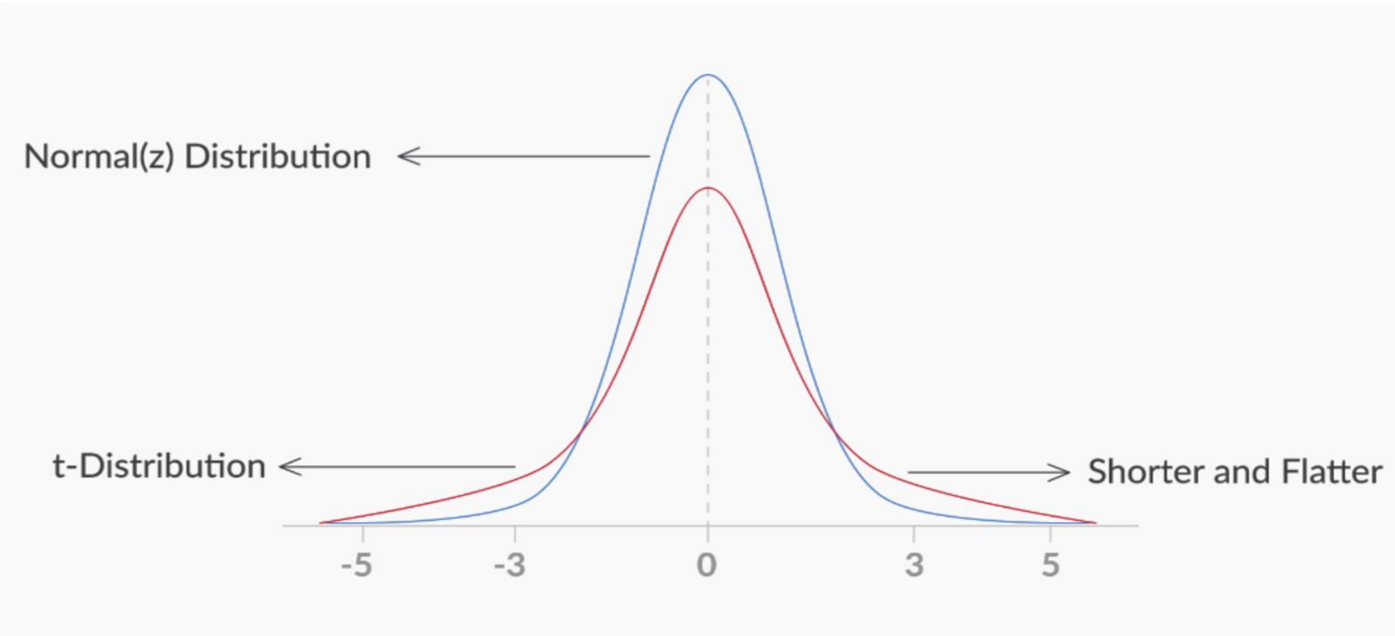


Figure 19 - t-distribution vs standard normal distribution

At a sample size beyond 30, the t-distribution becomes approximately equal to the normal distribution.

Each t-distribution is distinguished by what statisticians call degrees of freedom, which are related to the sample size of the data set. If your sample size is n, the degrees of freedom for the corresponding t-distribution is n -1. For example, if your sample size is 10, you use a t-distribution with 10 -1 or 9 degrees of freedom, denoted t_9. Smaller sample sizes have flatter t-distributions than larger sample sizes. And as you may expect, the larger the sample size is, and the larger the degree of freedom, the more the t-distribution looks like a standard normal distribution or the Z-distribution.

When T-Distribution is used?

The most important use of the t-distribution is that you can approximate the value of the **standard deviation of the population (σ)** from the **sample standard deviation (s)**. However, as the sample size increases more than 30, the t-value tends to be equal to the z-value. Thus, if you want to summarise the decision-making in a flowchart given in the following figure 20, this is what you would get.

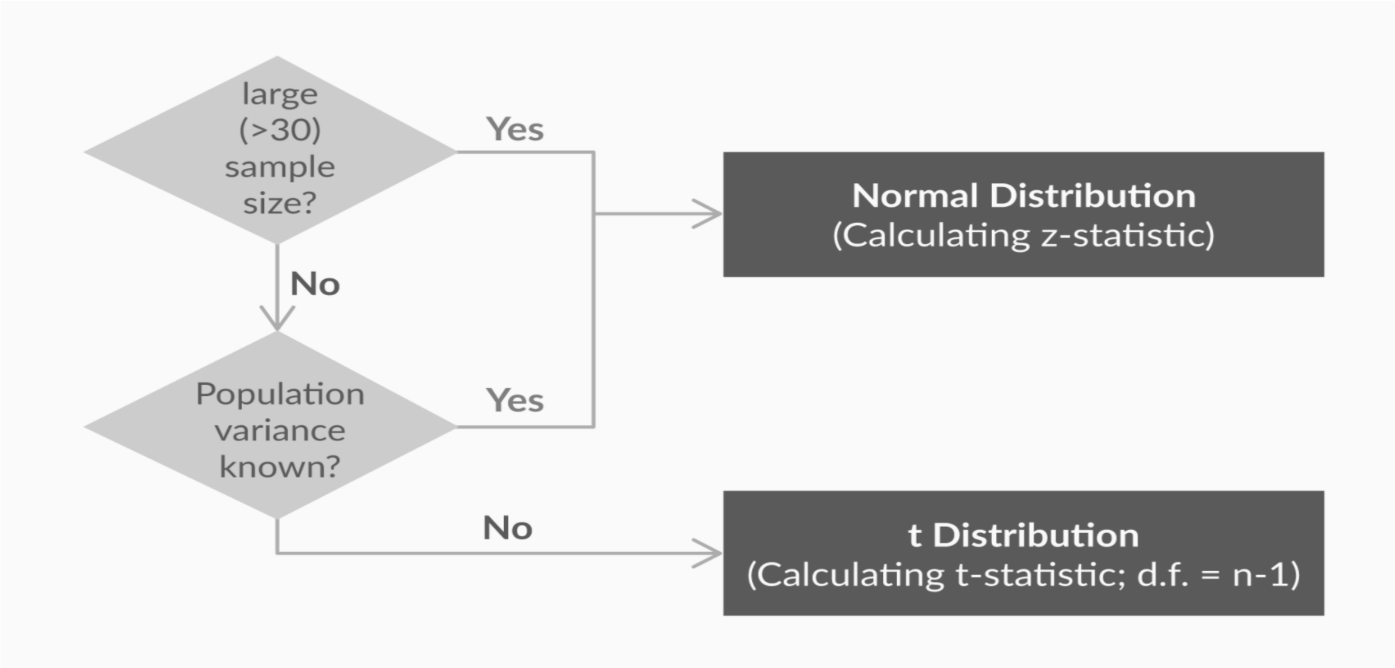


Figure 20 - Flowchart for deciding the test

If you look at how the method of **making a decision** changes if you are using the sample's standard deviation instead of the population's. If you recall the critical value method, the first step is as follows:

1. Calculate the value of Zc from the given value of α (significance level). Take it as 5% if not specified in the problem.

So, to find Zc, you would use the **t-table** instead of the z-table. The **t-table** contains values of Zc for a given degree of freedom and value of α (significance level). Zc, in this case, can also be called as t-statistic (critical).

Practically you would not need to refer to the z or the table when doing hypothesis testing in the industry. Going forward when you need to do hypothesis testing in demonstrations of Excel or R, you would use the term **t-test** since that is mostly performed in the industry. All calculations and results of a t-test are same as the z-test whenever the sample size ≥ 30.

Two-sample Mean Test

Two-sample mean test - paired is used when your sample observations are from the same individual or object. During this test, you are testing the same subject twice. For example, if you are testing a new drug, you would need to compare the sample before and after the drug is taken to see if the results are different.

Two-sample mean test - unpaired is used when your sample observations are independent. During this test, you are not testing the same subject twice. For example, if you are testing a new drug, you would compare its effectiveness to that of the standard available drug. So, you would take a sample of patients who consumed the new drug and compare it with those who consumed the standard drug.

Two-sample Proportion Test

Two-sample proportion test is used when your sample observations are categorical, with two categories. It could be True/False, 1/0, Yes/No, Male/Female, Success/Failure etc.

For example, if you are comparing the effectiveness of two drugs, you would define the desired outcome of the drug as the success. So, you would take a sample of patients who consumed the new drug and record the number of successes and compare it with successes in another sample who consumed the standard drug.

A/B Testing

A/B testing is a direct industry application of the two-sample proportion test sample.

While developing an e-commerce website, there could be different opinions about the choices of various elements, such as the shape of buttons, the text on the call-to-action buttons, the colour of various UI elements, the copy on the website, or numerous other such things.

Often, the choice of these elements is very subjective and is difficult to predict which option would perform better. To resolve such conflicts, you can use A/B testing. **A/B testing** provides a way for you to test two different versions of the same element and see which one performs better.

A/B testing is entirely based on the two-sample proportion test, as the two-sample proportion test is used when you want to compare the proportions of two different samples. You can use various tools to conduct A/B testing (or two-sample proportion test) like R, Optimizely etc.