Probability Experiment

- Let us consider one example where there is a bag of multiple balls comprising both Blue and red balls(3 B, 2 R). Suppose if we want to find out the probability of getting four red balls when a person tried selecting the ball four times from the bag. Each time she picks up the ball from the bag, notes down what is the color of the ball and she would again place the ball back to the bag. She continues this activity four times and each time notes down the color of the ball she picked.

- What is the probability that she picked red ball each time i.e. all the picked up balls being red color.

- We can theoretically find out this probability or we can do this test practically to find out the real probability.

- Let us do this test practically. We have arranged some 75 people to do this test. Each person picked up the ball from the bag four times. Each time they would place the ball back to the bag after selecting and again picks up the ball i.e the total number of balls available in the bag is going to be same each time you pick a ball.

Out of 75 people who took the test, 2 persons picked up all 0 red balls(All blue balls), 12 people picked up 1 red ball, 26 people picked up 2 red balls, 25 people picked up 3 red balls and 10 people picked up 4 red balls.

So probability would be as follows:

P(0) = 2/75 = 0.027
P(1) = 12/75 = 0.16
P(2) = 26/75= 0.347
P(3) = 25/75 = 0.33
P(4) = 10/75 = 0.133

EXPECTED VALUE OF MONEY WON AFTER 1 GAME

X can take two values: +150 and -10

EXPECTED VALUE OF MONEY WON AFTER 1 GAME

X can take two values: +150 and -10

$P(X = +150) = P(4 \text{ red balls}) = 0.133$

$P(X = -10) = P(0, 1, 2 \text{ or } 3 \text{ red balls}) = 0.027 + 0.160 + 0.347 + 0.333 = 0.867$

EXPECTED VALUE OF MONEY WON AFTER 1 GAME

X can take two values: +150 and -10

P(X = +150) = P(4 red balls) = 0.133

P(X = -10) = P(0, 1, 2 or 3 red balls) = 0.027 + 0.160 + 0.347 + 0.333 = 0.867

So, EV = (150 x 0.133) + (-10 x 0.867) = +11.28

Probability Distribution

x	P(x)
0	0.027
1	0.160
2	0.347
3	0.333
4	0.133

Figure 5 - Tabular Form of Probability Distribution

Expected Value

$$EV(X) = x_1 * P(X = x_1) + x_2 * P(X = x_2) + \ldots\ldots\ldots\ldots\ldots\ldots + x_n * P(X = x_n)$$

Another way of writing this is

$$EV(X) = \sum_{i=1}^{i=n} x_i * P(X = x_i)$$

- **Random Variable (X)** : A **random variable**, usually written **X**, is a **variable** whose possible values are numerical outcomes of a **random** phenomenon. There are two types of **random variables**, discrete and continuous

- **Probability Distribution**: A probability distribution is a statistical function that describes all the possible values and likelihoods that a random variable can take within a given range. This range will be between the minimum and maximum statistically possible values

- **Expected Value (EV)**: calculated by multiplying each of the possible outcomes by the likelihood each outcome will occur, and summing all of those values

Lets say probability of getting the red ball is denoted by p, then the probability of getting blue ball becomes 1-p.

P(x=4) = P * P * P * P = P^4
P(x=3) = P * P * P * (1-P) = 4 P^3 (1-p)

.

.

.

Goes on

PROBABILITY DISTRIBUTION FOR GENERAL PROBABILITY p

x	P(X=x)
0	$(1-p)^4$
1	$4p(1-p)^3$
2	$6p^2(1-p)^2$
3	$4p^3(1-p)$
4	p^4

BINOMIAL PROBABILITY DISTRIBUTION

x	P(X=x)
0	${}^{n}C_{0}(p)^{0}(1-p)^{n}$
1	${}^{n}C_{1}(p)^{1}(1-p)^{n-1}$
2	${}^{n}C_{2}(p)^{2}(1-p)^{n-2}$
3	${}^{n}C_{3}(p)^{3}(1-p)^{n-3}$
.	.
.	.
.	.
.	.
n	${}^{n}C_{n}(p)^{n}(1-p)^{0}$

So, the formula for finding **binomial probability is** given by –

$$P(X = r) = {}^nC_r (p)^r (1 - p)^{n-r}$$

Where n is no. of trials, p is probability of success and r is no. of successes after n trials.

However there are some conditions that need to be followed in order for us to be able to apply the formula.

1.Total number of trials is fixed at n

2.Each trial is binary, i.e., has only two possible outcomes - success or failure

3.Probability of success is same in all trials, denoted by p

Cumulative Probability

In the previous example, we only discussed the probability of getting an exact value. For example, we know the probability of X = 4 (4 red balls). But what if the house wants to know the probability of getting $\leq$ 3 red balls, as the house knows that for $\leq$ 3 red balls, the players will lose and they will make money?

CUMULATIVE PROBABILITY DISTRIBUTION

x	$F(x) = P(X \leq x)$
0	0.0256
1	0.1782
2	0.5238
3	0.8694
4	1.0000

- What is the probability of people buying a product in a retail store for exactly worth of Rs. 1680.5 in a day?

- What is the probability of people buying a product in a retail store for exactly worth of Rs. 1680.5 in a day?
- Ans: 0 . Probability for a continuos variable is always zero. For continuos variable we talk probability in range

Product revenue	Probability
200-500	0.25
500-1000	0.20
1000-1500	0.10
1500-2000	0.2
2000-2500	0.15
2500-3000	0.10

- **Continuous Variables - Probability**
- From the given probability data, can you find the cumulative probability for X = 200, i.e. P(X≤200)?

?

- From the given probability data, can you find the cumulative probability for X = 2500, i.e. P(X≤1500)?

?

- **Continuous Variables - Probability**
- From the given probability data, can you find the cumulative probability for X = 200, i.e. P(X$\leq$200)?

Ans : 0

- From the given probability data, can you find the cumulative probability for X = 2500, i.e. P(X$\leq$1500)?

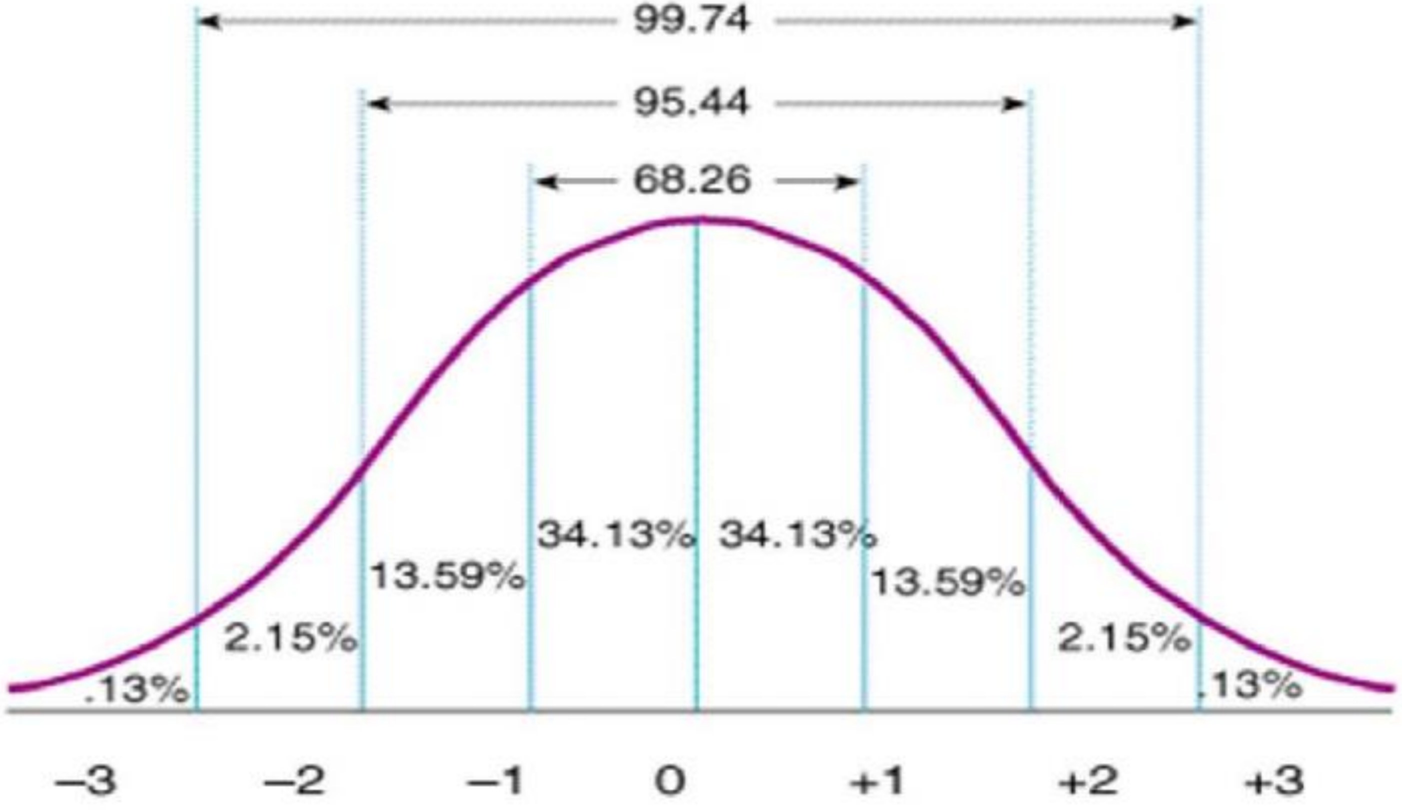
Ans: 0.9

Product revenue	Cumulative Probability
200-500	0.25
500-1000	0.45
1000-1500	0.55
1500-2000	0.75
2000-2500	0.9
2500-3000	1

- Considering you work at a retail store and you want to analyze the revenue of a product on week days. So, you create a list of revenue on each week day. Based on this data, they created a cumulative probability distribution for X, where X = revenue on weekdays.
- Now, based on the data, you conclude that the cumulative probability, $F(1680.5) = 0.3$. In this case, which of the following statements is correct?
- $P(X< 1680.5) = 0.3$
- $P(X\leq 1680.5) = 0.3$

(Remember that revenue is a continuous variable.)

Standard Normal Distribution



- Normal distribution with $\mu = 0$ and SD = 1

- All data that is normally distributed follows the **1-2-3 rule**. This rule states that there is a -
- **68%** probability of the variable lying **within 1 standard deviation** of the mean
- **95%** probability of the variable lying **within 2 standard deviations** of the mean
- **99.7%** probability of the variable lying **within 3 standard deviations** of the mean

FINDING PROBABILITY FOR NORMAL VARIABLE X

Mean (μ) = 35
Standard deviation (σ) = 5

$P(25 < X < 45) = ?$

Mean (μ) = 35
Standard deviation (σ) = 5

$P(25 < X < 45) = P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 95\%$

Normal Probability Distribution

- We do this using the following formula

$$z = \frac{x - \mu}{\sigma}$$

x = the normally distributed random variable of interest
μ = the mean for the normal distribution
σ = the standard deviation of the normal distribution
z = the z-score (the number of standard deviations between x and μ)

Basically, it tells you **how many standard deviations away from the mean** your random variable is. As you just saw, you can find the cumulative probability corresponding to a given value of Z, using the **Z table**

Test1

The regulatory authority selects a random tablet from Batch Z2. Based on previous knowledge, you know that Batch Z2 has a mean paracetamol level of 510 mg, and its standard deviation is 20 mg.

What is the probability that the tablet that has been selected by the authority has a paracetamol level below 550 mg?

Test2

Let's define X as the amount of paracetamol in the selected tablet. Now, X is a normally distributed random variable, with mean μ = 510 mg and standard deviation σ = 20 mg.

Now, you have to find the probability of X being less than 550, i.e. $P(X<550)$.

Converting this to Z, you get $P(X<550) = P(Z<\{550-510\}/20) = P(Z<2) = 0.977$, or 97.7%.

CENTRAL LIMIT THEORY

Sample vs Population

A population is a collection of people, items, or events about which you want to make inferences. It is not always convenient or possible to examine every member of an entire population. For example, it is not practical to take the opinion of all the voters in an poll survey. It is possible, however, to take the opinion of some people taken from that population. This subset of the population is called a sample.

A sample is a subset of people, items, or events from a larger population that you collect and analyse to make inferences. To represent the population well, a sample should be randomly collected and adequately large.
If the sample is random and large enough, you can use the information collected from the sample to make inferences about the population.

Central Limit theorem exactly talks about applying this sample information to the overall population

Sample vs Population

The central limit theorem will be useful whenever we need to find out the mean of the population given the statistics of the sampling distribution. It states that given a distribution with a mean μ and variance σ^2, the sampling distribution of the mean approaches a normal distribution with a mean (μ) and a variance σ^2/N as N, the sample size, increases. The amazing and counter-intuitive thing about the central limit theorem is that no matter what the shape of the original distribution, the sampling distribution of the mean approaches a normal distribution. Furthermore, for most distributions, a normal distribution is approached very quickly as N increases. Keep in mind that N is the sample size for each mean and not the number of samples.

In other words, if we repeatedly take independent random samples of size n from any population, then when n is large, the distribution of the sample means will approach a normal distribution.

How large is large enough? Generally speaking, a sample size of 30 or more is considered to be large enough for the central limit theorem to take effect. The closer the population distribution is to a normal distribution, the fewer samples needed to demonstrate the theorem. Populations that are heavily skewed or have several modes may require larger sample sizes

Sample vs Population

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$$

Where:

σ^2 = variance of the population (pronounced sigma squared)

x_i = the measurement of each data unit in the population

μ = the population mean

n = the size of the population

Sample vs Population

Standard Deviation - Sample

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Where:

s = the standard deviation of the sample

x_i = the measurement of each data unit in the sample

$\bar{x}$ = the sample mean

n = the size of the sample (the number of data values)

Sample vs Population

Notation

The mean of the sample means

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

The standard deviation of sample mean

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

(often called the standard error of the mean)

Sample vs Population

Central Limit Theorem

1. The distribution of sample $\bar{x}$ will, as the sample size increases, approach a normal distribution.
2. The mean of the sample means is the population mean μ.
3. The standard deviation of all sample means is $\sigma / \sqrt{n}$.

Estimating Mean Using CLT

Suppose if we have an example of employee population of 30000 and sampling distribution of 100 employees with sampling mean X = 36.6. Then population mean, i.e. Average age of all 30,000 employees \bar{X} = 36.6 (sample mean) +/- some margin of error.

FINDINGS FROM THE SAMPLING DISTRIBUTION

$P(36.6 - 2 < \mu < 36.6 + 2) = 95.4\%$

Probability associated with the claim is called **confidence level** (Here it is 95.4%)

Maximum error made in sample mean is called **margin of error** (Here it is 2 minutes)

Final interval of values is called **confidence interval** {Here it is the range – (34.6, 38.6)}

Estimating Mean Using CLT

So, to summarise, let's say you have a sample with sample size n, mean $\bar{X}$ and standard deviation S.

Now, the **y% confidence interval** (i.e. the confidence interval corresponding to y% confidence level) for μ would be given by the range:

$$\text{Confidence interval} = \left(\bar{X} - \frac{Z^*S}{\sqrt{n}}, \bar{X} + \frac{Z^*S}{\sqrt{n}}\right),$$

where, **Z* is the Z-score associated with a y% confidence level**. In other words, the population mean and sample mean differ by a **margin of error** given by $\frac{Z^*S}{\sqrt{n}}$.

Some commonly used Z* values are given below:

Confidence Level	Z*
90%	±1.65
95%	±1.96
99%	±2.58

Figure 6 - Z* Values for Commonly Used Confidence Levels

Estimating Mean Using CLT

You want to take some samples, measure the amount of paracetamol, and test if the manufacturing process is running successfully. You have the resources and time to take a sample of 100 tablets and measure the paracetamol content in each.

For the 100 tablets sampled by you, you find that the mean paracetamol content is 530 mg and the standard deviation is 100 mg.

Now, you want to know what the average content is for all the tablets in the plant. You are thinking of reporting the average as a confidence interval, for which you are 95% confident.

Estimating Mean Using CLT

What is the MOE (margin of error) for 95% confidence level?

Estimating Mean Using CLT

What is the MOE (margin of error) for 95% confidence level?

If X is the defined as the paracetamol content, then for this sample of X, sample mean ¯XX¯ *= 530 mg, sample standard deviation S = 100 mg and sample size n = 100. Also, for 95% confidence interval, Z* is 1.96. Now, you know that the margin of error =* Z∗S√n=1.96∗100√100Z∗Sn=1.96∗100100 *= 19.6*

Estimating Mean Using CLT

What is the confidence interval for 90% confidence level?

Estimating Mean Using CLT

What is the confidence interval for 90% confidence level?

As you know, sample mean ¯XX¯ *= 530, S = 100 and n = 100. Also, for 90% confidence interval, Z* is 1.65. Now, you know the confidence interval is* (¯X−Z∗S√n,¯X+Z∗S√n)(X¯−Z∗Sn,X¯+Z∗Sn). *Putting in the values, you can calculate the confidence interval as (513.5, 546.5).*

HYPOTHESIS TESTING

Hypothesis testing

is used to confirm your conclusion (or hypothesis) about the population parameter . Through hypothesis testing, you can determine whether there is enough evidence to conclude if the hypothesis about the population parameter is true or not.

The actual test begins by considering two **hypotheses**.
They are called the **null hypothesis** and the **alternative hypothesis**. These hypotheses contain opposing viewpoints.

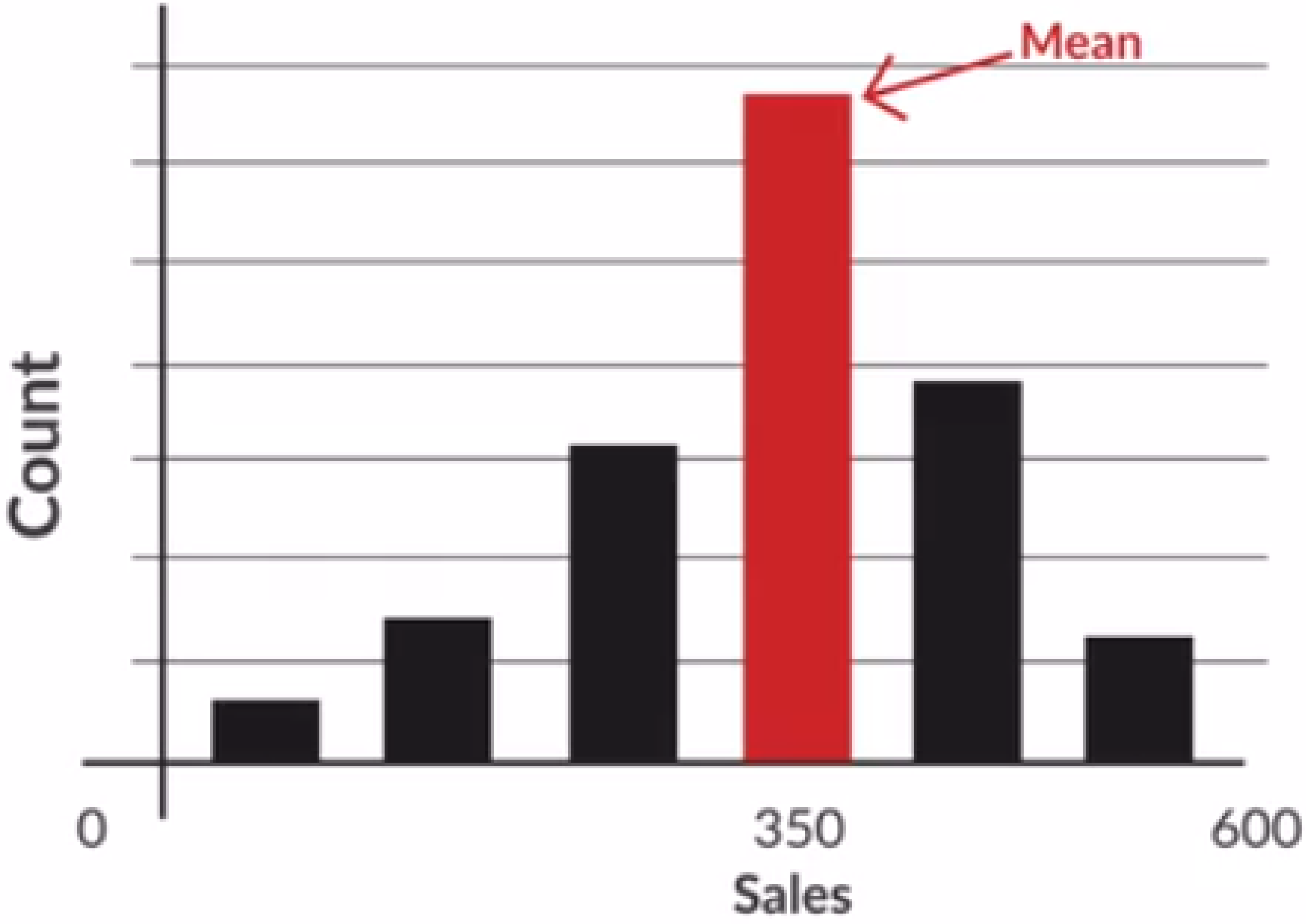
H_0: **The null hypothesis:** It is a statement about the population that either is believed to be true or is used to put forth an argument unless it can be shown to be incorrect beyond a reasonable doubt.

H_a: **The alternative hypothesis:** It is a claim about the population that is contradictory to H_0 and what we conclude when we reject H_0.

In the Beverages example, if you fail to reject the null hypothesis, what can you conclude from this statement?

The null hypothesis is that the average fertilizer content is less than or equal to 2.5 ppm. Since you fail to reject the null hypothesis, you can conclude that Beverages do not contain excess fertilizer. Please note that fertilizer you can only fail to reject the null hypothesis, you can never accept the null hypothesis

fail to reject the null hypothesis # accepting the null hypothesis


Mean
Count
0
350
600
Sales

The null hypothesis always has the following signs: = OR ≤ OR ≥

The alternate hypothesis always has the following signs: ≠ OR > OR <

The average commute time for an me to and from institute is at least 15 minutes.
What will be the null and alternate hypotheses in this case if the average time is represented by μ?

The null hypothesis is always formulated by either = or ≤ or ≥ whereas the alternate hypothesis is formulated by ≠ or > or <. In this case, the average time taken was greater than or equal to 35 minutes. So, that becomes the null hypothesis. Less than 35 minutes becomes the alternate hypothesis.

MAKING A DECISION

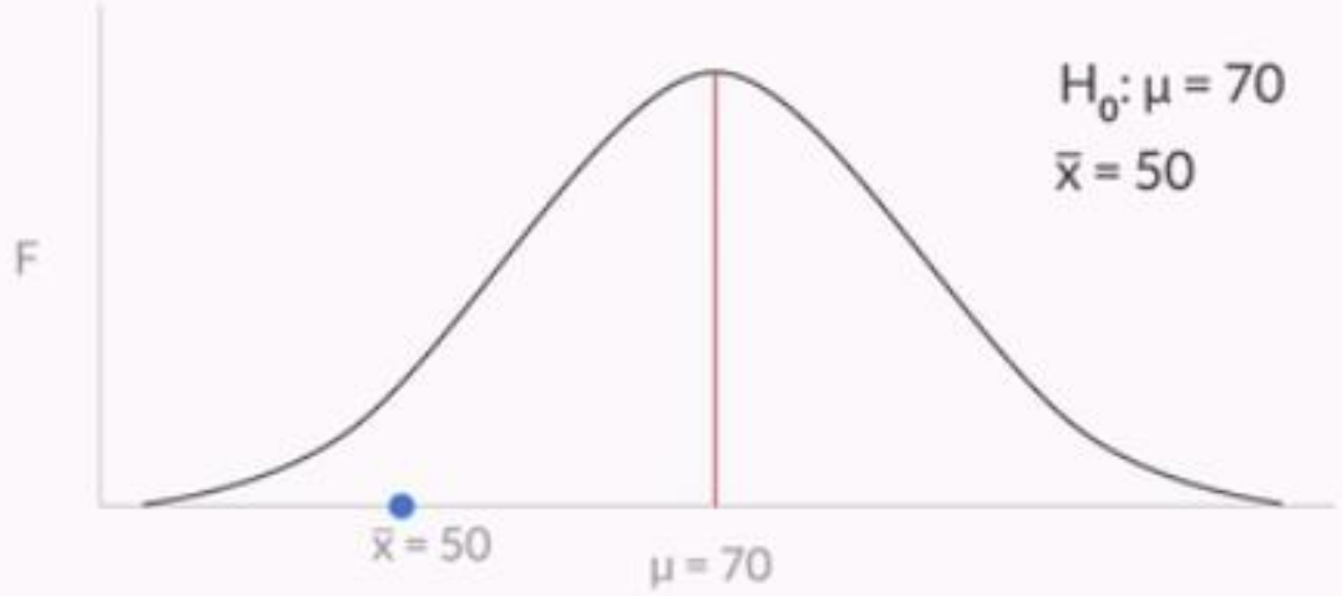
Ram's claim: Average score in basket ball = 70

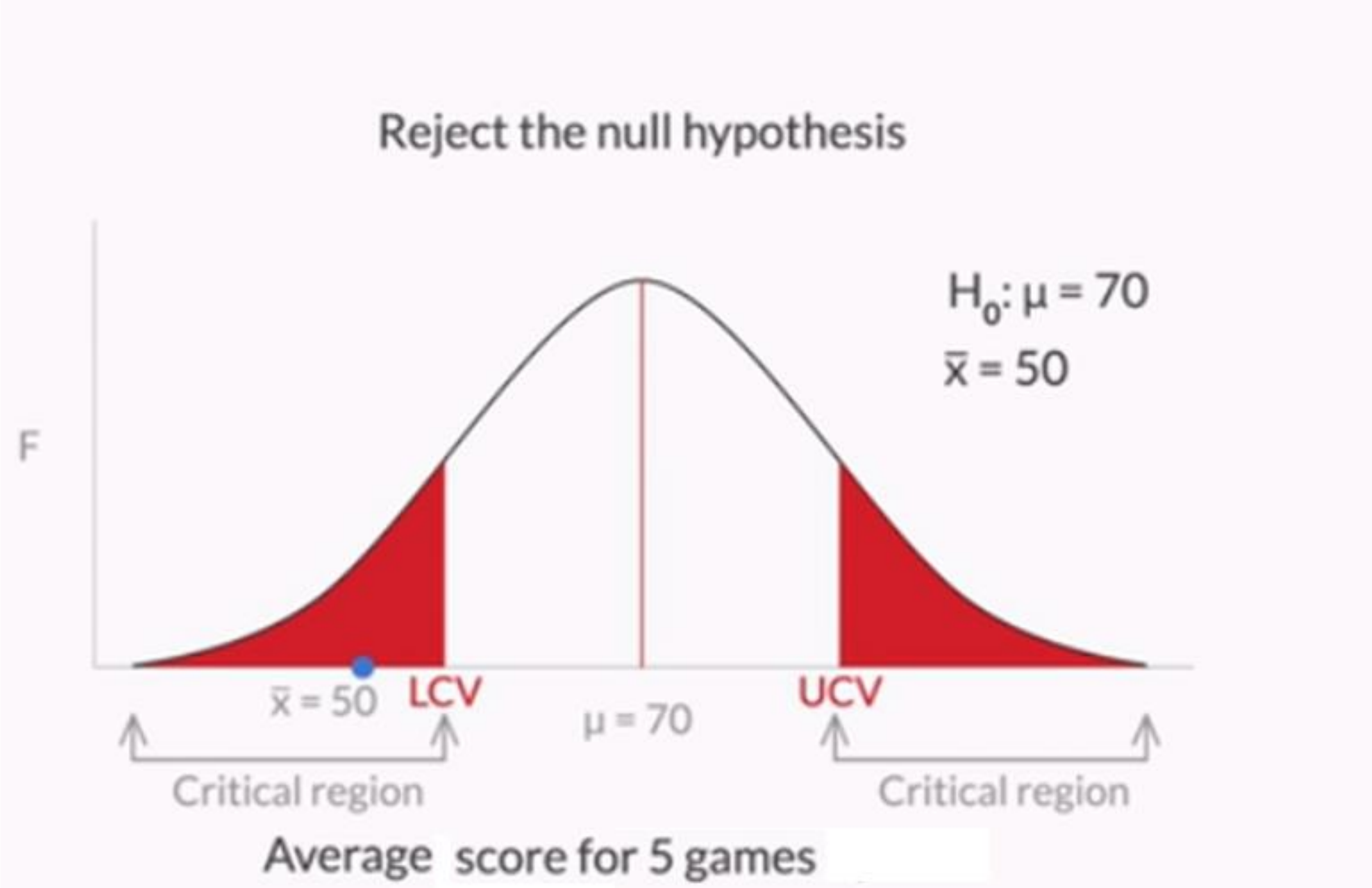
Over 5 games of basket ball:

1) Avg score = 20 → Less likely to believe his claim

2) Avg score = 65 → More likely to believe his claim

MAKING A DECISION

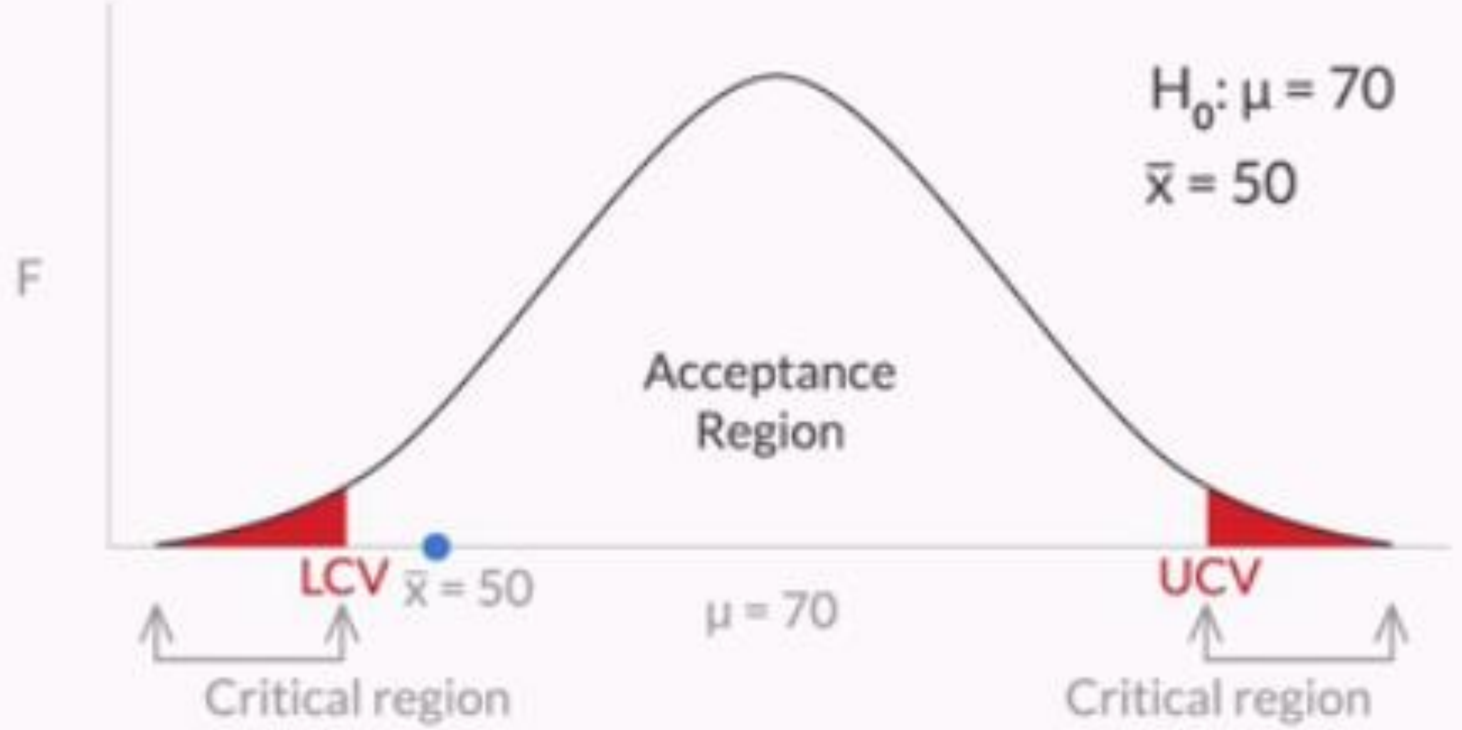




Reject the null hypothesis
H_0: μ = 70
x̄ = 50
F
x̄ = 50
LCV
μ = 70
UCV
Critical region
Critical region
Average score for 5 games

MAKING A DECISION

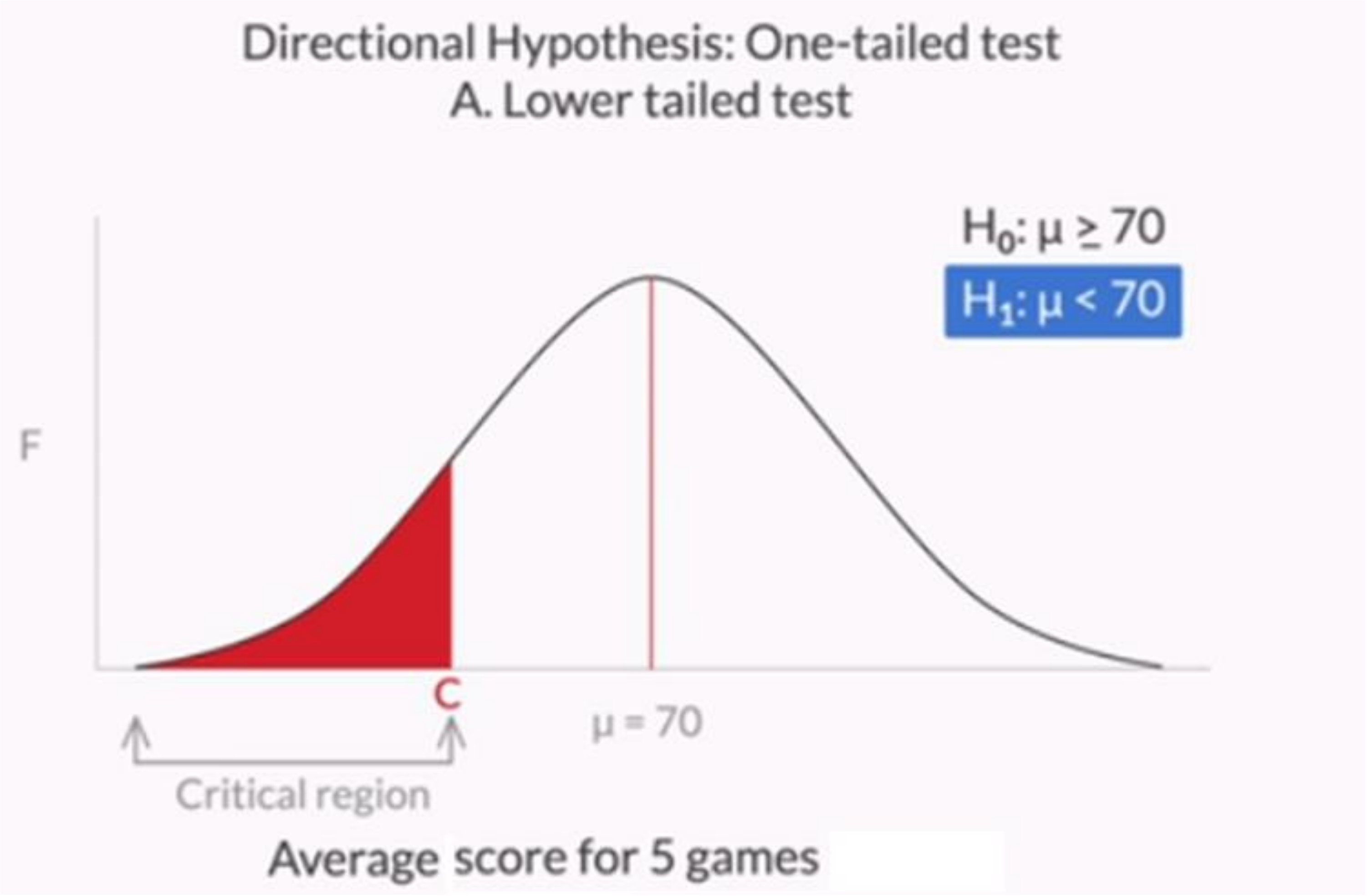
Fail to reject the null hypothesis



Making a decision

If your sample mean lies in the acceptance region, then

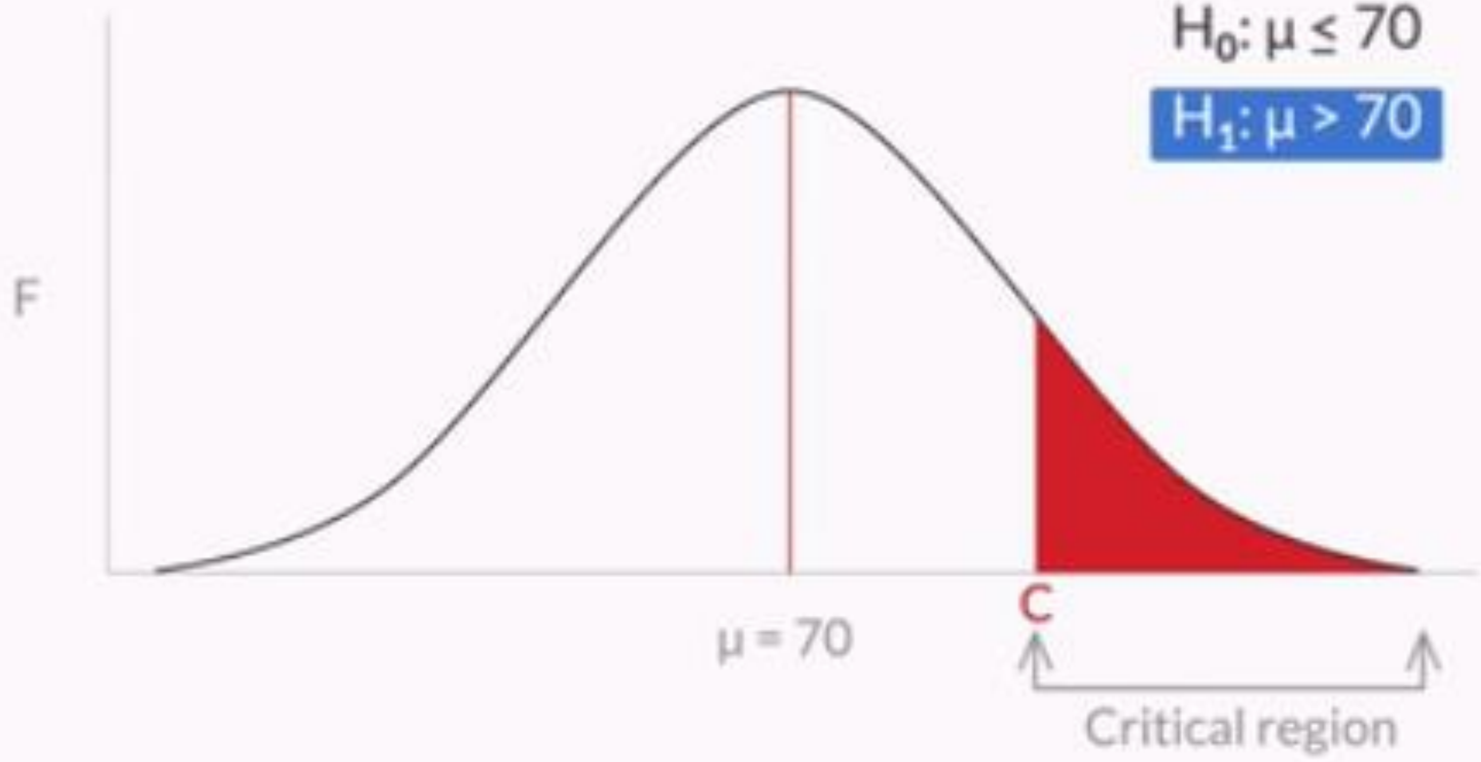
If your sample mean lies in the acceptance region, you fail to reject the null hypothesis because it is not beyond the critical point and you can consider that sample mean is equal to the population mean statistically.


Directional Hypothesis: One-tailed test
A. Lower tailed test
H_0: μ ≥ 70
H_1: μ < 70
F
C
μ = 70
Critical region
Average score for 5 games

MAKING A DECISION

Directional Hypothesis: One-tailed test

B. Upper tailed test



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.**

$\neq$ 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

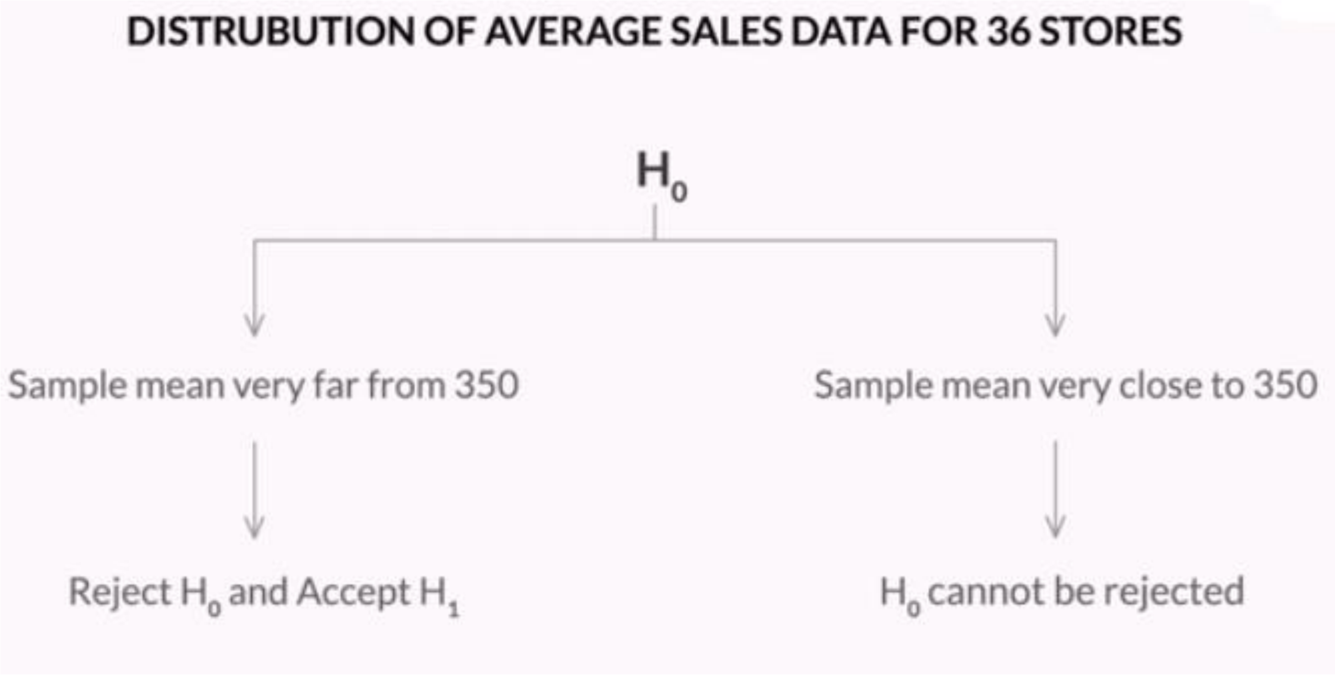
$>$ in H_1 → Upper-tailed test → Rejection region on **right side** of distribution

The average commute time for me to and from institute is at least 15 minutes.

If this hypothesis has to be tested, select the type of the test and the location of the critical region

For this situation, the hypotheses would be formulated as H_0: $\mu \geq 15$ minutes and H_1: $\mu < 15$ minutes. As $<$ sign is used in alternate hypothesis, it would be a lower-tailed test and the rejection region would be on the left side of the distribution.

In our previous example of no of umbrellas in rainy season



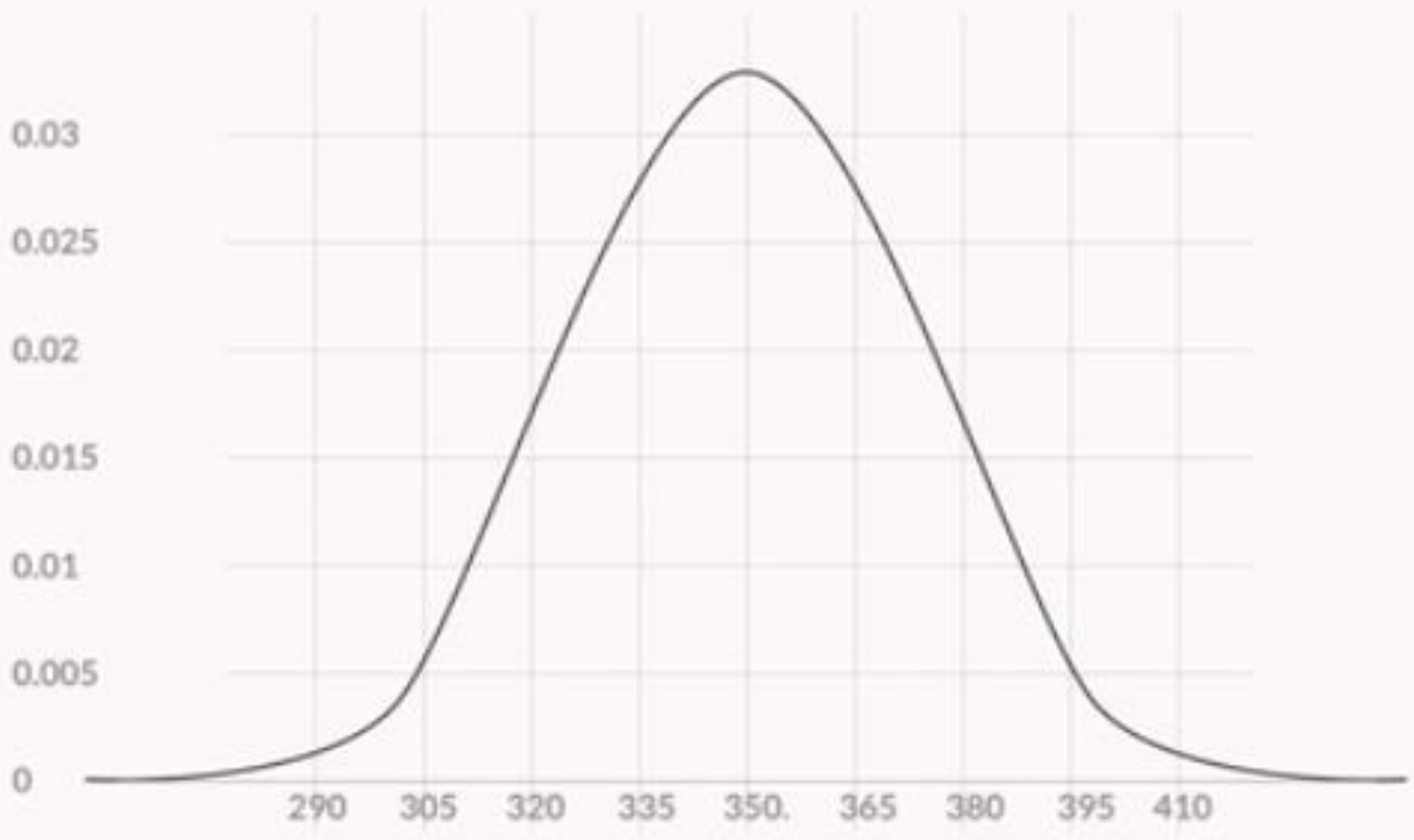
Lets consider:

What will be the standard deviation of the **distribution of sample means** if the population has a mean of 350 and a standard deviation of 90

Then it gives:

a sample mean of 370.16, and a sample size of 36?

SAMPLING DISTRIBUTION OF $\overline{X}$



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

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

$$= \frac{90}{\sqrt{36}}$$

$$\sigma = 15$$

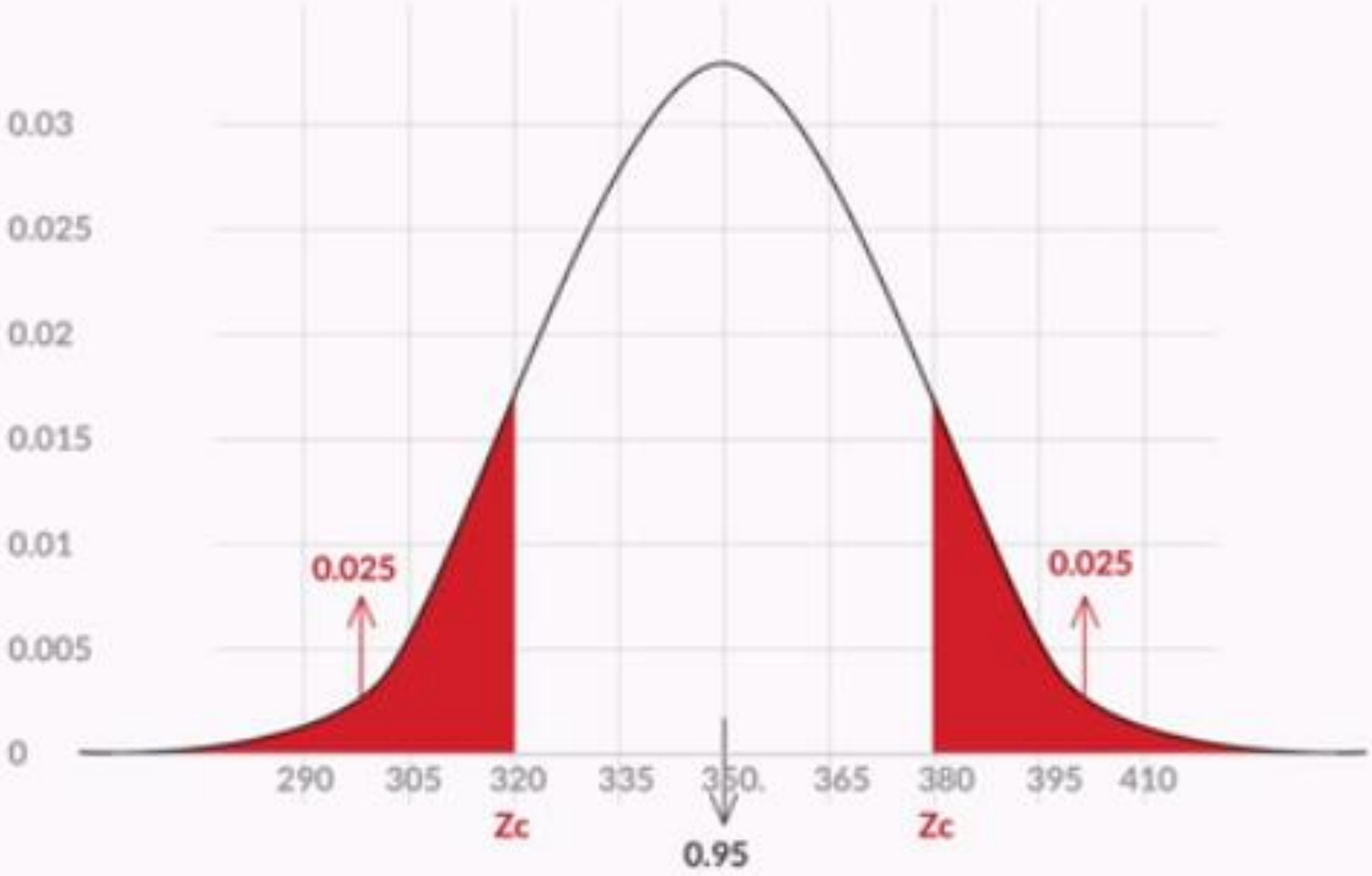
Conducting H Testing using Critical Value

If we decide the critical point as 380, what is the probability that this critical point is correct?

We are not sure. Lets take the Significance level here(Similar to Confidence Level)

So lets consider the 5% of this value as bias or error. Alpha = 5. Which means we will reject null hypothesis testing in 5% of cases and 95% cases we fail to reject null hypothesis testing when it is true

SAMPLING DISTRIBUTION OF $\bar{X}$



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

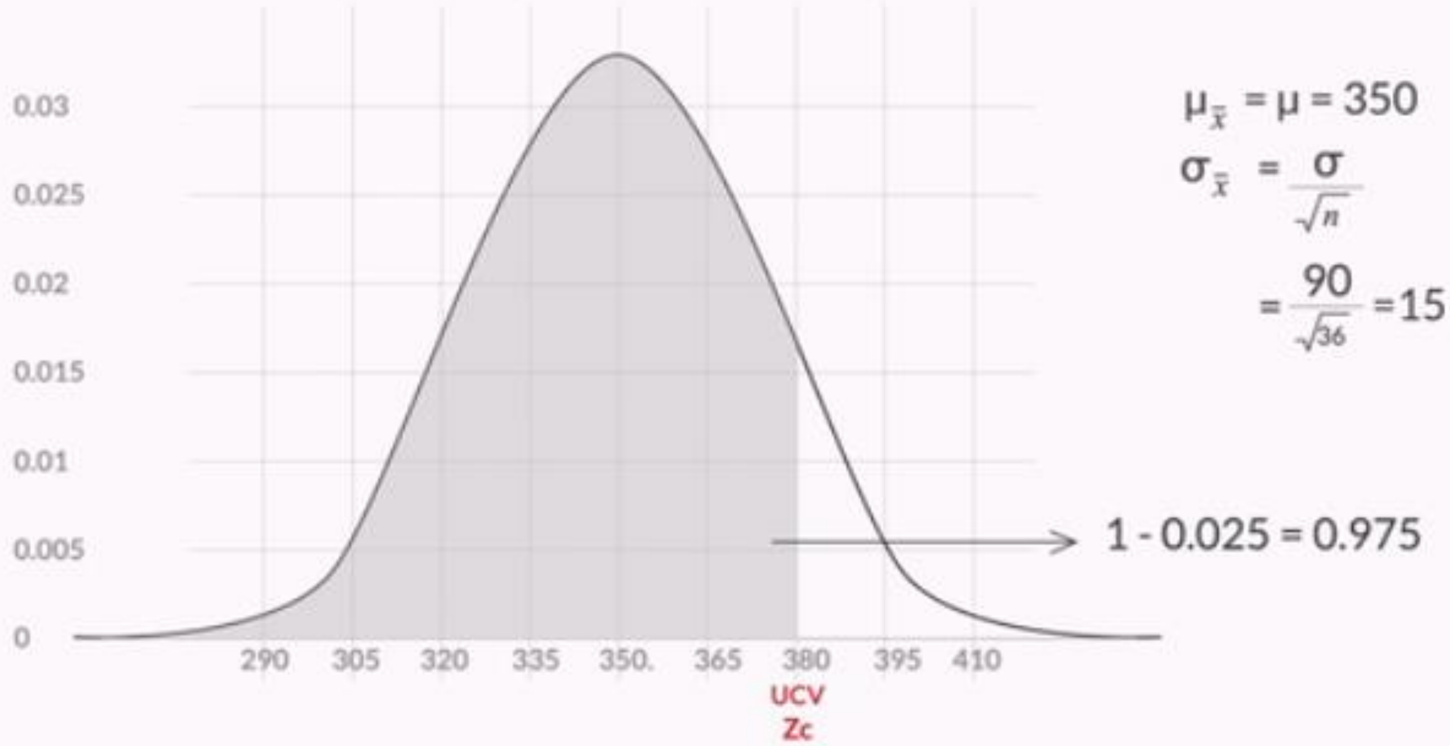
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$= \frac{90}{\sqrt{36}}$$

$$\sigma = 15$$

$$\alpha = 0.05$$

SAMPLING DISTRIBUTION OF $\bar{X}$



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

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

$$= \frac{90}{\sqrt{36}} = 15$$

First, you define a new quantity called α, which is also known as the significance level for the test. It refers to the proportion of the sample mean lying in the critical region. For this test, α is taken as 0.05 (or 5%).

Then, 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.

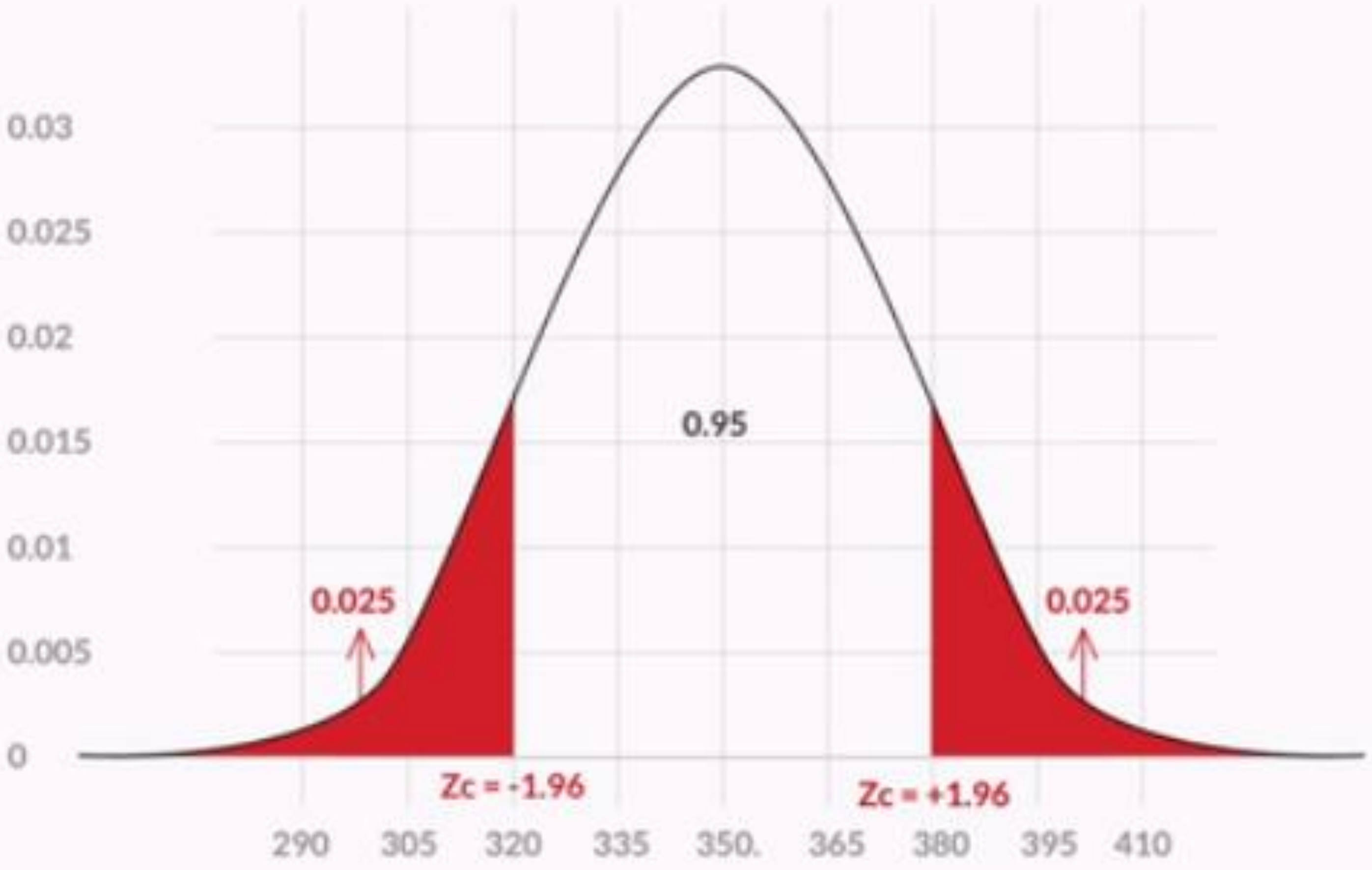
Area of rejection region

What will be the area of the critical region on the right-hand side of the distribution if the significance level (α) for a two-tailed test is 3%?

Here, value of α is 0.03 (of 3%), so the area of the rejection region would be 0.03 and the area of the acceptance region would be 0.97. In addition, since this is a two-tailed test, the area of the critical region on the right-hand side would be half of 0.03, i.e. 0.015.

What would be the area of the critical region on the right-hand side of the distribution if the significance level (α) for an upper-tailed test is 3%?

SAMPLING DISTRIBUTION OF $\bar{X}$

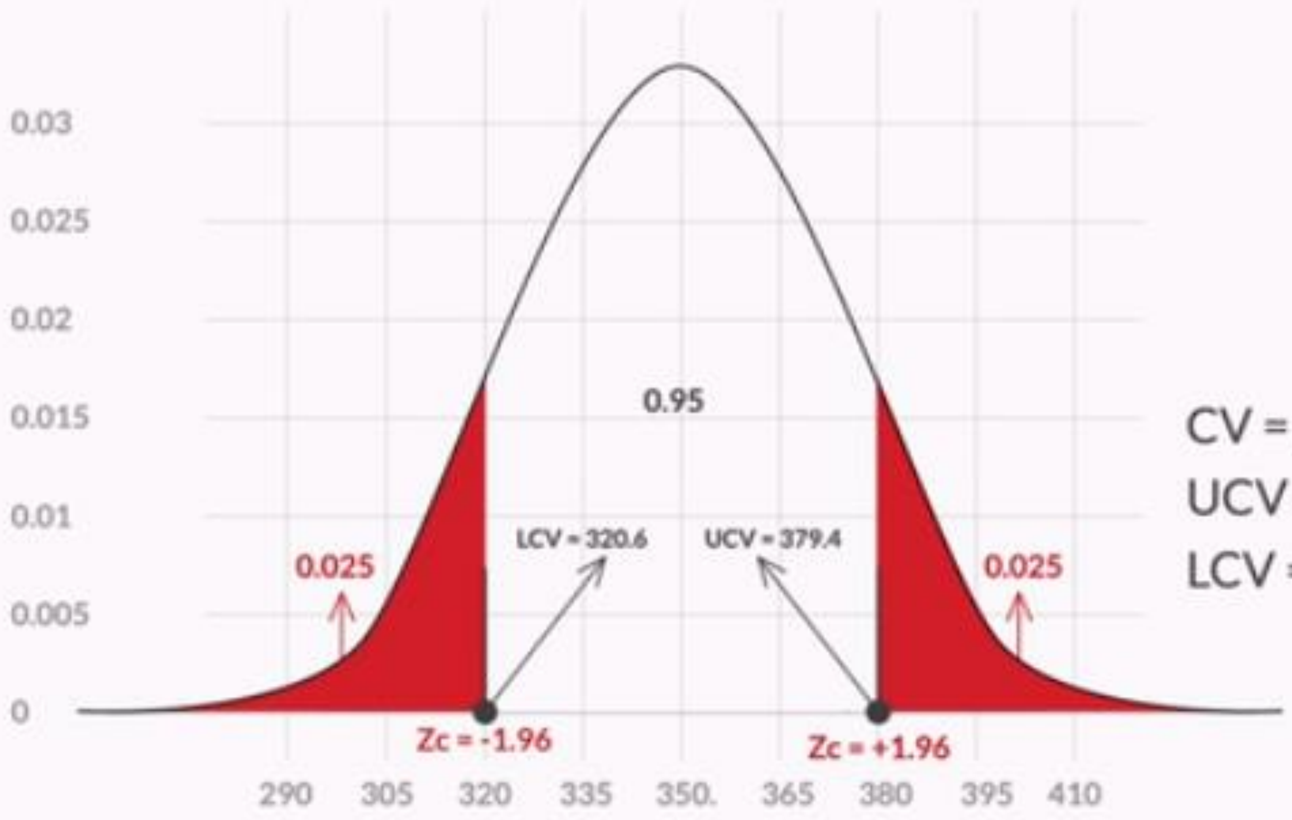


$$\mu_{\bar{x}} = \mu = 350$$

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

$$= \frac{90}{\sqrt{36}} = 15$$

SAMPLING DISTRIBUTION OF $\overline{X}$



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

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

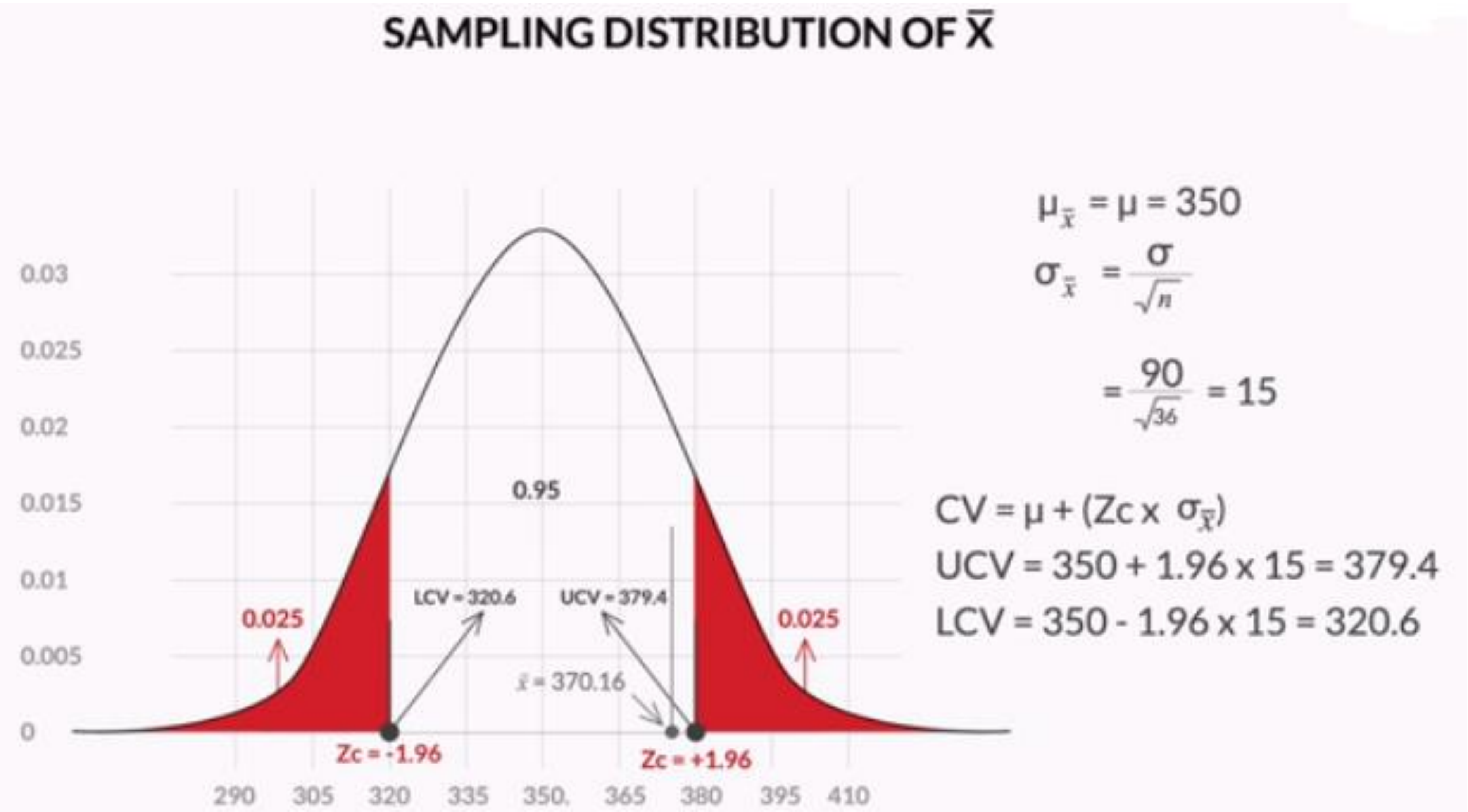
$$= \frac{90}{\sqrt{36}} = 15$$

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

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

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

What about the sample mean we've got earlier 370.16? It seems it lies in the acceptance region



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

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

Calculate the critical values (UCV and LCV) from the value of Zċ.
Make the decision on the basis of the value of the sample mean x with respect to the critical values (UCV AND LCV).

A manufacturer claims that the average life of its product is 36 months. An auditor selects a sample of 49 units of the product, and calculates the average life to be 34.5 months. The population standard deviation is 4 months.

Test the manufacturer's claim at 3% significance level using the critical value method.

First, you need to **formulate the hypotheses** for this two-tailed test, which would be:

$H_0: \mu = 36$ months and $H_1: \mu \neq 36$ months

Now, you need to follow the three steps to **find the critical values and make a decision**.

1st step: Calculate the value of Zc from the given value of α (significance level).
Calculate the z-critical score for the two-tailed test at 3% significance level.

For 3% significance level, you would have two critical regions on both sides with a total area of 0.03. So, the area of the critical region on the right side would be 0.015, which means that the area till UCV (cumulative probability of that point) would be 1 - 0.015 = 0.985. So, you need to find the z-value of 0.985. The z-score for 0.9850 in the z-table is 2.17 (2.1 on the horizontal axis and 0.07 on the vertical axis).

2nd step: Calculate the critical values (UCV and LCV) from the value of Zc.
Find out the UCV and LCV values for Zc = 2.17.
μ = 36 months σ = 4 months n (Sample size) = 49

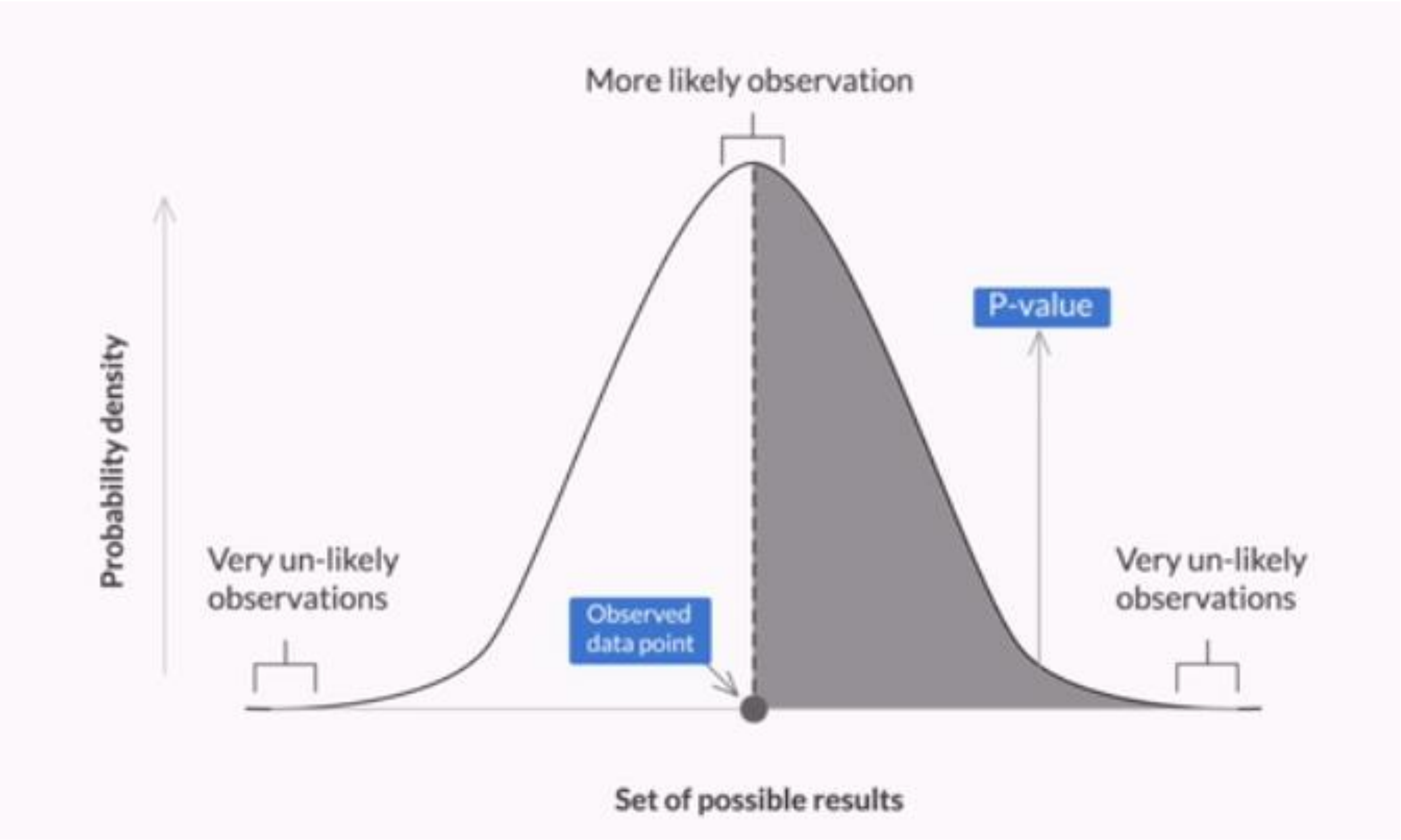
The critical values can be calculated from μ ± Zc x (σ/√NN) as 36 ± 2.17(4/√4949) = 36 ± 1.24 which comes out to be 37.24 and 34.76.

3rd step: Make the decision on the basis of the value of the sample mean ¯xx¯ with respect to the critical values (UCV AND LCV).
What would be the result of this hypothesis test?

UCV = 37.24 months LCV = 34.76 months Sample mean (¯xx¯) = 34.5 months

The UCV and LCV values for this test are 37.24 and 34.76. The sample mean in this case is 34.5 months, which is less than LCV. So, this implies that the sample mean lies in the critical region and you can reject the null hypothesis.

P-Value


More likely observation
Probability density
P-value
Very un-likely
observations
Very un-likely
observations
Observed
data point
Set of possible results

p-value as the **probability of the null hypothesis** being accepted (this should more formally be stated as "probability of failing to reject the null hypothesis"). This statement is not technically correct (or formal) definition of p-value, but it is used for better understanding of the p-value.

Higher the p-value, higher is the probability of failing to reject a null hypothesis. On the other hand, lower the p-value, higher is the probability of the null hypothesis being rejected.

After formulating the null and alternate hypotheses, the steps to follow in order to **make a decision** using the **p-value method** are as follows:

Calculate the value of z-score for the sample mean point on the distribution
Calculate the p-value from the cumulative probability for the given z-score using the z-table
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.

Example: z-score for sample point = + 3.02

Cumulative probability of sample point = 0.9987

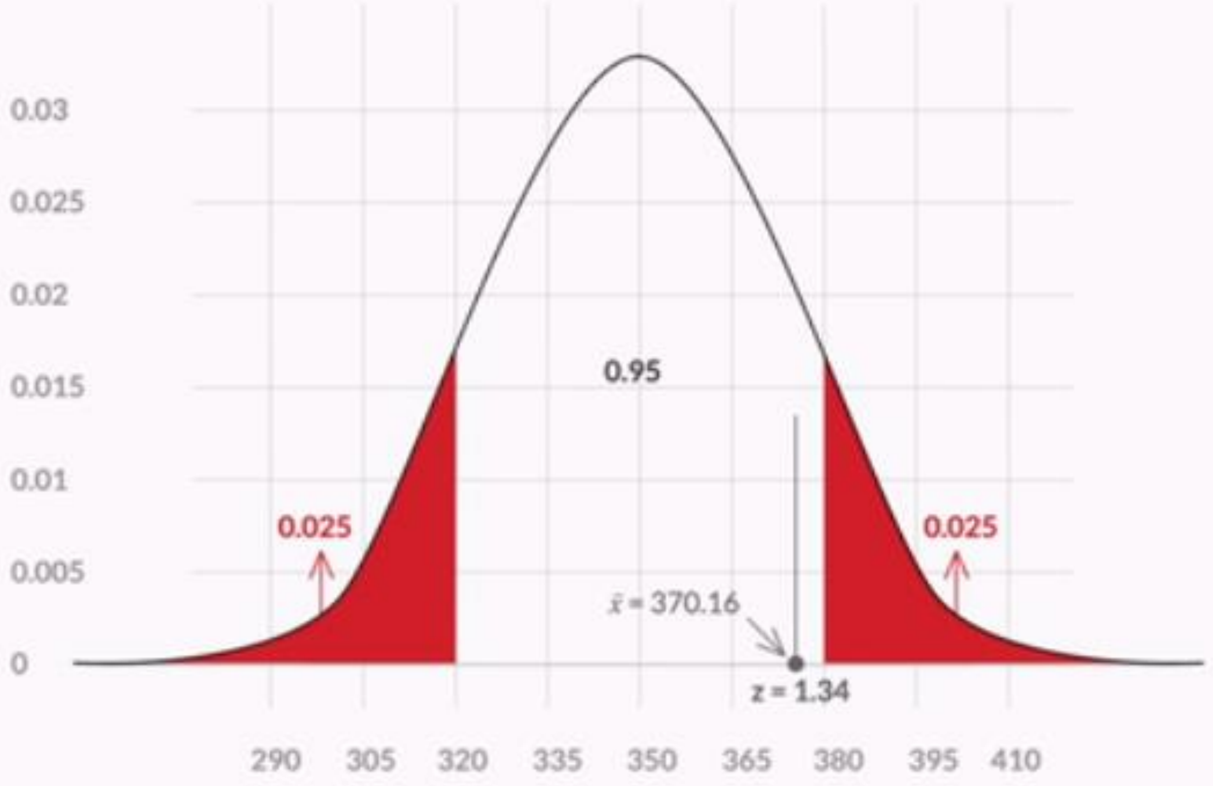
For one-tailed test → p = 1 - 0.9987 = 0.0013
For two-tailed test → p = 2 (1 - 0.9987) = 2 * 0.0013 = 0.0026

Example: z-score for sample point = -3.02

Cumulative probability of sample point = 0.0013

For one-tailed test → p = 0.0013
For two-tailed test → p = 2 * 0.0013 = 0.0026

SAMPLING DISTRIBUTION OF $\overline{X}$



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

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

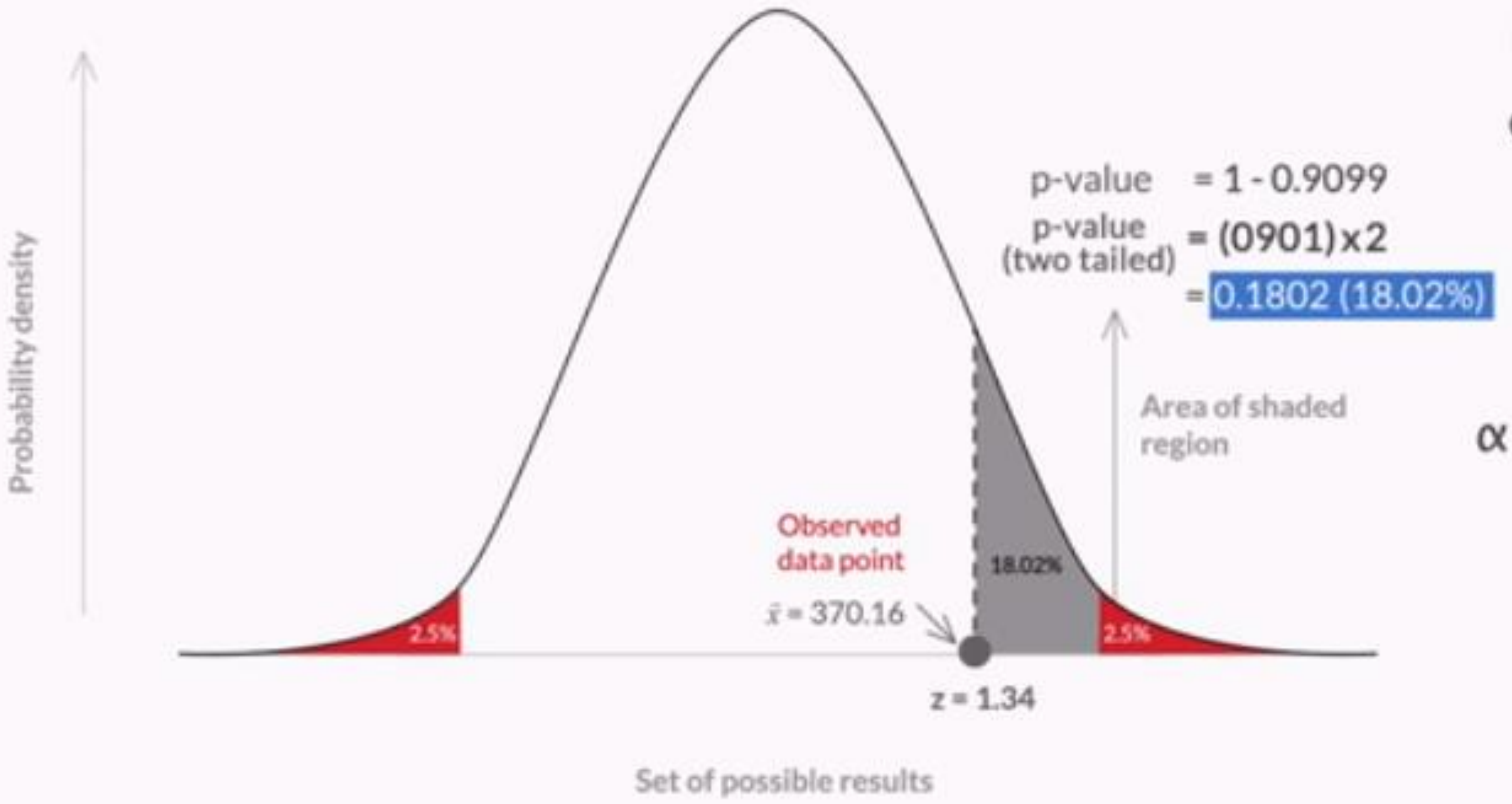
$$= \frac{90}{\sqrt{36}} = 15$$

$$Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$= \frac{370.16 - 350}{15}$$

$$= 1.34$$

SAMPLING DISTRIBUTION OF X



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

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

$$= \frac{90}{\sqrt{36}} = 15$$

$$\alpha = 0.05\ (5\%)$$

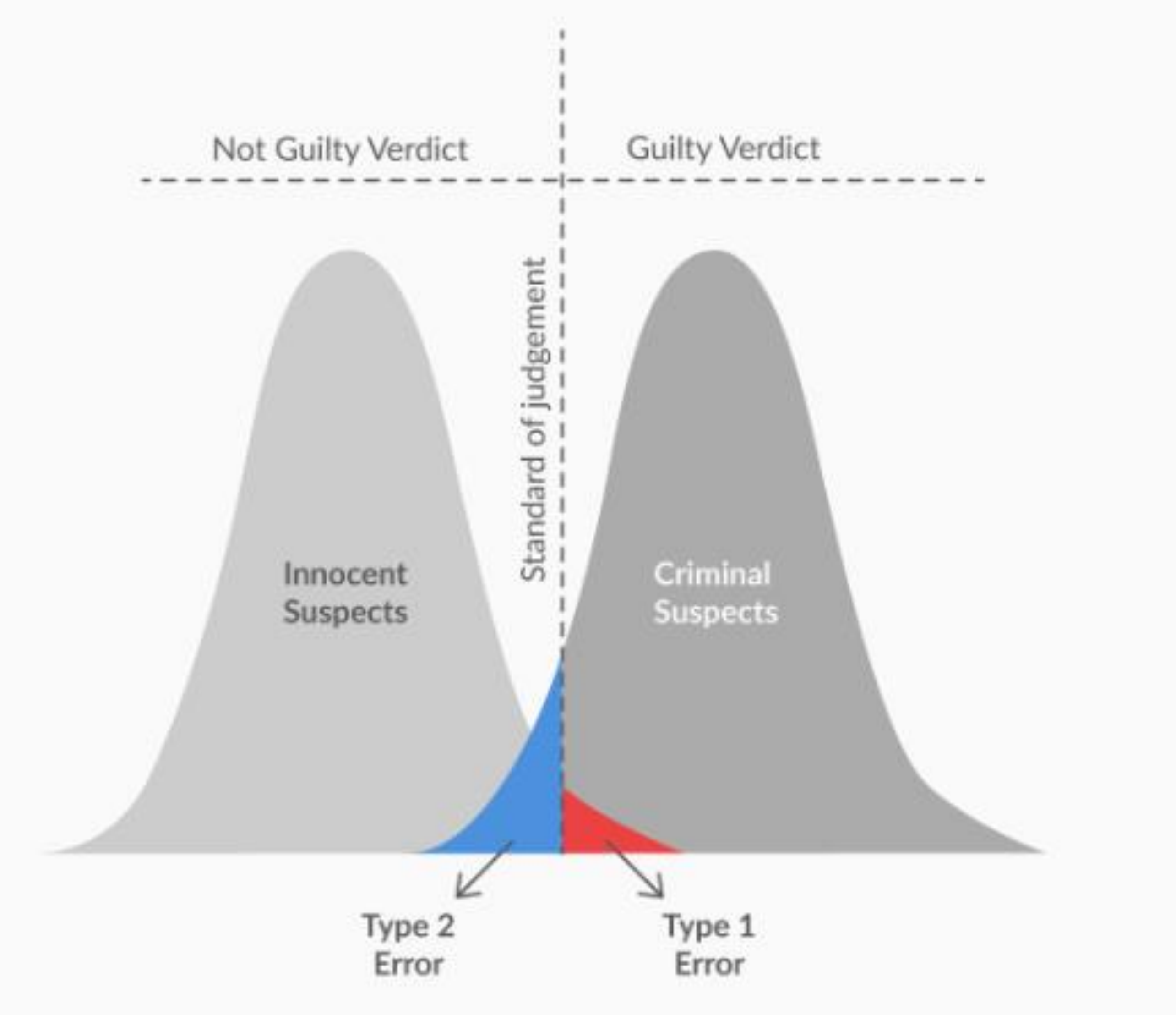
When you perform a hypothesis test, there are four possible outcomes depending on the actual truth (or falseness) of the null hypothesis H_0 and the decision to reject or not. The outcomes are summarized in the following table:

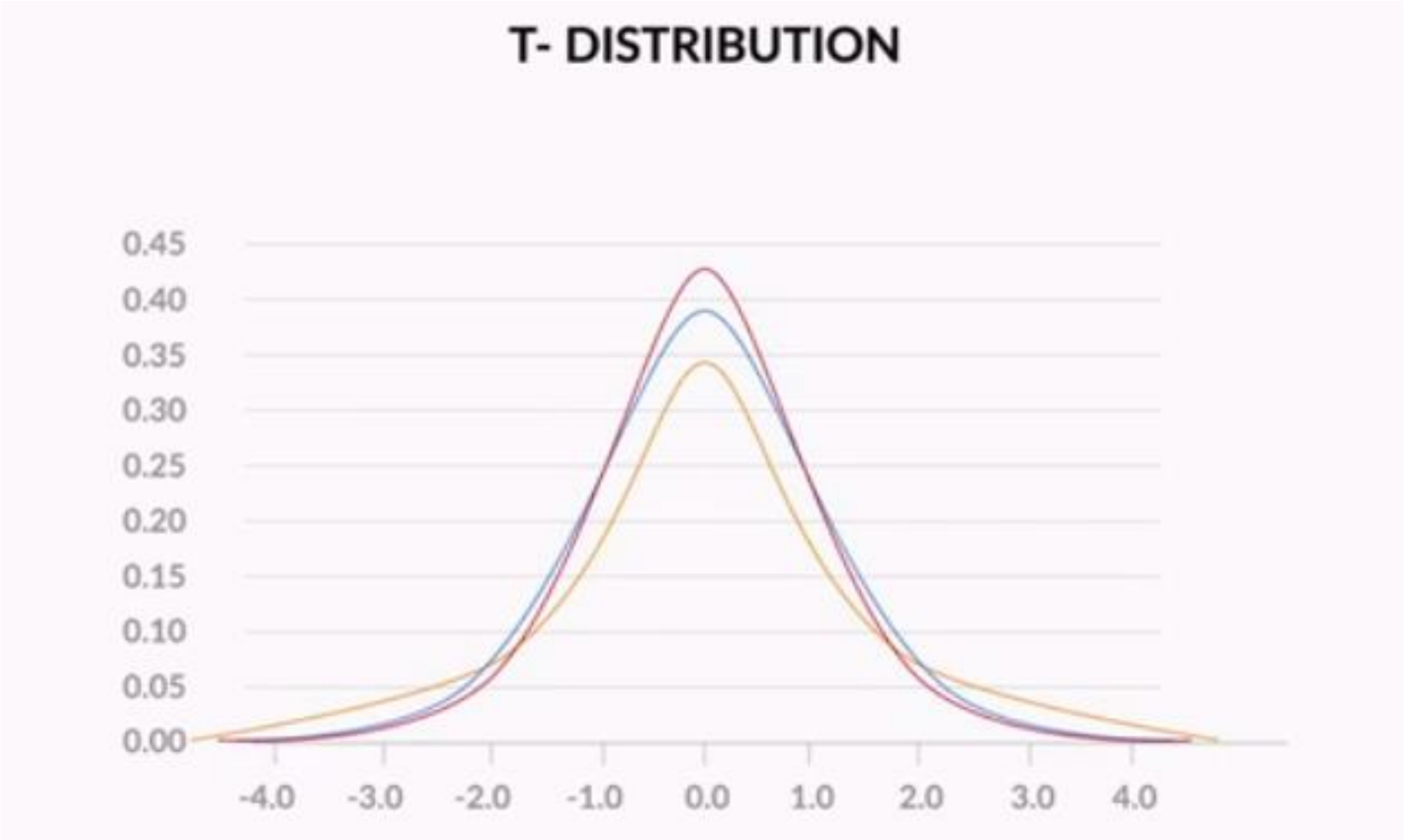
ACTION	***H_0* IS ACTUALLY**	**…**
	True	False
Do not reject *H_0*	Correct Outcome	Type II error
Reject *H_0*	Type I Error	Correct Outcome

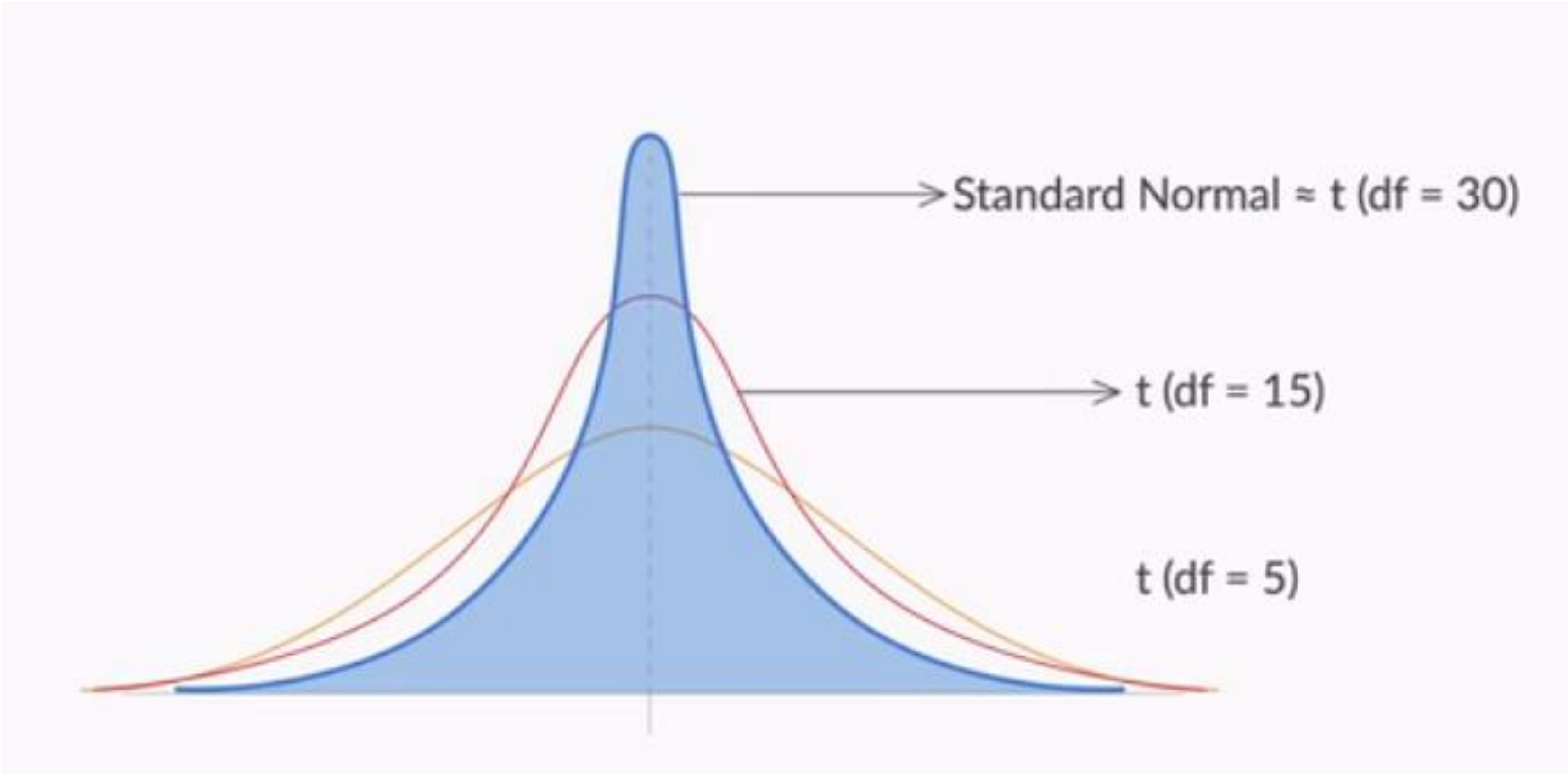
The four possible outcomes in the table are:

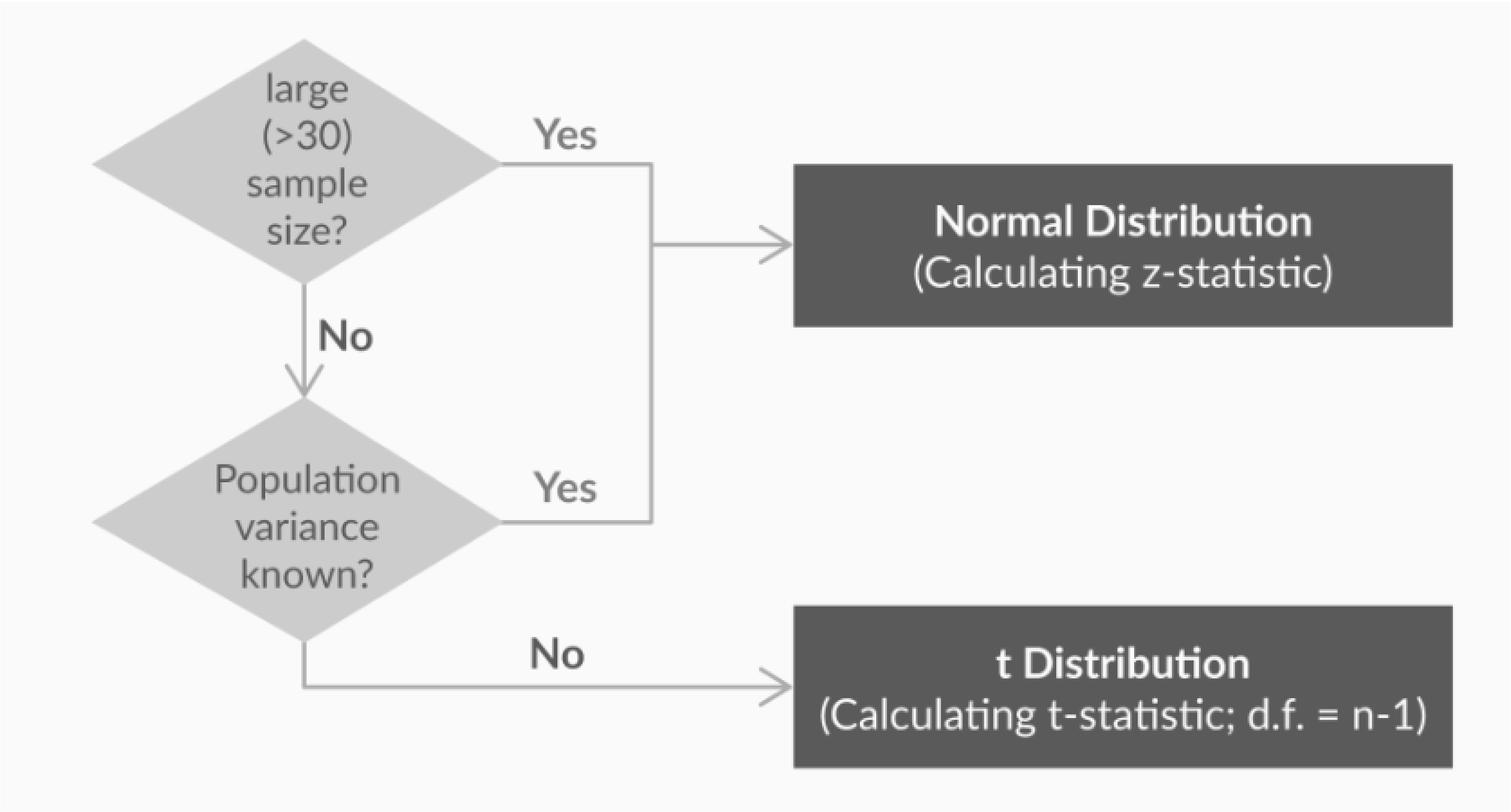
The decision is **not to reject *H_0*** when ***H_0* is true (correct decision).**
The decision is to **reject *H_0*** when ***H_0* is true** (incorrect decision known as a**Type I error**).
The decision is **not to reject *H_0*** when, in fact, ***H_0* is false** (incorrect decision known as a **Type II error**).
The decision is to **reject *H_0*** when ***H_0* is false** (**correct decision** whose probability is called the **Power of the Test**).

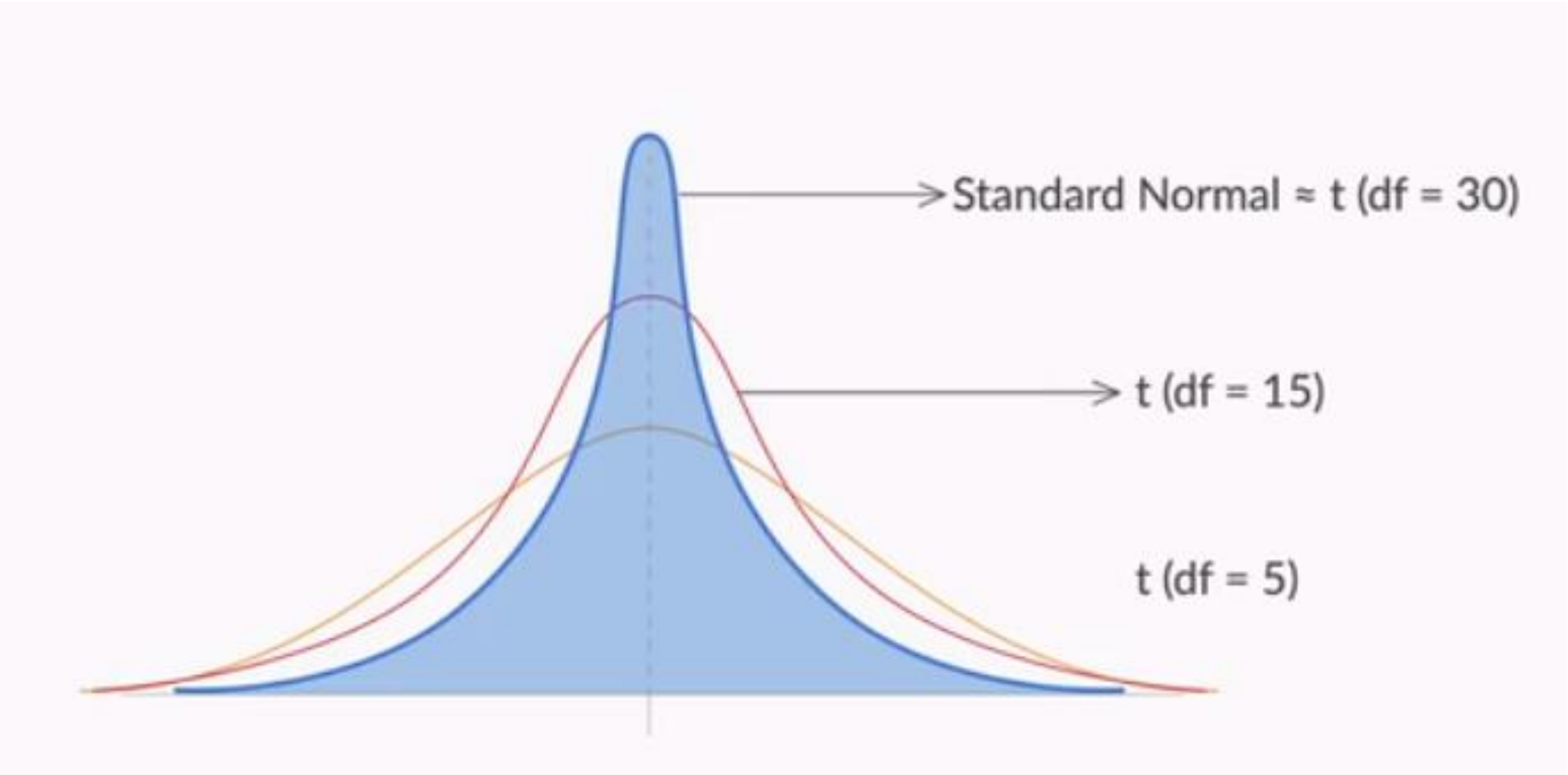
Each of the errors occurs with a particular probability. The Greek letters α and β represent the probabilities.

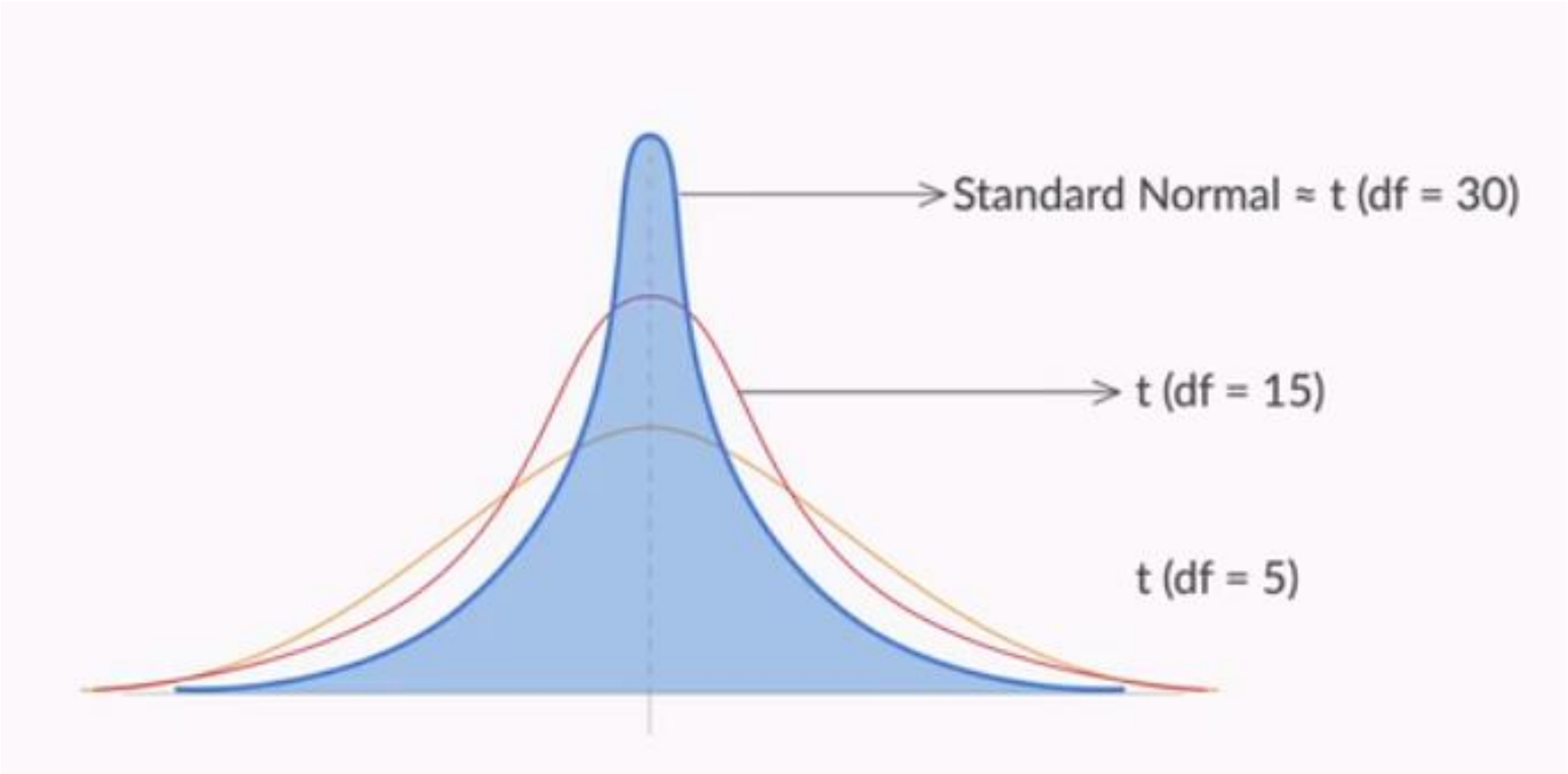

Not Guilty Verdict
Guilty Verdict
Standard of judgement
Innocent Suspects
Criminal Suspects
Type 2 Error
Type 1 Error


T- DISTRIBUTION
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-4.0
-3.0
-2.0
-1.0
0.0
1.0
2.0
3.0
4.0


Standard Normal ≈ t (df = 30)
t (df = 15)
t (df = 5)


large (>30) sample size?
Yes
No
Population variance known?
Yes
No
Normal Distribution
(Calculating z-statistic)
t Distribution
(Calculating t-statistic; d.f. = n-1)


Standard Normal ≈ t (df = 30)
t (df = 15)
t (df = 5)



Standard Normal ≈ t (df = 30)
t (df = 15)
t (df = 5)