INNOMATICS
RESEARCH LABS

Hypothesis Testing

Hypothesis Testing

- Also called as Statistical Inference
- Family of statistical methods used to identify whether a sample of observed data can be used to accept or reject a predefined hypothesis.
- Hypothesis Testing is applied in many domains, mainly in research but also a key method in online marketing (AB testing).

Intuitive Example:

- In a box of 100 apples (call this **population**), we take a **sample** of 8 apples.
- This years’ sample of 8 apples has 5 rotten apples (62%) and last years’ sample of 8 apples had 4 rotten apples (50%).
- We want to use **hypothesis testing** to determine whether the percentage of rotten apples in the box of 100 apples is larger this year than last year.

Hypothesis Testing Procedure

Step 1: state hypothesis H_0 as alternative to the assertion, H_A is the alternative,

(“assertion”) or the claim

H_0 : (Null Hypothesis) … expected result

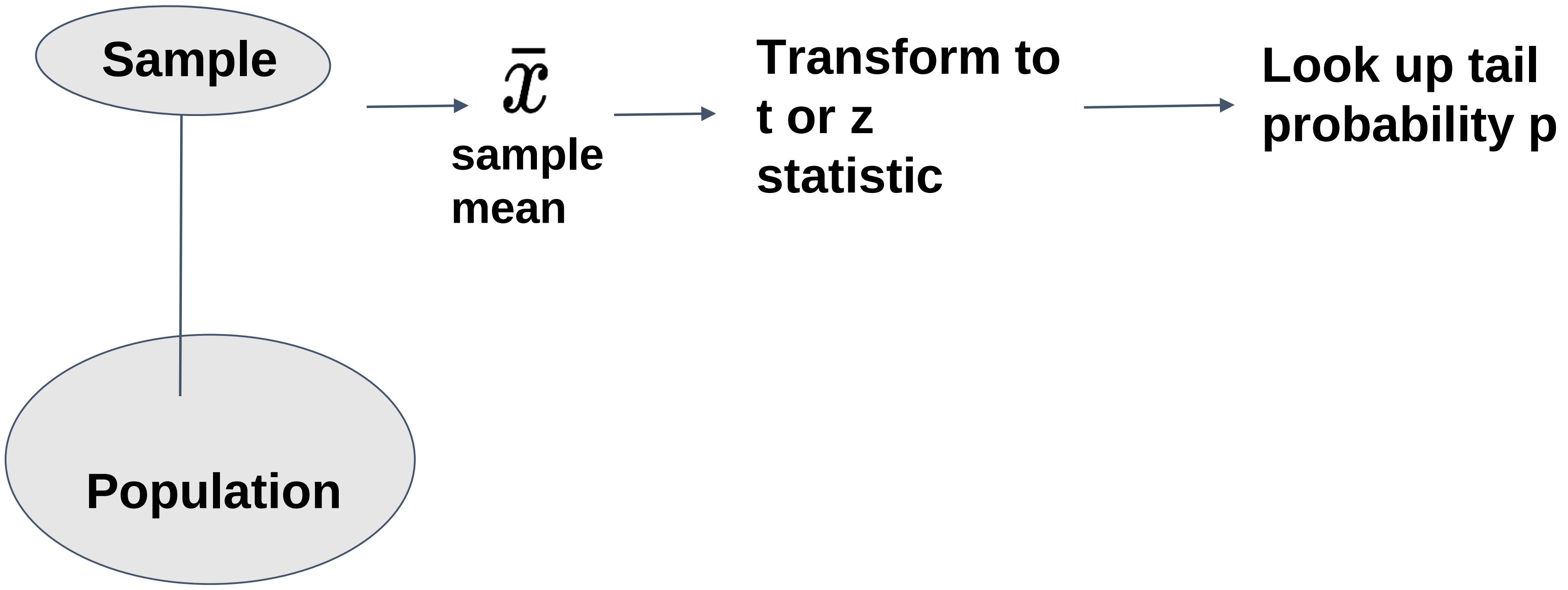
H_A : Alternative Hypothesis … the assertion or claim

The **null hypothesis** **always** states that the sample mean is equal to the expected or advertised mean. The assertion or **alternative** hypothesis **always** states that the sample mean is different than the expected mean.

Step 2: Verify the conditions of inference

Step 3: Calculate the test statistic (t-statistic or z-statistic) and the p-value

(probability in the tail(s))



Step 4: state your decision … **If p ≥ α, fail to reject null, if p < α reject the null hypothesis**

Because p < α, we reject the null hypothesis. There is sufficient evidence to conclude that the measurement is biased.

Because p > α, we fail to reject the null hypothesis. There is not sufficient evidence to conclude that the measurement is biased.

Assumptions for Inference	
Randomness	**Data is collected randomly**
Samples are Independent	**Population ≥ 10n** **1. Population is at least 10 times the sample size or 2. sampling with replacement**
Normality	**N ≥ 30 or the parent distribution must be normal - e.g., population parameter is normally distributed: the weight of gold bars normally distributed. This assumption should validated be if using a t-distribution.**

Intuition on what is hypothesis testing

- Hypothesis test is a **statistical test** used to determine if an assumption made on sample is true for a population.
- Now, for every belief there exist an exact opposite, a contradictory belief.
- In hypothesis testing, your belief is called alternate hypothesis while the contradictory belief is called null hypothesis.
- Here, you don't try to directly prove alternate hypothesis (your belief).
- Rather you assume that the null hypothesis (contradictory belief) is true and then work backwards to disprove it

Intuitive example

- Consider two colleges A and B. Containing heights of students.
- Take a random sample of 50 heights from each college (the actual population of each college can be anything, but we take only 50 sample values).
- Now, let us calculate the means (sample means) of both the samples. Let the sample means be X and Y respectively.
- Let us compute the difference in sample means now.
- Let X-Y = d.
- **Now, “d” is the actual computed value that is calculated from the sample means. This cannot be wrong.**
- Let us assume d=15 cm.

Intuitive example contd.

- That is, let us assume the difference in sample means that we have calculated be equal to 15 cm.
- **This 15 cm is the observed value.**
- Now , we want to know if the same or different or no difference in means will happen for the actual population of A and B. That is if we want to know that the difference between the **POPULATION** MEANS of colleges A and B will be 15cms or greater than 15 cm or less than 15 cm or 0.
- How do we know this?
- We use **Hypothesis Testing** to find out the likelihood of this possibility.

Terminologies revisited

Some simple terms to know.

- **Null Hypothesis (Ho):** Null hypothesis is the outcome which is ideal or which we want to see. In this case, null hypothesis can be
  - "There is no difference between the population means of college A and college B"
- **Alternative Hypothesis :** Alternative hypothesis is something that says that there is a difference in the statistical difference, which means in this case, alternative hypothesis can be
  - "There is a difference between the population means of college A and college B".

What do we want to find?

- We want to find out the probability of getting a value of (d >= 15 cms) if there is NO difference of means in the actual **POPULATION** heights of A and B.
  - **Note**: **d** is the difference of means of **SAMPLE** heights of A and B.
- In other words, what is the probability of **observing a difference** of 15 cms of more in the **SAMPLE** heights ,**IF** there is **no difference** in the actual **POPULATION** heights,
- This can be represented like this: **P(d>=15cm | Ho)**
  - This probability value is called the **p-value.**
- Usually, as a thumb-rule if **the p-value is more than 5%** we **fail to reject** the null hypothesis and if it’s **less than 5%** , we **reject** the null hypothesis.

Solving the hypothesis case

- Let us see both the cases.
  - **Consider a p-value of 0.02:**
- A p-value of 0.02 means that there is a 2% chance of observing a sample difference of 15 cm or more considering there is no difference in population means (or considering Ho=true)
- 2% chance is relatively small, hence we reject the Null Hypothesis.
  - This means, in simple words, there is no (or very less) chance of observing a sample difference 15 cm of more if there is no actual difference in population means.
  - But, like I mentioned in the beginning , the 15 cm of sample difference is already made. That is the reality. Hence, it cannot be wrong. **Therefore, we reject the Null Hypothesis or, we reject our assumption which says that there is no difference in the population means.**

Solving the hypothesis case contd.

- **Consider a p-value of 0.7:**

- A p-value of 0.7 means that there is 70% chance of observing a sample difference of 15 cm or more considering there is no difference in population means (or considering Ho=true)
- 70% is a high or a significant value. Hence, we cannot reject the null hypothesis.
- This means, in simple words, there is a high chance of observing a sample difference 15 cm of more if there is no actual difference in population means.
- **Hence, we fail to reject our Null Hypothesis.**

Example 2

- Suppose, I've applied for a typing job and I've stated in my resume that my typing speed is 60 words per minute on an average.
- My recruiter may want **to test my claim.** If he finds m**y claim to be acceptable, he will hire me otherwise reject my candidature.**
- So he asked me to type **a sample** letter and found that my speed is 54 words a minute.
- Now, he can **decide on** whether to hire me or not [assuming that I meet all other eligibility criteria].
- This procedure depicts hypothesis testing in layman's terms.

Example 2 contd.

- ***In statistical terms:***
  - **Hypotheses:**

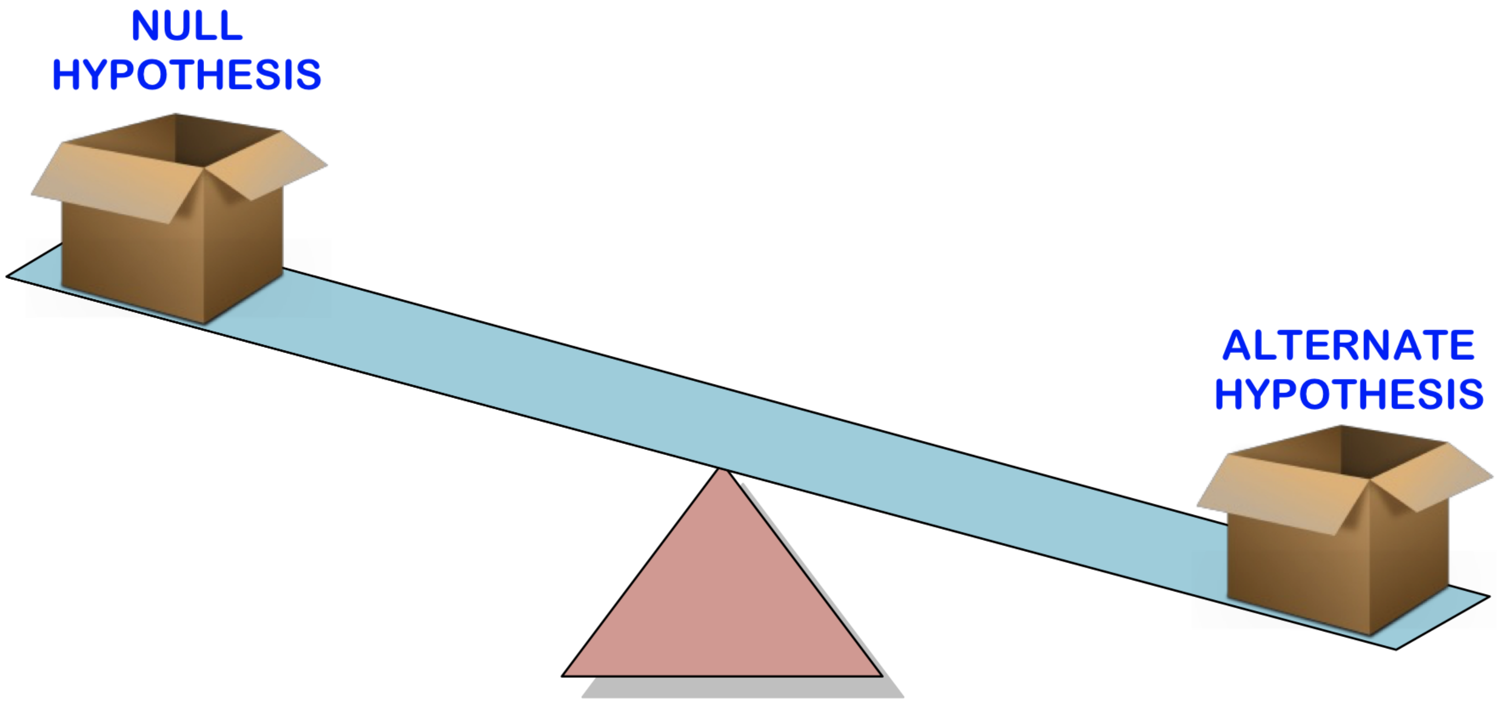
  "my typing speed is 60 words per minute on an average" is a **hypothesis to be tested**, called null hypothesis. [obviously, the alternating hypothesis is "my typing speed is **not** 60 words per minute on an average"]

  - **Population & Sample:** My average typing speed is population parameter and my sample typing speed is sample statistics.

Level of significance:

The criteria of accepting /rejecting my claim is to be decided by the recruiter [the researcher]. For example, he may decide that an error of 5 words is ok to him so he would accept my claim between 55 to 65 words/minutes. In that case, my sample speed 54 words/minute will conclude to reject my claim. And the decision will be "I was making a false claim“. But if the recruiter extends his acceptance region to +/- 7 words [that is 53 to 67 words], he would be accepting my claim.

- So, to conclude, Hypothesis testing is a process to test claims about the population on the basis of sample.


NULL
HYPOTHESIS
ALTERNATE
HYPOTHESIS

Hypothesis testing:

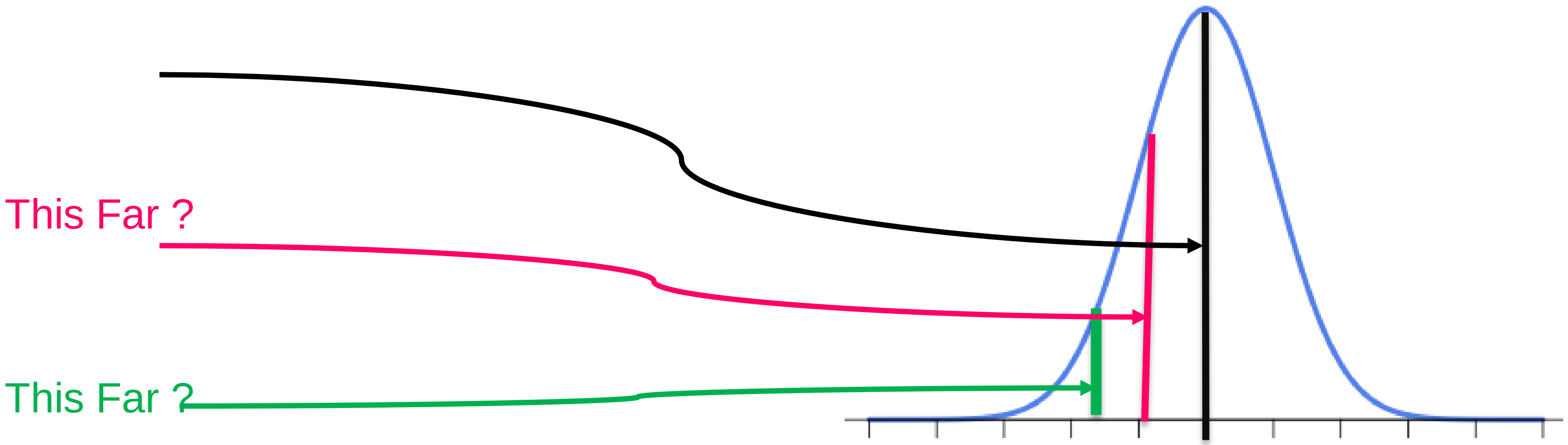
A Sporting Analyst claims that Lionel Messi have an average score of 8.3/10 in Club Matches.

You doubt that claim and take a random sample of 30 matches and you find a mean score of 6.4/10, with a sample standard deviation of 1. Can you reject the claim by the Sporting Analyst?

Hypothesis Testing Process

Considering variations in samples, how far away from 8.3/10

is acceptable to you as expected variation and when do you say “enough is enough; this is too far”?



Claim or Expectation, say, mean score = 8.3 /10

Step 1: Decide on the hypotheses

Average score on the test is 8.3/10.

This is called Null Hypothesis and is represented by H_0.

In this case, H0 : μ = 0.83

If Null Hypothesis is rejected based on evidence, an Alternate Hypothesis, H_1, needs to be accepted. **We always start with the assumption that Null Hypothesis is true.**

In this case, H1: μ < 0.83

Examples of Hypotheses

- Two hypotheses in competition:
  - H0 : The NULL hypothesis, usually the most conservative.
  - H1 or HA : The ALTERNATIVE hypothesis, the one we are actually interested in.
- Examples of NULL Hypothesis:
  - The coin is fair
  - The new drug is no better (or worse) than the placebo
- Examples of ALTERNATIVE hypothesis:
  - The coin is biased (either towards heads or tails)
  - The coin is biased towards heads
  - The coin has a probability 0.6 of landing on tails
  - The drug is better than the placebo

Step 2: Choose your statistics

Sample size = 30

Normal distribution is a good approximation

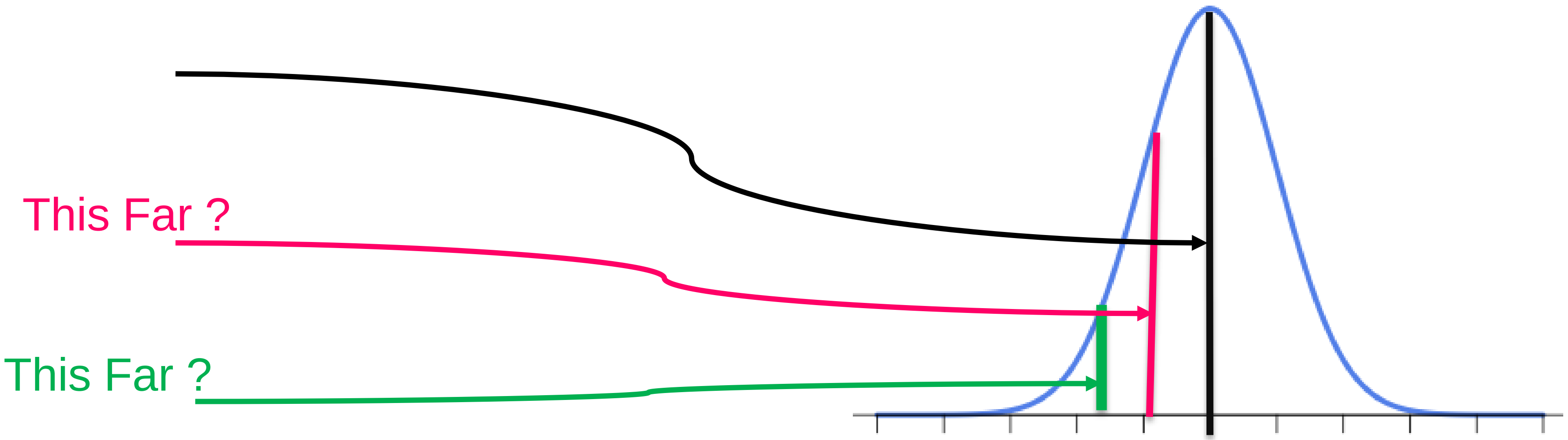
$$Std\ Err = \frac{s}{\sqrt{n}} = \frac{1.0}{\sqrt{30}} = 0.182$$

X ~ N(0.7, 0.182^2) = N(0.7, 0.033)

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{0.64 - 0.83}{0.182} = -1.04$$

Step 3: Specify the significance Level

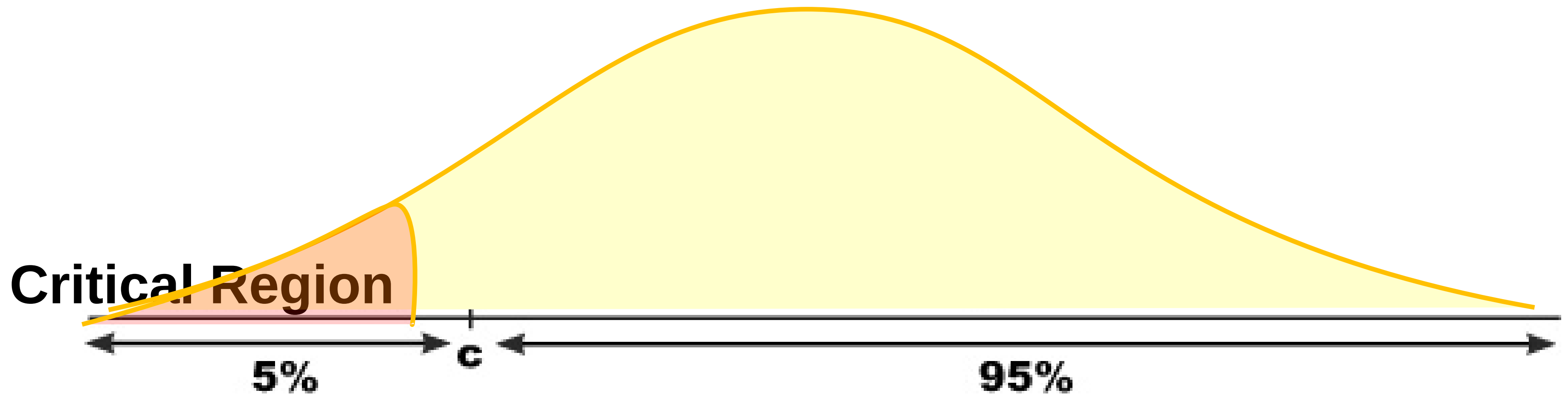
First, we must decide on the Significance Level, α. It is a measure of how unlikely you want the results of the sample to be before you reject the null hypothesis, H_0.



Claim or Expectation, say, mean score = 7 /10

Step 4: Determine the critical region

If X represents the sample mean score, the critical region is defined as $P(X < c) < \alpha$ where $\alpha = 5\ \%$.



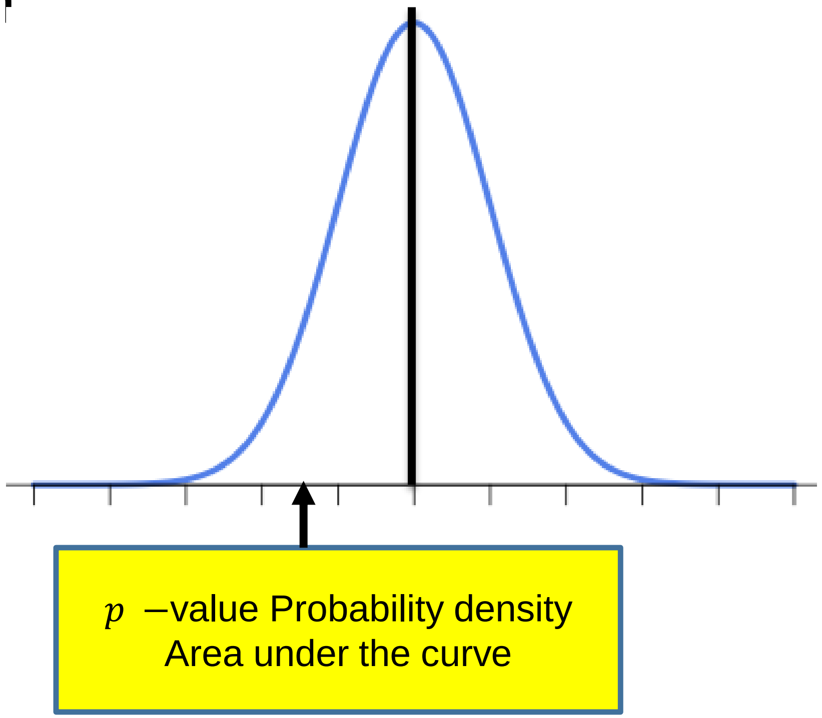
Recall that in a 95% CI, there is a 5% chance that the sample will not contain the population mean. Hence if the sample falls in the critical region, the null hypothesis that 0.7 is the mean score is rejected.

That is the reason 5% or 0.05 is called the Significance Level. In a 99% CI, 0.01 is the Significance Level.

Step 5: Find the p − value

p-value is the probability of getting a value up to and including the one in the sample in the direction of the critical region

It is a way of taking the sample and working out whether the result falls within the critical region of the hypothesis test.



Essentially, this is the value used to determine whether or not to reject the null hypothesis.

Step 5: Find the p −value

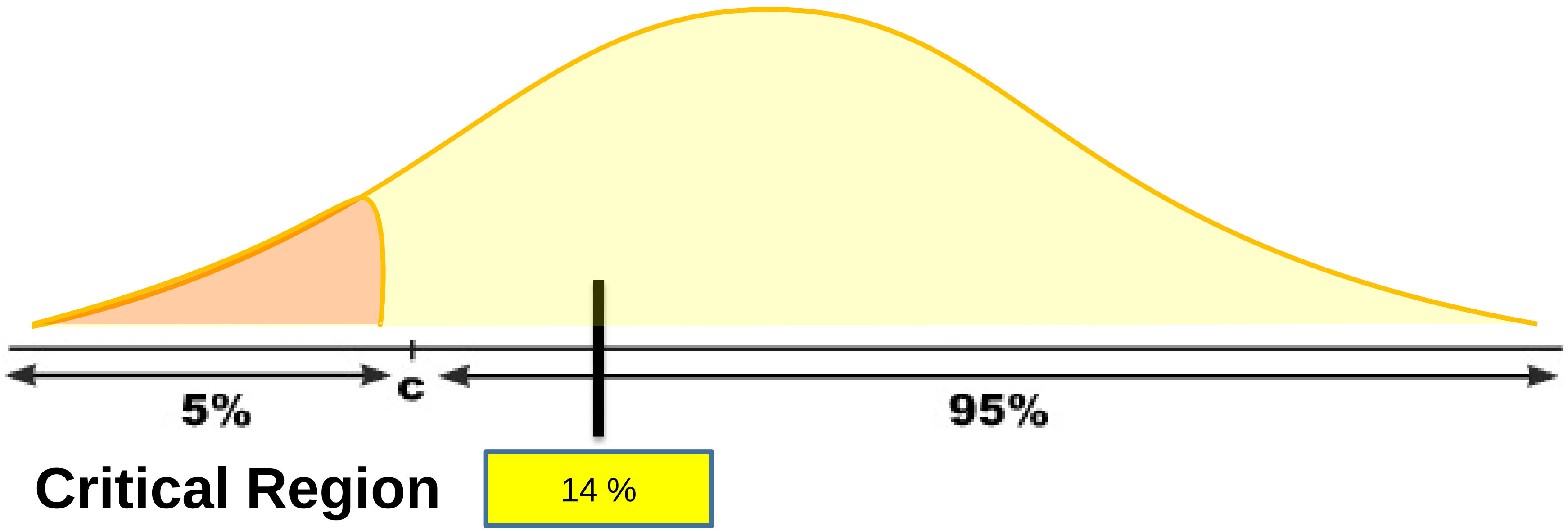
In our sample, we found a mean score of 6.4/10. This means our *p*-value is P(X = 0.64), where X is the distribution of the mean scores in the sample.

If P(X = 0.64) < 0.05 (Significance Level), it indicates that 0.64 is inside the critical region, and hence H0 can be rejected.

Given that Z = -1.04 , P(X $\leq$ 0.64) = 0.149

So there is a 15% probability of find a mean score of 6.4/10 or less.

Step 6: Is the sample result in the critical region ?



Step 7: Make your decision

There isn't sufficient evidence to reject the null hypothesis and so, the claims of the principal are accepted.

Would your conclusion be any different if the same average score of 6.4/10 was found from a sample of size 300 ?

What are the null and alternate hypotheses ?

$$H_0 : \mu = 0.83$$
$$H_1 : \mu < 0.83$$

What is the test statistics ?

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{0.64 - 0.83}{\frac{1}{\sqrt{300}}} = -3.3$$

p-value = P(Z < -3.3) = 0.0005

What is your decision ?

Since the p-value (0.0005) is less than the Significance Level of 0.05, the null hypothesis can be rejected.

Attention Check

- In hypothesis testing, do you assume the null hypothesis to be true or false?

- If there is sufficient evidence against the null hypothesis, do you accept it or reject it?

Attention Check

In hypothesis testing, do you assume the null hypothesis to be true or false?

True.

If there is sufficient evidence against the null hypothesis, do you accept it or reject it?

Reject it.

Attention Check

If the *p*-value is less than 0.05 for the above significance level, will you accept or reject the null hypothesis?

Do you need weaker evidence or stronger to reject the null hypothesis if you were testing at the 1% significance level instead of the 5% significance level?

Attention Check

If the *p*-value is less than 0.05 for the above significance level, will you accept or reject the null hypothesis?

Reject it.

Do you need weaker evidence or stronger to reject the null hypothesis if you were testing at the 1% significance level instead of the 5% significance level?

Stronger.

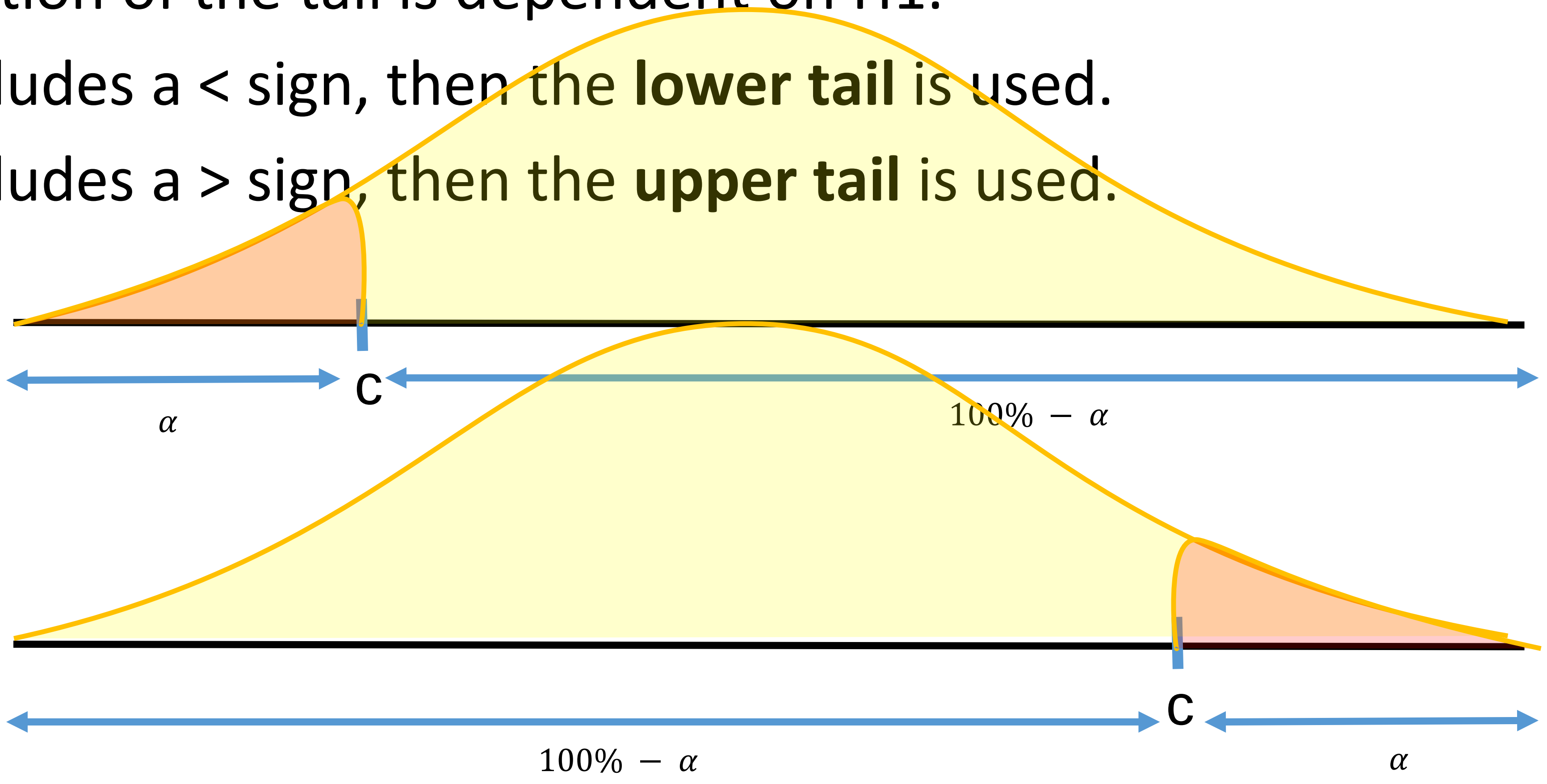
Critical Region Up Close

One-tailed tests

The position of the tail is dependent on H1.

If H_1 includes a < sign, then the **lower tail** is used.

If H_1 includes a > sign, then the **upper tail** is used.

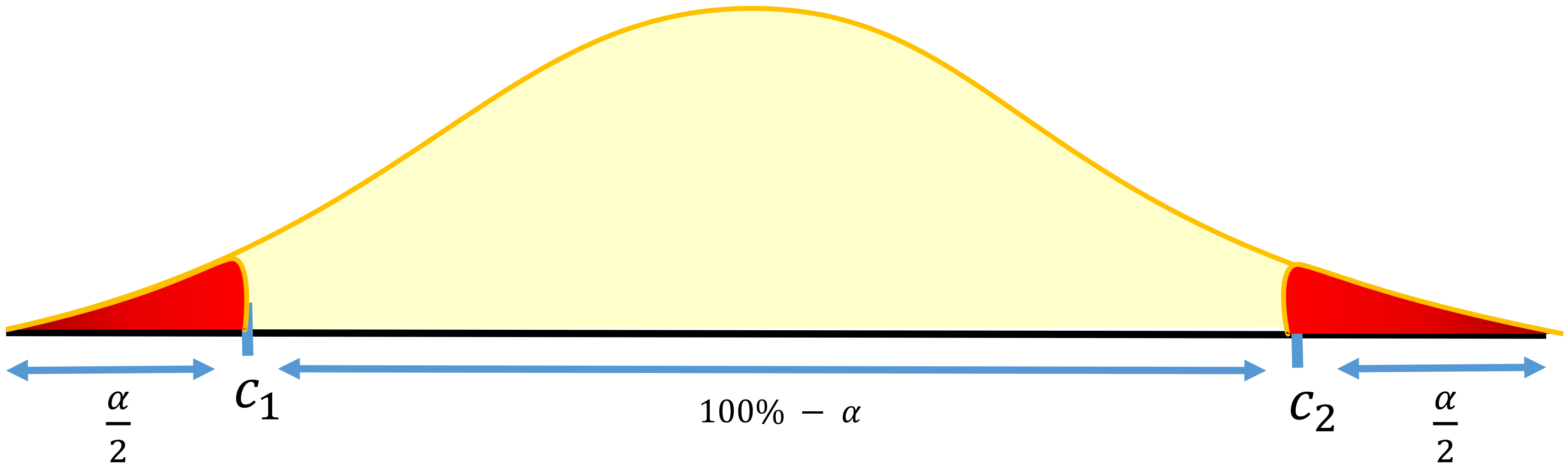


Critical Region Up Close

Two-tailed tests

Critical region is split over both ends. Both ends contain $\frac{\alpha}{2}$, making a total of α.

If H_1 includes a ≠ sign, then the two-tailed test is used as we then look for a change in parameter, rather than an increase or a decrease.



Critical Region Up Close

For each of the scenarios below, identify what type of test you would require.

- Average test score problem as discussed till now.

- If we were checking whether the average is significantly different from 7/10, i.e., H1: $\mu \neq 0.7$.

- The coin is biased.

- The coin is biased towards heads with probability 0.8.

Critical Region Up Close

For each of the scenarios below, identify what type of test you would require.

- Average test score problem as discussed till now.

One-tailed/Lower-tailed

- If we were checking whether the average is significantly different from 7/10, i.e., H1: $\mu \neq 0.7$.

Two-tailed test

- The coin is biased.

Two-tailed test

- The coin is biased towards heads with probability 0.8.

One-tailed/Upper-tailed

The Missing Link in the Interview

Q. What is the probability of getting 15 or more heads out of 20 coins?

A.

$$P(X \geq 15) = P(X = 15) + P(X = 16) + P(X = 17) + P(X = 18) + P(X = 19) + P(X = 20)$$

$$= 0.021$$

What can you now say about the coin being biased or not?

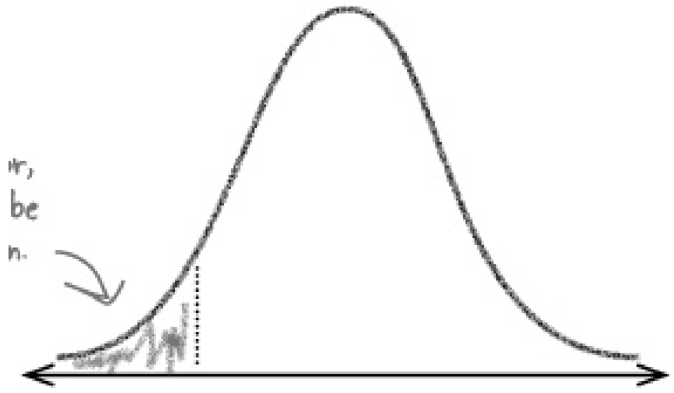
The hypothesis test doesn’t answer the question whether the coin is biased or not; it only states whether the evidence is enough to reject the null hypothesis or not ***at the chosen significance level***.

Errors

- Type I: We reject the NULL hypothesis incorrectly
- Type II: We “accept” it incorrectly

		State of Nature	
		Null True	Null False
Action	Fail to reject null (negative)	Correct decision True Negative $P(\text{accept } H_0 \mid H_0 \text{True})$	Type II error(β) False Negative $P(\text{Accept } H_0 \mid H_0 \text{ False})$
	Reject null (positive)	Type II error(α) False Positive $P(\text{Accept } H_0 \mid H_0 \text{ True})$	Correct decision (Power) Sensitivity /Recall $P(\text{accept } H_0 \mid H_0 \text{False})$

Probability of Getting Type I Error



α

$$\boldsymbol{P(Type\ I\ error) = \alpha}$$

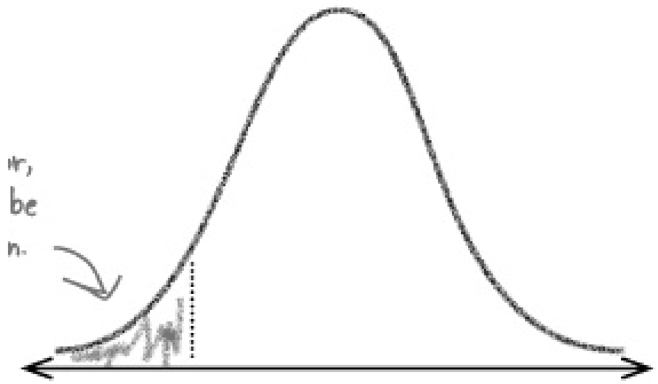
		State of Nature	
		Null True	Null False
Action	Fail to reject null (negative)	Correct decision True Negative $P(accept\ H_o \mid H_o True)$	Type II error(β) False Negative $P(Accept\ H_o \mid H_o\ False)$
	Reject null (positive)	Type II error(α) False Positive $P(Accept\ H_o \mid H_o\ True)$	Correct decision (Power) Sensitivity /Recall $P(accept\ H_o \mid H_o False)$

Probability of Getting Type II error

$$P(Type\ II\ error) = \beta$$

To find β

1. Check that you have a specific value for H1.



2. Find the range of values outside the critical region of the test. If the test statistic has been standardized, it needs to be de-standardized for the purpose.

3. Find the probability of getting this range of values, assuming H1 is true. In other words, find the probability of getting the range of values outside the critical region, but this time using the test statistic described by H1 and not H0.

A new miracle drug claims that it cures common cold and it has had a success rate of 90%. You conduct a random sample test with 100 patients and you find that 80 of them are cured.

At 5% significant level, do you reject or accept the claim by the drug company?

What are the null and alternate hypotheses ?

$$H_o: p = 0.9$$
$$H_1: p < 0.9$$

What is the test statistics ?

X ~ B(100,0.9)

Since np>5 and nq>5, Normal distribution can be used instead.

X ~ N(np, npq)

X ~ N(90, 9)

What is the probability of 80 or fewer getting cured?

$$Z = \frac{80.5 \ - 90}{\sqrt{9}} = -3.17$$

p-value = P(Z < -3.17) = 0.0008

Probabilities of Errors in our Example

P(Type | error) = 0.05

<u>To calculate P(Type II error)</u>

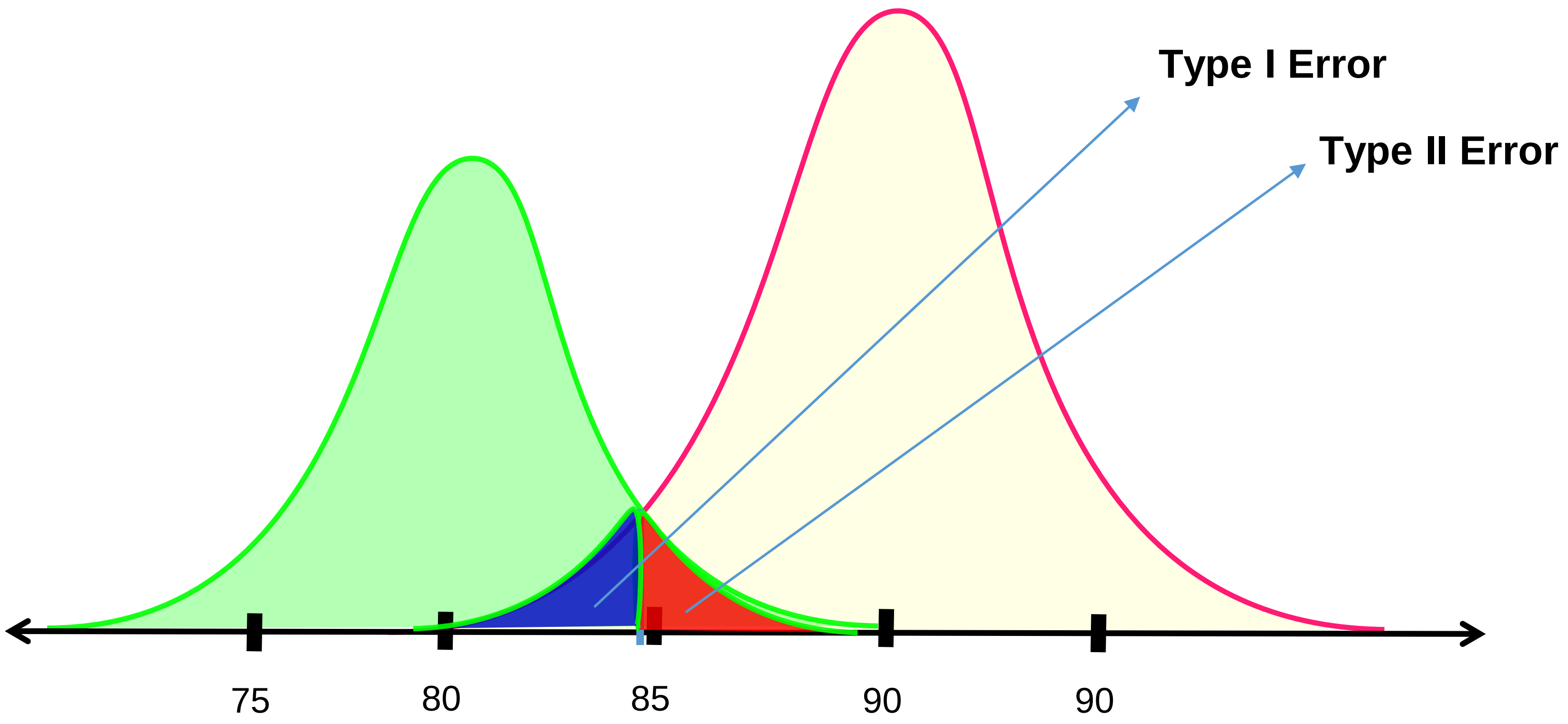
$H_o: p = 0.9$
$H_1: p = 0.8$

P(Z < c) = 0.05 for 5% Significance value. From probability tables, c = -1.64.

To de-standardize and find values outside the critical region, $\frac{X-90}{\sqrt{9}} \geq -1.64;$

X = 85.08, i.e., we would accept null hypothesis if 85.08 or more people had been cured.

Probabilities of Type I and Type II Errors



Probabilities of Errors in Our Example

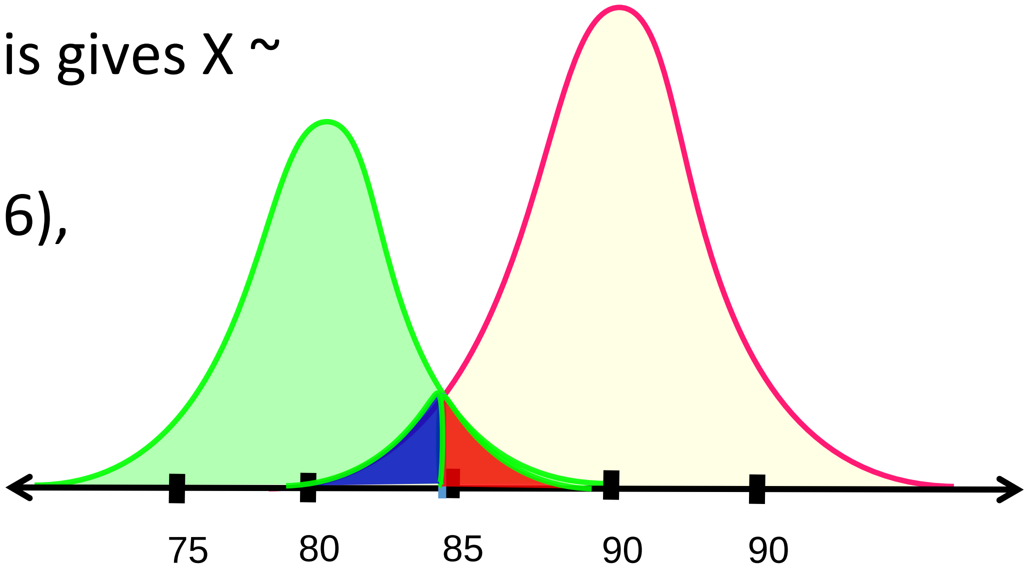
Finally, we need to calculate $P(X \geq 85.08)$, assuming H1 is true.

X ~ N(np, npq) where n=100 and p=0.8. This gives X ~ N(80, 16).

To calculate P(X = 85.08) where X ~ N(80, 16),

We find

$$z = \frac{85.08 - 80}{\sqrt{16}} = 1.27$$



P(Z = 1.27) = 1 – P(Z < 1.27) = 1-0.8980 = 0.102

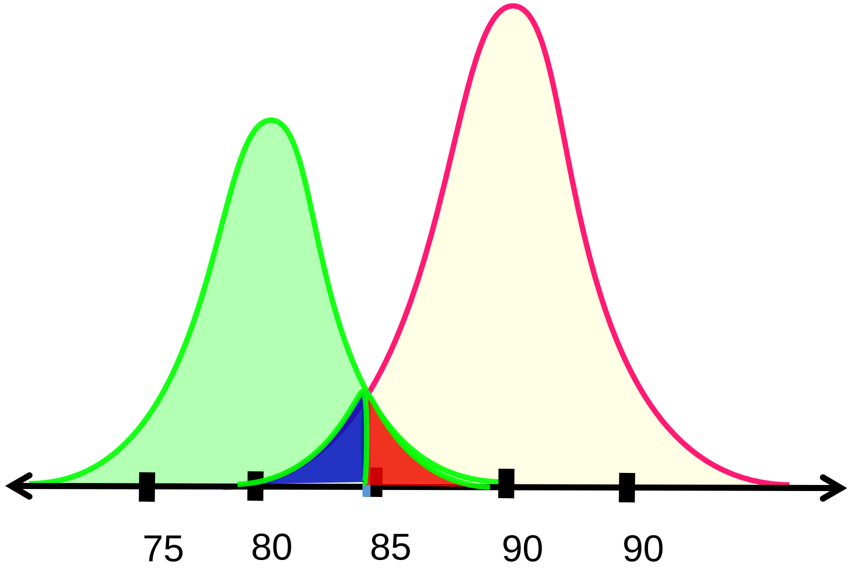
P(Type II error) = 0.102

The probability of accepting the null hypothesis that 90% are cured when its actually 80% is 10.2%.

Power of Hypothesis Test

We reject null hypothesis correctly when it is false.

It is actually the opposite of Type II error, and therefore,



Power = 1 – ß = 1-0.102 = 0.898, i.e., the probability that we will make the correct decision in rejecting the null hypothesis is 89.8%.

THANK YOU