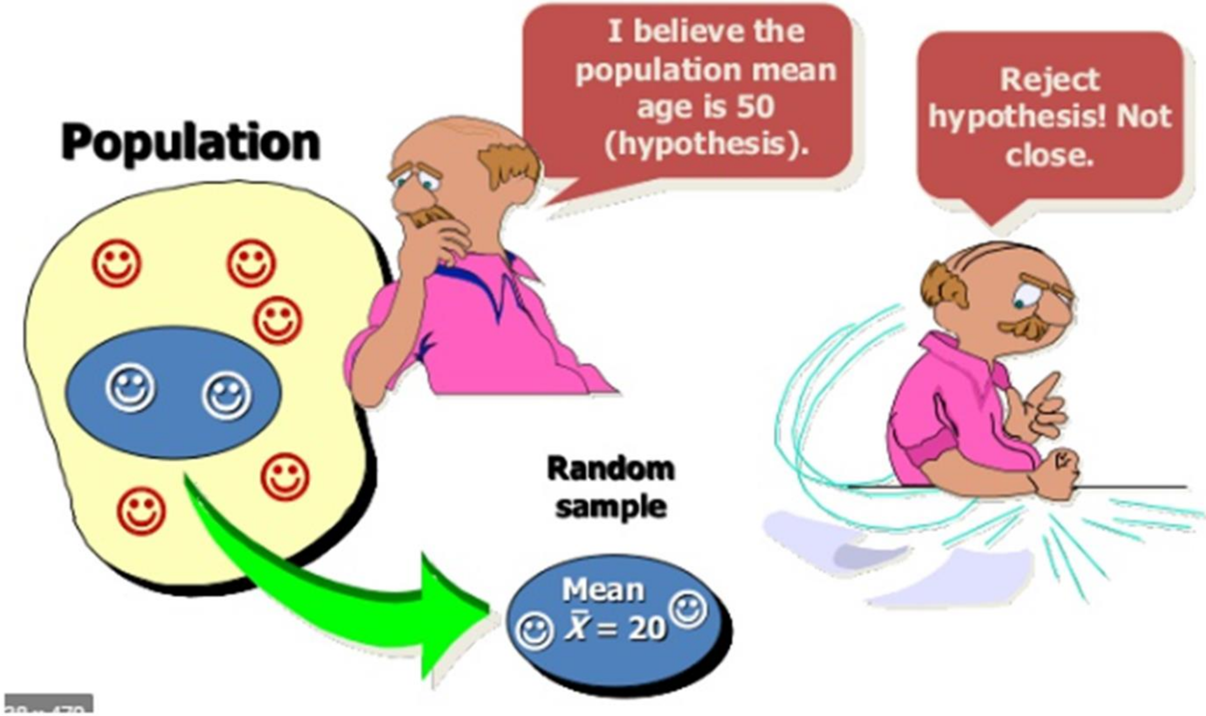
HYPOTHESIS TESTING

- Or ‘Significance Testing’
- These techniques are used to compare a **RANDOM SAMPLE to a POPULATION**

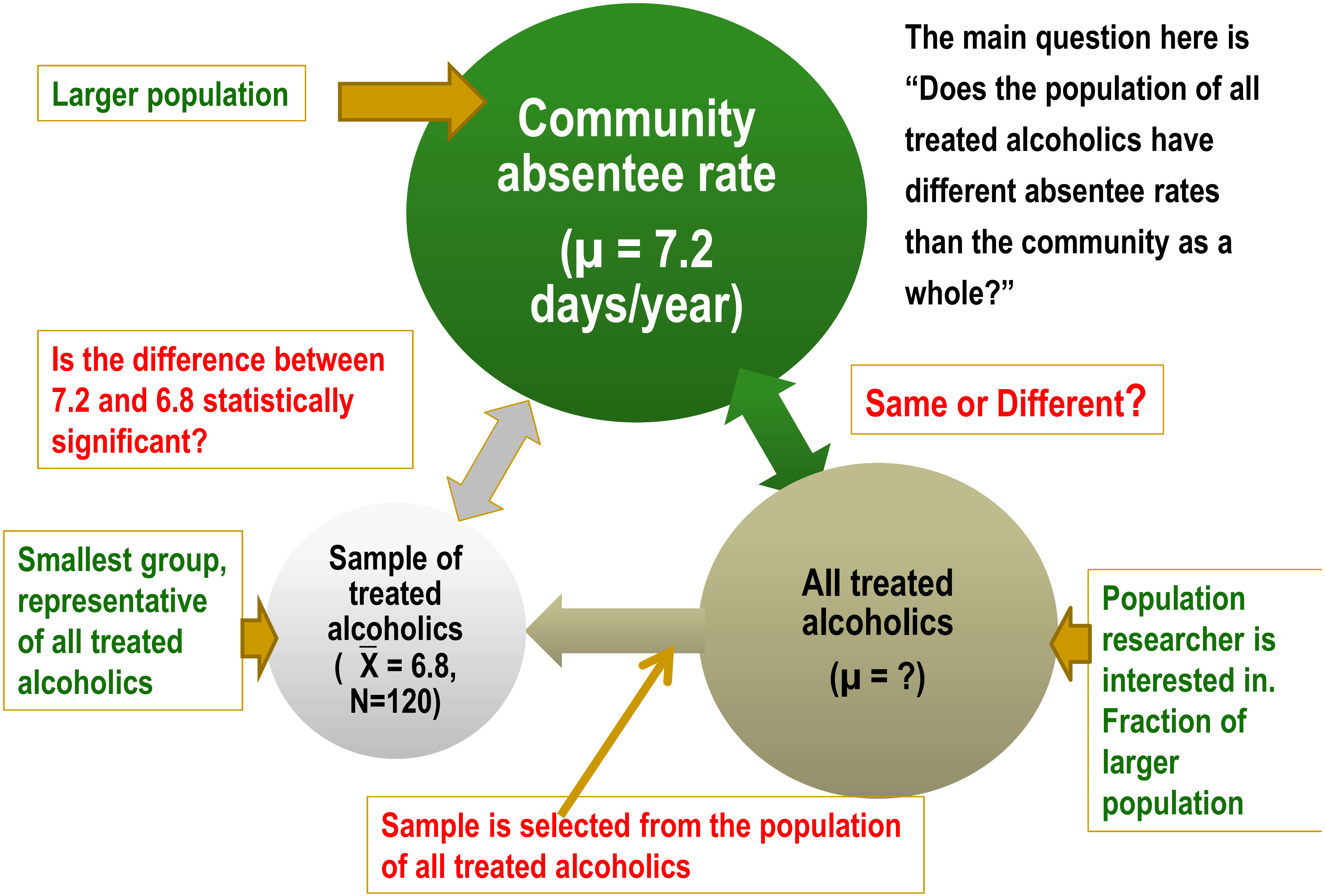


- We are **not interested in the sample per se** but in the larger group from which it has been selected
- We want to know **whether the group represented by the sample differs from the population parameters on a specific statistic??**

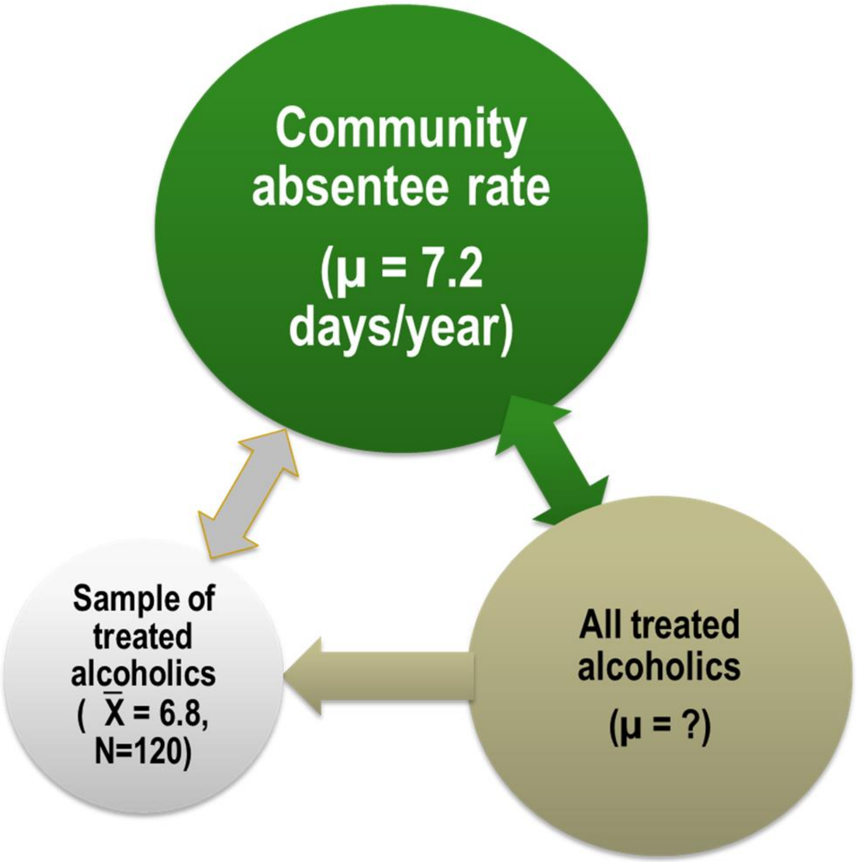
Example

- A researcher is assessing the effectiveness of a rehabilitation program for alcoholics in the city. She draws a random sample of 120 from the list of all clients and questions them.
- She finds that her sample misses fewer days of work than workers in the city as a whole.
- She will use a test to see whether the workers in her sample are more reliable than workers in the community as a whole

THE QUESTIONS



POSSIBLE EXPLANATIONS



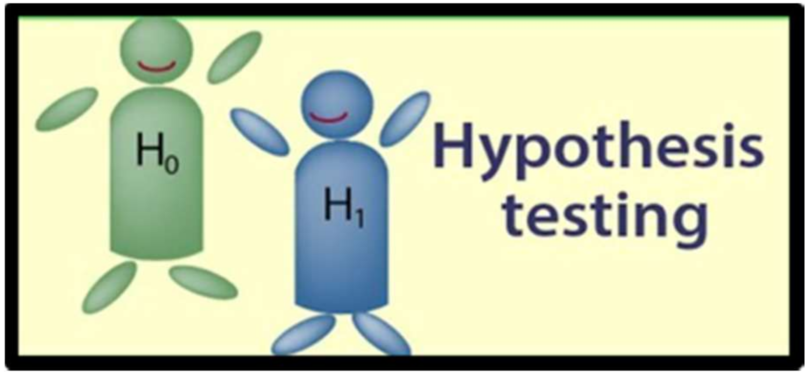
The researcher wants to know what caused the observed difference between the sample mean of 6.8 and community mean of 7.2?

There are TWO possible explanations for this difference

1. The observed difference is 'statistically significant'. The population of all treated alcoholics is different from the community as a whole. The sample did not come from a population with a mean absentee rate of 7.2 days
2. 'Null Hypothesis' – Observed difference between sample mean and community mean was caused by mere random chance. There is no difference between treated alcoholics and the community

Which explanation is correct???

THE DECISION MAKING PROCESS

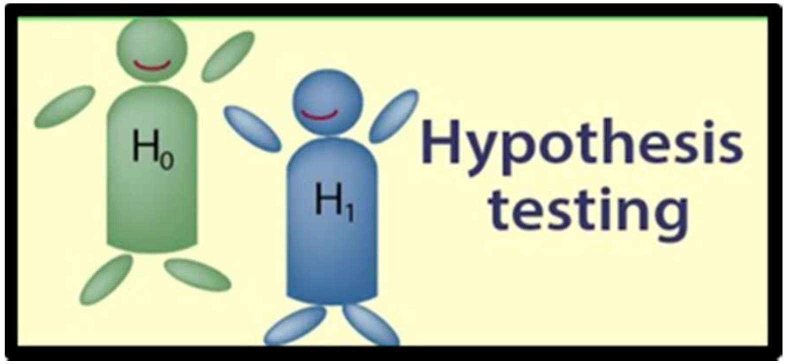


- We set up a decision making procedure **SO CONSERVATIVE** that one of the explanations can be chosen.
- In such a way **that the probability of choosing the wrong explanation is very low**
- **We begin with an assumption** that the second explanation - the NULL HYPOTHESIS – is correct
- Symbolically it can be stated as

$$H_0 : \mu = 7.2 \text{ days per year}$$

(*The population of treated alcoholics is not different from the community as a whole and has a μ of 7.2*)

THE DECISION MAKING PROCESS



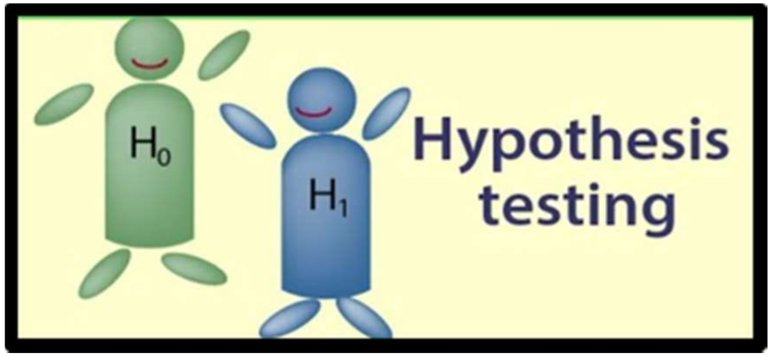
Ho : μ = 7.2 days per year

(*The population of treated alcoholics is not different from the community as a whole and has a μ of 7.2*)

- We now find the probability of getting the observed sample outcome ($\bar{X} = 6.8$).
- We set up a **DECISION RULE**
- If the odds of getting the observed difference are less than 0.5 (5/100), we will reject Ho
- If this is true (we reject Ho), the observed difference (7.2 days Vs. 6.8 days) would be a very rare event .

IN HYPOTHESIS TESTING WE ALWAYS BET AGAINST RARE EVENTS.

THE DECISION MAKING PROCESS



Ho : μ = 7.2 days per year

(*The population of treated alcoholics is not different from the community as a whole and has a μ of 7.2)*

- We estimate the probability of getting the observed sample outcome (X = 6.8), if Ho is correct.
- We use the sampling distribution of all possible sample outcomes
- Recall the <u>CENTRAL LIMIT THEOREM</u>
- N= 120; We can assume that the sampling distribution is normal in shape
- Recall the <u>FIRST THEOREM</u>
- Hence Mean of the Sampling Distribution = Population Mean; SE = (σ/√120)
- The observed sample outcome ($\bar{X}$ = 6.8) is one of thousands of sample outcomes

THE DECISION MAKING PROCESS

Ho : μ = 7.2 days per year

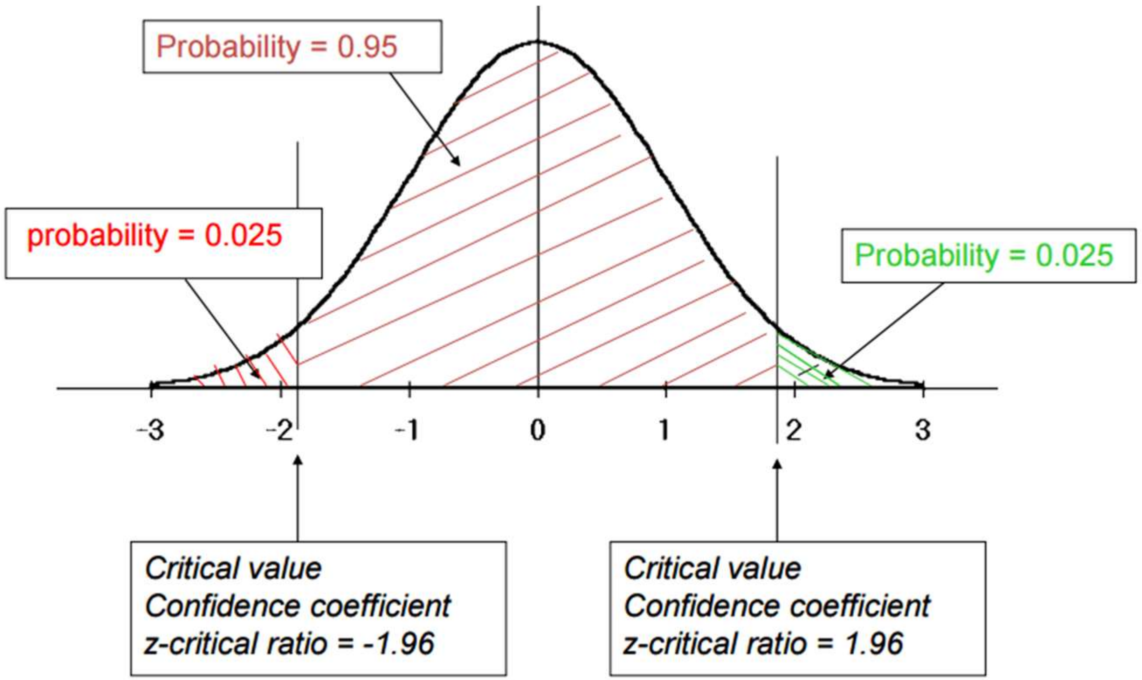
(The population of treated alcoholics is not different from the community as a whole and has a μ of 7.2)

- We now ascertain **the probability of getting a sample mean = 6.8 from the population with μ = 7.2**

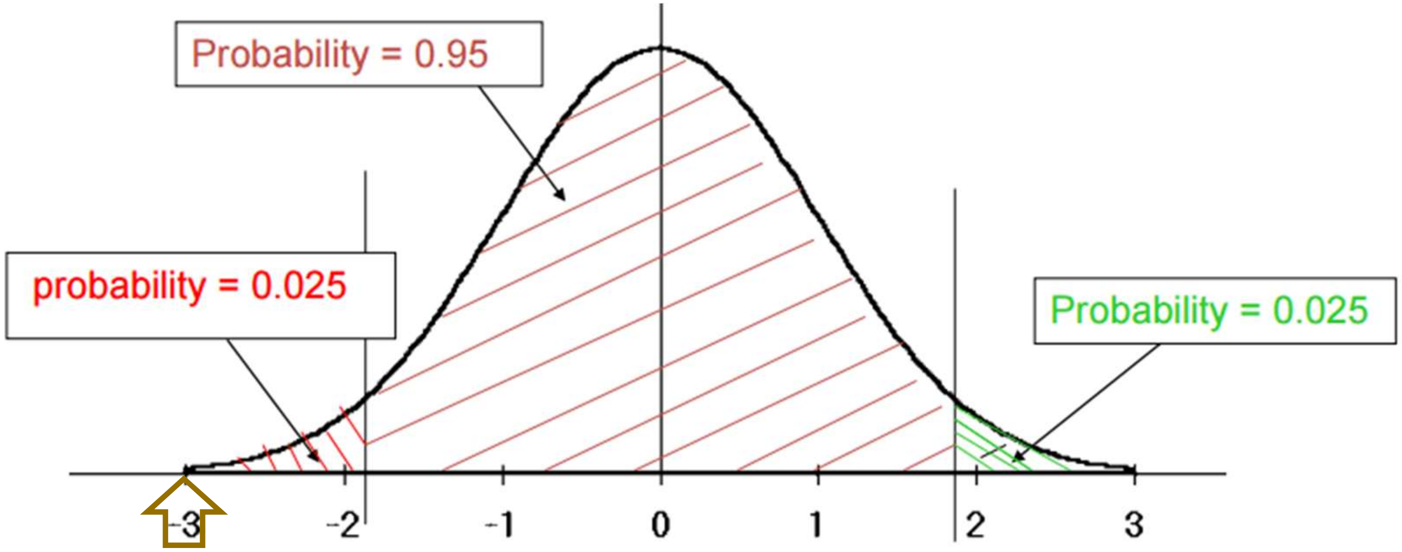
The probability of 0.05 is divided equally into upper and lower tails of the distribution

We find that the Z-Score equivalent of this area is ± 1.96

Any sample outcome falling in the red or green shaded area has a probability < 0.05. Such an outcome would be rare and will cause to reject Ho.



THE DECISION MAKING PROCESS

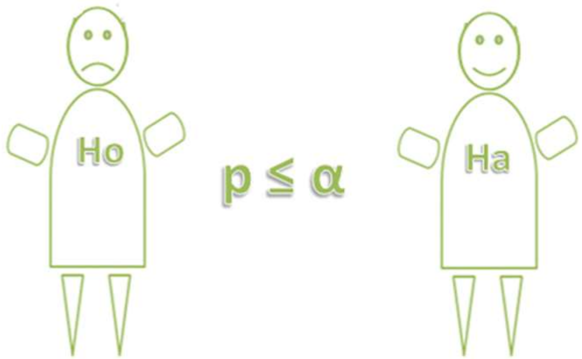


6.8 will lie somewhere here   μ = 7.2   Z = ($\bar{X}_i$ - μ) / SE

- Translate the sample outcome ($\bar{X}$ = 6.8) into Z- score and locate where it falls on the curve
- We see that the sample outcome of 6.8 falls in the RED shaded area.
- Thus if Ho is true, the probability of getting a sample outcome of 6.8 is less than 0.05
- The researcher may reject the NULL HYPOTHESIS

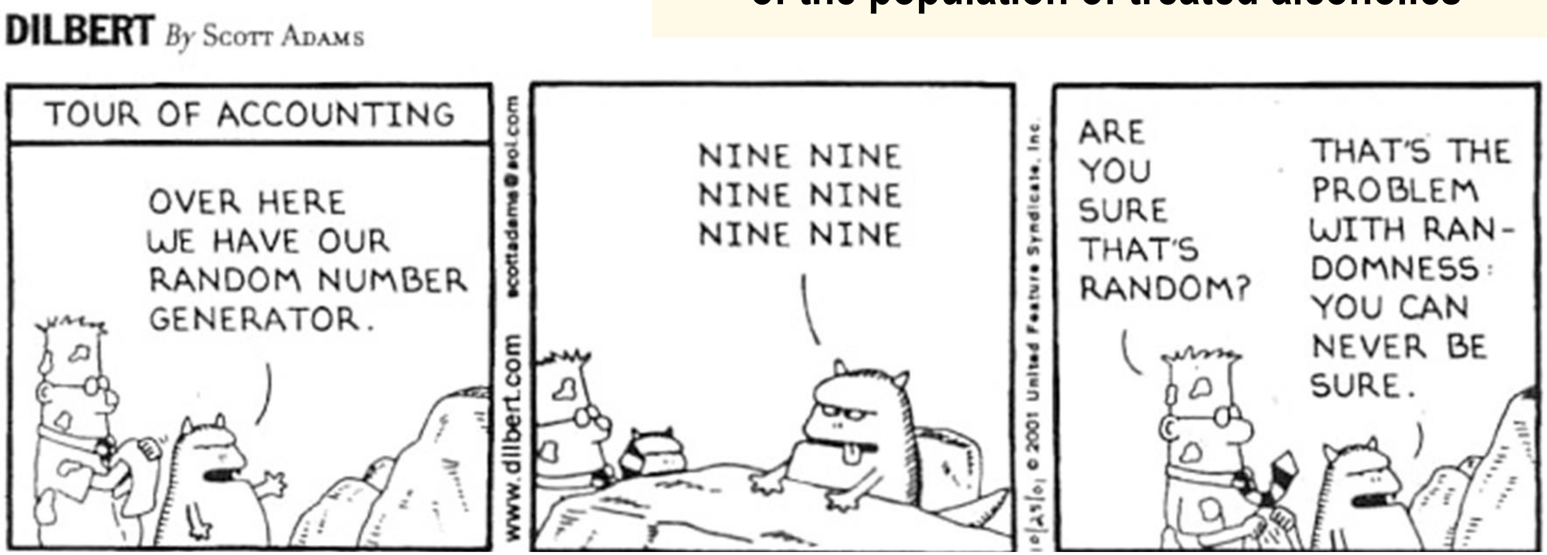
CONCLUSION

The sample does not come from a population that has a mean of 7.2 days of absence



We have used a sample to represent a larger group, and have compared the characteristics of SAMPLE with those of POPULATION

- Remember that our decision is based on the assumption that the sample is RANDOM
- It also assumes that the sample is representative of the population of treated alcoholics



CONCLUSION

The sample does not come from a population that has a mean of 7.2 days of absence

- There is a risk that the decision to reject the H_0 is incorrect if the sample happens to be one of the few that is unrepresentative of the population of treated alcoholics
- We can estimate the probability of taking a wrong decision
- This probability is 0.05 meaning that if we repeated the same test an infinite number of times, we would wrongly reject the null hypothesis only 5 times out of every 100.

Concepts and Procedures

THE FIVE STEP MODEL

OF HYPOTHESIS

TESTING

HYPOTHESIS TESTING

- **What do you understand by " Hypothesis" ?**

A proposition whose truth and falsity is capable of being tested

- **Hypothesis testing is a fundamental way in which inferences about a population are made from a sample**

- **The researcher's belief is stated in Research Hypothesis (H_1), a statement that directly contradicts the NULL HYPOTHESIS**

- **The NULL HYPOTHESIS is always a statement of NO DIFFERENCE. Its exact form will vary depending on the test being conducted**

STEPS

1. Making assumptions and meeting test requirements
2. Set up a null hypothesis –Concept of no difference

   Accepting Null hypothesis keeps us with the default option i.e. the sample statistics is no different from the population statistic

   Rejecting null hypothesis occurs when we find significant difference from the population statistic
3. State an alternative hypothesis – one-sided/ two sided
4. Gather evidence in favor of or against the hypothesis
5. The role of statistics is to inform us regarding precisely how unusual it would be to obtain our sample if the null hypothesis were true.

STEP-1

1. **Making assumptions and meeting test requirements**

THREE ASSUMPTIONS have to be satisfied

1. **The sample has to be selected according to the rules of EPSEM**
2. **The variable being tested is interval – ratio in level of measurement**
3. **The sampling distribution of sample means is normal in shape (can be satisfied by using large samples)**

STEP-2

2. Stating the Null Hypothesis

The NULL hypothesis is the central element in any test of hypothesis because the entire process is aimed at rejecting or failing to reject Ho.

Very often, the researcher's goal is to support the research hypothesis by rejecting the NULL HYPOTHESIS

Research hypothesis is stated as (continuing with the previous example)

$$(H_i: \mu \neq 7.2)$$

It is enclosed in parentheses to emphasize that it has no formal standing in the hypothesis-testing process (except for in choosing between one-tailed and two-tailed tests). It serves as a reminder what the researcher believes to be the truth.

STEP-3

3. Selecting the Sampling Distribution and Establishing the Critical Region

By assuming that the Null Hypothesis is true, we can attach values to the Mean and SD of the sampling distribution and measure the probability od any specific sample outcome.

The sampling distribution is different for different tests.

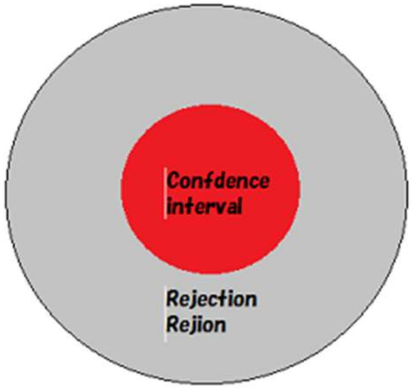
The CRITICAL REGION or the REGION OF REJECTION consists of the areas under the sampling distribution that include unlikely sample outcomes. It is demarcated by the CRITICAL VALUES of the TEST STATISTIC at the selected LEVEL OF CONFIDENCE (α)

In our example of sampling distribution of sample means

Sampling distribution – Z distribution

$\alpha = 0.05$

Z (Critical) = ± 1.96



STEP-4 Computing the Test Statistic

The likelihood of rejecting a true hypothesis is referred to as *significance level*

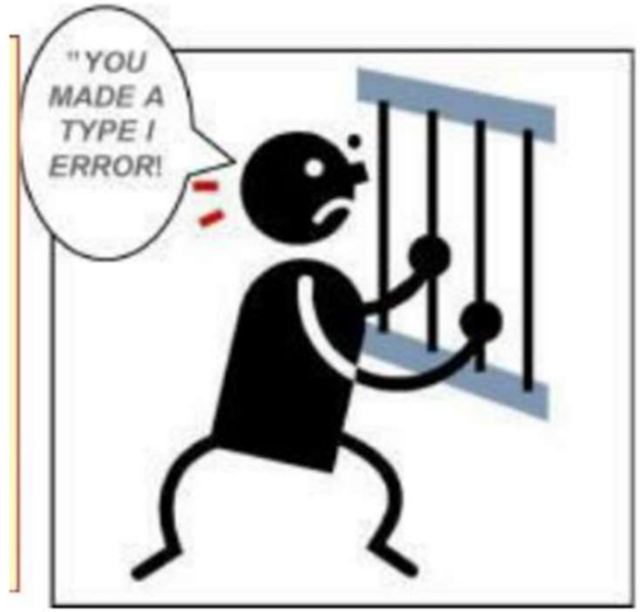
To evaluate the probability of the sample outcome, we CONVERT THE SAMPLE VALUE into test statistic (Z- Score in our example).

- Choose a significance level - Common values are 0.01, 0.05 and 0.10
- Choose a test statistic and find the *observed value* of the test statistic.
- Determine the *critical value* of the test statistic and compare with the observed value
- *Critical values* are those values of the test statistic where we are on the knife edge between acceptance and rejection. If the observed test statistic is slightly to one side of the critical value, we accept H_0; if it is slightly to the other side, we reject H_0.

STEP-5

4. Making a Decision and Interpreting the Results

- **HOW LARGE** must the observed value of test statistic be before we reject H*o*???
- **HOW OFTEN** would we observe a value as high as our observed value under H*o*???



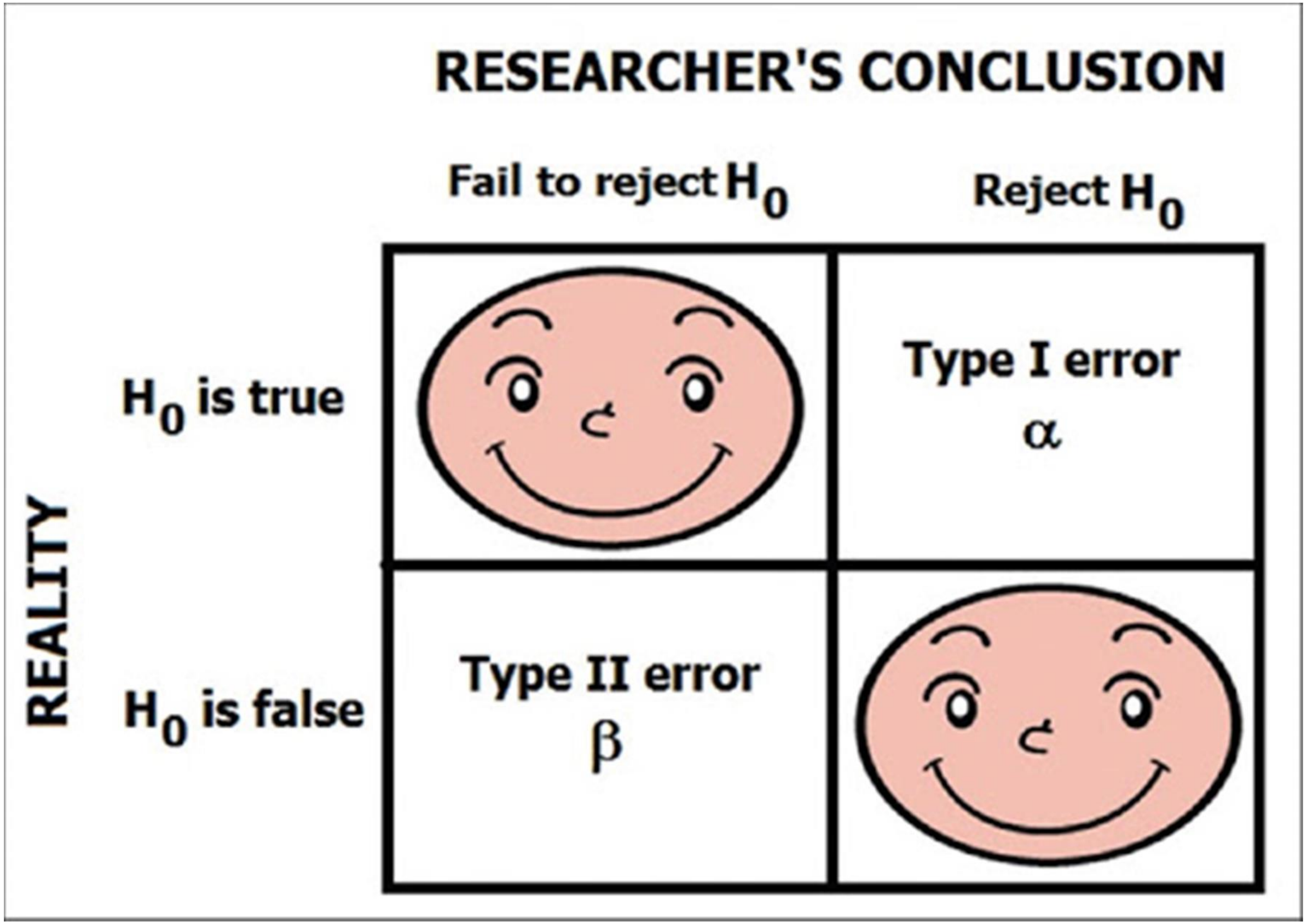
- **Decide to accept or reject H_0**

  H_0 accepted if the likelihood of a result more extreme than the one we observed is more than 5% - the result is not uncommon.

  H_0 rejected if experiment indicates that result is uncommon. Probability of encountering extreme result than the one we got is less than 5 %.

- If observed value > critical value at chosen α (say 0.05), reject H_0

Situation	Decision	Interpretation
The test statistic is in the critical region	Reject Ho	The difference is statistically significant
The test statistic is NOT in the critical region	Fail to reject Ho	The difference is not statistically significant


RESEARCHER'S CONCLUSION
Fail to reject H0
Reject H0
REALITY
H0 is true
H0 is false
Type I error
α
Type II error
β

Interpretation

Ho : μ = 7.2 days per year

(*The population of treated alcoholics is not different from the community as a whole and has a μ of 7.2*)

- **Interpret the decision in terms of the original question.**

For example, in previous example

Z (obs)> Z (Critical)

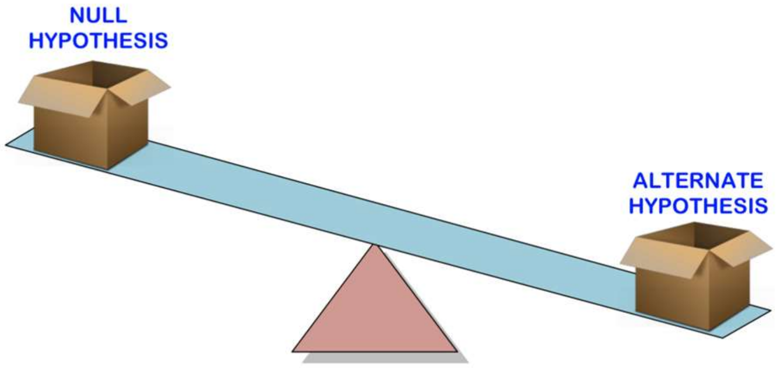
Researcher's decision would be to reject the Ho.

The decision says that

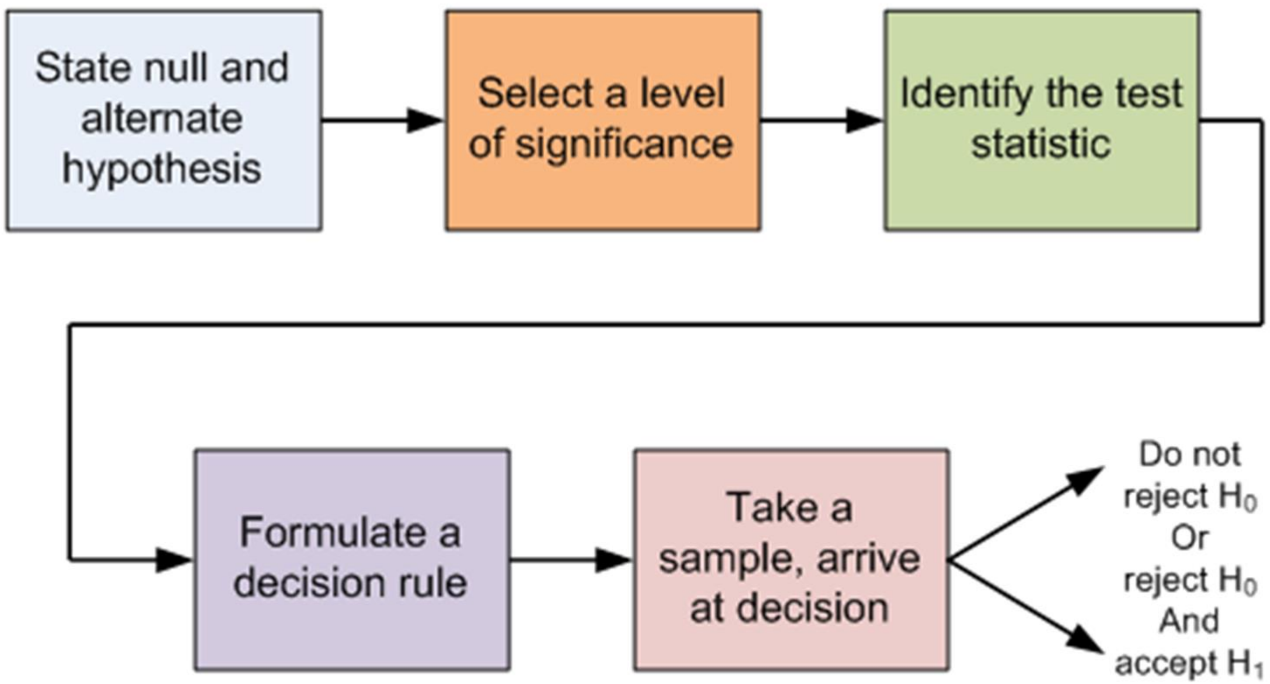
1. **Treated alcoholics do not have a mean absentee rate of 7.2 days.**
2. **There is a difference between them and the community.**
3. **Difference between the sample mean and the community mean is statistically significant at selected α.**
4. **The observed difference is unlikely to be caused by random chance alone**

- p- value tells us how likely a result more extreme than the one we observed would be if the null hypothesis were true

- For e.g. A p-value of 0.0139 in a two sided alternative (two-tailed probability distribution) implies that, if null hypothesis were true, the probability of occurrence of a value as high or higher than the corresponding observed value is 2(0.0139) = 0.0278



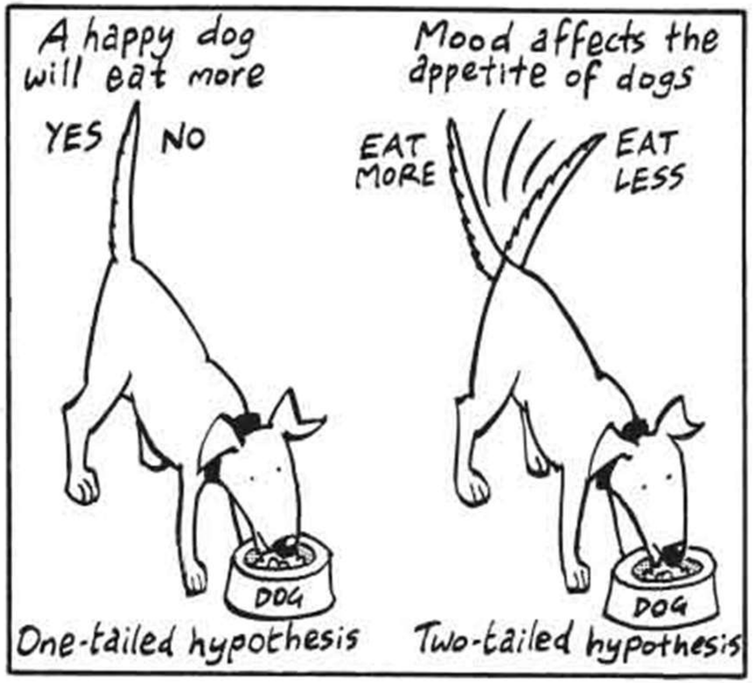
- This five step model serves as a common framework for decision making in all hypothesis testing procedures
- The exact nature and method of expression for decisions is different for different research questions



TWO CRUCIAL DECISIONS

In the procedure of hypothesis testing, the researcher has to make two decisions regarding the test:

1. One - tailed or two- tailed test
2. Alpha level

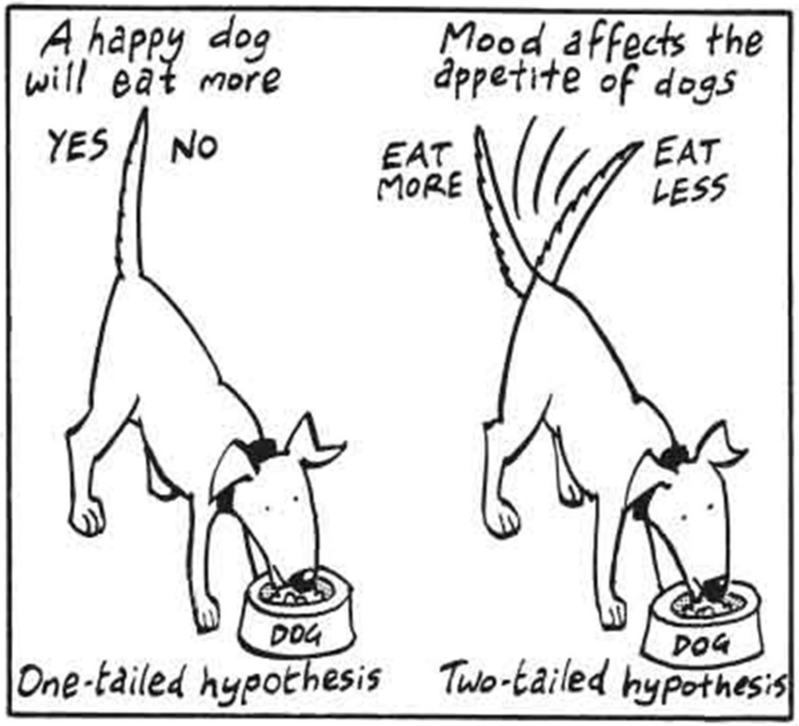


TYPES OF SIGNIFICANCE TESTS

Choice between a one – and two- tailed test is based on researcher's expectations about the population from which the sample was selected.

What the researcher believes to be 'the truth'??

These expectations are reflected in the research hypothesis (H_1) / alternative hypothesis (H_a)



Thus choice of the form of test depends on the relationship between Ho and H_1

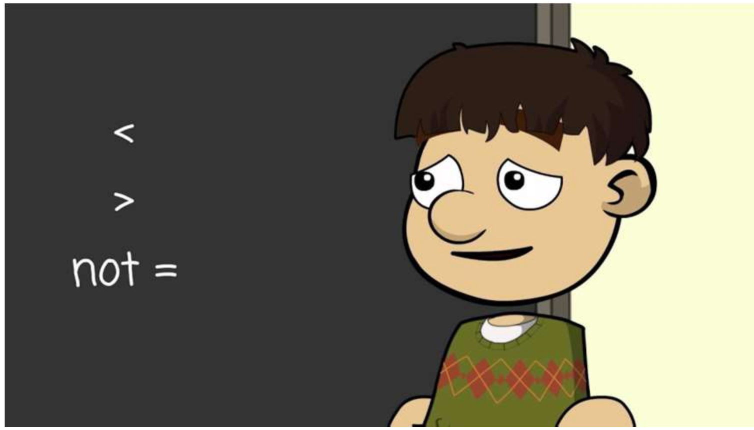
One Tailed & Two Tailed Tests

Consider the example of the population of treated alcoholics we have been discussing..

Ho = All treated alcoholics have the same absentee rate as the workers' community (7.2 days/Yr)

What does the researcher believe (H_a)?

1. **The population of treated alcoholics has LESS absenteeism (Their population mean is < 7.2)**
2. **….MORE absenteeism ($\mu > 7.2$)**
3. **He/ she might be unsure about the direction of difference (may be high/ may be low) but the population mean is NOT EQUAL to the value stated in the Ho ($\mu \neq 7.2$).**



Two Tailed Test

Researcher will be equally concerned with the possibility that the true population value is greater than or less than the value specified in the NULL HYPOTHESIS

The research hypothesis will be (For example)

$$H_1: \mu \neq 7.2$$

ONE Tailed Test

If the direction of difference can be predicted, or if the researcher is concerned only with the differences in one direction one tailed test is used.

One tailed test may take one of the two different forms:

1. **If the researcher believes that the true population value is greater than the value specified in the null hypothesis, H1 would use the '>' (greater than) symbol.**

   For eg. $H_1: \mu > 7.2$

   Researcher predicts that treated alcoholics had higher absenteeism than the community

2. **If the researcher predicted that treated alcoholics had lower absentee rates than the community, H1 would have been**

   $H_1: \mu < 7.2$

When to use One Tailed Test ?

Appropriate when programs designed to solve a problem or improve a situation are being evaluated.

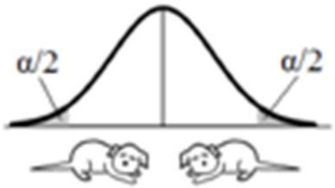
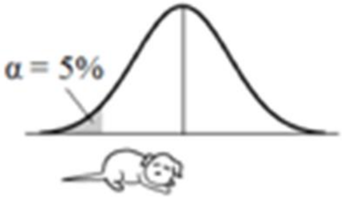
In such situations the researcher may focus only on the outcomes (direction) that would indicate that the program is a success.

Consider that a program designed to reduce unemployment is being evaluated. What form of one-tailed test will be used?

*H1 : "Unemployment rates for youth benefitted from the program are (*Greater than or Less than?) *unemployment rates in the community."*

TYPE OF TEST AND THE 5 STEP MODEL

- The choice of one-tailed or two-tailed test determines what we do with the critical region in STEP 3.

Tails of the Test	
2-tailed	
1- tailed, right-tailed	
1-tailed, left-tailed	

- In a two tailed test we split the critical region equally into the upper and lower tails of the sampling distribution.

- In a one-tailed test we put the entire area in either the upper or lower tail

TYPE OF TEST AND CRITICAL REGION

One-Tailed Test (Left Tail)	Two-Tailed Test	One-Tailed Test (Right Tail)
$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X < \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X \neq \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X > \mu_0$

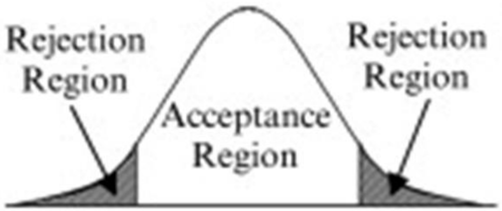
If H_1 includes < (less than) symbol, the entire critical region goes to the lower tail

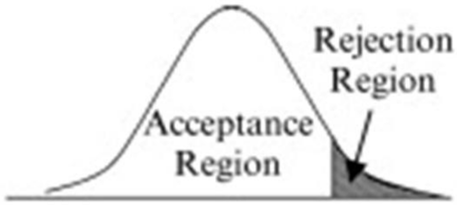
If H_1 includes $\neq$ (not equal to) symbol, the critical region is split among both the tails.

If H_1 includes > (greater than) symbol, the entire critical region is placed in the higher tail

Critical Values – Z Distribution

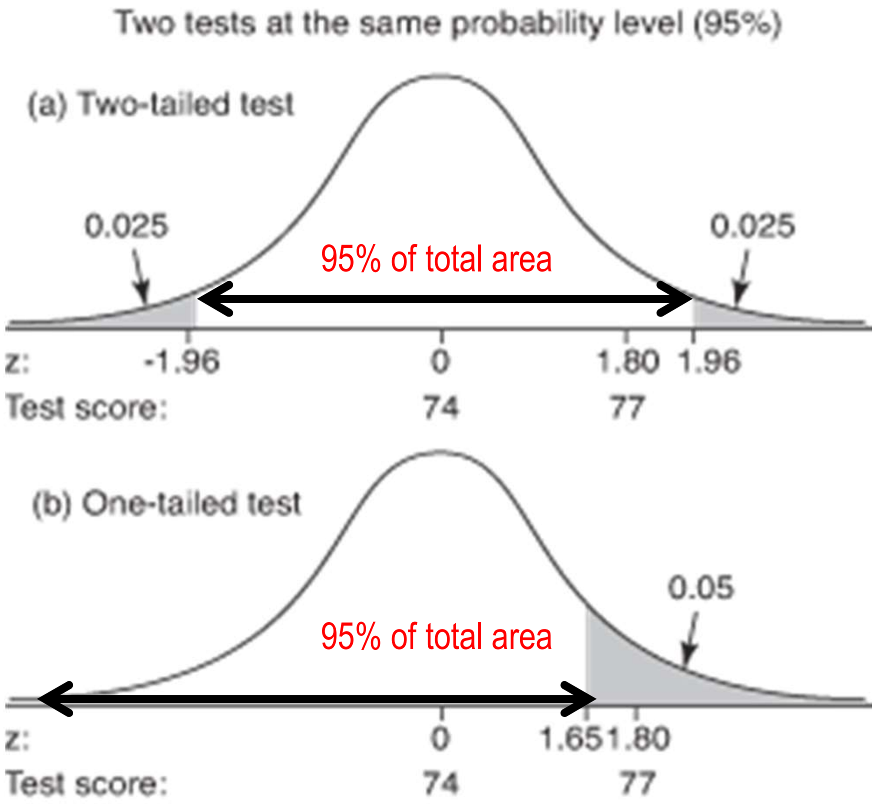
One-Tailed Test (Left Tail)	Two-Tailed Test	One-Tailed Test (Right Tail)
$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X < \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X \neq \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X > \mu_0$





If H1 uses	The test is	Concern is on	Z (critical) is
≠	Two -tailed	Both tails	**± 1.96**
<	One- tailed	Lower tail	**- 1.65**
>	One tailed	Upper tail	**+ 1.65**

Z Distribution, α-level and Critical Values



Thus one tailed test is more likely to reject H_0 without changing the alpha level.

At the given α- level (0.05)

In a two tailed test Z (Critical) = ± 1.96

In a one-tailed test Z (Critical) upper tail = + 1.65

In a one-tailed test Z (Critical) lower tail = - 1.65

Thus at a given α level the critical Z – values for one tailed tests are closer to the mean of the sampling distribution.

1- & 2-tailed test, α-level and Z scores

Alpha	2-tailed value	One-tailed value	
		Upper tail	Lower tail
0.10	± 1.65	+ 1.29	- 1.29
0.05	± 1.96	+ 1.65	- 1.65
0.01	± 2.58	+ 2.33	- 2.33
0.001	± 3.32	+ 3.10	- 3.10
0.0001	± 3.90	+ 3.70	- 3.70

One tailed test is used only when

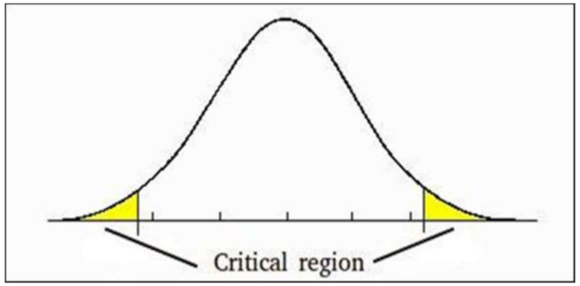
- The direction of difference can be confidently predicted
- The researcher is concerned only with differences in one tail of the sampling distribution

Selecting an alpha level

- Alpha level plays a crucial role in hypothesis testing.
- When we assign a value to alpha, we define what we mean by 'unlikely' sample outcome.
- If the probability of observed sample outcome is lower than the alpha level, we reject the NULL HYPOTHESIS.
- Besides to this , alpha level is the probability that the decision to reject the null hypothesis (if the test statistic falls in the critical region) will be incorrect (Type 1 error).

How to make a reasonable decision about the value of alpha?

Selecting an alpha level



- **When we select an alpha level we divide the sampling distribution into two sets of sample outcomes.**

1. **Critical region :** **includes all unlikely or rare sample outcomes**
2. **Non-critical region:** **Consists of all other sample outcomes which are not rare**

A lower α- level reduces the size of the critical region and moves it further from the mean of the sampling distribution

Thus lower α-levels make it less likely that we will commit a Type 1 error

Selecting an alpha level

- However a small α level increases the risk of falsely accepting the Ho - committing TYPE 2 error (β error)

Both errors cannot be minimized simultaneously

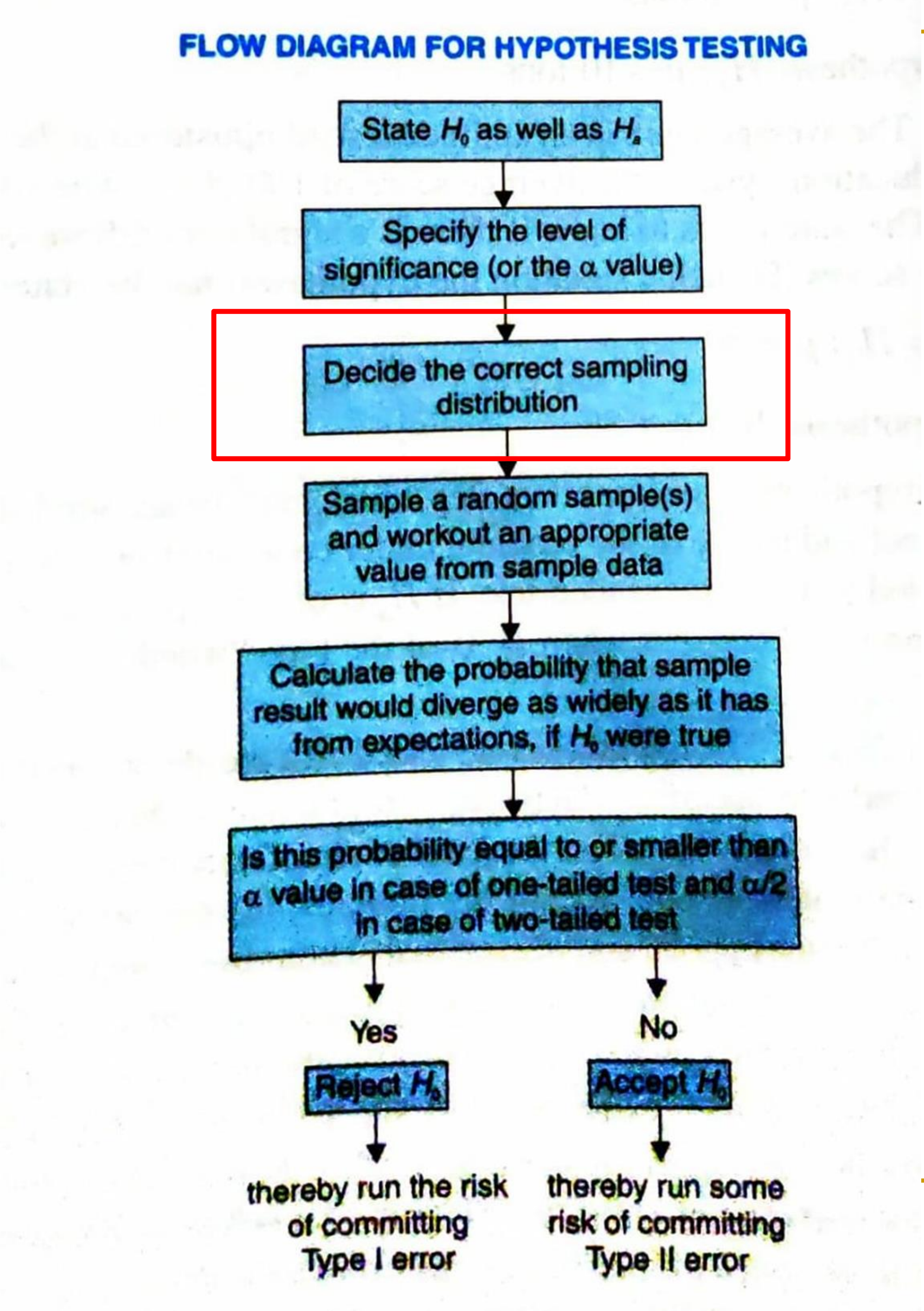
- Lower α level will minimize the probability of Type 1 error
- Higher α level will minimize the probability of Type 2 error

Both errors have to be balanced

Normally in social science research we want to minimize the Type 1 error – lower α levels will be used (0.05, 0.01, 0.001)

0.05 has become a standard indicator of significant result. However there is no reason.

SELECT THE α-LEVEL THAT SEEMS MOST REASONABLE IN TERMS OF GOALS OF THE RESEARCH PROJECT



TYPES OF TESTS

Tests may broadly be classified as:

1. **Parametric tests** – Standard tests of hypothesis
2. **Non-Parametric tests** – Distribution free tests of hypothesis

PARAMETRIC TESTS

Usually assume certain characteristics of the parent population from which we draw samples. These are pre-requisite `:

- Observations come from a normal population
- Sample size is large
- Assumptions about population parameters – mean, variance
- Require measurement at least at interval scale
- Z-test, t- test, F-test
- Chi-square test as a good ness of fit

NON- PARAMETRIC TESTS

- Do not depend on any assumption about the parameters of parent population
- Assume only nominal or ordinal data
- For e.g. Chi-square test as a test of independence

TESTS OF SIGNIFICANCE

- There are many tests available to test the null hypothesis
- All the tests require random samples (or at least reasonably random samples)
- The most important factors that determine choice of a test are
  1. The descriptive statistics we are testing
  2. The number of samples from which inferences are being made
  3. Whether we have dependent or independent samples

TESTS OF SIGNIFICANCE

Descriptive Statistic and number of samples	Test of Significance
One sample mean	z-test for a mean (population variance known) t-test for a mean (population variance unknown)
Two independent sample means	t-test for equality of means
More than two independent sample means	ANOVA F-test for the equality of means
Two dependent sample means	t-test for the mean difference
Frequency table for one sample (multinomial scale)	Chi-square test for goodness-of-fit
Cross-tabulation for two or more independent samples	Chi-square test for independence

Source: George Argyrous (Statistics for Research)