Test of Hypothesis

Bijay Lal Pradhan, PhD

Hypothesis

Two words

Hypo: Under   Thesis: reasoned theory

Theory which is not fully reasoned

Tentative answer of the research question

Imaginative idea or guess depending upon previous accumulated knowledge which can be put to test to determine its validity.

Generally specify relationship between variables or with specific value
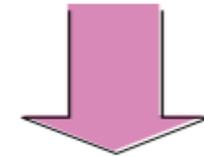
Test of Hypothesis

Assume the population mean age is 50. (Null Hypothesis)

Population
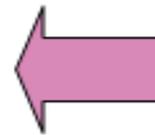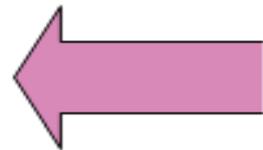
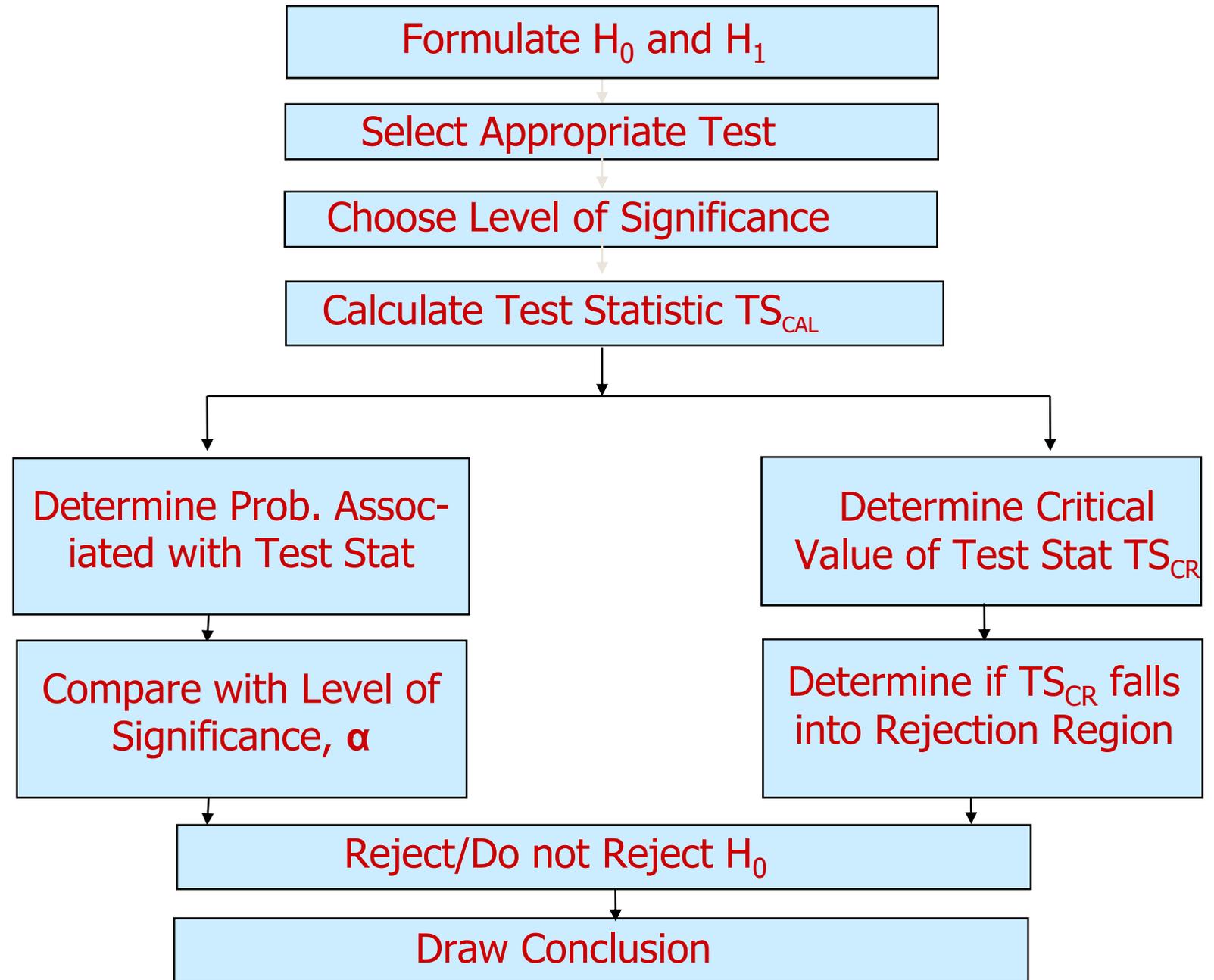Is $\overline{X} = 20 \cong \mu = 50$?

The Sample Mean Is 20

Sample

REJECT

Null Hypothesis

Steps in Test of Hypothesis

Formulate H_0 and H_1

Select Appropriate Test

Choose Level of Significance

Calculate Test Statistic TS_{CAL}

Determine Prob. Associated with Test Stat

Determine Critical Value of Test Stat TS_{CR}

Compare with Level of Significance, α

Determine if TS_{CR} falls into Rejection Region

Reject/Do not Reject H_0

Draw Conclusion
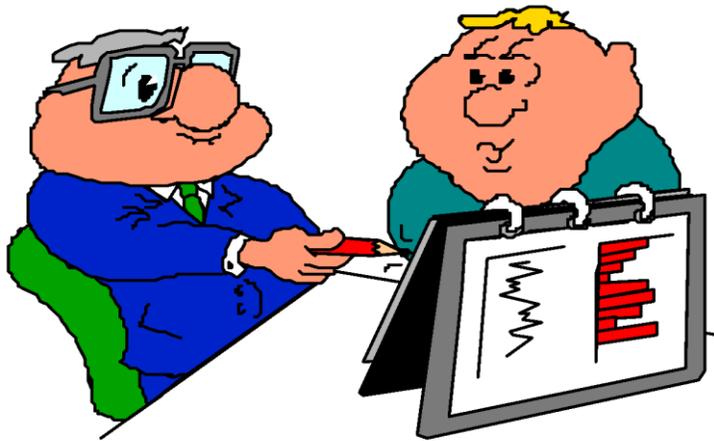
Interpreting the p-value…

- The smaller the p-value, the more statistical evidence exists to support the alternative hypothesis.
- If the p-value is less than 1%, there is *overwhelming evidence* that supports the alternative hypothesis.
- If the p-value is between 1% and 5%, there is a *strong evidence* that supports the alternative hypothesis.
- If the p-value is between 5% and 10% there is a *weak evidence* that supports the alternative hypothesis.
- If the p-value exceeds 10%, there is *no evidence* that supports the alternative hypothesis.

Interpreting the p-value…

Overwhelming
Evidence
(Highly Significant)

Strong
Evidence
(Significant)

Weak
Evidence
(Significant)

No Evidence
(Not Significant)

0   .01   .05   .10

Conclusions of a Test of Hypothesis...

- If we reject the null hypothesis, we conclude that there is enough evidence to infer that the alternative hypothesis is true.

- If we fail to reject the null hypothesis, we conclude that there is not enough statistical evidence to infer that the alternative hypothesis is true. This does not mean that we have proven that the null hypothesis is true!

Hypothesis testing (Z,t F, Anova, Chi square)

- Z test
  - Test of significance of single mean
  - Test of significance of double mean
  - Test of significance of single proportion
  - Test of significance of double proportion

- T test
  - Test of significance of single mean
  - Test of significance of double mean (independent)
  - Test of significance of pair t-test (dependent two mean)

Hypothesis testing (Z,t F, Anova, Chi square)

- F test
  - Test of significance of more than two mean (ANOVA)
  - Test of significance of standard deviation

- Chi Square test
  - Test of significance of independence of attribute.
  - Test of goodness of fit.

Significance of mean / s

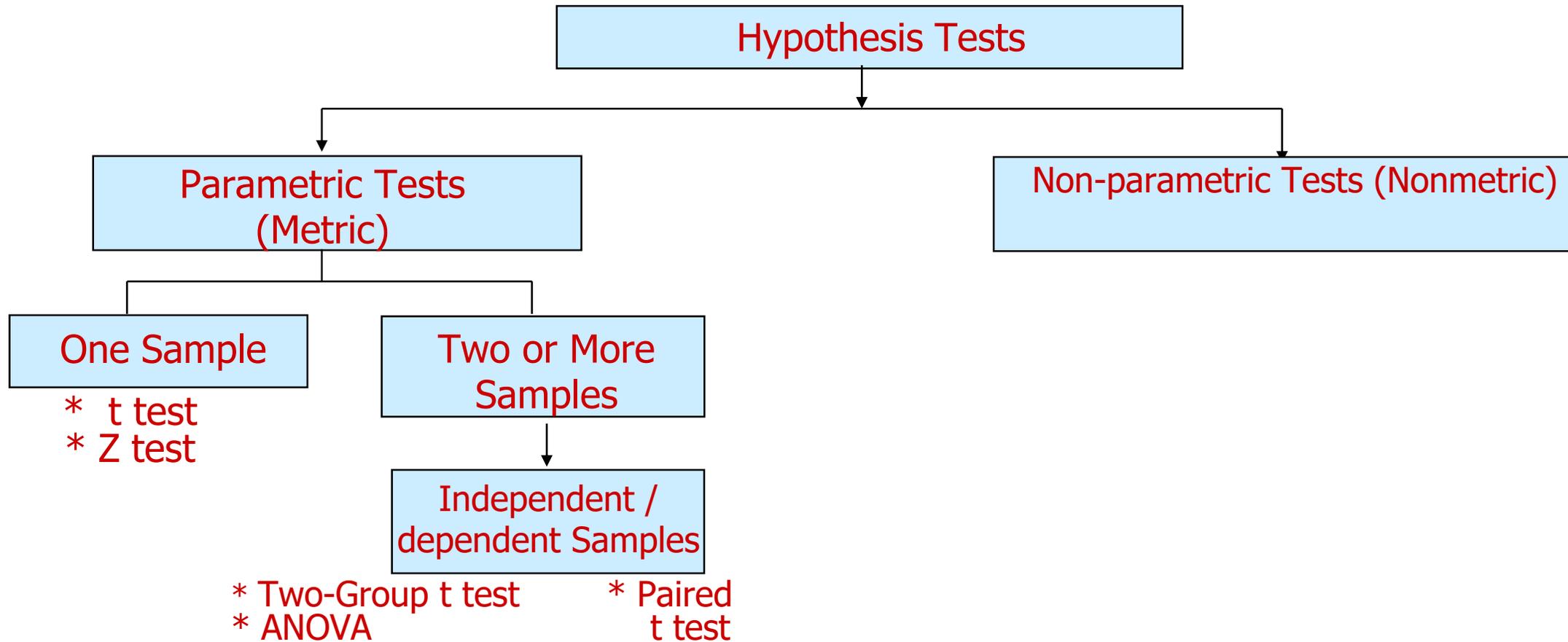| Single Mean (with specified value)	Double Mean	More than two Means
• Sample size more than 30 (Z- test) • Sample size less than or equals to 30 (T- test)	• Sample size more than 30 (Z- test) • Sample size less than or equals to 30 (T- test)	• ANOVA • Analysis Of Variance

Independent T test

Paired T test (dependent Observations)

Hypothesis Testing for Differences

```
                        ┌─────────────────────────┐
                        │    Hypothesis Tests     │
                        └─────────────────────────┘
                                     │
            ┌────────────────────────┴────────────────────────┐
            ▼                                                  ▼
┌─────────────────────────┐                  ┌──────────────────────────────────┐
│    Parametric Tests     │                  │  Non-parametric Tests (Nonmetric) │
│        (Metric)         │                  │                                   │
└─────────────────────────┘                  └──────────────────────────────────┘
            │
   ┌────────┴──────────┐
   ▼                   ▼
┌──────────────┐  ┌──────────────┐
│ One Sample   │  │ Two or More  │
│              │  │  Samples     │
└──────────────┘  └──────────────┘
                         │
   *  t test            ▼
   * Z test      ┌──────────────────┐
                 │  Independent /   │
                 │ dependent Samples│
                 └──────────────────┘

      * Two-Group t test        * Paired
      * ANOVA                       t test
```
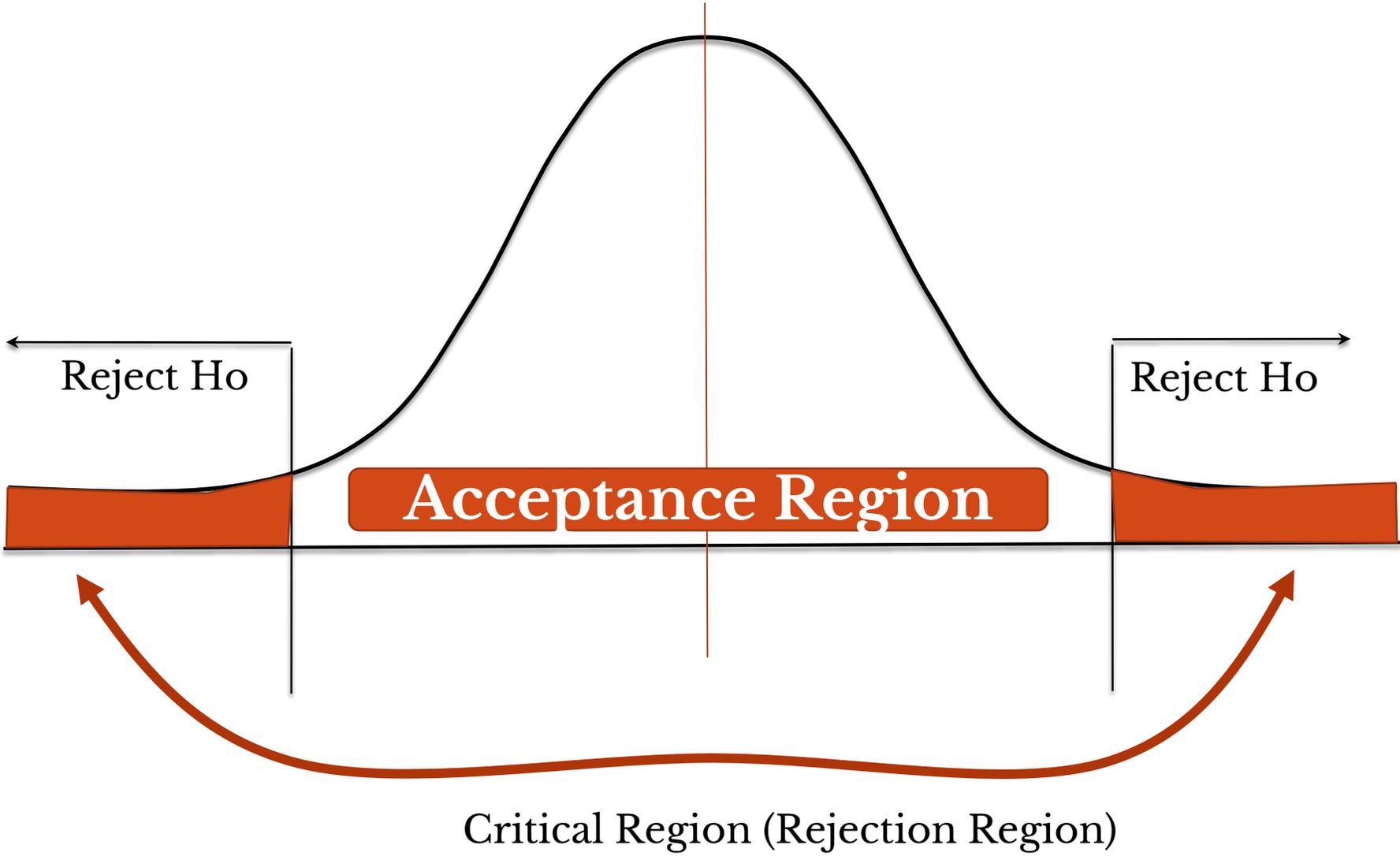
Central Limit Theorem

When sampling from a population with mean μ and finite standard deviation σ, the sampling distribution of the sample mean will tend to a normal distribution with mean μ and standard deviation σ/√n   as the sample size becomes large (*n* >30).
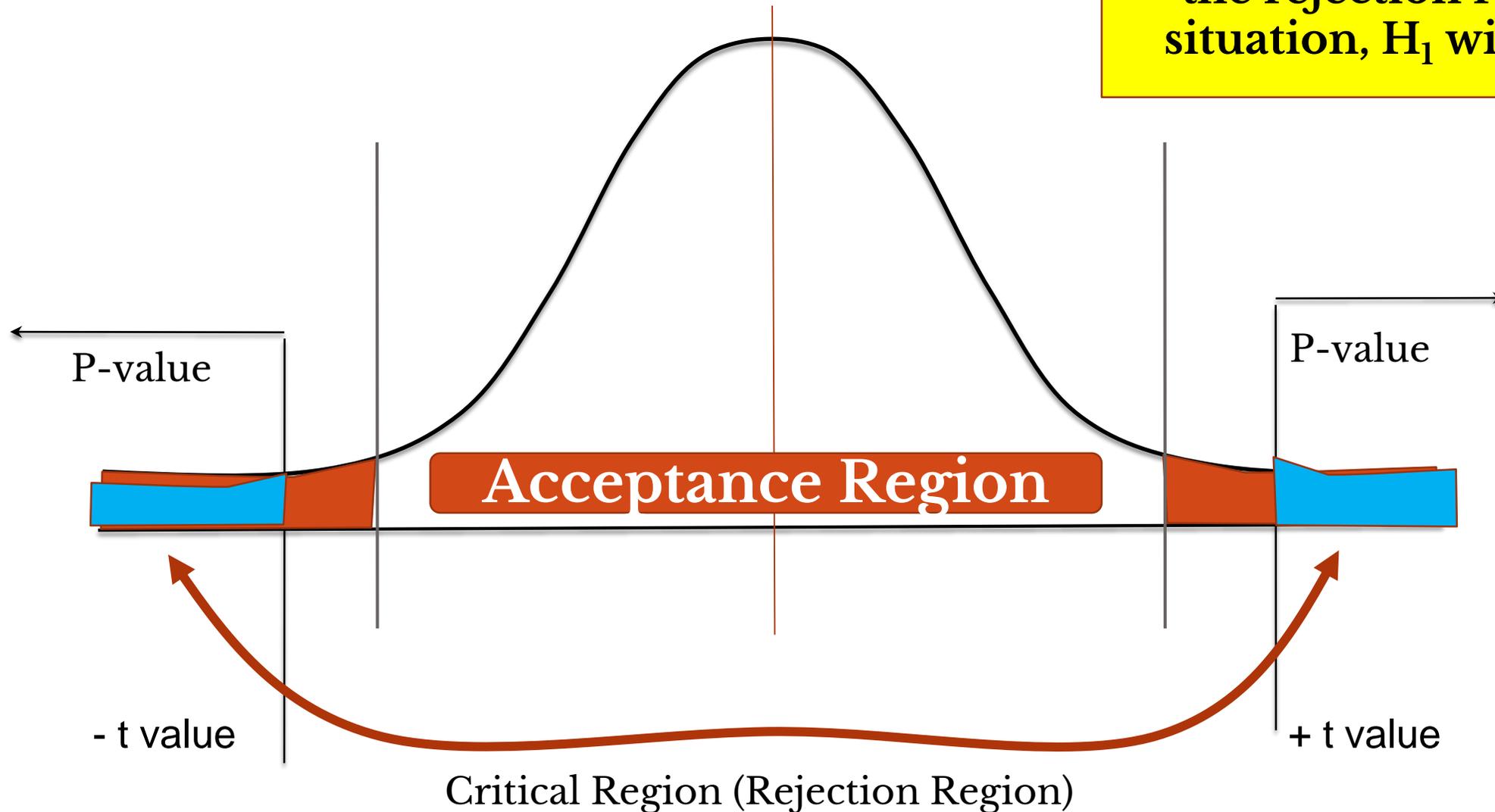
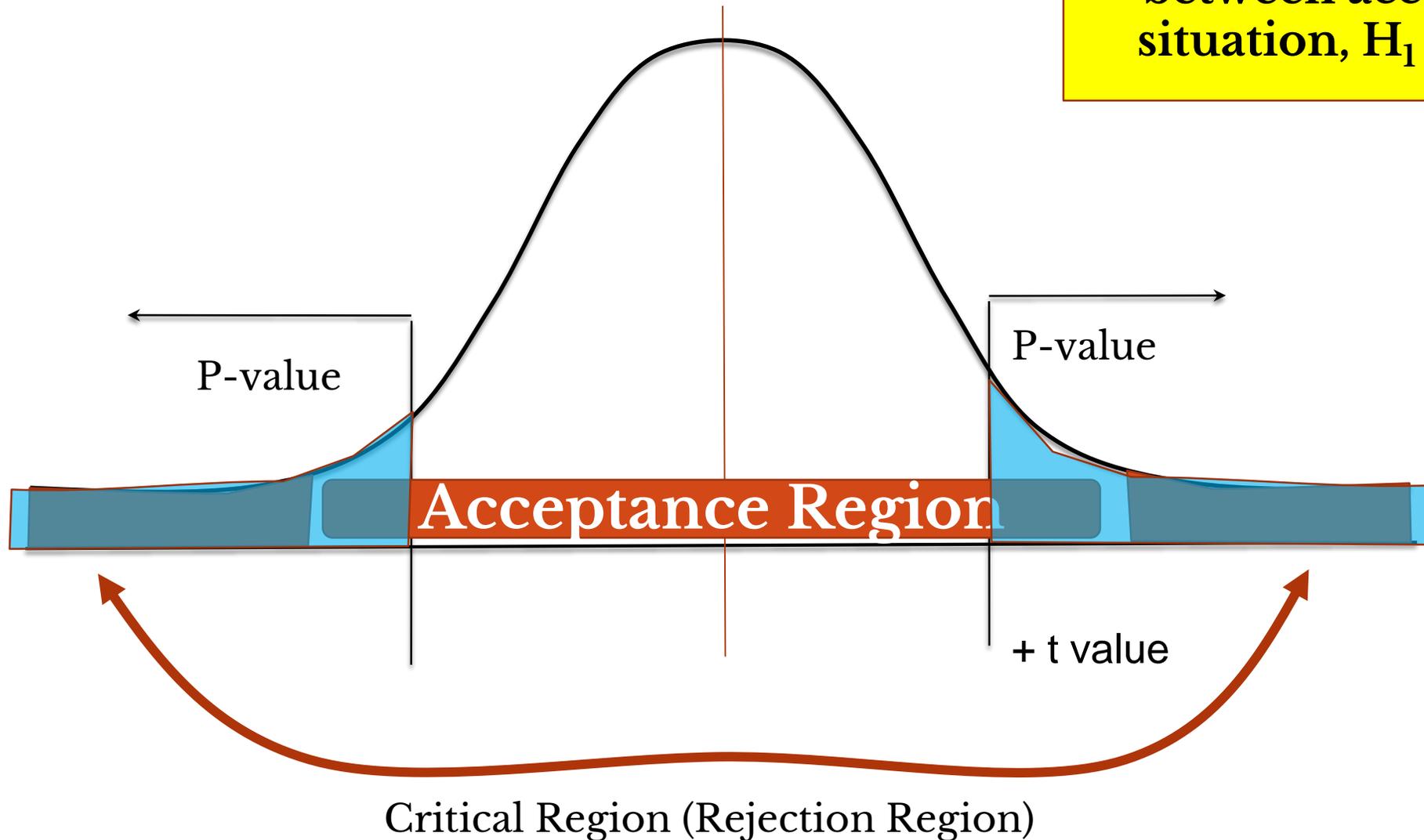For "large enough" *n*: $\bar{X} \sim N(\mu, \sigma^2/n)$

Decision Rule

Reject Ho

Reject Ho

Acceptance Region

Critical Region (Rejection Region)

Decision Rule

P-value

P-value

Acceptance Region

+ t value

Critical Region (Rejection Region)

One-Tail Tests…

One-Tail Test (left tail)	Two-Tail Test	One-Tail Test (right tail)
$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$

P-value

Acceptance Region

Rejection Region
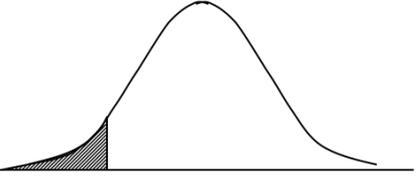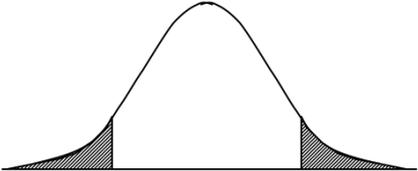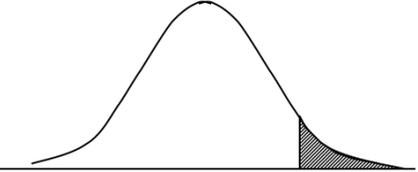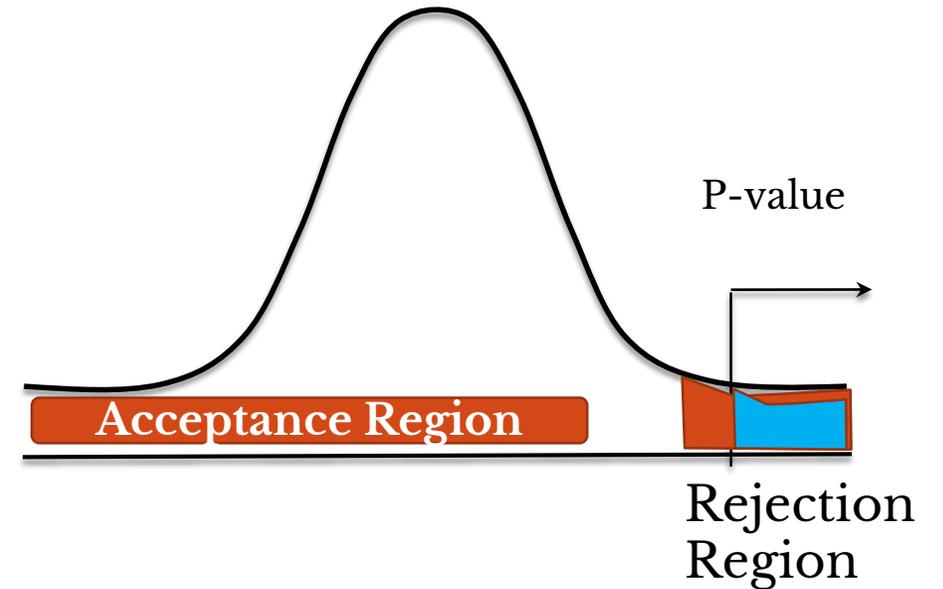
t-test, two means, with equal variance

Test Statistics

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S*\sqrt{(1/n_1 + 1/n_2)}} \text{ with df} = n_1 + n_2 - 2 \text{ where } S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

Score under two different condition with same person
a=c(12,16,8,6,4,8)
b=c(7,12,17,5,12,11)
R command
t.test(a,b,var.equal=TRUE)

Things to be consider
shapiro.test(a)
shapiro.test(b)
var.test(a,b)

If the normality does not follows then use Wilcoxon test
wilcox.test(a,b, paired=FALSE)

t-test, two means, with unequal variance

Test Statistics

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(S_1^2/n_1 + S_2^2/n_2)}} \quad \text{with} \quad df = \frac{(s_1^2/N_1 + s_2^2/N_2)^2}{(s_1^2/N_1)^2/(N_1 - 1) + (s_2^2/N_2)^2/(N_2 - 1)}$$

```
aa=c(10,18,15,38,25,27,45)
ab=c(22,26,32,24,27,29,35)
shapiro.test(aa)
shapiro.test(ab)
var.test(aa,ab)
t.test(a,b, var.equal = FALSE)
```

T test in our data

#Bring CS_Bank data in your environment
#attach your data so that it will be your active dataframe
attach(data) and see the mean and sd
aggregate(income~sex,data,function(s) c(mean=mean(s),sd=sd(s)))
Test the normality of income in the two different sex
mi=income[sex==1]   shapiro.test(mi)
mf=income[sex==2]   shapiro.test(fi)
If the data is normal then check equality of variance
var.test(income~sex)
If the data is normal and variance is equal use
t.test(income~sex,equal.var=TRUE)

T test in our data

#Bring CS_Bank data in your environment
#attach your data so that it will be your active dataframe
attach(data) and see the mean and sd
aggregate(income~sex,data,function(s) c(mean=mean(s),sd=sd(s)))
Test the normality of income in the two different sex
mi=income[sex==1]   shapiro.test(mi)
mf=income[sex==2]   shapiro.test(fi)
If the data is normal then check equality of variance
var.test(income~sex)
If the data is normal and variance is equal use
t.test(income~sex,equal.var=TRUE)
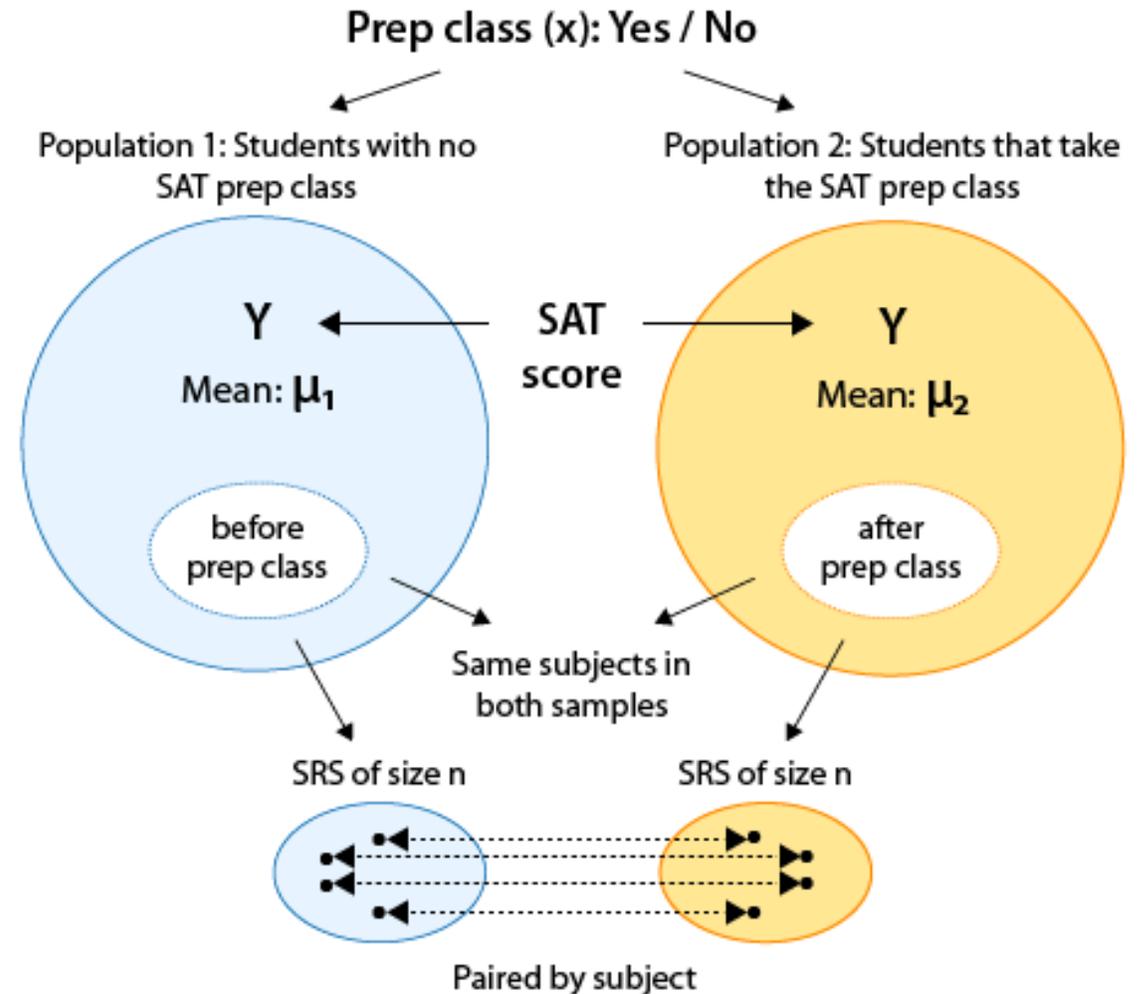
T test in our data

If the data is not normal then use
t.test(income~sex,equal.var=TRUE)

Significance difference of two means (dependent observations) Pair data

The idea behind the paired t-test is to **reduce** this **two-sample situation**, where we are comparing two means, **to a single sample s**a s**ituation** where we are doing inference on ingle mean, and **then use a simple t-test**.

we can easily reduce the raw data to a set of **differences** and conduct a **one-sample t-test**.

Prep class (x): Yes / No

Population 1: Students with no SAT prep class

Population 2: Students that take the SAT prep class

Y
Mean: μ_1

SAT score

Y
Mean: μ_2

before prep class

after prep class

Same subjects in both samples

SRS of size n

SRS of size n

Paired by subject

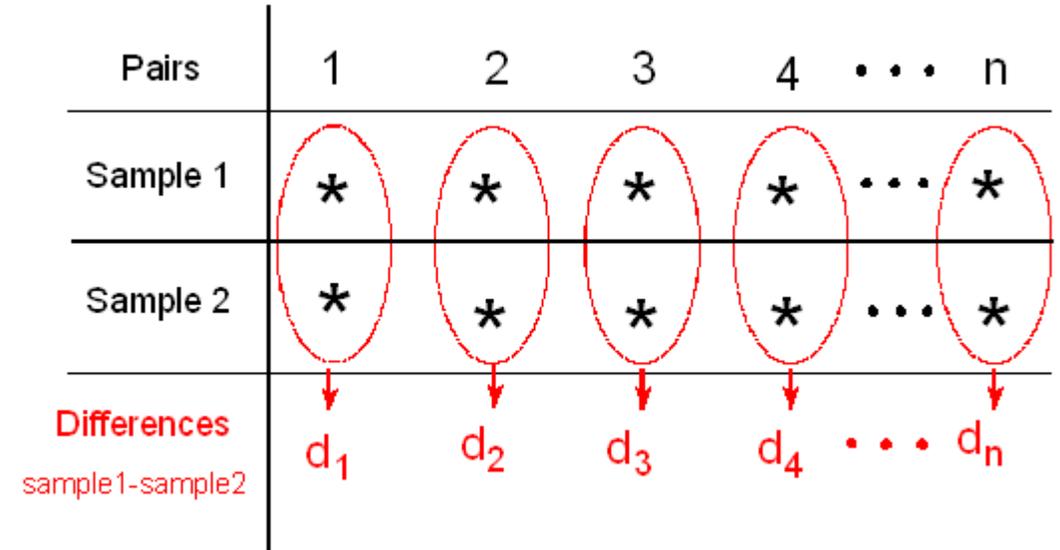| Pairs	1	2	3	4	· · ·	n
Sample 1	*	*	*	*	· · · ·	*
Sample 2	*	*	*	*	· · · ·	*

Pair - t -test

Ho: $\mu_d = 0$

There is no significance difference between sample

(The difference is not significantly differ than zero)

Ha: $\mu_d \neq 0$ or $\mu_d > 0$ or $\mu_d < 0$

The score under two different situation with same person
ba=c(12,16,8,6,4,8)
bb=c(7,12,17,5,12,11)
diff1=ba-bb
shapiro.test(diff1)
t.test(ba,bb,paired = TRUE)

Practice

In a manufacturing company the new modern manager is in a belief that music enhances the productivity of the workers. He made observation on eight workers for a week and recorded the production before and after music was installed. From the data given below can one conclude that productivity has been changed due to music?

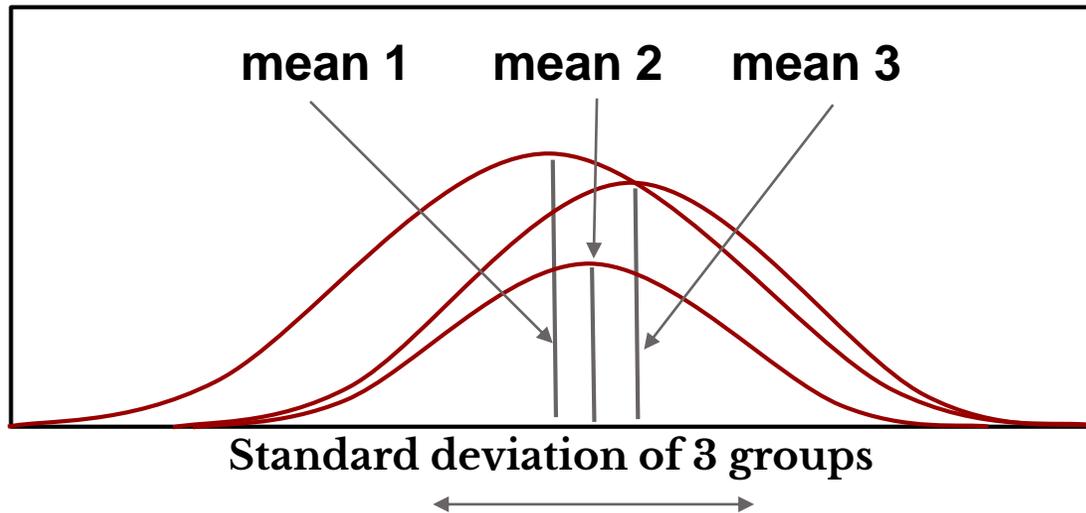| Employee	1	2	3	4	5	6	7	8
Without music	220	202	226	190	200	215	208	210
With music	236	190	240	200	220	205	212	215

wom=c(220,202,226,190,200,215,208,210)
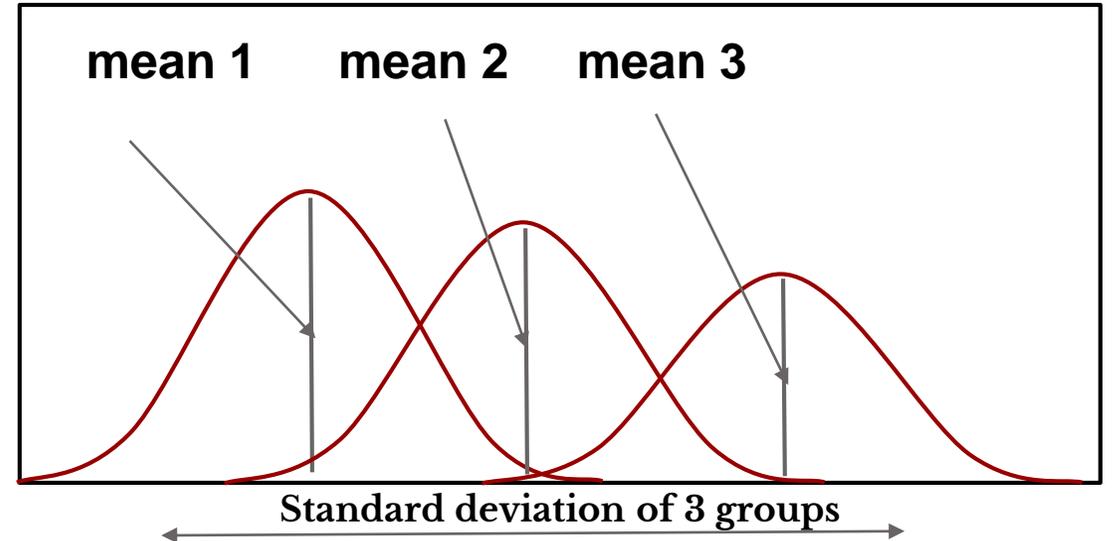wm=c(236,190,240,200,220,205,212,215)

Ans t = -1.4183, df = 7, p-value = 0.1991

How would we know or decide whether there is a real difference or not?

3 samples from Same population

mean 1   mean 2   mean 3

Standard deviation of 3 groups

3 samples from different population

mean 1   mean 2   mean 3

Standard deviation of 3 groups

To decide, we can compare our observed variance in means to what we would expect to get on the basis of chance given no true difference in means.

ANOVA

- Analysis of variance is used to test the claim that three or more population means are equal

- This is an extension of the two independent samples t-test

- The response variable is the variable you're comparing

- The factor variable is the categorical variable being used to define the groups

One way ANOVA

Conditions or Assumptions

- **The data are randomly sampled**

- **The variances of each sample are assumed equal**

- **Pop from where data drawn are normally distributed**

One way ANOVA

- The null hypothesis is that the means are not significantly different.

- The alternative hypothesis is that at least one of the means is different

  Think about the average productivity of three types of human resources, but one of these things is not like the other.

  They don't all have to be different, just one of them.

- The ANOVA doesn't test that one mean is less than another, only whether they're all equal or at least one is different.

Example

- The classroom is divided into three rows: front, middle, and back
- The instructor noticed that the students were away from him, the more likely they were not listening the class resulting less score.
- He wanted to see if the students away from did worse on the exams

Example

A random sample of the students in each row was taken. The score for those students on the second exam was recorded

- Front: 82, 83, 97, 93, 55, 67, 53

- Middle: 83, 78, 68, 61, 77, 54, 69, 51, 63

- Back: 38, 59, 55, 66, 45, 52, 52, 61

Data
f=c(82, 83, 97, 93, 55, 67, 53)
m=c(83, 78, 68, 61, 77, 54, 69, 51, 63)
b=c(38, 59, 55, 66, 45, 52, 52, 61)

R code

```
f=c(82, 83, 97, 93, 55, 67, 53)
m=c(83, 78, 68, 61, 77, 54, 69, 51, 63)
b=c(38, 59, 55, 66, 45, 52, 52, 61)
```
Combine all three vectors
```
dt=c(f,m,b)
```
Get the size of each vectors
```
n=c(length(f),length(m),length(b))
```
Make categories for combined vectors
```
gr=rep(c("a","b","c"),n)
```
Make dataframe
```
data6=data.frame(dt,gr)
anova1=aov(data6$dt~data6$gr)   #Use command aov
anova(anova1) #see ANOVA table for decision
```

Result

Source	SS	df	MS	F	P-value
Between	1901.5	2	950.76	5.896	0.009
Within	3386.3	21	161.2		
Total	5287.8	23			

ANOVA in our data

attach(data) ## attach data before your work

Check mean sd and number of obs for each category of edu

aggregate(income~edu,data,function(s) c(mean=mean(s),sd=sd(s),n=length(s)))

Test the homogeneity of variance

bartlett.test(income~edu)

Test the normality of the data

shapiro.test(income[edu==1])

shapiro.test(income[edu==2]) and so on

Use aov function to operate anova

anova1=aov(income~edu)

anova(anova1)

Note:
Before aov check whether
str of edu is of factor or not
If not then convert it into
factor
edu=as.factor(edu)

Measurement Metric / Non-metric

Sometimes we do not want to know the mean score of some population but rather the percentage of some population that does something.

A Department of traffic police, for example, might want to know the proportion of small vehicles that pass through Nagdhunga.

Test for a Population Proportion

Test for a Population Proportion *P*

prop.test(x=225, n=800, p=0.5, alternative="less")

Test of two population proportion P1 and P2

prop.test(x=c(225,450), n=c(500,1200))

More then two proportion

prop.test(x=c(225,450,345), n=c(1500,1500,1500), p=c(0.2,0.4,0.4))

Test of attribute (chi square, fisher)

chisq.test(sex,marital)

fisher.test(sex,marital)

Test using frequency table

```
        [,1] [,2]
[1,]   14   16
[2,]   25   21
```

matrix1=matrix(c(14,25,16,21),2,2)

chisq.test(matrix1)

fisher.test(matrix1)