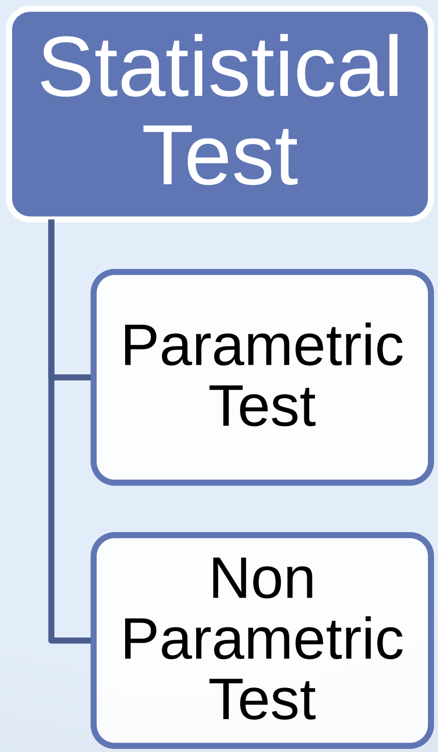
Parametric Tests

Dr Heena Sharma

Contents

Statistical Test

- These are intended to decide whether a hypothesis about distribution of one or more populations should be rejected or accepted.
- These may be:



These tests the statistical significance of the:-

1) Difference in sample and population means.
2) Difference in two sample means
3) Several population means
4) Difference in proportions between sample and population
5) Difference in proportions between two independent populations
6) Significance of association between two variables

System for statistical Analysis

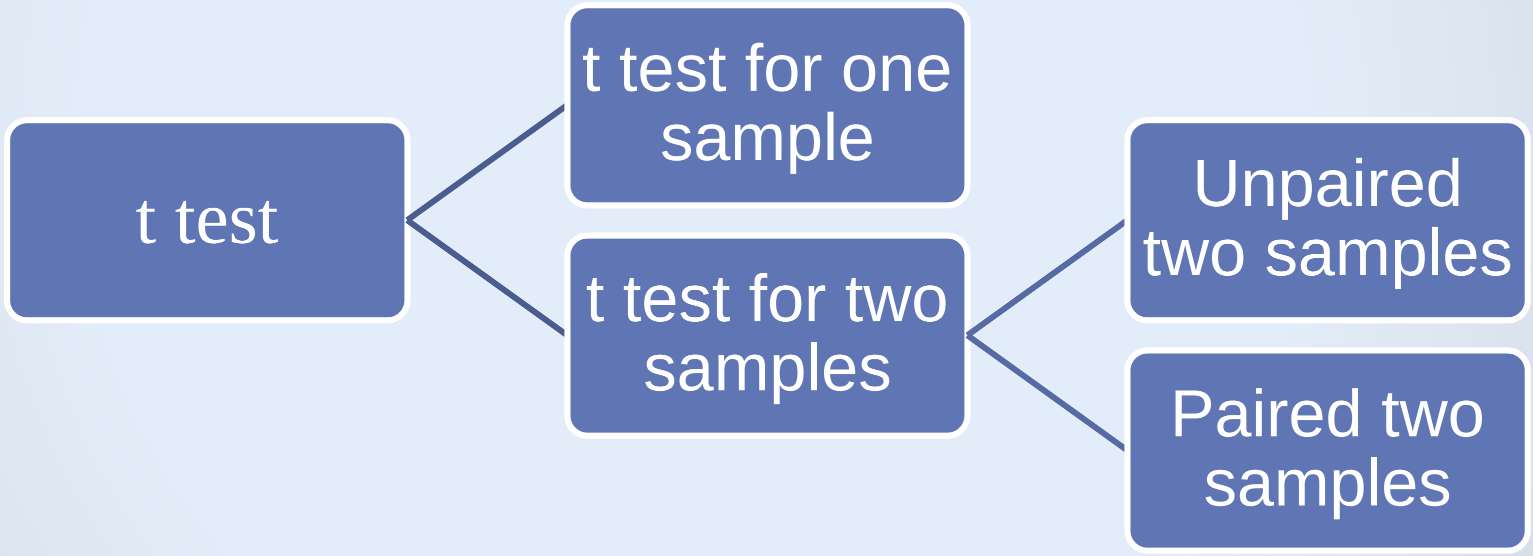
- State the Research Hypothesis
- State the Level of Significance
- Calculate the test statistic
- Compare the calculated test statistic with the tabulated values
- Decision
- Statement of Result

Parametric Tests

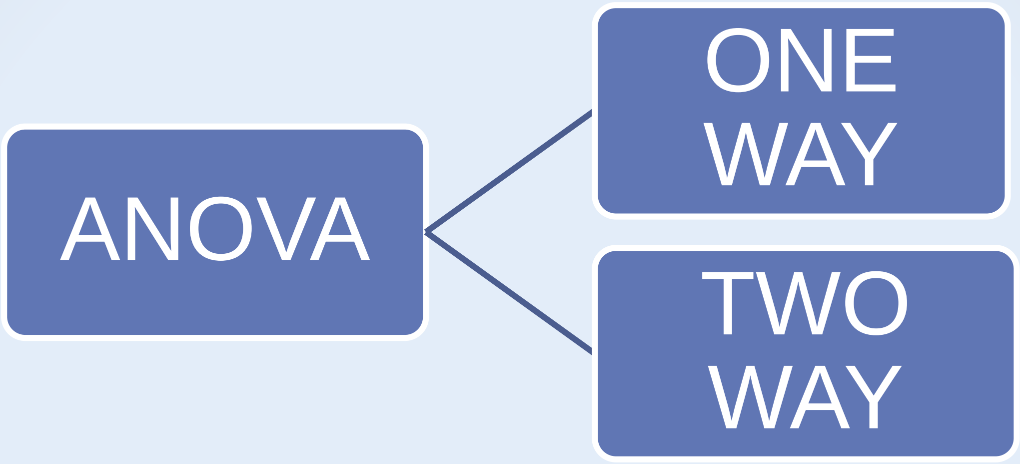
- Used for Quantitative Data
- Used for continuous variables
- Used when data are measured on approximate interval or ratio scales of measurement.
- Data should follow normal distribution

Parametric Tests

1. t test (n<30)

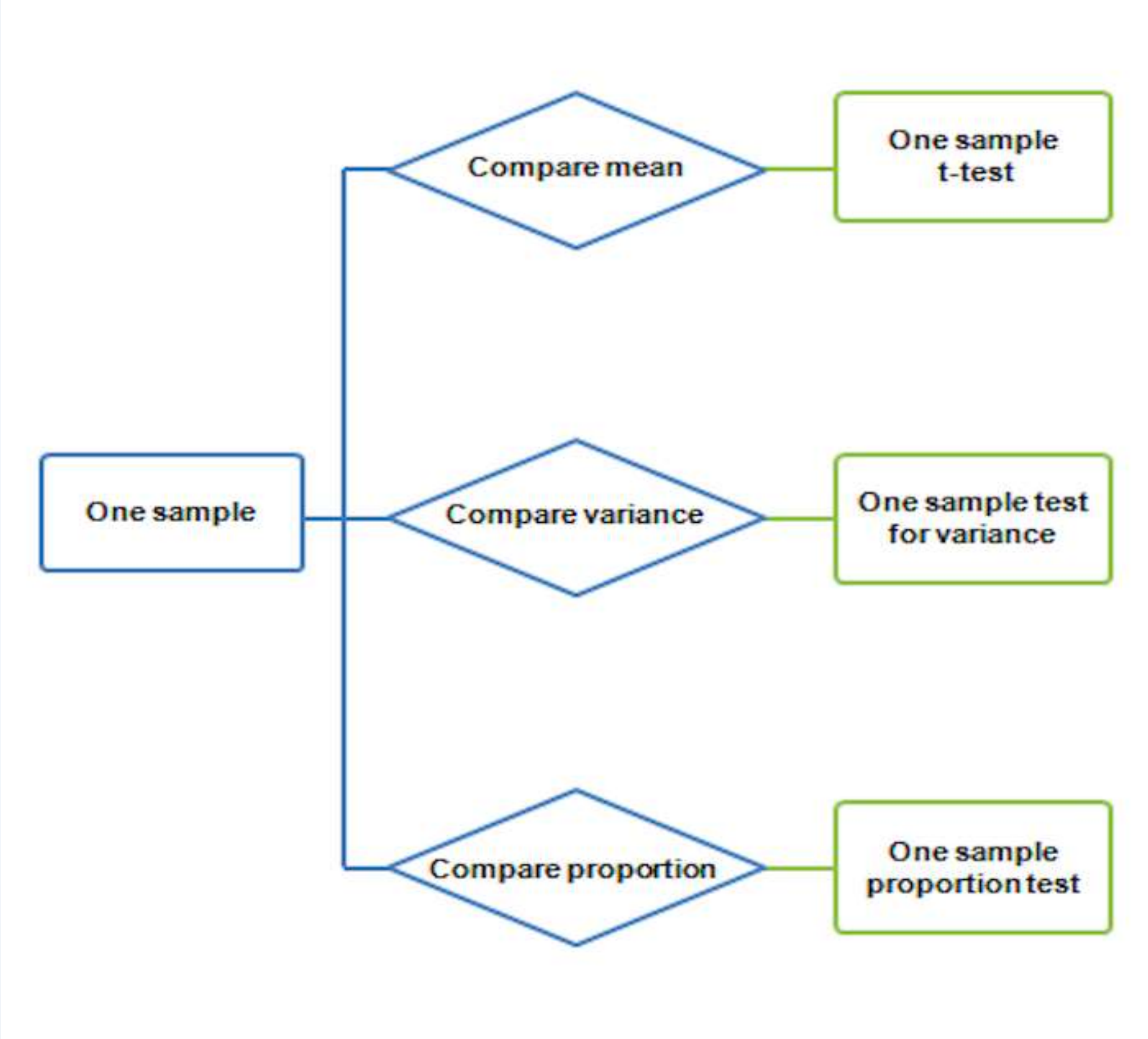


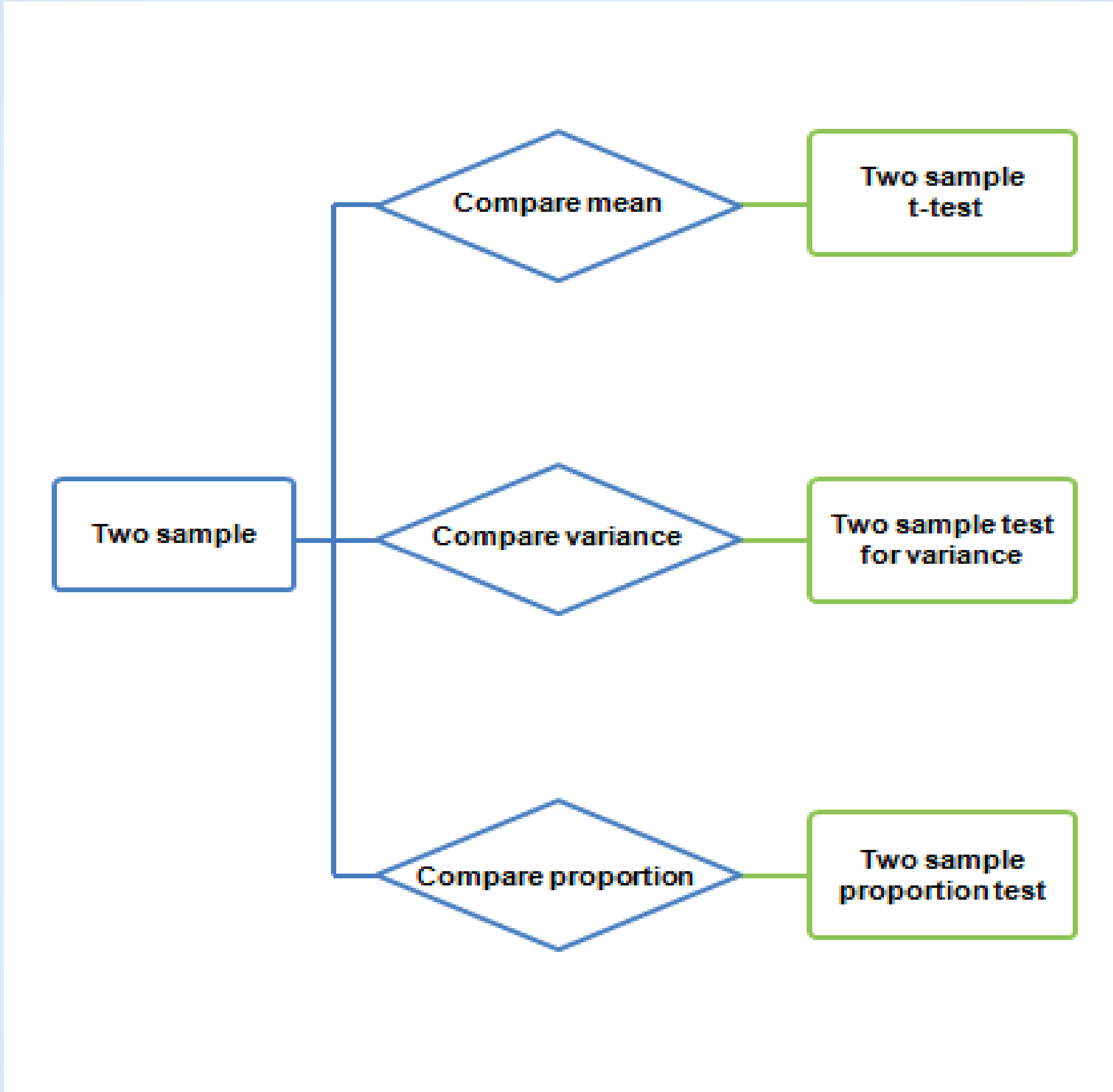
2. ANOVA (Analysis of Variance)



3. Pearson’s r Correlation

4. Z test for large samples (n>30)


One sample
Compare mean
Compare variance
Compare proportion
One sample t-test
One sample test for variance
One sample proportion test


Two sample
Compare mean
Compare variance
Compare proportion
Two sample t-test
Two sample test for variance
Two sample proportion test

Parametric tests

STUDENT'S T-TEST

- Developed by Prof W.S Gossett in 1908, who published statistical papers under the pen name of 'Student'. Thus the test is known as Student's 't' test.

- Indications for the test:-

1. When samples are small
2. Population variance are not known.

Uses

1. Two means of small independent samples

2. Sample mean and population mean

3. Two proportions of small independent samples

Assumptions made in the use of ‘t’ test

1. Samples are randomly selected
2. Data utilised is Quantitative
3. Variable follow normal distribution
4. Sample variances are mostly same in both the groups under the study
5. Samples are small, mostly lower than 30

- A t-test compares the difference between two means of different groups to determine whether that difference is statistically significant.

- Student’s ‘t’ test for different purposes
  - ✓ ‘t’ test for one sample
  - ✓ ‘t’ test for unpaired two samples
  - ✓ ‘t’ test for paired two samples

ONE SAMPLE T-TEST

- When compare the mean of a single group of observations with a specified value

- In one sample t-test, we know the population mean. We draw a random sample from the population and then compare the sample mean with the population mean and make a statistical decision as to whether or not the sample mean is different from the population.

Calculation

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

where $\bar{x}$ = sample mean

μ = Population mean

$S/\sqrt{n}$ = standard error

$$s^2 = \frac{\text{Sum } (x - \bar{x})^2}{n-1}$$

where x = element of sample

$\bar{x}$ = sample mean

$n-1$ = degrees of freedom

- Now we compare calculated value with table value at certain level of significance (generally 5% or 1%)

- If absolute value of ‘t’ obtained is greater than table value then reject the null hypothesis and if it is less than table value, the null hypothesis may be accepted.

EXAMPLE

Research Problem : Comparison of mean dietary intake of a particular group of individuals with the recommended daily intake.

DATA: Average daily energy intake (ADEI) over 10 days of 11 healthy women

sub	1	2	3	4	5	6	7	8	9	10	11
ADEI (KJ)	5260	5470	5640	6180	6390	6515	6805	7515	7515	8230	8770

Mean ADEI value = 6753.6

SD ADEI value = 1142.1

When can we say about the energy intake of these women in relation to a recommended daily intake of 7725 KJ ?

Research Hypothesis

State null hypothesis and alternative hypothesis:

H_0 = there is no difference between population mean and sample mean

OR

$H_0 : \mu = 7725$ KJ

H_1 = there is a difference between population mean and sample mean

OR

$H_1 : \mu \neq 7725$ KJ

- Set the level of significance

  $\alpha = .05, .01 \text{ or } .001$

- Calculate the value of proper statistic
- $$t = \frac{\text{sample mean} - \text{hypothesized mean}}{\text{standard error of sample mean}}$$

- $$\frac{6753.6 - 7725}{1142.1/\sqrt{11}} = -0.2564$$

State the rule for rejecting the null hypothesis:

Reject H_0 if $t \geq$ +ve Tabulated value

OR

Reject H_0 if $t \leq$ - ve Tabulated value

Or we can say that $p<.05$

In the above example we have seen

t=- .2564 which is less then 2.23

P value suggests that the dietary intake of these women was significantly less than the recommended level (7725 KJ)

Two Sample 't' test

A. Unpaired Two sample 't'- test

- Unpaired t- test is used when we wish to compare two means

- Used when the two independent random samples come from the normal populations having unknown or same variance

- We test the null hypothesis, that the two population means are same i.e $\mu_1 = \mu_2$ against an appropriate one sided or two sided alternative hypothesis

Assumptions

- The samples are random & independent of each other
- The distribution of dependent variable is normal.
- The variances are equal in both the groups

FORMULA

Test statistic is given by

$$t = \frac{\text{Mean1} - \text{Mean2}}{\text{SE (Mean1} - \text{mean2)}}$$

$$\text{SE(Mean1} - \text{mean2)} = S\sqrt{[1/n_1+1/n_2]}$$

$$S = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1+n_2-2)}$$

Where S_1^2 and S_2^2 are respectively called SD's of first and second group

Research Problem

A study was conducted to compare the birth weights of children born to 15 non-smoking with those of children born to 14 heavy smoking mothers.

Non-smoking mothers (n=15)	Heavy smoking mothers (n=14)
3.99	3.18
3.79	2.84
3.60	2.90
3.73	3.27
3.21	3.85
3.60	3.52
4.08	3.23
3.61	2.76
3.83	3.60
3.31	3.75
4.13	3.59
3.26	3.63
3.54	2.38
3.51	2.34
2.71	

- Research Hypothesis : State null hypothesis and alternative hypothesis
- H_0 = there is no difference between the birth weights of children born to non-smoking and smoking mothers
- H_1 = there is a difference between the birth weights of children born to non-smoking and smoking mothers
- Set the level of significance

α = .05, .01 or .001

- Calculate the value of proper statistic
- State the rule for rejecting the null hypothesis
- If $t_{cal} > t_{tab}$ we can say that $P < .05$ then we reject the null hypothesis and accept the Alternative hypothesis.

Decision

- If we reject the null hypothesis so we can say that children born to non-smokers are heavier than children born to heavy smokers.

PAIRED TWO-SAMPLES T-TEST

- Used when we have paired data of observations from one sample only, when each individual gives a pair of observations.

- Same individuals are studied more than once in different circumstances- measurements made on the same people before and after interventions

Assumptions

- The outcome variable should be continuous

- The difference between pre-post measurements should be normally distributed

FORMULA

$$t = \frac{\bar{d}}{SD/\sqrt{n}}$$

Where,

d = difference between x_1 and x_2

$\bar{d}$ = Average of d

SD = Std. deviation for the difference

n = sample size

Research Problem

A study was carried to evaluate the effect of the new diet on weight loss. The study population consist of 12 people have used the diet for 2 months; their weights before and after the diet are given

Patient no.	**Weight (Kgs)**	
	Before Diet	After Diet
1	75	70
2	60	54
3	68	58
4	98	93
5	83	78
6	89	84
7	65	60
8	78	77
9	95	90
10	80	76
11	100	94
12	108	100

Research Hypothesis

- State null hypothesis and alternative hypothesis
- H_0 = There is no reduction in weight after Diet
- H_1= There is reduction in weight after Diet
- Further Analysis through Statistical software SPSS as same as previous example

Decision

- If we reject the null hypothesis then there is a statistically significant reduction in weight

How do we compare more than two groups means ??

Example :

Treatments : A, B, C & D

Response : BP level

How does t-test concept work here ?

A versus B	B versus C
A versus C	B versus D
A versus D	C versus D

so the chance of getting the wrong result would be:

1 - (0.95 x 0.95 x 0.95 x0.95) = 26%

- Instead of using a series of individual comparisons we examine the differences among the groups through an analysis that considers the variation among all groups at once.

- i.e. ANALYSIS OF VARIANCE

Analysis of Variance(ANOVA)

- Given by Sir Ronald Fisher
- The principle aim of statistical models is to explain the variation in measurements.
- The statistical model involving a test of significance of the difference in mean values of the variable between two groups is the student's,'t' test. If there are more than two groups, the appropriate statistical model is Analysis of Variance (ANOVA)

Assumptions for ANOVA

1. Sample population can be easily approximated to normal distribution.

2. All populations have same Standard Deviation.

3. Individuals in population are selected randomly.

4. Independent samples

- ANOVA compares variance by means of a simple ratio, called F-Ratio

$$F = \frac{\text{Variance between groups}}{\text{Variance within groups}}$$

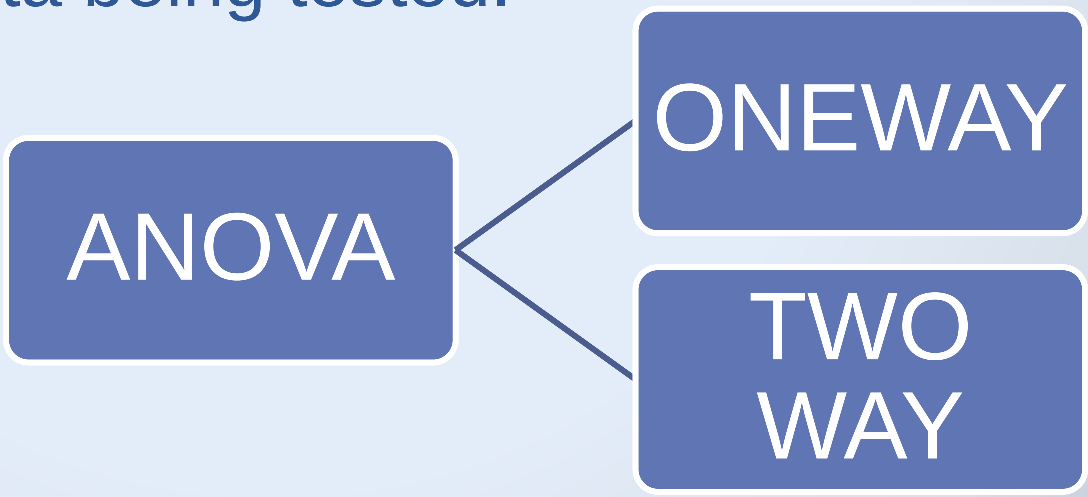
- The resulting F statistics is then compared with critical value of F (critic), obtained from F tables in much the same way as was done with ‘t’

- If the calculated value exceeds the critical value for the appropriate level of α, the null hypothesis will be rejected.

- A F test is therefore a test of the Ratio of Variances F Tests can also be used on their own, independently of the ANOVA technique, to test hypothesis about variances.

- In ANOVA, the F test is used to establish whether a statistically significant difference exists in the data being tested.

- ANOVA can be



- One Way ANOVA
  - If the various experimental groups differ in terms of only one factor at a time- a one way ANOVA is used

e.g. A study to assess the effectiveness of four different antibiotics on S Sanguis

- Two Way ANOVA
  - If the various groups differ in terms of two or more factors at a time, then a Two Way ANOVA is performed

e.g. A study to assess the effectiveness of four different antibiotics on S Sanguis in three different age groups

Pearson's Correlation Coefficient

- Correlation is a technique for investigating the relationship between two quantitative, continuous variables

- Pearson's Correlation Coefficient(r) is a measure of the strength of the association between the two variables.

Assumptions Made in Calculation of ‘r’

1. Subjects selected for study with pair of values of X & Y are chosen with random sampling procedure.

2. Both X & Y variables are continuous

3. Both variables X & Y are assumed to follow normal distribution

Steps

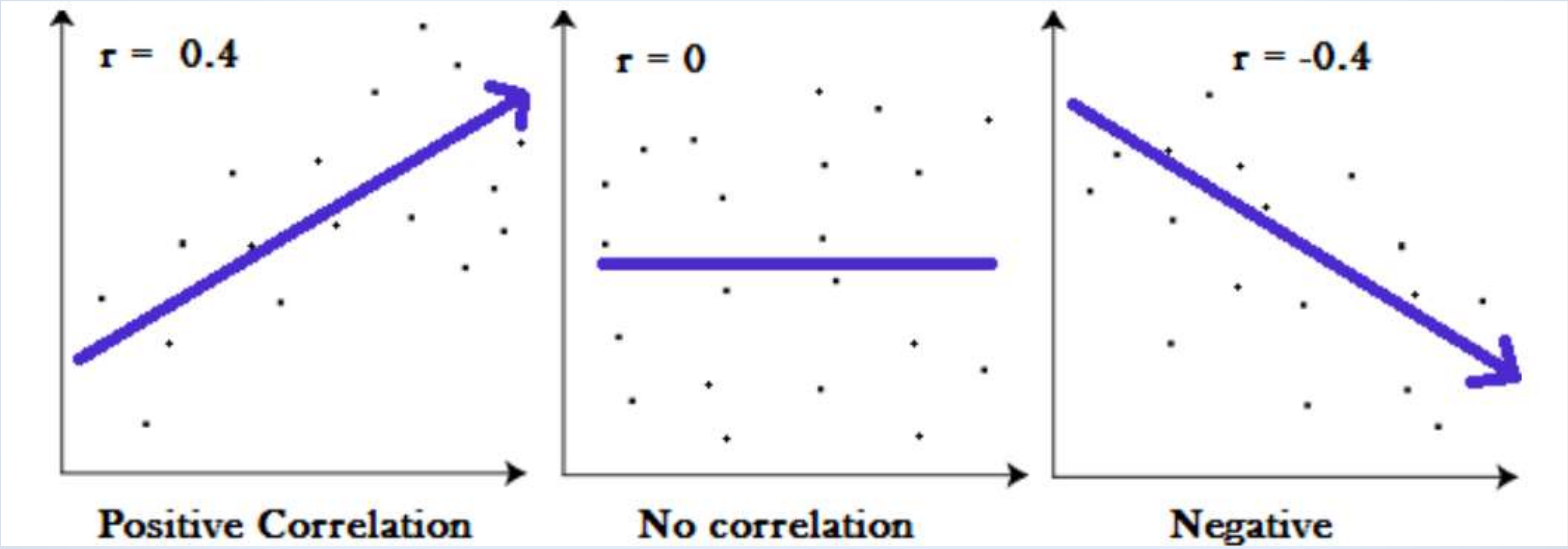
- The first step in studying the relationship between two continuous variables is to draw a scatter plot of the variables to check for linearity.
- The correlation coefficient should not be calculated of the relationship is not linear
- For correlation only purposes, it does not matter on which axis the variables are plotted

- However, conventionally, the independent variable is plotted on X axis and dependent variable on Y-axis

- The nearer the scatter of points is to a straight line, the higher the strength of association between the variables.

Types of Correlation

- Perfect Positive Correlation $r=+1$
- Partial Positive Correlation $0<r<+1$
- Perfect negative correlation $r=-1$
- Partial negative correlation $0>r>-1$
- No Correlation


r = 0.4
r = 0
r = -0.4
Positive Correlation
No correlation
Negative

Z Test

- This test is used for testing significance difference between two means (n>30).

- Assumptions to apply Z test
  - ✓The sample must be randomly selected
  - ✓Data must be quantitative
  - ✓Samples should be larger than 30
  - ✓Data should follow normal distribution
  - ✓Sample variances should be almost the same in both the groups of study

- If the SD of the populations is known, a Z test can be applied even if the sample is smaller than 30

Indications for Z Test

- To compare sample mean with population mean
- To compare two sample means
- To compare sample proportion with population proportion
- To compare two sample proportions

Steps

1. Define the problem

2. State the null hypothesis (H0) & alternate hypothesis (H1)

3. Find Z value

$$Z = \frac{\text{Observed mean-Mean}}{\text{Standard Error}}$$

4. Fix the level of significance

5. Compare calculated Z value with the value in Z table at corresponding degree significance level.

- If the observed Z value is greater than theoritical Z value, Z is significant, reject null hypothesis and accept alternate hypothesis

Z-PROPORTIONALITY TEST

- Used for testing the significant difference between two proportions,

$$z = \frac{P_1 - P_2}{SE(P_1 - P_2)}$$

Where, P_1 = Prop. rate for I^{st} population

P_2 = Prop. rate for II^{nd} population

Where, SE $(P_1 - P_2)$ is defined SE of difference

One tailed and Two tailed Z tests

- Z values on each side of mean are calculated as +Z or as -Z.

- A result larger than difference between sample mean will give +Z and result smaller than the difference between mean will give -Z

- E.g. for two tailed:
  - In a test of significance, when one wants to determine whether the mean IQ of malnourished children is different from that of well nourished and does not specify higher or lower, the P value of an experiment group includes both sides of extreme results at both ends of scale, and the test is called two tailed test.

- E.g. for single tailed:
  - In a test of significance when one wants to know specifically whether a result is larger or smaller than what occur by chance, the significant level or P value will apply to relative end only e.g. if we want to know if the malnourished have lesser mean IQ than the well nourished, the result will lie at one end (tail)of the distribution, and the test is called single tailed test

Conclusion

- ✓ Tests of significance play an important role in conveying the results of any research & thus the **choice of an appropriate statistical test is very important** as it decides the fate of outcome of the study.
- ✓ Hence the emphasis placed on tests of significance in clinical research must be tempered with an understanding that they are **tools for analyzing** data & should never be used as a substitute for knowledgeable interpretation of outcomes.

References

- Sundaram KR, Dewivedi SN, Sreenivas V. Medical statistics, Principles and methods;BI Publications New Delhi.
- Glaser AN. High Yeild Biostatistics 2nd Edition. Jaypee Brothers Medical Publisher Ltd.
- Dixit JV. Principles and practice of Biostatistics 3rd Edition Bhanot Publications.
- Rao KV. Biostatistics, A manual of statistical method for use in Health, Nutrition and Anthropology. Jaypee Publications
- Mahajan BK. Methods in biostatistics. 7 th edition. Jaypee publications


THANK YOU