

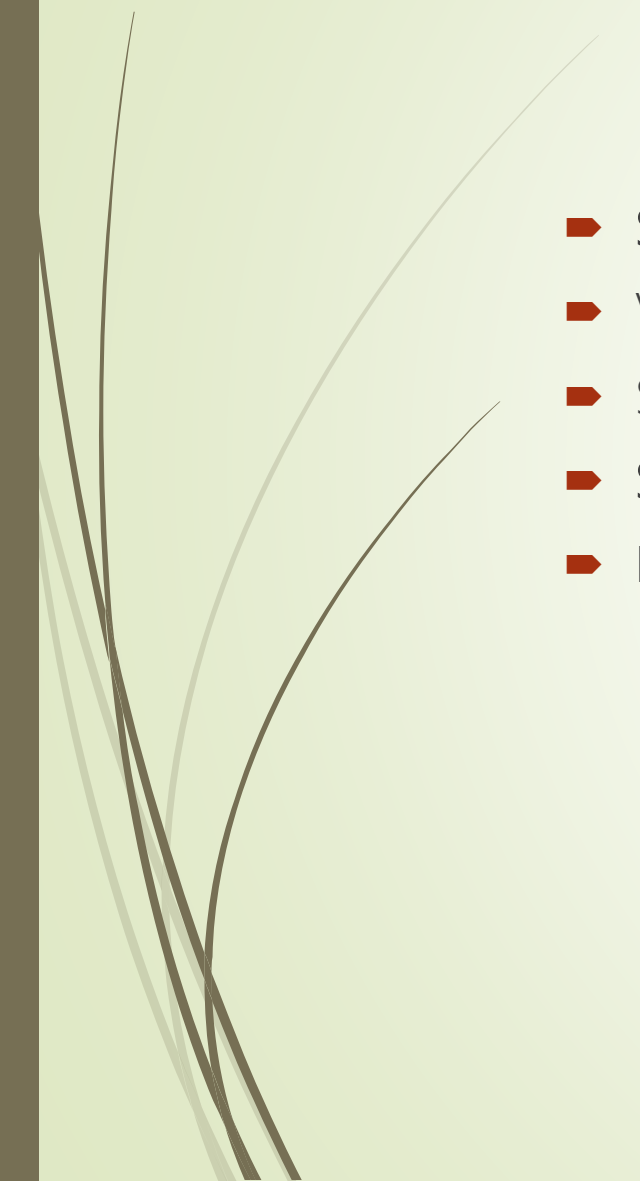


# DATA COLLECTION, TABULATION, PROCESSING AND ANALYSIS OF DATA

BY  
J.ROBINSON RAJA  
M-6006



# NEED FOR DATA COLLECTION

- Scientific research
  - Verify the hypothesis
  - Substantiate various arguments in research findings
  - Solution to problem
  - Helps in drawing generalization
- 

# OTHER METHODS OF DATA COLLECTION

- Warranty cards
- Distributor / store audits
- Pantry audits
- Consumer panel
- Mechanical devices
- Eye camera, Pupilometric camera
- Psychogalvanometer, Motion picture camera

**WARRANTY CARD**

Send this copy to: **Monkey Tools \* P.O. Box 123 \* Monkeyvale, FL 33123**

Product Model Number / Name: \_\_\_\_\_

Date of Purchase: \_\_\_\_\_ Where Purchased: \_\_\_\_\_

Print Name: \_\_\_\_\_

Address: \_\_\_\_\_

E-mail: \_\_\_\_\_

My signature below acknowledges that I have read, fully understand, and accept this limited warranty agreement.

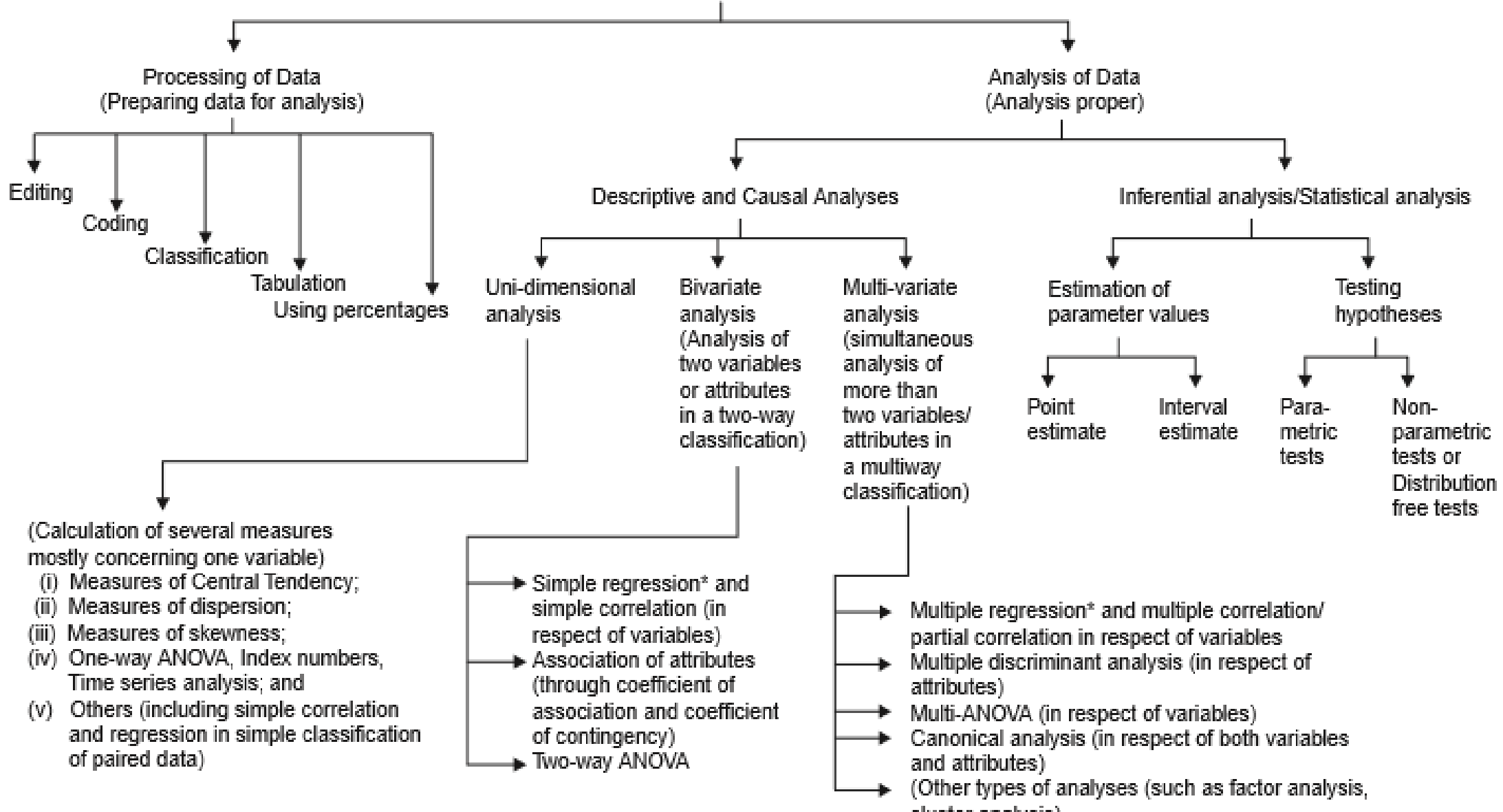
Signature: \_\_\_\_\_ Date \_\_\_\_\_

# Appendix

(Summary chart concerning analysis of data)

## Analysis of Data

(in a broad general way can be categorised into)





# NEED FOR PROCESSING AND ANALYSING

- Essential for scientific study
- To make the raw data meaningful,
- To test null hypothesis,
- To obtain the significant results,
- To draw some inferences or make generalization, and
- To estimate parameters



# EDITING

- ▶ Detect **errors and omissions** and correct them
- ▶ Assure that data are **accurate, consistent, complete, homogenous**
- ▶ **Field editing**-what the latter has written in **abbreviated and/or in illegible form** at the time of recording the respondents' responses.
- ▶ **Difficult** for others to **decipher**.
- ▶ As soon as possible after the interview.
- ▶ **Must not correct errors of omission**

# CENTRAL EDITING

- All forms -Completed and returned to the office.
- May correct the obvious errors (entry in the wrong place).
- On missing replies, sometimes determine the proper answer by **reviewing the other information** in the schedule.
- At times, the respondent can be contacted for clarification.
- The editor must strike out the answer if the same is inappropriate and he has no basis for determining the correct answer or the response. In such a case an editing entry of 'no answer' is called for.
- All the **wrong replies, which are quite obvious**, must be dropped from the final results, especially in the context **of mail surveys**.



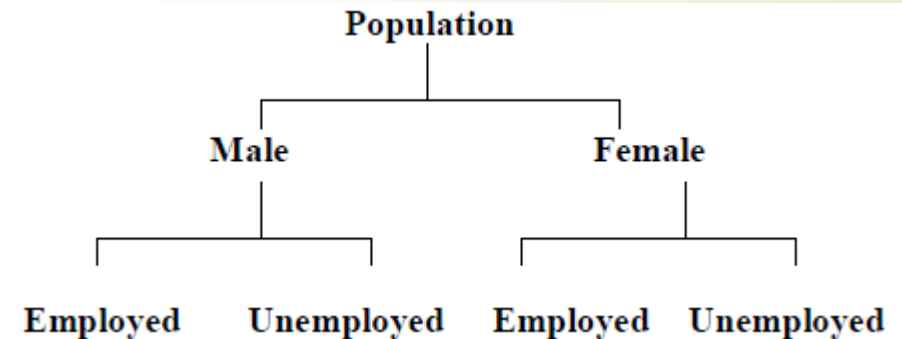
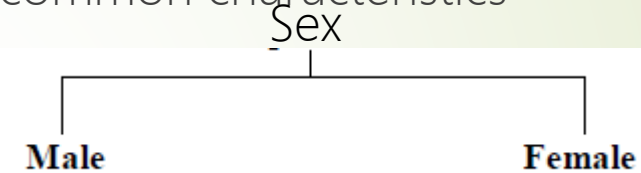
# CODING

- Assigning **numerals** or other symbols to answers so that responses can be put into a **limited number** of **categories** or classes.
- Such classes should be appropriate to the research problem under consideration.
- **Exhaustive , mutually exclusive**
- Uni-dimensionality
- Critical information required for analysis.
- **Coding decisions** -designing stage of the questionnaire.
- Easy tabulation



# CLASSIFICATION

- Arranging data in groups or classes – common characteristics
- Types of classification:
- Attributes- simple and manifold
- Class intervals- inclusive, exclusive
- 11-20, 21-30
- 10-20, 20-30
- How many classes ?





# TABULATION

- Orderly arrangement of data in columns and rows.
- conserves space.
- comparison.
- summation of items
- detection of errors and omissions.
- statistical computations

# Components Of Table

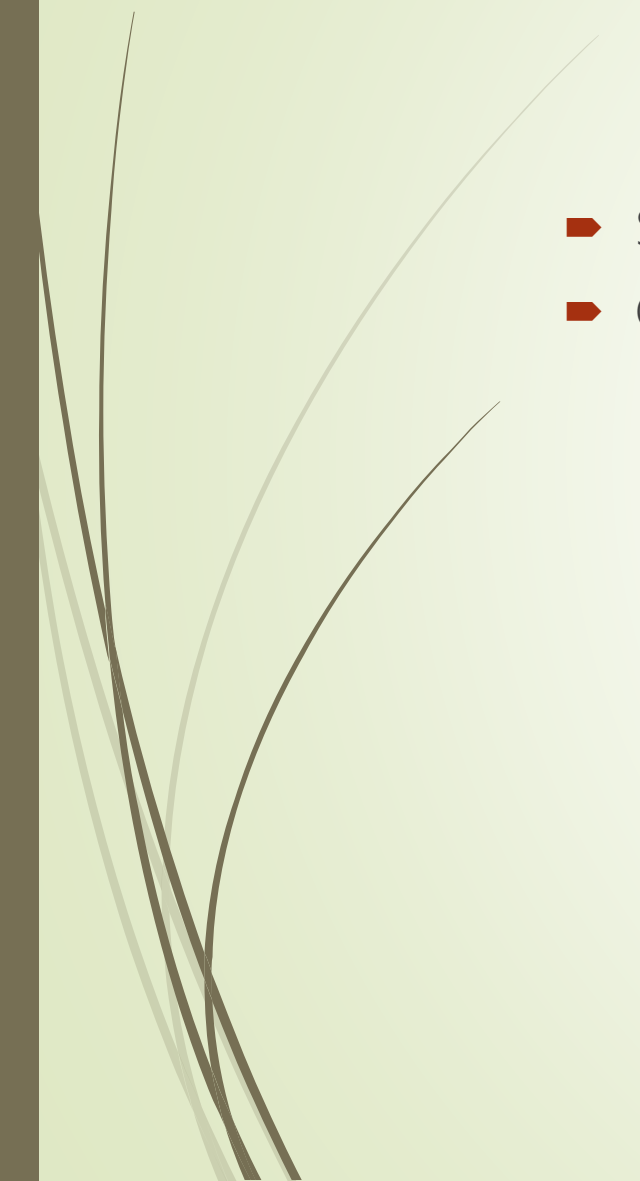
1. Table number
2. Title of the table
3. Caption / Box head
4. Stub
5. Body / Field
6. Head note
7. Foot note
8. Source data

Stub headings	Caption				Total (rows)
	Subhead		Subhead		
	Column- head	Column head	Column- head	Column head	
Stub Entries					
Total (columns)					

Foot note :  
Source note:



# TYPES OF TABLES

- ▶ Simple and complex
  - ▶ General and special purpose
- 

# PRINCIPLES OF TABULATION

- ▶ Clear, concise and adequate title.
- ▶ Distinct number to facilitate reference
- ▶ Column headings and row headings of table should be clear and brief.
- ▶ Units of measurement under each heading or subheading must always be indicated.
- ▶ Explanatory footnotes –beneath the table
- ▶ Source-just below the table.
- ▶ Abbreviation, ditto should be avoided



# PROBLEMS IN PROCESSING

- DK response
- Respondent may not know the answer or researcher fail in obtaining the appropriate information.
- DK response in separate category

# ANALYSIS OF DATA



- ▶ Analysis means **categorising, ordering, manipulating, and summarising** of data to obtain answer to research questions.
- ▶ Descriptive analysis
- ▶ Correlation analysis
- ▶ Causal analysis / regression analysis

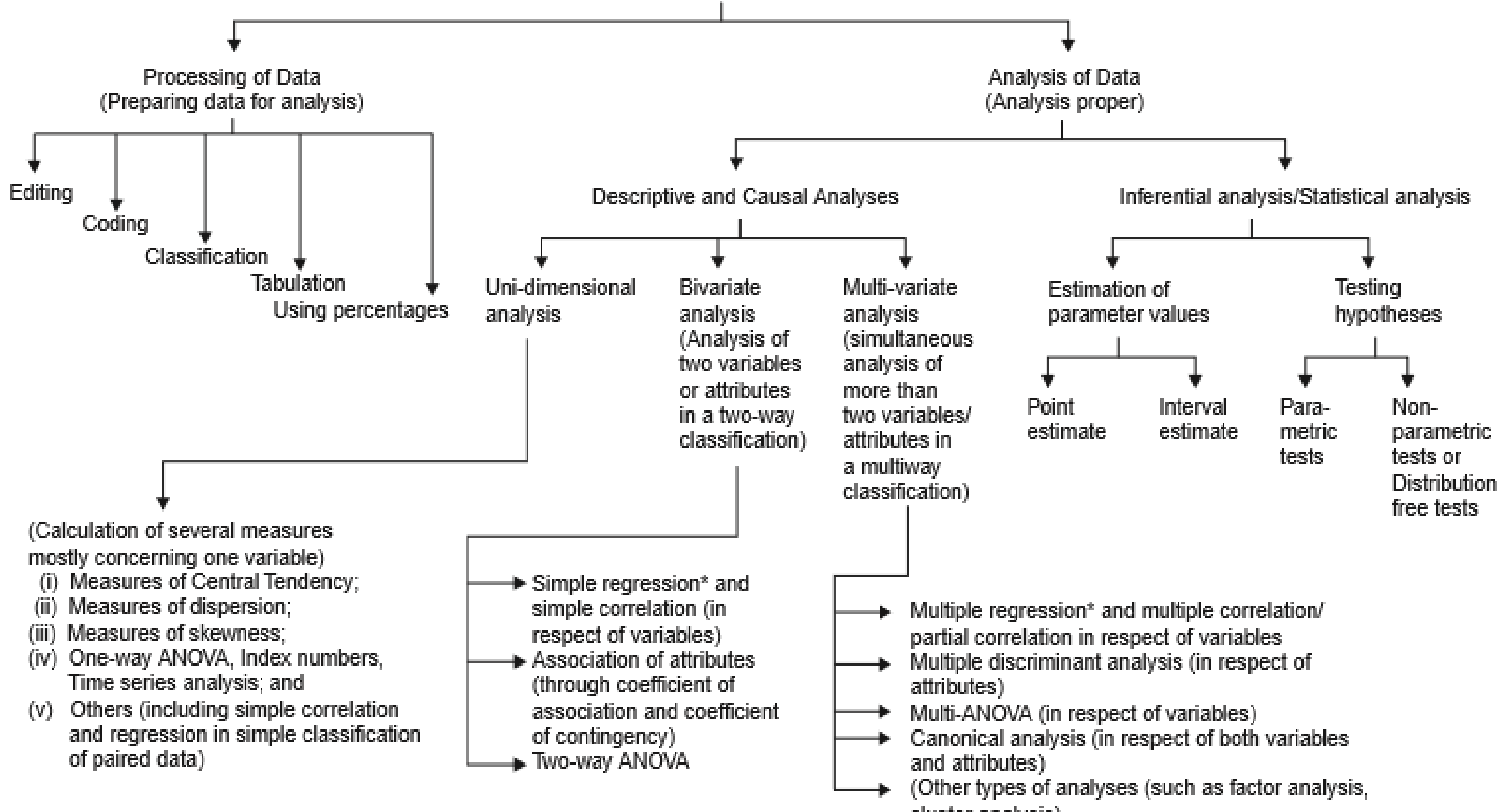


# Appendix

(Summary chart concerning analysis of data)

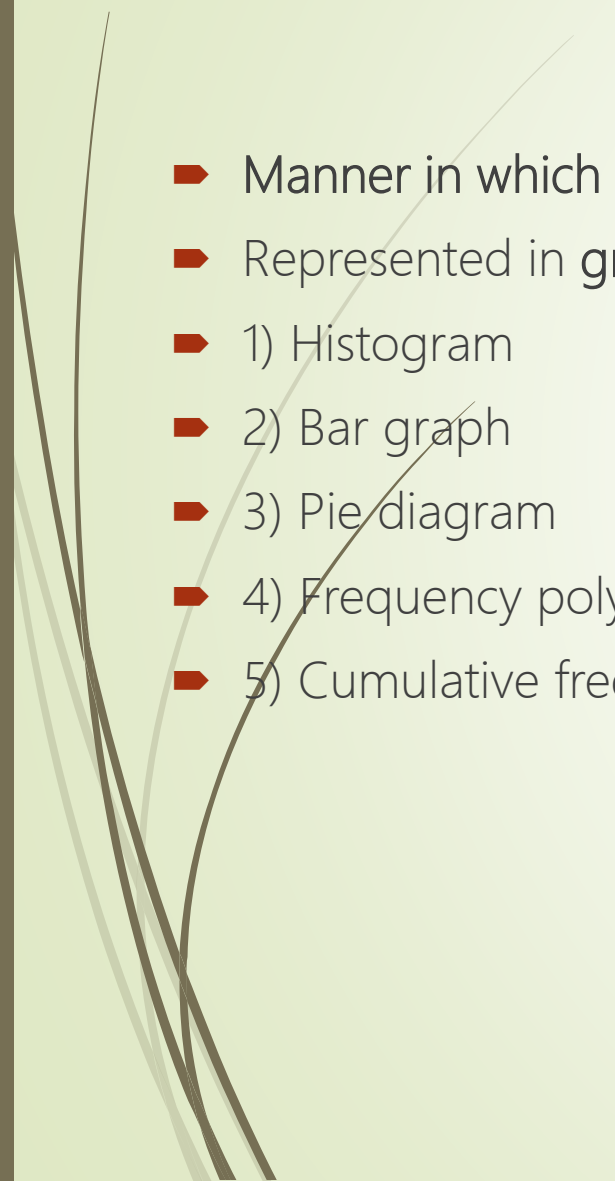
## Analysis of Data

(in a broad general way can be categorised into)

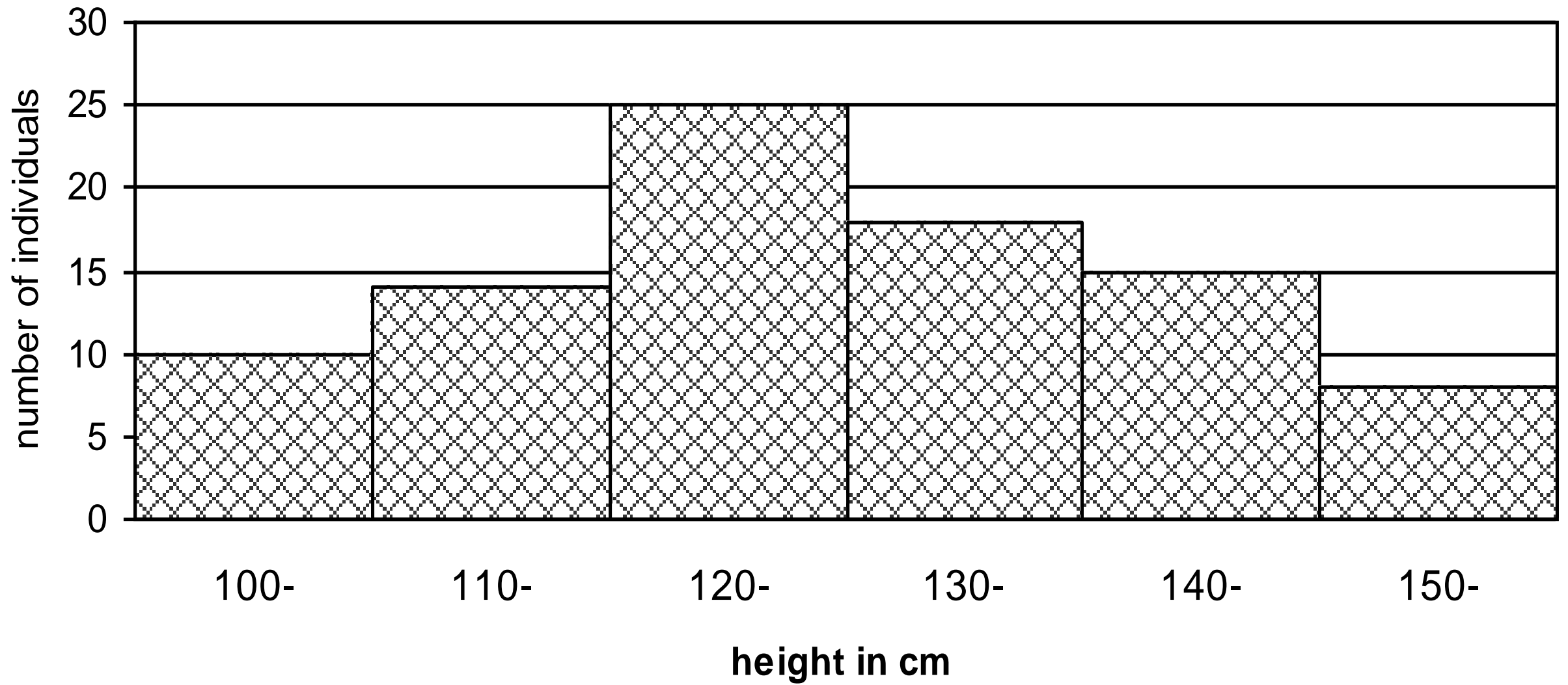




# FREQUENCY DISTRIBUTION

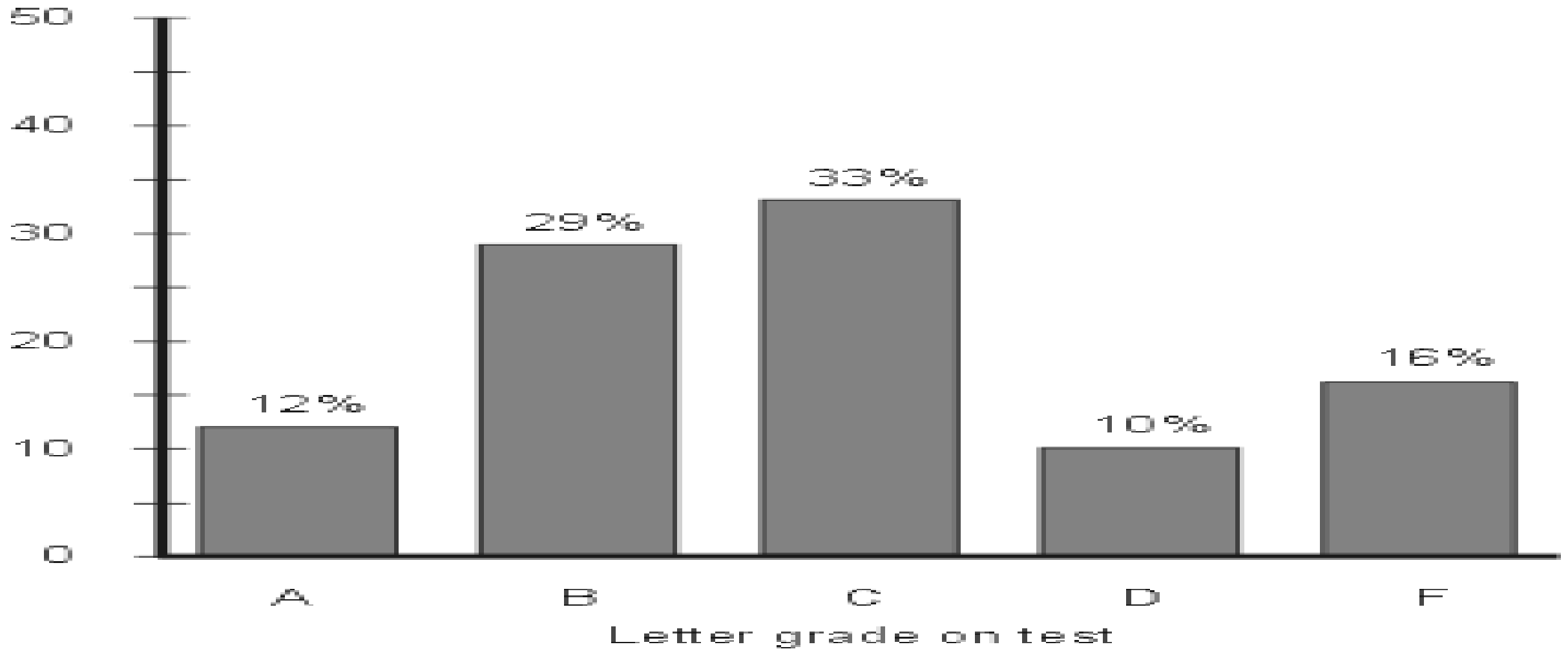
- Manner in which different frequencies are distributed over different classes.
  - Represented in graphical form.
  - 1) Histogram
  - 2) Bar graph
  - 3) Pie diagram
  - 4) Frequency polygon
  - 5) Cumulative frequency curve / ogive curve
- 

Distribution of studied group according to their height



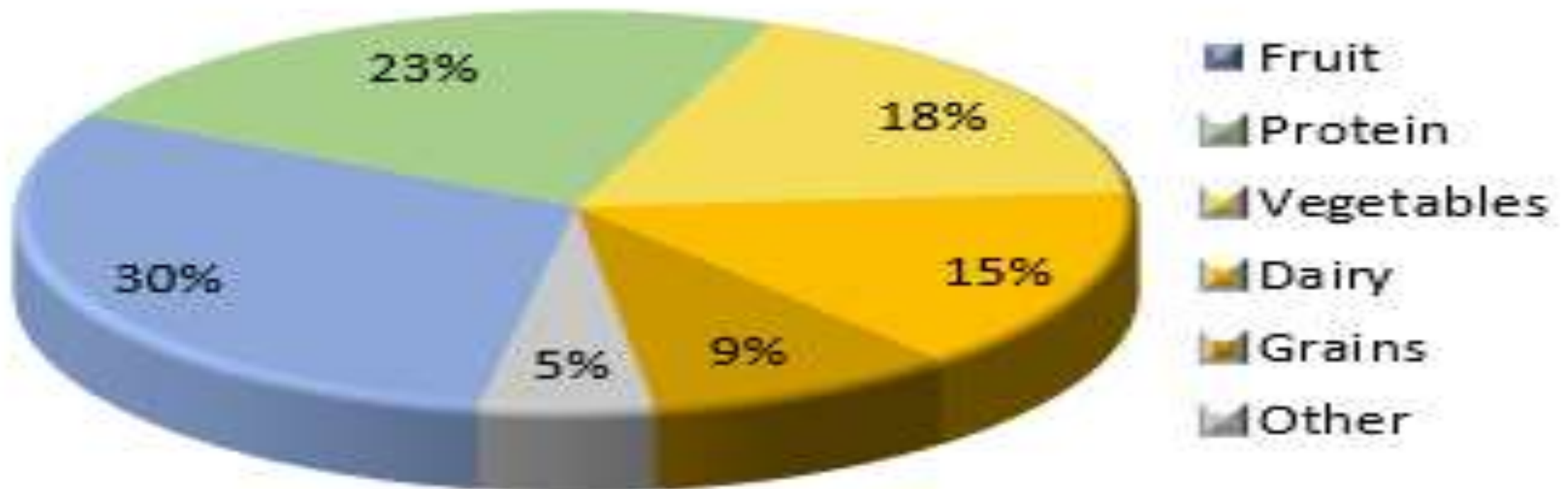
# BAR DIAGRAM

Percentage of students

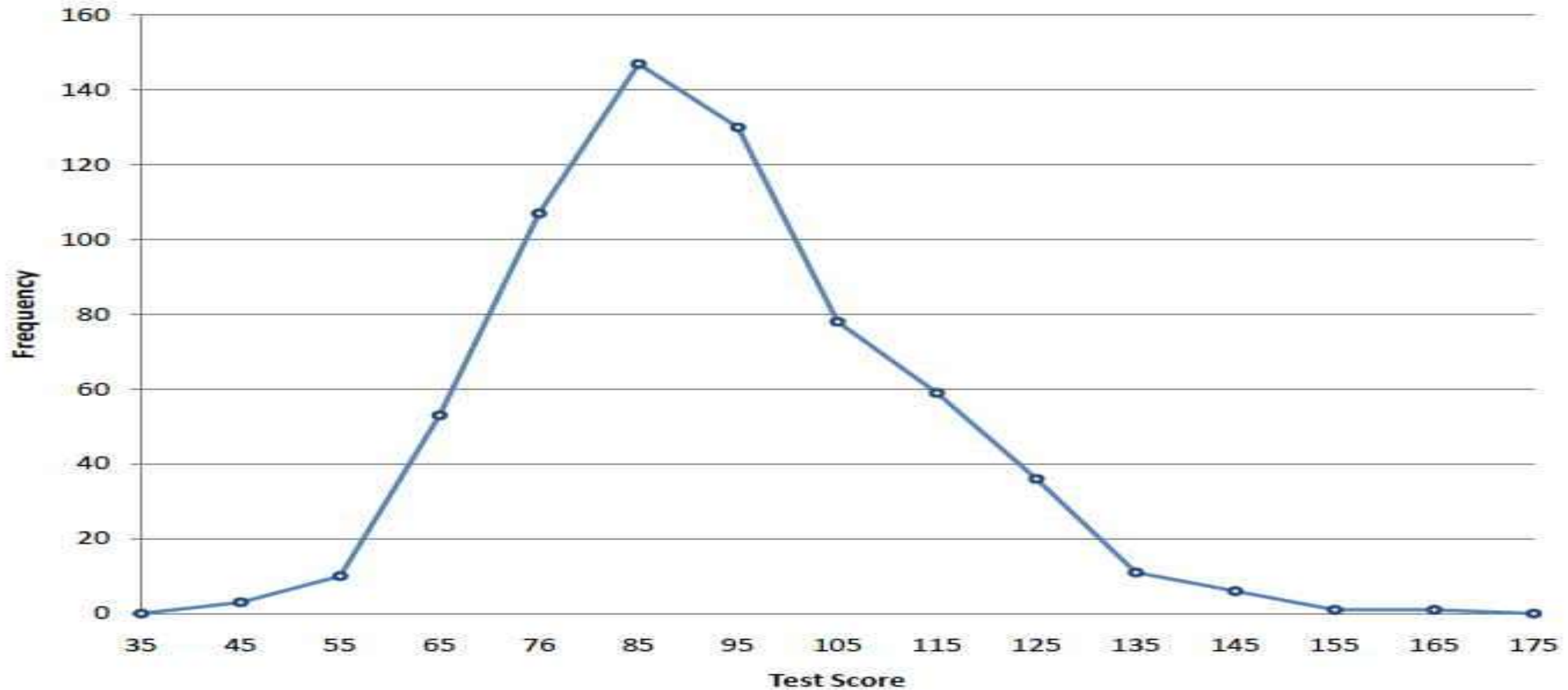


# PIE GRAPH

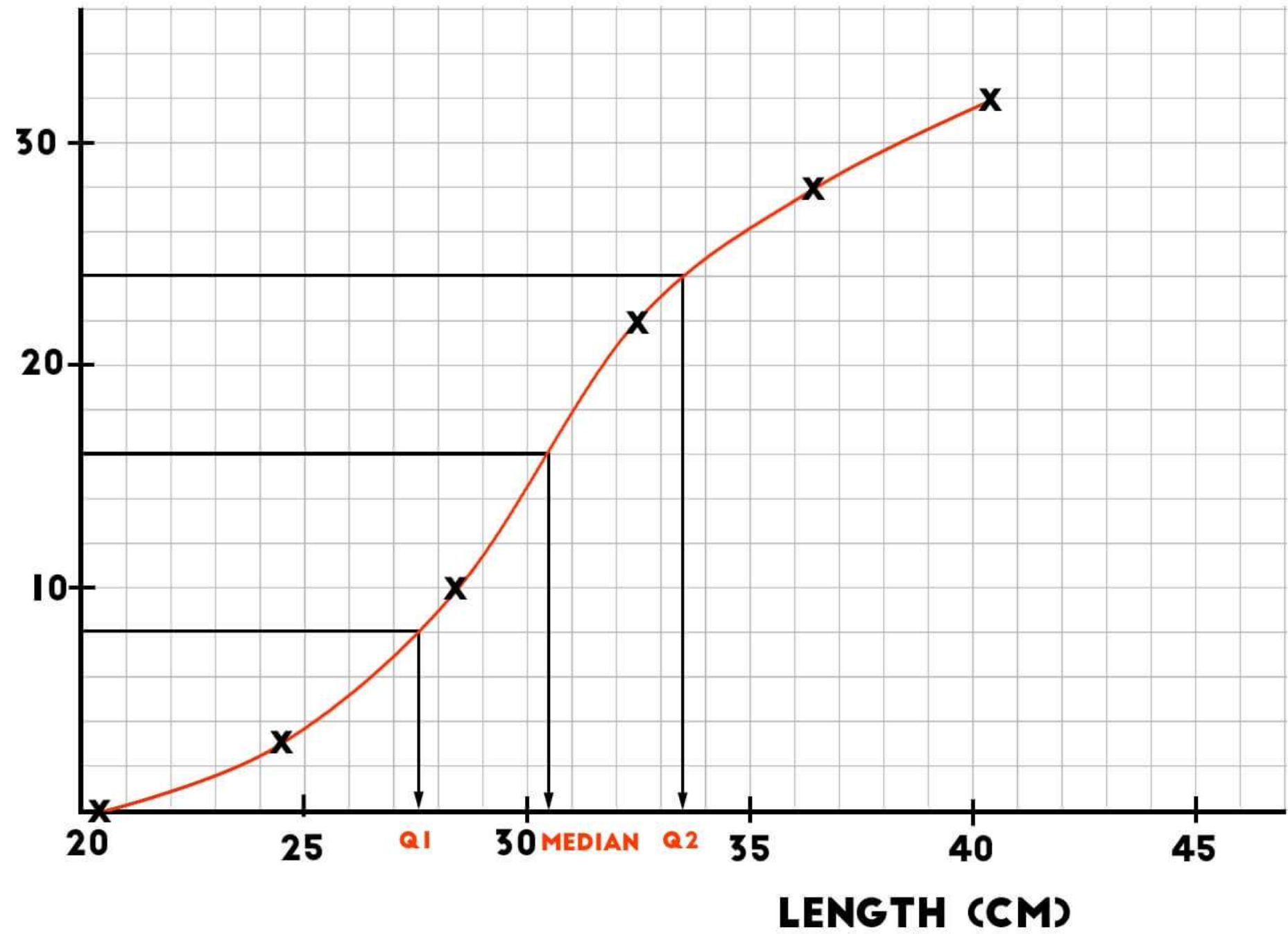
## Recommended Diet



# FREQUENCY POLYGON



# CUMULATIVE FREQUENCY





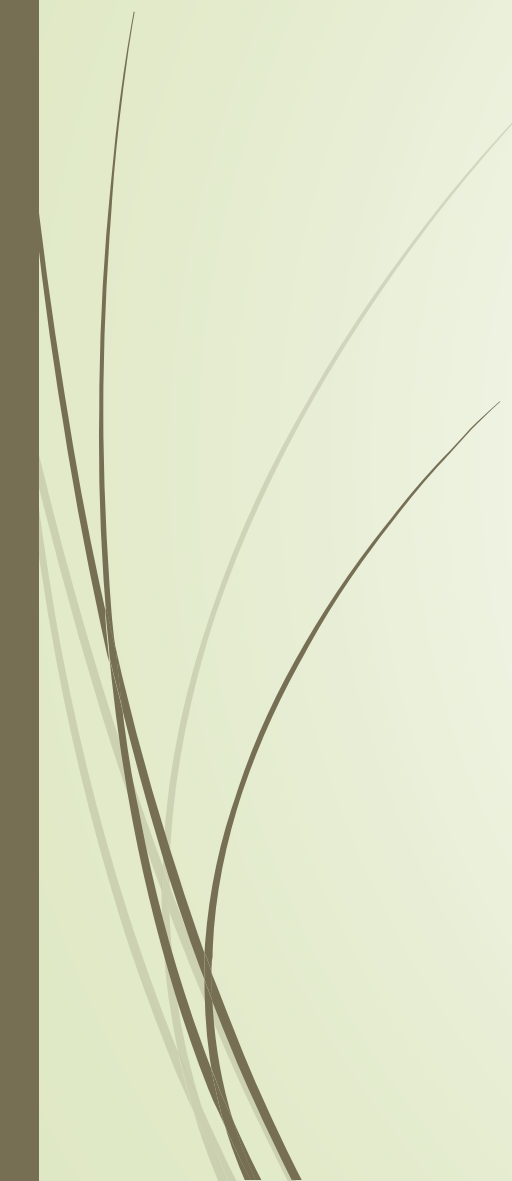
# MEASURES OF CENTRAL TENDENCY (MCT)

- AM
- GM
- HM
- MEDIAN
- MODE



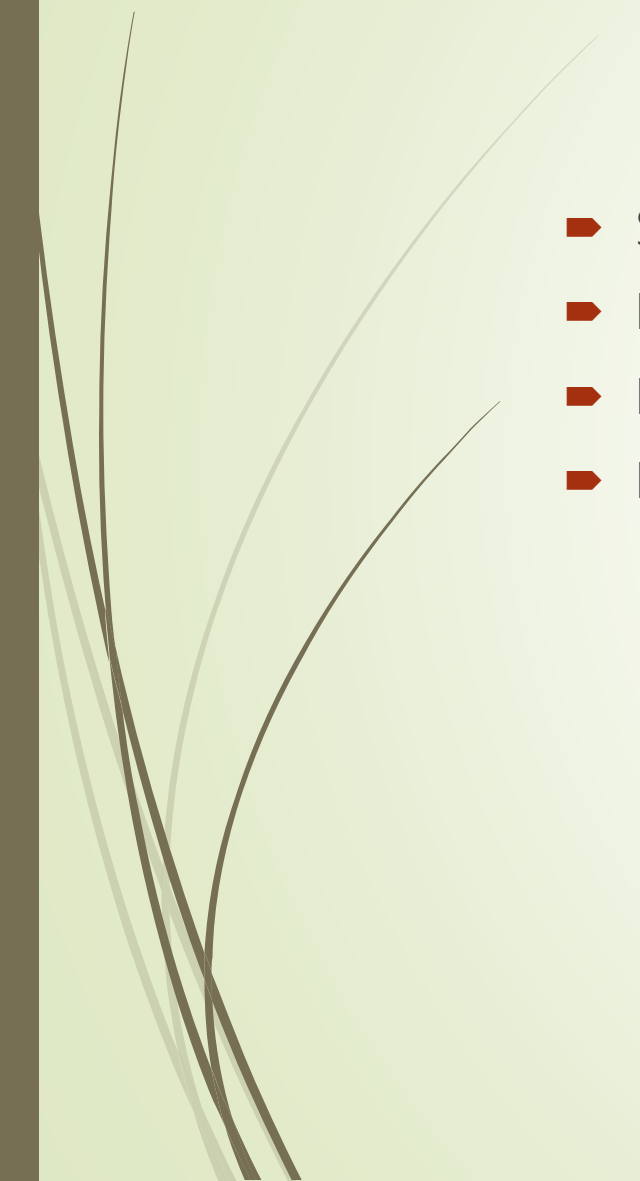


# MEASURES OF DISPERSION

- RANGE
  - QUARTILE DEVIATION
  - MD
  - SD
  - CV
- 

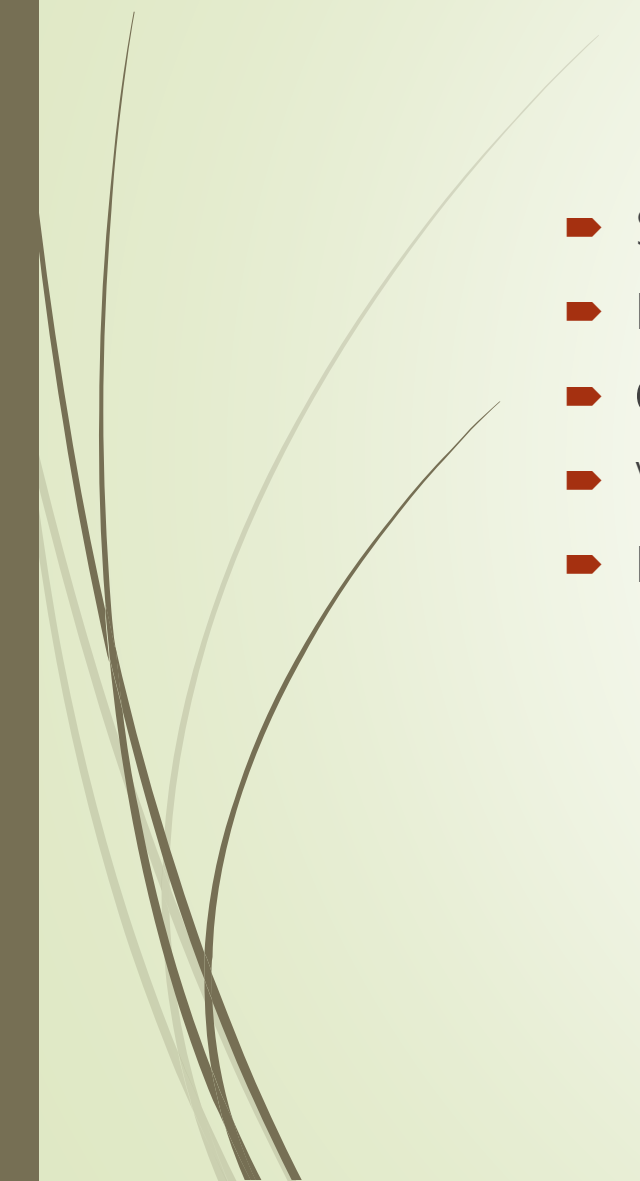


# SKEWNESS AND KURTOSIS

- Skewness- asymmetry
  - Positively skewed distribution
  - Kurtosis- peakedness
  - Leptokurtic
- 



# PARAMETRIC OR STANDARD TEST

- ▶ Sample is large
  - ▶ Population have normal distribution
  - ▶ Observations – independent
  - ▶ Variables- interval or ratio scale
  - ▶ Eg. t- test, z test, F test
- 

# NON- PARAMETRIC TEST

- Distribution free test
- Normal distribution- doubtful
- Small sample size
- Data – ranks or different between pairs
- Eg. Chi square test, The Mann- whitney U test, kendall W, rho

# IMPORTANT ANALYSES

t -Test:

- ▶ Comparing the means of two independent groups
- ▶ Sample size  $< 30$
- ▶ Paired t-test:

ANOVA:

- ▶ Statistical approach for determining the means of two or more populations are equal.

One way ANOVA

ANCOVA



# CORRELATION

- Intensity of association between 2 variable
- Ranges from + 1 to -1

## BISERIAL CORRELATION:

- One variable is **continous** and other is **dichotomous**.

## POINT BISERIAL CORRELATION:

- One of the 2 variables is a **genuine dichotomy** ( can't measured on interval scale)



# REGRESSION

- Determination of statistical relationship between two or more variables
- 



# DIFFERENT ANALYSIS

## PATH ANALYSIS:

- ▶ Clear picture of the direct and indirect effects of the independent variables on the dependent variable.

## DISCRIMINANT FUNCTION ANALYSIS:

- ▶ Necessary to find out how populations actually discriminate or differ

## CANONICAL ANALYSIS:

- ▶ Explaining several dependent variables by a set of independent variables.

## FACTOR ANALYSIS:

- ▶ Resolve large set of measured variables by few categories.





# CONTINUED

## META ANALYSIS:

- Set of techniques for **summarising** the findings of many studies

## CLUSTER ANALYSIS:

- Classify objects into small number of **mutually exclusive** and **exhaustive** groups.
- Better understanding of buyer behaviour

## MULTIDIMENSIONAL SCALING:

- Data reduction technique
- Uncover the hidden structure of a set of data.



# NON-PARAMETRIC TESTS

## CHI-SQUARE TEST:

- ▶ Testing the agreement between a given hypothesis and observed data
- ▶ Test goodness of fit

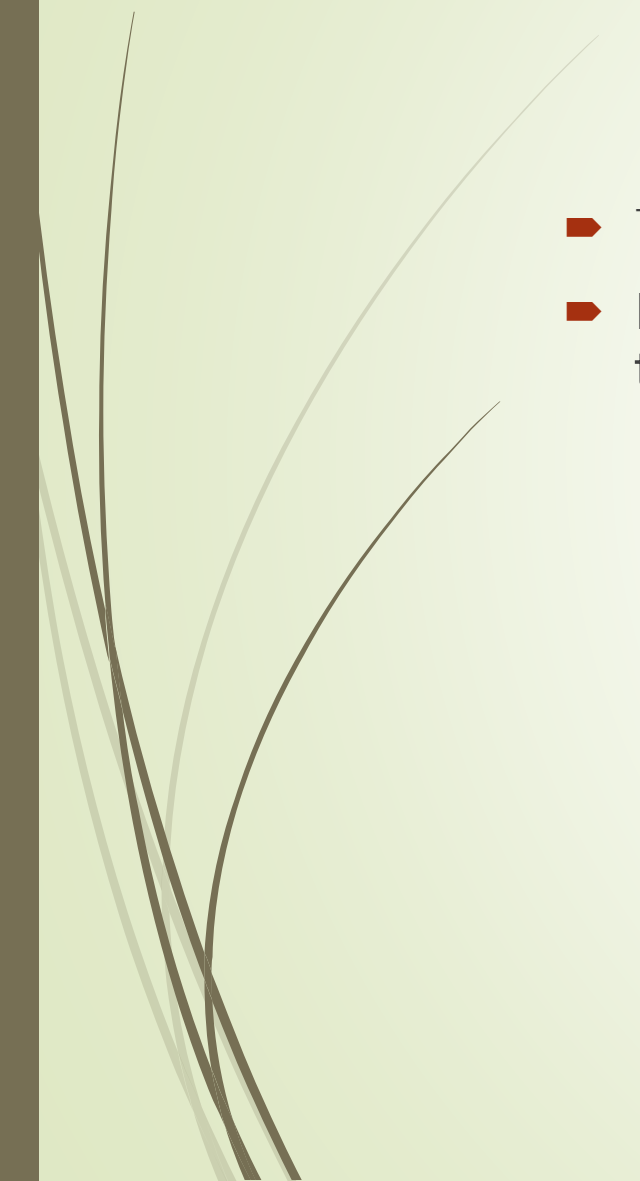



# MCNEMAR TEST

- Significance of change
  - Research involve before and after situation and data is nominal
  - Effectiveness of a particular treatment.
- 

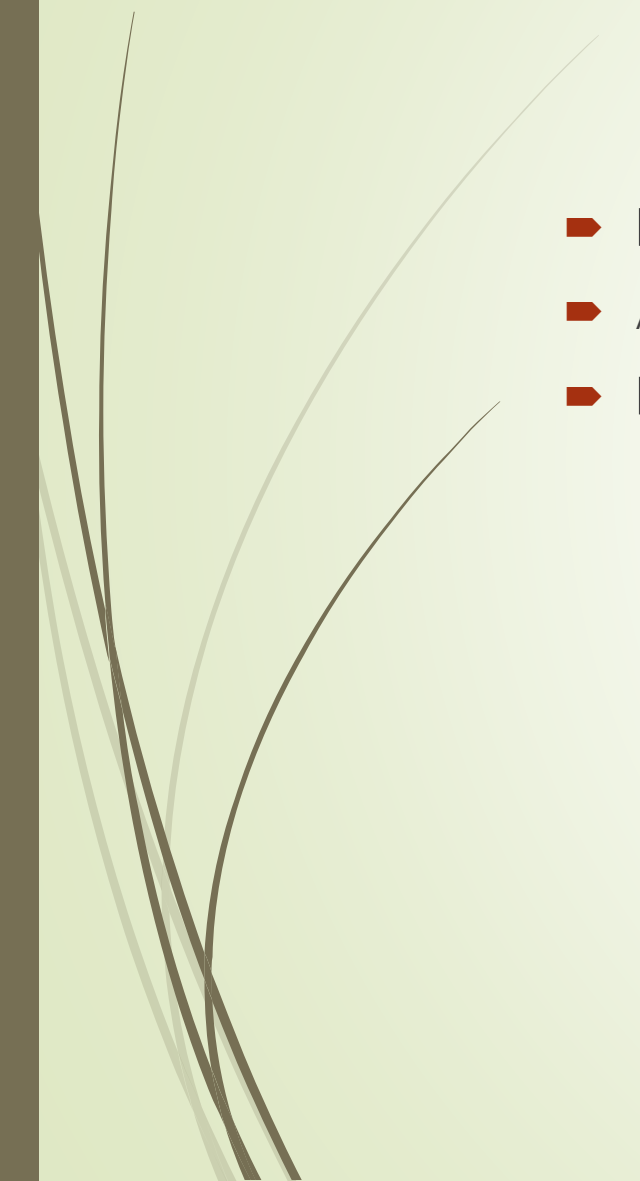


# MANN WHITNEY U TEST

- ▶ Test whether two **independent groups** have been drawn from **same population**.
  - ▶ Powerful test, alternative to parametric t test when researcher wants to **avoid t test assumptions**
- 



# KRUSKAL WALLIS – ONE WAY ANOVA

- ▶ H test.
  - ▶ Applied to more than 2 samples of observation
  - ▶ Nonparametric substitute for simple ANOVA
- 



# FRIEDMAN TWO WAY ANOVA

- ▶ Testing the null hypothesis of difference between treatments by ranking the data

## KENDALL COEFFICIENT OF CONCORDANCE:W

- ▶ Measures extent of association among several sets of ranking of events, objects
- ▶ Studies of cluster of variables



# OTHER MEASURES

## INDEX NUMBERS:

- ▶ Changes in various economic and social phenomena
- ▶ Approximate indicators

## TIME SERIES ANALYSIS:



# REFERENCES

- ▶ Research methods in social sciences and extension education by G.L.Ray & Sagar Mondal
- ▶ Research methodology – methods and techniques (2<sup>nd</sup> revised edition) by C.R. Kothari.
- ▶ Essentials of research design and methodology by Geoffrey Marczyk, David DeMatteo, and David Festinger.
- ▶ Research methodology by Ranjit kumar
- ▶ Fundamentals of research methodology and statistics by Yogesh kumar singh.





# QUIZ



- 1) Warranty cards is an example for
  - a) Primary data
  - b) Secondary data
  - c) both
  - d) none
- 2) Commonly used program for statistical analysis in social sciences is
  - a) SAS
  - b) SPSS
  - c) GENSTAT
  - d) STATA
- 3) Which of the following is not an example for standard test ?
  - a) T test
  - b) chi square test
  - c) Z test
  - d) F test
- 4) Which test is suitable for small sample size
  - a) Parametric
  - b) Non parametric
  - c) Standard
  - d) None
- 5) Vertical bar diagram without gap between the bars is
  - a) Ogive
  - b) frequency polygon
  - c) Histogram
  - d) Bar diagram



# QUIZ



6) Mode is obtained from

- a) Ogive   b) Histogram   c) frequency curve   d) Lorenz curve

7) Average which is applicable to all sorts of data

- a) AM   b) GM   c) HM   d) Median

8) Quickest measure of centrality is

- a) AM   b) GM   c) Mode   d) Median

9) Which is used for measuring intelligence

- a) Median   b) Mode   c) AM   d) GM

10) Which test is used for the significance of changes ?

- a) Chi square   b) Mann Whitney U   c) Mc Nemar   d) t test



THANK YOU