

RESEARCH METHODOLOGY & BIostatISTICS

By

Dr. Michael Chigunta - PhD

House Rules

- **Class Attendance is a MUST**
- **Free Participation**
- **No Opinion is Wrong**
- **No Phones**
- **Take Notes**
- **Ask at any point**
- **Any Other?????**

Course Objectives

- 1. Apply biostatistics concepts and tools critical in biomedical analysis of research data**
- 2. Demonstrate knowledge of how to summarize data and use statistical methods.**
- 3. Utilize epidemiology concepts and methods in disease control**
- 4. Demonstrate knowledge and skills of scientific enquiry and research methods**
- 5. Demonstrate knowledge and skills in project proposal development and research writing**

Defining Research

- ✓ **RE:** Regarding or About!!!!
- ✓ **Search:** Attempt to find something
- ✓ **Systematic:** Planned or Ordered procedure

Introduction to Research

- ✓ To go about seeking.
- ✓ To investigate systematically.
- ✓ Any gathering of data, information and facts for the advancement of knowledge.
- ✓ A process of steps used to collect and analyse information to increase our understanding of a topic or issue.

Purpose of Research

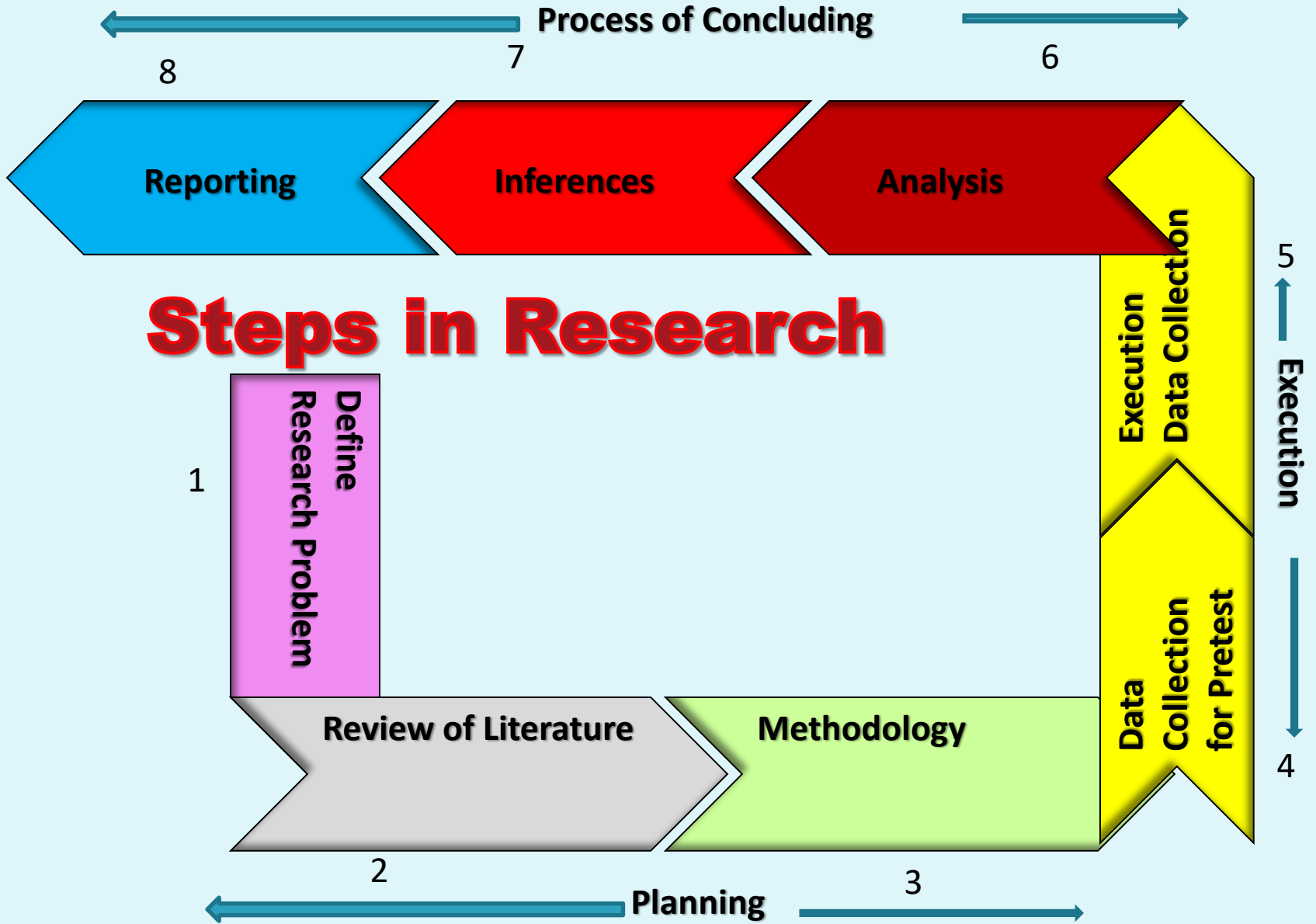
- ❑ Documentation, discovery & application to Communities of Practice:
 - Establish or confirm facts.
 - Reaffirm the results of previous work.
 - Solve new or existing problems.
 - Support theorems.
 - Develop new theories.
 - Test the validity of instruments, procedures, or experiments

Research Paradigms



Research Worldviews

1. **Quantitative** (Numbers Count/Matter)
2. **Qualitative** (Experiences/ Perceptions / Feelings etc Matter)
3. **Mixed Methods (Combo)**

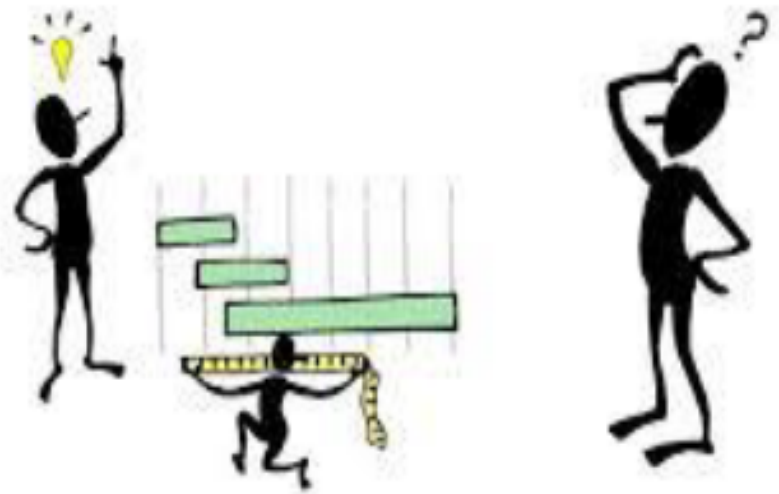




VARIABLES

- ❑ Variables are the things we try to understand, that are measured or data collected about, and whose value can vary from subject to subject.
- ❑ Variables can be defined as “discrete” or “continuous”.

What is a Variable?



Variables are closely linked to scales of measure!

Operationalisation:

- Operationalization is process that defines fuzzy concepts and allows them to be measured, empirically and quantitatively.

[Strictly defining **variables** into measurable factors:

For predictive studies

Correlational with a distinct time order or sequence to the variables:

Independent variable

(Manipulated: Presumed Cause)

Dependent variable

(Presumed Effect - *outcome*) is the response that is measured.

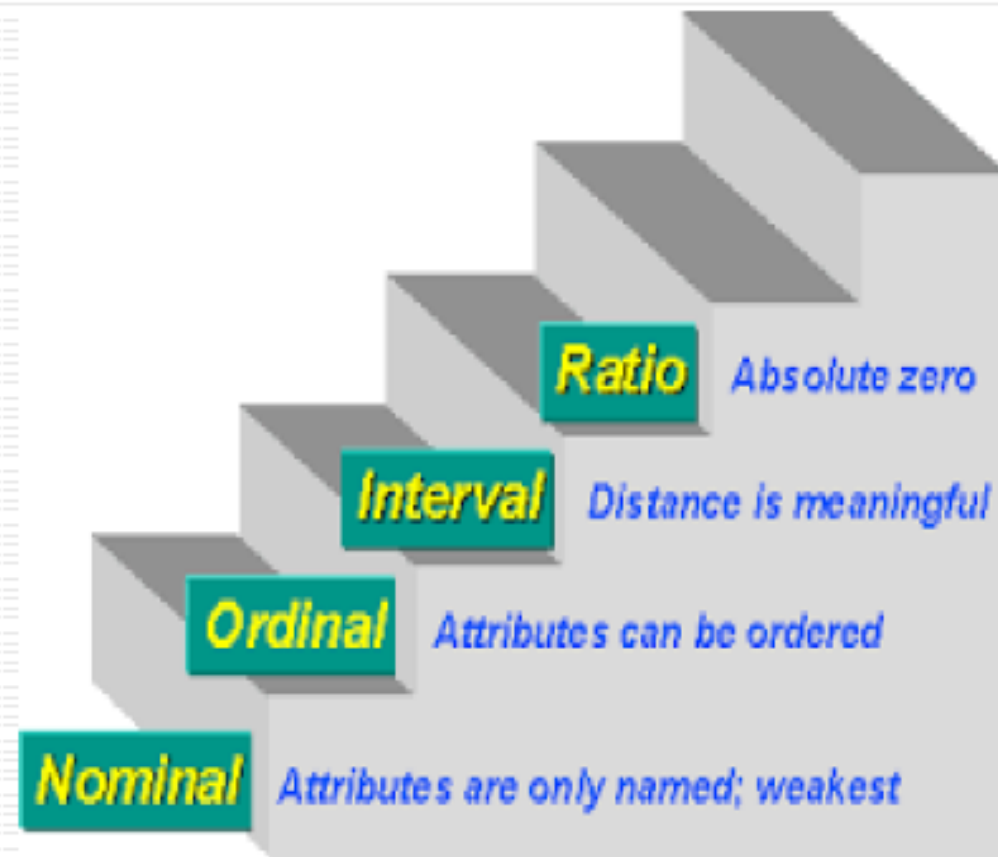
Confounding variable (May not control for but can have effect on outcome)

Predictor variable (the one that precedes the other and we have information about it and use it to predict)

Criterion variable (the second other 'thing' being predicted).

LEVELS OF MEASUREMENT

- ❑ **Categorically scaled:**
 - **Nominal** (*the numerical value just names the attribute uniquely*)
 - **Ordinal** (*the numbers can be rank ordered*)
- ❑ **Continuously scaled:**
 - **Interval** (*distance between numbers has meaning*)
 - **Ratio** (*has absolute zero and meaningful fractions can be constructed*)



The levels of measure influence what type of analyses can be done!

HYPOTHESIS

- A hypothesis is a statement about the relationship between two or more **variables**. It must have three components:
 - The variables
 - The population
 - The relationship between the variables

[State **null** AND **alternative** hypothesis]

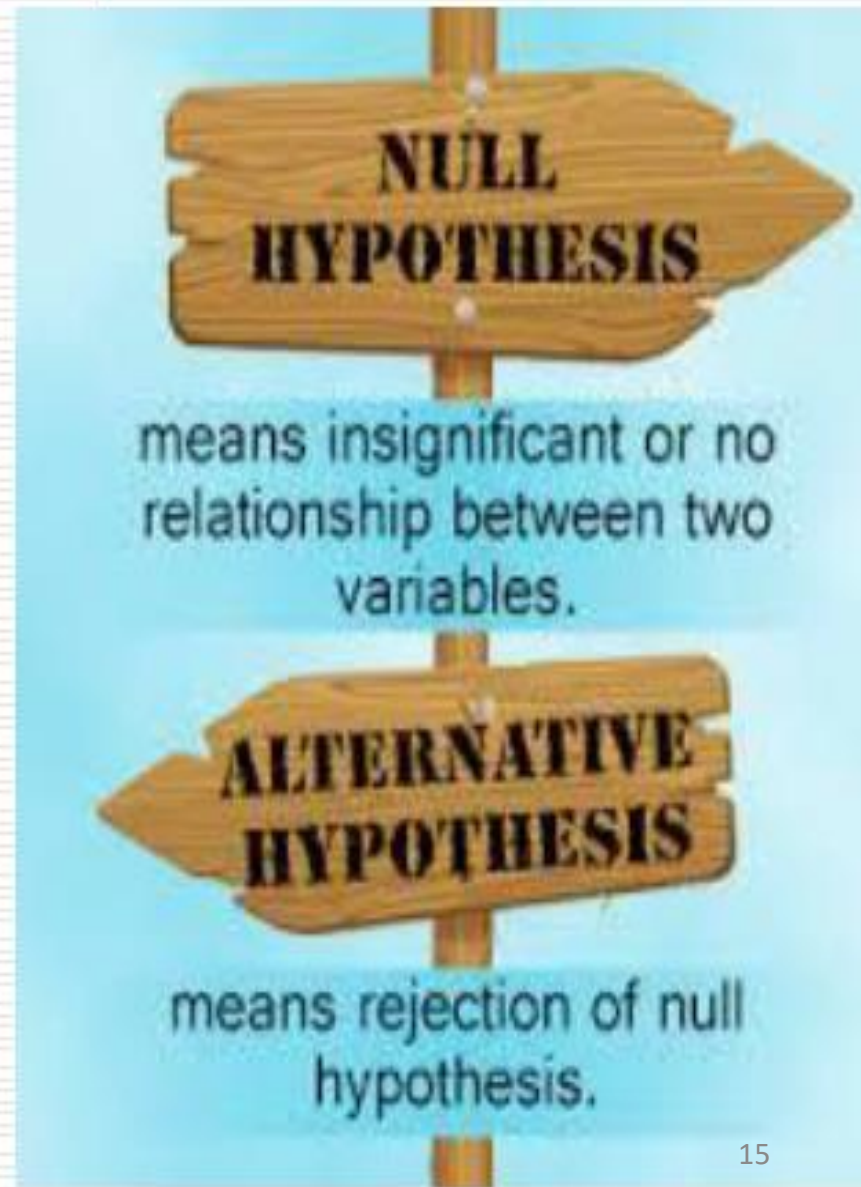


**"I've narrowed it to two hypotheses:
it grew or we shrunk."**

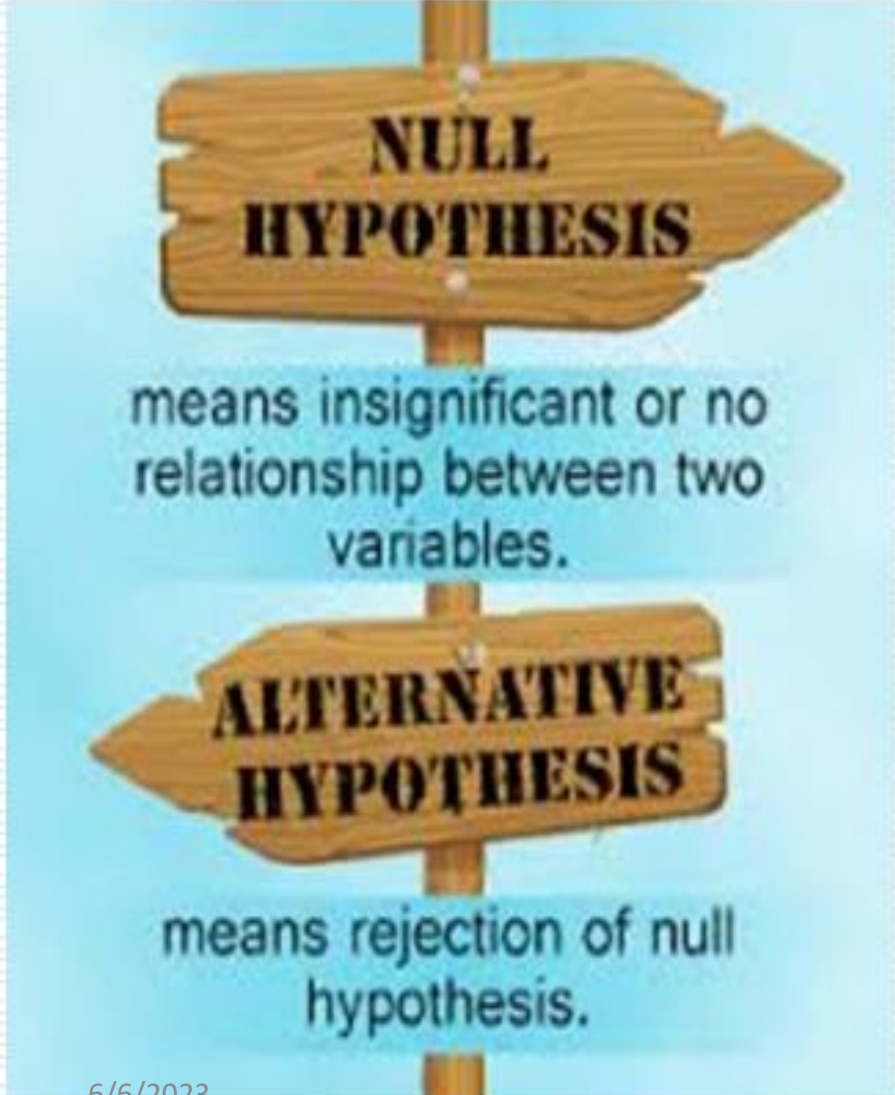
The Null Hypothesis (H_0)

Null hypothesis refers to a general statement or default position that **there is no** relationship between measured phenomenon.

[sample observations result purely from chance]



The Alternative Hypothesis (H_1 or H_a)



NULL HYPOTHESIS

means insignificant or no relationship between two variables.

ALTERNATIVE HYPOTHESIS

means rejection of null hypothesis.

- The **alternative hypothesis** is a statement of what a hypothesis is set up to establish.
 - Opposite of null **hypothesis**.
 - Only reached if H_0 is rejected.
 - Frequently desired conclusion

[sample observations are influenced by some non-random cause]

Interpreting Hypothesis Testing

Developing

- ❑ **A good hypothesis must be based on a good research question**

Interpreting

- ❑ If statistical significance is shown the null hypothesis is rejected. If not significant the null hypothesis is accepted.

RESEARCH OBJECTIVES

Research objectives are declarative statements that describe the reasons (intent) and specific actions that will be done to measure the variables. Usually Overall Objective and Specific Objectives.

[What the researcher will be able to achieve at the end of the study]



The Research Questions are Transformed into Research Objectives!

Research Objectives

- ❖ Research Objectives should be clear and achievable.
- ❖ Generally, they are written as statements, using Active VERBs.

(For example, 'to describe ...', 'to determine ...', 'to establish ...', 'to assess, to investigate -----'etc.)

(SMART)

Objectives should infer in the end of the study

The Initial Research Process

Research
Problem

Research
Question

Hypothesis

Research
Objectives

STEP-2

REVIEW OF LITERATURE

Review of literature

❖ What ?

❖ Why ?

❖ Where ?

What ?

REVIEW OF LITERATURE

**Literature Review is the documentation
of published and unpublished work
from secondary sources of data
in the areas of specific interest to the researcher.**

Why ? - PURPOSE OF REVIEW

- ❖ **To find out already investigated problems and those that need further investigation.**
- ❖ **To formulate researchable hypothesis.**
- ❖ **To gain a background knowledge**
- ❖ **To identify data sources**
- ❖ **To learn how others structured their reports.**

Where ?

SOURCES OF LITERATURE

- ❖ **Books and Journals**
- ❖ **Databases**
 - ❖ **Bibliographic Databases**
 - ❖ **Abstract Databases**
 - ❖ **Full-Text Databases**
- ❖ **Govt. and NGO Records & Reports**
- ❖ **Internet**
 - ❖ **On line journals: ww.articalbase.com**
 - ❖ **E. Databases – Popline, Medline**
- ❖ **Research Dissertations / Thesis**

How to search

- ..\Conducting a literature search using PubMed..mp4

Step-4 & 5

Data Collection and Data Management

Sources of Data

- **Primary** –Own generated data
- **Secondary** –Already generated data
 - ❖ **Published**
 - ❖ **Non-Published**

Primary Vs Secondary source of Data

Primary data

- Need to be generated
- First hand information
- Questionnaire
- Purpose served
- Analysis as per purpose
- Require more time and money

6/6/2023

Secondary data

- Readily available
- Second hand information
- Not need of questionnaire
- Purpose served ?
- Descriptive
- Less expensive

Type of Data Collection Methods

- ❖ **Interview**

 - ❖ **Personnel**

 - ❖ **Telephonic**

- ❖ **Observation**

- ❖ **Experimental**

- ❖ **Interview and Observation**

- ❖ **Observation and Experimental**

- ❖ **Interview , Observation and Experimental**

Forms of questions(Open Vs Closed)

Open ended

- Possible responses are not given.
- Mean, SD, Median
- For seeking opinions, attitudes ,perceptions
- Provides in depth info.
- Experience of investigator and analyst required

Close ended

- Categories are given already coded
- Proportion
- For eliciting factual information
- Not so depth
- Investigator's bias
- Ease of answering,
- Easy to analyse

Considerations in formulating questionnaire

(Questionnaire/Interview schedule)

- ❖ Use simple and everyday language
- ❖ Do not use ambiguous questions(?!?)
- ❖ Do not ask leading questions
- ❖ The order of questions:
- ❖ Guideline for filling an instrument, pen-pencil
- ❖ **Pre testing**

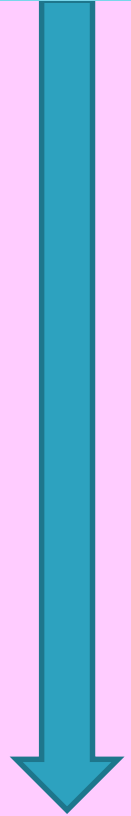
Validity of a Research Instrument

Ability of an instrument to measure what it is designed to measure being measured

- ❖ **Establish the logical link between the questions and objectives**
- ❖ **Items/questions cover the full range of issue/attitude being measured**

Steps

1. Decide the information required.
2. Define the target respondents.
3. Method(s) of reaching target
4. Decide on question content.
5. Develop the question wording.
6. Put questions into a meaningful order.
7. Check the length of the questionnaire.
8. Pre-test the questionnaire.
9. **Develop the final survey form**



Tabulation – Content of Table

- ❖ **Table No. Sequence in the text**
- ❖ **Title of Table –short, clear and self explanatory to say about what the table is ?**
- ❖ **Body of Table –consist of rows and columns**
 - ❖ **Rows – 1st row shows headings of columns**
 - ❖ **1st column shows headings of rows**
 - ❖ **rest of rows and columns are showing data as per required**
 - ❖ **number of rows and columns should be limited to maintained simplicity of table**
 - ❖ **source of data (if it is other than the present study) should be written just below the body of table**
- ❖ **Source of Data ?**
- ❖ **Foot Note - written just below the body of table, if there is any hidden information**
- ❖ **Inferences –summary value of table**

Types of Tables

❖ As per purpose

- ❖ General tables –about Socio-demographic profile
- ❖ Specific tables –about Aims and objectives

❖ As per originality

- ❖ Original tables-from original Data
- ❖ Derived tables –from original tables

❖ As per Construction

- ❖ Simple tables- showing one variable at one time
- ❖ Complex tables – showing > one variable at one time

Home Insert Page Layout Formulas Data Review View PDF

PivotTable Table Picture Clip Art Shapes SmartArt Column Line Pie Bar Area Scatter Other Charts Hyperlink Text Box Header & Footer WordArt Signature Line Object Symbol

Insert PivotTable

Summarize data using a PivotTable.

PivotTables make it easy to arrange and summarize complicated data and drill down on details.

Press F1 for more help.

	C	D	E	F	G	H	I	J	K	L
	Age	Age Groups	Sex	Duration of diabetes in yrs	Duration of DM Groups	Occupation	Education	Income per capita	SES	Patients Satisfaction with their Health
5	40	(2) 31-45 Years	F	5	(1) 1-5 Years	(1) Unemployed	illiterate	24000	class II	(4) Satisfied
6	35	(2) 31-45 Years	F	3	(1) 1-5 Years	(1) Unemployed	Primary	57000	class I	(2) Dissatisfied
7	55	(3) 46-64 Years	F	5	(1) 1-5 Years	(1) Unemployed	Graduate	16666	class III	(3) neither satisfied nor dissatisfied
8	48	(3) 46-64 Years	F	5	(1) 1-5 Years	(1) Unemployed	Middle	15000	class III	(3) neither satisfied nor dissatisfied
9	30	(1) 16-30 Years	M	6	(2) 6-10 Years	(6)Professional	PG	300000	class I	(3) neither satisfied nor dissatisfied
10	28	(1) 16-30 Years	M	2	(1) 1-5 Years) Semi Profession	Graduate	60000	class I	(4) Satisfied
11	23	(1) 16-30 Years	F	5	(1) 1-5 Years	(1) Unemployed	Graduate	150000	class I	(3) neither satisfied nor dissatisfied
12	35	(2) 31-45 Years	F	2	(1) 1-5 Years	(1) Unemployed	Graduate	60000	class I	(3) neither satisfied nor dissatisfied
13	53	(3) 46-64 Years	F	7	(2) 6-10 Years) Semi Profession	PG	60000	class I	(2) Dissatisfied
14	50	(3) 46-64 Years	M	5	(1) 1-5 Years	(6)Professional	Graduate	60000	class I	(4) Satisfied
15	50	(3) 46-64 Years	F	10	(2) 6-10 Years	(1) Unemployed	illiterate	60000	class I	(3) neither satisfied nor dissatisfied
16	54	(3) 46-64 Years	M	1.5	(1) 1-5 Years	(3) Semi skilled	Sec..	48000	class I	(4) Satisfied
17	58	(3) 46-64 Years	M	15	(3) 11 & Above Years	(2) Unskilled	Primary	12000	class IV	(3) neither satisfied nor dissatisfied
18	49	(3) 46-64 Years	F	20	(3) 11 & Above Years	(1) Unemployed	illiterate	60000	class I	(4) Satisfied
19	55	(3) 46-64 Years	M	5	(1) 1-5 Years	(6)Professional	Graduate	60000	class I	(2) Dissatisfied
20	54	(3) 46-64 Years	M	15	(3) 11 & Above Years	(2) Unskilled	Primary	24000	class II	(2) Dissatisfied
21	50	(3) 46-64 Years	M	2	(1) 1-5 Years) Semi Profession	HSC	60000	class I	(3) neither satisfied nor dissatisfied
22	35	(2) 31-45 Years	M	5	(1) 1-5 Years	(6)Professional	Graduate	75000	class I	(4) Satisfied
23	35	(2) 31-45 Years	F	10	(2) 6-10 Years	(1) Unemployed	Middle	80000	class I	(3) neither satisfied nor dissatisfied
24	59	(3) 46-64 Years	M	15	(3) 11 & Above Years	(1) Unemployed	Sec..	60000	class I	(2) Dissatisfied
25	32	(2) 31-45 Years	M	2	(1) 1-5 Years	(2) Unskilled	Graduate	40000	class II	(3) neither satisfied nor dissatisfied
26	45	(2) 31-45 Years	M	2	(1) 1-5 Years) Semi Profession	illiterate	40000	class II	(4) Satisfied
27	40	(2) 31-45 Years	M	6	(2) 6-10 Years	(6)Professional	PG	84000	class I	(3) neither satisfied nor dissatisfied
28	25	(1) 16-30 Years	M	4	(1) 1-5 Years	(1) Unemployed	Grad			(3) neither satisfied nor dissatisfied
29	35	(2) 31-45 Years	F	6	(2) 6-10 Years	(1) Unemployed	illit			(3) satisfied

Wireless Network Connection is now connected

Connected to: Akesh Architects
Signal Strength: Low

Tabulation

Microsoft Excel interface showing a PivotTable titled "Patients Satisfaction with their Health". The PivotTable is structured as follows:

Count of Patients Satisfaction with their Health	Duration of DM Groups			Grand Total
Patients Satisfaction with their Health	(1) 1-5 Years	(2) 6-10 Years	(3) 11 & Above Years	Grand Total
(1) Very Dissatisfied	8		3	11
(2) Dissatisfied	46	22	10	78
(3) neither satisfied nor dissatisfied	51	36	11	98
(4) Satisfied	37	15	8	60
(5) Very Satisfied	1	1	1	3
Grand Total	143	74	33	250

The PivotTable Field List on the right shows the following configuration:

- Report Filter: Patients Satisfaction with their Health
- Column Labels: Duration of DM Groups
- Row Labels: Patients Satisfaction with their Health
- Values: Count of Patients Satisfaction with their Health

Two red arrows point to the "Duration of DM Groups" and "Patients Satisfaction with their Health" fields in the PivotTable Field List.

Diagrammatic Presentations

- ❖ Bar
 - ❖ Simple
 - ❖ Multiple
 - ❖ Component
- ❖ Pie
- ❖ Line
- ❖ Pictogram
- ❖ Spot Map

- ❖ Histogram
- ❖ Frequency Polygon
- ❖ Cumulative Frequency Polygon
- ❖ Scatter Diagram
- ❖ Box and Whisker
- ❖ Correlation Diagram

Qualitative Data

Quantitative Data

Exercise MC 2003.xls [Compatibility Mode] - Microsoft Excel

Home Insert Page Layout Formulas Data Review View PDF

PivotTable Table Picture Clip Art Shapes SmartArt Column Line Pie Bar Area Scatter Other Charts Hyperlink Text Box Header & Footer WordArt Signature Line Object Symbol

S No.	Name	Age	Age Groups	Sex	Duration of DM Groups	Occupation	Education	Income per capita	SES	Patients Satisfaction with their Health	
1	Vimla	40	(2) 31-45 Years		(1) 1-5 Years	(1) Unemployed	illiterate	24000	class II	(4) Satisfied	
2	Radha	35	(2) 31-45 Years		(1) 1-5 Years	(1) Unemployed	Primary	57000	class I	(2) Dissatisfied	
3	Anamika sharma	55	(3) 46-64 Years		(1) 1-5 Years	(1) Unemployed	Graduate	16666	class III	(3) neither satisfied nor dissatisfied	
4	Gulab devi	48	(3) 46-64 Years		(1) 1-5 Years	(1) Unemployed	Middle	15000	class III	(3) neither satisfied nor dissatisfied	
5	V K Gupta	30	(1) 16-30 Years	M	(2) 6-10 Years	(6)Professional	PG	300000	class I	(3) neither satisfied nor dissatisfied	
6	Rajendra Prasad	28	(1) 16-30 Years	M	(1) 1-5 Years) Semi Profession	Graduate	60000	class I	(4) Satisfied	
7	Meenal Singh	23	(1) 16-30 Years	F	(1) 1-5 Years	(1) Unemployed	Graduate	150000	class I	(3) neither satisfied nor dissatisfied	
8	Sheela Shrivastav	35	(2) 31-45 Years	F	2	(1) 1-5 Years	(1) Unemployed	Graduate	60000	class I	(3) neither satisfied nor dissatisfied
9	Shashi Jain	53	(3) 46-64 Years	F	7	(2) 6-10 Years) Semi Profession	PG	60000	class I	(2) Dissatisfied
10	Kaluram	50	(3) 46-64 Years	M	5	(1) 1-5 Years	(6)Professional	Graduate	60000	class I	(4) Satisfied
11	Rampyari devi	50	(3) 46-64 Years	F	10	(2) 6-10 Years	(1) Unemployed	illiterate	60000	class I	(3) neither satisfied nor dissatisfied
12	Hukam Singh	54	(3) 46-64 Years	M	1.5	(1) 1-5 Years	(3) Semi skilled	Sec..	48000	class I	(4) Satisfied
13	Mohd. Ali	58	(3) 46-64 Years	M	15	(3) 11 & Above Years	(2) Unskilled	Primary	12000	class IV	(3) neither satisfied nor dissatisfied
14	Saroj	49	(3) 46-64 Years	F	20	(3) 11 & Above Years	(1) Unemployed	illiterate	60000	class I	(4) Satisfied
15	Sameer	55	(3) 46-64 Years	M	5	(1) 1-5 Years	(6)Professional	Graduate	60000	class I	(2) Dissatisfied
16	Ram Prasad	54	(3) 46-64 Years	M	15	(3) 11 & Above Years	(2) Unskilled	Primary	24000	class II	(2) Dissatisfied
17	Sitaram	50	(3) 46-64 Years	M	2	(1) 1-5 Years) Semi Profession	HSC	60000	class I	(3) neither satisfied nor dissatisfied
18	Sukhdeep	35	(2) 31-45 Years	M	5	(1) 1-5 Years	(6)Professional	Graduate	75000	class I	(4) Satisfied
19	Radha Devi	35	(2) 31-45 Years	F	10	(2) 6-10 Years	(1) Unemployed	Middle	80000	class I	(3) neither satisfied nor dissatisfied
20	Rajkamal	59	(3) 46-64 Years	M	15	(3) 11 & Above Years	(1) Unemployed	Sec..	60000	class I	(2) Dissatisfied
21	Samundar Singh	32	(2) 31-45 Years	M	2	(1) 1-5 Years	(2) Unskilled	Graduate	40000	class II	(3) neither satisfied nor dissatisfied
22	Keshuram	45	(2) 31-45 Years	M	2	(1) 1-5 Years) Semi Profession	illiterate	40000	class II	(4) Satisfied
23	Kailash Rathi	40	(2) 31-45 Years	M	6	(2) 6-10 Years	(6)Professional	PG	84000	class I	(3) neither satisfied nor dissatisfied
24	Sameera Khan	25	(1) 16-30 Years	M	4	(1) 1-5 Years	(1) Unemployed	Graduate	60000	class I	(3) neither satisfied nor dissatisfied
25	Rebhi	35	(2) 31-45 Years	F	6	(2) 6-10 Years	(1) Unemployed	illiterate	20000	class III	(2) Dissatisfied

Pie

Insert a pie chart.

Pie charts display the contribution of each value to a total.

Use it when values can be added together or when you have only one data series and all values are positive.

Diagrammatic Presentations

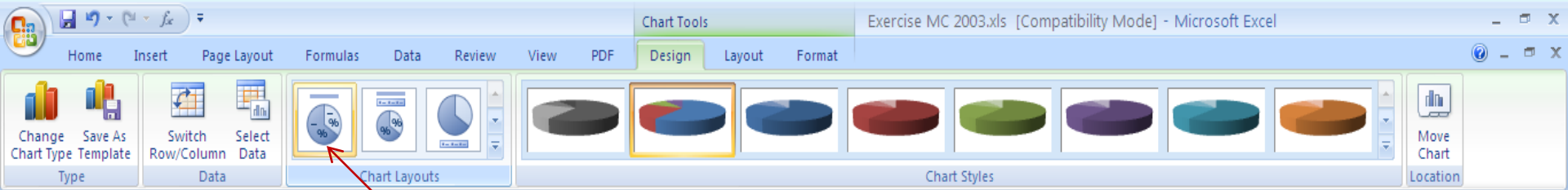
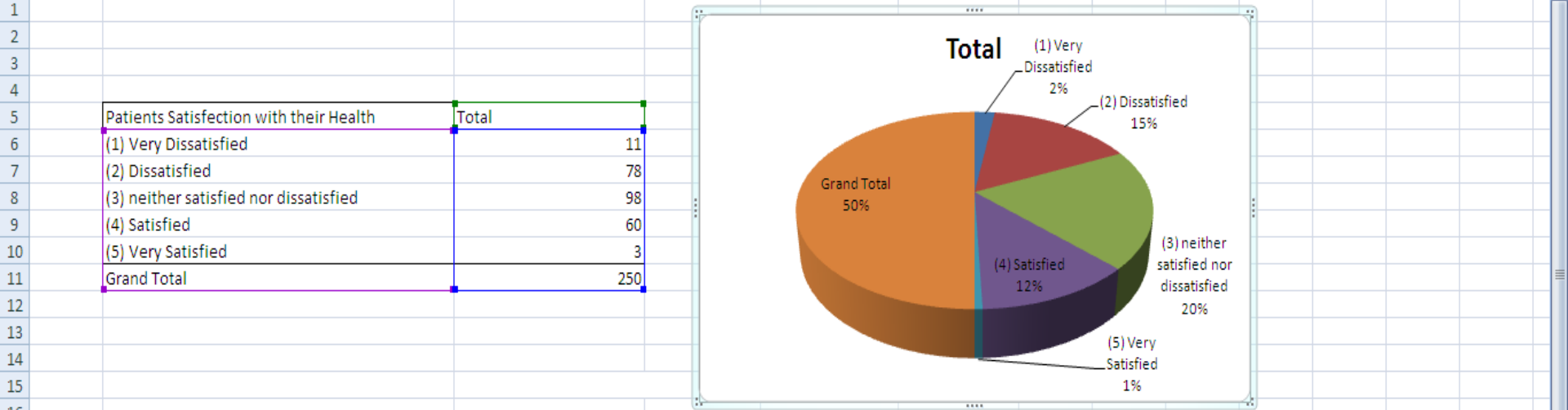


Chart 3

A B C D E F G H I J K L M



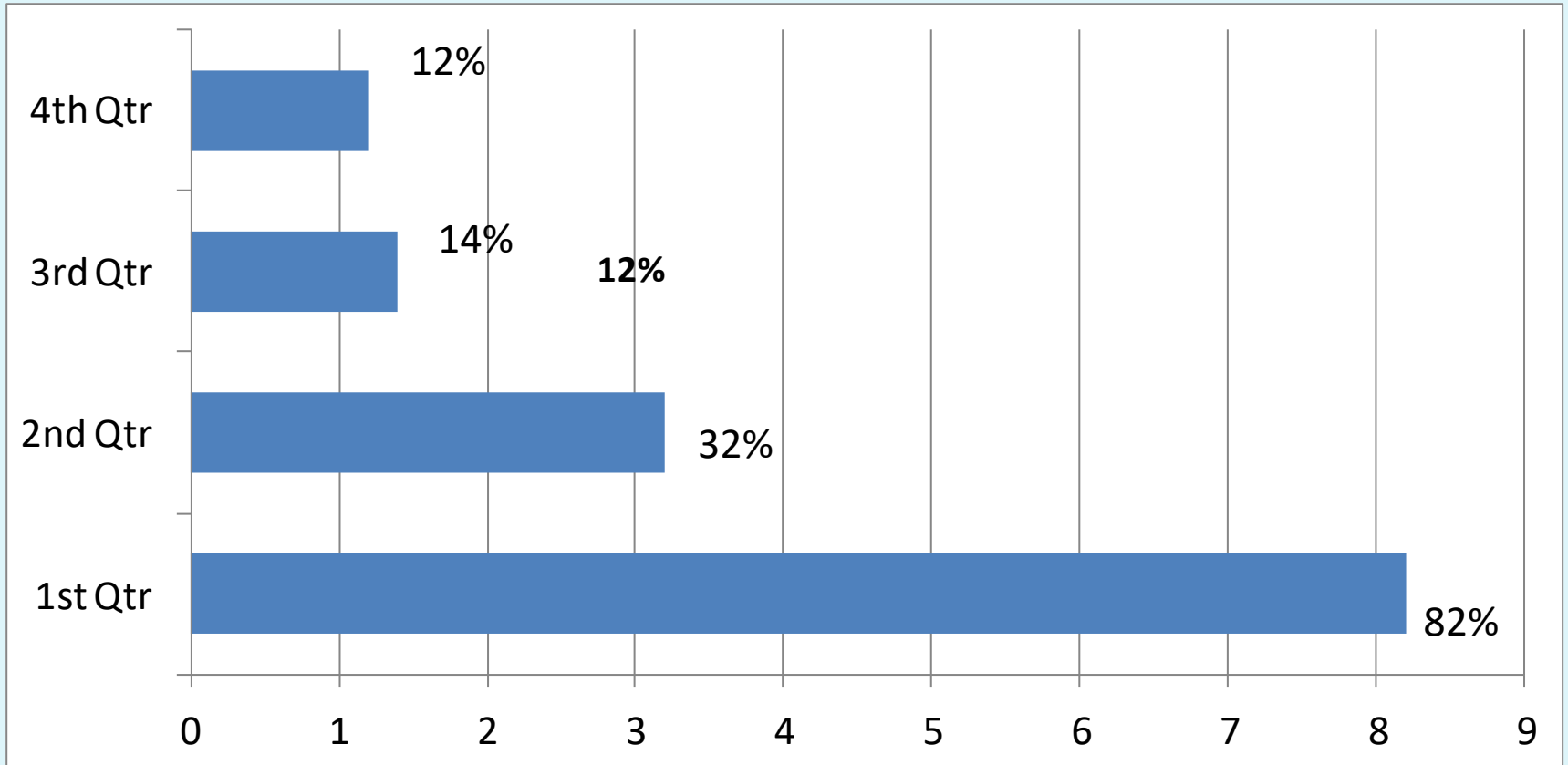
Patients Satisfaction with their Health	Total
(1) Very Dissatisfied	11
(2) Dissatisfied	78
(3) neither satisfied nor dissatisfied	98
(4) Satisfied	60
(5) Very Satisfied	3
Grand Total	250

Patients Satisfaction with their Health	Duration of DM Groups			Grand Total
	(1) 1-5 Years	(2) 6-10 Years	(3) 11 & Above Year	
(1) Very Dissatisfied	8	3	11	
(2) Dissatisfied	46	22	78	
(3) neither satisfied nor dissatisfied	51	36	98	
(4) Satisfied	37	15	60	
(5) Very Satisfied	1	1	3	
Grand Total	143	74	250	

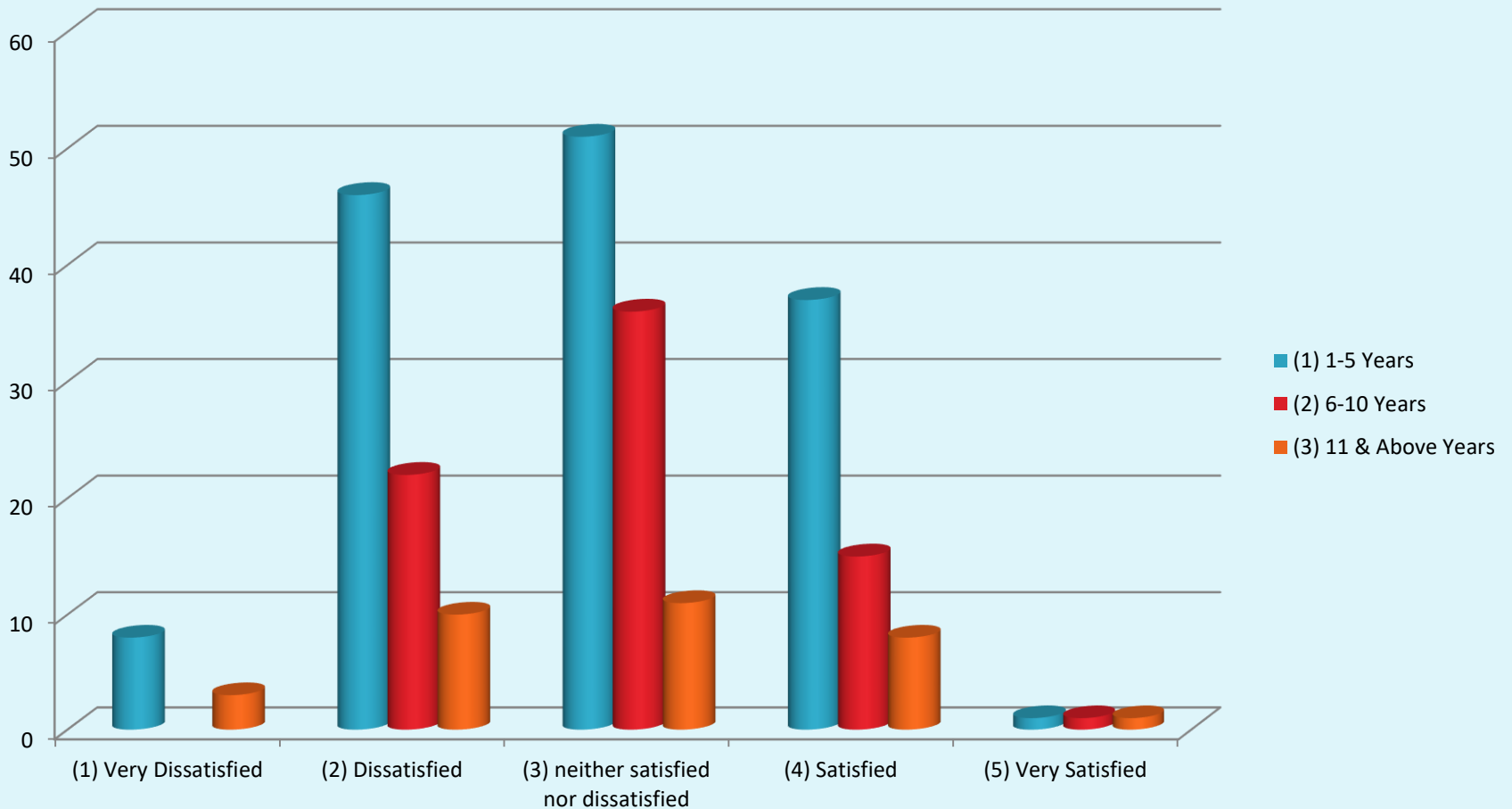
Chai Squira -7.663 at 8 DE P Value = .467 LS = NS

Select destination for new data or choose Paste

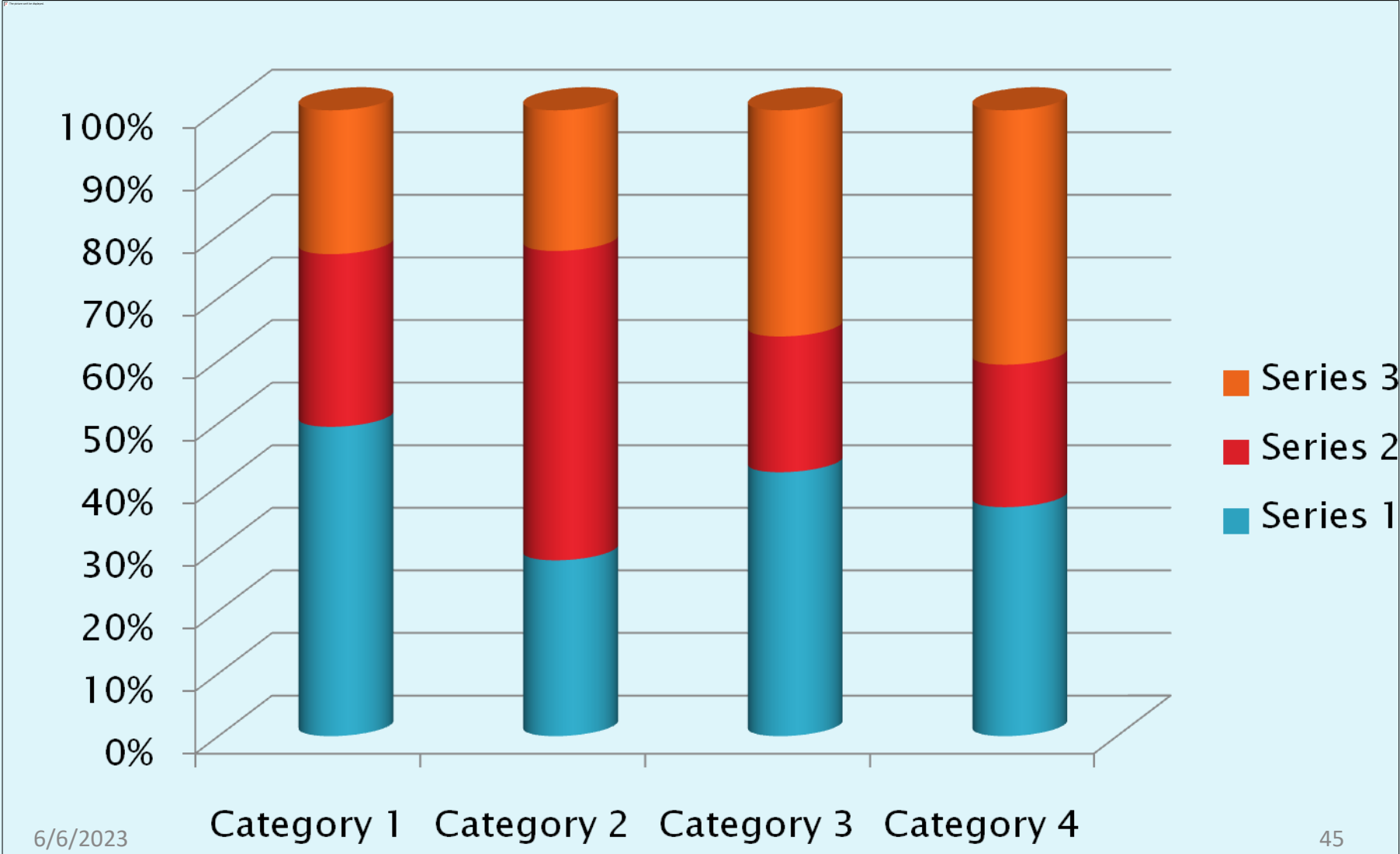
Simple Bar diagram



Multiple Bar diagram

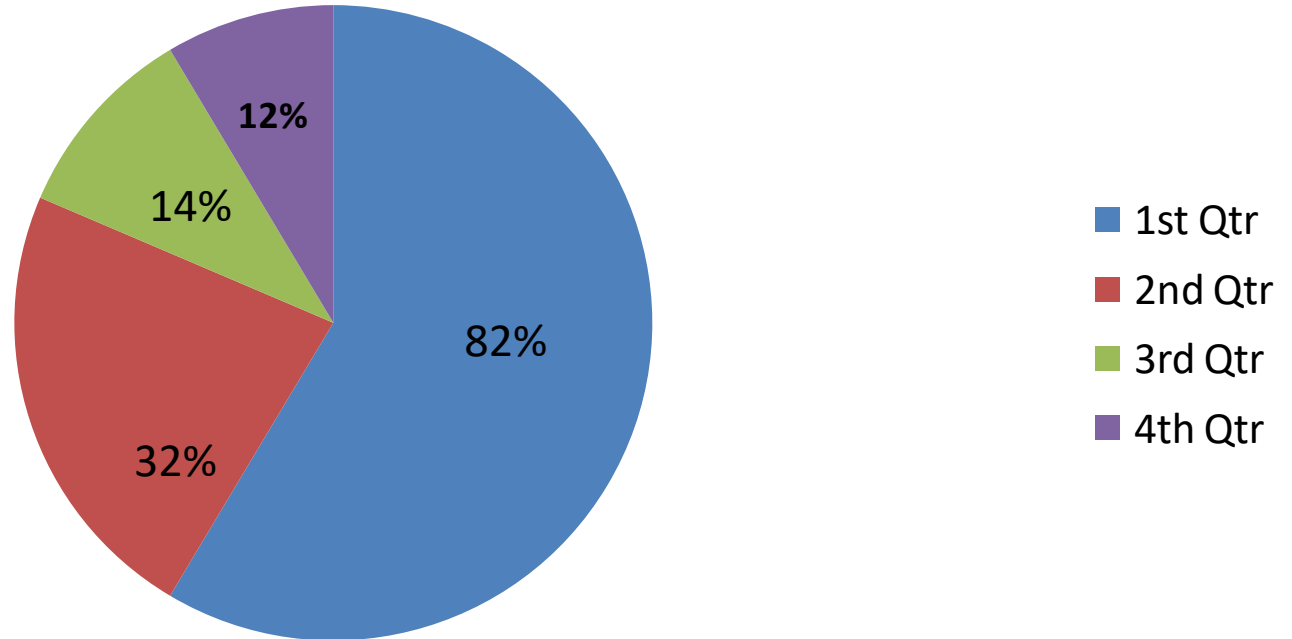


Component Bar diagram

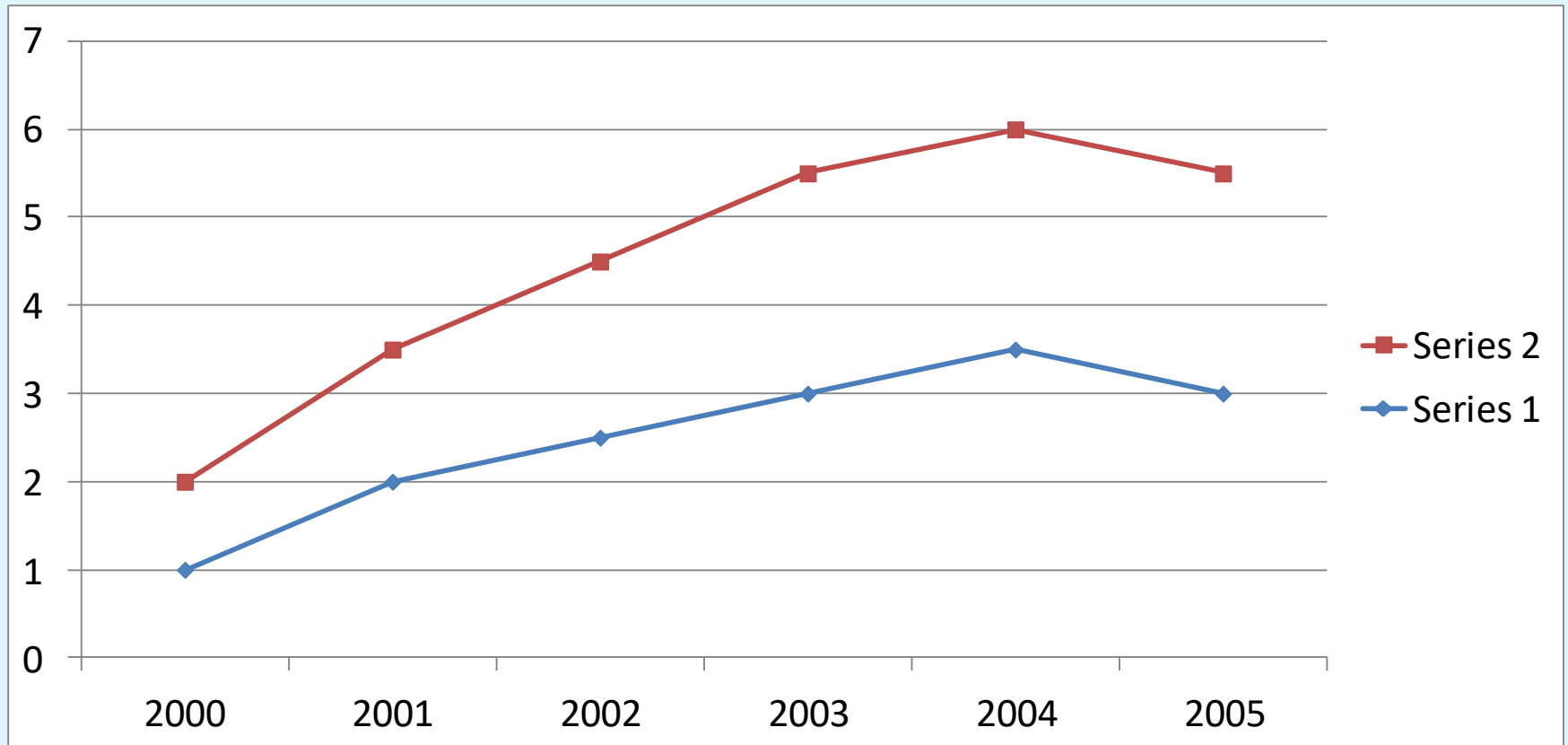


Pie diagram

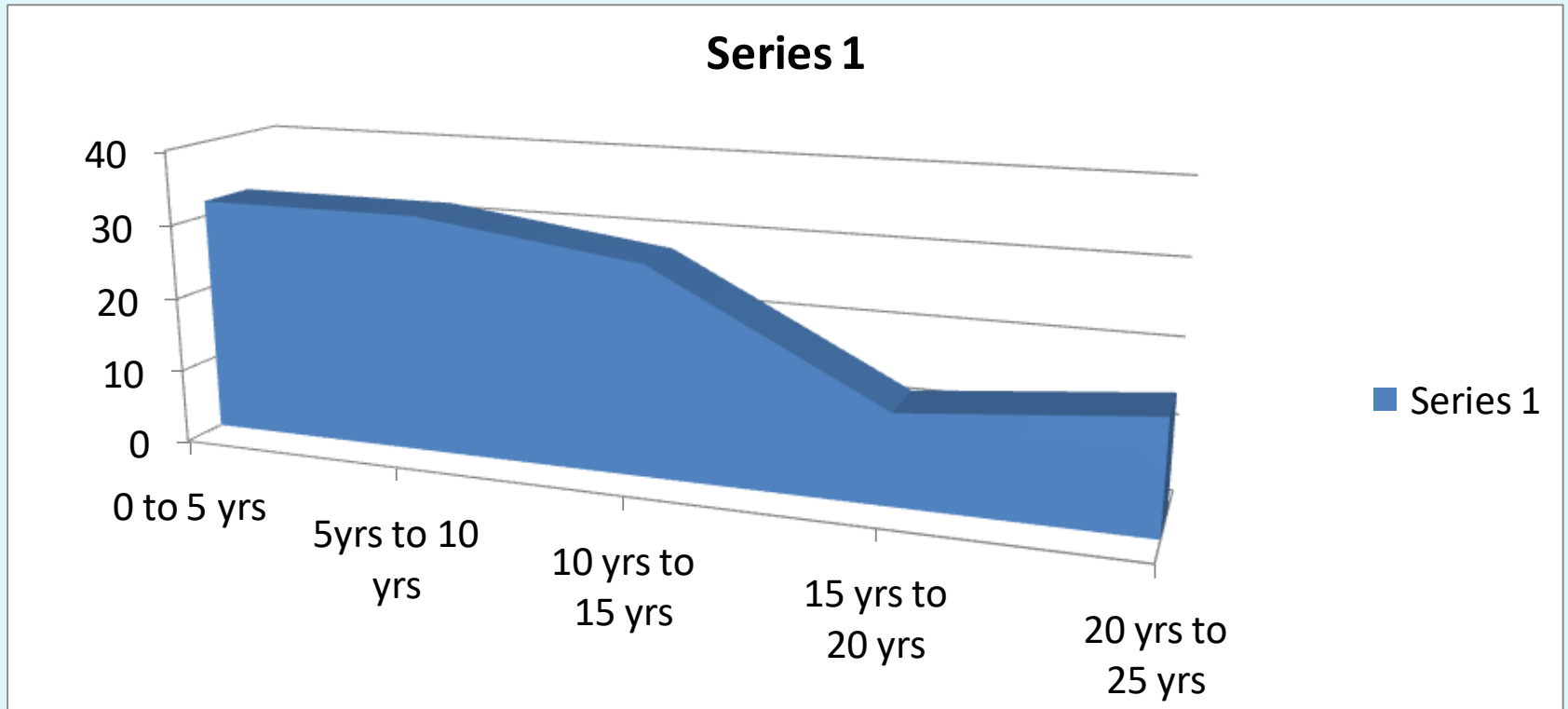
Proportion of Pie = (Proportion of that variable)(360)Degree



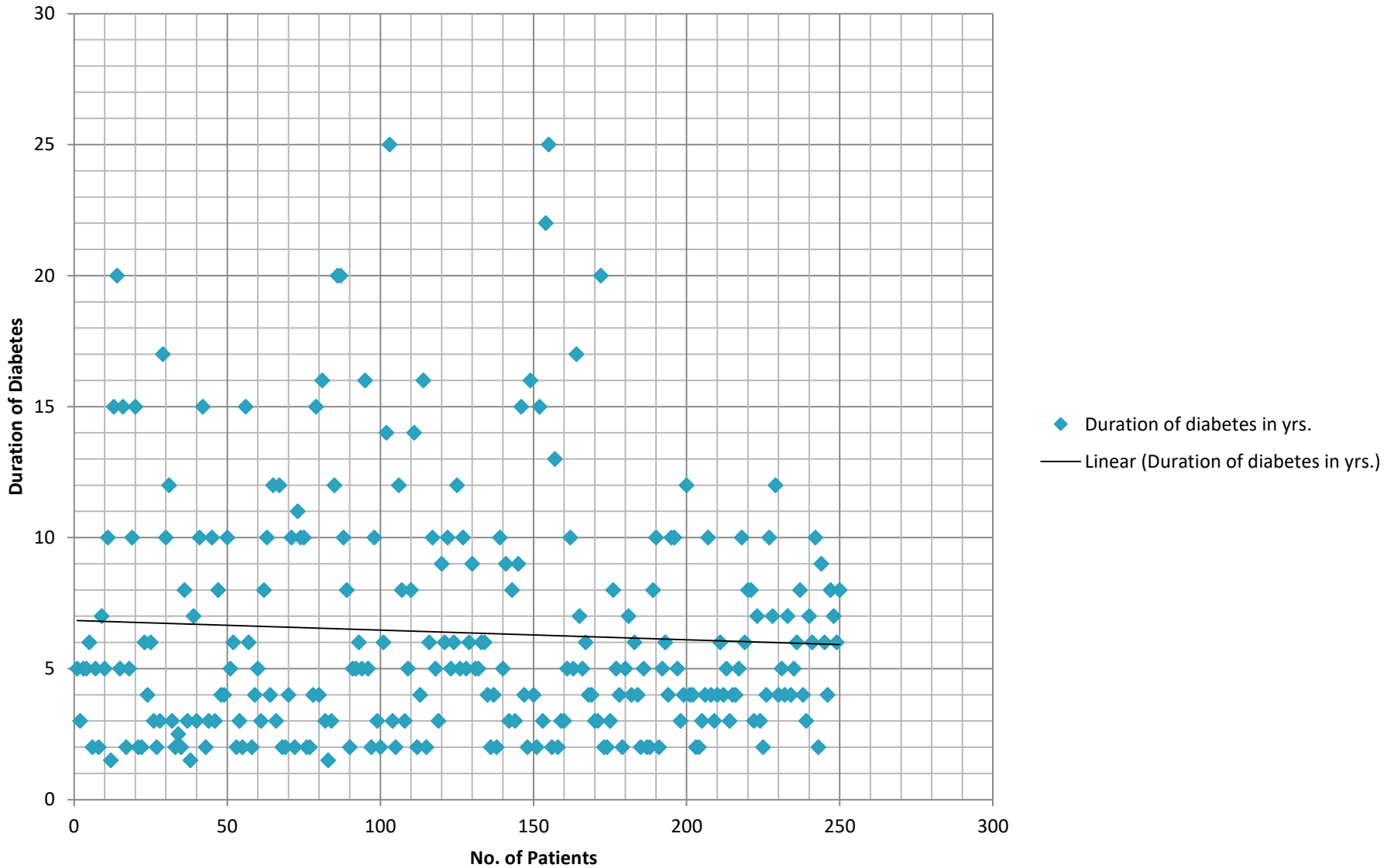
Line diagram



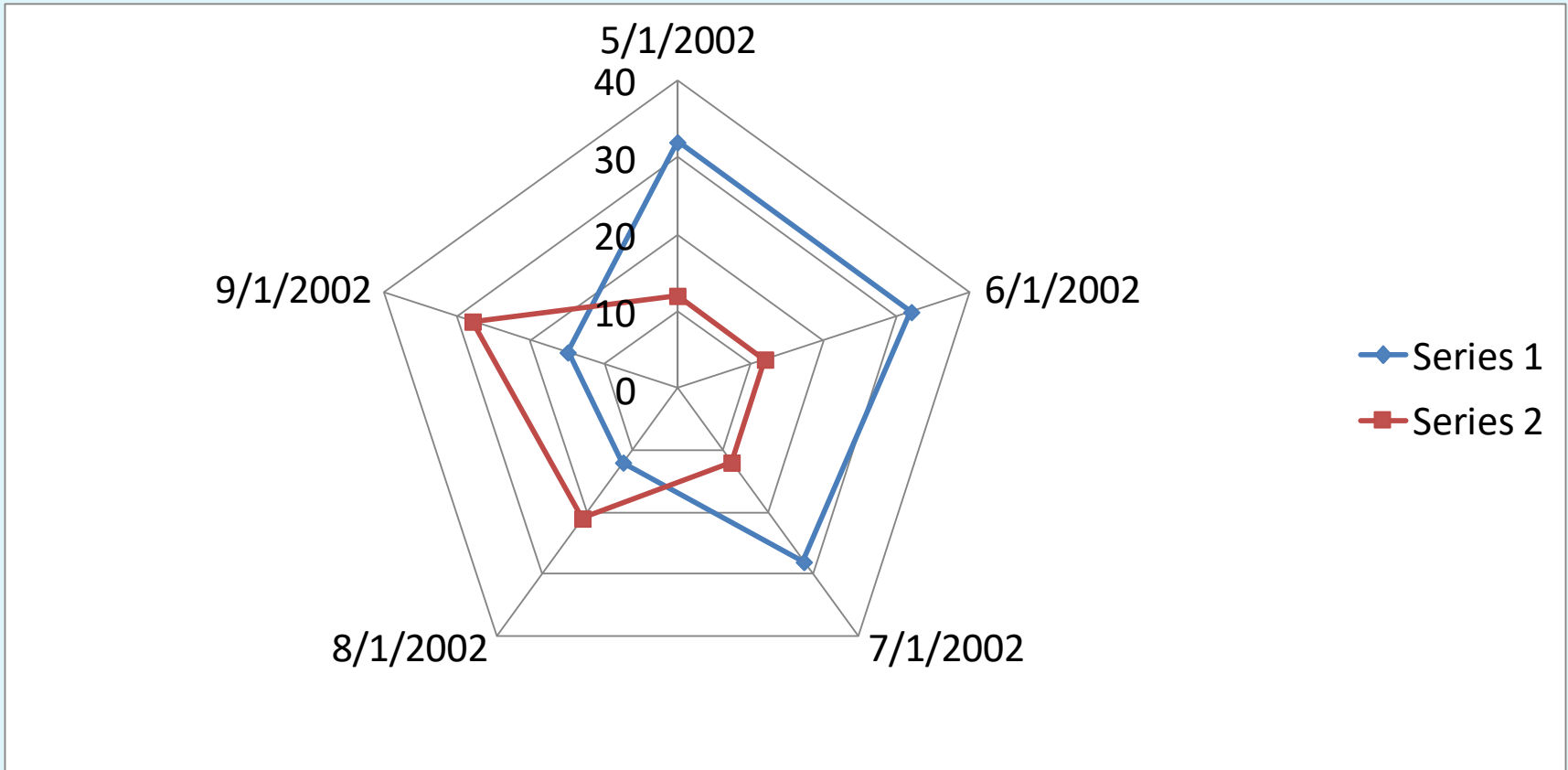
Histogram (Area Diagramme)



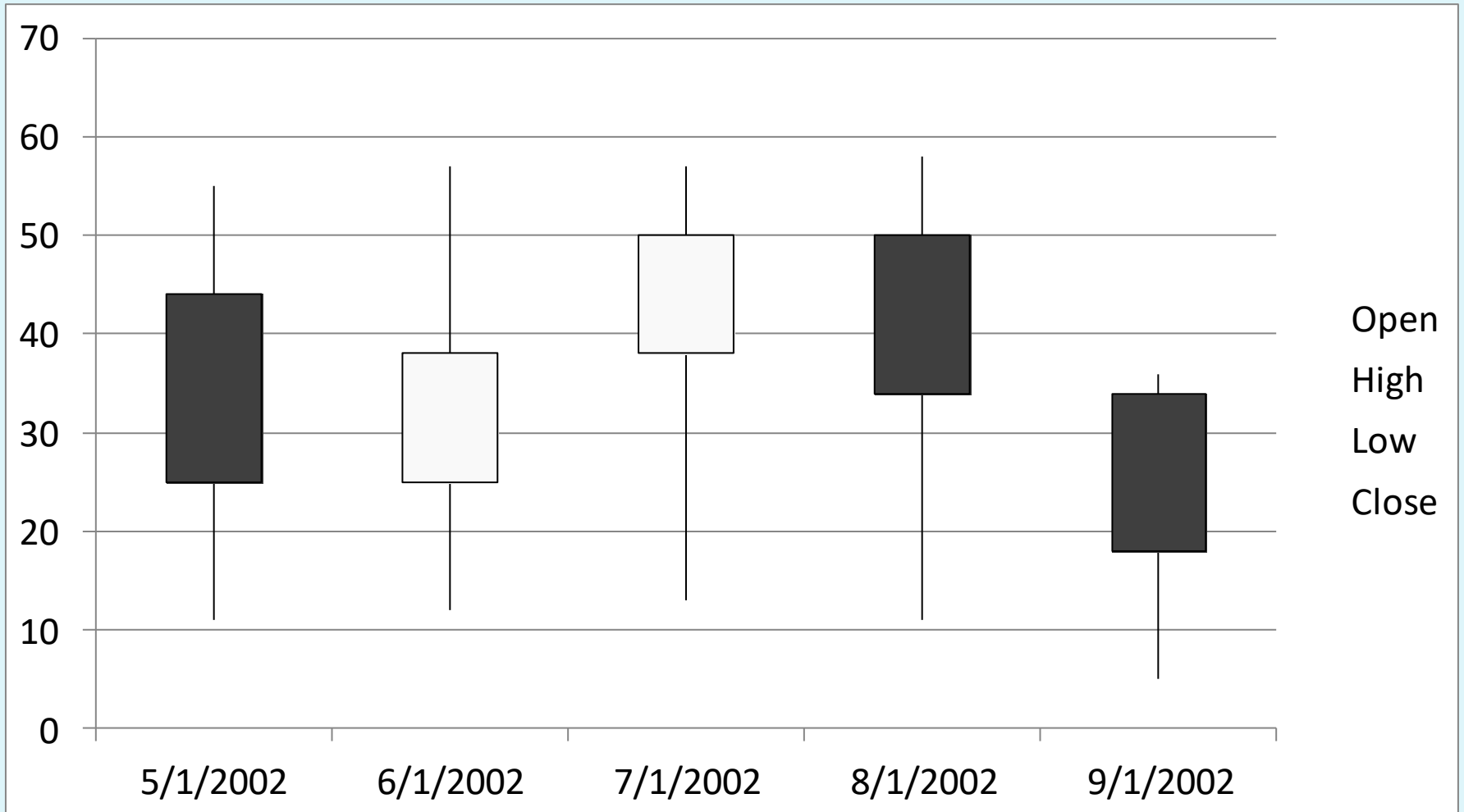
Scatter Diagram



Radar diagram



Box & Whisker



Step-8

Reporting

Steps of Report Writing

- ❖ Title of Project
- ❖ Abstract
- ❖ Introduction
- ❖ Aims & Objectives
- ❖ Methodology
- ❖ Observations-Compilation, Classification & Presentation of data with analysis and inferences
- ❖ Discussion
- ❖ Conclusions
- ❖ Recommendations
- ❖ Limitations
- ❖ Acknowledgment
- ❖ Bibliography

Discussion

- ❖ **Explanation of findings**
- ❖ **Logic and reasoning for the results as it appears**
- ❖ **Compare and contrast with findings of other researchers**
- ❖ **Based on objectives of the study**
- ❖ **Should answer the research question**
- ❖ **Scope & limitations of the study**

Recommendations & conclusions

- **Based on our findings**
- **Limited to objectives of the study**
- **Policy implications**
- **Relevance should be emphasized**
- **Should be exclusively limited to observations**

Managerial and financial aspects

- ❖ **Protocol development**
- ❖ **Time line/Gantt chart**
- ❖ **Peer review**
- ❖ **Development of tools**
- ❖ **Training in data collection**
- ❖ **Budget/ financial accounting**
- ❖ **Quality control**
- ❖ **Monitoring & Evaluation**

Time Line/Gant chart/log Fram



Proposal Vs Report

- **List the key Sections and or Chapters**



Computer in Statistics

Web sites related to Statistics

- <http://stattrek.com>
- <http://vassarstat.net>
- <http://www.scribd.com>
- <http://www.statistixl.com>
- <http://statisticscalculators.com>
- <http://stat.ubc.ca/~rollin/stats/ssize/>



•
.....
.....

Computer Softwares in Statistics

- **Microsoft Excel**
- **SPSS**
- **Epi info**
- **Epi tab**
- **Mini tab**
- **Graph Pad**
- **Primer**
- **Medcal**
-



Important Terms to Remember

For comparison with Gold Standard:

- ❖ Sensitivity
- ❖ Specificity
- ❖ PPV
- ❖ NPV
- ❖ Prevalence
- ❖ Incidence
- ❖ Rate
- ❖ Ratio
- ❖ Risk Ratio

Sensitivity and Specificity

Status based on gold standard test

		Diseased	Normal
Observation in new test	Test positive	True positive a	False positive b
	Test negative	False negative c	True negative d

$$\text{Sensitivity} = a / (a+c)$$

$$\text{PPV} = a / (a+b)$$

$$\text{Specificity} = d / (b+d)$$

$$\text{NPV} = d / (c+d)$$

- **END**

Referencing