Kenya Medical Training College
Department of Clinical Medicine
Year Two Semester One
Data Analysis and Presentation
15th October 2020

Willis J. Opalla

Statistical Data

- Learning Objective

  To analyze, interpret and present statistical data.

Learning Outcomes

- By the end of this session, you should be able to
  1. Explain data processing.
  2. Explain data analysis.
  3. Explain the importance of data analysis.
  4. Explain the different ways of presenting data.
  5. Present, analyze and interpret data using frequency tables, charts and graphs.

Data Analysis and Presentation

- Introduction:
  - Data is a systematic record of numbers, words, measurements, observations or descriptions of an item being studied.
  - When classified according to manner of presentation then data is either
    a) Grouped data or
    b) Ungrouped data.
  - Grouping or ungrouping of data is part of the processing that data undergoes after collection.

Ungrouped and Grouped Data

- Classification based on how data are presented:

- **Ungrouped data** – Data that has not been organized into groups. Also called as raw data.

Data	Frequency
2	8
3	4
5	6
7	7
8	2
9	5

- **Grouped data** - Data that has been organized into groups (into a frequency distribution).

Data	Frequency
2 - 4	5
5 - 7	6
8 - 10	10
11 - 13	8
14 - 16	4
17 - 19	3

Data Processing

- Data are raw facts that do not have much meaning to the user and may include numbers, letters, symbols, sound or images.
- Data processing entails transforming raw data into meaningful output i.e. information.
- Data processing can be done manually using pen and paper or electronically using computers.

Steps for Data Processing

- Data processing entails
  - Editing,
  - Coding,
  - Classification and
  - Tabulation of the collected data.

Step 1: Editing

- Examining the collected raw data to detect errors or omissions and make corrections if possible.
- Involves carefully scrutinizing the completed questionnaires or schedules.
- Done to ensure

1. accuracy of data and consistency with other gathered facts.
2. completeness and uniformity in entering.
3. proper arrangement to facilitate coding and tabulation.

Step 2: Coding

- Assigning numerals or other symbols to responses so that the responses can be put into a limited number of categories or classes.
- The classes should resonate with the research problem being studied.
- A specific response should be placed in only one class or category cell.
- Coding facilitates efficient analysis.
- Codes should be decided on by the time the questionnaire is being developed.

Step 3: Classification

- Arranging data into homogenous groups or classes on the basis of common characteristics.
- The entire data is placed into specific groups that are determined by certain common characteristics.
- Classification Approaches:
  1. Classification according to attributes.
  2. Classification according to class intervals.

Step 4: Tabulation

- Arranging the mass of collected data in some concise and logical order.
- The process of summarizing raw data and displaying it in a statistical table for further analysis.
- An orderly arrangement of data in columns and rows.
- Tabulation can be done manually or by use of electronic devices.

Importance of Tabulation

1. It conserves space and reduces explanatory and descriptive statements to a minimum.
2. It facilitates the process of comparison.
3. It facilitates the summation of items and detection of errors and omissions.
4. It provides a basis for various statistical computations.

Principles of Tabulation

1. Every table should have a clear, concise title.
2. Every table should be assigned a distinct number to facilitate reference.
3. The columns are known as *captions* whereas rows are known as *stubs*.
4. Headings of captions and stubs should be clear and brief.
5. The units of measurement under each heading or subheading (e.g. years for age) must be indicated.

Data Analysis

- Data analysis implies computation of measures and scrutiny of the data to search for patterns of relationship that may exist among data groups.
- In analysis, relationships or differences supporting or conflicting with the original or new hypothesis are subjected to *statistical tests of significance* to determine with what validity, conclusions can be made regarding the data.

The Statistical Tests

- Statistical tests help make decisions on the basis of observed patterns in the collected data.
- Choice of a statistical test depends on the structure of data, the distribution of the data and type of variable.
- The statistical tests include:
  - t-test
  - Z-test
  - chi-square test
  - ANOVA test, binomial test, one sample median test etc.

Importance of Data Analysis

- Data analysis organizes, structures, summarizes and presents the data in forms that make it easy to interpret into useful information**.**
- Therefore data analysis involves
  1. Organization of data
  2. Structuring of data
  3. Summarization of data
  4. Presentation of data
  5. Interpretation of data into information.

Importance of Data Analysis Cont..

- Once data is collected and sorted using analytical tools, the results are interpreted to make decisions.
- The end results are presented as a summarized text, or as a visual in the form of a chart or graph.
- Data presentation in a text, tables and graphs

1. makes an article easy to understand,
2. attracts and sustains the interest of readers,
3. efficiently presents large amounts of complex information.

Data Presentation Methods

- For raw data to be effectively informative, it may be presented in any of the three ways:

1. Textual Method– a narrative description of the data gathered.
2. Tabular Method– a display of information in columns (captions) and rows (stubs).
3. Graphical Method– an illustration of the data in forms of graphs or charts.

Data Presentation Methods

- Text is the principal method for explaining findings, outlining trends and providing contextual information.
- A table is best suited for representing individual information and represents both quantitative and qualitative information.
- A graph displays data at a glance, facilitates comparison and can reveal trends and relationships within the data, e.g. changes over time, frequency distribution and correlation.

Specific Data Presentation Methods

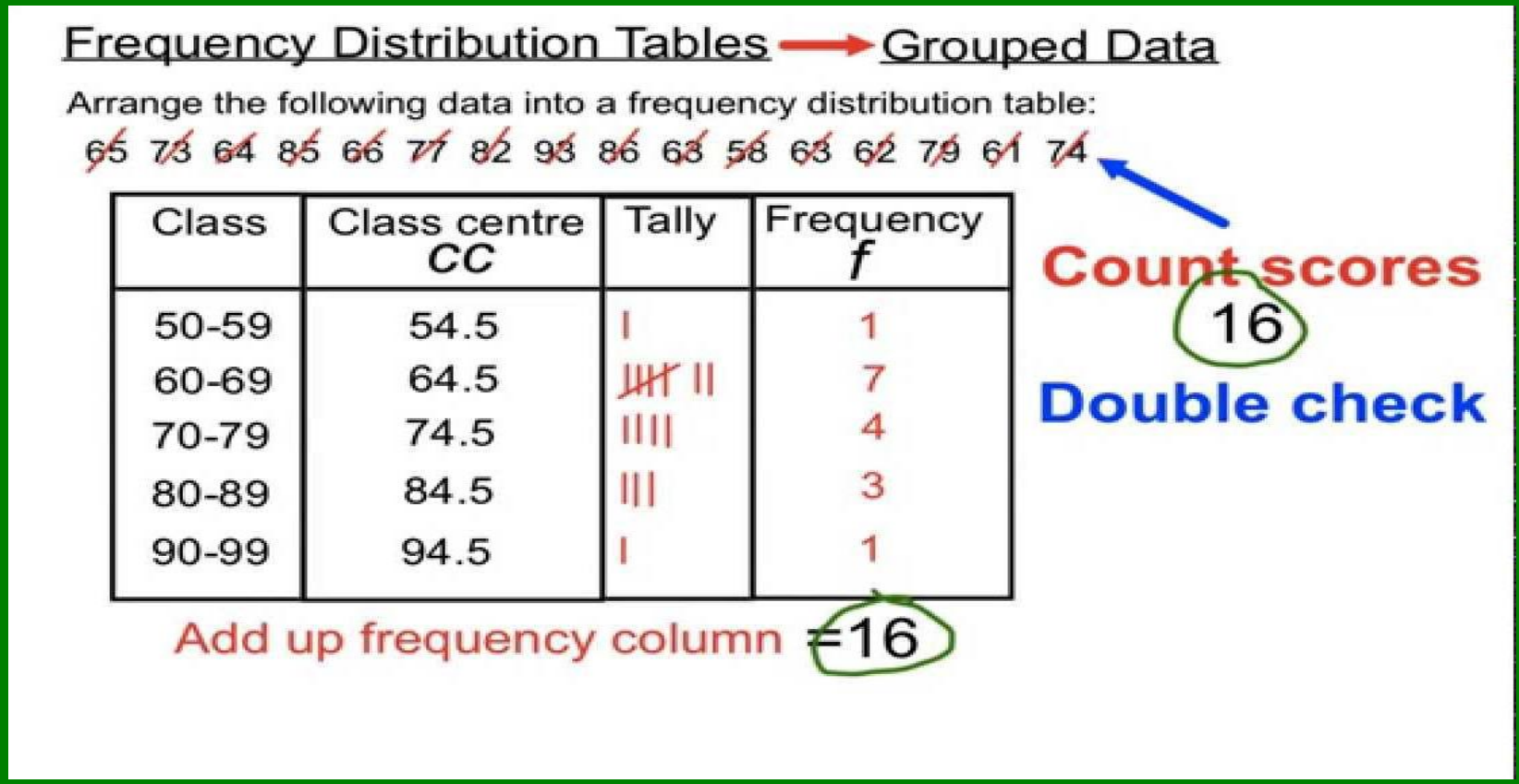
1. Frequency Distribution Table
2. Histogram
3. Polygon
4. Bar graph
5. Pie chart
6. Box and Whisker plots
7. Stem and Leaf

Frequency Distribution Table

- Family planning methods preferred by some 405 women.

Method	Number
Abstinence	14
Condoms	47
Injectables	1
Norplant	1
Pill	35
None	307
Total	**405**

Frequency Distribution Table



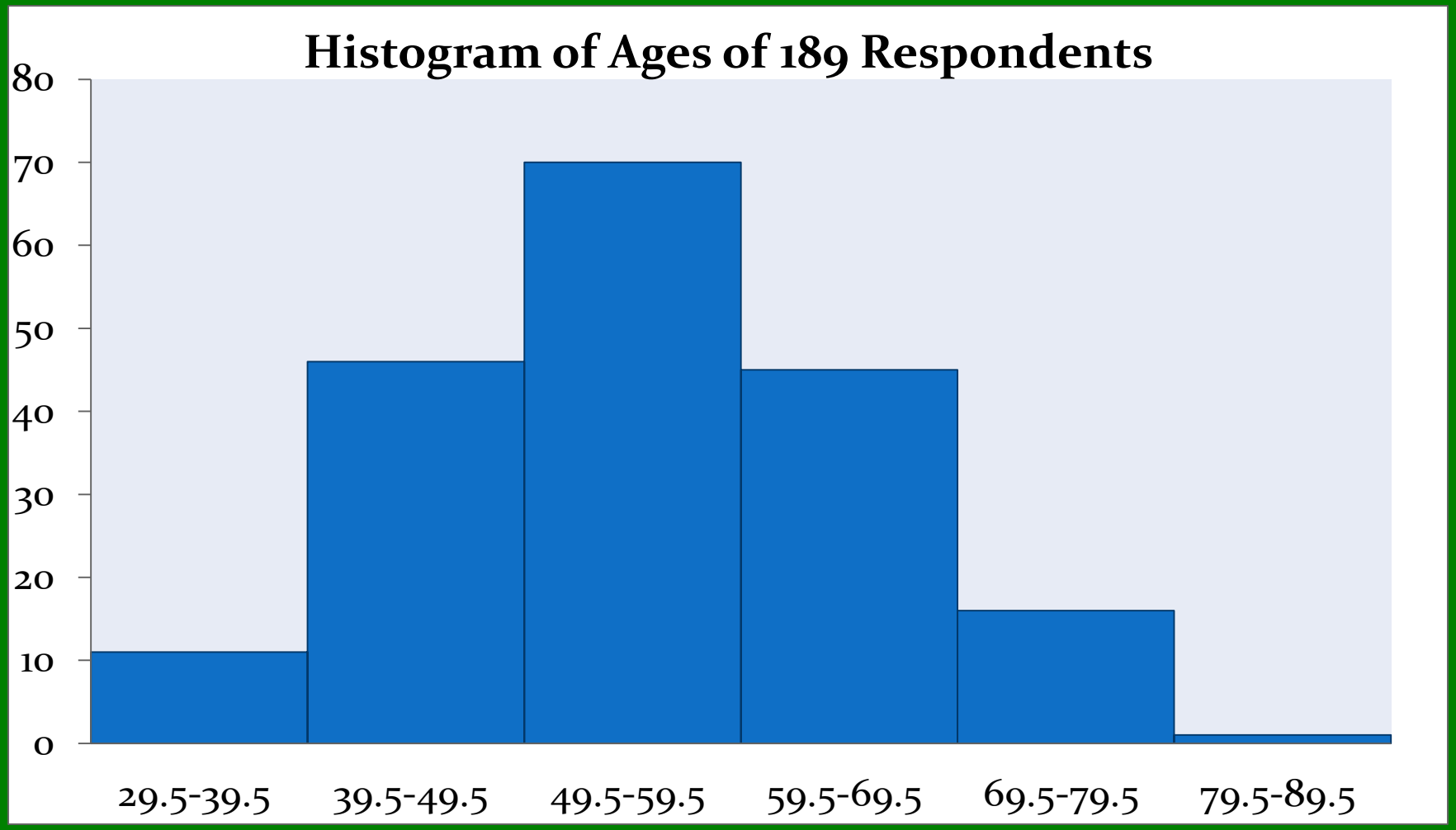
Class	Class centre *CC*	Tally	Frequency *f*
50-59	54.5	I	1
60-69	64.5	𝍸 II	7
70-79	74.5	IIII	4
80-89	84.5	III	3
90-99	94.5	I	1

Histogram

- The ages of 189 respondents may be displayed graphically in the form of *histogram*, which is a special type of bar graph

True Class Limits	Frequency
29.5-39.5	11
39.5-49.5	46
49.5-59.5	70
59.5-69.5	45
69.5-79.5	16
79.5-89.5	1
Total	189

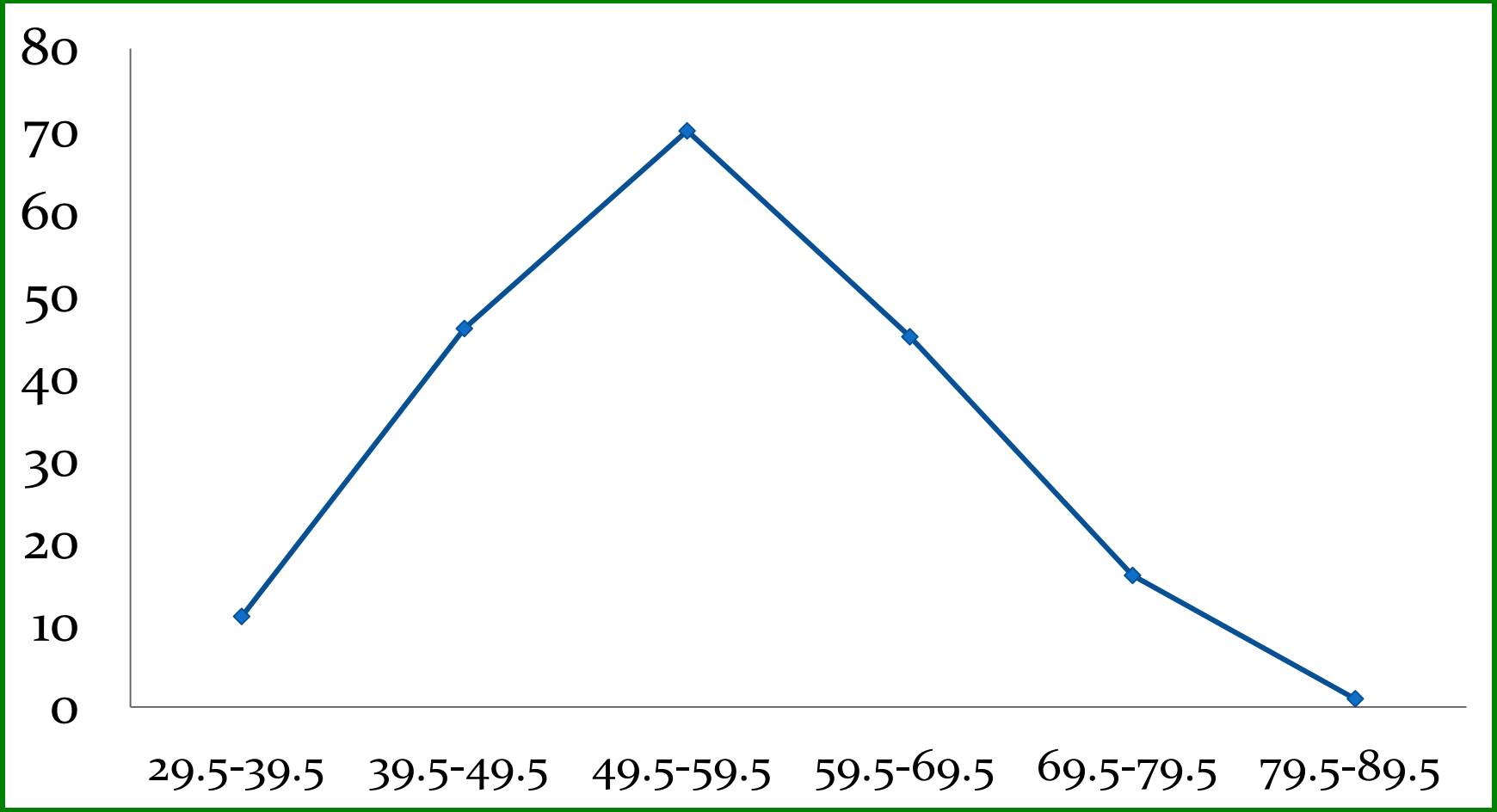
Histogram


Histogram of Ages of 189 Respondents
80
70
60
50
40
30
20
10
0
29.5-39.5
39.5-49.5
49.5-59.5
59.5-69.5
69.5-79.5
79.5-89.5

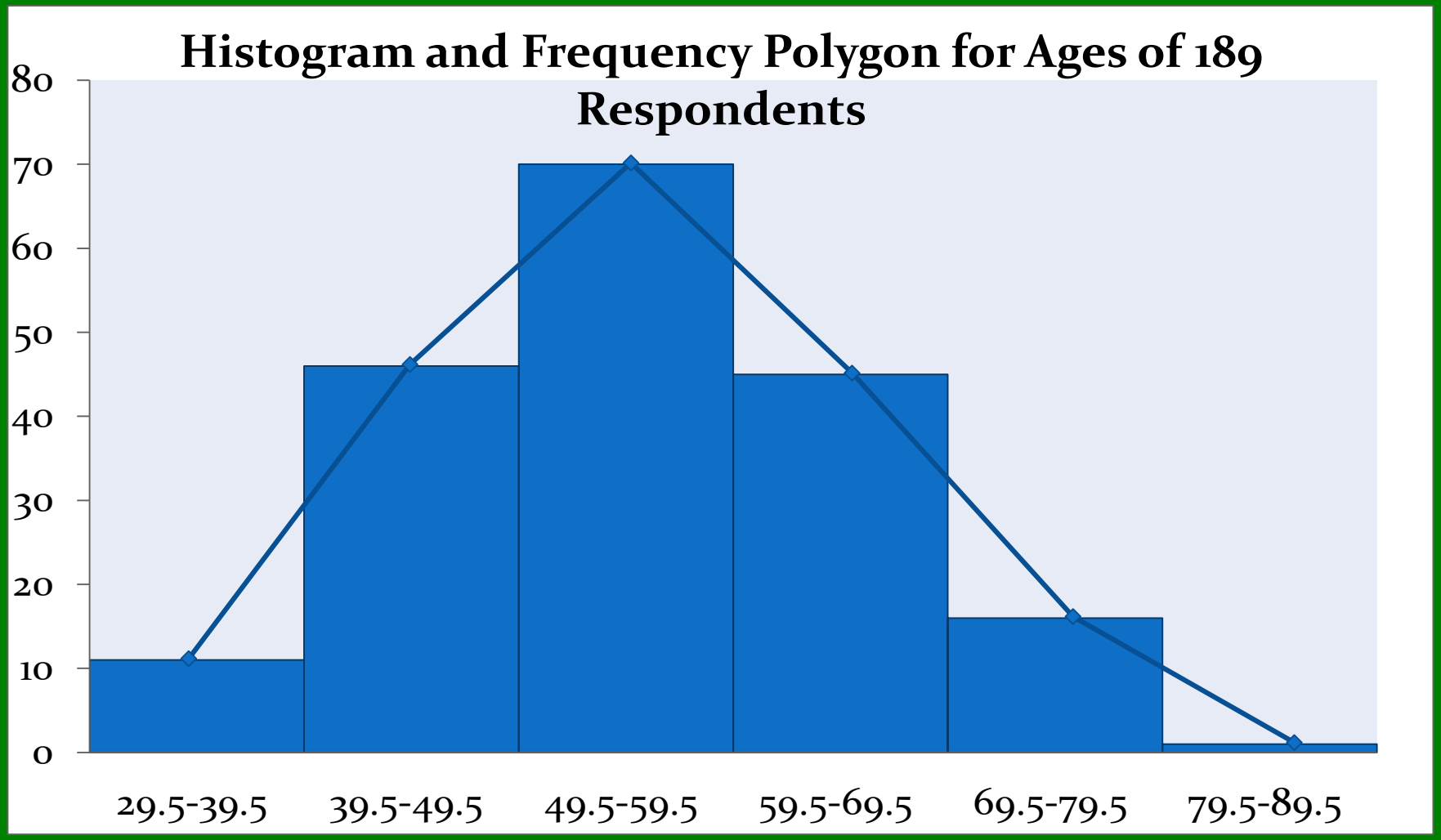
Frequency Polygon

1. A frequency distribution can be diplayed graphically in another different way with a *frequency polygon*, which is a special kind of a graph.
2. To draw a polygon, a dot is first placed on the midpoint of each class interval represented on the horizontal axis of a graph.

Frequency Polygon for 189 Respondents



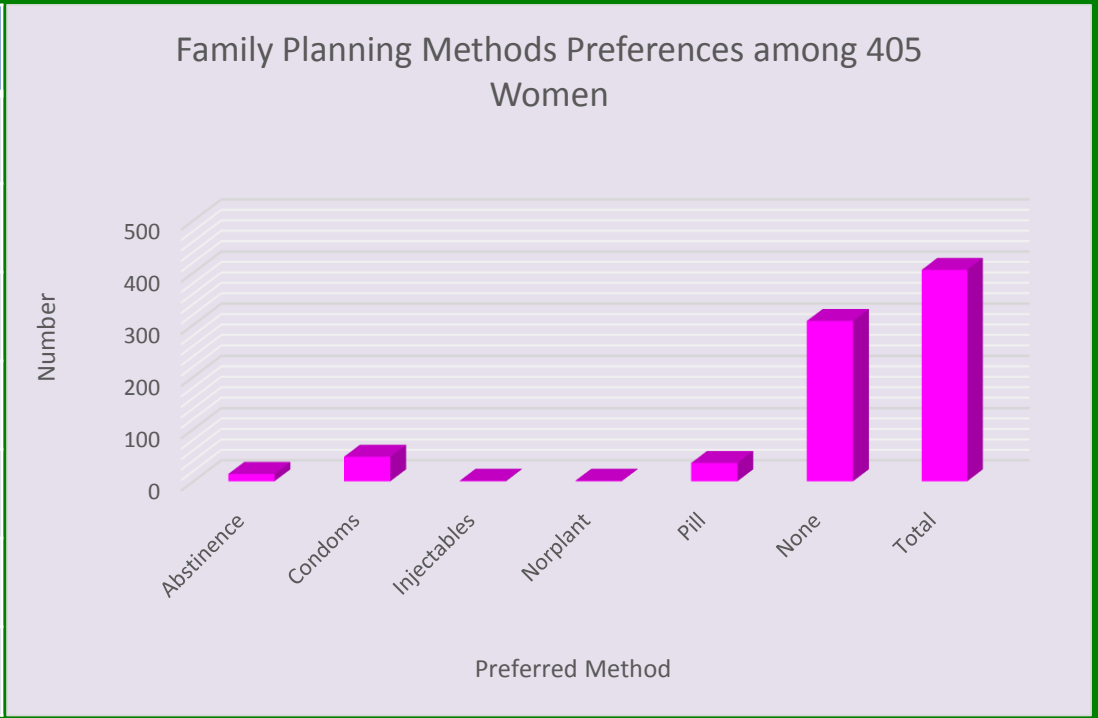
Histogram and Frequency Polygon



Bar Graph

- Family Planning Methods Preferences of some 405 women.

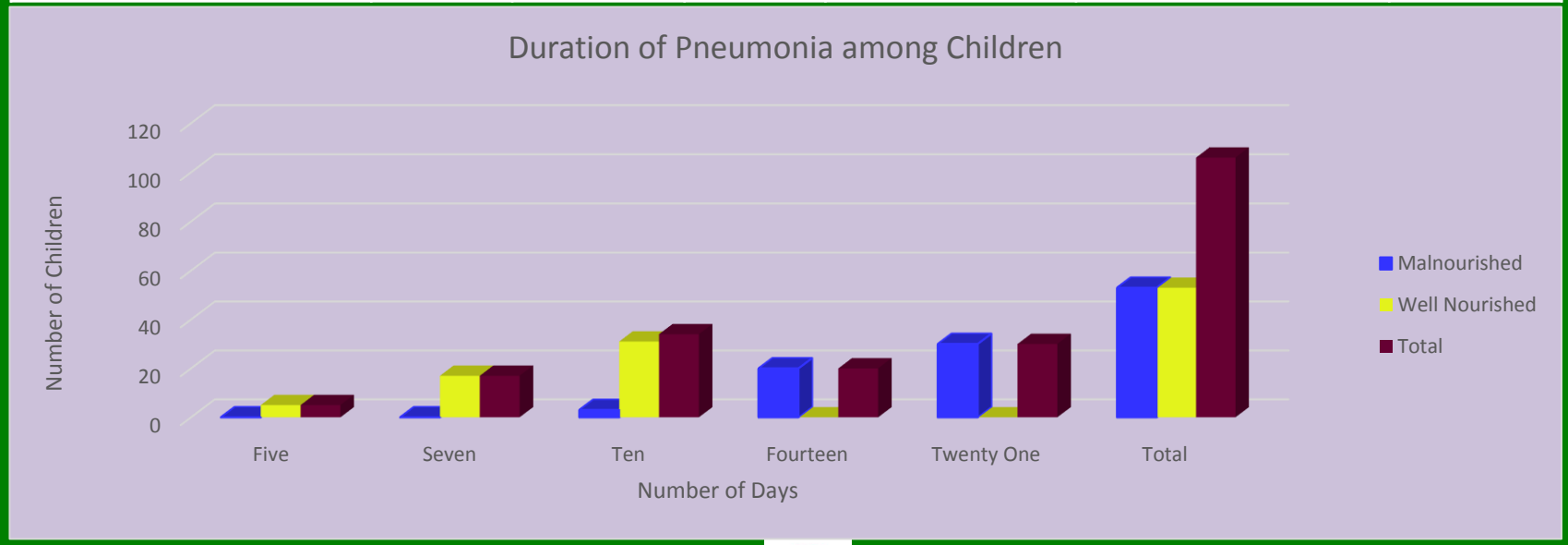
Method	Number
Abstinence	14
Condoms	47
Injectables	1
Norplant	1
Pill	35
None	307
Total	**405**



Bar Graph

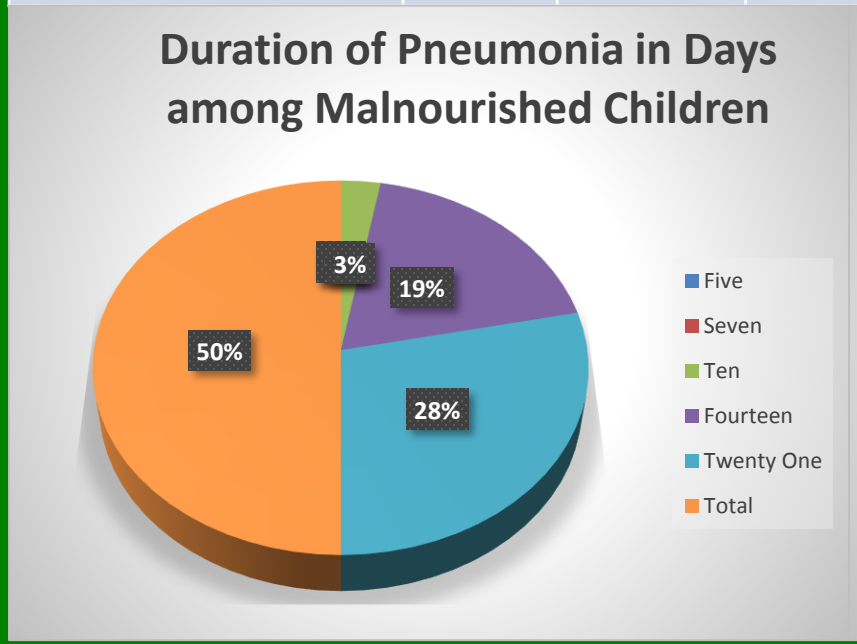
- Length of pneumonia in days among 106 children of varying nutritional status.

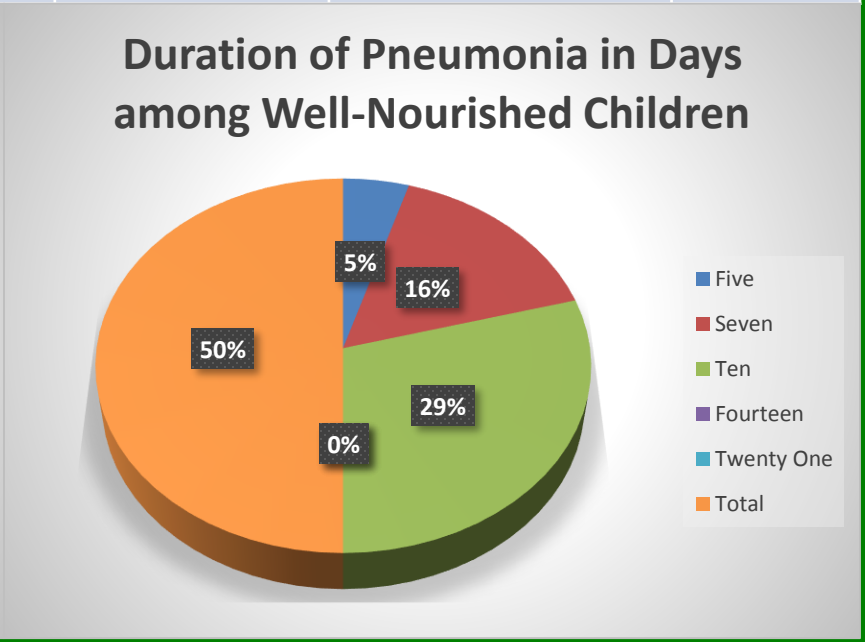
	Five	Seven	Ten	Fourteen	Twenty One	Total
Malnourished	0	0	3	20	30	53
Well-nourished	5	17	31	0	0	53
Total	5	17	34	20	30	106



Pie Chart

	Five	Seven	Ten	Fourteen	Twenty One	Total
Malnourished	0	0	3	20	30	53
Well-nourished	5	17	31	0	0	53
Total	**5**	**17**	**34**	**20**	**30**	**106**

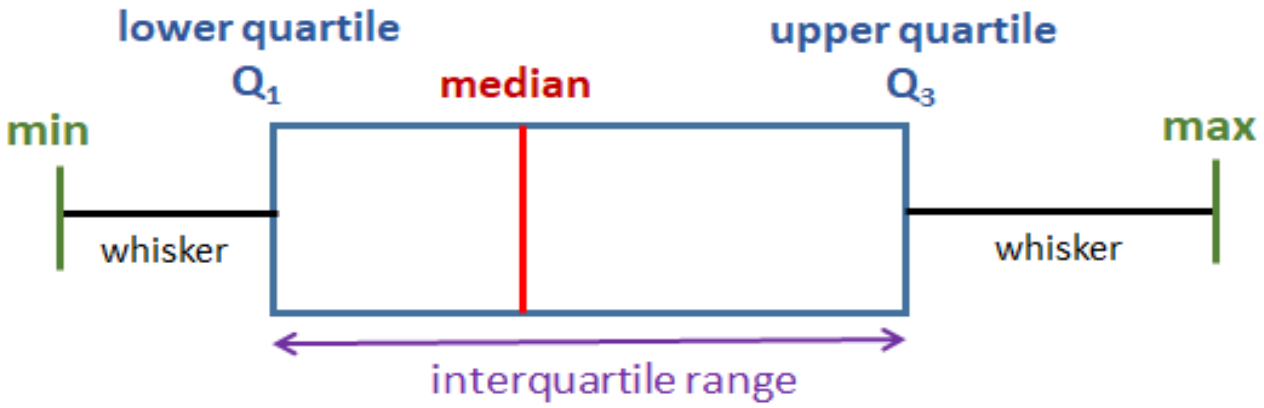




Box and Whisker Plots

Box and Whisker Plot

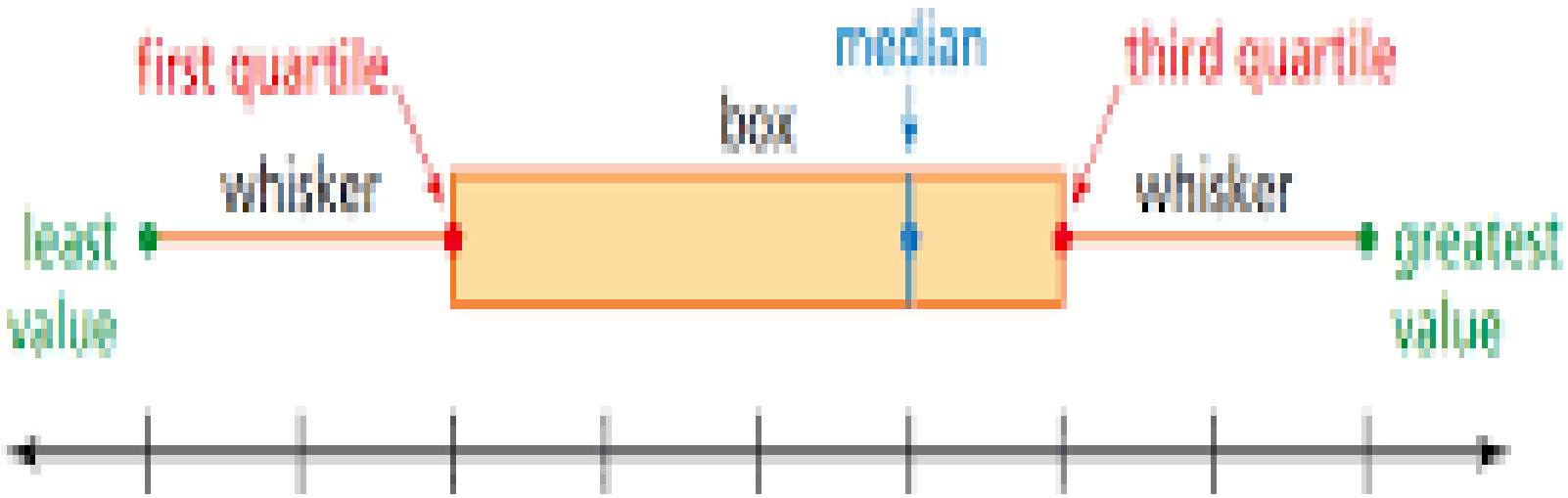
A box and whisker plot (also called a box plot) shows the five-number summary of a set of data: **minimum**, **lower quartile**, **median**, **upper quartile**, and **maximum**.



Box and Whisker Plots

- A box and whisker plot displays a data set along a number line using medians.
- Quartiles divide the data set into 4 equal parts.
- The median (2^{nd} Quartile) divides the data set into 2 halves.
- The median of the lower half is the 1^{st} quartile.
- The median of the upper half is the third quartile.

Box and Whisker Plots



The five numbers that make up the box-and-whisker plot are called the **five-number summary** of the data set.

Stem and Leaf Display

- A properly constructed stem and leaf display
  - provides information regarding the range of the data set,
  - shows the location of the highest concentration of measurements, and
  - reveals the presence or absence of symmetry.

Stem and Leaf Display

- An advantage of the stem leaf over the histogram is the fact that it preserves the information contained in the individual measurements.
- Such information is lost when measurement is assigned to the class intervals of a histogram.

Worked Example: Stem and Leaf Display

- In the display,

1. The stem is the column of the unique elements of the data after the last digit has been removed e.g in the data series containing the numbers 57, 38, 24, 79, the stems are 5, 3, 2 and 7. The stems are arranged in a column in the ascending numerical order.

Worked Example: Stem and Leaf Display

2. The final digits are the leaves of each column and are placed in a row next to the appropriate column and sorted in an ascending numerical order. e.g. for 57, 38, 24, 79, the leaves are 4, 7, 8, 9.

The stem-and-leaf plot therefore involves splitting each individual data into two.

Stem and Leaf Display

- Stem and Leaf Display for a data set of

30, 34, 35, 37, 37, 38, 38, 38, 38. 40, 40. 42. 42, 43, 43, 43, 43, 43, 43, 44, 44, 44, 44, 44, 44, etc…… and 82.

Stem	Leaf
3	045778888
4	0022333333444444555666667777778888899999
5	000111222233333333444444555666677777888899999
6	000111111111222222233444444444455666678899
7	011111111111111111123567888
8	2

Stem and Leaf Display

- Stem-and-Leaf displays are most effective with relatively small data sets.
- As a rule they are not suitable for use in annual reports or other communications to the general public.
- They are primarily of value in helping researchers and decision makers understand the nature of their data.
- Histograms are more appropriate for externally circulated publications.

Stem and Leaf Display: Exercise

- Which of these two is the correct stem and leaf display for the following data set? 5, 5, 6, 7, 7, 8, 9, 9, 11, 23, 420.

A.

Stem	Leaf
0	5 5 6 7 7 8 9 9
1	1
2	3
42	0

B.

Stem	Leaf
0	5 5 6 7 7 8 9 9
1	1
2	3
4	20

Assignment

Ungrouped Data	Grouped Data	
	Score	Frequency
96 67 28 32 66 65 69 33 98 96 76 49 52 64 76 83 92 93 68 52 79 81 30 68 69 83 86 43 45 39 83 75 66	20-29	3
	30-29	14
96 92 75 83 76 83 85 62 37 65 63 44 62 31 36 38 42 39 83 87 56 58 88 89 93 42 53 69 90 55 66 49 52		
42 32 38 42 40 40 42 89 65 73 81 83 59 82 75 82 86 23 35 76 83 85 83 92 75 89 66 91 83 34 36 27 90 69		
	Total	**100**

Assignment

- Complete the table in slide 40 by grouping the ungrouped data into class intervals as shown in the first and second classes. Proceed to fill in all the empty spaces in the column for frequency.

Summary

- Data analysis involves scrutiny of data to search for patterns of relationship or differences in order to determine the validity of conclusions made.
- Once raw data is processed, it may be presented through textual, tabular or graphical method.

References

- In, J and Lee, S. (2017) Statistical Data Presentation, *Korean Journal of Anesthesiology*. 70(3):pp 267–276.

- Kothari, C. R., (2004) *Research Methodology, Methods and Techniques*, 2nd ed., New Age International Publishers, New Delhi.