Chapter 3

Graphical and Tabular Statistical Data Summarizing Tools

3

Graphical and Tabular Statistical Data Summarizing Tools

Felix Kutsanedzie[1]; Sylvester Achio[1]; Selassie Kwasi Diaba[2]

[1]Accra Polytechnic, Accra, Ghana
[2]Anglican University College of Technology, Nkorazan, Sunyani–Ghana

Abstract

Data collected in every experiment or survey that has been conducted must be summarized using the appropriate statistical tool for the purpose of making a good presentation of the findings by the researcher. This when rightly done would ensure that the right interpretation is given to the data for the information to be communicated easily by the researcher to his targeted audience in the scientific community. This chapter seeks to explain the various tools used in summarizing qualitative and quantitative data collected during an experiment or a survey study.

Keywords

Tables, Qualitative, Quantitative, Data, Graphs

3.1 Concepts

Concepts are abstractions which are mental representations of real and agreed upon phenomena based on experience. Examples of mental representations of real phenomena include buildings, cars, animals; and those of agreed-upon phenomena – marriage, justice, trust, faith etc.

3.2 Constructs

Constructs are also abstractions which are theoretical creations of observations which cannot be seen directly or indirectly. Examples include intelligence, leisure, taste, etc. The constructs are almost the same as concepts which are agreed-upon phenomena.

Since concepts and constructs are abstractions or ideas that are within the mind of the researcher, they remain nebulous or vague to others until they are reduced to measurable units referred to as variables before others can appreciate and understand them the same way as the researcher.

3.3 Measurement and Variables

Measurement is a symbol assigned to a phenomenon or variable to give / denote / represent / describe it with a value. Thus the measurable label of a construct or a concept or phenomenon is referred to as a variable. When a variable is measured, its value is obtained. In terms of measurements, all things that can be measured are classified into three:

- Direct observables: all items that can be seen – length, volume, weight etc.

- Indirect observables: data from questionnaires - age, gender, income, marital status, religion etc.

- Constructs: variables that based on theoretical creations which cannot be seen directly or indirectly – leisure, pleasure, taste, love, etc.

To enable variables to be measured, there must be *theoretical definitions of measurement of the variable,* i.e. measurement of the variable as defined in books or in literature; and *operational definition of measurement of the variable*, i.e. how the researcher intents to measure the variable based on the work being carried out.

It should also be noted that concepts or constructs (phenomena) must be free to vary if they are to be measured as *variables* or else are regarded as *constants.*

There are four ways by which all variables can be measured. These include *"metering" (using devices), counting, ordering (ranking or referencing) and classifying (grouping without ranking).*

With all research studies excluding the descriptive ones, there are always two categories of variables that must be identified – *independent variables (IV)* and the *dependent variables (DV).* Once these two variables are established in a research, one can posit a cause-effect relationship. The cause – effect relationship implies that a change in the dependent variable causes a change in the independent variables. However if all the variables in the research are all independent variables, one cannot posit this cause-effect relationship in the study, but the other variables outside the ambit of the research may depend on the independent variables. Independent variables are classified into Active and Attribute independent variables. The Active Independent variables are those whose levels are controlled or manipulated by the researcher whereas the Attribute Independent variables are those whose levels are controlled by the subjects involved in the research (natural characteristics such as sex, age, race etc.). There are also Intervening variables (extraneous variables) which involve uncontrolled variables that may account for changes in the dependent Variables (DV). Controlled variables are variables that are controlled and statistically measured and accounted for in order not to affect the DV.

3.4 Levels of Measurement

Having understood what variables are and the different types, it becomes necessary to know the levels of measurement that can be applied to them to enable the use of the appropriate statistical tools for their measurements. There are basically four levels of measurement into which every variable falls. These include nominal (classificatory), ordinal (ranking), interval, and ratio which together are represented by the acronym NOIR.

Nominal Level (classificatory scale)

This level of measurement applies to all variables that are mutually exclusive and collectively exhaustive – which implies that it puts all items into one and only one group at a time and once an item is captured in a group it is excluded from belonging to another group at the same time and thus the chance of being found in a different group at that same time exhausted. It is primarily used for classifying hence the name classificatory scale). Assuming gender is being considered as a variable, all items are to be considered or grouped into either a male or female and not both (hermaphrodite) at the same time. Again considering a variable like religion – Christianity, Islam, pagan; race – African, Asian, American etc.; marital status – single, married; body colour- dark, white, brown, yellow etc. It should however be noted the nominal level of measurement is used in only classifying and not ranking and they make no mathematical sense or cannot be subjected to any mathematical operation to make sense. All variables that this level of measurement applies to are considered qualitative except when there is the issue of coding where they are given codes in order to be counted.

Ordinal (Ranking Scale) Level of Measurement

The ordinal levels of measurement of variables are similar to those of the nominal or classificatory scale but the distinctive feature that distinguishes the two is the fact that the ordinal level involves ranking or ordering of items under consideration. Example is grading of students into: first class, second class, and third class. It thus not only put students into classes but also ranks them. More examples can be given as taste of a sumptuous meal – palatable, more palatable, most palatable, classifying types of shoes in terms of their quality – good, better, best. Range of Students Marks such as 31 – 40, 41 – 50, 51 – 60. The ordinal level of measurement also makes no mathematical sense or cannot be subjected to any mathematical operations but rather used in ordering items. Where ordinal levels of measurement are applied to a variable, the data collected are mostly qualitative but can be quantitative when values of variables are coded.

Interval Level of Measurement

This is applied to variables that are ordered based on equal intervals. They make mathematical sense or can be subjected to mathematical operations such as addition and subtraction. It has no absolute zero value. Which means that it has no true zero value and any zero value assign to a variable based on the interval level of measurement is arbitrary. For example, the fact that a Celsius thermometer is used to measure the temperature of a body to be 0 ℃ does not imply that there body has no heat in it but rather the instrument used for measuring the temperature could not pick up the latent heat (hidden) of the body. Therefore should a different thermometer scale such as the Kelvin or the thermodynamic scale be used the value would not be recorded. It thus implies that the zero (0) value recorded on the Celsius scale is arbitrary or not the true zero. However every unit on the on Celsius thermometer is equal in value and the differences between temperatures are equal and mathematically meaningful irrespective of the scale used - (20 ℃ - 15 ℃) = (293.15K – 288.15K) = 20. Thus examples of variables that the interval level of measurement can be subjected to include temperature, intelligence quotient (IQ) etc. Interval measurements can assume negavtive values.

Ratio Level of Measurement

This is applied to all variables that can be subjected to all mathematical operations such as addition, subtraction, multiplication, division etc. to make sense. They all have true zero values or true absolute values. It has equal units and can also be ordered. Examples include weight, height, amount of money, distance, volume etc.

3.5 Problems Associated with Data Collected

When variables are measured or responses are obtained – that is to say data is collected. It must be noted that there are problems associated with data collected in research studies with regards to their reliability and validity. The data collected in a study can be classified into qualitative and quantitative. However since qualitative data is mostly classificatory, validity and reliability is less of concern to qualitative data. The same cannot be said of quantitative data.

Validity of data refers to the extent or the degree to which an instrument used in measuring or collecting data on a concept reduced to variable would actually measure to exactness. It actually refers to the accuracy of the measurement. Thus validity of data depends on whether one is measuring the intended concept rightly or using the right instrument. If the right instrument is not being used, then the data can not be validated.

Reliability of data refers to the measure of variations in data collected on a particular concept (reduced to variable) over a repeated number of trials. It is thus synonymous to consistency of the collected data. Reliability can be systematic – a change in a subject value overtime or repeated trials or random – which results from sampling error.

3.6 Measure of Reliability of Collected Data

Retest Correlation

To measure reliability of the data collected depends on the data type. The retest correlation involves using the data collected over repeated trials for the same variable (concept) and then finding the correlation between the first collected data and the subsequent. If the correlation value is 1, it means it is a perfect correlation hence the more consistent the data obtained, if the correlation reduces towards zero, it means the data is less consistent.

Kappa coefficient

It is used to measure the reliability of nominal variables over a repeated collection of data. It values range from – 1 to 1.

For the measurement of the reliability of variables of average scores of two or more items on multi questionnaire or inventory, and the alpha reliability is used.

Measurement error

During a study or an experiment, errors emanate in measurement due to sampling, subjects or experimenter or researcher effects as well as on occasions where there is reduced reliability and validity. Errors due to sampling arise when the right sampling techniques are not deployed. There are also errors due to the subjects or experimental materials and researcher or experimenter effects which results in bias. Last but not the least; errors arise when there is reduced reliability and validity.

3.7 Tools for Summarizing Data Collected

Table 3.1 *Data Summarizing Tools.*

			Frequency distribution
			Relative frequency distribution
			Percent frequency distribution
		Tabular Method	*Crosstabulation*
	Qualitative Data		
		Graphical Method	*Bar graph*
			Pie chart
Data			
			Frequency distribution
			Relative frequency distribution
			Cumulative frequency distribution
			Crosstabulation
			Cumulative Relative frequency distribution
			Stem-and-leaf display
		Tabular Method	
	Quantitative Data		
		Graphical Method	*Dot plot*
			Histogram
			Ogive
			Scatter
			Diagram

Qualitative Data Summarizing Tools

The data collected from research is basically classified into twofold: qualitative and quantitative data. For both types of data, the researcher must be able to identify the type of data collected – whether qualitative or quantitative before the right statistical tools can be deployed for summarizing and analyzing them.

3.8 Tabular Method

Frequency Distribution

The frequency table is used for summarizing both qualitative and quantitative data. It is a descriptive statistical tool used in counting the number of times a particular data (responses) is obtained for a measured variable. Applications like Excel and SPSS can be used to find the frequency distribution of data collected in order to summarize it. Taking for example data collected on the colours of different cars observed on a road during a defined period as follows: green, green, red, green, yellow, white, black, white, yellow and red. The frequency distribution table for the data above is shown below:

Table 3.2 *Frequency Table.*

Colour of Cars	**Frequency(f)**
Yellow	2
Red	2
White	2
Green	3
Black	1
Total	$\sum f = 10$

3.9 Relative Frequency Distribution

The relative frequency distribution table is a type of frequency distribution which is expressed as unity or where the total frequency is equal to 1. Thus it is the ratio of the frequency of each data categories to the total frequency.

Taking the same data on the colour of cars, the relative frequency table for the data is drawn below:

Table 3.3 *Relative Frequency Table.*

Colour of Cars	Frequency(f)	Relative frequency (Rfq)
Yellow	2	$^2/_{10} = {}^1/_5 = 0.2$
Red	2	$^2/_{10} = {}^1/_5 = 0.2$
White	2	$^2/_{10} = {}^1/_5 = 0.2$
Green	3	$^3/_{10} = 0.3$
Black	1	$^1/_{10} = 0.1$
Total	∑f =10	∑Rfq = 1

Percent Frequency Distribution

This is a frequency distribution table in which the frequencies of responses or data collected are expressed as percentage. Using the same data on the colours of cars, the percentage frequency table is shown below:

Table 3.4 *Percentage Frequency Table.*

Colour of Cars	Frequency(f)	Percent frequency (%fq)
Yellow	2	$^2/_{10}\, x\ 100 = {}^1/_5 = 20$
Red	2	$^2/_{10}\, x\ 100 = {}^1/_5 = 20$
White	2	$^2/_{10}\, x 100 = {}^1/_5 = 20$
Green	3	$^3/_{10}\, x\ 100 = 30$
Black	1	$^1/_{10}\, x\ 100 = 10$
Total	∑f =10	∑%fq = 100

Crosstabulation

Cross tabulation is a tabular summary of data on two variables. The data on one variable is put in the column and other in a row. It helps to describe the relationship between the variables. Assuming one needs to draw a crosstabulation on Mechanical Engineering students in Accra Polytechnic with the following data: first year class – 70 good, 30 poor; second year class – 50 good, 40 poor; and third year class – 10 good, 90 poor. A cross tabulation of this data can be done as shown below:

Table 3.5 *Crosstabulation of Data on two variables.*

	Mechanical Engineering Students		
Performance Rating	**Year 1**	**Year 2**	**Year 3**
Good	70	50	10
Poor	30	40	90
Total	100	90	100

3.10 Qualitative Data

Graphical Method

Bar Graph

For summarizing of qualitative data, the bar graph or chart is used instead of the histogram. This is because the frequency of the data is represented on the vertical axis and on the horizontal axis the label of data items whose frequencies are indicated on the vertical axis. The items on the horizontal axis are descriptive in nature because the axis gives no indication of measurement in terms of counting or quantifying but rather classifying. This implies that the tool is used purely or purposely for describing a data set and not to establish quantity. It is distinguished from histogram by the space left between the various drawn bars. Hence it is not appropriate tool for summarizing quantitative data. A bar chart of the data on colour of cars is represented below:

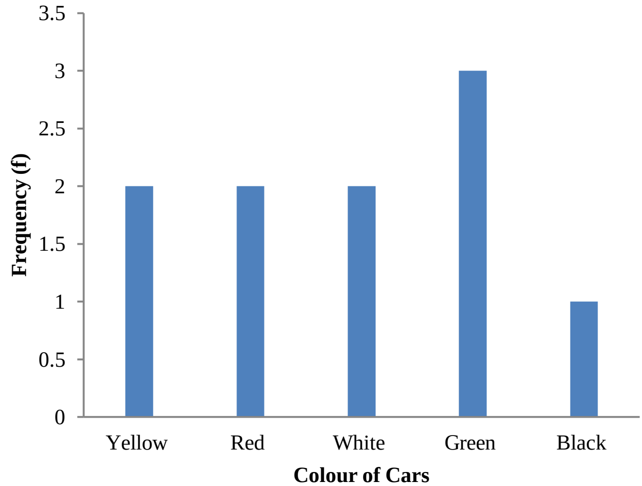


Figure 3.1 *Colour of Cars Observed on a Road within a Period.*

Pie Chart

The pie chart is used for summarizing qualitative. It involves the frequency of the data in a circle - distributing the data according the total angle distribution within a circle which is 360°. It can be expressed in the degrees or degrees subsequently converted into percentage as seen in the related table. The data on the colour of cars can be represented with a pie chart as shown below:

Table 3.6 *Table on the conversion of frequencies into degrees for drawing of a pie chart.*

Colour of Cars	**Frequency**	**Expressing data in degrees and percentage**	
Yellow	2	$\frac{2}{10}\ x\ 360° = 72°,$	$\frac{72°}{360°}\ x\ 100 = 20\%$
Red	2	$\frac{2}{10}\ x\ 360° = 72°,$	$\frac{72°}{360°}\ x\ 100 = 20\%$
White	2	$\frac{2}{10}\ x\ 360° = 72°,$	$\frac{72°}{360°}\ x\ 100 = 20\%$
Green	3	$\frac{3}{10}\ x\ 360° = 108°,$	$\frac{108°}{360°}\ x\ 100 = 30\%$
Black	1	$\frac{1}{10}\ x\ 360° = 36°,$	$\frac{36°}{360°}\ x\ 100 = 10\%$
Total	10	360°	100%

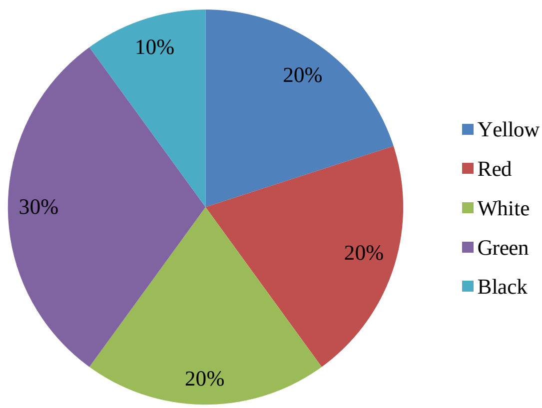


Figure 3.2 *Pie Chart.*

Quantitative Data Summarizing Tools

Tabular Tools

Frequency Distribution

With summarizing quantitative data, the frequency distribution is one of the tools that can be used. Taking for example count of cars that have covered measured distances within a specific time. The frequency distributions – the frequency tables such as relative frequency table and the percentage frequency table can be drawn as follows:

Table 3.7.1 *Frequency distribution table.*

Distances Covered	Number of cars (frequency)
10	2
20	2
30	2
40	3
50	1
	$\sum f = 10$

Table 3.7.2 *Relative frequency distribution table.*

Distances Covered	Number of cars (frequency)	Relative frequency
10	2	$\frac{2}{10} = 0.2$
20	2	$\frac{2}{10} = 0.2$
30	2	$\frac{2}{10} = 0.2$
40	3	$\frac{3}{10} = 0.3$
50	1	$\frac{1}{10} = 0.1$
	$\sum f = 10$	

Table 3.7.3 *Percent frequency distribution table.*

Distances Covered	Number of cars (frequency)	Percent frequency
10	2	$\frac{2}{10}\ x\ 100 = 20$
20	2	$\frac{2}{10}\ x\ 100 = 20$
30	2	$\frac{2}{10}\ x\ 100 = 20$
40	3	$\frac{3}{10}\ x\ 100 = 30$
50	1	$\frac{1}{10}\ x\ 100 = 10$
	$\sum f = 10$	

Cumulative frequency distributions

The cumulative frequency distribution such as cumulative relative frequency and cumulative percentage frequency are used in summarizing quantitative data. This is because cumulative frequency is found by adding the frequencies preceding classes or group to those succeeding classes or groups. Since it is only quantitative data that can be added to make mathematical sense, it presupposes that cumulative frequency table must be a quantitative tool. The cumulative frequency distribution tables of the data above are drawn as follows:

Table 3.8 *Cumulative frequency distribution table.*

Distance	Frequency	Cumulative frequency	Cumulative relative frequency	Cumulative Percentage frequency
10	2	2	0.2	20
20	2	4	0.4	40
30	2	6	0.6	60
40	3	9	0.9	90
50	1	10	1.0	100
	$\sum f = 10$	$\sum Cf = 10$	$\sum Rf = 1$	$\sum \%f = 100$

Cross tabulation

Cross tabulation can also be used to summarize quantitative data. Taking for instance the data on two car brands, X and Y used to cover a distance with their respective fuel consumptions, a cross table can be drawn below:

Table 3.8.1 *Crosstabulation table on cars and the distances covered.*

Distance Covered / Km	Car Brands and their Fuel Consumption in Km/hr	
	X	Y
10	5	6
20	8	9
30	10	15
40	15	20
50	18	30

3.11 Graphical Tools

3.11.1 Stem and Leaf Display

This is a quantitative data summarizing tool where the data or values of the collected variables are arranged or displayed in a pattern mimicking a stem and it leaves to enable easy summarization of the data for analysis. The values or numbers are separated into stems and leaves, where the 'stems' are first identified for each value, and then the 'leaves' of the value identified and placed close to their respective identified 'stems'. For example, taking the following numbers as values of variables within a data set obtained for a study: 52, 54, 53, 61, 65, 93, 98, 102, 44, 48, 31. The stem and leaf display of this data set can be represented as below:

Stem	Leaf				
3	1				
4	4	8			
5	2	3	4		
6	1	5			
9	8				
10	2				

Figure 3.3 *Stem and leaf display.*

3.11.2 Dot Plot

This is one of the graphical tools used in summarizing quantitative data. It is a graph where dots are used to represent values on top of a single graded or calibrated horizontal line or axis. Taking for instance the data set: 5, 6, 6, 7, 8, 10, 10, 10, 12, 13, 15, 15, 6, 6, 7, 1, 1, 1; the dot plot for such a data set is drawn as fig 3.4.

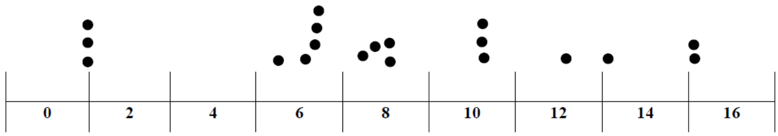


Figure 3.4 *Dot Plot.*

3.11.3 Histogram

The histogram is a tool used for summarizing quantitative data. It looks different from the bar chart in that there are no spaces between its bars and its horizontal and vertical axis are numerical whereas the y –axis of the bar chart always represent the frequency of a data which is numerical but the horizontal axis whether it is numerical or not is only used descriptively. For example data collected on class scores such as: 20, 25, 20, 20, 30, 10, 20, 30, 40, 50, 30 can be represented using the histogram as below:

Table 3.8.2 *Summary of data for drawing of histogram.*

Class Scores	Frequency
10	1
20	4
30	2
40	1
50	1
	∑f = 9

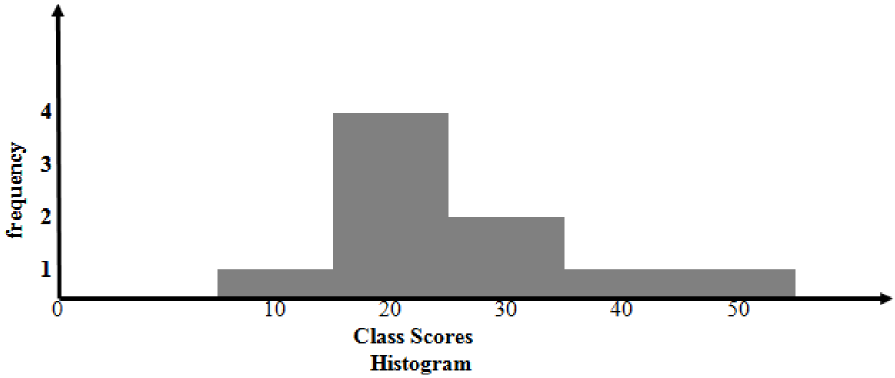


Figure 3.5 *Stem and leaf display.*

3.11.4 Ogive

Ogive is also a tool used for summarizing quantitative data. It is a graph obtained by plotting the cumulative frequency of a data set against its raw scores or marks or values of the data set. The cumulative frequency is always on the y-axis as it depends on the scores or the values of the data set. Taking the data used in the plot of the histogram, the ogive can be drawn as below:

Table 3.8.3 *Table for summarizing scores used for drawing the ogive.*

Class Scores	Frequency
10	1
20	4
30	2
40	1
50	1
	$\sum f = 9$

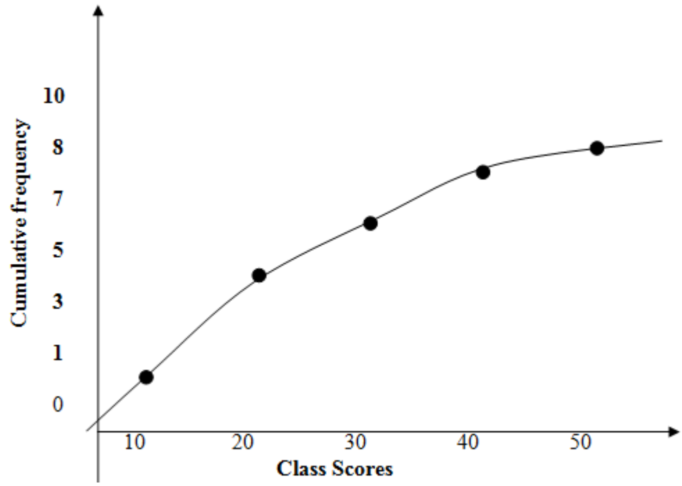


Figure 3.6 *Ogive.*

3.11.5 Scatter Diagram

The scatter diagram is used as a graphical tool for summarizing quantitative data, aiding in finding the relationship between two or more variables. With the scatter diagram there is no line made through the plots. Thus the coordinates are plotted and the pattern used to determine the relationships between the plotted variables on which data have been collected in a study.

Taking for example the data below:

Table 3.8.4 *Data for drawing the relationship between X and Y Coordinates.*

X	Y
1	5
2	10
3	15
4	20
5	25

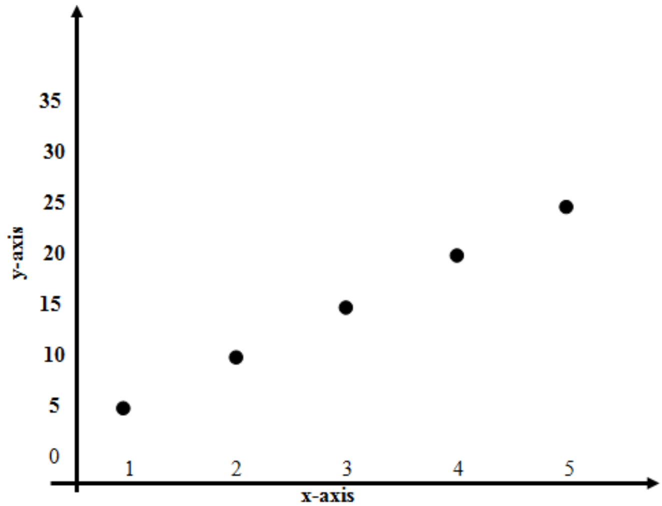


Figure 3.7 *Scatter Diagram.*

Various relationships can be established with scatter diagrams. Below are some of the identified relationships that can be established using the scatter diagram for various collected data sets

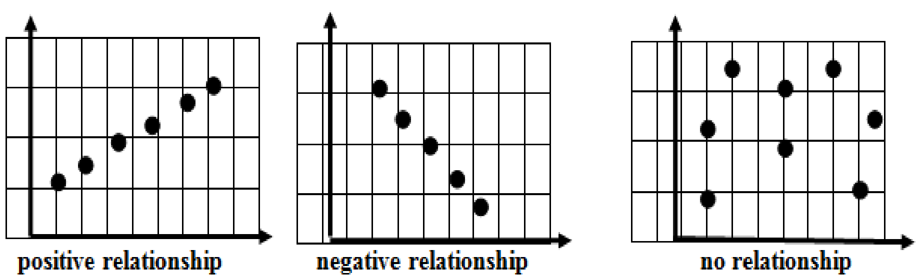


Figure 3.8 *Box and Whisker.*

3.11.6 Box and Whisker

The box and whisker plot is used in summarizing quantitative data. it looks like a number line with a box in the middle divided into three sections, representing the first quartile, second quartile and the third quartile starting from left to right. Also attached to the box are whisker, the tail one to the left indicating the minimum value and frontal one to the right the maximum value. Taking for instance the data below: 2, 2, 4, 4, 1, 3, 2, 4, 5, 6. The box and whisker can be drawn as follows:

The values are arranged first in ascending order shown below:

1,2,2,2,3,4,4,4,5,6,6,6

$$The\ position\ of\ the\ First\ Quartile\ Mark\ = (\frac{1}{4}\ x\ 12)^{rd} = \ 3^{rd}$$

Therefore looking at the numbers arranged in an ascending order, the 2 is the mark that occupies the third position.

Therefore the first Quartile Mark = 2

$$The\ position\ of\ the\ Second\ Quartile\ Mark\ = (\frac{2}{4}\ x\ 12)^{th} = \ 6^{th}$$

Hence from the number arranged in an ascending order the 6^{th} position is occupied by the numbers 4 and 4 counting from both sides. Therefore to find the 6^{th} position. The average of both numbers must be used.

$$Thus\ average\ of\ the\ numbers\ = (\frac{4+4}{2}) = \ 4$$

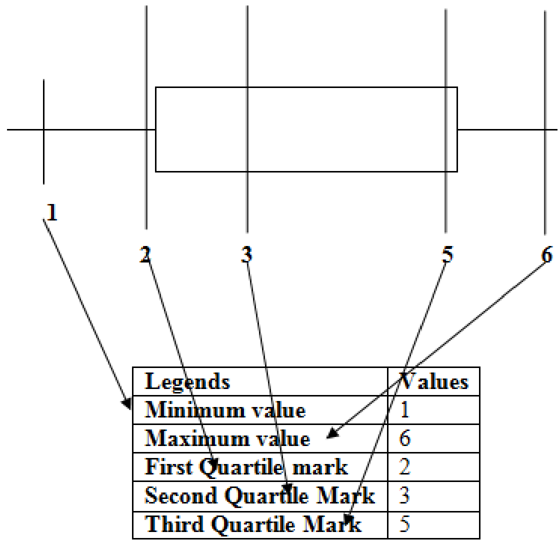
Hence the Second Quartile Mark = 4

$$The\ position\ of\ the\ Third\ Quartile\ Mark\ = (\frac{3}{4}\ x\ 12)^{th} = \ 9^{th}$$

Thus the 9^{th} position is occupied by 5, hence the Third Quartile Mark = 5

From the given data set, the minimum value = 1; and the maximum value = 6

This computed information from the data set can be used to draw the box and whisker plot as shown Figure 3.9:



Legends	Values
Minimum value	1
Maximum value	6
First Quartile mark	2
Second Quartile Mark	3
Third Quartile Mark	5

The box and whisker plot for the data set

Figure 3.9 *Stem and leaf display.*

Bibliography

[1] Fetterman, D. M. (1989). *Ethnography: Step by Step*. Applied Social Research Methods Series, Vol. 17. Newbury Park, CA: Sage.

[2] Guba, E. G., and Lincoln, Y. S. (1981). *Effective Evaluation*. San Francisco, CA: Jossey-Bass.

[3] Lincoln, Y. S., and Guba, E. G. (1985). *Naturalistic Inquiry*. Beverly Hills, CA: Sage.

[4] Patton, M. Q. (1990). *Qualitative Evaluation and Research Method*, 2nd Ed. Newbury Park, CA: Sage.

[5] Yin, R. K. (1989). *Case Study Research: Design and Method*. Newbury Park, CA: Sage.

[6] Sofaer, S. (1999) Qualitative research methods: what are they and why use them? Special Supplement on Qualitative Methods in Health Services Research, December 1999, Part II. *Health Serv. Res.* 34: 1101-1118.

[7] Shortell SM. The emergence of qualitative methods in health services research. Special Supplement on Qualitative Methods in Health Services Research, December 1999, Part II. *Health Serv. Res.* 34: 1083-1090.

[8] Edmunds, H. (2000). *The Focus Group Research Handbook.* Lincolnwood, IL: NTC Publishing Group.

[9] Templeton, J. F. Focus Group: *A Strategic Guide to Organizing Conducting and Analyzing the Focus Group Interview*. Chicago, IL: The McGraw-Hill Companies 1996.

[10] Krueger, R. A., Casey, M. A. (2000). *Focus Groups: A Practical Guide for Applied Research.* Thousand Oaks, CA: Sage Publications.

[11] Morgan, D. L. (1998). *Focus Groups as Qualitative Research* (2nd edition). Thousand Oaks, CA: Sage Publications.

[12] Strauss, A., Corbin, J. (1998). *Basics of Qualitative Research*. (2nd edition). Thousand Oaks, CA: Sage Publications.