Data Analysis: Practical Session

Instructions for the practical:

Summary Statistics

Example 01: Open the "heights.xlsx" file. The file contains heights of 150 students.

- Measures of location
 - 1. Calculate the mean height of a student.
 - Hint: AVERAGE (data_range)
 - 2. Calculate the median height of a student. Hint: MEDIAN (data range)
 - 3. Find the mode of the heights.
 - Hint: MODE (data range)
 - 4. Find the sum of heights. Hint: SUM (data_range)
- Measure of dispersion
 - 1. Find the minimum height and maximum height of students. Hint: MIN (data range)
 - MAX (data_range)
 - Find the interquartile ranges of the heights (Q1, Q2, Q3) Hint: QUARTILE (data_range, quart)
 - Find the standard deviation of the heights Hint: STDEV (data_range)

Displaying Quantitative data (Continuous case)

1. Obtain a histogram of heights.

Hints:

- Select data range
- Go to Insert -> Insert Statistic Chart -> Histogram
- 2. Define data intervals appropriately.

Hints:

- Double click on the x axis of the histogram. Then Format Axis pane will appear on the right hand side.
- Tick a mark on **Bin width** and type appropriate bin width (for ex: type 5)
- Enter
- 3. Add a chart title called "heights of students" to the histogram.

Hint: Click on the Chart title and type.

4. Add Axis titles, named x axis as "Height intervals" and y axis as "Frequency".

Hint:

- Click on the chart. Then a green color plus sign (Chart Elements) will appear on the right.
- Click on it and tick a mark in front of **Axis Titles**.

Centre for Health Informatics, Biostatistics and Epidemiology Faculty of Medicine, University of Kelaniya

- Click on the x-axis title and type "Height intervals"
- Click on the y-axis title and type "Frequency"
- 5. Add data labels
 - Hint: Use Chart Elements icon on the right hand side of the graph.
- 6. Remove gridlines
 - Hint: Use **Chart Elements** icon on the right hand side of the graph.

Exercise 01:

- 1. Open the "Marks.xlsx" file that contains the mathematics marks of 100 students in a school.
- 2. Find the average mathematics mark of a student.
- 3. Find median and mode of marks.
- 4. Calculate quartiles of the mathematics marks. (Q1,Q2,Q3) Compare Q2 value with median.
- 5. Calculate standard deviation.
- 7. Obtain a histogram of mathematics marks.
- 8. Add a chart title called "Mathematics Marks of Students".
- 9. Add Axis titles, named x axis as "Marks" and y axis as "Frequency".
- 10. Add data labels.
- 11. Remove gridlines.

Displaying Quantitative data (Discrete case)

Example 02: Open the "accidents.xlsx" file. Data set contain the number of accidents experienced by 80 machinists in a certain industry over a period of one year.

1. Create a frequency table of number of accidents.

Hints:

- Select the data range.
- Go insert -> Tables -> Pivot Table
- Select Existing Worksheet and give a location. (ex. Click on an empty cell)
- Click OK
- In the Pivot Table Fields pane select and drag no of family members to ROWS area.
- Again select and drag no of family members to VALUES area.
- Click drop down list in the VALUES area and select Value field settings.
- Select **count** and click **OK**.
- 2. Obtain a bar chart to compare no of accidents.

Hints:

- Select frequency table
- insert -> Insert Column or Bar chart -> 2-D Column
- 3. Add a Chart title called "No of accidents over a year"
 - Hint: Proceed the same path as "heights" example
- 4. Add Axis titles, named x axis as "No of accidents" and y axis as "Frequency".

Centre for Health Informatics, Biostatistics and Epidemiology Faculty of Medicine, University of Kelaniya

- 5. Add data labels
- 6. Remove gridlines

Exercise 02:

- 1. Open the "family.xlsx" file that contains the number of family members lived in each flat of a large housing complex.
- 2. Create a frequency table of number of family members
- 3. Obtain a bar chart to compare different no of family members.
- 4. Add a Chart title called "No of family members in a house"
- 5. Add Axis titles, named x axis as "No of family members" and y axis as "Frequency".
- 6. Add data labels
- 7. Remove gridlines

Displaying categorical data

Example 03: Open the "Smolking.xlsx" file. File contains the details of age, gender and smoking preference of 65 individuals.

1. Create a frequency table for gender.

Hints:

- Select data range
- Go insert -> Tables -> Pivot Table
- Select Existing Worksheet and give a location. (ex. Click on an empty cell)
- Click OK
- Select and drag gender to **ROWS** area.
- Again select and drag gender to VALUES area.
- 2. Create a frequency table for smoking preference.
- 3. Create a two-way frequency table for gender and Smoking preference. Hints:
 - Select data range
 - Go insert -> Tables -> Pivot Table
 - Select Existing Worksheet and give a location. (ex. Click on an empty cell)
 - Click OK
 - Select and drag gender to **ROWS** area.
 - Again select and drag gender to VALUES area.
 - Select and drag smoking to **COLUMNS** area.
- 4. Obtain a bar chart of gender and smoking preference.
 - Select frequencies
 - insert -> Insert Column or Bar chart -> 2-D Column
- 5. Add a Chart title called "Gender & smoking preference"
- 6. Add data labels.
- 7. Remove gridlines

Centre for Health Informatics, Biostatistics and Epidemiology Faculty of Medicine, University of Kelaniya Exercise 03: Open the "university.xlsx" file that contains the details of 60 students who applied for university entrance.

- 1. Create a frequency table for gender.
- 2. Create a frequency table for district.
- 3. Create a frequency table for university admission.
- 4. Create a two-way frequency table for gender and university admission.
- 5. Create a two-way frequency table for district and university admission.
- 6. Create three-way frequency table to compare district, gender and university admission. Hints:
 - Select data range
 - insert -> Pivot Table
 - Go insert -> Tables -> Pivot Table
 - Select Existing Worksheet and give a location. (ex. Click on an empty cell)
 - Click OK
 - Select and drag gender to **ROWS** area.
 - Also select and drag district to **ROWS** area.
 - Select and drag university admission to COLUMNS area.
 - Again select and drag gender to VALUES area.
- 7. Obtain a bar chart of gender and university admission.
- 8. Obtain a bar chart of district and university admission.
- 9. Obtain a bar chart to compare gender, district and university admission.