An Introduction to Statistics using Minitab
Minitab Lab 1 Exercises

Datasets for these examples can be accessed at:

http://personal.strath.ac.uk/david.young/SQA/

Question 1

The `Activity` file contains data relating to health care commissioned by the Scottish government to address health inequalities.

(i) For each variable in the dataset, decide which are quantitative and which are qualitative.

(ii) Produce a table to examine the distribution of patient activity.

(iii) Use the **Graph** option to produce a pie chart of the activity levels. Use the various options to change the titles and format of the chart.

(iv) Use the **Graph** option to produce a histogram of the pulse rates. Comment on the distribution of the data and produce a table of the most appropriate statistics to summarise the location and spread of the pulse rates.

(v) Compute descriptive statistics for the pulse rates for each level of activity and comment on any differences between the three groups.

Question 2

The data below show rainfall levels for each month (mm) for a location in the USA:

```
SamSamWater Climate Tool

Name of location (approximately): Palm Springs, CA 92264, USA
Latitude: 33.82790 (decimal degrees)
Longitude: -116.57257 (decimal degrees)
Altitude: (m above mean sea level)

Average precipitation (in mm or liter per m) for this location is listed
in the table below.

Month       Rainfall (mm)
January      99
February     82
March        93
```

```
April        58
May          24
June         8
July         19
August       27
September    22
October      24
November     66
December     82
```

Produce a bar chart of this data to illustrate the trend over time – the data are available in the Excel worksheet `Rainfall`.

Note: This data is available at `http://www.samsamwater.com/climate/index.php`. Download rainfall data at another location from this website and produce a similar plot.

Question 3

Crime data were downloaded for Tayside police force areas on SIMD and crime counts (which relates to selected recorded offences, not all crimes committed in the area). These are available in the Excel file `Crimes` and were downloaded from:

`http://simd.scotland.gov.uk/publication-2012/download-simd-2012-data/`

Produce a plot to illustrate the relationship between SIMD and crime rates and comment on the relationship.

Question 4

People who are concerned about their health may prefer hotdogs that are low in salt and calories. The `Hotdogs` datafile contains data on the sodium and calories contained in each of 54 major hotdog brands. The hotdogs are classified by type: beef, poultry, and meat (mostly pork and beef, but up to 15% poultry meat).

(i) Produce a table to show the number and percentage of each type of hotdog in the data set.

(ii) Produce a plot to compare the calories between the different types of hotdog and give a subjective impression of any differences. (Hint: This could be done using a boxplot of calories for each type of hotdog on the same graph, which allows a good visual comparison. Use **Graph > Boxplot** and select the option **With Groups**. The **Graph variable** is **Calories** and in the **Categorical variables** box enter **Type**)

(iii) Produce a plot to investigate any association between the amount of sodium in the hotdogs and the calorie content and interpretation the relationship. (Hint: A scatterplot is the appropriate graphical display to use to subjectively assess the association between two quantitative variables, i.e. **Sodium** and **Calories**).

Saving Data and Output

In the **File** command there are two options for saving a Minitab file:

1. **Save Current Worksheet** saves only the data. There are options to **Save as Type** which allows you to save the data as a Minitab worksheet or as an Excel worksheet.

2. **Save Project** - this option saves the whole Minitab file exactly as it is, including all the output in the session window. Once re-opened, the data sheet will be available along with all the previous output and graphs.

Outline Solutions

Question 1

To open an Excel file in Minitab, use the following commands:

File > Open Worksheet

In the command box, select **Excel (*.xls, *.xlsx)** and select the Excel file to be opened. To produce a table of activity levels choose:

Stat > Tables > Tally Individual Variables.

In **Variables**, enter *Activity*. Under **Display**, check **Counts** and **Percents**. Click **OK**.

In the session window the output table should be:

```
Tally for Discrete Variables: Activity

Activity  Count  Percent
   A lot     21    23.08
Moderate     61    67.03
  Slight      9     9.89
     N=      91
```

Interpretation: You have 9 subjects with slight activity level, representing 9.89% of all the individuals in the study. You have 61 subjects with moderate activity level.

Note: Each dialogue box in Minitab has a help button which links to a useful description of the command being executed. Throughout these exercises, make use of the help function to clarify any points that are not clear.

Note: Graphs can be copied by right clicking in the graph box and selecting **Copy Graph**. This can then be pasted into a Word document. Alternatively you can save the graph in various formats using the **Save Graph As** option in the main menu.

The histogram of pulse rates appears slightly skewed (see Figure 1) so the appropriate measures of location and spread are the median and IQR.

Use **Stat > Basic Statistics > Display Descriptive Statistics** to produce summary statistics for numerical variables. Select **Pulse** in the **Variables** box and **Activity** in the **By variables** box to display the descriptive statistics shown below. Note that you should comment on the differences between the mean values *or* the differences between the median values as appropriate for the distribution of the pulse rates.
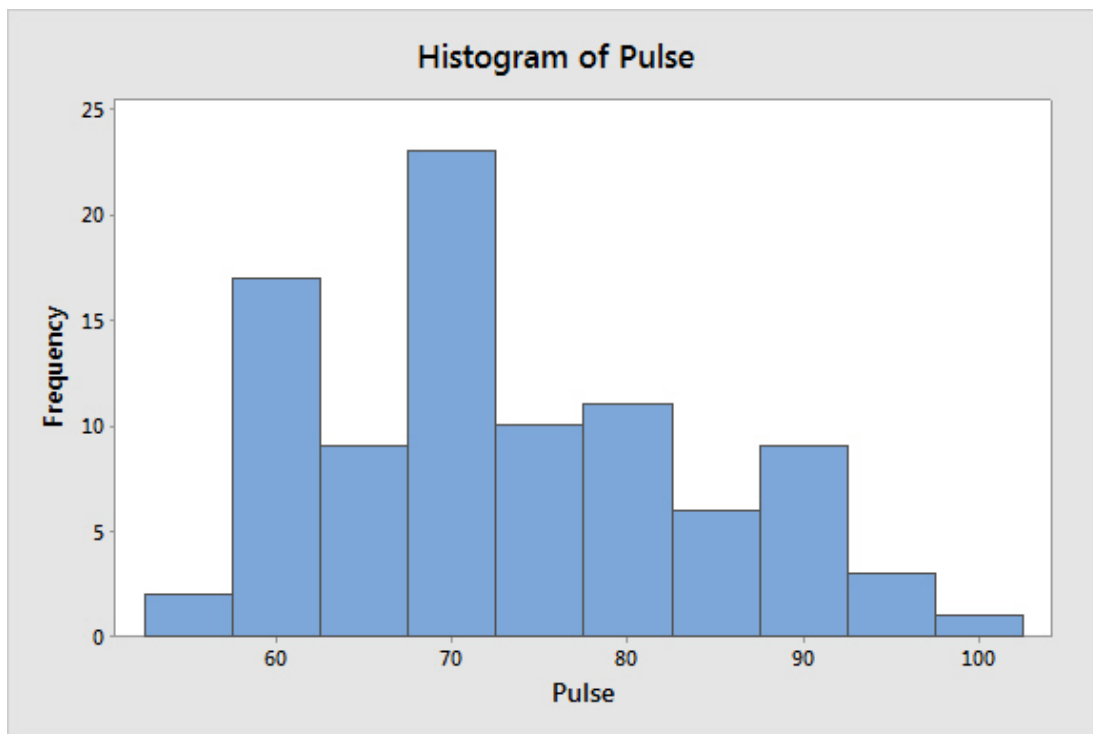
Figure 1: Histogram of pulse rates

```
Descriptive Statistics: Pulse

Variable  Activity   N    Mean  StDev  Minimum     Q1  Median     Q3  Maximum
Pulse     Slight     9   79.56  10.48    62.00  70.00   82.00  90.00    90.00
          Moderate  61   72.74  10.98    54.00  65.00   70.00  80.00   100.00
          A lot     21   71.57   9.63    58.00  63.00   70.00  77.00    92.00
```

Question 2

Using **Graph > Bar Chart** and under the option **Bars represent** select **Values from a table**. Choose the simple bar chart and enter **Rainfall** in the **Graph variables** box and **Month** in the **Categorial variables** box. This produces the bar chart in Figure 2.

Question 3

The scatter plot in Figure 3 is the appropriate way to display the relationship between two quantitative variables. There is a clear association between the two variables: as SIMD increases, the number of crimes decreases.
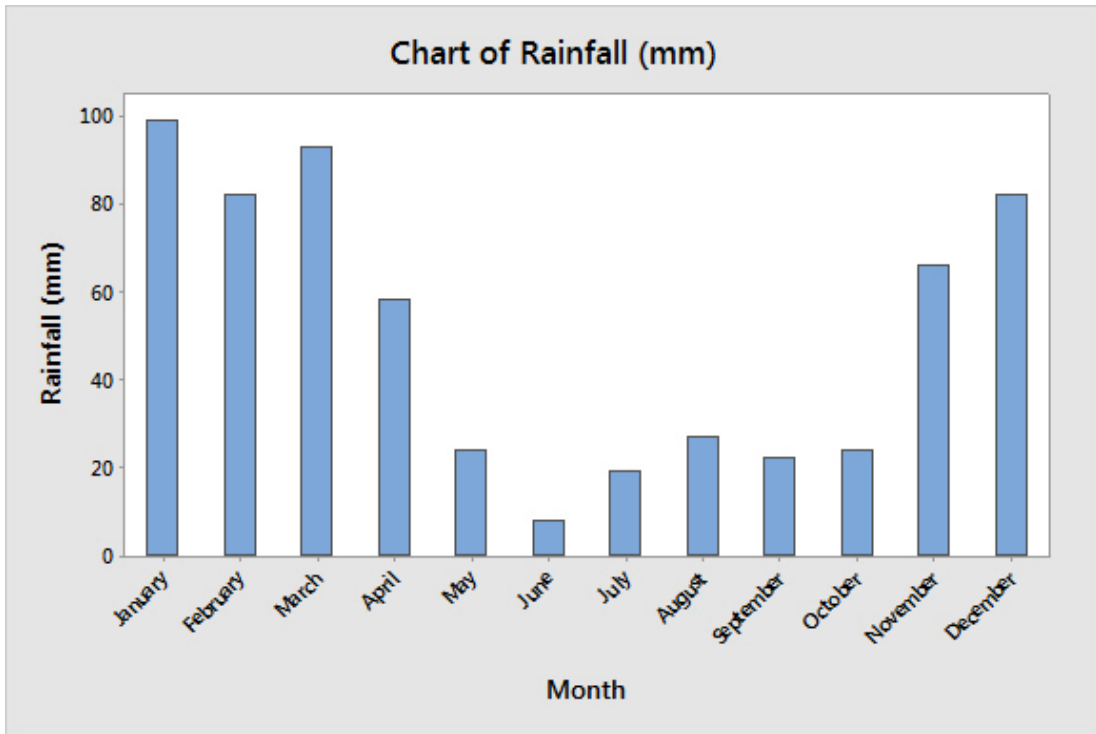
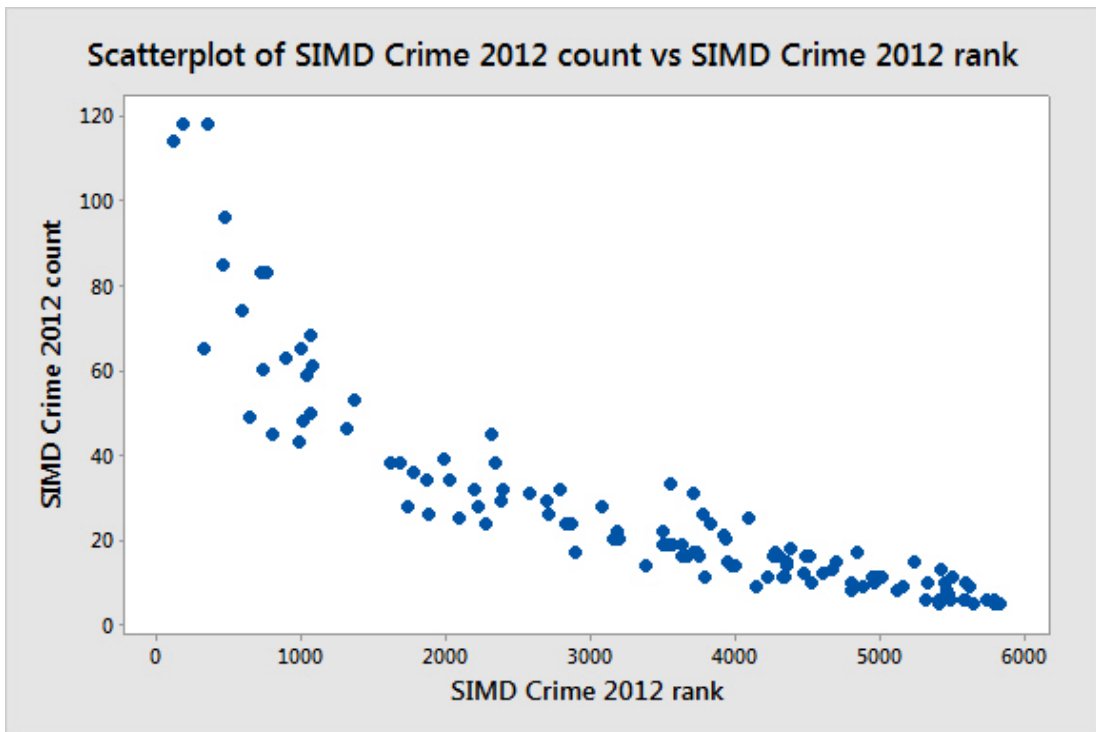Figure 2: Bar chart of rainfall for Palm Springs, USA

Figure 3: Relationship between SIMD and crime rates

Question 4

(i) The distribution of hotdog type is shown:

```
Tally for Discrete Variables: Type

  Type  Count  Percent
  Beef     20    37.04
  Meat     17    31.48
Poultry    17    31.48
   N=      54
```

(ii) The boxplot is shown in Figure 4. This shows that poultry hotdogs have fewer calories and there is little difference between beef and meat hotdogs in terms of their calorie content.
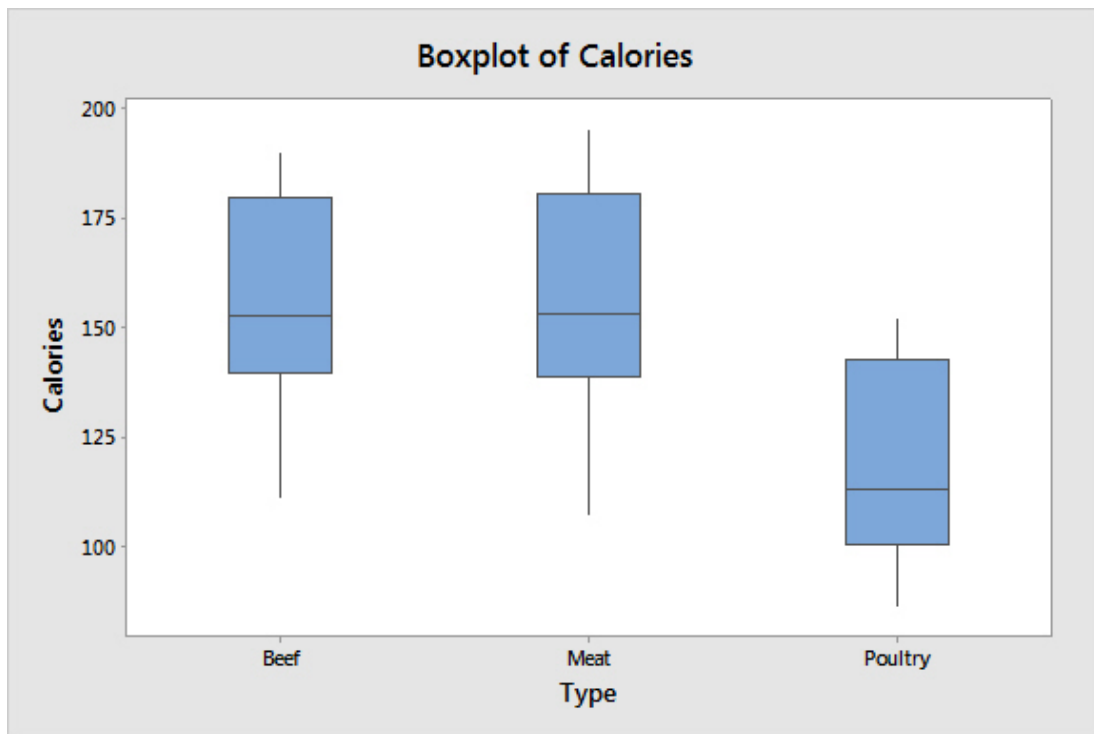

Figure 4: Boxplot comparing the calories between hotdogs

(iii) This should be a scatter-plot as in Question 3. There is some indication that both increase together, but not a strong linear relationship between the two variables.