

9.1 Bar Charts

The instructions here are given for **Excel**. Unfortunately the interface in **LibreOffice** is slightly different from Excel so the instructions for LibreOffice are given below. Ensure you save your answers in the Exercise sheets for your submission.

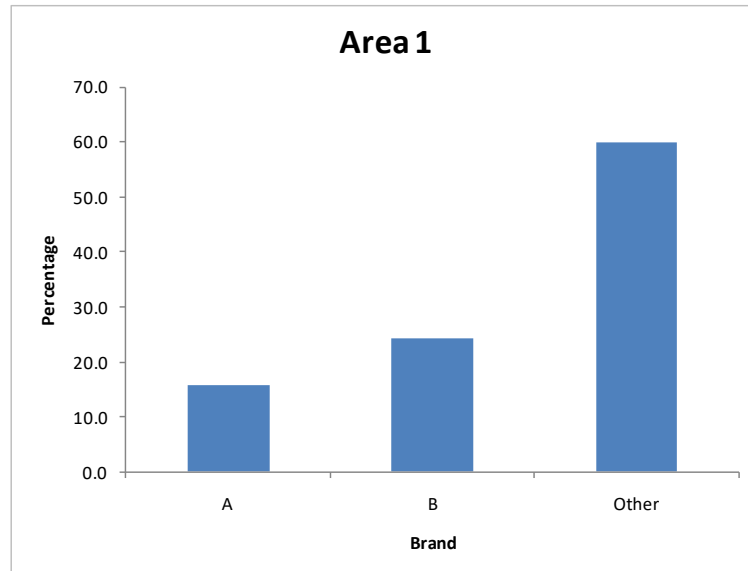
Example 9.1

Consider Data Set D (see the Data Annexe). We produce a percentage frequency bar chart for the breakfast cereal brand preferences for demographic Area 1.

1. Open the Excel workbook **Exa 9.1D.xlsx** from the Examples folder. This contains the relevant data, together with the frequency and percentage frequency distributions.
2. Highlight the cells containing the Brands and percentage frequencies (D15:E17).
3. From the **Insert** menu bar tab, select the **Column** icon from the **Charts** group. A drop down menu of charts appears. Select the first chart on the first row – the “clustered column” chart.
4. A bar chart will appear. You should also find that three new **Chart Tools** menu bar tabs have appeared. Choose the **Layout** menu bar tab, select **Chart Title** from the **Labels** group and then select **Above Chart**. A text box will appear at the top of the chart. Edit the title by typing “Area 1” into the box.
5. Now from the **Layout** menu bar choose **Axis Titles** from the **Labels** group, then **Primary Horizontal Axis Title** and **Title Below Axis**. A text box will appear below the horizontal axis; edit this by clicking in the box - change the label to “Brand”. Similarly, by selecting from the **Labels** group: **Axis Titles/Primary Vertical Axis Title/Rotated Title**, create a label reading “Percentage” for the vertical axis.
6. Now select **Gridlines** from the **Axes** group on the **Layout** menu bar, then **Primary Horizontal Gridlines** and **None** to remove the unsightly horizontal lines from the graph.
7. Now select **Legend** from the **Labels** group on the **Layout** menu bar. As we have only one data series, we do not require a legend! Select **None**.
8. Click anywhere near the edge of the chart area to select the chart. Drag the chart so that it's top left-hand corner in cell G3. Click on the chart, and drag the handle at the lower right corner to cell N25, hence re-sizing the chart. Click back in the worksheet to de-select the chart.

Your completed chart should resemble that in the figure below:

Figure 9.1: Percentage Frequency Bar Chart for Area 1 Brand Preferences



It is clear from the chart that Brand A is least preferred, followed by Brand B, whilst the majority of respondents preferred some other brand.

You might want to experiment with your chart to further customise it. For example, right click on the vertical axis and select **Format Axis** and then **Number** to reformat the axis numbering to remove the redundant decimal places.

Exercise 9.1

Open the Excel workbook in **Exe 9.1D.xlsx** from the Exercises folder. This contains the percentage frequencies together with the bar chart just created in the above example. Add a percentage frequency bar chart showing the brand preferences in Area 2, using the same format as that employed for the Area1 results in the above example. Drag your new chart so that it lies alongside that for Area 1.

Briefly interpret your findings. What do these results tell you about the patterns of brand preferences for each of the two demographic areas?

- In area 1, 60% prefer “other” brand as compared to 15.7% of brand A and 24.3% of brand B.
- Similarly, area two presents a higher preference at rate 21.1 for brand A while a 33.3% preference towards brand B. In addition, a more significant rate of 45.6% prefers “other” brands.

Interpretation:

- According to Area 1, “Other” brand preference is high thus showing that there are many choices besides brands A and B.
- The second area indicates a comparatively lower preference of “other” brands against the identified brands A and B. This implies that “other” brands may be less preferred by the consumers.
- The implication is that the two regions have different brands’ tastes. In area 1, there is a broader preference in contrast to area 2 that appears to be specific about some brands A and B.

Example 9.2

Consider again Data Set D (see the Data Annexe). We produce a percentage frequency clustered column bar chart for the breakfast cereal brand preferences for the two demographic Areas 1 and 2.

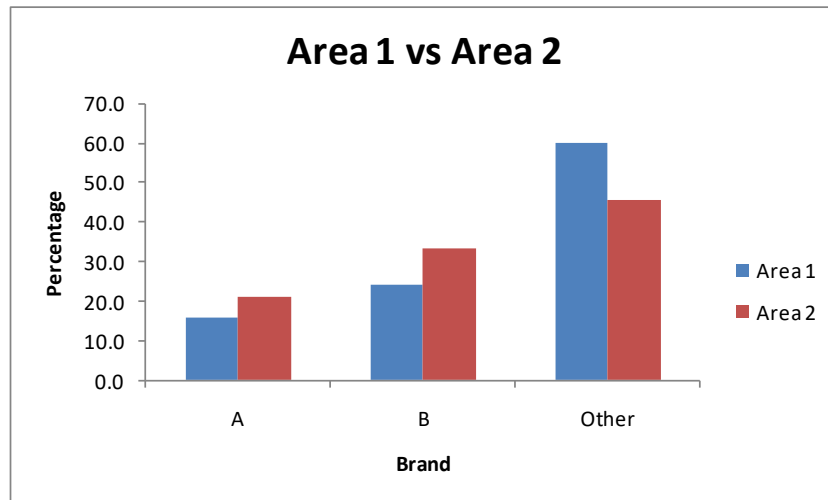
1. Open the Excel workbook **Exa 9.2D.xlsx** from the Examples folder. This contains the relevant data, together with the frequency and percentage frequency distributions.
2. Highlight the cells containing the Areas, Brands and percentage frequencies (D14:F17).
3. From the **Insert** menu bar tab, select the **Column** icon from the **Charts** group. A drop down menu of charts appears. Select the first chart on the first row – the “clustered column” chart.
4. The bar chart will appear. Choose the **Layout** menu bar tab, select **Chart Title** from the **Labels** group, then select **Above Chart**. A text box will appear at the top of the chart. Edit the title by typing “Area 1 vs Area 2” into the box.
5. Now from the **Layout** menu bar choose **Axis Titles** from the **Labels** group, then **Primary Horizontal Axis Title** and **Title Below Axis**. A text box will appear below the horizontal axis; edit this by clicking in the box - change the label to “Brand”. Similarly, by selecting from the **Labels** group: **Axis Titles/Primary Vertical Axis Title/Rotated Title**, create a label reading “Percentage” for the vertical axis.
6. Now select **Gridlines** from the **Axes** group on **Layout** menu bar, then **Primary Horizontal Gridlines** and **None** to remove the unsightly horizontal lines from the graph.
7. Click anywhere near the edge of the chart area to select the chart. Drag the chart so that it’s top left-hand corner in cell H3. Click on the chart, and drag the handle at the lower right corner to cell O25, hence re-sizing the chart. Click back in the worksheet to de-select the chart.

Your completed chart should resemble that in the figure below:

It is clear from the chart that in both Areas, Brand A is least preferred, followed by Brand B, whilst even more respondents preferred some other brand. However, it is now very clear that

Brand A and Brand B preferences were both higher in Area 2 than in Area 1, whilst the percentage of respondents who preferred other brands was lower in Area 2.

Figure 9.2: Percentage Frequency Clustered Column Bar Chart for Area 1 and Area 2 Brand Preferences



You might again want to experiment with your chart to further customise it. For example, as before, right click on the vertical axis and select **Format Axis** and then **Number** to reformat the axis numbering to remove the redundant decimal places.

Exercise 9.2

Open the Excel workbook in **Exe 9.2E.xlsx** from the Exercises folder. This contains the frequency distributions for Data Set E (see the Data Annexe) to which has been added the corresponding percentage frequency distributions. Complete a percentage frequency clustered column bar chart showing the heather species prevalences in the two different locations.

Briefly interpret your findings.

- At location A, there are 46.4 percent “abundant”, 39.3 percent “sparse” and 14.3 percent “absent”.
- Compared to location B, the graph shows a distinct pattern with maximum values of “Absent” (45.5%) and “Sparse” (31.8%) while the lowest value is recorded as “Abundant” (22.7%).
- Interpretation:
 - “Abundant” observations seem more frequent for location A compared to “Absent” observations.
 - The situation in location is different – “absent” counts dominate followed by “abundant”.
- This implies that there is a major difference in the distributions of observations at these two sites. There are more “Abundant” observations in location A whereas there is an increased number of “Absent” in location B.

9.2 Histograms

Example 9.3

Consider the dietary data contained in Data Set B (see the Data Annexe). We construct a relative frequency histogram for the weight losses of those individuals who underwent Diet A.

1. Open the Excel workbook **Exa 9.3B.xlsx** from the Examples folder. This contains the relevant data, together with the summary statistics and an added text template.
2. Before constructing our histogram, we need to inspect the sample minimum, sample maximum and sample range for the Diet A data. This will inform our choice of classes. Thus, in cell F7, enter the formula **=MIN(B2:B51)**. In cell F8, enter the formula **=MAX(B2:B51)**, and in cell F9, enter the formula **=F8-F7**. Centre these cells for neatness, and if necessary, round them to three decimal places..

Now, our sample range is just about 12. This will give either four classes each of width 3 (too few!) or six classes of width 2 (OK!). So we'll use six classes each of width 2.

Since the sample minimum is -1.715 and the sample maximum is 10.062, we'll start at -2, that is, with the class (-2, 0]. The remaining classes are therefore (0, 2], (2, 4], ..., (10, 12]. Notice that the last class has to go up to 12 so that we capture the largest observation (the sample maximum of 10.062). In fact, we now have seven classes.

In order to construct a frequency distribution, Excel needs to be told our upper class boundaries.

3. Enter the upper class boundaries of 0, 2, 4, ..., 12 in the respective cells H4 to H10 (under the heading **UCB**) and centre these cells for neatness.

We now use the Excel *array function* **FREQUENCY()** to obtain the various class frequencies. An array function assigns values to a range of cells simultaneously, so there are special rules governing how to set up such a function.

4. Select the cell range I4 to I10 (under the heading Frequency). This is where we want to define our array function. With these cells still selected, click in the formula bar at the top of the screen.
5. Now enter the array formula **=FREQUENCY(B2:B51,H4:H10)** *but do not press Enter or click OK..* The first range in this formula tells Excel where the relevant data are, whilst the second range tells Excel where the required upper class boundaries are.

Now press **Ctrl-Shift-Enter** (this is how an array formula must be entered). You should see the frequencies 1, 3, 10, 15, 15, 5, 1 appear in the cells under the Frequency heading. (If you make an error here, press the **Escape** key to reset and try again.)

In order to obtain a relative frequency distribution, we'll need to know the total number of observations. But we already know this – it's the sample size n in cell F3. However, just as a check, we'll add up the various class frequencies.

6. In cell H11 type: Total, and in cell I11 enter the formula **=SUM(I4:I10)**. Centre and embolden these cells. The total number of observations should check out as 50.

Now, to ensure that our histogram bars will meet at the class boundaries, each must be plotted at the corresponding **class mark** – the mid-point of each class. This can be obtained by simply averaging the upper and lower class boundaries.

Thus for the class (-2, 0] the class mark is $(-2 + 0)/2 = -1$
 (0, 2] the class mark is $(0 + 2)/2 = 1$
 (2, 4] the class mark is $(2 + 4)/2 = 3$

... and so on. The class marks increase by 2 each time.

7. Enter the class marks of -1, 1, 3,..., 11 in the respective cells K4 to K10 (under the heading **Class Mark**) and centre these cells for neatness.

We now enter the various relative frequencies (proportions of observations) into the cells L4 to L10 (under the heading **Relative Frequency**).

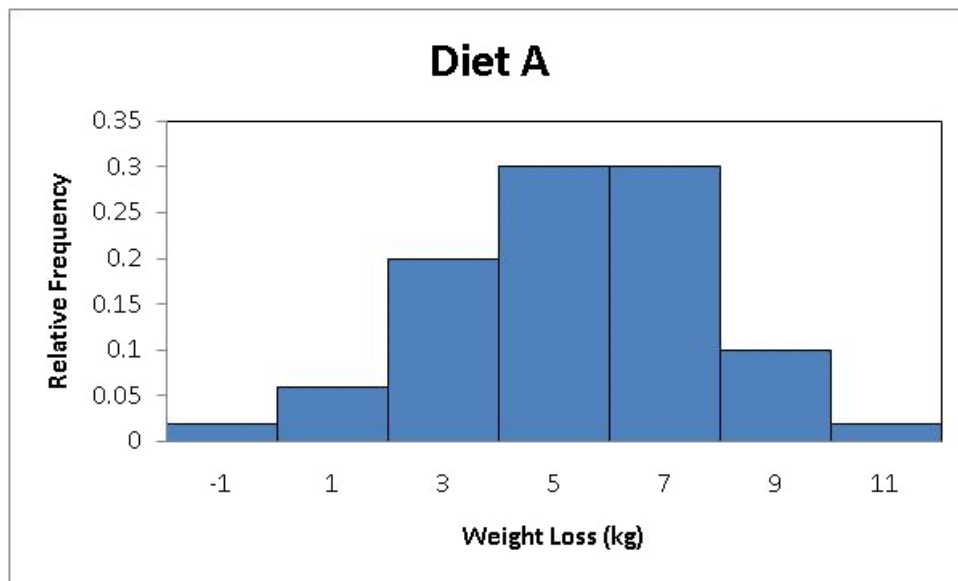
8. In cell L4, enter the formula **=I4/I\$11**. Copy and paste this cell into the cells L5:L10. Format the cells L4:L10 to two decimal places. As a check on calculations, copy and paste cells H11:I11 into K11:L11. As anticipated, the total relative frequency is 1 (all the observations have been accounted for).

We now construct a simple bar chart, plotting the relative frequencies against the class marks, following the procedure outlined in Example 9.1. This will constitute our histogram.

9. Highlight the cells containing the relative frequencies, **L4:L10**. Select the **Insert** menu bar tab, choose **Column** from the **Charts** group then select the first chart on the first row – the “clustered column” chart.
10. In the **Chart Tools/Design** menu bar, find the **Chart Layouts** group, scroll down and select **Layout 8** (where there are no gaps between the bars).
11. Again from the **Chart Tools/Design** menu bar, choose **Select Data** from the **Data** group. Below **Horizontal (Category) Axis Labels**, click on **Edit**. Click on the small red arrow at the right of the **Axis label range** box and highlight cells **K4:K10**. The category axis labels (-1, 1, ...) should appear to the right of the box. **Click OK** and then **click OK** again. Correct labels should appear on the horizontal axis.
12. Double click on the Chart Title and edit it to read "Diet A". Edit the Axis Titles in the same way, making the horizontal one read "Weight Loss (kg)" and the vertical one "Relative Frequency".
13. Ensure the chart is still selected (just **click** on it). Select the **Chart Tools/Layout** menu bar tab, from the **Current Selection** group at the left-hand side, choose **Series 1** from the drop down menu, then select **Format Selection**. In the dialog box, select **Border Colour** from the left-hand menu and check the **Solid Line** button. In the drop down colour menu choose black, then click **Close**. Click anywhere back in the worksheet to see the effect of this. Your histogram is complete.
14. Drag the chart so that its top left-hand corner is in cell D13. Click on the chart, and drag the handle at the lower right corner to cell K35, hence re-sizing the chart.

Your completed histogram should resemble that in the figure below:

FIGURE 9.3: RELATIVE FREQUENCY HISTOGRAM FOR DIET A WEIGHT LOSS



For those individuals who underwent diet A, the weight loss distribution is unimodal and fairly symmetrical, with perhaps a hint of negative skewness.

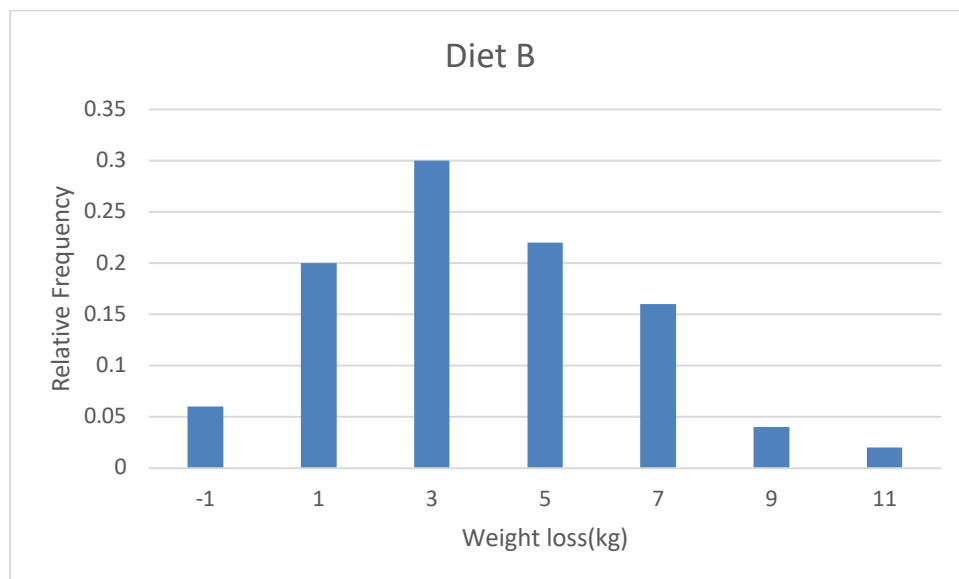
As before, you might again want to further customise your chart. For example, right click on the vertical axis and select **Format Axis** and then **Number** to reformat the axis numbering to display exactly two decimal places.

Note finally that it is possible to construct a histogram in Excel using the Analysis ToolPak (by selecting the **Data** menu bar tab, then **Data Analysis** from the **Analysis** group, and then picking the **Histogram** option). This is not recommended – the resulting histogram is mislabelled on its horizontal axis!

Exercise 9.3

Open the Excel workbook in **Exe 9.3B.xlsx** from the Exercises folder. This contains the relative frequency histogram for the Diet A weight loss produced in Example 9.3 together with some of the Diet B weight loss summary statistics. Add a relative frequency histogram of the weight loss for Diet B, where possible using the same classes as those employed for the Diet A results in the above example.

Briefly interpret your histogram. What do these results tell you about the patterns of weight loss for each of the two diets?



- Mean weight loss in Diet A is 5.341 against 3.71 of Diet B which indicates the fact that on an average individuals following Diet A lost more weight than those who followed the other diet.
 - The standard deviation for Diet A was 2.536 and less than that of Diet B which was 2.77, meaning that those who followed Diet A experienced consistent weight loss results which were mostly centered on the average than with those following Diet B.
 - The values for Diet A (11.777) are significantly larger than those for Diet B (10) and represent more variance among Diet A participants' respective weight losses.
 - It seems like diet A had a higher frequency of weight loss than the other one, because in classes 4, 5 and 6 there is weight loss. Whereas, in the classes 1, 2 and 3 diet B did not show so high frequency.
- ☐ In totality, these results imply that Diet A led to greater average amounts of weight loss as well as larger spread of the amount of weight loss, than did Diet B. Even though diet

B showed more compact grouping around moderate levels of weight loss and less cases of substantial weight loss that was observed