

Chapter 3

Graph Construction

Aim: Upon completing the chapter, the learner should be able to create the graphs that are most commonly used for data description.

3.1 Introduction

To create a graph, we click on the **Graphics** option on the taskbar (Figure 3.1). After we do this, the following dropdown menu appears, listing a series of possible graphs that can be constructed.

Afterward, the user clicks the type of graph or plot needed; a new window with the different specifications available for this type of graph will be displayed. Once the specifications are provided, the user must choose one of the following two options for obtaining the graph that he or she desires: *Submit* or *OK*. If *Submit* is chosen, the requested graph will be displayed, with the graph window remaining open (enabling the user to explore other specifications); choosing *OK* brings up the requested graph but the graph window remains closed.

3.2 Box Plot

To construct a box plot, the user should click the *Box Plot* option after clicking *Graphics* (Figure 3.2). Afterward, a quantitative variable must be defined. For example, to obtain the box plot for the variable *bmi* of the previous database, insert *bmi* in the space provided; in addition, the user has the option of writing a title in the *Title* option (Figure 3.3).

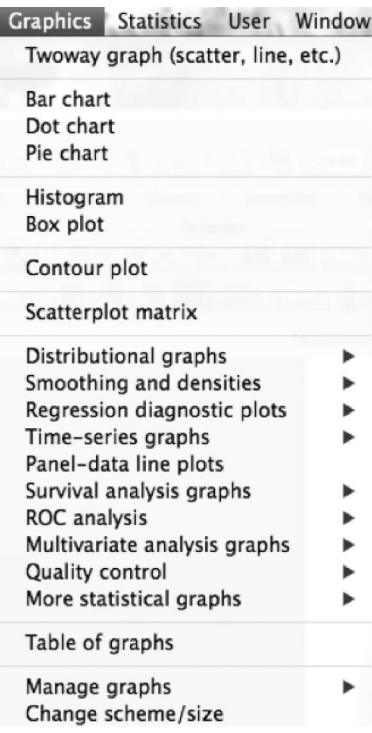


Figure 3.1 Graphics options.



Figure 3.2 Box plot window.

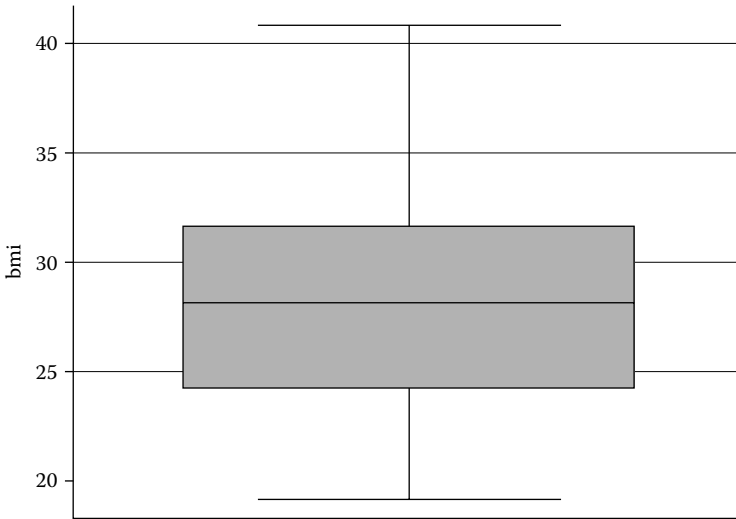


Figure 3.3 Graph box for bmi.

3.3 Histogram

Another commonly used chart is the histogram, which shows the frequency distribution of the variable of interest using abutting rectangles, and in which the height of each rectangle corresponds to the frequency of subjects within certain limits of the variable (these limits are the base of each rectangle). For example, to create a histogram of the variable with four rectangles using the *Graphics* window, the user needs to click the Histogram option and write the name of the variable, bmi (Figure 3.4). At this point, the user has the option of specifying the number of rectangles in the space labeled *Number of bins and*, in addition, has the option to include the normal density plot (see Figure 3.5).

The *normal* option will show a curve of the normal probability distribution over the histogram. This tells us how far away the distribution of the variable of interest is from the normal distribution.

3.4 Bar Chart

To construct a bar chart, the user clicks the *Bar Chart* option after clicking *Graphics* and sets the specifications for this type of graph (Figure 3.6). For example, to show the mean of the bmi by sex, the user needs to define the requested statistics (i.e., mean), and the variable that identifies the subgroup (for the purposes of this example, the variable *sex*) in the window that is opened when the *By* button is depressed (see Figure 3.7).

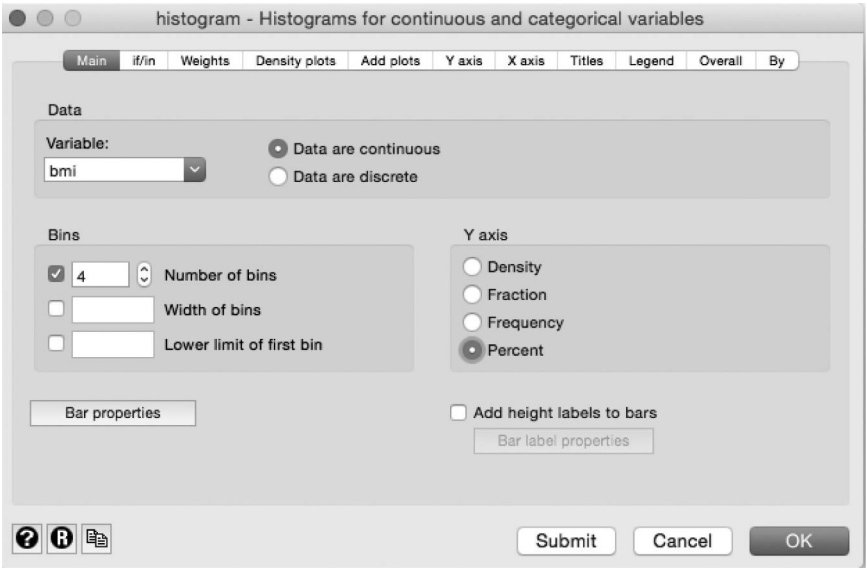


Figure 3.4 Histogram window.

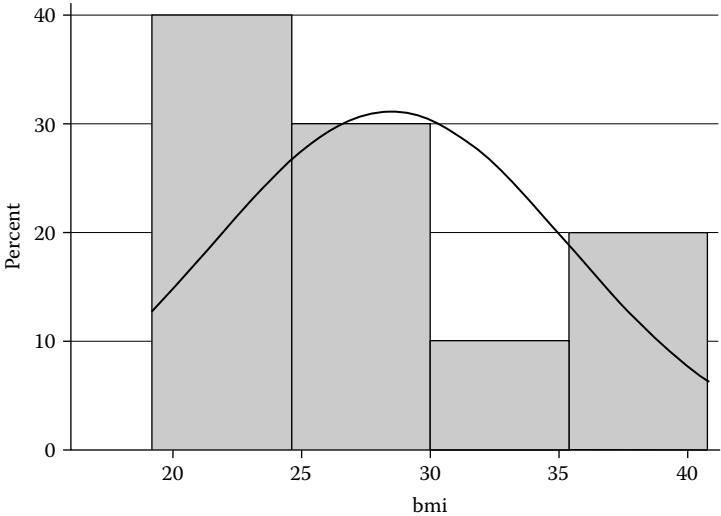


Figure 3.5 Histogram for bmi.

The results show that the mean of the bmi in women is higher than it is in men. The next chapter will demonstrate the procedure that is used to determine whether this sort of difference is statistically significant.

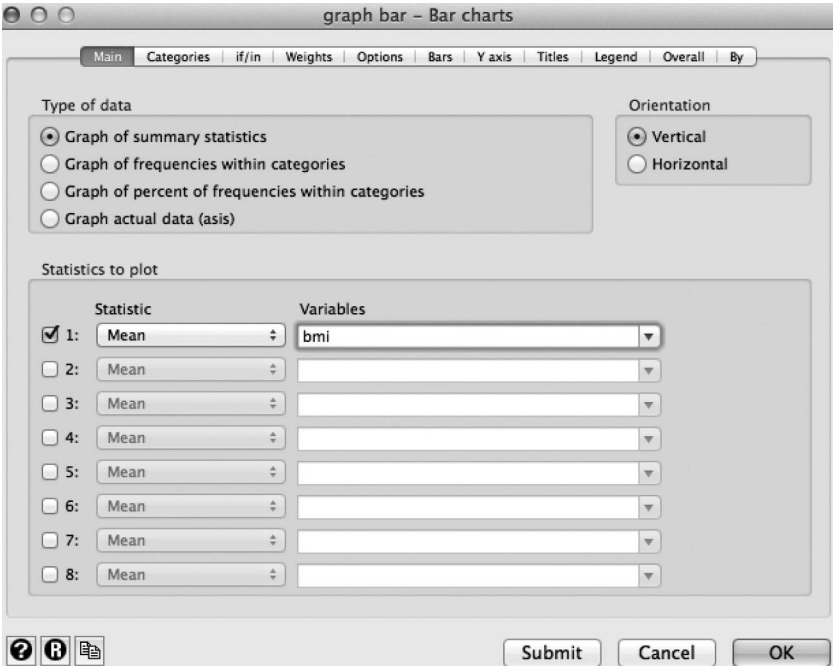


Figure 3.6 Bar chart window.

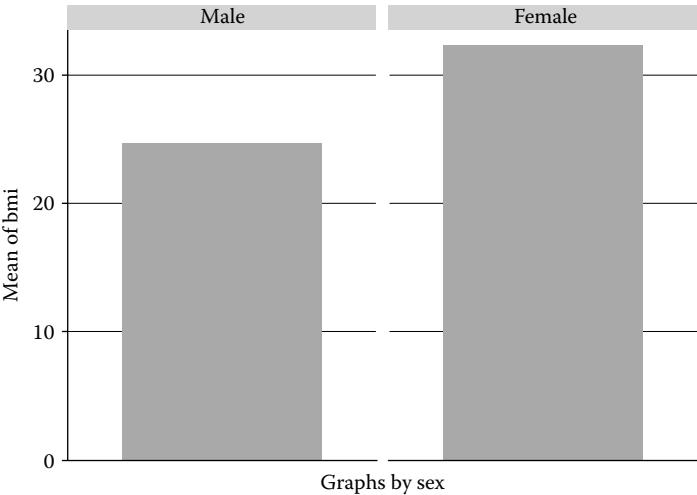


Figure 3.7 Graph bar: Mean bmi by sex.

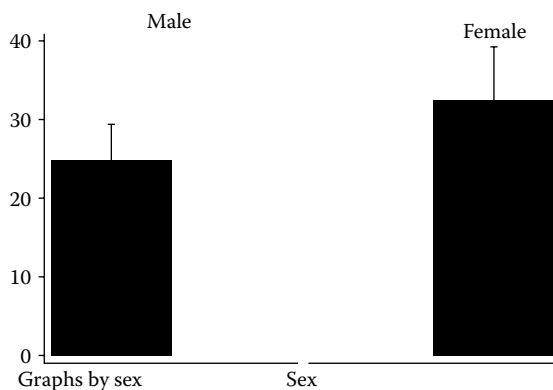


Figure 3.8 Mean bmi with one standard deviation by sex.

Another type of bar chart that the user might want to create is one in which the standard deviation is added (see [Figure 3.8](#)). The next sequence of commands can be used for this purpose:

```
sort sex
gen mbmi=bmi
gen sbmi=bmi
collapse (mean) mbmi (sd) sbmi, by(sex)
gen hbmi = mbmi + sbmi
twoway (bar mbmi sex) (rcap hbmi mbmi sex), yscale(range(0 40))
xlabel(none) by(sex, noxrescale) by(,legend(off))
```

The *collapse* command is used to summarize a set of data using statistics, such as mean, sum, median, and percentiles. These statistics can be computed overall or for each category of specific variables, previously sorted. In the last sequence of commands, we computed the mean and standard deviation of the variable *bmi* for each category of variable *sex*. The *twoway* command is used to create different plots in the same graph. In the previous example, we used *bar* for graph bars and *rcap* for capped spikes in the same graph.