Lab 3 Part 1: SPSS Graphing

Lab Objectives:

- Create a variety of graphs in SPSS
- Edit graphs using the chart editor
- Save chart templates
- Export data

Exercise 1: Graphing in SPSS

SPSS has the capability to create many types of charts as can be seen by clicking on the **Graphs** menu bar and selecting **Chart Builder**. Graphs can also be created by using options available in some dialog boxes for analyses. For example, histograms can be created from the Graphs menu or from **Analyze, Descriptive Statistics,** and **Frequencies**. In this exercise, we will focus on the Graphs menu. We'll begin with **frequency distributions**.

Recall that a frequency distribution plots the number of occurrences or counts for each value of a single variable. For categorical variables a frequency distribution is displayed in a bar graph, while for numerical variables a frequency distribution is displayed in a histogram.

Bar graphs:

Let's start by looking at the frequency of social problems from the data in the file Appendix D. This file contains a variety of information about a group of individuals, including their score on an ADD test, an IQ test, their GPA in grade 9, whether they were observed to have social problems, and whether they went on to drop out of high school.

- **Open** *appendixd.sav.*
- Select **Graphs**, and **chart builder.**
- At the next dialog box, in the lower left under gallery, select Bar and click on simple bar (the top left icon). Drag and drop this to the top right Chart preview area. You will notice that an element properties box pops up, you can close this.
- Under variables, click on social problems, and drag this to the X-axis in the chart preview. The Y-axis should indicate "count". Click OK. The graph will appear in the output viewer window. We will use this output window throughout this exercise, so don't close it.
- To edit the graph, double click on it and the **Chart Editor** window will open. The Chart Editor has several Menu bars and Icons. Take a moment to review them by clicking on the menus and reading the commands that appear, and putting your cursor over the icons and reading the descriptions that appear to the bottom left hand corner.
- To edit virtually anything on a chart, you can double click it and relevant dialog boxes will appear. Let's try it.

→ Double click the labels under the bars, and a dialog box will appear. Click through all of the tabs to see the various items you can change. Close the dialog box.

→ Double click on the bars and a dialog box will appear. Click through all of the tabs to see the various items you can change. Under "fill & border" change the fill to dark blue. Click apply, then close the dialog box.

→ To edit the title, click in the Axis title, then click again a second time to see a cursor show up. Adjust the axis title so that it is capitalized.

→ Change the title of the Y-axis from Count to Frequency.

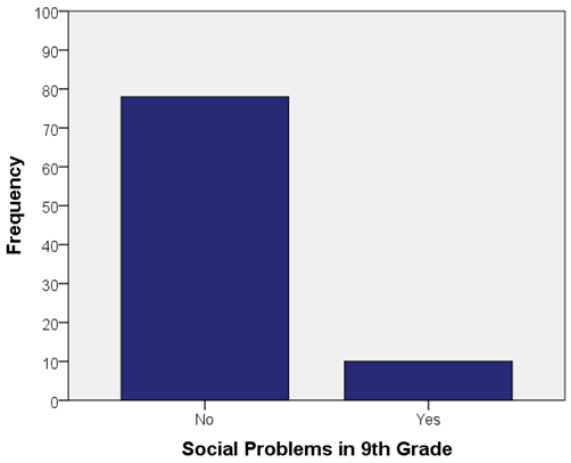
→ To label the actual frequency value inside the bars, click on the bars, then click on the icon for show data labels. Labels should appear within each bar showing the number of observations in each group.

→ Double click on the Y-axis and then the scale tab. Adjust the axis so it goes from 0 to 100, with a major increment of 10. At the bottom of this tab, change the upper margin to 0%.

→ Adjust the font size of your two axis titles to 16 pt, by clicking on one and selecting 16 from the "preferred size" dropdown menu in the editor. Increase the font used in the axis labels to 12 pt.

→ Turn the data labels within the bars back off.

- Check that your final graph looks like this:



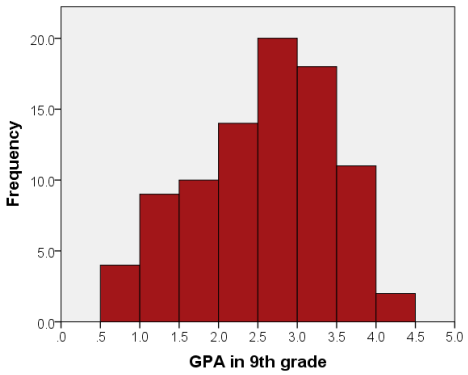
- Return to your data file, and select **Graphs**, and **chart builder** again**.**
- Select simple bar again. Drag and drop this to the top right Chart preview area. This time, do not close the element properties box pops up.
- In the element properties window, highlight "Bar 1" in the top section. In the statistics section, click on the dropdown menu under Statistic (where it currently says count), and have a look at the

other options that you can select. Not all of these will apply to this data set now, but may come in handy later. Select "percentage" and re-make the social problems graph. Have a look at the new chart in the chart output window, and note that you can scroll up to view your previous chart to compare.

Histograms:

Histograms are essentially a frequency distribution for ranges of values rather than individual values. A histogram is a perfect way to display the GPA data from the Appendix D file.

- Select **Graphs**, chart builder, **histogram**, and simple histogram. Drag this simple histogram to the top right chart preview window.
- Under variables, select GPA in 9th grade and drag this to the X-axis in the chart preview window. Click Ok. The chart should appear in your output window.
- Double click on the graph to open the chart editor.
- Double-click on the bars and then select **binning**. Under X-axis, select custom. Note that you can change the number of intervals, or the interval width. Change the number of intervals to 8. Click apply, then close.
- Double click the X-axis, and then adjust the scale. Set the lower and upper margins to 0. Change the major increment to 0.5. Click on number format and change the number of decimal places to 1. Click apply, then close.
- Change the bars to dark red.
- Increase the font sizes so they will be more visible when the graph is shrunk down to a small size.
- Click on the "hide legend" button to hide the descriptive statistics that are shown to the right of the graph.
- Review your histogram and make any other changes you feel would improve its presentation.
- Your final graph should look similar to this:



Boxplots:

Boxplots are useful to illustrate the dispersion of data. They are frequently used with data that compare one numerical variable and one categorical variable. Let's create a boxplot using the same data file.

- Select Graphs, chart builder, box plot.
- Select the first option from the gallery, simple box plot. Drag this to the top right chart preview window. Drag the variable GPA in the 9th grade to the Y-axis. Select the variable "dropped out of high school" and drag this to the X-axis. This will allow us to compare the GPA in ninth grade for those who did and did not ultimately drop out of high school. Click OK.
- Review the box plot, and remind yourself of what the different components of the box plot indicate.
- Repeat the previous step, except this time compare the GPA in ninth grade between males and females by selecting the gender variable.
- Open the chart editor and delete the labels on the outlying points, which correspond to the row that contains the outlying data.
- Using what you have learned in the previous examples, edit the box plot to improve its appearance.

Bar Graphs:

Another way to visually compare the data from different groups is with a bar graph. When comparing one numerical variable and one categorical variable, bar graphs do not show the data as well as other methods (strip chart, box plot, multiple histogram), but they are frequently used in scientific publications. Let's create one from the same example as above so we can compare them.

- Select Graphs, chart builder, bar, and simple bar. Drag this to the top right chart preview window.
- Drag the variables appropriately to compare mean GPA in the 9th grade between those who did and who did not ultimately drop out of high school.
- Note that the chart builder automatically graphs the mean GPA in 9th grade for both groups. Look in element properties window, under "Bar 1" and "Statistic" and review the different statistics that can be graphed instead of mean.
- Click OK to complete the bar graph.
- Notice the biggest difference between this and the boxplot on the previous page is that the boxplot gives you a sense of variability and the shape of the data (i.e. skew).

A **grouped bar graph** is used frequently used to compare two categorical variables. Let's use a grouped bar graph to compare the frequency counts of males and females who dropped out of high school.

- Select Graphs, chart builder, bar, and clustered bar (the second option). Drag this to the top right chart preview window.
- Drag gender to the X-axis, and "dropped out of high school" to the box in the upper right of the preview window where it says "cluster on X". Click OK.
- Open the chart editor. Double click on the bars, then under bar options, reduce the bars width to 85% and the clusters width to 80%. Click apply, then close.
- When there are multiple independent variables, you have to decide which one should be displayed on the X axis and which one should be used as the categories. Let's review the alternate way these data could be graphed. Double-click on the chart to open the properties window. Click on the **variables** tab. From the drop-down menu next to drop out of high school, change from X cluster to X axis. This should automatically change gender to X cluster. Click apply, then close.
- Review the new layout illustrated in the figure.
- Remember from class that grouped bar graphs are typically graphed with bars grouped by categories of the explanatory variable, and different categories of the response variable indicated by different colours or shades. Which graph do you think is best?
- Close the chart editor.

A grouped bar graph can also be used to compare two categorical variables, for the value of a numerical variable. This time, we'll graph the mean grade point average based on both gender and whether or not the student ultimately dropped out.

- Select Graphs, chart builder, bar, and clustered bar (the second option). Drag this to the top right chart preview window.
- Drag mean GPA in 9^{th} grade to the Y-axis, gender to the X-axis, and "drop out of high school" to the box in the upper right of the preview window where it says "cluster on X". Click OK.
- Double click on the graph to activate the Chart Editor.
- Double-click on the bars. Under bar options, decrease the width of the bars to 80% so that they stand apart from one another. Click apply, then close.
- Use what you learned above to edit the graph until you are happy with it. Try selecting the bars so that you can change the colour of one series only.

Scatterplots

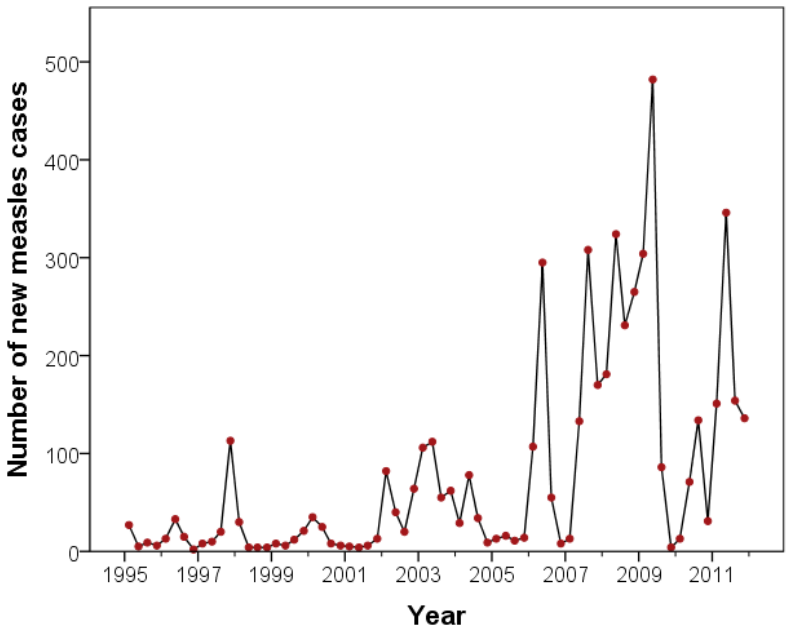
Scatterplots are typically used to visualize relationships between two numerical variables. In Appendix D, we may hypothesize that there is a positive association between IQ and GPA. Let's create this scatterplot to check.

- Select Graphs, chart builder, scatter/dot, and simple scatter (first option). Drag this to the top right chart preview window.
- Drag GPA to the Y-axis, and IQ score to the X-axis. Click OK.
- Next we can add a line of best fit to better illustrate the relationship. Double-click on the chart to activate the chart editor. Click on the menu icon to add a fit line. Ensure that linear is selected.
- Double-click on the data points and change them to solid fill.
- Edit the graph to improve its appearance.

Line Graphs

Line graphs are most commonly used to display differences over time. Let's make a line graph to display the data we looked at in class showing the number of new measles cases over time.

- Open the file *measles outbreak.sav*
- Select Graphs, chart builder, line, and simple line (first option). Drag this to the top right chart preview window.
- On the Y-axis put the variable corresponding to the number of confirmed cases. On the X-axis put the variable for year by quarter.
- Open the chart editor and click on the appropriate menu icon to add markers.
- Change the markers to size 4, with a dark red fill and dark red border.
- For the X-axis, change the number of decimal places to zero. Change the maximum to 2012 and the major increment to 2. Uncheck the "auto" box next to minimum, but leave it set to the year 1995.
- Click on the labels & ticks tab, select the checkbox to display minor ticks, with 1 minor tick per major ticks.
- Change the grey background of the graph to white. Edit the font size and the rest of the graph as needed.
- Verify that your graph looks similar to this one:



- Go to Sakai and answer the questions about your graphs. All of your graphs should still be visible in your output window. You can return to them and double-click on them if needed.

Exercise 2: Saving and applying chart templates

If you do a lot of graphing and develop a style you particularly like, you can save it as a template for future graphs. This can save a lot of time editing.

- Select one of the graphs you made today that you really like. Open it in the chart editor.
- Select file, save chart template.
- Note that you can save either all of the settings, or only various aspects of the settings. Select the box next to "all settings". Then click continue to save the file to your folder.
- The next time you are creating a similar graph you can apply this template so that several adjustments to your graph will be applied at once. Try this now by returning to your data and recreating the graph for which you saved the template. Open this chart, and select file, apply chart template. Verify that the changes were applied.

Exercise 3: Exporting data

- Return to your output window. Select file, export. Under objects to export select all. Verify that Word/RTF file is selected from the dropdown. Change the file name and save location if desired. Click OK. Open this file and scroll through it. All of your charts should be saved as pictures within

this file. The file also contains information about the data sets that were used. Close this file and return to your output window.

- In your output window, select file, export again. Under objects to export select all. From the dropdown select "none (graphics only)". Click OK.
- Navigate to your save folder. You should see several image files, each corresponding to one of the graphs you made this class.
- If you think you may want to work with your graphs more later within SPSS, you can also save the output file itself. In the output window select file, save, and save it with a descriptive name in an appropriate location.

As you can see, SPSS is capable of creating many different chart types, and each type has many options. Furthermore, charts can be edited tremendously after they are created using the Chart Editor. It is important that you are comfortable using SPSS for graphing for the rest of this term. Don't be afraid to play around with graphing the data. Try out several different options along the way, even those that we have not covered.

Additional Practice

ADD Score:

Using the appendix D file, create a **clustered boxplot** to compare the ADD score of males and females who did and did not drop out of high school.

- Answer the question on Sakai about this data.

Diet breadth:

The data in the file "*diet breadth.xls*" are from an ecological study of the rainforest community in El Verde in Puerto Rico. Diet breadth is the number of types of food eaten by an animal species. Import this data into SPSS and visualize the frequency distribution using the appropriate type of graph. Adjust the X axis so it ranges from 0 to 70 with a major increment of 10, and use a binning interval width of 5.

- Answer the two questions on Sakai about this data.

Lab 3 Part 2: SPSS Descriptive Statistics

Lab Objectives:

- Calculate descriptive statistics for continuous and categorical data
- Edit output tables

Exercise 1: Descriptive statistics

Most descriptive statistics are calculated using the **Analyze** menu, followed by **Descriptive Statistics**, and the **Descriptives**, **Frequencies,** or **Explore** commands. When calculating descriptives for more complex designs, including more than one independent variable, you can also use the **Compare Means/ Means** or the **Descriptive Statistics/ Crosstabs** command which allow you to calculate descriptive statistics of subgroups.

It is always important to take a moment to think about the type of data you are using and what descriptive statistics will be most useful given the type. For numerical data, you typically report measures of central tendency and measures of variability. It is often useful to observe the frequency distributions or histograms of continuous distributions to note if they are normal or skewed. For categorical data you typically report the frequency or proportion of each value.

Descriptive Statistics:

Let's begin by calculating descriptive statistics for some of the data in the Appendix D file.

- Open *appendixd.sav*
- Ensure that you are viewing the data labels (e.g. male, female) rather than the value labels (e.g. 1, 2). If the values are visible rather than the labels, click the appropriate button to change this.
- In the **Analyze** menu, select **Descriptive Statistics** and then **Descriptives**.
- Select each of the numerical variables – ADD score, IQ score, Grade in 9th grade English, and GPA in 9th grade – and transfer them to the variables list by clicking the right arrow.
- Click on the **options** button. Put check marks next to Mean, Std. deviation, Variance, and S.E. mean. Uncheck the rest. Note that you can change the display order of your table. If variable list is selected, the variables will be displayed in the order that you added them to the list in the previous window, which is OK for right now. Click **Continue**, and **OK**.
- Double click the table so you can edit it. A formatting toolbar and "pivoting trays" window pops up. Note that the menu at the top of the output windows also changes. As was the case with graphs, SPSS has many options to edit statistics in tables as well. Let's try some of them.
- Close the pivoting trays window. Under the pivot menu, select **Transpose Rows and Columns.** Note the change in your table. However, as this is not the conventional way to display statistics, repeat this again to return to the original orientation.

- Select **Format, Table Looks**. Scroll through the various options for changing the overall look of the table. Many of these contain too much distracting formatting. Select Academic, and click OK.

- Now click on **Format, Table properties**. Take a moment to view all of the options in this dialog box. "General" allows you to specify the width of row and column labels. "Notes" allows you to choose numeric or alphabetic labels, and subscript or superscript as the position for those labels for use as footnotes. "Cell formats" allows you to change the font style and size, color, and the alignment. "Borders" allows you to add or remove borders around rows, columns, and cells. "Printing" allows you to select options such as rescaling tables to fit on paper.

- Under Cell Formats, select "row labels" using the area dropdown menu. Change the text to size 10 and bold. Repeat this for the area "column labels", and also set the column labels to align right. Then select the area "data" and change to size 10, but do not bold. After you have made these changes, click OK.

- Highlight all of the data in your table (all of the numbers within your table), except for the N values. Right-click and select cell properties. Under format value, decrease the number of decimals to 2. Click OK.

- Double click on the row label "Grade in 9th grade English class" and edit the text so it says "English grade" instead. Do the same for GPA in 9th grade so it says only "GPA".

- Highlight the word mean, and change its justification to centre so that it is positioned above both the statistic and the Std. Error value.

- Check that your final table looks like this:

Descriptive Statistics

	N	Mean		Std. Deviation	Variance
	Statistic	Statistic	Std. Error	Statistic	Statistic
ADD score	88	52.60	1.32	12.42	154.31
IQ score	88	100.26	1.38	12.98	168.61
English grade	88	2.66	.10	.95	.89
GPA	88	2.46	.09	.86	.74
Valid N (listwise)	88				

Frequencies:

Now, we'll use the frequencies command to help us examine the distributions of the same continuous variables.

- Select **Analyze/Descriptive Statistics/Frequencies**.

- Select ADD score, IQ score, Grade in 9th grade English, and GPA in 9th grade and transfer them to the variables list by clicking the right arrow

- Uncheck "display frequency tables", as this is not useful for continuous numerical variables. Click OK on the message that pops up.

- Click on Statistics. Under percentile values, select quartiles. Under central tendency, select mean, median and mode. Under dispersion, select std. deviation, variance, and SE mean. Then click continue.
- Click on Charts, and select "histograms", then click continue.
- Click OK.
- Have a look at the frequency table and the histograms that pop up. If necessary, transpose the frequency table so that the variables are presented in rows, and the statistics in columns.
- Look at the distribution for English grade. This histogram triggers your memory and you realize that students received grades of A, B, C, D or F depending on their original numeric mark (e.g. students with marks between 80 and 100 received a grade of A). These letter grades were coded 4, 3, 2, 1 or 0 respectively. This means English grade would be more appropriately categorized as a categorical ordinal variable.
- Click on variable view, and edit the English grade variable to incorporate this new information. Change the measure, and add the above information about the grades using value labels, as you learned in last lab.
- As before, you can edit the tables or the graphs by double clicking on them. One difference we have seen between the **Descriptives** and **Frequencies** options is that **descriptives** only include mean for measures of central tendency, whereas **Frequencies** include the mean, median, and mode. Further, **Descriptives** does not have any built in graphing options, but **Frequencies** does.

Now let's use Frequencies to describe categorical data.

- Select **Analyze/Descriptive Statistics/Frequencies**.
- This time, put gender, repeated a grade, level of English class, Grade in English class, social problems, and drop out status in the variable list.
- Select **Display frequency table**. Since there is a finite number of values for these variables, we want to know how many people fit in every category.
- Click on **Statistics** and unselect all of the options. Then click **Continue**.
- Next, click on **Charts**. Click on **Bar chart** and select **Percentages** as the Chart Values. Click **Continue** and then **Ok**.
- Take a moment to review the output.
- Notice that the frequency tables include a column labeled **Percent** and another labeled **Valid percent**. This is an important distinction when you have missing cases. Valid percent is the percentage out of all the known cases, while percent is the percentage out of total cases. For example, imagine a sample of 100 students. Fifty cases are women, 40 are men, and 10 are

missing the data. The percent of men would be 40%, but the valid percent of men would be 44.4%.

- Notice that the frequency table for grade in English class is organized from F to A. This is because F was assigned a value of 0, while A was assigned a value of 4.

- Return to analyze, descriptive statistics, frequencies and select only Grade in 9th grade English. Click on Format. Select order by "descending values". Click Continue then OK. Check that the table is now presented with the frequencies ordered from A to F.

➢ Go to Sakai and answer the question about the frequency data.

Crosstabs:

Sometimes we need to know the number and percent of cases that fall in multiple categories. This is useful when we have multiple categorical variables in a data set. For example, in the data set we have been using, I'd like to know what percent of dropout and nondropout students had social problems. We'll use crosstabs to calculate this.

- Select **Analyze/Descriptive Statistics/Crosstabs.**

- Select social problems for **Rows** and dropped out for **Columns**.

- Click on **Cells** and select **Observed** for **Counts**, and select **Row**, **Column**, and **Total** under **Percentages**. The click **Continue**.

- Select **Display clustered bar charts**. Then, click **Ok**.

- View the output. Edit the width of the chart columns so that the descriptions appear on one line rather than spread over two. This makes the data easier to see.

- Both the table and the graph show that of those youth with social problems, an equal number did and did not ultimately drop out of high school.

Explore:

Another useful tool for descriptive statistics is the explore option. It is often used to quickly screen a data set. It provides a variety of descriptive statistics for a numerical variable, including **95% confidence interval**. It can be used for one numerical variable alone, or you can view the statistics for this variable according to another grouping/factor.

- Select **Analyze/Descriptive Statistics/Explore**.

- Click on statistics and ensure that the confidence interval for the mean is set to 95%.

- Click on plots. Check histogram, and uncheck stem and leaf if it is selected. Under boxplot, "factor levels together" should be selected.

- The numerical variable you are interested in must go under "dependent list". Add IQ score to the dependent list, and click OK.
- Review the output. You should see a table of descriptive statistics, a histogram, and a box plot.
- Now let's view these statistics for IQ score based on gender. Select analyze, descriptive statistics, explore. Add IQ score to the dependent list, and gender to the factor list.
- Review the output. You should see a table of descriptive statistics for IQ score for males and females separately. There will be separate histograms for males and females, and a box plot illustrating median, quartile data, and any extreme values for males and females.

➢ Go to Sakai and answer the next question.

Compare Means:

Now, let's consider another way to describe a continuous variable at different levels of a categorical variable. Let's begin by comparing ADD symptoms for males and females.

- Select **Analyze/Compare Means/Means**. Notice this is the first time we haven't selected Descriptive Statistics.
- Select ADD score for the Dependent List and Gender for the Independent List. Click Options. Notice that mean, number of cases and standard deviation are already selected under statistics. Look through the other statistics you can add. Add Standard error of the mean, then click continue and then Ok.
- Review the output. It shows you the mean, standard deviation, and standard error of the mean for ADD score, broken down by males and females.

Let's try another more complicated example. This time, let's compare statistics for ADD symptoms broken down by gender **and** whether or not a child had social problems.

- Select **Analyze/Compare Means/Means**.
- Just like before, select ADD score for the Dependent List, and gender for the Layer 1 Independent List. Then click **Next**. Select social problems as the Layer 2 Independent List. Select the same statistics under Options and then click Continue and Ok.
- Review the output. Notice that this table gives you the ADD score statistics for gender independent of social problems under "total", as well as the ADD score statistics for each gender with or without social problems.
- Save the data file in case we want to return to it later.

➢ Go to Sakai and answer the question ADD scores.

Additional Practice

Caffeine in your coffee

The data in the file *caffeine.xls* shows the amount of caffeine in a 16-oz cup of coffee obtained from various vendors. Import this data into SPSS. For context, doses of caffeine over 25 mg are enough to increase anxiety in some people, and doses over 300 mg are enough to significantly increase heart rate in most people. Red Bull contains 80mg of caffeine per serving. Analyze this data using the appropriate descriptive statistics function to view the mean amount of caffeine in a 16-oz coffee and the 95% confidence interval.

View the caffeine data in a histogram. Adjust the scale of the X-axis so it has a lower and upper margin of 0%, a minimum of 140, a maximum of 260, and a major increment of 20.

Now import the data contained in the file *caffeine-starbucks.xls*. This file has data on six 16 oz cups of coffee sampled on six different days from the same Starbucks location. Use SPSS to calculate the mean and standard error for these data.

- Answer the next three questions on Sakai.

Hemoglobin levels

Import the data from *Hemoglobin.xls* into SPSS. This file contains data on blood hemoglobin level from three populations living at high-altitudes (Andes, Ethiopia, and Tibet) and a sea-level population from the USA. Use the appropriate command to view the mean, standard deviation, and standard error of the mean for hemoglobin concentration (g/dL), according to population.

- Answer the last two questions on Sakai.