Fundamental Statistic

By the Author: Dr. Bryan R. Burnham

2.4 Graphs for Frequency Distributions of Quantitative Variables

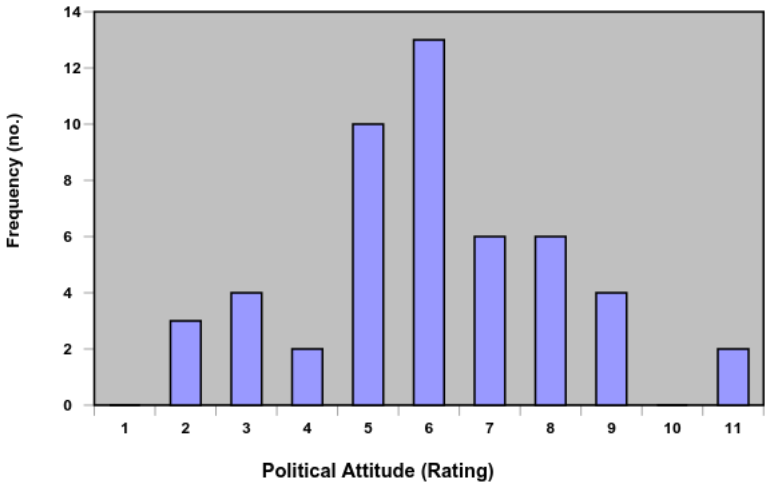
When data are measured on a quantitative variable one of three different graphs are generally used to display the data: **histograms**, **polygons**, and **line plots**. There are some rules for when each of these graphs is used, but which graph type is used is a little less important than displaying data in a clear and simple manner. For each of the following examples, we use the frequency distribution of the Political Attitude data, which is reproduced below:

Political Attitude	11	10	9	8	7	6	5	4	3	2	1
F	2	0	4	6	6	13	10	2	4	3	0

People use the terms *histogram* and *bar graph* interchangeably, but they are not the same graph. Bar graphs are used when the variable on the Abscissa is qualitative, whereas histograms are used when the variable on the Abscissa is quantitative. In bar graphs the bars above each category do not touch, but in histograms the bars do touch. For histogram, each bar is centered above its level on the abscissa and the width of each bar represents real limits around each value. The height of each bar represents the magnitude of the dependent variable, which in this case is the frequency of each value.

To create a histogram set up the ordinate and the abscissa by giving each axis a proper title. Next, label the entries on each axis. Next, extend a column up to a

point above each value on the abscissa that corresponds to the frequency of that value on the abscissa. Again, make sure that the bars touch.

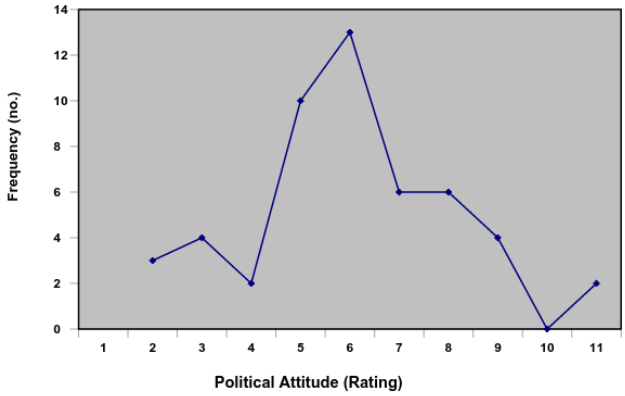


It is important to place the smallest values on the left side of the abscissa nearest the ordinate. The point where the abscissa and the ordinate meet, called the **origin**, represents the smallest values. This way, as distance increases from the origin on either axis, it indicates greater and greater values.

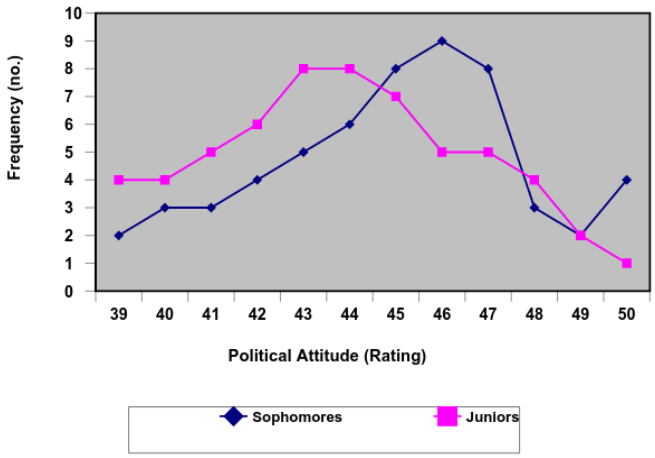
A polygon uses the height of lines between data points to represent the magnitude of a dependent variable, instead of columns as in bar graphs and histograms. Polygons are used when the variable on the abscissa is quantitative and is general used when the variable is continuous. Remember from Section 2.2, continuous data is where there are theoretically an infinite number of possible measurements between any two points; hence, fractional values are possible. The use of a line to connect data points in a polygon graph is supposed to represent continuity between the data points. The graph below presents the Political Orientation data as a polygon.

You'll notice that I added a score of 12 to the abscissa and gave it a frequency of zero. This is because frequency polygons are always closed on the abscissa where

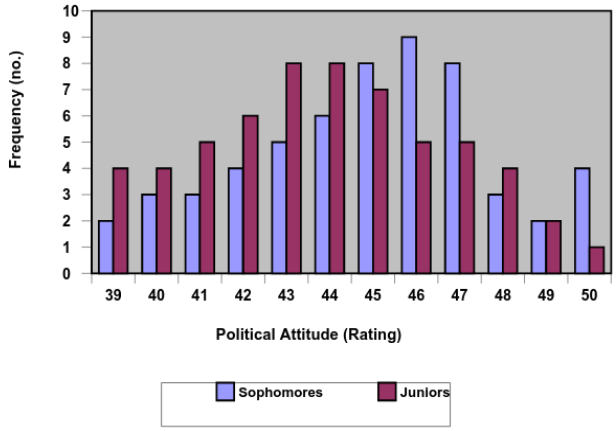
the line extends from the lowest value with a frequency down to the abscissa. In this case, the rating of 1 had a frequency of zero, so the left side of the graph was already closed, but the rating of 11 had a frequency of 2, so an additional value above 11 and with a frequency of was needed in order to "close" the polygon. Similar to the polygon is the **line plot**, which is simply a frequency polygon that is not closed.



Line plots and histograms are also for presenting data from several groups within the same graph, that is, if several groups were measured on the same variable, these groups' data can be plotted within the same graph. For example, say we measure the time, in minutes, it takes students to complete an exam. We plot the data by sophomores and juniors, which are the two college classes in this particular course.



Here's a histogram showing the frequencies of sophomores and juniors who completed.



2.5 Graphs for Frequency Distributions of Qualitative Variables

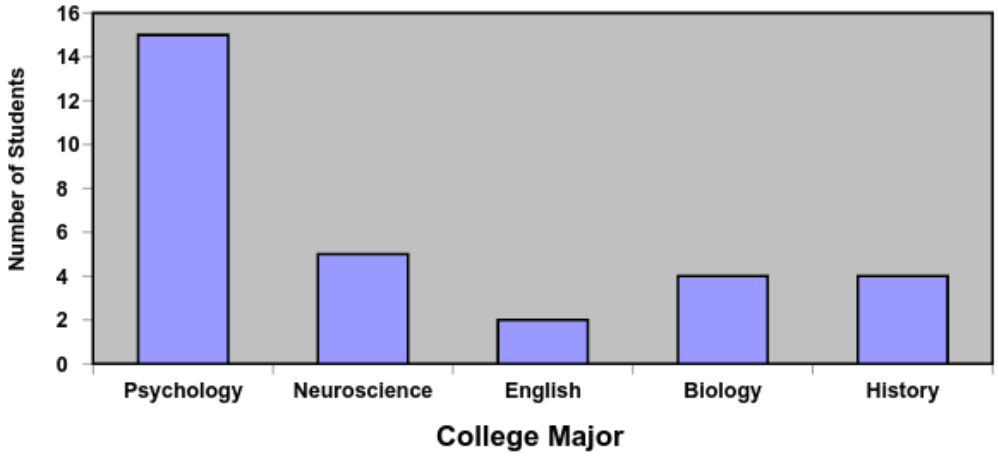
If the variable displayed on the abscissa is qualitative, a **bar graph** is used. Bar graphs and histograms look nearly identical, but histograms are used when the variable on the abscissa is quantitative and bar graphs are used when the variable on the abscissa is qualitative. Also, when using a histogram the bars touch, but with bar graphs the bars do not touch. This is supposed to indicate a qualitative (categorical) difference between levels of the independent variable. An example of a bar graph of the data is below:

Say we want to know the numbers of each college major in a particular class with 30 students. We ask each student his/her primary major and list each major with the number of students in that major. In this case, each 'value' for the variable College Class is qualitative and the data are on a nominal scale. When creating a frequency distribution for qualitative data, most steps are identical to those when creating a frequency distribution with quantitative data:

College Class	Psychology	Neuroscience	English	Biology	History
f	15	5	2	4	4

n = 30

The only difference between frequency distributions for quantitative data and for qualitative data is you do not include cumulative frequency, cumulative relative frequency, and cumulative percentages for qualitative data.



Homework

A. Use the frequency distribution table below to complete exercises:

Score	*f*	*rf*	*cf*	*crf*
5	5	.2	25	1.00
4	3	.12	20	.8
3	7	.28	17	.68
2	5	.2	10	.4
1	3	.12	5	.2
0	2	.08	2	.08

1. Create a histogram for the absolute (raw) frequencies.
2. Create a line plot for the absolute (raw) frequencies.
3. Create a line plot for the relative frequencies.
4. Create a line plot for the cumulative frequencies above.

5. Create a polygon for the cumulative relative frequencies above.

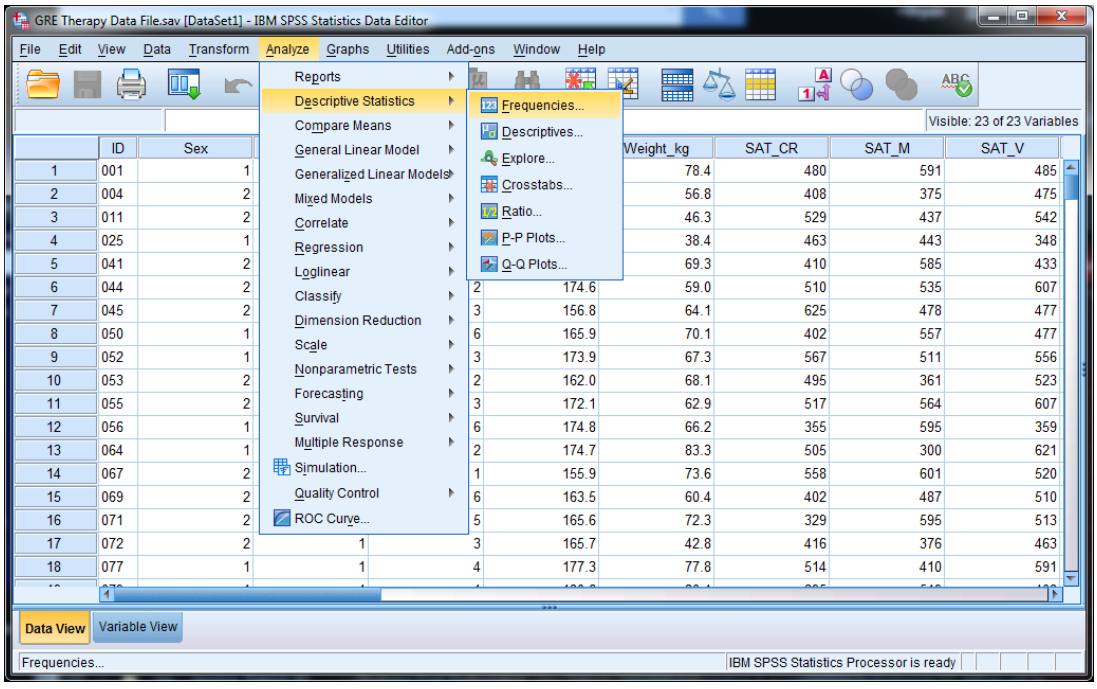
B. A principal in a small school measured the intelligence of fifth-grade students in her school. The intelligence test scores for those students were as follows:

129	99	98	113	103	128	102	110	80	105
93	98	109	109	100	111	106	96	108	90
104	94	92	119	127	89	95	92	105	108
83	100	107	106	101	118	84	119	105	111
118	106	122	120	102	117	103	117	103	88

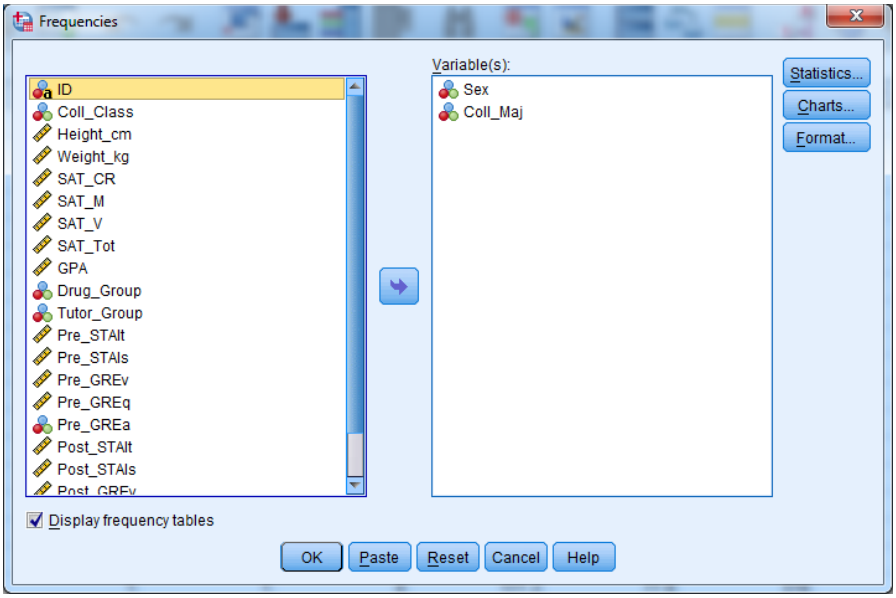
Calculatr the (*f*, *rf*, *cf*, *crf*). Draw a stem and leaf plot for the set of scores.

2.6 Frequency Distributions in SPSS

The following uses the GRE Therapy Data file. This data file is based on a hypothetical study examining the influences of a study-aid drug and types of tutoring on performance on the Graduate Record Examinations (GREs). To request frequency distributions in SPSS, from the Analyze menu, select Descriptive Statistics and the Select Frequencies (see below):



Assume we want to know the frequencies for each of the college majors (Coll_Maj) in the data set and the frequencies of the males and females (Sex); two qualitative variables. In the Frequencies window that opens, move Coll_Maj and Sex from the left to the right and click he OK button.



The resulting output is below, where it can be seen that there were 109 males (45.4% of the subjects) and 131 females (54.6% of the subjects). Additionally, you can see there were students from six different college majors in the data set, with each major represented roughly the same among the n = 240 students in the sample.

Frequency Table

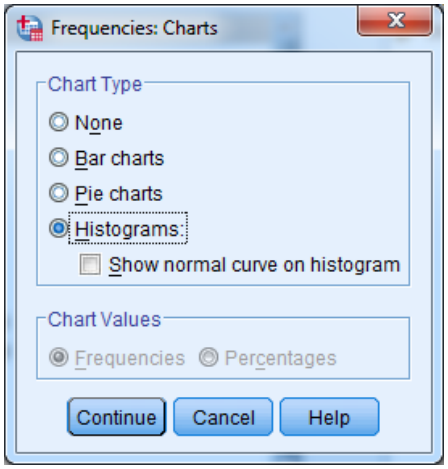
Sex

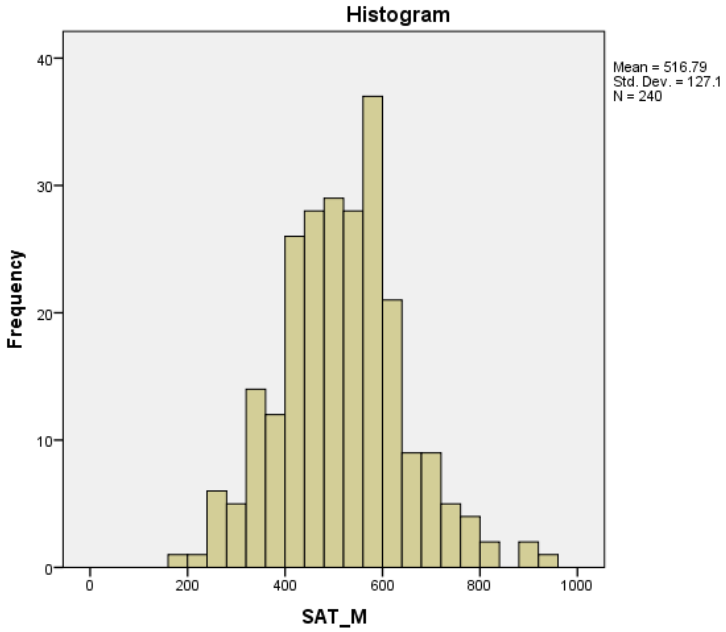
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Males	109	45.4	45.4	45.4
	Females	131	54.6	54.6	100.0
	Total	240	100.0	100.0	

Coll_Maj

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Psychology	39	16.3	16.3	16.3
	History	40	16.7	16.7	32.9
	Biology	40	16.7	16.7	49.6
	Communications	30	12.5	12.5	62.1
	English	38	15.8	15.8	77.9
	Mathematics	53	22.1	22.1	100.0
	Total	240	100.0	100.0	

If you want to ask for frequencies from a quantitative variable, such as SAT Math scores (SAT_M), the procedure is the same; however, it is good to also request a histogram of the data (click the Charts button and select Histograms, see right), because quantitative variables can often take on many values and the frequency of each value will be low. The histogram will "group" adjacent values together to show you the frequencies are similar values, making it much easier to interpret the frequencies (see below).




Histogram
Mean = 516.79
Std. Dev. = 127.1
N = 240
Frequency
SAT_M
0
10
20
30
40
0
200
400
600
800
1000