TAKE IT TO THE MAT

A NEWSLETTER ADDRESSING THE FINER POINTS OF MATHEMATICS INSTRUCTION

Southern Nevada Regional Professional Development Program

May 2003 — Middle School Edition

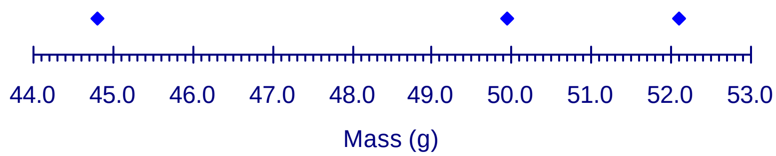
rpdp.ccsd.net

In the March and April editions of *Take It to the MAT*, we graphed numerical data using line plots, stem-and-leaf plots, and histograms. In this last issue of the 2002–2003 school year, we will look at one other graph to display numerical data, the *boxplot* or *box-and-whisker plot*.

Masses (g)		
44.8	49.5	50.3
47.3	49.7	50.5
48.1	49.7	***50.7***
48.3	49.8	50.8
48.5	**49.9**	51.2
48.7	**50.0**	51.2
48.7	50.1	51.4
49.0	50.1	51.5
49.2	50.2	51.8
49.2	50.2	**52.1**

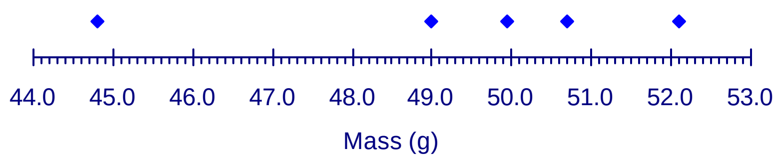
The table at right shows the masses of 30 bags of colored candies, the same data we used in the April issue. We will begin by summarizing the data with three numbers: the minimum value, the maximum value, and the median. The minimum and maximum masses are easy to find, 44.8 g and 52.1 g respectively. The median is the middle value in the data set when sorted in numerical order. Since there is no one middle value, we will average the 15th and 16th observations to find the median. The median is thus, $\frac{49.9\text{g} + 50.0\text{g}}{2} = 49.95\text{g}$. So far, we have what is commonly called the three-number summary: minimum = 44.8 g, median = 49.95 g, maximum = 52.1 g.

These three numbers give us an idea of how the data is cut in half. Half of the observations are between 44.8 g and 49.95 g, (in red) half are between 49.95 g and 52.1 g (in green). The lower half seems more spread out than the upper half. Let's plot those values on a number line to see the spread visually.

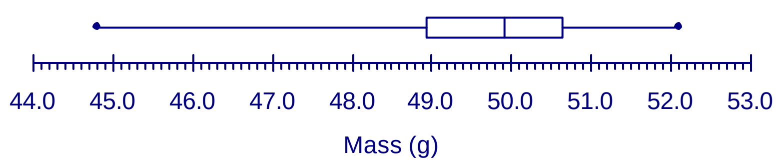


Now that the data is cut into halves, we will divide each of the halves in half again. We do that by finding the middle value of the observations below the median and the middle value of the observations above the median. For the lower half, the middle value of the fifteen observations from 44.8 g to 49.9 g is 49.0 g. The median of the upper half is 50.7 g. The values 49.0 g and 50.7 g are known, respectively, as the first and third *quartiles*. Quartiles divide the data into four equal-sized groups. (What is the second quartile?) When the quartiles are combined with the minimum, maximum, and median, the *five-number summary* results. The five-number summary of our data set is 44.8 g, 49.0 g, 49.95 g, 50.7 g, and 52.1 g.

Plotting the first and third quartile on the number line, we now have the graph at right. Each of the distances between consecutive points contains one-fourth of the data. Just as the median divided the data in half, the quartiles cut it into quarters. Thus, one-fourth of the data are between 44.8 g and 49.0g, one-fourth are between 49.0 g and 49.95 g, and so on.



When we have these five points on the number line, we make a rectangle with ends at the first and third quartile, then divide it in two pieces at the median. Then, we draw line segments from the ends of the rectangle to the minimum and maximum. The result is the boxplot, or box-and-whisker plot.



The purpose of a boxplot is to get a quick visual impression of the *center*, *spread*, and *shape* of the data. In the case of our data, it is centered at about 50 g; half of the data is greater and half is less. The middle half of the data ranges from about 49.0 g to 50.7 g. The lower half of the data is much more spread out than the upper half, but the 2nd and 3rd quarters of the data (the left and right portions of the box) show a roughly equal spread.

The boxplot of a single data set doesn't reveal a lot more information that a line plot, stem-and-leaf plot, or a histogram. Its real power is in comparing two or more sets of data. That's a topic we'll address next year.