TAKE IT TO THE MAT

A NEWSLETTER ADDRESSING THE FINER POINTS OF MATHEMATICS INSTRUCTION

Southern Nevada Regional Professional Development Program
April 2003 — Middle School Edition

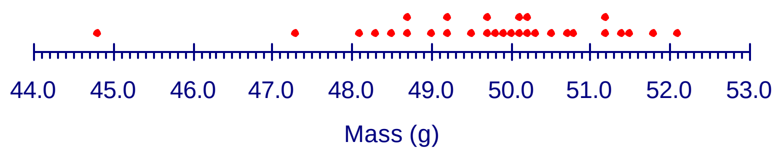
rpdp.ccsd.net

In the March edition of *Take It to the MAT*, we looked at one method of graphing numerical data: *the line plot* or *dot plot*. The line plot is especially useful when data sets are reasonably small and *discrete*, that is, the data may only take on certain values. For example, in a bag of colored candies one may get 19 or 20 or 21 candies. One may not get 19.5 candies or 20.432 candies. This newsletter addresses what to do when data is not discrete.

Numerical data that is not discrete is called *continuous*. Think of the possible values of data as being on a continuum—no matter what two values one considers, there is always another one in between. Continuing the colored candies example, the net mass of the candies is a continuous numerical variable. One bag may have a mass of 49.3 g, another 49.4 g, and it is certainly possible that a bag having a mass of 49.37 g could exist. This is not to say every value is *possible*. Masses cannot be negative, nor is it likely a small bag of candies would have a mass of 1000 grams—there are limits on how much candy will fit in a paper wrapper!

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 masses were found using an electronic balance that displays the mass to tenths of a gram.) Once we have the raw data, our first task is to graph it. We could make a line plot; it is shown below the table.



The line plot tells us some things, but not everything. While we can visually see the overall spread of the data, a concentration around 50 g, and that there is a potential outlier, we can't really see its shape. Of the thirty data values, twenty-four are unique! If we had measured to the hundredth of a gram, there may not have been any duplicates.

What we need to do is organize the data in some way that allows us to see its shape. One way to organize numerical data with a lot of values is with a *stem-and-leaf* plot, or *stemplot*. Each datum has a stem, a value that indicates how the data will be grouped, and a leaf, the remaining part of the datum. In this example, our stems will be the whole numbers of grams and the leaves will be tenths of grams. A vertical line separates the stems and leaves and the leaves are spaced equally. The final stem plot is shown at right. (Note the proper inclusion of a key.)

44|8
45|
46|
47|3
48|13577
49|02257789
50|011223578
51|22458
52|1

Key:
Stem: Units
Leaf: Tenths
44|8 represents 44.8 g

The stem plot indicates a mound-shaped distribution. Most of the data is concentrated in the range of 49.0–50.8 grams and nearly all of it is between 48.1–51.8 grams. There is a possible outlier at 44.8 grams. To see the shape, turn the paper 90º counter-clockwise.

The stem plot is actually a form of a *histogram*, a type of bar graph used to display numerical data where the data are grouped into intervals called *classes*. In this case, the classes are from 44.0–44.9 g, 45.0–45.9 g, etc. The histogram shows the frequencies of the data in each class, reveals its shape and outliers, but does not retain the original data values like the stem plot. A histogram of our data is shown at right. Values on the borders of the bars fall into the bar to the right, i.e. 49.0 grams is in the highlighted bar that ranges from 49.0–50.0 grams.