TAKE IT TO THE MAT

A NEWSLETTER ADDRESSING THE FINER POINTS OF MATHEMATICS INSTRUCTION

Southern Nevada Regional Professional Development Program

March 2003 — Middle School Edition

rpdp.ccsd.net

This issue of *Take It to the MAT* continues to explore the topic of data. The January and February issues focused on *categorical* or *qualitative* data, data that can be described by some attribute or be labeled in some way. We used the example of charting the colors of individual candies in a bag of candy. In this issue, we will continue using candies to examine *quantitative* or *numerical* data.

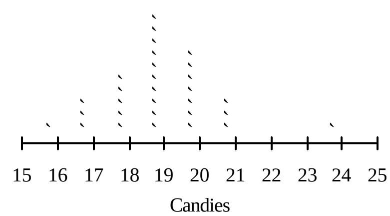
Numerical data is data where the variable of interest comes from counting or some measurement. The example we will use this time is not a single bag of candies, but several bags of candies. Imagine that each student in a class has a bag of the same kind of candies and we are interested in how many candies are in each bag. Each student opens a bag and counts how many candies are in it. The data for a class of 30 is shown in the table at right and has been sorted for clarity.

Candies		
16	19	20
17	19	20
17	19	20
17	19	20
18	19	20
18	19	20
18	19	21
18	19	21
18	19	21
19	20	24

How are the data distributed? What is the most common number of candies? What is the range of the candies one can expect to get in a single bag? Are there any unusual data?

One may answer these questions by calculating the mean, median, mode, range, or some other value. Let's resist that impulse to make calculations. The first thing to do when one has data is to draw a picture of it—make a graph.

Let's start with the easiest of all graphs of numerical data, the *line plot*, sometimes called a *dot plot*. We start with a number line that includes the entire spread of the data. Since the data ranges from 16 candies to 24 candies, we'll use that section of the number line. Place an "X", or a dot, or some other symbol to represent a single bag of candies. (Little post-it notes work on a whiteboard work very well.) As discussed in previous issues, it's important to maintain equal spacing of the x's or dots. If equal spacing isn't maintained, a false impression of the relative frequency of the candy counts will occur.



Let's go back to our original questions. How are the data distributed? What is the most common number of candies? What is the range of the candies one can expect to get in a single bag? Are there any unusual data? These questions can much more easily be answered by looking at the graph than the table.

The general principle is to describe the center, spread, and shape of the distribution, and look for anything unusual or interesting. In the graph above, the distribution is roughly *mound-shaped* and centered around 19 candies, which is also its most common value. The data spreads out from 15 candies to 24 candies, and 24 appears to be an *outlier*. We can expect to get 17 to 21 candies from a single bag. Those values account for nearly all the observations.

What can one do with this information? When might this information be useful? How did the outlier occur? What questions might one ask the candy manufacturer? These are several of the many questions that can be posed once the data is collected and graphed. Can you think of others?

(Note: There are inconsistencies in the naming of graphs. Some texts and programs do not illustrate a *line plot* as it is in this newsletter, but as what others call a *line graph*, a series of points connected with line segments. It's not necessary to be overly concerned about these differences in terminology. It's more about how we display data and what it tells us than what the graph is named.)