TAKE IT TO THE MAT

A NEWSLETTER ADDRESSING THE FINER POINTS OF MATHEMATICS INSTRUCTION

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

rpdp.ccsd.net

Welcome back to school! This issue of *Take It to the MAT* begins the fifth year of the Regional Professional Development Program's newsletter on mathematics instruction. Good luck in the upcoming school year. Let's get started!

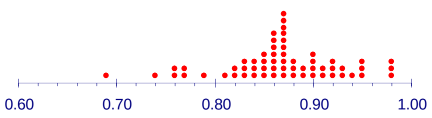
In the January through May 2003 issues of *Take It to the MAT*, we looked at how to construct various types of graphs for categorical and numerical data. The goal was always to get a graph that gives a clear, concise, purposeful, and accurate picture of the data. In this and upcoming editions, we'll address using graphs to make comparisons between sets of data.

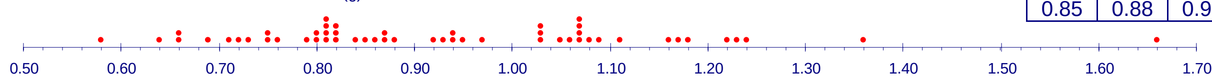
Table 1 lists the masses of the fifty-three individual candies from a single bag of "crispy" chocolate candies. Table 2 contains the masses of sixty "plain" chocolate candies. How can we compare these two groups of data graphically?

In order to answer that question, we must consider what types of graphs would be appropriate for each individual set of data. Since the data is numerical, we could use any one of a line plot, a stem-and-leaf plot, a histogram, or a box (-and-whisker) plot.

Table 1: Crispy Masses (g)		
0.58	0.82	1.05
0.64	0.82	1.06
0.66	0.82	1.07
0.66	0.84	1.07
0.69	0.85	1.07
0.71	0.86	1.07
0.72	0.87	1.08
0.73	0.87	1.09
0.75	0.88	1.11
0.75	0.92	1.16
0.76	0.93	1.17
0.79	0.94	1.18
0.80	0.94	1.22
0.80	0.95	1.23
0.81	0.97	1.24
0.81	1.03	1.36
0.81	1.03	1.66
0.81	1.03	

Table 2: Plain Masses (g)		
0.69	0.86	0.89
0.74	0.86	0.89
0.76	0.86	0.90
0.76	0.86	0.90
0.77	0.86	0.90
0.77	0.86	0.90
0.79	0.86	0.91
0.81	0.87	0.91
0.82	0.87	0.92
0.82	0.87	0.92
0.83	0.87	0.92
0.83	0.87	0.93
0.83	0.87	0.93
0.84	0.87	0.94
0.84	0.87	0.95
0.84	0.87	0.95
0.85	0.87	0.95
0.85	0.88	0.98
0.85	0.88	0.98
0.85	0.88	0.98





Whether looking at a single set of data, or comparing several sets, we must keep in mind to look for certain characteristics of the data: center, spread, shape, and outliers. Also, if we are to accurately compare multiple groups, we must use the same scale on our graphs. If the two plots are not scaled similarly, it is more difficult to compare and contrast characteristics of the data.

Let's first look at line plots of the two data sets. The most striking characteristic is the difference in spread. The crispy candies' masses are clearly more spread out than the plain candies'. The range is almost four times as large for crispy (1.08 g) as the plain (0.29 g). The plain candies' masses appear to be centered at about 0.86 grams; there are about as many candies with smaller masses than that value as there are larger masses. The crispy candies appear to have a median mass of about 0.95 grams. We could find the exact median by looking at the data table, but for now we're just getting an estimate from the graph. By looking at the two graphs' centers we see that the crispy candies tend to have greater mass, on average, than the plain.

When considering outliers in the data, the crispy candy with mass 1.66 grams sticks out like a sore thumb. What happened there? Outliers should be explained, if possible. Was it really that massive? Or were two smaller candies melted together? Did we record the mass incorrectly? When outliers are present, investigation is warranted.

Lastly, let's look at shape. The distribution of plain candy masses is fairly symmetrical about its center. The crispy's distribution is much more difficult to see because of the great spread of the data (1.08 g) with respect to the precision of our measurement (0.01g). Of the 53 data values, 38 are unique; no value occurs more than 4 times. Hence, line plots are problematic if many data values are possible and also occur. Consider that when measuring to the hundredth of a gram, there are over 100 possible values over the range of our data and we have only half that many actual observations.

So, the line plot tells us some things, but not everything. Next time, we'll look at another way to examine this data: the stem-and-leaf plot. Maybe it will tell us more than the line plot.