TAKE IT TO THE MAT

A NEWSLETTER ADDRESSING THE FINER POINTS OF MATHEMATICS INSTRUCTION

Southern Nevada Regional Professional Development Program
February 2004 — Middle School Edition

rpdp.ccsd.net

We have spent a considerable amount of ink during the last several months comparing distributions of numerical data using line plots, stem-and-leaf plots, and histograms. In this issue of *Take It to the MAT*, we will conclude our discussion of comparing numerical distributions with *box-and-whisker plots*.

People often ask, "What is a box-and-whisker plot good for? Where is it used? Why have I never seen one?"

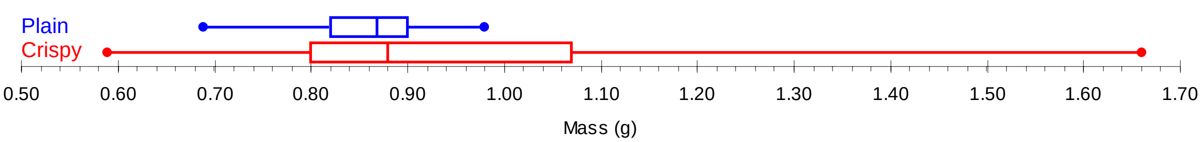
Those are valid questions. It's true that we don't see box-and-whisker plots, also called *boxplots*, in *USA Today*. That's because they are a fairly young tool in data analysis, being created in the early 1960's by John Tukey. It's also because their power is not in picturing single data sets, but in comparing two or more sets of numerical data. Let's use boxplots to compare the distributions of masses of our plain and crispy candies.

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**

First, we need to find the five-number summary for each of the data sets. (For more details, see the May 2003 issue of *Take It to the MAT*.) For the crispy candies, the five-number summary is (0.58 g, 0.80 g, 0.88 g, 1.07 g, 1.66 g); for plain candies, (0.69 g, 0.84 g, 0.87 g, 0.90 g, and 0.98 g).

Now, construct boxplots for the two distributions on the *same scale*.



The detail of the data is lost. All we can see are the extremes (minimum, maximum), the median, and the quartiles. But what those 5 numbers tell us is very useful. The medians are almost equal. On average, the masses of plain and crispy candies are pretty close. The spread is a different matter. Overall, crispy candy masses are much more variable than the masses of plain candies. Even when comparing the middle half of each group—those observations within the "box"—we can see that crispy candies have greater variability. Lastly, the masses of plain candies seem to be distributed fairly symmetrically while the crispy candy masses appear to be right-skewed.

The boxplot is a simple graphical representation whose power lies in its use for comparing distributions of numerical data. It filters out the clutter of the raw data, is useful for data sets of different sizes (as we have here), and quickly indicates center, spread, and shape.

For even more on the power of comparative boxplots, see the November 2003 issue of *Take It to the MAT — High School Edition*.