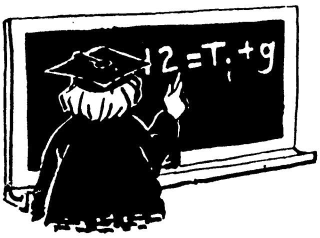

CLASSROOM CAPSULES

Edited by
Frank Flanigan

A Classroom Capsule is a short article that contains a new insight on a topic taught in the earlier years of undergraduate mathematics.

Please submit manuscripts prepared according to the guidelines on the inside front cover to:

Frank Flanigan
Department of Mathematics and Computer Science
San Jose State University
San Jose, CA 95192

Using Median Splits to Motivate Learning

David P. Doane, Oakland University, Rochester, MI

"I know the mean is 14.0 and the standard deviation is 2.4, but what does it *tell* me?" Such a comment reveals the need for concrete reference points. A few statistics (e.g., skewness) can be compared with known distributions, and a standardized sample can be compared with normal percentiles. But the *median split* offers a simple way to make a univariate statistics assignment both more meaningful and more interesting. The idea is to divide the sample on Y into two subgroups based on values of another variable X suspected of being causally related to Y. Students are then asked to compare central tendency, dispersion, and shape for these two subgroups.

For example, to investigate whether infant mortality rates [*Hammond Almanac*, 1982, p. 271] are related to per capita income [U.S. Bureau of the Census, *Statistical Abstract of the U.S., 1982–1983*, p. 429] we divide the fifty states into two equal groups, those above and those below the median income. Then, we split infant mortality rates into two groups corresponding to the low-income and high-income states and prepare simple summary statistics for the two groups. We present our descriptive statistics side-by-side, as shown in Table 1, rounding heavily so the presentation is effective [A. S. C. Ehrenberg, The problem of numeracy, *American Statistician* 35 (1981) 67–71].

Table 1. Infant Mortality Rates Per 1,000 Births

	Low-Income States ($n = 25$)	High-Income States ($n = 25$)
Mean	14.0	12.9
Standard Deviation	2.4	1.2
1st Quartile	12.2	11.9
2nd Quartile	14.1	13.0
3rd Quartile	15.8	13.8
Skewness	0.2	0.2
Kurtosis	2.3	2.4

Students quickly learn to interpret such a table. For example, they would note that the samples have similar shapes but the high-income states have lower infant mortality rates and exhibit less variation. They might infer that differences of this magnitude would be hard to detect in practice, and that many poorer states have lower infant mortality rates than richer states, despite the difference in means. Such insights are a first step toward recognizing "important" as opposed to "significant" differences, and toward distinguishing variation within samples from variation between samples.

Simple histograms or back-to-back stem-and-leaf plots (see Figure 1) tell the novice that no high-income state has an infant mortality rate exceeding 16, while six of the low-income states do. Differences in modality and dispersion are apparent without any formal tests.

Low Income States (Income < $9333)		High Income States (Income > $9333)
7 6	18	
3	17	
6 4 1	16	
4 1	15	7
8 8 3 1 1	14	0 3 4 7
5 5 1 1	13	0 0 0 2 3 6 7 8 8
9 7	12	0 5 5 5 6
7 6 4	11	1 1 2 2 6 8
4 4 4	10	

Figure 1. Stem-and-Leaf Plot of Infant Mortality Rates (15|7 represents infant mortality of 15.7 per 1000 births)

Although descriptive statistics precedes inference, the first contact with real data can motivate subsequent topics since median splits invite questions such as whether observed differences are significant or how to explain variation within samples. It is only a short step to formal two-sample tests such as Student's t or Kruskal-Wallis, or to demonstrate the equivalence of the t-test for two means and regression on a binary (a new variable that is 0 for states below the median income and 1 for states above the median income). The search to explain variation in infant mortality leads a natural thought progression toward a multivariate regression model, starting right from the first assignment.