ST 516: Foundations of Data Analytics

Resistance to Outliers

Resistance
    The Sample Median
    t-based Procedures

The Sample Median

A statistical procedure is said to be **resistant** to outliers (or simply, resistant) if it doesn't change very much when the values of one or a few observations change.

- You may have learned that the sample median is resistant to outliers. Consider the following example:

2.0	2.3	2.4	3.1	3.2	3.4	3.5	4.3	4.4	37

  Notice that the number 37 is substantially different from the other numbers—perhaps this was a data entry error, and the number should have been recorded as 3.7.

Example, continued

The median of the 10 numbers from the previous page is 3.3 (the average of the 5th and 6th observations when the observations are placed in increasing order).

If, in fact, the 37 should have been 3.7, the median of the corrected dataset remains 3.3.

By contrast, the sample mean is not resistant to outliers—in the same example, with 37 in the dataset, the sample mean is 6.6; without it, the sample mean is 3.2.

The t-based Procedures

Since the t-based procedures are based on sample means, they are not particularly resistant to outliers.

- We would like to emphasize that if a statistical procedure is not resistant to outliers, that's not necessarily a bad thing.

- If you believe that an outlier really is a legitimate observation from a particular population, then your analysis should reflect that.

- The next slide provides some guidelines for handling outliers.

Statistical Sleuth Display

Display 3.6   **Examination Strategy**   **p. 66**