Exploration 6A: Data Summaries and Sensitivity

Open the data file C06 ExplorationA.xls. This file contains 1000 observations from five random distributions of data. Each data set contains values in the range of 0 – 100. In this exploration, you are going to investigate two aspects of the data. You will look at the error in estimating the mean and standard deviation of data from different distributions (symmetric, positively skewed, negatively skewed, bimodal and uniform) and you will explore how these errors are affected by the way the data is summarized.

1. Compute the actual values for the mean and standard deviation of each of the distributions (symmetric, positively skewed, negatively skewed, bimodal, and uniform).
2. Below are two different summaries of the data into frequency tables, one using bins of width 10, and the other with width 25. Reproduce these summary tables using your software.
3. 3. Use the techniques of this section to estimate the mean and standard deviation from each of the data summaries. Enter the results in the table below.
4. Once you have recorded the results of your calculations in the table below, think about how our assumptions work when estimating the mean and standard deviation. For which types of data are these estimates most accurate? Why? For which are the estimates least accurate? Why? Keep in mind, these errors may seem small; typically the most error you get with these data is about 0.5 to 1.0, but that's about a 1% to 2% error in estimating these statistics! How does the bin width affect the accuracy of the estimates?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Symmetric | Positively Skewed | Negatively Skewed | Bimodal | Uniform |
| Actual Mean |  |  |  |  |  |
| Mean (bin width = 10) |  |  |  |  |  |
| Mean (bin width = 25) |  |  |  |  |  |
| Actual St Dev |  |  |  |  |  |
| St Dev (bin width = 10) |  |  |  |  |  |
| St Dev (bin width = 25) |  |  |  |  |  |

