How Many Textbooks??

During World War II, Allied forces noticed that all German tanks were labeled with a serial number. After observing numerous tanks, it became obvious that the tanks were numbered systematically. Thus, the allies were able to predict the total number of German tanks and their locations using the observed serial numbers. This became known as the German Tank Problem.

In this assignment, your job is to develop a method to estimate the total number of AP Statistics books (or any other textbook) at our school based on a random sample of books and the assumption that the books are numbered sequentially starting from 1.

For example, suppose a random sample of n = 7 books gives the following numbers: 10, 38, 59, 61, 74, 90, 94. How can you use this information to estimate the total number of books (N)? One possible method would be to take the median value and double it. In this example, median • 2 = 61 • 2 = 122. Thus, our estimate for the total number of books is N = 122.

For this activity you will create three other statistics to estimate the total number of books. Remember, a statistic is any quantity computed from the values in a sample. You may use any combination of the summary statistics we already know (mean, median, min, max, quartiles, IQR, standard deviation, etc.) or invent your own. The goal is to find a relatively *simple* statistic that reliably predicts the total number of books.

To determine which of your three statistics gives the best predictions, you will perform a simulation to generate sampling distributions for each of your three statistics. For the purposes of the simulation, assume that there are 100 books total (N = 100) and that you will be taking samples of size 7 (n = 7). For each run (do 25 runs), generate 7 random numbers from 1-100 and compute the values of each of your three statistics.

What you need to turn in:

- An introduction describing the choice of your 3 statistics
- Three dotplots showing the sampling distributions for each of your statistics. Make sure they are on the same scale so they can be easily compared.
- A verbal description/comparison of the three distributions.
- A description of which statistic you think is best and why.