Session S.2.2
Floyd Bullard

The German Tank Problem

This is a classic problem that has been used by high school and college teachers for years to introduce students to the idea of sampling variability. It is inherently engaging, it is based on a real historical situation, it requires students to think creatively and logically, and it introduces not only sampling variability but also the ideas of bias, low variability, and robustness, and how we judge sample statistics.

During World War II, the Allies had spies in the field who estimated the number of tanks the Germans had based on their observations. The Allied forces were later able to capture a small number of German Mark V tanks, and it was discovered that they had serial numbers on them that almost surely were part of a consecutive series from the same manufacturing plant. British mathematicians used the serial numbers to estimate the number of Mark V tanks the Germans had. After the war, it was learned that the mathematicians' estimate was much better than that of the spies, for the Germans had been repainting their tanks regularly to make their numbers appear greater.

In this classroom activity, students are to take the role of the British mathematicians and develop a method for estimating the number of tanks that are in a population given a random sample of "serial numbers" drawn from the set {1, 2, 3, …, N}, where N is the unknown population parameter to be estimated.[1] Groups of students should draw 7 numbers from a bag of numbers for which the teacher knows the population value. (In the graphs that follow, N=344 was used because that's how many squares I happened to cut out of the cardstock paper I had available.) Alternately, the teacher may use the random number generating feature of a calculator to generate sets of numbers for students. *e.g.*, on the TI-83, one may enter randInt(1,344) and hit enter seven times to generate a list of numbers on the screen that may then be shown to students. I find it preferable to have students actually draw numbers from a bag because it gives them the experience of random sampling. Calculator technology is exceptionally useful but should be avoided if there is a hands-on alternative whenever there is a danger of the technology becoming mere button-pushing. However, drawing numbers from a bag does take longer.

Students should be instructed that after drawing their seven numbers (and returning them to the bag) they are to come up with an estimate of N, the number of "tanks" in the population, and also a clear statement of the method they used. Their method should be applicable to any set of seven randomly selected numbers. Groups that work quickly should be encouraged to come up with a second or even third method and estimate.

---

[1] The actual population of tank serial numbers during WWII had an unknown upper *and lower* bound. The one parameter problem is more useful in the classroom.

After students have had sufficient time to develop their methods with the teacher moving around to assess their progress, their methods and estimates should then be put in a table on the board. It is not unusual for different groups to come up with the same method (*e.g.*, twice the sample mean is a common method). In that case, the method may be listed only once with the different estimates it produces listed beside it.

I find that students will typically then want to know the true value of N without me having to prompt them. At any rate, this is now the time to tell them. The following is then a useful question to lead a discussion: "The true value of N is 344. So which of these methods is best?"

The "obvious" answer is the method that happened to produce an estimate closest to 344, but some students are likely to pick up on the fact that the method may have just gotten "lucky" with a particular sample. In fact, the question above contains an undefined word that should be discussed with the class: *best*. What do we mean by the "best" method? The one that's exactly right most often? The one that's within 50 tanks of being right most often? Many statisticians think of the "best" estimator as the one that, among all unbiased methods, has the smallest standard deviation.[2]

After posing the question, the teacher should lead a class discussion in which the following points come out:

- You can't judge a method ("estimator") by how it performs on one random sample only. You have to judge it by how it performs over many random samples. In other words, you have to look at the distribution of its estimates over many random samples. Simulation (especially with technology) is useful for looking at that distribution.
- It is desirable that an estimator be *unbiased*. That is, you would like for the mean of its estimates over many random samples to equal the parameter (N) being estimated.[3]
- It is desirable that an estimator have low variability, perhaps measured by standard deviation. That is, you would like the estimates it produces over many random samples to be relatively close to one another.
- The combination of unbiasedness and low variability makes an estimator that comes close to the desired target for a large number of possible random samples.

In my experience, it is common for students to learn that statistics should be unbiased and have low variability but not be able to express *why* these are important, perhaps because they do not really understand why. They should remember that the ultimate goal of the estimator is to estimate a parameter given a *single* random sample. Since they only get one shot, they want their method to be one that works well for many

---

[2] Other criteria for "best" are possible, the most common being the method with the smallest mean squared error over all possible samples. While not beyond most students, this is more advanced than an introductory course requires, and is certainly not necessary for the discussion surrounding this activity.

[3] Some statisticians prefer *median-unbiased* to *mean-unbiased*. Although there are good reasons for this, the mathematics involved in median-unbiased statistics is often difficult.

possible random samples, thus minimizing their risk of being far from correct. The ability of an estimator to work well for many different random samples is sometimes referred to as its *robustness*.

Students will likely want to know what the British mathematicians did. They used the statistic that has the minimal variance among all unbiased estimators, which for this activity would be $\frac{8}{7}\times\max$, where max is the largest value in the sample. Occasionally a student group will come up with this method. Students typically don't like it when they see its skewed distribution but this is a good occasion for pointing out that distributions need not be normal or even symmetric: all that is desirable is that they come close to the target for many samples.

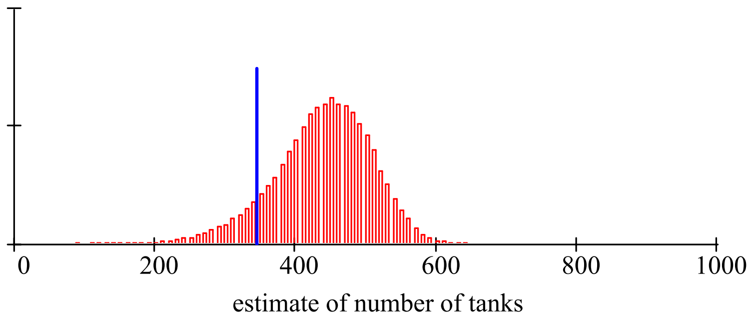
Below is a list of some common student-generated estimators and their distributions. Since many teachers will not have access to technology that makes rapid simulation of distributions possible (*e.g.*, Mathcad was used to create the graphs below), these are included in a format here that can be reproduced by teachers for classroom use. However, it is not unusual for students to come up with methods that are not in this list. The vertical line in each histogram is placed at N=344, the parameter value used for the simulations.

Some possible methods:

1. Take the sample mean and add three times the sample standard deviation. ($\overline{X}+3s_X$)
2. Take the sample mean and add two times the standard deviation. ($\overline{X}+2s_X$)
3. Take two times the mean and add one standard deviation. ($2\overline{X}+s_X$)
4. Double the sample mean. ($2\overline{X}$)
5. Double the sample median ($2Med$)
6. Take 1.5 times the inner quartile range and add that to the third quartile. ($1.5\times IQR+Q3$, where $IQR=Q3-Q1$)
7. Take the largest number in the sample as the estimate of how many tanks there are in all. (*max*)
8. Take six times the sample standard deviation. ($6s_X$)
9. Take three and a half times the standard deviation. ($3.5s_X$)
10. Take the largest value and add the smallest value. (*max* + *min*)
11. Take $\frac{8}{7}$ times the largest value. ($\frac{8}{7}\times\max$)

These pages show the distribution of the described statistic over 50,000 simulations. Each simulation represents the computation of the statistic when a sample of 7 observations is taken from the integers 1 to 344. The vertical line in each case is placed at 344—the true value that the statistic is shooting for.

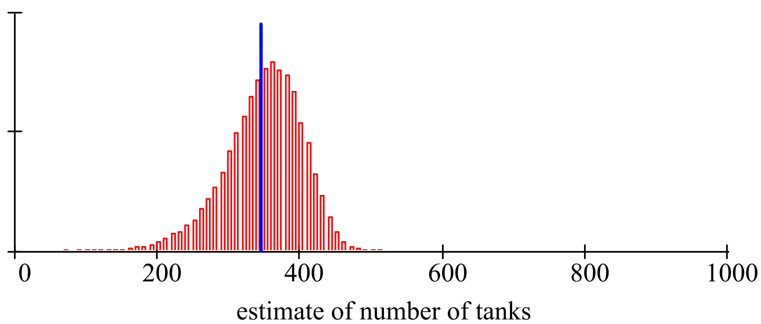
Method 1. $(\bar{X} + 3s_X)$



mean(X) = 442.471

stdev(X) = 68.421

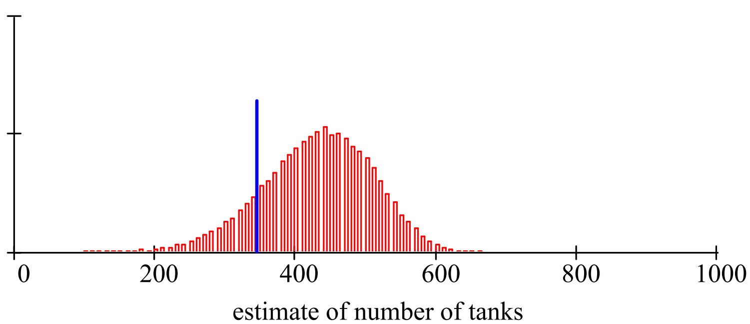
Method 2. $(\bar{X} + 2s_X)$



mean(X) = 352.489

stdev(X) = 53.696

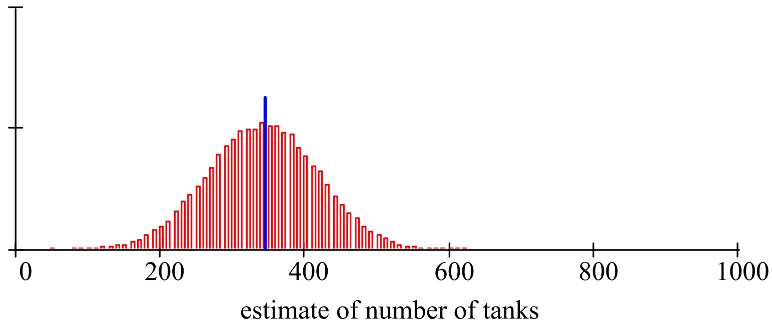
Method 3. $(2\bar{X} + s_X)$



mean(X) = 435.149

stdev(X) = 76.908

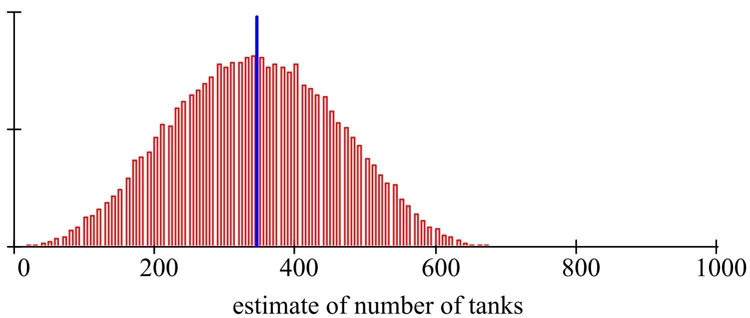
Method 4. $(2\overline{X})$



mean(X) = 344.882

stdev(X) = 75.03

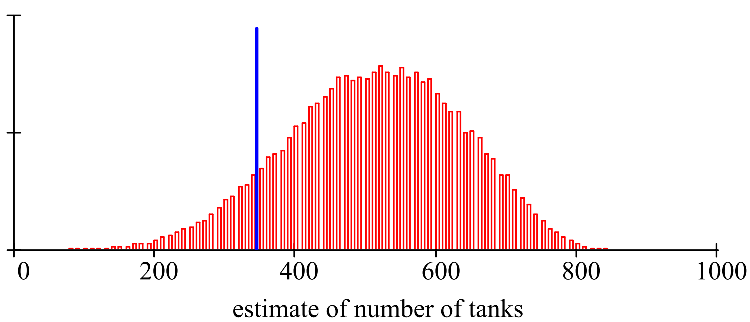
Method 5. $(2Med)$



mean(X) = 345.102

stdev(X) = 114.541

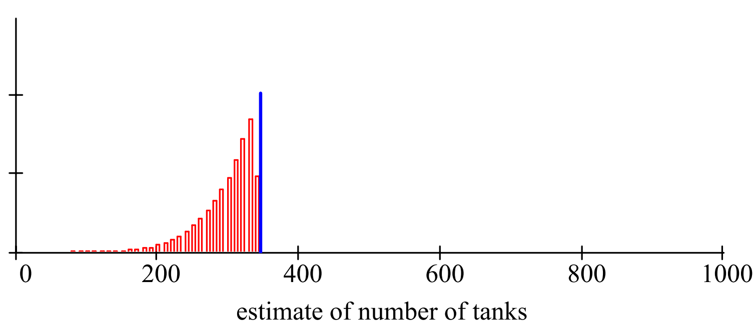
Method 6. $(1.5 \times IQR + Q3$, where $IQR = Q3 - Q1)$



mean(X) = 516.294

stdev(X) = 121.88

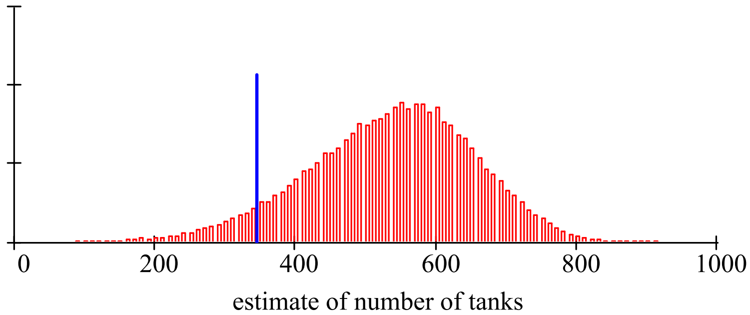
Method 7. (max)



mean(X) = 301.511

stdev(X) = 37.828

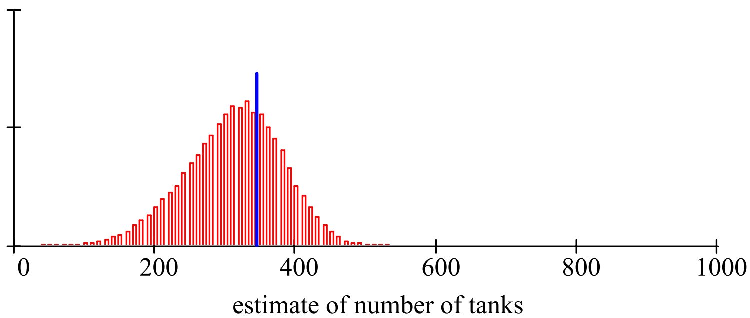
Method 8. $(6s_X)$



mean(X) = 539.683

stdev(X) = 116.253

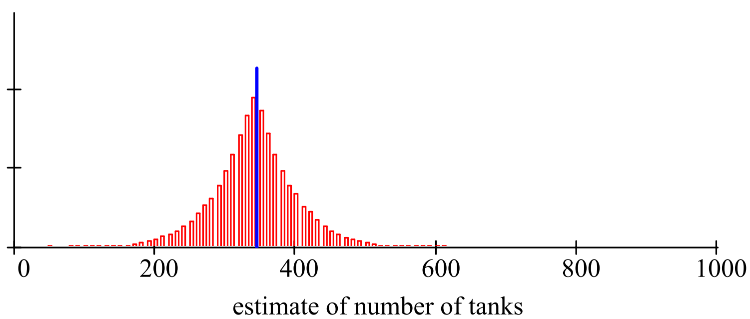
Method 9. $(3.5s_X)$



mean(X) = 315.311

stdev(X) = 67.782

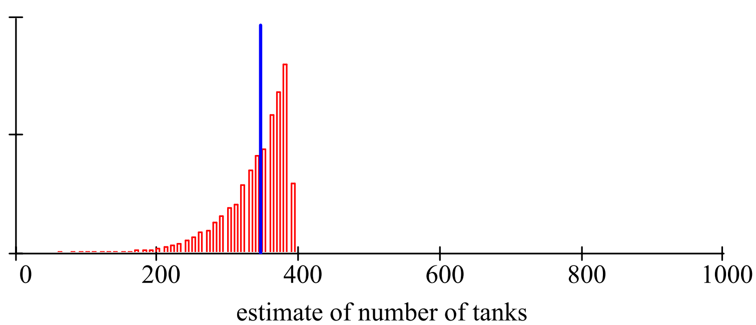
Method 10. $(\max + \min)$



mean(X) = 344.925

stdev(X) = 57.632

Method 11. $(\frac{8}{7} \times \max)$



mean(X) = 344.677

stdev(X) = 43.382

Session number	Trial number 1	2	3	4	5	Totals
1	1	6	6	6	3	**22**
2	11	0	6	3	3	**23**
3	3	2	0	0	1	**6**
4	3	8	8	4	1	**24**
5	10	3	6	6	8	**33**
6	6	0	0	0	2	**8**
7	1	1	2	5	2	**11**
8	5	1	1	2	2	**11**
9	0	0	4	2	0	**6**
						144

Miller's hypotheses were:

H_0 : Scipio is *not* learning where the platform is.

H_A : Scipio *is* learning where the platform is.

Consider Miller's research question: *Is Scipio learning?* How strong is the evidence provided by his data that Scipio is, in fact, learning to find the middle platform? Choose an inference procedure that you think is appropriate and answer Miller's question. Be prepared to share with the rest of the class what inference procedure you chose, why you chose it, and what conclusion it led you to.

This data set and an article describing a classroom activity similar to the one presented in this session at Beyond The Formula will soon be available at apcentral.collegeboard.com.