## Introduction to Estimation Activity

**Problem**: You're in a land war and the enemy has armored tanks. You don't know how many. Your intelligence agents have told you that their tanks have sequential serial numbers, starting with 1. So, when you capture a tank with serial number 357, you know that the enemy has at least 357 tanks. Our troops have captured 5 tanks and we can read the serial number on each. Your task is to estimate how many tanks the enemy has. Here are their serial numbers: 342, 075, 139, 659, 582

**Simple answer**: Do the sample, look at the numbers on our 5 captured tanks and then give an estimate of the answer the question.

**Applied statistics course answer**: Look through the book to find the method / formula to answer a question like this and then plug in our 5 numbers. In a good applied statistics course, we will also learn what assumptions are needed and some methods for checking whether they are met.

**Theoretical statistics course answer**: Use the basic ideas of probability and statistics to derive the appropriate method / formula. When approximations or simplifications are made, be quite clear about how those are used in the mathematical derivations of the techniques. The insights gained from this should improve one's ability to use the standard statistical techniques, check the assumptions, and modify the techniques when necessary.

(If you were the person in charge, which type of answer would you want?)

**Exploration -- An Outline of Questions**

1. Find a theoretical distribution to use to model the individual data points from this population. What is an appropriate one?
2. Give at least two different possible summary statistics that would be reasonable to use to predict the number of tanks in the population.
3. What is the sampling distribution of each statistic?   
   Can you find the sampling distribution exactly? Can you simulate it if you can’t find it exactly?
4. What is the mean and variance of the sampling distribution of each statistic?   
   Can you find the mean and variance of the sampling distribution exactly? If not, can you estimate them from your simulation of the sampling distribution?
5. Based on your analysis of the sampling distributions, does one of these statistics seem to be a better estimator than the other?
6. Does your analysis suggest a modification of any of these statistics that would make them even better estimators?
7. In reality, often you will have a sample size larger than 5. What if the sample size here was 60? How would that change your answers to the above questions?
8. Your supervisor is interested in not only a point estimator, but also an interval estimator. Could you obtain an interval estimator for each of your point estimators? How?
9. What would be a reasonable way to judge whether one interval estimator is better than another interval estimator?

To explore this:

Investigate these three estimators, summarizing each simulation of the sampling dist’n with

* a histogram
* the mean
* the variance

1. 
2. 
3. 

## Discuss

## What distribution should you assume for your data?

## When you simulate, what will you use for the number of enemy tanks in the population?

## What properties should estimators have?

## What does simulation tell us about each of these three estimators’ properties?

## Based on simulation, which estimator would you choose?

## What if n = 60, rather than n = 5.

## Redo 2 and 3 above for the larger n.