

Introduction to Estimation Activity

Purpose:

Start this part of the course thinking about estimation in a broad sense.



Assignment:

Think about the question and think of at least two reasonable candidates for a good estimator.

Number of “enemy” tanks

- 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.

Strategies to form an answer:

Simple answer:

Do the sample

Look at the numbers on our 5 captured tanks

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.

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.

Strategies, continued

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.

What would YOU do – among these three?

What would I do?

- “In the moment” I’d try to look it up in an applied statistics book, if I could. Even if I could just find something “nearly right.”
- I wouldn’t trust theoretical work in done in a hurry.
- If I expected to have to do this again, and if it were important, I’d start working on the theoretical approach – asking for help if needed.

Suggested estimators?

1. The sample mean
2. The sample maximum
3. Twice the sample mean
4. The sample maximum plus 30
5. The sample maximum times 1.2
6. The sample mean plus $3 \cdot \text{StDev}$

My thoughts

1. The sample mean **Unreasonable**
2. The sample maximum **Close, somewhat too small**
3. Twice the sample mean **Maybe**
4. The sample maximum plus 30 **Maybe, but why 30?**
5. The sample maximum times 1.2 **Maybe, but why 1.2?**
6. The sample mean plus $3 \cdot \text{StDev}$ **Maybe, but why 3?**

Toward Theoretical Stat: Outline of Questions 1

1. What is an appropriate mathematical model for the population?
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?

Toward Theoretical Stat: Outline of Questions 2

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 either or both of these statistics that would make them even better estimators?

Toward Theoretical Stat: Outline of Questions 3

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?

What should you do now?

- Think about what you learned about simulating sampling dist'n's of statistics.
- You don't have to know math stat to answer several of those questions.
- But, if you want to consider an estimator like the sample mean plus $3 \cdot \text{StDev}$, then you'll have to write a script to approximate its sampling dist'n.

Which you could do!

What should you have noticed?

- We can have multiple estimators for parameter.
 - It's possible to investigate the properties of those estimators in more than one way: simulation and, in some cases, with theory.
 - It might be a good idea to have an overview of Math Stat methods for finding good estimators.
-