

**Tank Inference**

Today, we’ll be helping the Allied war effort by tackling a real military intelligence problem from WWII. For this problem, we will be taking a **sample** from a **population**. We will use a calculation from the sample to estimate a **parameter** from that population.

Key sources:

* Ruggles, R.; Brodie, H. (1947). "An Empirical Approach to Economic Intelligence in World War II". *Journal of the American Statistical Association.* 42 (237): 72. doi:10.1080/01621459.1947.10501915.
* "Gavyn Davies does the maths – How a statistical formula won the war". The Guardian. 20 July 2006. https://www.theguardian.com/world/2006/jul/20/secondworldwar.tvandradio

As the Allied forces prepared for their D-Day invasion of Europe, one question commanders had was: How many tanks do the Nazis manufacture per month? Some intelligence officials provided estimates based on spies and reconnaissance. However, a group of statisticians had a different strategy.

**Serial No.**

**July: 121**



**Serial No.**

**July: 121**

Since the Germans were…well…German, they had an incredibly organized manufacturing process. They labeled various tank parts with serial numbers in the order of their production. Let’s say German tanks with the following July serial numbers were captured during combat. Assume it’s a random sample of the tanks produced:

**Serial No.**

**July: 121**

**038, 072, 121, 158, 206**

1. Identify the population, the sample, and the parameter we are attempting to estimate.

2. What statistic would you calculate to estimate the total number of tanks produced that month? Describe your method, why you chose it, and show your result (using the data above).

Sample: **038, 072, 121, 158, 206**

3. The best estimation methods for this problem take into account that serial numbers are uniformly distributed. This fact is helpful in multiple ways. For example, uniform distributions are symmetric, so the maximum will be approximately double any valid measure of center. Calculate the estimated maximum using the following two methods:

Method 2: $ \hat{N}=2(sample median)$

Method 1:$ \hat{N}=2\overbar{x}$

4. Another consequence of the uniform distribution is that the expected “gaps” between the sampled numbers should be similar (assuming a random sample). So, the maximum should be one “gap” away from the highest number in the sample. Calculate the estimated maximum using this method:

Method 3: $\hat{N}=m+\frac{m}{n}$ (m is the maximum of the sample and n is the sample size)

5. The allied forces used a modified version of Method 3. Method 3 tends to overestimate N slightly. Subtracting by 1 provides you with the “minimum-variance unbiased estimator.” Don’t worry about the particulars of what that means. Calculate the estimated maximum using this method:

Method used during the war: $\hat{N}=m+\frac{m}{n}-1$

|  |  |
| --- | --- |
| **Statistical Method** | $$\hat{N}$$ |
| **1:** 2$\overbar{x}$ |  |
| **2:** 2(*median*) |  |
| **3:** $m+\frac{m}{n}$ |  |
| **Used:** $m+\frac{m}{n}-1$ |  |

6. Fill in the table with all the calculated estimates for $\hat{N}$

7. Military intelligence, using reconnaissance, estimated the Nazi tank production was about 1,400 per month. According to records recovered after the war, the Nazis ***actually produced 245 tanks per month***. Why do you believe the statistical inference did better than the military intelligence?

What is a Sampling Distribution? Day 1

Important ideas:

Check Your Understanding

The statisticians in Washington, D.C. considered several other statistics to use estimate of the total number of German tanks *N*. Here are simulated sampling distributions for three statistics that the statisticians considered, using samples of size *n* = 7. The blue line is at *N*, the total number of German tanks. The shorter green line segments mark the mean of each simulated sampling distribution.



1. Do any of these statistics appear to be unbiased? Justify.
2. Which of these statistics do you think is best? Explain your reasoning.
3. Explain how the Allies could get a more precise estimate of the number of German tanks using the statistic you chose in part 2.