Gavyn Davies does the maths

How a statistical formula won the war

Here is a story about mathematical deduction that I love, mainly because it is said to be true, and because it had an impact (albeit small) on the outcome of the second world war. It is the story of how a simple statistical formula successfully estimated the number of tanks the enemy was producing, at a time when this could not be directly observed by the allied spy network.

By 1941-42, the allies knew that US and even British tanks had been technically superior to German Panzer tanks in combat, but they were worried about the capabilities of the new marks IV and V. More troubling, they had really very little idea of how many tanks the enemy was capable of producing in a year. Without this information, they were unsure whether any invasion of the continent on the western front could succeed.

One solution was to ask intelligence to guess the number by secretly observing the output of German factories, or by trying to count tanks on the battlefield. Both the British and the Americans tried this, but they found that the estimates returned by intelligence were contradictory and unreliable. Therefore they asked statistical intelligence to see whether the accuracy of the estimates could be improved.

The statisticians had one key piece of information, which was the serial numbers on captured mark V tanks. The statisticians believed that the Germans, being Germans, had logically numbered their tanks in the order in which they were produced. And this deduction turned out to be right. It was enough to enable them to make an estimate of the total number of tanks that had been produced up to any given moment.

The basic idea was that the highest serial number among the captured tanks could be used to calculate the overall total. The German tanks were numbered as follows: 1, 2, 3 ... N, where N was the desired total number of tanks produced. Imagine that they had captured **five** tanks, with serial numbers **20**, **31**, **43**, **78** and**92**. They now had a sample of **five**, with a maximum serial number of **92**. Call the sample size S and the maximum serial number M. After some experimentation with other series, the statisticians reckoned that a good estimator of the number of tanks would probably be provided by the simple equation (M-1)(S+1)/S. In the example given, this translates to (**92-1**)(**5**+1)/**5**, which is equal to **109.2**. Therefore the estimate of tanks produced at that time would be **109**

By using this formula, statisticians reportedly estimated that the Germans produced **246** tanks per month between June 1940 and September 1942. At that time, standard intelligence estimates had believed the number was far, far higher, at around **1,400**. After the war, the allies captured German production records, showing that the true number of tanks produced in those **three** years was **245** per month, almost exactly what the statisticians had calculated, and less than **one fifth** of what standard intelligence had thought likely.

Emboldened, the allies attacked the western front in 1944 and overcame the Panzers on their way to Berlin. And so it was that statisticians won the war - in their own estimation, at any rate.