War: Crash Course Statistics #42

Crash Course: Statistics
https://youtube.com/watch?v=rRhHY7Mh5T0
https://nerdfighteria.info/v/rRhHY7Mh5T0

Hi, I'm Adriene Hill, and Welcome back to Crash Course Statistics.

Statistics and probability have been used in applications beyond the ones we usually think of like research science and business analytics. One of the most consequential applications of statistics is helping countries survive and win wars.

Today, we'll talk about how people applied statistics to break codes, locate sunken submarines, and even predict the next big war. INTRO Our first story is pretty well known in the fields of computer science and statistics. In World War II, the Germans used what looked like a complicated typewriter to encode their messages.

These machines, called Enigmas, allowed the Germans to type in messages and receive encoded ones back. You may have done some simple encoding in your childhood. Like, if you wanted to send a message to your friend that says "I like, like Alex" but you don't want Alex or anyone else for that matter to be able to read the message.

You could create a key so that each letter is represented by another letter like this: If you wanted to write "I like" you'd find those letters in the top row, and write down their counterparts. "I like" becomes "Y tymf" which makes no sense...unless you have the key. Your entire message would go from this: "I like, like Alex." To this: "Y tymf, tymf ptfw." So you're safe to deliver your message. But sometimes decoding messages has much higher stakes than protecting crushes.

Like when there's a war going on. So by necessity, the keys--or methods of encryption--are much more complex. During the Enigma Encryption process, a letter was sent through three rounds of encoding--similar to how we encoded our message about Alex.

But the enigma had three Rotors, or wheels, doing the encoding. And the enigma machines would rotate the wheels systematically after EVERY LETTER. A letter that appeared in the original message twice could get encoded as two totally different letters.

There were 26 settings on each wheel, one setting for every letter in the alphabet. So there were 17,576 possible starting settings (just for the wheels!) making it impossible to figure out a message by manually trying each start point. If you wanted to decode a message, you needed to know how those wheels were set.

The Germans also had multiple wheel options AND plugboards, making things even more complicated. Alan Turing and his team developed a technique called Banburismus for deciphering messages from the German Navy - which exploited the fact that sometimes pairs of messages would have chunks of text within them that had been encoded with the same settings. They used a very time-consuming method to find these pairs.

Every intercepted message got hole punched, in order, into paper that was lined with horizontal alphabets. Then, one message was placed on top of another message, so a person could see how often holes overlapped. Why?

Well, two messages that were encoded with different Enigma settings would only have letters that matched by random chance. The German navy had a primary Enigma that they were using known as "Dolphin." Two messages encoded by the same Dolphin settings had a 1/17 chance of having randomly matching letters. If two messages were encoded using DIFFERENT settings, there's a 1/26 change of having randomly matching letters.

So, more matches than 1/26 would be increasing evidence that the messages were encoded using the same settings. The Enigma codebreakers used that knowledge to determine whether two intercepted messages were more likely to be encoded using the same or different settings. They were also able to use other knowledge in the decoding process.

Like, the team already knew that 90% of Enigma messages contained the German word "ein," which can mean "one," "a," or "an." Plus, there were phrases about the weather that were getting repeated often in messages. 'Cuz, boats. When Turing and his team determined it was 50 times more likely that the messages were encoded with the same settings than not, they considered it almost certain they'd found a match. They had a machine called "the bombe" that could automatically cycle through a bunch of those wheel settings in order to decode messages.

But it took a LONG time to go through all of the possibilities, so being able to narrow them down was a necessary step. As Mike Lee and Benedict King put it in their article in The Conversation, "Turing's crucial Bayesian insight was that certain messages were much more likely than other messages." All this knowledge helped the team figure out how the Enigma's wheels were set when it encoded a given message. Using Bayesian reasoning helped Turing's team crack the Enigma code, and limited the amount of settings they had to test by hand.

Some historians think cracking the Enigma may have shortened the War by 2-3 years, saving millions of lives. In WWII, German U-boats were systematically taking down Allied ships, including unarmed merchant ships with supplies. While some ships escaped unharmed like the Empress of Scotland which carried Turing from New York back to Europe the Allied forces suffered many losses.

Locating the U-boats was not an easy task, but the mathematician B. O. Koopman used Bayesian reasoning.

Koopman would first ask experts where the U-boat was likely headed. With limited time and resources, prior information and beliefs about the U-boat were important. Koopman commented that: "Police will patrol localities of high incidence of crime.

Public health officials will have ideas in advance of the likely sources of infection and will examine them first." And he wanted to do the same with the German U-boats. Using signals from the ship, Koopman was able to target a 236 mile radius for planes to search. But that's still big.

He would assign a 50-50 probability that the U-boat was inside the circle, then he would use all of the military information that he had access to in order to update those beliefs. That way he could make the best decisions with whatever information he currently had. Think about the last time you lost your keys.

You could plot out a grid that represented your apartment, and you could assign a probability that your keys are in each 1 foot by 1 foot square based on the likelihood of possible ways you misplaced them. So maybe your keys fell out of your bag, which would put them somewhere in this square. Or maybe your cat got into your bag and dumped its contents onto the floor.

Then they'd be in this square. Or maybe you left them in your jacket pocket. Then they'd be here.

Based on how likely you think these scenarios are...and the knowledge that your cat loves to push things off of tables... the best guess is that the cat knocked over your bag again...you can use Bayesian reasoning to create a probability map of where your keys are most likely to be. You could also include information about how likely you are to find your keys if you searched for them in that square. Keys that fell behind the refrigerator might be hard to find

even if you did search there.

It'd also be really hard to find your keys if they went down a drain outside your door. Combining all this information would leave you with a map of your apartment that tells you the best places to search. This same theory--called Bayesian Search Theory-- was also applied by John Craven to find a missing nuclear submarine in 1968.

Craven collected experts' opinions on what happened to the USS Scorpion, and used Bayesian Search Theory to create a map of where the sub would likely be found. And it worked! Craven found the sub right next to where he expected it.

Often in war it's also essential to know approximately how MANY of these vehicles exist. Your strategy might be different if your enemy had 1000 tanks than if they had only 200. During WWII, Allied forces used traditional techniques such as spying and interrogating captured German soldiers and estimated that the Germans were producing about 1,400 tanks a MONTH.

But that seemed high. Luckily, the Allies had already captured some tanks with serial numbers on them. So they used some clever math to estimate the actual total number of tanks.

Assuming that the tanks' serial numbers went in order which was a reasonable assumption they could use the range of the serial numbers to estimate how many there were. For example, if we found 4 tanks with the serial numbers 7, 17, 47, and 65. We'd know there are at LEAST 65 tanks.

But it's possible there are 67 tanks. Or 102 tanks. Or 500 tanks.

We need a way to estimate what the most likely maximum is. There are many ways to do this, but one simple one is to use this formula, where m is the maximum serial number you observed ours is 65 and n is the number of observations you made. We made 4.

So our best guess at how many tanks there are based on the data we collected is 80.25 we'll round that to 80. Because you can't have .25 tanks. When the Allies used similar techniques, they estimated that there were 256 tanks being made per month.

A much more accurate estimate. The Germans were actually making about 255 tanks a month at the time. And note to self: when fighting a war, do not use sequential serial numbers - unless you're fighting raccoons - they can't read.

Jumping forward in time, to today. Some researchers use statistical models to predict when the next big war will be. It has been a long time since the last major World War.

Aaron Clauset of the University of Colorado has set out to examine other stretches of peace. And this isn't as simple as just counting the years between major wars and calculating the average time of peace. Clauset looked for trends and correlations that might predict the number of years between major conflicts..

He found that across history, huge stretches of peace were not unusual. In fact it was downright common to see 100 to 140 years of peace following a large scale war. This long stretch of time without large-scale world war is more rule than exception.

Statistics has many important applications. War being one particularly high stakes application. Mathematicians and Statisticians played a huge role in WWII, and they continue to be a part of defense departments and military planning to this day.

Out of necessity, we often make huge strides in the fields of math and statistics during wars. They force us to solve problems we may not have needed to solve in times of peace. Things like the Bayesian Search Theory that Koopman worked is also used in times of peace like in helping us find missing planes.

And the code breaking done by Turing and his team was not only important in introducing statisticians to Bayesian inference, but it provided foundations for future code breaking and encryption work that's being done today. Thanks for watching, I'll see you next time.