Subject: statistical significance
From: Joseph Picone <picone@temple.edu>
Date: 1/14/23, 7:50 AM
To: ECE 8527 <temple_engineering_ece8527@googlegroups.com>

One of the most important things we will discuss this semester is the concept of statistical significance. So let me introduce this concept using a sport analogy regarding an American football game today:

"What to watch for: The 49ers beat the Seahawks in two previous meetings this year by a combined score of 48–20. Teams that have won the first two meetings and then hosted the third meeting are 14–9 in the postseason matchup since the 1970 merger."

The author is arguing that the odds are very good that the 49ers will beat the Seahawks because the 49ers won two previous meetings this year.

But is a 14–9 edge for a team in this situation really a conclusive argument. This is not too far from a 50/50 chance given the small sample size.

I think we all agree that if the history said 23–0, we would agree the 49ers should win. Also, if the history was 12–11, we would agree this was not evidence the 49ers should win. For what difference (14–9, 13–10, 15–8, etc.) would we be 95% sure the 49ers would win.

We will learn to answer this question by stating that "we are 80% confident that the 14–9 difference is statistically significant", meaning the 49ers are favored.

We will learn how to compute the confidence level, 80% in the above statement, and to determine at what confidence level 14–9 becomes a statistically significant difference. There is a chance, after all, that 14–9 could have been a normal random outcome for this experiment, and within the realm of what we call experimental error.

This is a topic generally referred to as experimental design, and it is an important part of machine learning.

–Joe