Predicting success in european soccer based on tactical spatiotemporal data

By Ryan Lussier

Hypothesis testing comparing shooting positions

During our exploratory data analysis it appeared that events involving Mo Salah were biased toward the left side of the field. Let's take a closer look and explore the question: is Mo Salah's mean y position when he shoots the ball different from everyone else? To evaluate this hypothesis we will assume Mo Salah is independent of the rest of the field and perform an independent t-test. To start, let's establish our null and alternative hypotheses.

> H0: Mo Salah shoots the ball from the same y position as everyone else.
> H1: Mo Salah shoots the ball from a different y position than everyone else

Before we get to the t-test, however, let's take a look at the distributions.
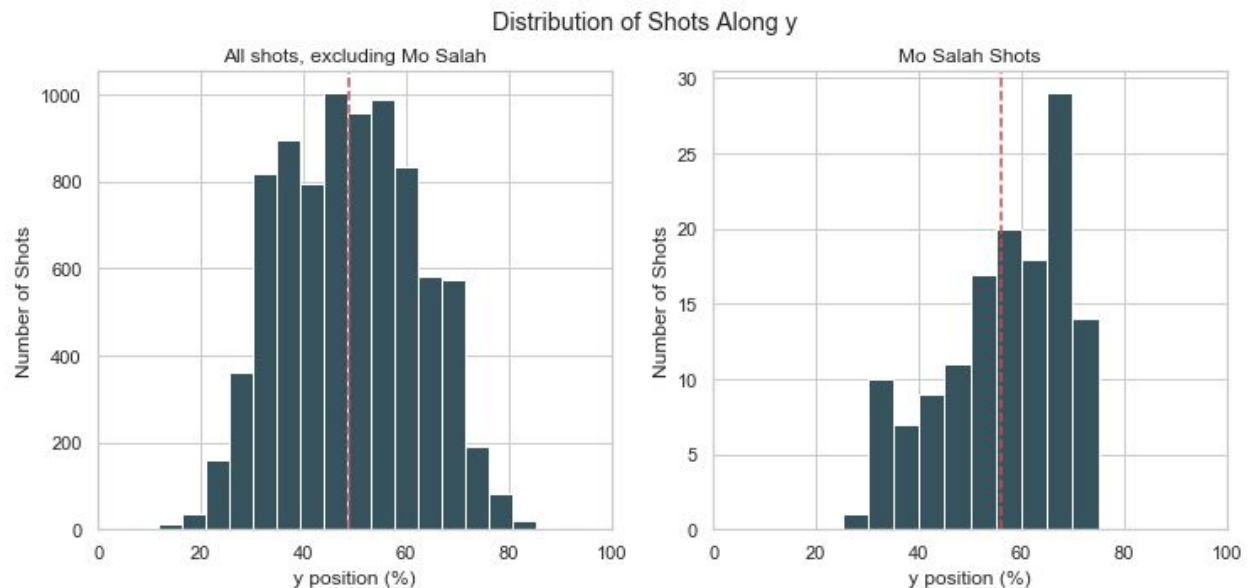
In the histograms above, 0 on the x-axis represents the right sideline if you are facing the goal and 100 represents the left sideline. The positions are shown as a percentage of the field width. It appears that the mean of all players excluding Mo Salah is close to 50% while Mo is biased towards the left side of the goal.

Now lets run the t-test to see if this difference is significant. Typically normality is assumed when running a t-test. This assumption seems to hold for all shots, but when looking at Mo exclusively the distribution is not as normally distributed. In this case we will overlook the less than ideal distribution because t-tests are robust to violations of the normality assumption. To read more about this robustness, check out this link.

For the t-test we will utilize the ttest_ind function from scipy.stats which yields the following results.

t statistic: 6.140147247988889
p value: 8.610519188761847e-10

The results of this test show a p value which is essentially 0 which makes a strong case for rejecting the null hypothesis. Let's take this a step further and compare the t-statistic to the critical t score. For this case we'll set our confidence level to 0.95.

critical t score: 1.6450339959007674

If we compare the t-statistic with the critical t score we find that 6.14 > 1.65 and we reject the null hypothesis. Thus we can conclude with 95% certainty that on average Mo Salah shoots the soccer ball from a different y position than everyone else. One possible explanation for this may be that Salah prefers his left foot and thus plays to the left side of the goal where it is easier to control the ball and cross back toward the goal or the center of the field.