Sports Analytics

Guided Notes 01/05/23

Recall:

R^2 tells us how successful a linear regression is – the higher the R^2, the better the explanatory variable is for predicting the response.

The **least squares regression line (LSRL)** is the line that is the best fit straight line to a scatterplot:

$$\widehat{Y} = mX + b,$$

where m is the slope, and b is the y-intercept.

- The slope tells us how much we expect Y to change for each *unit increase* in the X variable.
- The intercept tells us what value we would predict for Y if $X = 0$.

We used the slope of various LSRL's to get a sense of the run-scoring values of various offensive events. This method has some limitations. (The example for 3B's from last time provides an indication of this.) There's a difference between finding the actual average value of an event and finding the value one should place on the event when trying to evaluate players.

Today we use the work of Lindsey to determine the actual average values of various events more directly. Central to the method is the probabilistic notion of *expected value*.

Def: The **expected value** of a situation with random outcomes is the **average value** that would result from the situation if it was repeated many, many times. To compute expected value, we do the following:

1. List all possible ____________________ .
2. Find the ____________________ for each of the outcomes.
3. Multiply each outcome by its probability.
4. Add up all of the results.

Ex: Suppose that a basketball team scores 3 points on 15% of its possessions, 2 points on 35% of its possessions, 1 point on 5% of its possessions, and 0 points on the rest. Find the expected value (or average value) of an offensive possession for this team.

By analyzing thousands of base-out situations in MLB, Lindsey estimated the probabilities of all possible game situations along with the expected number of runs for each and the probability of scoring for each. We can use his work to determine the average values of various offensive events. We can also use it to analyze various strategies.

Ex: Use Lindsey's results to estimate the cost of a batter striking out with the bases loaded and nobody out.

Class: If runners are on 1st and 2nd with nobody out, estimate the value of a successful sacrifice bunt. (Do this both in terms of expected runs and in terms of probability of scoring.)

Class: If runners are on 1st and 2nd with nobody out, estimate the value of an unsuccessful sacrifice bunt. (Do this both in terms of expected runs and in terms of probability of scoring.)

Class: What is the value of attempting a sacrifice bunt in this situation if the probability of a successful sacrifice bunt is 60%? (Do the calculations both in terms of expected runs and in terms of probability of scoring.)

Ex: Estimate the average run value of a HR.

Class: Estimate the average cost of striking out.