**Exploring Logistic Regression Through Cricket**

Cricket is a game watched and played by billions of people across the world. Second in global popularity only to football (soccer), it is extremely popular in South Asia, Australia, Africa, and Europe.

**The Rules of Cricket**

Cricket is played on a rectangular pitch inside an oval boundary. In the pitch, on either side, there is a vertical **wicket**, made up of three vertical wooden "stumps" with two wooden bails (small blocks of wood) resting atop the stumps. There are two teams, each with 11 players (a batting team and a bowling team).

**The bowling team:** The bowler bounces a small leather ball at the batter. The goal of the bowler is to get the batter out and ultimately knock the bails off of the stumps.

**The batting team:** The goal of the batting team is to score as many runs as possible and not let the bowler hit the bail off of the stumps. There are two batters at a time for the batting team. After the batter hits the ball, the two batters will attempt to switch sides. Each batter will go until they get out, when at that time, they will switch with another teammate.

**Scoring Runs:** Each time the two batters switch sides without getting out counts as a run scored. For example, if you hit the ball and are able to run to the other wicket and back without getting out, that would score two runs for your team. If the batter hits the ball on a bounce to the oval boundary, it counts as four runs scored. If the batter hits the ball over the oval boundary on a fly, it counts as six runs.

**Outs:** Below are a few common ways in which a batter can get out

* If the batter hits the ball and it is caught in the air (caught out)
* If a player on the bowling team throws the ball and it hits the bails off the stumps before the batter can cross the line while trying to score a run (run out)
* If the bowler throws the ball past the batter and it knocks a bail off the wicket
* If the batter gets hit in the legs by a ball that would have hit the wicket (Note: if the batter gets hit and the ball wasn't ruled as one that would have hit the wicket, the batter is not out)

A screenshot of a sports game

Description automatically generatedThe duration of cricket matches can range from hours to days, depending on what format is being played, and scores are often high as each individual batter usually scores many runs on multiple pitches before getting out.

**Data**

The asia\_cup data set includes data from each cricket match played in all Asia Cup Tournaments from 1984 (the first one) to 2022. The Asia Cup is a tournament now takes place every two years, alternating host cities in different countries throughout Asia. To the right is a description of the variables in the data set you will be using.

1. **Provide the R code that when executed, would run a logistic regression that predicts the result of the match using runs scored, sixes, highest score, and extras given.**

glm(Result ~ `Run Scored` + `Sixes` + `Highest Score` + `Given Extras`,

family = binomial,

data = asia\_cup)

1. **Given the R output below, give the logistic regression equation, interpret all of the coefficients in context, and for the predictors, determine whether or not they are significant at the 0.05 level.**

**A screenshot of a computer

Description automatically generated**

**Intercept = -4.814**

the predicted log odds that a team wins the cricket match when they score zero runs, zero sixes, bat first, and give up zero runs is predicted to be -4.814

**Run Scored = 0.018**

for each additional run scored by a team, we predict the estimated log odds that team wins the cricket match to increase by 0.018, assuming all else remains constant. It is a good predictor because it is significant at the 0.05 level as the p value is 4.85x10^-6

**Sixes = 0.147**

For each additional six scored by a team, we predict the estimated log odds that team wins the cricket match to increase by 0.147, assuming all else remains constant. It is a good predictor because it is significant at the 0.05 level as the p value is 0.021

**Selection = 4.440**

The estimated difference in predicted log odds that a team wins the cricket match between a team that batted first and a team that bowled first is 4.440, with the team that bowled first having a higher likelihood of winning the match. It is a good predictor because it is significant at the 0.05 level as the p value is 4.92x10^-5

**Given Extras = 0.007**

For each additional extra run a team gives up, we predict the estimated log odds that team wins the cricket match to increase by 0.007, assuming all else remains constant. It is not a good predictor because it is not significant at the 0.05 level, as the p value is 0.724.

**Run Scored:Selection = -0.018**

The estimated difference in log odds that a team wins the cricket match for each additional run scored, between a team that bats first and a team that bowls first is -0.018. With the team that bowls first less likely to win the match. It does make a good predictor because the p value is significant at the 0.05 level (0.0003).

1. **Write the logit form and the probability form of the multiple logistic regression model above and interpret what they are predicting.**

**Logit form:**

**Where logit(win) = the estimated log odds of a team winning the cricket match**

**Probability form:**

**Where P = the probability of a team winning the cricket match**

1. **Find the estimated log odds of a team that starts the match batting wins given they score 265 runs, 8 sixes, and give up 5 extras.**

-4.814+0.018(265)+0.147(8)+4.440(0)+0.007(5)-0.018(0) = **1.167**

1. **Find the estimated probability of that same team winning the cricket match**

e^(-4.814+0.018(265)+0.147(8)+4.440(0)+0.007(5)-0.018(0))/(1+ e^(-4.814+0.018(265)+0.147(8)+4.440(0)+0.007(5)-0.018(0)) ) = **0.76260232361 = 76.26%**