

# Project Proposal: MATH 532 Regression II

## Due: Friday, Mar 15, 2019

Sara Bey and Jayanth Sivakumar

March 12, 2019

**Questions:** Can we predict our chances of getting a certain hand in a poker game? How often can we expect to get a hand containing one pair, two pairs, or even a royal flush? If we play multiple games, can our past hands tell us anything about what we might draw in the next game? How many of our predictions are correct?

**Data Set:** We will use the Poker Hand Data Set provided by the UCI Machine Learning Repository to carry out our analysis. This data set contains examples of poker hands, consisting of five cards, drawn from a standard deck of 52 cards. Each card in the hand is described using its suit and rank, which gives a total of 10 predictors for a given poker hand. The suit of each card is represented by  $S1, \dots, S5$ , and the rank is represented by  $C1, \dots, C5$ . The suit is given by an ordinal value of 1 – 4, which represent Hearts, Spades, Diamonds, and Clubs, respectively. The rank is given by a numerical value of 1 – 13, representing Ace, 2, 3, ..., Queen, King, respectively. The last piece of information in the data set is called the *Class* of the poker hand. The *Class* is represented by an ordinal value between 0 and 9, where 0 corresponds to having nothing in hand, 1 corresponds to one pair, 2 to two pairs, 3 to three of a kind, 4 to a straight, 5 to a flush, 6 to a full house, 7 to four of a kind, 8 to a straight flush, and 9 to a royal flush. There are a total of 1025010 total poker hands, 1000000 in the test set and 25010 in the training set, provided in the Poker Hand Data set.

**Hypothesis:** We hypothesize that it will more likely to get a hand that contains nothing, or even one pair, than it will be to get a hand that contains four of a kind, a flush or even a royal flush. If the game and the draws are made at random, we should not be able to predict our future hands based off of what we have had in hand in previous rounds of poker.

**Analysis:** We will first start by providing some visual representations of the poker hands provided in the data set to get an idea of how many of each *Class* (type of poker hand) are present among the instances. We will then use the deviance to analyze the goodness of fit and odds ratio, in order to determine the odds of getting each type of poker hand. We can use this model to make predictions about what will be in our hand during any given round of poker. We can then use the training set provided in the data to fit a model that predicts the probability of getting each poker hand with *Class* values from 0 to 9. We can use this model to see how well it fits the test set. We can make confidence intervals for these predictions to see how accurate they are. We can then take our data one column at a time, for instance looking at the first cards suit and rank, and see if we can make some predictions about which card we may get next, and overall which hand we may end up with. Lastly, we can use multinomial regression to see how accurately our 10 variables can predict the response variable *Class*.

**Rationale:** We are interested in investigating the accuracy of a model that can predict your next card given your current card as this will give some insight as to what type of hand a player may end up with. We are also interested in the accuracy of the model that takes the 10 predictor variables, the suit and rank for each of the 5 cards, and predicts the *Class*, or overall poker hand at the end of the round as we want to see how many instances in the next round are accurately determined based on the prediction of the current round. The accuracy of these models is important, as someone playing poker would be interested in getting a better hand, with a *Class* value of 9, in order to win more money.

**Citation:** <https://archive.ics.uci.edu/ml/datasets/Poker+Hand>