Modeling Kobe

Daniel Byrne

April 12, 2019

## Introduction

Kobe Bryant marked his retirement from basketball by scoring 60 points in his final game as a member of the Los Angeles Laker team on Wednesday, April 12, 2016. Starting to play professional basketball at the age of 17, Kobe earned the sport’s highest accolades throughout his long career. Using 20 years of data on Kobe’s shots made and shots missed, we explore potential models that attempt to predict whether or not his shot went in. This type of model could be used in building a simulation or video game in which a model can be used to predict whether or not a shot went in.

## Data

The original data set contains the location and circumstances of every shot attempted by Bryant during his 20-year career. Your task is to predict whether the basket went in (shot\_made\_flag = 1) or missed (shot\_made\_flag = 0). The data for estimation is in project2Data.xlsx.

## Questions

We explored several models trying to establish the odds Kobe’s shot would fall in. We summarized that the odds of the shot being good would decreae relatively lineraly as a function of the distance from the basket.

The dataset contained several datpoints related to distance which we explored individually and in concert with other variables. Two predictors, y\_loc and shot\_distance were highly colinear, and often were interchangeable in a model. Most often however they interfered with each other when used together in the same model.

## loc\_y shot\_distance  
## loc\_y 1.000000 0.818124  
## shot\_distance 0.818124 1.000000

An analysis of the correlation matrix between the two predictors yielded a colineraity of about .8.

Alogistic regressionmodel consisting of the predictors shot\_distance, combined\_shot\_type, and shot\_type appeared to yield the most signifigant results.

##   
## Call:  
## glm(formula = shot\_made\_flag ~ shot\_distance + combined\_shot\_type +   
## shot\_type, family = "binomial", data = kobe)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.2953 -1.0223 -0.8908 1.2956 1.6256   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.488234 0.226523 6.570 5.04e-11 \*\*\*  
## shot\_distance -0.017137 0.003066 -5.589 2.29e-08 \*\*\*  
## combined\_shot\_typeDunk 1.071499 0.255857 4.188 2.82e-05 \*\*\*  
## combined\_shot\_typeHook Shot -1.250658 0.286975 -4.358 1.31e-05 \*\*\*  
## combined\_shot\_typeJump Shot -1.590542 0.226130 -7.034 2.01e-12 \*\*\*  
## combined\_shot\_typeLayup -1.219195 0.228348 -5.339 9.34e-08 \*\*\*  
## combined\_shot\_typeTip Shot -2.109243 0.283277 -7.446 9.63e-14 \*\*\*  
## shot\_type3PT Field Goal -0.171709 0.048310 -3.554 0.000379 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 35325 on 25696 degrees of freedom  
## Residual deviance: 33474 on 25689 degrees of freedom  
## AIC: 33490  
##   
## Number of Fisher Scoring iterations: 5

## 2.5 % 97.5 %  
## (Intercept) 1.04425637 1.93221067  
## shot\_distance -0.02314623 -0.01112698  
## combined\_shot\_typeDunk 0.57002858 1.57296873  
## combined\_shot\_typeHook Shot -1.81311758 -0.68819745  
## combined\_shot\_typeJump Shot -2.03374826 -1.14733591  
## combined\_shot\_typeLayup -1.66674848 -0.77164147  
## combined\_shot\_typeTip Shot -2.66445564 -1.55402947  
## shot\_type3PT Field Goal -0.26639391 -0.07702393

From the results in the table above, shot\_distance, combined\_shot\_type, and shot\_type all appear to be statistically signifigant.

## Wald test:  
## ----------  
##   
## Chi-squared test:  
## X2 = 501.9, df = 5, P(> X2) = 0.0

The chi-squared test statistic of cstwald$result$chi2["chi2"], with cstwald$result$chi2["df"] degrees of freedom is associated with a p-value << 0 indicates that the overall effect of combined\_shot\_type is statistically significant.

## Wald test:  
## ----------  
##   
## Chi-squared test:  
## X2 = 12.6, df = 1, P(> X2) = 0.00038

Furthermore the chi-squared test statistic of 12.6, with 1 degree of freedom and a p-value .0038 indicates that the overall effect of shot\_type is statistically significant as well.

## Waiting for profiling to be done...

|  |  |  |  |
| --- | --- | --- | --- |
|  | OR | 2.5 % | 97.5 % |
| (Intercept) | 4.4292644 | 2.8892509 | 7.0488882 |
| shot\_distance | 0.9830094 | 0.9771044 | 0.9889199 |
| combined\_shot\_typeDunk | 2.9197519 | 1.7421680 | 4.7674689 |
| combined\_shot\_typeHook Shot | 0.2863165 | 0.1610940 | 0.4976780 |
| combined\_shot\_typeJump Shot | 0.2038151 | 0.1281629 | 0.3122020 |
| combined\_shot\_typeLayup | 0.2954679 | 0.1850569 | 0.4547038 |
| combined\_shot\_typeTip Shot | 0.1213298 | 0.0685775 | 0.2087289 |
| shot\_type3PT Field Goal | 0.8422243 | 0.7661417 | 0.9258792 |

For a one unit increase in shot\_distance, the odds of Kobe making his shot decreases by 2%.