DS 6372

Black Mamba

Applying Binomial Linear Regression models to predict success of Kobe's field goals

Hannah Kosinovsky, Yat Leung, and Javier Saldaña
4-10-2019

Kobe Shot Selection

Introduction

At the age of 18, Kobe Bryant was a basketball sensation straight out of high school. Having practiced with the Los Angeles Lakers at the age of 17, his parents signed a contract on his behalf. The Lakers and the Hornets organized a trade in which Kobe Bryant would be drafted by the Hornets and then traded to the Lakers before the beginning of the season. Kobe then became the youngest guard ever drafted into the NBA. From that point on, Kobe played his entire career with the Los Angeles Lakers. He went on to earn astonishing records and accolades in professional sports. His talent and production has earned Kobe a legendary status among other great NBA legends. His skills has been studied by many in an attempt to try to replicate his performance to other players.

In this project, we have the ability to study the production of Kobe Bryant's career. With the given dataset, we look to determine whether distance affected the odds ratio and probability of a successful field goal. In addition, we look to see if that same relationship varies when in a playoff game versus a regular season game. Finally, we take the model and, using stepwise forward selection, attempt to predict whether a field goal would be successful based on the given variables.

Data Description

The dataset used for this project is sourced from the [Kaggle Kobe Bryant Shot Selection Competition](). The data set includes every single field goal attempt by Kobe Bryant during his 20-year career with the Los Angeles Lakers. Out of the 30,697 field goal attempts, 5000 (19.5% of entire data set) were extracted into a test data set. The remaining 25,697 attempts were used to train the models.

The data set contains 29 variables, which range between continuous and discrete variables. The variables pertain to information regarding the environment, location, field goal description, and opponents. The variables of interest will be *shot_made_flag* (describes whether a shot was successful), *shot_distance* (distance from shot to basket), and *playoffs* (describes if regular season game or playoff game). Considering each attempt is independent of each other, we will assume independence for this study.

Detailed Data Content Summary

As mentioned above, this dataset has 25697 observations and 29 variables.

Variable	**Type**	**Value Summary**	**Description**
recID	Num	vals 1-30692	unique id for record
action_type	Chr	ex:("Jump Shot" "Tip Shot")	type of shot taken
combined_shot_type	Chr	ex:("Jump Shot" "Layup)	type of combined shot taken
game_event_id	Num	values 2-653	id of event assoc. with shot
game_id	Num	vals 20000012-49900088	id of a specific game
lat	Num	vals 33.2533-34.0883	latitude of shot attempt
loc_x	Num	vals -250-248	(x)location of shot attempt by grid
loc_y	Num	vals -44-791	(y)location of shot attempt by grid
lon	Num	vals -118.5198- -118.0218	longitude of shot attempt
minutes_remaining	Num	vals 0-11	num minutes remaining in period during shot
period	Num	vals 1-7	period of game during shot attempt
playoffs	Num	vals 0 or 1	binary: whether or not game is playoff

season	Chr	ex("2000-01")	game season ie 2000-2001 when shot occurred
seconds_remaining	Num	vals 0-59	num seconds remaining in period during shot
shot_distance	Num	vals 0-79	distance in feet of shot attempt
shot_made_flag	Num	vals 0 or 1	binary: whether or not shot was made
shot_type	Chr	ex("2PT Field Goal")	technical term for type of shot
shot_zone_area	Chr	ex("Left Side(L)")	zone location on court of shot attempt
shot_zone_basic	Chr	ex("Mid-Range")	approximate zone on court of shot attempt
shot_zone_range	Chr	ex("8-16 ft.")	approximate zone on court of shot attempt
team_id	Num	single val: 1610612747	distance in feet of shot attempt zone location
team_name	Chr	unique val:"Los Angeles Lakers"	NBA team league name
game_date	Num	MMDDYY format	date game occurred
matchup	Chr	ex("LAL @ POR")	the two teams playing each other
opponent	Chr	ex("POR")	the opposing team of the game
shot_id	Num	vals 2-30697	id of shot attempt within game
attendance	Num	vals 11065-20845	audience number during game
arenatemp	Num	vals 64-79	temperature of arena during game
avgnoisedb	Num	vals 88.56-102.43	Average noise in db during game

Using this dataset, no causal inference can be made, since the data set was obtained through observing Kobe Bryant's 20-year professional career. Therefore, the findings can only aid in predicting if he will make the next shot or not and aren't applicable to other players.

Exploratory Data Analysis

Since our analysis is based on logistic regression, our variables don't have to be normality distributed to fit in the model, therefore no transformations are necessary.

In this analysis, multicollinearity between variables is of concern so correlation between variables is examined. From Figure 1, we can see that loc_y and shot_distance as well and lat and shot_distance are highly correlated and therefore we don't need to include all those variables in the models we create. For our analysis we decided to remove lat from any models we fit to avoid multicollinearity between explanatory variables.

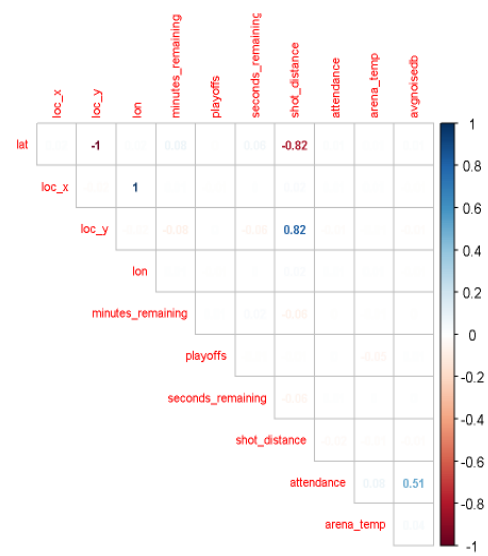
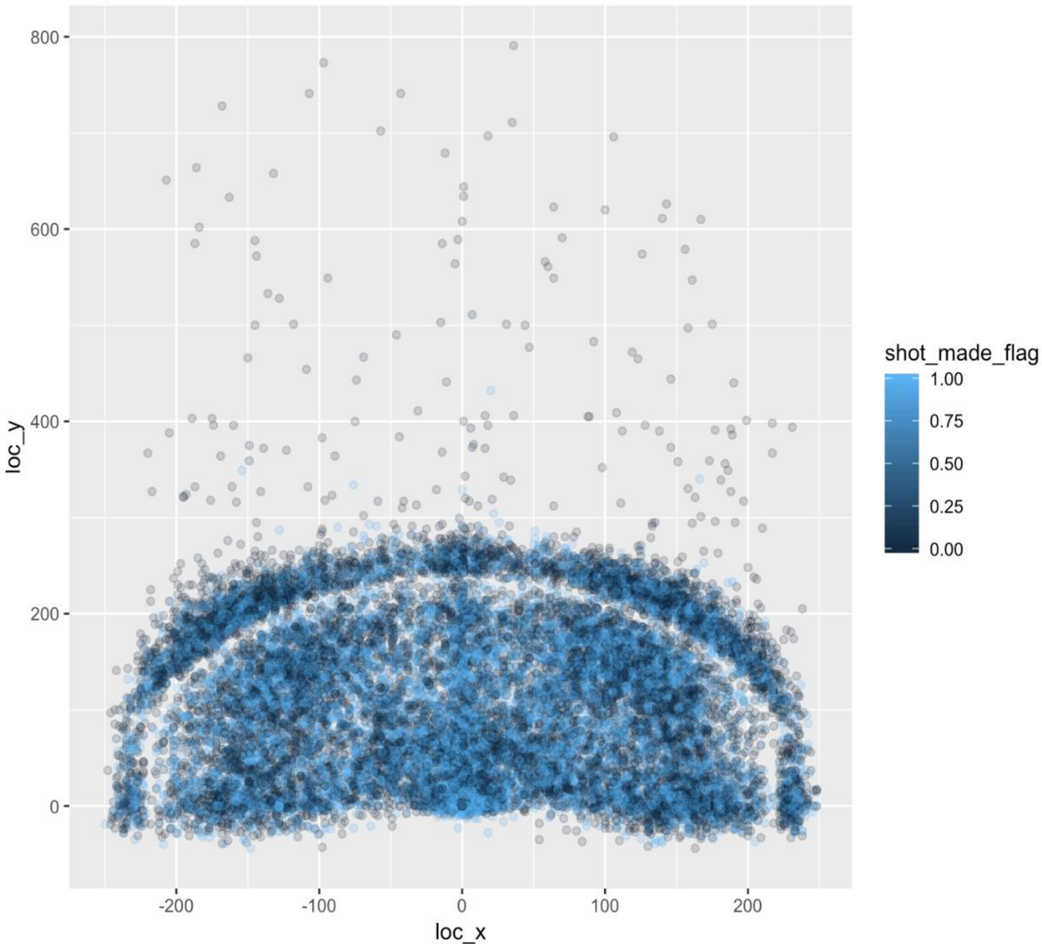


Figure 1 - Correlation Matrix

Since our dataset has over 1000 data points and we anticipate the central limit theorem to apply. The main analysis of this project consists of exploring the relationship between the distance and whether a shot was made. For an initial glance at this data, we decided to visualize the court distance locations versus the shot made flag variable. This can be viewed in Figure 2 below:



From looking at figure 2, we can preliminarily see a visual cluster closer around where the hoop would be at location (0,0) but the dark blue(shot not made) and light blue(shot made) indicators exist at all points of the court and don't appear clustered in a particular way.

Figure 2 - Plot of shots made and location

Analysis:

Build models to provide arguments and evidence for or against the propositions below:

Odds of Field Goal Made v. Distance from Basket

For our first hypothesis, we look to compile a model in which we will look to distance and determine if its effect is significance and what the odds are of a successful field goal. The first model looks at all the variables (regardless of collinearity) and presents an odds ratio of 1.010 with a 95% confidence interval of 0.991 and 1.029. This would suggest that with regards to distance, the odds of a successful field goal is 1.010 with a 95% confidence interval between 0.991 and 1.029. This would suggest the mean odd of a successful field goal increase by 1.010 times with every unit increase of distance. Most importantly, we point toward the fact that the p-value is above 0.05, which would suggest the distance variable is not significant in this model. The results are shown in Figure 3 below.

Full Model

Odds Ratio Estimates and Wald Confidence Intervals			
Odds Ratio	Estimate	95% Confidence Limits	
shot_distance	1.010	0.991	1.029

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
action_type	46	1834.8480	<.0001
combined_shot_type	0	.	.
matchup	69	101.6259	0.0065
opponent	0	.	.
season	11	15.2362	0.1719
shot_type	1	0.0268	0.8700
shot_zone_area	5	34.0368	<.0001
shot_zone_basic	6	20.2411	0.0025
shot_zone_range	2	15.9026	0.0004
arena_temp	1	3.8245	0.0505
attendance	1	77.3901	<.0001
avgnoisedb	1	0.9266	0.3357
game_date	1	0.5975	0.4395
minutes_remaining	1	12.0619	0.0005
period	1	24.4270	<.0001
playoffs	1	0.1074	0.7432
seconds_remaining	1	10.2733	0.0013
shot_distance	1	1.1187	0.2902

Simplified Model

Odds Ratio Estimates and Wald Confidence Intervals			
Odds Ratio	Estimate	95% Confidence Limits	
shot_distance	0.969	0.952	0.986

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
action	1	276.5885	<.0001
period	1	4.7552	0.0292
playoffs	1	0.0937	0.7595
season	11	13.9284	0.2370
time_remaining	1	10.3643	0.0013
shot_distance	1	12.0334	0.0005
shot_type	1	0.0294	0.8638
shot_zone_area	5	14.7072	0.0117
shot_zone_basic	6	13.1413	0.0408
shot_zone_range	2	16.4571	0.0003
game_date	1	0.0005	0.9827
opponent	31	27.9176	0.6254
attendance	1	75.5634	<.0001
arena_temp	1	9.2452	0.0024
avgnoisedb	1	1.8231	0.1769
home_away	1	0.0861	0.7693

Figure 3 - Odds Ratio Estimate output from SAS

When looking at the simplified model, we see that the time variables have been collapsed into time_remaining. This was done by converting the minutes_remaining variable to seconds and then combining the columns. The action_type was also collapsed and merged with combined_shot_type since both were indicators of whether the distance was short or other. Matchup was also reduced to 2 levels instead of the 70 it was previously under. This was performed since the opponent element already captures who the opposing team is, which would indicate the only value the matchup column would provide would be whether the game was home or away. When we performed the binomial regression under this simplified model, we see that the odds ratio, while still relatively close, now illustrates that the mean odds of a successful basket are no longer increasing. Furthermore, the distance is now significant (p-value < 0.05) in the simplified model.

Comparing both models, the simplified model would actually make the more sense. It is much more probable for the odds of a successful basket to actually decrease as the distance increases. While this is especially true for basketball players, athletes like Kobe Bryant are excellent players because they know how to play effectively within their means, which would make them the most efficient players as well. Kobe was well known for creating space between him and the defense in order to be able to improve his odds of making the field goal.

Probability of Kobe making a shot respect to the distance

For this problem we want to use logistic regression since the response variable is whether or not Kobe makes the shot (binary). We can use a generalized linear model to predict the membership in the target group (whether or not a shot is made). Our goal is to model the probability that shot_made_flag = 1. Our x variable is distance so our p(x) modeled in the logistic regression model is the probability of a shot made at distance x.

When looking at the simple logistic regression model estimates in Figure 4, the p-values for the estimates of the intercept and slope are extremely small and therefore significant.

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.3681	0.0224	270.3278	<.0001
shot_distance	1	-0.0441	0.00141	983.1629	<.0001

Figure 4 - Coefficient Estimate Output from SAS

Since these are significant, we know that the probability of making a shot differs with various distances. In this case, the slope is negative, so the larger the distance is, the less chance there is that the shot is made.

Relationship between the distance and odds of making the shot in the playoffs.

To determine the relationship between distance from basket and the odds of making the shot, we approached the problem by analyzing the quantitative continuous variables with component analysis. Without eliminating any potential multicollinear variables, the Scree plot below illustrated that approximately 52% of the variances can be explained with the first two dimensions. On the right, we plotted the variables and their contributions to each dimension to the left. (Figure 5 & 6)

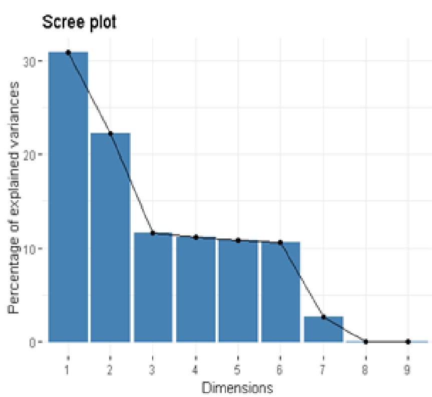


Figure 5 - Scree Plot

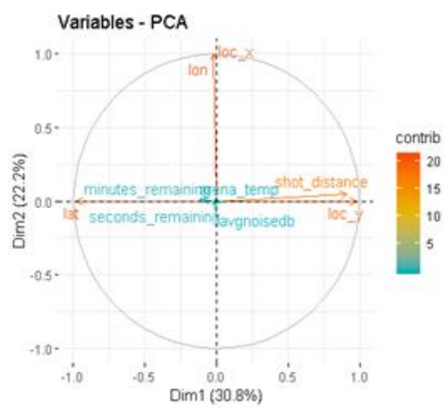


Figure 6 - Variables PCA

Our assumption that lat and loc_x were multicollinear. With lat removed from the set, we refitted the Scree plot resulted in 48% of the variances can be explained with the first two dimensions. (Figure 7 & 8)

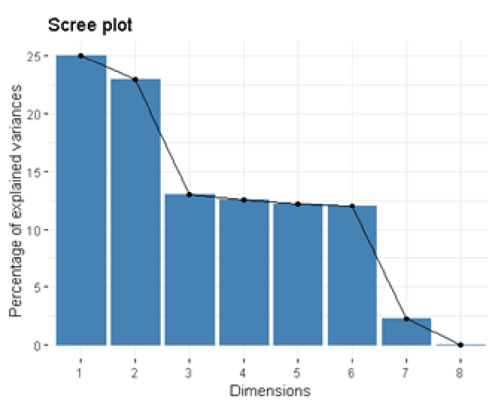


Figure 7 - Scree plot without lat

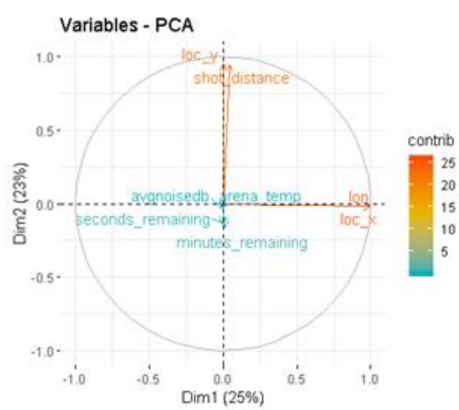


Figure 8 - Variables PCA without lat

We performed a Pearson's Chi-squared test for variables independence for the qualitative variables *action_type* and *combined_shot_type*. Our null assumption was that the two categorical variables were independent whereas the alternative would be that the pairs were not independent therefore exhibiting co-linearity. Since the resulting p-value was less than α 0.05, we rejected the null hypothesis that these two variables were independent. (Table 1)

Table 1

Categorical variables		**p-value**
action_type	combined_shot_type	< 2.2e-16

X-squared = 128480, df = 270

In our logistic model, we fitted the coordinations (loc_x, loc_y, lon), distance (shot_distance), playoffs factor, and the distance factors (shot_zone_area, shot_zone_basic, shot_zone_range) then we performed a stepwise procedure to further tapering the variables for a more parsimonious model. Note that loc_x, loc_y, lon and playoffs were eliminated from the stepwise procedure

Base logistic model

```
glm(shot_made_flag~loc_x+loc_y+lon+playoffs+shot_distance+shot_zone_area+shot
_zone_basic+shot_zone_range,family=binomial(link="logit"),data=kobe)
```

Stepwise logistic model

```
glm(formula = shot_made_flag ~ shot_distance + shot_zone_area + 
shot_zone_basic + shot_zone_range, family = binomial(link = "logit"), data = 
kobe)
```

Odds Ratio

	Odds ratio	2.5 %	97.5 %
(Intercept)	1.962484e-05	3.713612e-85	1.037088e+75
shot_distance	9.594545e-01	9.447178e-01	9.744210e-01
shot_zone_areaCenter(C)	7.100505e+04	1.344892e-75	3.748791e+84
shot_zone_areaLeft Side Center(LC)	6.974884e+04	1.321102e-75	3.682456e+84
shot_zone_areaLeft Side(L)	5.799215e+04	1.098375e-75	3.061877e+84
shot_zone_areaRight Side Center(RC)	7.491551e+04	1.418964e-75	3.955233e+84
shot_zone_areaRight Side(R)	5.987611e+04	1.134059e-75	3.161341e+84
shot_zone_basicBackcourt	8.958763e+03	1.678990e-76	4.780221e+83
shot_zone_basicIn The Paint (Non-RA	1.138011e+00	9.275034e-01	1.396295e+00
shot_zone_basicLeft Corner 3	1.343206e+00	1.005611e+00	1.794135e+00
shot_zone_basicMid-Range	1.091381e+00	9.565173e-01	1.245259e+00
shot_zone_basicRestricted Area	1.918734e+00	1.436307e+00	2.563198e+00
shot_zone_basicRight Corner 3	1.127703e+00	8.730093e-01	1.456700e+00
shot_zone_range24+ ft.	NA	NA	NA
shot_zone_range8-16 ft.	9.383795e-01	8.265318e-01	1.065363e+00
shot_zone_rangeBack Court Shot	NA	NA	NA
shot_zone_rangeLess Than 8 ft.	6.155917e-01	4.917847e-01	7.705673e-01

Table 2

Both *shot_zone_range24+ ft.* and *shot_zone_rangeBack Court Shot* did not have a odds ratio and the 95% CIs because these two variables were perfectly correlated. The odds of Bryant making the shot increased when the shot zone was between 8-16 ft (OR=9.38^{-01}, 95% CI=[8.27^{-01},1.07]) and when the shot zone was less than 8 ft (OR=6.16^{-01}, 95% CI=[4.92^{-01},7.71^{-01}]). (Table 2)

Prediction

Using the stepwise model, we predicted Bryant's odds of making the shot (shot_made_flag, binary 1/0). Because the predicted values were the probability values of making the shot (between 0 and 1), therefore we had to explicitly recode the probability values using a threshold of 0.5 (values > 0.5 would be classified as 1 else into 0).

Sensitivity and Specificity

The sensitivity (True Positive Rate) based on our stepwise model was 0.826588 and the specificity (True Negative Rate) was 0.3344963 (See Figure 10 for the confusion matrix). The calculated area under the curve was 0.6144 (Figure 9) and the calculated log loss function at 13.57258.

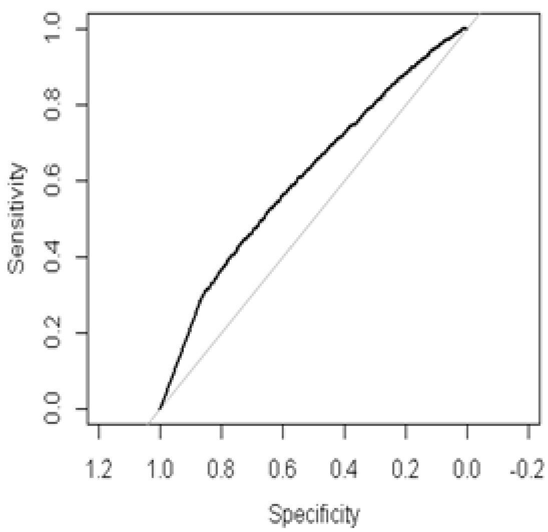


Figure 9 - Sensitivity v. Specificity

```
                 actual_values
predicted_values     0     1
               0 11764  7630
               1  2468  3835
```

Figure 10 - Confusion Matrix

Code

SAS Code

```
ods graphics on;
/*Perform binomial linear regression using all variables*/
title 'Binomial Linear Regression on Kobe Field Goals Data with All Relevant
Variables';
proc logistic data = KOBEDATA outmodel=model1 plots=all;
class shot_made_flag action_type combined_shot_type matchup opponent season
shot_type shot_zone_area shot_zone_basic shot_zone_range / param = ref;
model shot_made_flag (event='1') = action_type combined_shot_type matchup
opponent season shot_type shot_zone_area shot_zone_basic shot_zone_range
arena_temp attendance avgnoisedb game_date minutes_remaining period playoffs
seconds_remaining shot_distance / lackfit ctable;
oddsratio shot_distance;
run;

/*Perform binomial linear regression using combined variables*/
title 'Binomial Linear Regression on Kobe Field Goals Data with Combined
Variables';
proc logistic data = KOBEDATA outmodel=model2 plots=all;
class shot_made_flag action  playoffs home_away opponent season shot_type
shot_zone_area shot_zone_basic shot_zone_range / param = ref;
model shot_made_flag (event='1') = action period playoffs season
time_remaining shot_distance shot_type shot_zone_area shot_zone_basic
shot_zone_range game_date opponent attendance arena_temp avgnoisedb home_away
/ lackfit ctable;
oddsratio shot_distance;
run;

/*Use Stepwise Regression to find ideal model to predict shots*/
title 'Stepwise Regression on Kobe Field Goals Data';
proc logistic data = KOBEDATA outest=betas outmodel=model3 plots = all;
class shot_made_flag action_type combined_shot_type matchup opponent season
shot_type shot_zone_area shot_zone_basic shot_zone_range / param = ref;
model shot_made_flag(event='1') = action_type combined_shot_type matchup
opponent season shot_type shot_zone_area shot_zone_basic shot_zone_range
arena_temp attendance avgnoisedb game_date minutes_remaining period playoffs
seconds_remaining shot_distance
      / selection = stepwise details lackfit ctable;
output out=pred p=phat lower=lcl upper=ucl predprob=(individual
crossvalidate);
```

```
run;
ods graphics off;

*print betas and predictions;
proc print data=betas;
title2 'Parameter Estimates and Covariance Matrix';
run;

proc print data=pred;
title2 'Predicted Probabilities and 95% Confidence Limits';
run;

/*Predict shot using Model 1*/
proc logistic inmodel = model1;
title 'Kobe Field Goal Predictions Based on Binomial Linear Model_1';
score data = KOBE_PREDS out=KobePreds1;
run;
proc print data = KobePreds1;
run;

/*Predict shot using Model 2*/
proc logistic inmodel = model2;
title 'Kobe Shot Predictions Based on Binomial Linear Model_2';
score data = KOBE_PREDS out=KobePreds2;
run;
proc print data = KobePreds2;
run;

/*Predict shot using Model 3*/
proc logistic inmodel = model3;
title 'Kobe Shot Predictions Based on Binomial Linear Model_3 (Stepwise
Regression)';
score data = KOBE_PREDS out=KobePreds3;
run;
proc print data = KobePreds3;
run;

/*Perform simple logistic regression for prob of shot made*/
PROC LOGISTIC DATA = Kobe DESCENDING;
MODEL shot_made_flag = shot_distance / LACKFIT CTABLE;
TITLE 'Kobe Shot Data';
RUN;
```

R code

```
#load data
require(openxlsx)
Kobe = read.xlsx(xlsxFile="~/Downloads/Project2_2/project2Data.xlsx", sheet =
1, startRow = 1, colNames = TRUE)

str(Kobe)
require(ggplot2)

#visualize relationship bw shot made and distance
ggplot(data = Kobe)+
  geom_point(aes(x = loc_x, y = loc_y, color = shot_made_flag),alpha = 1 / 5)

library(readxl)
library(tidyverse)
library(aod)
library(caret)
library(glmnet)
library(corrplot)
library(MASS)
library(regclass)
library(FactoMineR)
library(factoextra)
library(pROC)
library(psych)
library(MLmetrics)
xlsx_kobe <- read_excel("C:\\Users\\Yat\\Documents\\MSDS\\MSDS 6372\\Project
2\\project2Data.xlsx")

#Take out an ID column from when data was imported
kobe<-xlsx_kobe[,c(2:29)]

#Re-code the character columns into factors
kobe<-as.data.frame(unclass(kobe))
kobe$shot_made_flag<-as.factor(kobe$shot_made_flag)
kobe$shot_type<-as.factor(kobe$shot_type)
kobe$shot_zone_area<-as.factor(kobe$shot_zone_area)
kobe$shot_zone_basic<-as.factor(kobe$shot_zone_basic)
kobe$shot_zone_range<-as.factor(kobe$shot_zone_range)
kobe$game_date<-as.factor(kobe$game_date)
kobe$season<-as.factor(kobe$season)
kobe$period<-as.factor(kobe$period) #change period into factor?
kobe$playoffs<-as.factor(kobe$playoffs)
```

```
#subsetting variables (sans IDs)
kobe<-kobe[,c(1:2,5:20,22:25,27:28)]
kobe<-na.omit(kobe)

#data check
summary(kobe)
#numerical variables
kobe.NV<-kobe[,c(3:7,11:12,23:24)]
kobe.NV<-na.omit(kobe.NV)
fit<-prcomp(~., data=kobe.NV, cor=TRUE)
summary(fit) # print variance accounted for
loadings(fit) # pc loadings
plot(fit,type="lines") # scree plot
fit$scores # the principal components
biplot(fit,expand=10, xlim=c(-0.15, 0.05), ylim=c(-0.1, 0.05))
# Varimax Rotated Principal Components
# Extract, rotate and retain 5 PCs
component.retained <- principal(kobe.NV, nfactors=5, rotate="varimax")
component.retained
# Principal Axis Factor Analysis
axis.fit <- factor.pa(kobe.NV, 5)
axis.fit
#PCA
res.pca1 <- prcomp(kobe.NV, scale = TRUE)
fviz_eig(res.pca1)
fviz_pca_var(res.pca1,
             col.var = "contrib", # Color by contributions to the PC
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE     # Avoid text overlapping
             )
#numerical variables with lat removed
kobe.NV2<-kobe[,c(4:7,11:12,23:24)]
kobe.NV2<-na.omit(kobe.NV2)
fit2<-prcomp(~., data=kobe.NV2, cor=TRUE)
summary(fit2) # print variance accounted for
loadings(fit2) # pc loadings
plot(fit2,type="lines") # scree plot
fit2$scores # the principal components
biplot(fit2,expand=10, xlim=c(-0.15, 0.05), ylim=c(-0.1, 0.05))
# Varimax Rotated Principal Components
# Extract, rotate and retain 5 PCs
component.retained2 <- principal(kobe.NV2, nfactors=5, rotate="varimax")
```

```
component.retained2
# Principal Axis Factor Analysis
axis.fit2 <- factor.pa(kobe.NV2, 5)
axis.fit2
#PCA
res.pca2 <- prcomp(kobe.NV2, scale = TRUE)
fviz_eig(res.pca2)
fviz_pca_var(res.pca2,
             col.var = "contrib", # Color by contributions to the PC
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE     # Avoid text overlapping
             )
#action_type and combined_shot_type
tbl.1<-table(kobe$action_type, kobe$combined_shot_type)
chisq.test(tbl.1)
#shot_made_flag to distance, playoffs and coordinations
glm.fit<-
glm(shot_made_flag~loc_x+loc_y+lon+playoffs+shot_distance+shot_zone_area+shot
_zone_basic+shot_zone_range,family=binomial(link="logit"),data=kobe)

summary(glm.fit)
#Stepwise based on the first logit regression
stepwise<- stepAIC(glm.fit,direction="both",trace = FALSE)
summary(stepwise)
#odds ratio calculation using library(epiDisplay)
exp(cbind("Odds ratio" = coef(stepwise), confint.default(stepwise, level =
0.95)))
#load the project2Pred.xlsx
pred_kobe <- read_excel("C:\\Users\\Yat\\Documents\\MSDS\\MSDS 6372\\Project
2\\project2Pred.xlsx")
#Take out an ID column from when data was imported
pred_kobe<-pred_kobe[,c(2:29)]

#Re-code the character columns into factors
pred_kobe<-as.data.frame(unclass(pred_kobe))
pred_kobe$shot_made_flag<-as.factor(pred_kobe$shot_made_flag)
pred_kobe$shot_type<-as.factor(pred_kobe$shot_type)
pred_kobe$shot_zone_area<-as.factor(pred_kobe$shot_zone_area)
pred_kobe$shot_zone_basic<-as.factor(pred_kobe$shot_zone_basic)
pred_kobe$shot_zone_range<-as.factor(pred_kobe$shot_zone_range)
pred_kobe$game_date<-as.factor(pred_kobe$game_date)
pred_kobe$season<-as.factor(pred_kobe$season)
pred_kobe$period<-as.factor(pred_kobe$period) #change period into factor?
```

```r
pred_kobe$playoffs<-as.factor(pred_kobe$playoffs)

#subsetting variables (sans IDs)
pred_kobe<-pred_kobe[,c(1:2,5:20,22:25,27:28)]
pred_kobe<-na.omit(pred_kobe)
#dataframe for the shot_made_flag column in the pred_kobe
pred_shot_made_flag<-pred_kobe
predictions <- predict(stepwise, pred_shot_made_flag, type="response")
predictions<-as.data.frame(ifelse(predictions>0.5,"1","0"))
colnames(predictions) <- "Predicted shot_made_flag"
#actual and predicted values from model
threshold=0.5
predicted_values<-ifelse(predict(stepwise,type="response")>threshold,1,0)

actual_values<-stepwise$y

#confusion matrix using the training set
conf_matrix<-table(predicted_values,actual_values)
conf_matrix
#Sensitivity of the model
sensitivity(conf_matrix)
#specificity of the model
specificity(conf_matrix)
predicted_prob<-predict(stepwise,type="response")
roccurve <- roc(actual_values, predicted_prob)
plot(roccurve)
#AUC
auc(roccurve)
#Log Loss Function
LogLoss(predicted_values,actual_values)
```