**Part 1: Executive Summary**

**Project Problem/Purpose:**

The purpose of the following sections is to choose a logistic and linear regression model to predict both score and if a player is in the top 25. The analyses will be made using a random sample of 150 from the given data of over 1600 observations. These models can be used throughout a season of golf to make predictions about certain players and can give both players and sponsors insight into player performance.

**Assumption analysis:**

The initial linear model chosen did not follow two of the four assumptions, namely the constant variance and normality of residuals. A transformation could have been done to fix these, but a change in model was decided to be better. The updated model follows all assumptions while being easy to read and to use for estimation.

**Model analyses:**

The chosen linear model only uses the Putt, ARG, and OTT predictors. It shows that for an increase in any one area of these variables that score is expected to decrease. This follows what was expected and strengthens the validity of the model. Outliers were detected, but shown to not be significant so they were not removed from the linear model. The linear model also did not show any abnormalities such as multicollinearity and confidence intervals including 0 for slopes. The predictors used in the linear model were carried over to the logistic model. This model shows that for an increase in any one area of the predictor variables, the probability that the player is in the top 25 increases as well. This also falls within expectations. There were two models to choose from but the one with the probit link was chosen because it had a lower AIC value.

**Overall conclusion:**

Using these models, estimations were made for both score and predicting top 25 placement. The linear model estimated that the average overall score will be between -24.558 and -13.599, with the fitted value being about -19.1. It also estimated that an individual will have a score between about -30.514 and -7.643. The logistic model predicted that the probability of a player with the given values for each predictor being in the top 25 is nearly 100%. With both these estimates it is clear that if a player had these values for Putt, ARG, and OTT that they would be having a very good season.

**Comments:**

There are two major comments regarding this analysis. The first is with the chosen linear model. The initial model was determined to be better based on the values in Figures 1, 2, and 3, but the updated model was determined to be better because it would be easier to read and understand than a transformed initial model. The second is with the logistic model where the ARG predictor was determined to not be significant when added sequentially to the model. The model was not altered so that both models have the same predictors making a more cohesive analysis. The model was also kept because all predictors were determined significant when added last, as well as because each confidence interval for the slopes did not include 0.

**Part 2: Linear Regression Application**

**Introduction:**

This section will go through the process in which the chosen linear regression model was determined followed by a closer look at the model itself. A 5% significance level will be used at all stages of this analysis. This will include looking at any assumptions the model(s) might violate, outlier detection, and analysis if outliers are significant or not. This will be followed by a closer look at the chosen regression model and its slopes. Once all of this is examined, Score will be estimated using the given values for the Putt, ARG, APP, OTT, and T2G predictor variables. Keep in mind that the chosen model will not include all of these predictors.

**Initial Model Selection, Outlier Detection, Multicollinearity Check:**

The initial model chosen was based on two main things: automatic selection functions and Figures with different criteria. Based on Figures 1, 2, and 3, the model with two predictors-Putt and T2G-is the best. This was also the model chosen by an automatic selection function going forward and stepwise. Based on Figure 4 there are two observations that are considered outliers, 268 and 286. As shown in Tables 2 and 4, the outliers have little effect on the test decision for the model. This is also shown in Figure 5 where the two regression lines are very similar. Based on these Tables and the Figure, outliers were not removed from the initial model. Table 5 also shows that there is no multicollinearity between the predictors in the initial model since both VIF values are less than ten.

**Analysis on Initial Model:**

Based on Tables 2 and 3, both predictors are significant to the model no matter the order in which they are introduced. For this model to be considered good for prediction, it must also follow four assumptions: normality of its residuals, constant variance of its residuals, independence of its residuals, and that there is a linear relationship between the predictors and the response variable. As seen in Table 1 all the predictors have a linear relationship with the response of Score at a 5% significance level. Figure 8 suggests that the residuals of the initial model are independent since none of the vertical black lines cross the dotted blue lines. This is also shown in Table 10 where the Durbin-Watson test verifies that there is not sufficient evidence that the residuals are not independent, so this assumption is met. Figure 6 and Table 6 show that there is sufficient evidence that the residuals of the initial model do not follow a normal distribution. For the constant variance assumption Figure 7 may not immediately show that this assumption is violated, but according to Table 7 there is sufficient evidence that the residuals do not have constant variance at a 5% significance level. There is a possible solution to these violations, a log transformation of the response or the predictor variables. While this has the possibility of fixing these violations, all the variables have negative values which makes a log transformation impossible without adding a constant value to everything. With all of this in mind, a different model will be used to ensure simplicity and readability. Many other models were considered, but the model with three predictors-Putt, ARG, and OTT-struck a good balance. This model is ranked the 13th best based on the criteria in Figures 1, 2, and 3. The following section will show how each predictor is significant and that it follows all four assumptions.

**Analysis on Updated Model:**

Before looking at if the updated model follows all the assumptions, it must first show that all its predictors are significant. As shown in Tables 8 and 9, every predictor is significant both when added last to the model (as shown in Table 8) and when added sequentially to it (as shown in Table 9). A new model also means that outliers and multicollinearity must be considered again. Figure 12 shows that observations 262 and 286 are outliers. These outliers will once again not be removed because the test decisions in Tables 8 and 14 are the same. It should be noted that Figure 13 shows that the regression lines are slightly different for the same model with and without outliers. This means that predictions will be slightly different depending on the model used, but since each line has a positive linear relationship and the difference is small the outliers will still not be removed. Table 15 also shows that there is no multicollinearity present in the updated model since all VIF values are below 10. With this proved all the assumptions can be looked at. As stated in the initial model analysis, Table 1 shows that all the predictors have a linear relationship with the response variable of Score. Figure 9 suggests that the residuals of the updated model are not independent because the dotted blue line is crossed. This is in contrast to Table 11 where the Durbin-Watson test does not have sufficient evidence that the residuals are not independent at a 5% significance level. The assumption will be considered met because of the test, regardless of what Figure 9 suggests. Figure 10 suggests that the residuals of the updated model are from a normal distribution because all the points are within the blue bounds. This is supported by Table 12 which shows that with a p-value of 0.5054 that there is not sufficient evidence that the residuals do not come from a normal distribution, so this assumption is met. Figure 11 suggests that the residuals of the updated model have constant variance across all levels of each predictor because there is no discernable pattern in the plot. This is supported by Table 13 where the Breusch-Pagan test for constant variance has a p-value of .1184, which means there is not sufficient evidence that the residual variance is not the same across each predictor at a 5% significance level. Since the updated model follows all assumptions, it can be used to estimate values of Score. However the difference in R-squared values should be noted before predictions are made. Table 16 shows that the updated model explains approximately 47% of the variation in Scores while the initial model explains approximately 65%. While this difference seems significant, not having to do transformations makes the updated model easier to understand and make predictions with. This advantage makes up for the difference in R-squared values. With this acknowledged, Score can now be estimated with the given values of Putt, ARG, and OTT.

**Analysis of Regression Equation and Slopes:**

Table 18 shows that we are 95% confident that when controlling for ARG and OTT that one more unit in Putt score will change the Score by, on average, approximately -4.047 points to -2.601 points. It also shows that we are 95% confident that when controlling for Putt and OTT that one more unit in ARG score will change the Score by, on average, approximately -3.986 points to -1.265 points. It also shows that we are 95% confident that when controlling for Putt and ARG that one more unit in OTT score will change the Score by, on average, approximately -5.614 points to -2.864 points. All these intervals show that for a better score in different areas that the overall score will go down. Because a lower overall score is better, this is within expectations. Also none of these intervals include zero, so that also verifies that each variable is significant in predicting overall score. The overall regression equation will be Score = -2.1208 - -3.3237(Putt) - 2.6256(ARG) - 4.2395(OTT). With this Score will be estimated using the given values for Putt, ARG, and OTT.

**Estimation of Score:**

This section will use both confidence and prediction intervals to estimate score. The given values for the predictor variables are 0, 0, and 4 for Putt, ARG, and OTT respectively. It should be noted that a higher value in these variables is better. Based on Table 17 we are 95% confident that the average overall score will be between -24.558 and -13.599 for the given values of Putt, ARG, and OTT. Rounding to the nearest whole number the score will be between -25 and -14. Table 17 also shows that we are 95% confident that an individual with the given Putt, ARG, and OTT values will have a score of between -30.514 and -7.643. Rounding to the nearest whole number the individual will have a score from -31 to -8. Based on the regression equation in the last section and Table 17, the score is estimated to be about -19.

**Part 3: Logistic Regression Analysis**

**Introduction:**

This section will use the predictor variables in the linear model chosen in part 2 to create a logistic regression model to predict the Top25 variable. It will include a section on selecting a link, analysis on the chosen model and its slopes, and estimating Top25.

**Logistic Model Link Selection and Analysis:**

When choosing a logistic model two links were considered, logit and probit links. Using the three predictor variables in the chosen linear model-Putt, ARG, and OTT-two logistic models were made. Table 19 shows that every predictor is significant when added to the model last at a 5% significance level. This is contrasted by Table 20 that shows that the ARG variable is not significant when added after Putt. Tables 21 and 22 show that this trend is repeated for the logit link. Because the model with these predictors was determined to be the best in linear regression the same predictors will be used in the logistic model, but ARG not being significant sequentially should be kept in mind. Table 23 shows that the AIC value for the probit link is lower, so that is the model that will be used going forward.

**Logistic Model Equation and Slope Analysis:**

The regression equation will not be predicting the Top25 variable itself because the probit link gives a Z-score and not a probability. Table 24 shows that we are 95% confident that when controlling for ARG and OTT that one more unit in Putt score will change the Top25 Z-score by, on average, approximately 0.779 to 1.835. It also shows that we are 95% confident that when controlling for Putt and OTT that one more unit in ARG score will change the Top25 Z-score by, on average, approximately 0.155 to 1.381. It also shows that we are 95% confident that when controlling for Putt and ARG that one more unit in OTT score will change the Top25 Z-score by, on average, approximately 0.937 to 2.346. Since all these intervals do not include zero it supports the argument that all the predictor variables are significant. The overall regression equation is Z-score = -1.5777 + 1.2552(Putt) + 0.7326(ARG) + 1.5906(OTT). With this probability of being in the Top25 can now be estimated.

**Estimating Top25:**

The given values for each predictor are the same as in part 2, that is, Putt = 0, ARG = 0, and OTT = 4. This section will use both the regression equation along with a standard normal table and software output to determine the probability of being in the Top25. Using the regression equation in the last section the Z-score for Top25 will be approximately 4.78. Using a standard normal Table this corresponds to a probability of greater than 99.9% for being in the Top25. Table 25 draws the same conclusion where software was used to get the probability. Both of these approaches show that there is a very high, nearly 100%, probability that a person with the given values of Putt, ARG, and OTT will be in the Top25.

**Part 4: Appendix**

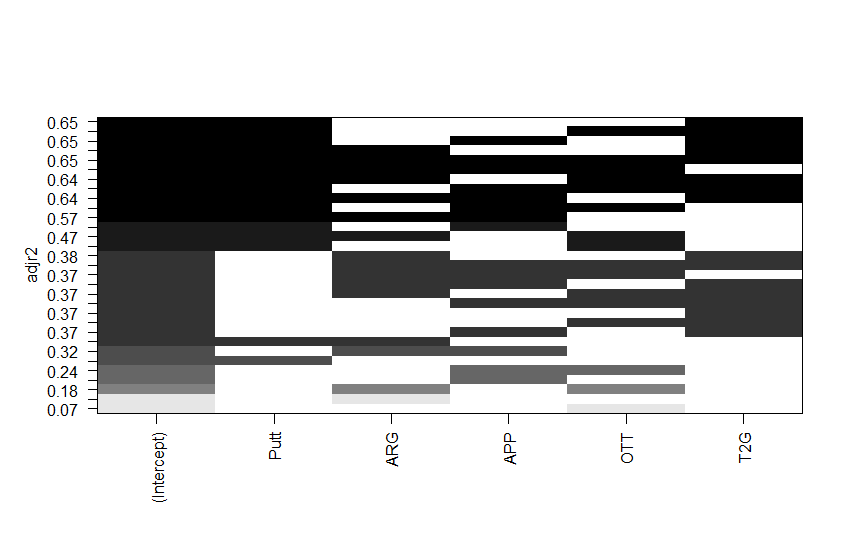
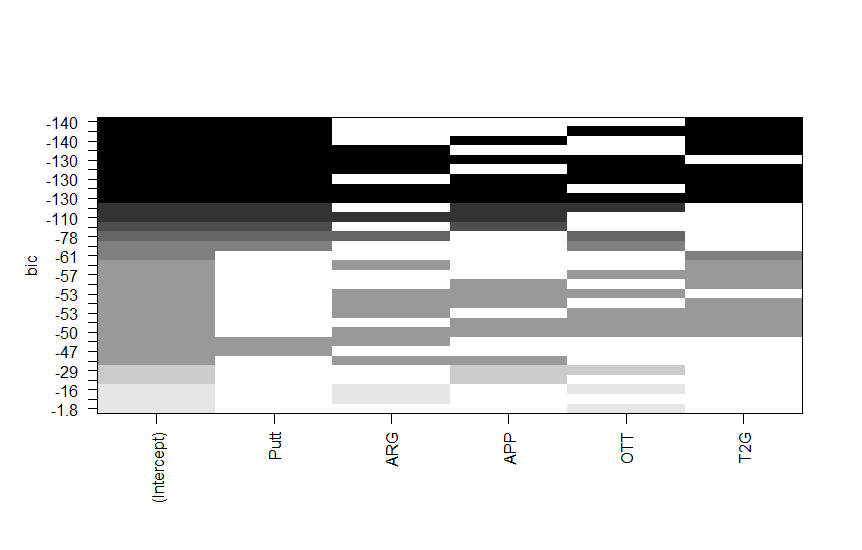
****

Figure 1: model selection Figure with adjusted R-squared criteria

Figure 2: model selection Figure with bic criteria

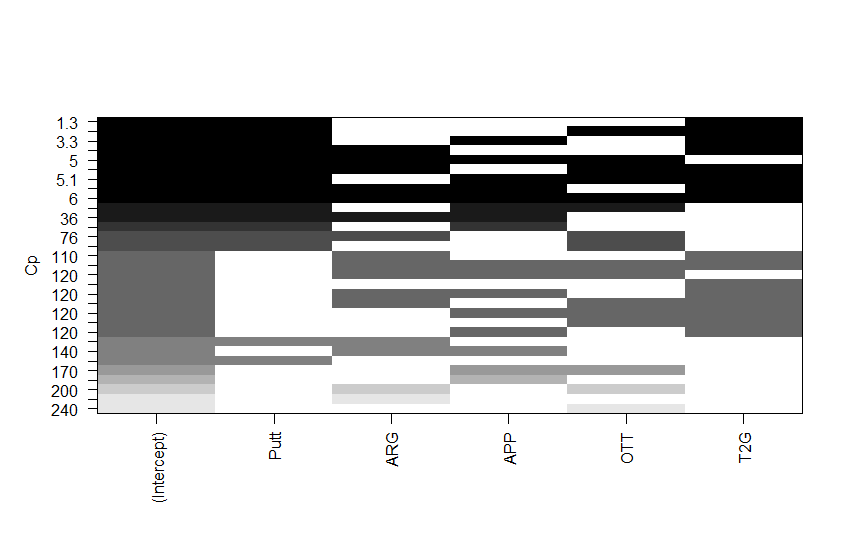


Figure 3: model selection Figure with cp criteria

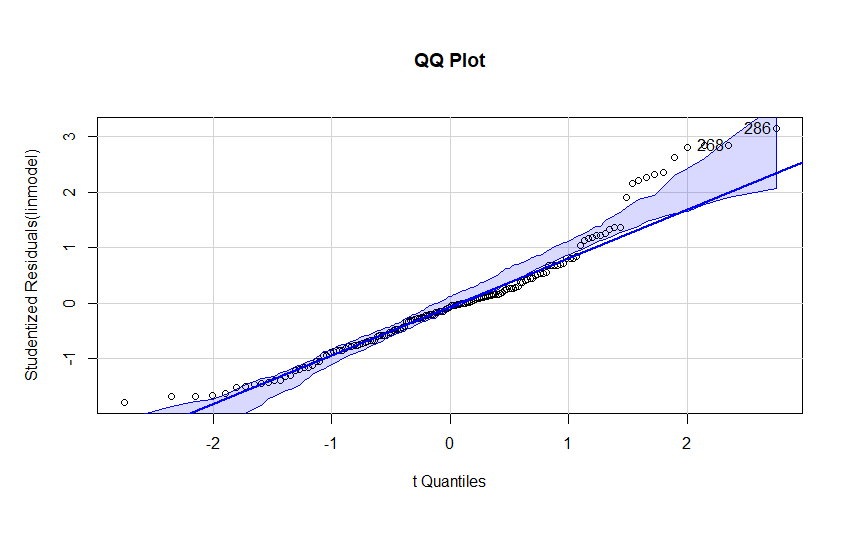


Figure 4: QQ plot of initial model

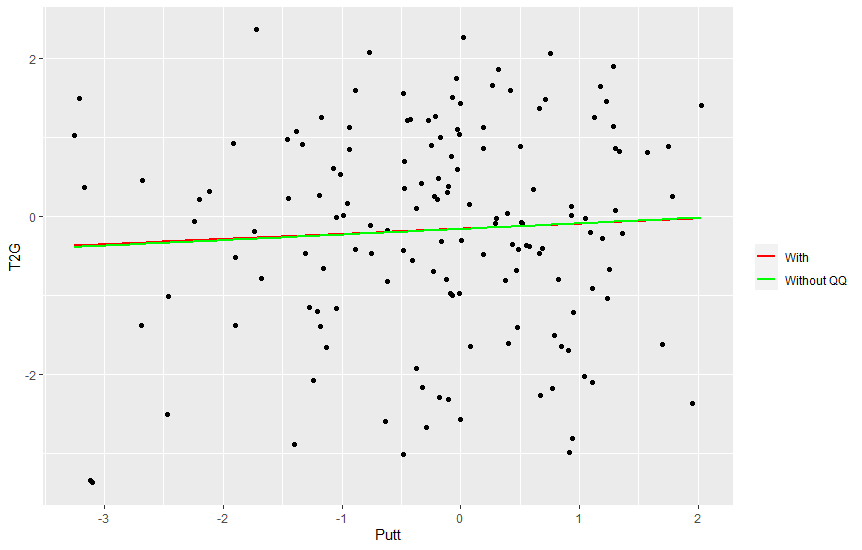


Figure 5: Plot of initial model with and without outliers

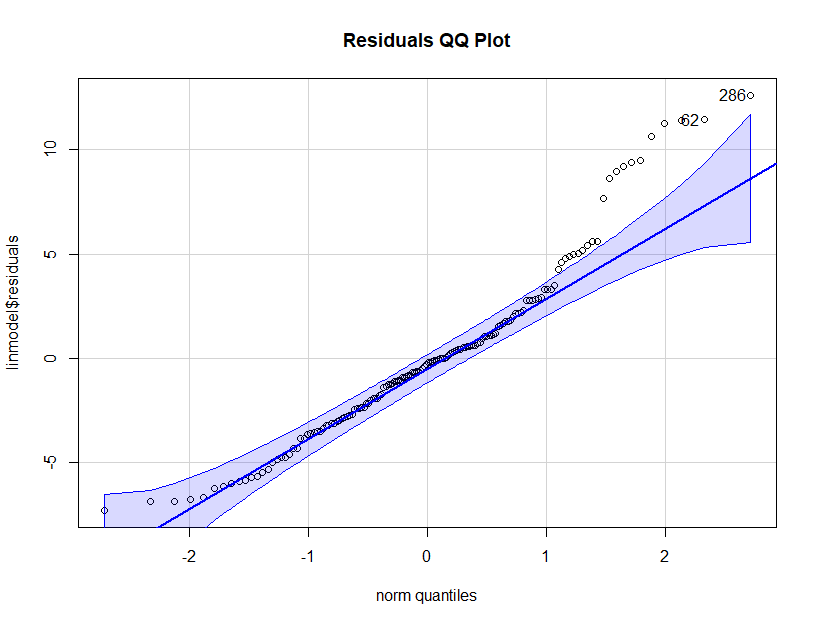


Figure 6: QQ plot of initial model’s residuals

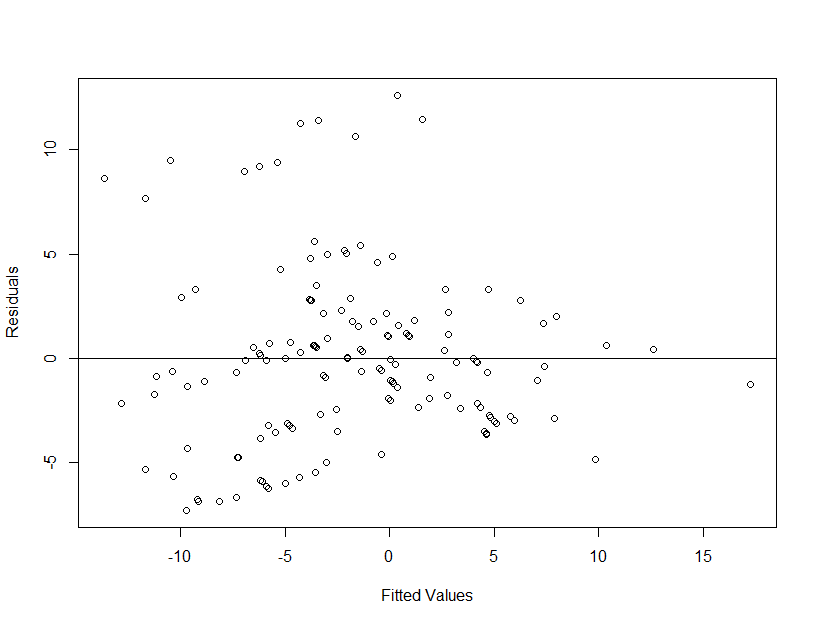


Figure 7: Residuals versus fitted values of initial model with horizontal line at y = 0

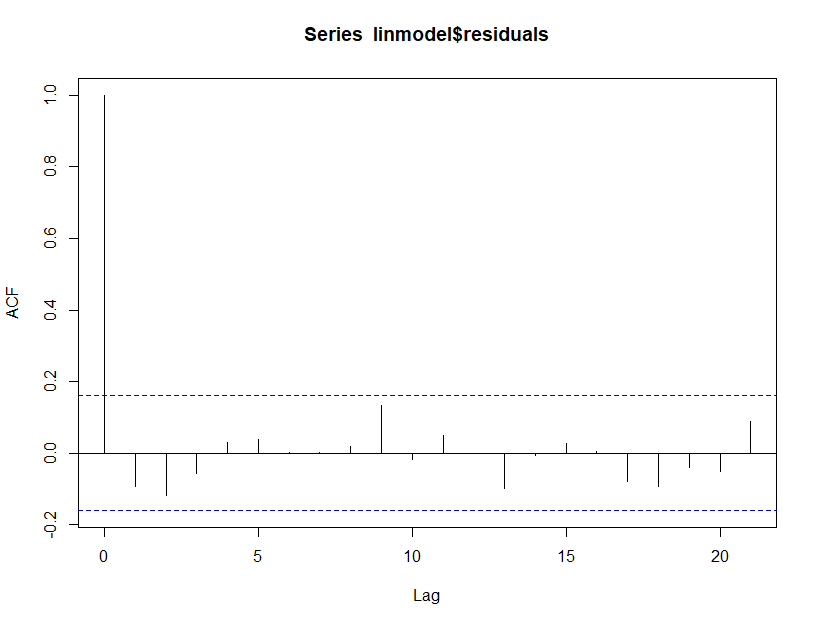


Figure 8: ACF plot of initial model

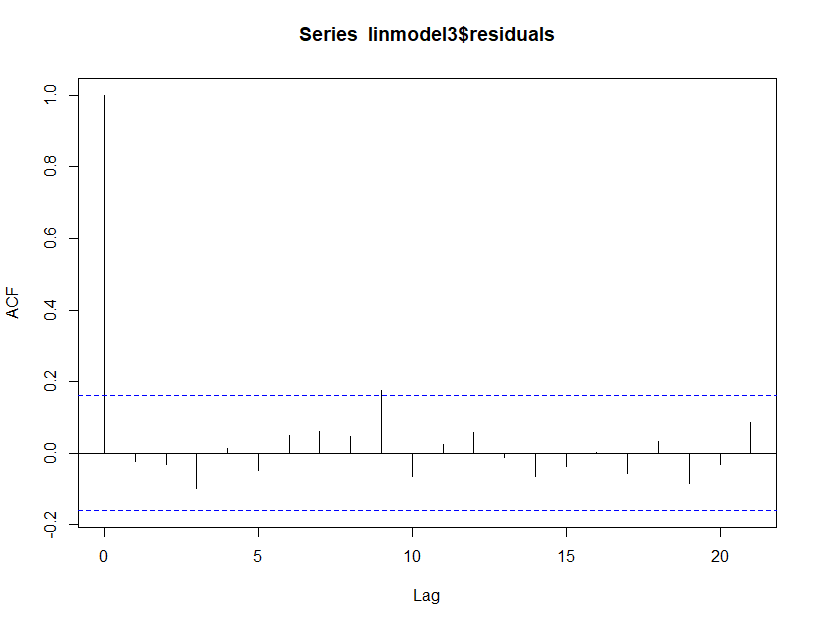


Figure 9: ACF plot of updated model

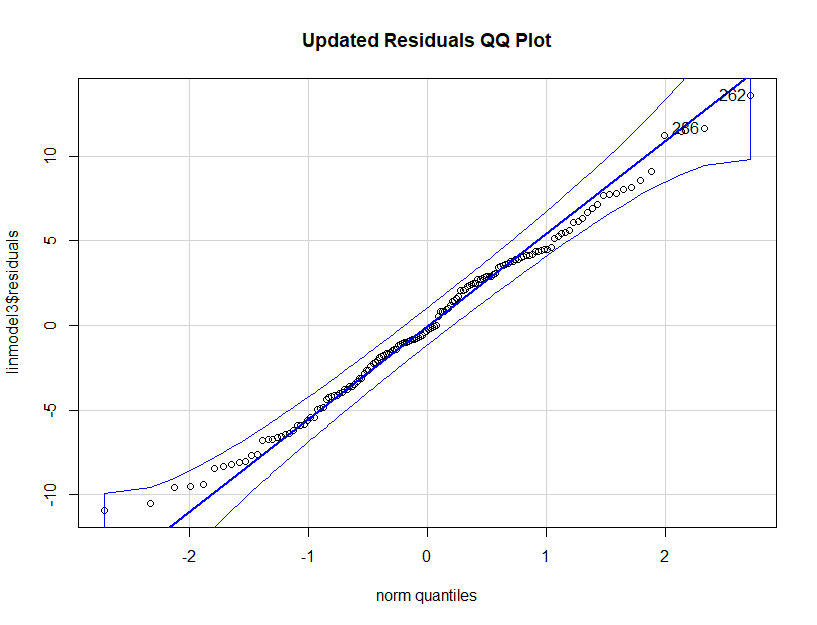


Figure 10: QQ plot of updated model’s residuals

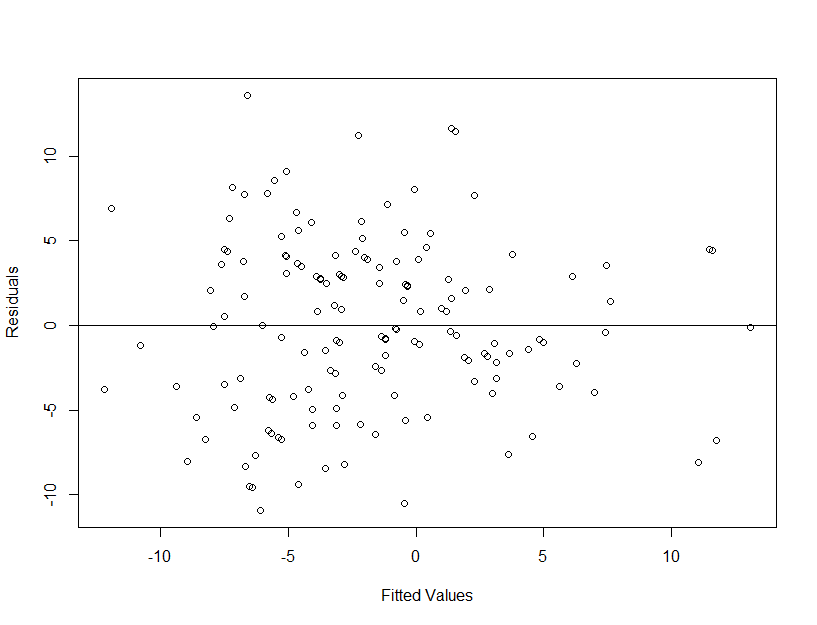


Figure 11: Residuals versus fitted values of updated model with horizontal line at y = 0

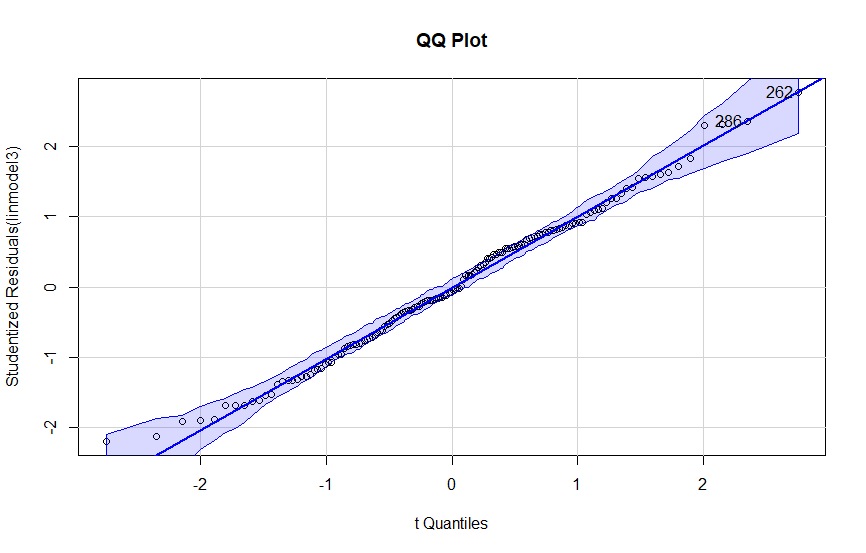


Figure 12: QQ plot of updated model

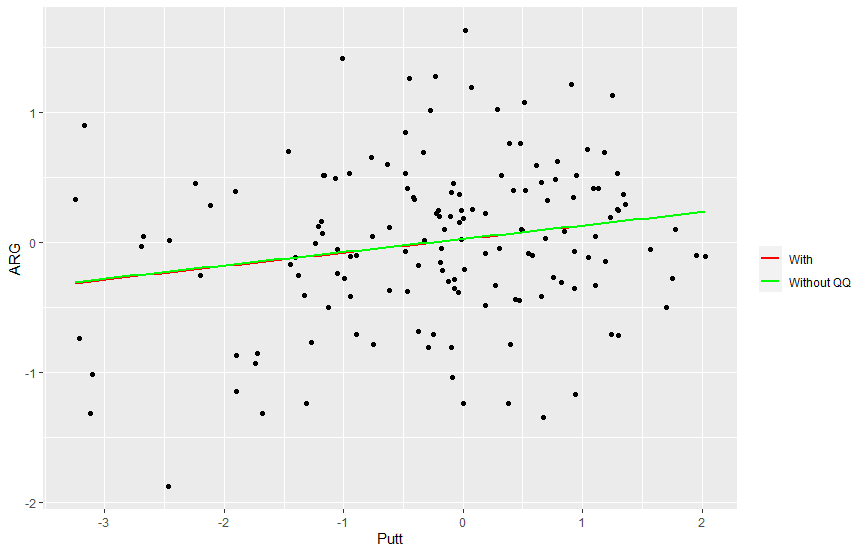


Figure 13: Plot of regression lines with and without QQ plot outliers

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Top25 | Score | Putt | ARG | APP | OTT | T2G |
| Top25 |  | 0.0000 | 0.0000 | 0.0908 | 0.0000 | 0.0011 | 0.0000 |
| Score | 0.0000 |  | 0.0000 | 0.0003 | 0.0000 | 0.0006 | 0.0000 |
| Putt | 0.0000 | 0.0000 |  | 0.0171 | 0.8049 | 0.1943 | 0.4840 |
| ARG | 0.0908 | 0.0003 | 0.0171 |  | 0.2266 | 0.0784 | 0.0000 |
| APP | 0.0000 | 0.0000 | 0.8049 | 0.2266 |  | 0.0176 | 0.0000 |
| OTT | 0.0011 | 0.0006 | 0.1943 | 0.0784 | 0.0176 |  | 0.0000 |
| T2G | 0.0000 | 0.0000 | 0.4840 | 0.0000 | 0.0000 | 0.0000 |  |

Table 1: Linearity matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard Error | t-value | p-value |
| Intercept | -2.9168 | 0.3449 | -8.458 | <0.001 |
| Putt | -3.1595 | 0.2913 | -10.845 | <0.001 |
| T2G | -3.0811 | 0.2566 | -12.005 | <0.001 |

Table 2: summary of initial linear model with Putt and T2G predictors

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Degrees of Freedom | Sums Squared | Mean Squared | F-value | p-value |
| Putt | 1 | 2278.1 | 2278.1 | 133.53 | <0.001 |
| T2G | 1 | 2458.8 | 2458.8 | 144.13 | <0.001 |
| Residuals | 147 | 2507.9 | 17.06 |  |  |

Table 3: ANOVA Table of initial linear model with Putt and T2G predictors

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard Error | t-value | p-value |
| Intercept | -3.0752 | 0.3283 | -9.366 | <0.001 |
| Putt | -3.1231 | 0.2761 | -11.311 | <0.001 |
| T2G | -3.1010 | 0.2440 | -12.709 | <0.001 |

Table 4: Summary of linear model with Putt and T2G predictors without outliers

|  |  |
| --- | --- |
|  | VIF Value |
| Putt | 1.003327 |
| T2G | 1.003327 |

Table 5: VIF Values for initial model

|  |  |
| --- | --- |
| Test Value | p-value |
| A = 2.0181 | <0.001 |

Table 6: Anderson-Darling normality test for initial model residuals

|  |  |  |
| --- | --- | --- |
| Test Value | Degrees of Freedom | p-value |
| BP = 10.141 | 2 | 0.006279 |

Table 7: Breusch-Pagan test for constant variance on initial model residuals

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard Error | t-value | p-value |
| Intercept | -2.1208 | 0.4236 | -5.007 | <0.001 |
| Putt | -3.3237 | 0.3658 | -9.086 | <0.001 |
| ARG | -2.6256 | 0.6883 | -3.815 | <0.001 |
| OTT | -4.2395 | 0.6957 | -6.094 | <0.001 |

Table 8: Summary of updated linear model with Putt, ARG, and OTT predictors

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Degrees of Freedom | Sums Squared | Mean Squared | F-value | p-value |
| Putt | 1 | 2278.1 | 2278.1 | 88.3164 | <0.001 |
| ARG | 1 | 242.8 | 242.8 | 9.4139 | 0.00257 |
| OTT | 1 | 957.8 | 957.8 | 37.1312 | <0.001 |
| Residuals | 146 | 3766.1 | 25.8 |  |  |

Table 9: ANOVA Table of updated linear model

|  |  |
| --- | --- |
| Test Value | p-value |
| DW = 2.1851 | 0.8711 |

Table 10: Durbin-Watson test for independence on initial model residuals

|  |  |
| --- | --- |
| Test Value | p-value |
| DW = 2.0236 | 0.5582 |

Table 11: Durbin-Watson test for independence on updated model residuals

|  |  |
| --- | --- |
| Test Value | p-value |
| A = 0.33482 | 0.5054 |

Table 12: Anderson-Darling normality test for updated model residuals

|  |  |  |
| --- | --- | --- |
| Test Value | Degrees of Freedom | p-value |
| BP = 5.8642 | 3 | 0.1184 |

Table 13: Breusch-Pagan test for constant variance on updated model residuals

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard Error | t-value | p-value |
| Intercept | -2.3014 | 0.4107 | -5.604 | <0.001 |
| Putt | -3.3772 | 0.3524 | -9.584 | <0.001 |
| ARG | -2.5756 | 0.6730 | -3.827 | <0.001 |
| OTT | -4.2568 | 0.6709 | -6.345 | <0.001 |

Table 14: Summary of updated linear model with Putt, ARG, and OTT predictors without QQ plot outliers

|  |  |
| --- | --- |
| Predictor | VIF Value |
| Putt | 1.046148 |
| ARG | 1.056217 |
| OTT | 1.027954 |

Table 15: VIF values for updated linear model with Putt, ARG, and OTT predictors

|  |  |
| --- | --- |
| Model | Adjusted R-Squared |
| Initial | 0.6491 |
| Updated | 0.4695 |

Table 16: Adjusted R-squared values for initial and updated linear models

|  |  |  |  |
| --- | --- | --- | --- |
| Interval Type | Fitted Value | Lower Bound | Upper Bound |
| Confidence | -19.07859 | -24.55769 | -13.59949 |
| Prediction | -19.07859 | -30.51428 | -7.642911 |

Table 17: 95% confidence and prediction interval for Score

|  |  |  |
| --- | --- | --- |
|  | Lower Bound | Upper Bound |
| Intercept | -2.957881 | -1.283660 |
| Putt | -4.046641 | -2.600794 |
| ARG | -3.985853 | -1.265318 |
| OTT | -5.614458 | -2.864454 |

Table 18: 95% confidence intervals for updated model slopes and intercept

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard Error | Z value | p-value |
| Intercept | -1.5777 | 0.2518 | -6.266 | <0.001 |
| Putt | 1.2552 | 0.2645 | 4.745 | <0.001 |
| ARG | 0.7326 | 0.3004 | 2.439 | 0.0147 |
| OTT | 1.5906 | 0.3530 | 4.506 | <0.001 |

Table 19: Summary of probit link logistic model

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Degrees of Freedom | Deviance | Residual Degrees of Freedom | Residual Deviance | p-value |
| NULL |  |  | 149 | 147.306 |  |
| Putt | 1 | 29.654 | 148 | 117.653 | <0.001 |
| ARG | 1 | 1.823 | 147 | 115.829 | 0.177 |
| OTT | 1 | 28.381 | 146 | 87.449 | <0.001 |

Table 20: ANOVA Table of probit link logistic model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard Error | Z value | p-value |
| Intercept | -2.7305 | 0.4814 | -5.672 | <0.001 |
| Putt | 2.1881 | 0.4864 | 4.498 | <0.001 |
| ARG | 1.2062 | 0.5339 | 2.259 | 0.0239 |
| OTT | 2.7490 | 0.6462 | 4.254 | <0.001 |

Table 21: Summary of logit link logistic model

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Degrees of Freedom | Deviance | Residual Degrees of Freedom | Residual Deviance | p-value |
| NULL |  |  | 149 | 147.306 |  |
| Putt | 1 | 29.1016 | 148 | 118.205 | <0.001 |
| ARG | 1 | 1.7118 | 147 | 116.493 | 0.1907 |
| OTT | 1 | 28.2184 | 146 | 88.274 | <0.001 |

Table 22: ANOVA Table of logit link logistic model

|  |  |
| --- | --- |
| Model | AIC |
| Probit | 95.449 |
| Logit | 96.274 |

Table 23: AIC values for probit and logit link logistic regression models

|  |  |  |
| --- | --- | --- |
|  | Lower Bound | Upper Bound |
| Intercept | -2.1384515 | -1.126672 |
| Putt | 0.7792129 | 1.834604 |
| ARG | 0.1550752 | 1.380854 |
| OTT | 0.9374175 | 2.345723 |

Table 24: 95% confidence intervals for logistic regression model slopes and intercept

|  |  |
| --- | --- |
|  | Fitted Probability |
| Putt = 0, ARG = 0, OTT = 4 | 0.9999991 |

Table 25: Fitted probability with given values of Putt, ARG, and OTT

**R Code Used:**

#loading packages

library(leaps)

library(car)

library(ggplot2)

library(MASS)

library(lmtest)

library(nortest)

library(car)

library(Hmisc)

# read in data and getting sample

projectdata <- read.csv("ProjectData.csv", sep=',', header = T)

set.seed(38273)

mydata <- projectdata[sample(1:nrow(projectdata), 150),]

#Selecting inital model

bestsub1=regsubsets(Score~Putt+ARG+APP+OTT+T2G,

data=mydata, nbest=10)

plot(bestsub1, scale="adjr2")

plot(bestsub1, scale="bic")

plot(bestsub1, scale="Cp")

#Automatic selection - Forward

null=lm(Score~1, data=mydata) #intercept only model

full=lm(Score~Putt+ARG+APP+OTT+T2G, data=mydata)#full model- Regresses y on all variables in dataset

step(null, scope=list(lower=null, upper=full),

direction="forward")

#Automatic selection - backward

step(full, data=mydata, direction="backward")

#Automatic selection - Stepwise

step(null, scope = list(upper=full), data=mydata, direction="both")

linmodel = lm(Score~Putt+T2G, data = mydata)

#looking for outliers, multicollinearity

qqPlot(linmodel, main="QQ Plot")

#create new data set with the QQplot outliers removed

mydata2 = mydata[-c(71,140), ]

#Run regression on new data set and compare the two models

linmodel2 = lm(Score~Putt+T2G, data = mydata2)

summary(linmodel)

anova(linmodel)

summary(linmodel2)

anova(linmodel2)

#plot regression lines from the full and reduced data sets

ggplot(mydata, aes(Putt, T2G, Score)) +

geom\_point() +

geom\_smooth(method="lm", se=F, aes(color="With")) +

geom\_smooth(data = mydata[-c(71,140), ], method="lm", se=F, aes(color="Without QQ"))+

scale\_colour\_manual(name='',values=c("red","green"))

vif(lm(Score ~Putt+T2G, data=mydata))

#Checking assumptions for model 1

summary(linmodel)

anova(linmodel)

#Normality check with QQ plot and Anderson-Darling Test

qqPlot(linmodel$residuals, main="Residuals QQ Plot")

ad.test(linmodel$residuals)

#Constant variance check with Res vs Fits plot and Breusch Pagan test

plot(linmodel$fitted,linmodel$residuals, ylab="Residuals", xlab="Fitted Values")

abline(0,0)

bptest(linmodel)#Breusch-Pagan for constant variance

#Linearity check

rcorr(as.matrix(mydata))

#Independence check

acf(linmodel$residuals)

dwtest(linmodel)

#Checking assumptions for updated model

linmodel3 = lm(Score ~ Putt+ARG+OTT, data = mydata)

summary(linmodel3)

anova(linmodel3)

#Normality check with QQ plot and Anderson-Darling Test

qqPlot(linmodel3$residuals, main="Updated Residuals QQ Plot")

ad.test(linmodel3$residuals)

#Constant variance check with Res vs Fits plot and Breusch Pagan test

plot(linmodel3$fitted,linmodel3$residuals, ylab="Residuals", xlab="Fitted Values")

abline(0,0)

bptest(linmodel3)#Breusch-Pagan for constant variance

#Linearity check

rcorr(as.matrix(mydata))

#Independence check

acf(linmodel3$residuals)

dwtest(linmodel3)

#Confidence intervals for updated slopes

confint(linmodel3)

#Estimation and Prediction for updated linear model

predict(linmodel3, interval="confidence", se.fit=T, newdata=data.frame(Putt = 0, ARG = 0, OTT = 4))

predict(linmodel3, interval="prediction", se.fit=T, newdata=data.frame(Putt = 0, ARG = 0, OTT = 4))

#Logistic Model Selection and prediction

probit <- glm(Top25 ~ Putt + ARG + OTT, family = binomial(link = "probit"), data = mydata)

summary(probit)

anova(probit, test="Chisq")

logit <- glm(Top25 ~ Putt + ARG + OTT, family = binomial, data = mydata)

summary(logit)

anova(logit, test="Chisq")

confint(probit)

predict(probit, newdata=data.frame(Putt=0, ARG = 0, OTT = 4), type="response",

se.fit=TRUE)