Maxime Maltais
Regression Analysis – Spring 2009
Student Project
August 25, 2011

Let's help Tiger

Introduction

For the past 12 years or so, Tiger Woods was to be the number one golfer on the PGA tour. Right now, he seems to really struggle with his golf game. By using a regression model on 2010 PGA tour players statistics, I will to try to find out what are the most important variables that explain the player's ranking (Fedex Cup) on tour. I will then give Tiger some recommendations on what he needs to improve to be the number one golfer again.

All the calculations and analysis were performed using R. Programming code is incorporated in the report. The Excel file joined to this report includes all the data used for the analysis.

Data

Source

All the data used for my regression analysis has been extract from the PGA tour website (www.pgatour.com) for years 2009 and 2010. There are thousand of statistics available on this website but here is a summary of the data I choose to import based on my intuition of what could be significant variables in trying to better understand a golfer's ranking. Year 2009 data will only be use to test the consistency of the model at the end of the process.

Dependant variable

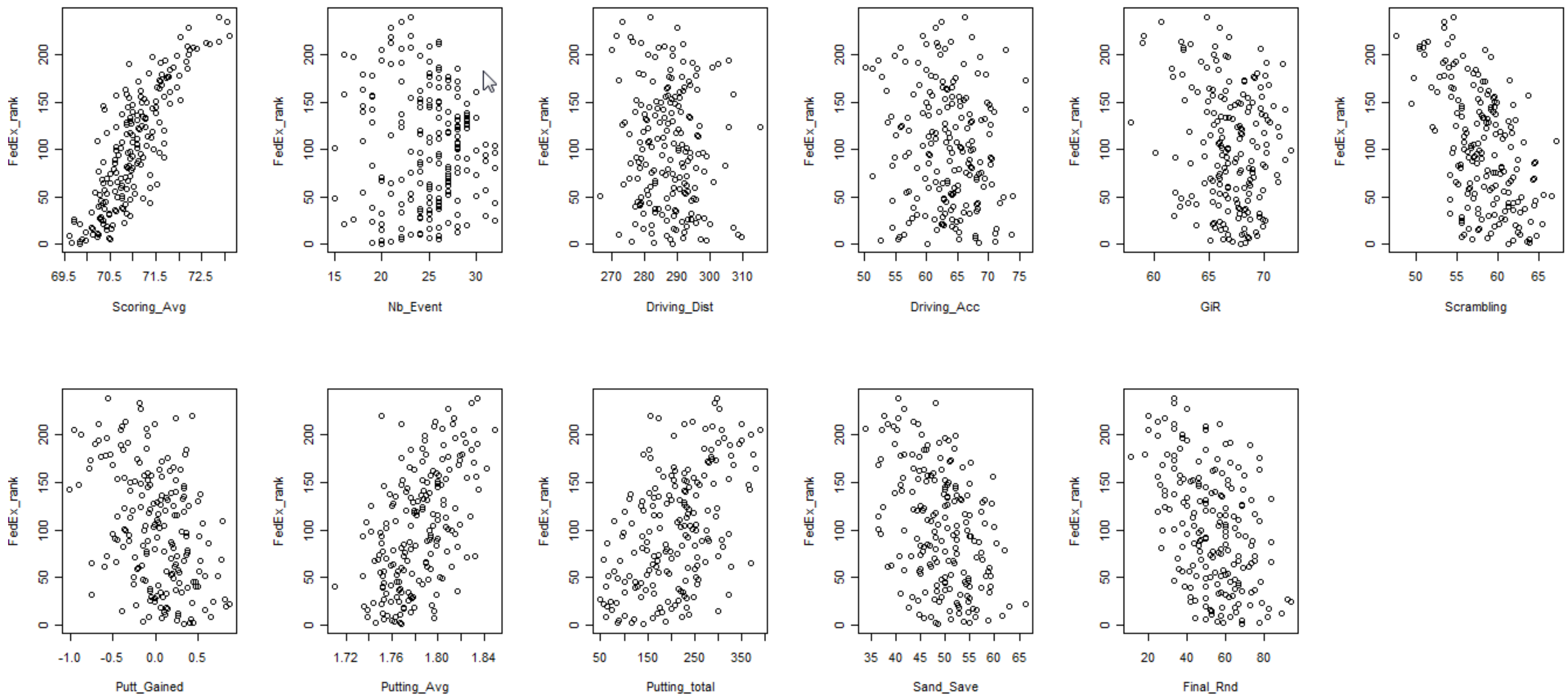
The dependant variable (Y) that I will try to explain is the 2010 Fedex Cup ranking. During the regular season, players accumulate point based on their result on each event. Larger events (Majors) give more points.

Independent Variables Description

Variable	Description
Scoring average	The weighted scoring average which takes the stroke average of the field into account. It is computed by adding a player's total strokes to an adjustment and dividing by the total rounds played. The adjustment is computed by determining the stroke average of the field for each round played. This average is subtracted from par to create an adjustment for each round. A player accumulates these adjustments for each round played.
Nb of events	Number of official events played
Driving distance	The average number of yards per measured drive. These drives are measured on two holes per round. Care is taken to select two holes which face in opposite directions to counteract the effect of wind. Drives are measured to the point at which they come to rest regardless of whether they are in the fairway or not.
Driving accuracy	The percentage of time a tee shot comes to rest in the fairway (regardless of club).
Greens in regulation (GiR)	The percent of time a player was able to hit the green in regulation (greens hit in regulation/holes played). Note: A green is considered hit in regulation if any portion of the ball is touching the putting surface after the GIR stroke has been taken.
Scrambling	The percent of time a player misses the green in regulation, but still makes par or better.
Putt Gained	The number of putts a player takes from a specific distance is measured against a statistical baseline to determine the player's strokes gained or lost on a hole. The sum of the values for all holes played in a round minus the field average strokes gained/lost for the round is the player's Strokes gained/lost for that round. The sum of strokes gained for each round are divided by total rounds played.
Putting average	The average number of putts per green in regulation. By using greens hit in regulation, it eliminate the effects of chipping close and one-putting in the computation.
Putting total	Total Putting is computed using 6 putting stats Putting from 3-5', Putting from 5-10', Putting from 10-15', Putting from 15-20', Putting from 20-25' and Three Putt Avoidance from > 25'. Each statistic is given a numerical weighting based on the frequency of putts attempted from each distance. The players rank in each of the statistics used is multiplied by the corresponding weigh factor, totalled, and divided by the number of statistics used to produce the Total Putting Value.
Sand save	The percent of time a player was able to get 'up and down' once in a greenside sand bunker (regardless of score). Note: 'Up and down' indicates it took the player 2 shots or less to put the ball in the hole from that point.
Final round performance	The percent of time a player's finish position improves or remains unchanged in the final round.

Data summary

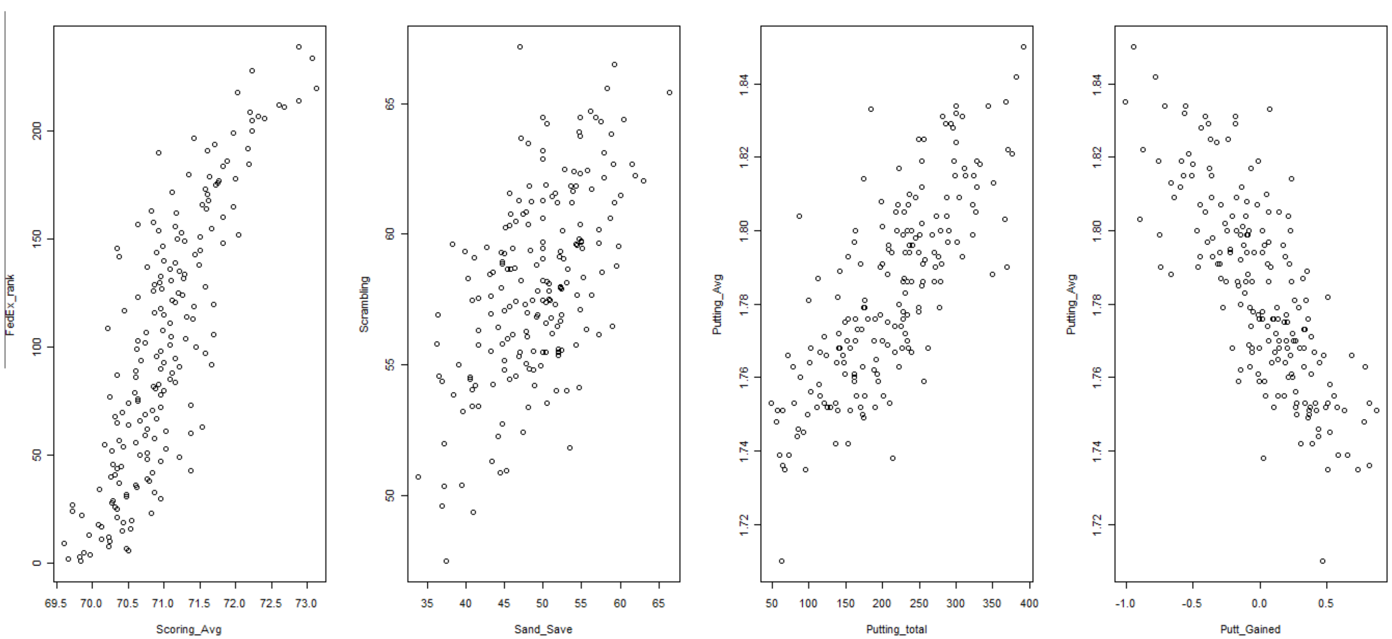
```
> summary(Rdata2010)
   FedEx_rank      Scoring_Avg       Nb_Event      Driving_Dist    Driving_Acc         GiR          Scrambling      Putt_Gained          Putting_Avg     Putting_total
 Min.   :  1.00   Min.   :69.61   Min.   :15.00   Min.   :266.4   Min.   :50.15   Min.   :57.89   Min.   :47.49   Min.   :-1.008000   Min.   :1.710   Min.   : 48.7
 1st Qu.: 49.75   1st Qu.:70.51   1st Qu.:22.00   1st Qu.:281.3   1st Qu.:60.27   1st Qu.:65.50   1st Qu.:55.55   1st Qu.:-0.222000   1st Qu.:1.764   1st Qu.:153.2
 Median : 97.50   Median :70.95   Median :25.00   Median :287.9   Median :63.70   Median :66.83   Median :58.04   Median : 0.029000   Median :1.781   Median :215.4
 Mean   :101.22   Mean   :71.01   Mean   :24.74   Mean   :287.5   Mean   :63.37   Mean   :66.91   Mean   :58.10   Mean   : 0.008474   Mean   :1.783   Mean   :208.8
 3rd Qu.:148.25   3rd Qu.:71.41   3rd Qu.:27.25   3rd Qu.:292.5   3rd Qu.:66.91   3rd Qu.:68.66   3rd Qu.:60.60   3rd Qu.: 0.267000   3rd Qu.:1.800   3rd Qu.:257.0
 Max.   :239.00   Max.   :73.12   Max.   :32.00   Max.   :315.5   Max.   :76.08   Max.   :72.49   Max.   :67.16   Max.   : 0.871000   Max.   :1.850   Max.   :391.3
   Sand_Save       Final_Rnd
 Min.   :33.78   Min.   :11.11
 1st Qu.:45.05   1st Qu.:41.67
 Median :50.00   Median :53.33
 Mean   :49.23   Mean   :53.15
 3rd Qu.:53.64   3rd Qu.:64.39
 Max.   :66.39   Max.   :94.12
```



Multicollinearity

If two or more predictor variables in a multiple regression model are highly correlated, the coefficient estimates may change erratically in response to small changes in the model or the data. Let see the correlation matrix of our variables.

	FedEx_rank	Scoring_Avg	Nb_Event	Driving_Dist	Driving_Acc	GiR	Scrambling	Putt_Gained	Putting_Avg	Putting_total	Sand_Save	Final_Rnd
FedEx_rank	1.00000000	0.8308450	-0.065898592	-0.10078798	-0.10400437	-0.17757213	-0.527327769	-0.41102897	0.53817007	0.49549196	-0.40280039	-0.44426144
Scoring_Avg	0.83084498	1.0000000	0.078263401	-0.15084838	-0.16858324	-0.38233200	-0.692479564	-0.46341308	0.51223966	0.47842103	-0.46653422	-0.47150163
Nb_Event	-0.06589859	0.0782634	1.000000000	-0.12158768	0.04707410	0.08832395	-0.006505892	0.03610452	-0.09168549	-0.02103184	0.05748463	-0.01051546
Driving_Dist	-0.10078798	-0.1508484	-0.121587681	1.00000000	-0.58487339	0.26281460	-0.232897205	-0.18096800	0.12417383	0.19118272	-0.19092288	0.09869765
Driving_Acc	-0.10400437	-0.1685832	0.047074101	-0.58487339	1.00000000	0.30407068	0.274668348	-0.06803420	0.06766742	0.03053523	0.04789468	0.02015971
GiR	-0.17757213	-0.3823320	0.088323950	0.26281460	0.30407068	1.00000000	0.087047334	-0.27684016	0.28073183	0.27472018	-0.09297680	0.12225906
Scrambling	-0.52732777	-0.6924796	-0.006505892	-0.23289720	0.27466835	0.08704733	1.000000000	0.51730034	-0.34182831	-0.51202395	0.58266164	0.32847076
Putt_Gained	-0.41102897	-0.4634131	0.036104517	-0.18096800	-0.06803420	-0.27684016	0.517300344	1.00000000	-0.76800501	-0.93147756	0.45493641	0.27616582
Putting_Avg	0.53817007	0.5122397	-0.091685488	0.12417383	0.06766742	0.28073183	-0.341828312	-0.76800501	1.00000000	0.77041108	-0.42673490	-0.25181920
Putting_total	0.49549196	0.4784210	-0.021031836	0.19118272	0.03053523	0.27472018	-0.512023949	-0.93147756	0.77041108	1.00000000	-0.46494558	-0.28228472
Sand_Save	-0.40280039	-0.4665342	0.057484625	-0.19092288	0.04789468	-0.09297680	0.582661643	0.45493641	-0.42673490	-0.46494558	1.00000000	0.26485693
Final_Rnd	-0.44426144	-0.4715016	-0.010515465	0.09869765	0.02015971	0.12225906	0.328470760	0.27616582	-0.25181920	-0.28228472	0.26485693	1.00000000



Based on the correlation matrix and 4 graphs above:

- I will not use sand save since it is highly correlated with scrambling and less with the Fedex ranking. Sand save is a part of the scrambling statistic. Its signal will be capture with Scrambling.
- I will use only Putting_total since all putting statistic are very correlated and this variable include the more information on putting by its definition.
- Even if Driving_Dist and Driving_Acc are correlated, I will kept them both because my intuition makes me believe that they can provide a different signal.

Also, we cannot really say to Tiger "Improve your scoring average", it is not very specific, he will ask us how and we will have no answer for him. Therefore we will not include this variable to explain the FedEx ranking even if they are highly correlated.

Regression model

I will use a Multiple Linear Regression model of the form

$$y = X\beta + \varepsilon,$$

Where,

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

With the assumption that:

1. Residuals are normally distributed, $\varepsilon \sim N(\mu=0,\sigma^2)$
2. The variance of the error is constant across observations (homoscedasticity)
3. Residuals are independent and not correlated $COV(\varepsilon_i, \varepsilon_j) \neq 0 \ \forall$ i,j
4. The predi**c**tors are linearly independent (multicollinearity).

Let's fit a first linear model that includes all the dependant variables to see if it seems to be globally adequate in our case.

```
> LM0=lm(FedEx_rank~Driving_Acc + Driving_Dist + Final_Rnd + GiR + Nb_Event + Putting_total  + Scrambling)
> summary(LM0)

Call:
lm(formula = FedEx_rank ~ Driving_Acc + Driving_Dist + Final_Rnd +
    GiR + Nb_Event + Putting_total + Scrambling)

Residuals:
     Min       1Q   Median       3Q      Max
-123.920  -35.115    4.875   32.309   98.354

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   1351.75866  192.07849   7.038 3.75e-11 ***
Driving_Acc     -2.31205    0.97021  -2.383 0.018189 *
Driving_Dist    -2.39140    0.60655  -3.943 0.000114 ***
Final_Rnd       -0.75985    0.21950  -3.462 0.000667 ***
GiR             -1.96215    1.64130  -1.195 0.233434
Nb_Event        -1.36782    0.86947  -1.573 0.117398
Putting_total    0.29934    0.05248   5.704 4.60e-08 ***
Scrambling      -4.70740    1.13223  -4.158 4.92e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 44.13 on 184 degrees of freedom
Multiple R-squared: 0.4994,     Adjusted R-squared: 0.4804
F-statistic: 26.23 on 7 and 184 DF,  p-value: < 2.2e-16
```

We see that R^2 is low, this shows us that the goodness of fit is not good, maybe a linear model is not adequate on the FedEx ranking or that our independent variables are not really good indicators of the ranking. Since we saw that the FedEx ranking is highly correlated with the Scoring average, let's see if a linear model on this variable would be better.

```
> LM0=lm(Scoring_Avg~Driving_Acc + Driving_Dist + Final_Rnd + GiR + Nb_Event + Putting_total  + Scrambling)
> summary(LM0)

Call:
lm(formula = Scoring_Avg ~ Driving_Acc + Driving_Dist + Final_Rnd +
    GiR + Nb_Event + Putting_total + Scrambling)

Residuals:
     Min       1Q   Median       3Q      Max
-1.05196 -0.21316  0.02799  0.23154  0.93914

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  89.0382728  1.4997342  59.369  < 2e-16 ***
Driving_Acc  -0.0174621  0.0075753  -2.305 0.022275 *
Driving_Dist -0.0247576  0.0047359  -5.228 4.63e-07 ***
Final_Rnd    -0.0063257  0.0017139  -3.691 0.000294 ***
GiR          -0.0786524  0.0128151  -6.137 5.03e-09 ***
Nb_Event      0.0139235  0.0067888   2.051 0.041687 *
Putting_total 0.0029228  0.0004097   7.133 2.18e-11 ***
Scrambling   -0.0889017  0.0088403 -10.056  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3445 on 184 degrees of freedom
Multiple R-squared: 0.752,      Adjusted R-squared: 0.7426
F-statistic: 79.71 on 7 and 184 DF,  p-value: < 2.2e-16
```

The adjusted R^2 is much better and the regression is globally significant, we can now conclude that a multiple linear regression on the Scoring average would be a better response variable then the FedEx Cup ranking.

Before continuing with the Scoring average as our response variable, let's make sure that we will be able to help Tiger with our better understanding of this statistic.

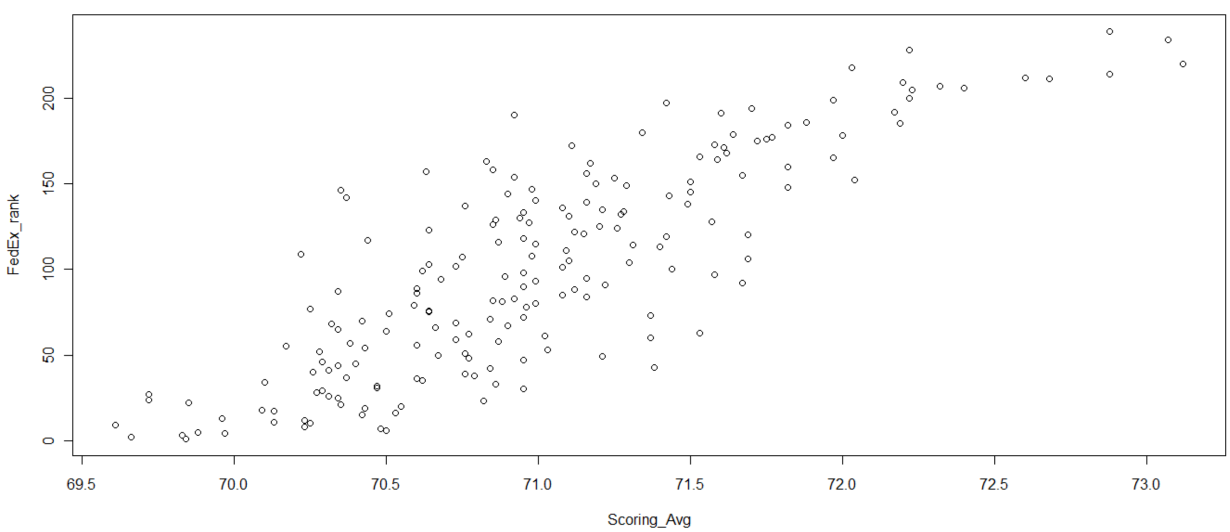
```
> LM=lm(FedEx_rank ~ Scoring_Avg)
> summary(LM)

Call:
lm(formula = FedEx_rank ~ Scoring_Avg)

Residuals:
    Min      1Q  Median      3Q     Max 
-86.302 -21.844   0.514  21.024  95.151 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -5216.915    258.435  -20.19   <2e-16 ***
Scoring_Avg    74.898      3.639   20.58   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 34.16 on 190 degrees of freedom
Multiple R-squared: 0.6903,     Adjusted R-squared: 0.6887 
F-statistic: 423.5 on 1 and 190 DF,  p-value: < 2.2e-16
```



The R^2 is high and the graph clearly indicates a linear relation.

Let's now try to find a model.

Model Selection

Approaches to find the best model:

1. Forward selection, starting with no variables in the model, trying out the variables one by one and including them if they are 'statistically significant' (using p-value, α = 5%).
   - using function *add1()* in R
2. Backward selection, starting with all candidate variables and testing them one by one for statistical significance, deleting any that are not significant (using p-value, α = 5%).
   - using function *drop1()* in R
3. Combination of the above, testing at each stage for variables to be included or excluded (using AIC as criteria). [*Chosen Approach*]
   - using function step*()* in R

Resulting model using Stepwise approach is presented below. Detailed results (steps) from R are presented in appendix.

```
> summary(LM)

Call:
lm(formula = Scoring_Avg ~ Scrambling + GiR + Putting_total +
    Driving_Dist + Final_Rnd + Driving_Acc + Nb_Event)

Residuals:
     Min       1Q   Median       3Q      Max
-1.05196 -0.21316  0.02799  0.23154  0.93914

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   89.0382728  1.4997342  59.369  < 2e-16 ***
Scrambling    -0.0889017  0.0088403 -10.056  < 2e-16 ***
GiR           -0.0786524  0.0128151  -6.137 5.03e-09 ***
Putting_total  0.0029228  0.0004097   7.133 2.18e-11 ***
Driving_Dist  -0.0247576  0.0047359  -5.228 4.63e-07 ***
Final_Rnd     -0.0063257  0.0017139  -3.691 0.000294 ***
Driving_Acc   -0.0174621  0.0075753  -2.305 0.022275 *
Nb_Event       0.0139235  0.0067888   2.051 0.041687 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3445 on 184 degrees of freedom
Multiple R-squared: 0.752,      Adjusted R-squared: 0.7426
F-statistic: 79.71 on 7 and 184 DF,  p-value: < 2.2e-16

>
```

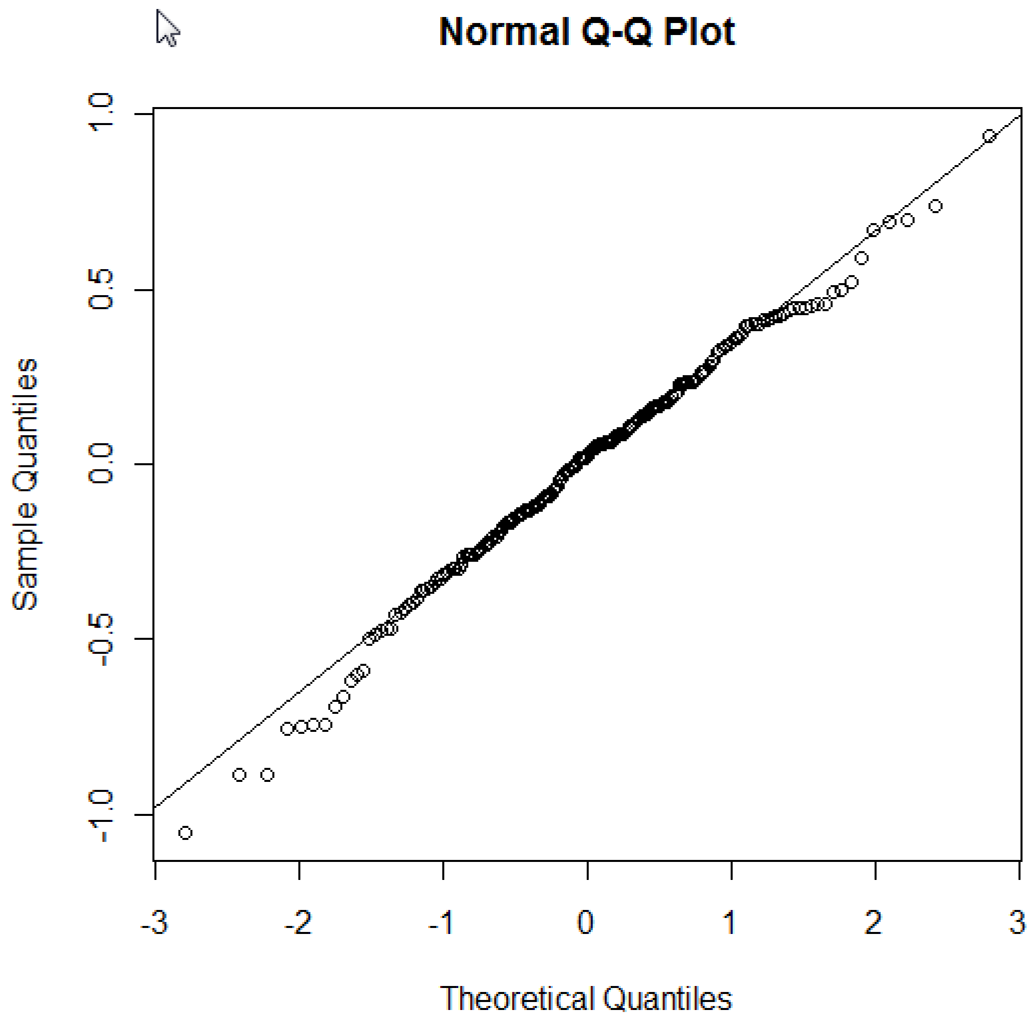
Let's validate that model.

Final Assessment of the model

To assess if the model is adequate, I will verify that my initial assumptions hold:

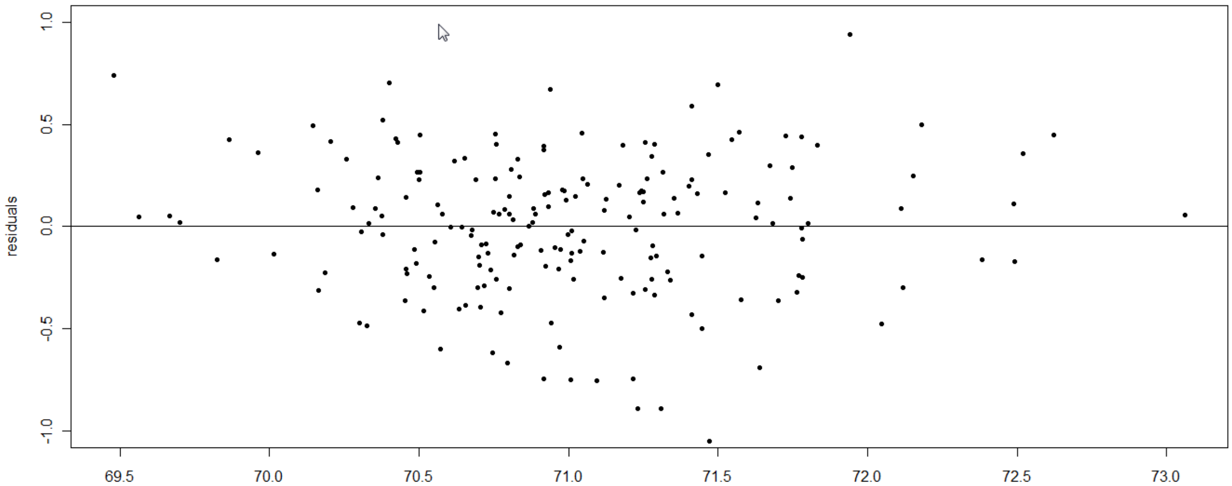
A. I will look at the QQ-Plot to see the normality of the residuals [assumption # 1: $\varepsilon \sim N(\mu=0,\sigma^2)$]
B. I will look at the residual graph to assess that the standard deviations of the error terms are constant and do not depend on the *x*-value and that they are uncorrelated [assumption # 2: $VAR(\varepsilon_i) = \sigma^2 \ \forall \ i$] [assumption # 3: $COV(\varepsilon_i, \varepsilon_j) \neq 0 \ \forall \ i,j$]
C. I will look at the Shapiro-Wilk test to see if residuals could be normally distributed.
D. I will look at the variance inflation factor (VIF) for multicollinearity (assumption #4)

<u>QQ Plot</u>



A QQ plot of the residuals shows slight normality, although there appears to be a slight deviation from normality in the tails.

Residuals vs Fitted



A plot of the residuals versus the fitted values show that the residuals appear uncorrelated and seem to have constant variance.

Shapiro-Wilk

```
> shapiro.test(residuals(LM))

        Shapiro-Wilk normality test

data:  residuals(LM)
W = 0.9904, p-value = 0.2295
```

Ho: The residuals are normally distributed
Ha: The residuals are not normally distributed

A Shapiro-Wilk normality test produces a p-value of .2295, which leads me to fail to reject the null hypothesis that the residuals are normally distributed.

Variance inflation Factor

```
> vif(LM)
   Scrambling          GiR Putting_total  Driving_Dist     Final_Rnd  Driving_Acc      Nb_Event
     1.668759     1.816983      1.670568      2.436075      1.211589     2.410971      1.054018
```

Usually, VIF of 5 or 10 and above indicates a multicollinearity problem, our model is ok.

Consistency

Finally, if we would have used 2009 data, the following result would be obtain.

```
> summary(LM)

Call:
lm(formula = Scoring_Avg ~ Scrambling + GiR + Putting_total +
    Driving_Dist + Final_Rnd + Driving_Acc + Nb_Event)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.84329 -0.29560 -0.02394  0.30295  0.90368 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)   90.4174647  1.7148896  52.725  < 2e-16 ***
Scrambling    -0.0664365  0.0104499  -6.358 1.72e-09 ***
GiR           -0.0919089  0.0178589  -5.146 7.08e-07 ***
Putting_total  0.0036252  0.0004921   7.367 6.63e-12 ***
Driving_Dist  -0.0301387  0.0058763  -5.129 7.68e-07 ***
Final_Rnd     -0.0097403  0.0019435  -5.012 1.31e-06 ***
Driving_Acc   -0.0274372  0.0101061  -2.715  0.00729 ** 
Nb_Event       0.0201346  0.0078217   2.574  0.01088 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 0.3998 on 175 degrees of freedom
Multiple R-squared: 0.6955,     Adjusted R-squared: 0.6833 
F-statistic:  57.1 on 7 and 175 DF,  p-value: < 2.2e-16 
```

We can see that the consistency of our model is fine.

Final recommendations to Tiger

So finally, for Tiger to be numero uno again, we can conclude that he should concentrate on his short game, scrambling and putting. Isn't it a surprising conclusion?

You're welcome Tiger, call me any time.

Appendix

```
>   LM=step(lm(Scoring_Avg ~ 1),Scoring_Avg ~ Nb_Event + Driving_Dist + Driving_Acc
+GiR+Scrambling+ Putting_total+ Final_Rnd  ,direction="both")
Start:  AIC=-147.61
Scoring_Avg ~ 1

                Df Sum of Sq    RSS     AIC
+ Scrambling     1    42.238 45.845 -270.99
+ Putting_total  1    20.161 67.922 -195.51
+ Final_Rnd      1    19.582 68.501 -193.88
+ GiR            1    12.876 75.207 -175.95
+ Driving_Acc    1     2.503 85.580 -151.15
+ Driving_Dist   1     2.004 86.079 -150.03
<none>                       88.083 -147.61
+ Nb_Event       1     0.540 87.543 -146.79

Step:  AIC=-270.99
Scoring_Avg ~ Scrambling

                Df Sum of Sq    RSS     AIC
+ GiR            1     9.206 36.639 -312.02
+ Driving_Dist   1     9.073 36.771 -311.33
+ Final_Rnd      1     5.880 39.964 -295.35
+ Putting_total  1     1.831 44.013 -276.82
+ Nb_Event       1     0.479 45.366 -271.01
<none>                       45.845 -270.99
+ Driving_Acc    1     0.045 45.800 -269.18
- Scrambling     1    42.238 88.083 -147.61

Step:  AIC=-312.02
Scoring_Avg ~ Scrambling + GiR

                Df Sum of Sq    RSS     AIC
+ Putting_total  1     7.176 29.463 -351.88
+ Driving_Dist   1     4.940 31.699 -337.83
+ Final_Rnd      1     4.552 32.087 -335.49
+ Driving_Acc    1     1.319 35.320 -317.07
+ Nb_Event       1     0.935 35.704 -314.99
<none>                       36.639 -312.02
- GiR            1     9.206 45.845 -270.99
- Scrambling     1    38.568 75.207 -175.95

Step:  AIC=-351.88
Scoring_Avg ~ Scrambling + GiR + Putting_total

                Df Sum of Sq    RSS     AIC
+ Driving_Dist   1    4.6405 24.823 -382.78
+ Final_Rnd      1    2.7175 26.746 -368.46
+ Nb_Event       1    1.3182 28.145 -358.67
+ Driving_Acc    1    0.7356 28.727 -354.73
<none>                      29.463 -351.88
- Putting_total  1    7.1761 36.639 -312.02
- Scrambling     1   13.3112 42.774 -282.30
- GiR            1   14.5504 44.013 -276.82

Step:  AIC=-382.78
Scoring_Avg ~ Scrambling + GiR + Putting_total + Driving_Dist

                Df Sum of Sq    RSS     AIC
+ Final_Rnd      1    1.7029 23.120 -394.43
+ Driving_Acc    1    0.8628 23.960 -387.57
+ Nb_Event       1    0.6545 24.168 -385.91
<none>                      24.823 -382.78
- Driving_Dist   1    4.6405 29.463 -351.88
                                         Page 1
```

```
- Putting_total  1     6.8764 31.699 -337.83
- GiR            1     9.3287 34.151 -323.53
- Scrambling     1    16.4376 41.260 -287.22

Step:  AIC=-394.43
Scoring_Avg ~ Scrambling + GiR + Putting_total + Driving_Dist +
    Final_Rnd

                Df Sum of Sq    RSS     AIC
+ Driving_Acc    1    0.7771 22.343 -398.99
+ Nb_Event       1    0.6456 22.474 -397.87
<none>                       23.120 -394.43
- Final_Rnd      1    1.7029 24.823 -382.78
- Driving_Dist   1    3.6259 26.746 -368.46
- Putting_total  1    5.4043 28.524 -356.10
- GiR            1    8.3209 31.441 -337.40
- Scrambling     1   13.6165 36.736 -307.52

Step:  AIC=-398.99
Scoring_Avg ~ Scrambling + GiR + Putting_total + Driving_Dist +
    Final_Rnd + Driving_Acc

                Df Sum of Sq    RSS     AIC
+ Nb_Event       1    0.4994 21.843 -401.33
<none>                       22.343 -398.99
- Driving_Acc    1    0.7771 23.120 -394.43
- Final_Rnd      1    1.6172 23.960 -387.57
- Driving_Dist   1    3.9018 26.244 -370.09
- GiR            1    4.0731 26.416 -368.84
- Putting_total  1    5.8669 28.210 -356.22
- Scrambling     1   12.3918 34.734 -316.27

Step:  AIC=-401.33
Scoring_Avg ~ Scrambling + GiR + Putting_total + Driving_Dist +
    Final_Rnd + Driving_Acc + Nb_Event

                Df Sum of Sq    RSS     AIC
<none>                       21.843 -401.33
- Nb_Event       1    0.4994 22.343 -398.99
- Driving_Acc    1    0.6308 22.474 -397.87
- Final_Rnd      1    1.6172 23.460 -389.62
- Driving_Dist   1    3.2443 25.088 -376.74
- GiR            1    4.4717 26.315 -367.57
- Putting_total  1    6.0404 27.884 -356.46
- Scrambling     1   12.0055 33.849 -319.23
> summary(LM)

Call:
lm(formula = Scoring_Avg ~ Scrambling + GiR + Putting_total +
    Driving_Dist + Final_Rnd + Driving_Acc + Nb_Event)

Residuals:
     Min       1Q   Median       3Q      Max
-1.05196 -0.21316  0.02799  0.23154  0.93914

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   89.0382728  1.4997342  59.369  < 2e-16 ***
Scrambling    -0.0889017  0.0088403 -10.056  < 2e-16 ***
GiR           -0.0786524  0.0128151  -6.137 5.03e-09 ***
Putting_total  0.0029228  0.0004097   7.133 2.18e-11 ***
Driving_Dist  -0.0247576  0.0047359  -5.228 4.63e-07 ***
Final_Rnd     -0.0063257  0.0017139  -3.691 0.000294 ***
                                    Page 2
```

```
Driving_Acc   -0.0174621  0.0075753  -2.305 0.022275 *
Nb_Event       0.0139235  0.0067888   2.051 0.041687 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3445 on 184 degrees of freedom
Multiple R-squared: 0.752,      Adjusted R-squared: 0.7426
F-statistic: 79.71 on 7 and 184 DF,  p-value: < 2.2e-16

>
```