Regression Analysis on the National Football League

Student Name

Southern New Hampshire University

MAT-300

[Course Instructor]

August 8, 2013

Introduction

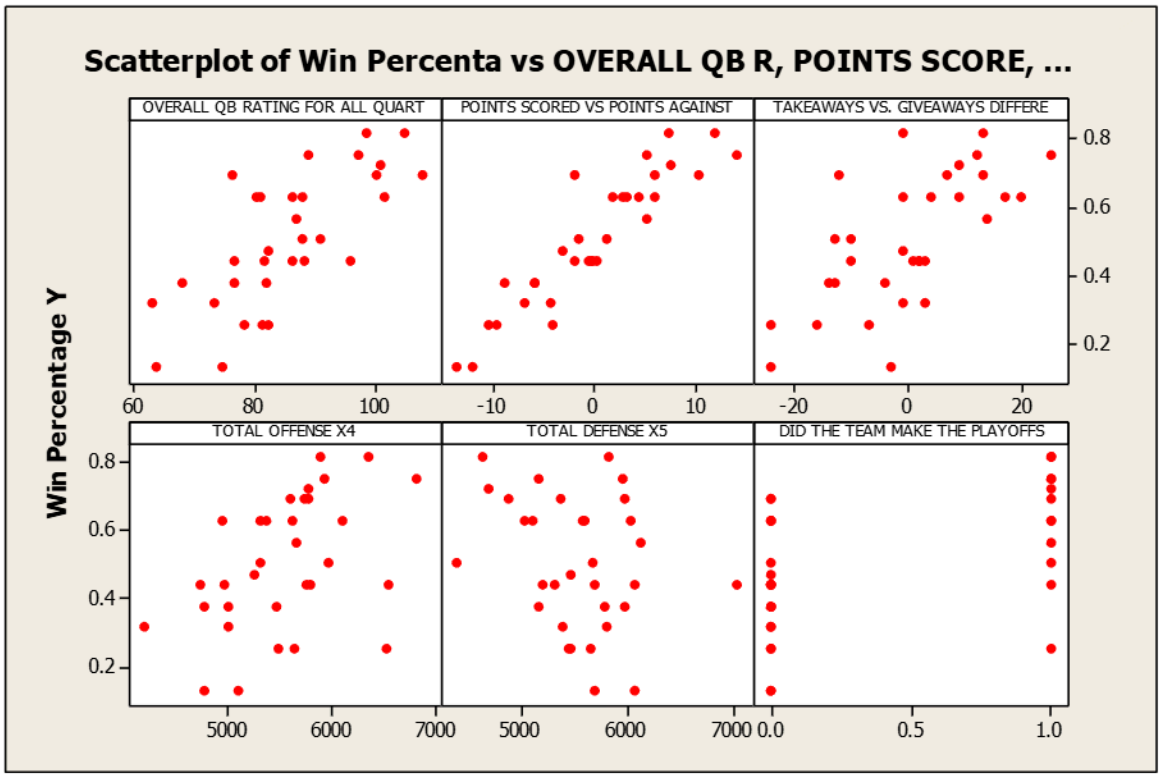
The purpose of this report was to do an analysis on the National Football League. What was analyzed is what parameters have the strongest influence on a team's Win Percentage. The parameters used to predict Win Percentage were: Total Offense, Total Defense, Overall Quarterback Rating, Average Points Scored vs. Average Points Against Differential, Takeaways vs. Giveaways Differential, and Did The Team Make The Playoffs Last Year. The data was collected from the stats and standings section on the ESPN website and considers all 32 National Football League teams in the analysis. The data collected includes only the 2012 regular season, with the exception of the qualitative predictor Did the Team Make the Playoffs Last Year was obtained from the 2011 standings section. The aspiration of this analysis was to identify what parameter or parameters has the strongest correlation to a team's Win Percentage.

Regression Model Building

The main objective of this analysis was to identify what are strong predictors of a team's win percentage in the National Football League. The first task was to select what parameters would be analyzed. The two main units of football are offense and defense, so the most essential aspects associated with each unit was selected. The most important part of the offense is total quarterback rating and total offense, and the most important part of the defense is total defense or yards allowed. Two additional parameters were selected that incorporate both the offense and defense; average points scored vs. average points against differential, and takeaways vs. giveaways differential. The final parameter analyzed was did the team make the playoffs the previous year. This parameter was chosen to see if making the playoffs is a good predictor of win percentage.

The next part of the analysis started off by collecting the appropriate data for all 32 National Football League teams from the ESPN website. The data obtained is displayed in the excel spreadsheet below, with scatterplots to graph the data.

TEAM	DID THE TI	OVERALL (	POINTS SC	TAKEAWA	TOTAL OFI	TOTAL DEF
Pittsburg	Yes	88.1	1.4	-10	5324	4413
Denver	Yes	105.3	12	-1	6366	4652
San Franci	Yes	101.2	7.7	9	5789	4710
Seattle	No	100.6	10.5	13	5610	4899
Chicago	No	80.4	6.1	20	4969	5050
Cincinnati	Yes	88.1	4.4	4	5323	5115
Houston	Yes	89.2	5.3	12	5953	5172
NY Jets	No	68.3	-5.8	-14	4787	5174
San Diego	No	88.5	0	2	4756	5223
Carolina	No	86.5	-0.4	1	5771	5329
Green Bay	Yes	108.3	6.1	7	5751	5388
Arizona	No	63.1	-6.7	-1	4209	5405
Detroit	Yes	81.5	-4	-16	6540	5458
St. Louis	No	82.7	-3.1	-1	5264	5482
Philadelpl	No	78.6	-10.3	-24	5665	5491
Minnesota	No	81.2	1.9	-1	5385	5600
Baltimore	Yes	86.4	3.4	9	5640	5615
Oakland	No	82.5	-9.6	-7	5504	5672
Dallas	No	91.3	-1.5	-13	5994	5687
Kansas Cit	No	63.8	-13.4	-24	5108	5704
Miami	No	76.7	-1.8	-10	4984	5708
Buffalo	No	82.2	-5.7	-13	5486	5806
Cleveland	No	73.6	-4.1	3	5028	5821
Atlanta	Yes	99.1	7.5	13	5906	5849
New Engla	Yes	97.7	14.1	25	6846	5972
Indianapo	No	76.4	-1.9	-12	5799	5988
Tennesse	No	76.9	-8.8	-4	5010	5999
Washingt	No	102.1	3	17	6131	6043
Tampa Bay	No	81.8	-0.3	3	5820	6078
Jacksonvil	No	74.7	-11.9	-3	4788	6088
NY Giants	Yes	87.2	5.3	14	5687	6134
New Orlea	Yes	96.4	0.4	2	6574	7042



The first order main effects model in general form:

$E(y)= \beta 0 + \beta 1x_1 + \beta 2x_2 + \beta 3x_3 + \beta 4\, x_4 + \beta 5x_5 + \beta 6x_6$

Where: y= Win Percentage, x_1= OVERALL QB RATING, x_2= POINTS SCORED VS. POINTS AGAINST DIFFERENTIAL, x_3= TAKEAWAYS VS. GIVEAWAYS DIFFERNTIAL, x_4= TOTAL OFFENSE, x_5= TOTAL DEFENSE, x_6= DID THE TEAM MAKE THE PLAYOFFS (0= no, 1=yes)

The data was then analyzed using the statistical software Minitab 16, the complete regression output is shown below.

Regression Analysis: Win Percenta versus OVERALL QB R, POINTS SCORE, ...

```
The regression equation is
Win Percentage Y = 0.432 + 0.00138 OVERALL QB RATING FOR ALL QUART
                   + 0.0304 POINTS SCORED VS POINTS AGAINST
```

```
- 0.00268 TAKEAWAYS VS. GIVEAWAYS DIFFERE
- 0.000039 TOTAL OFFENSE X4 + 0.000031 TOTAL DEFENSE X5
- 0.0174 DID THE TEAM MAKE THE PLAYOFFS

Predictor                            Coef     SE Coef      T      P
Constant                           0.4323      0.2391   1.81  0.083
OVERALL QB RATING FOR ALL QUART  0.001383    0.002491   0.56  0.584
POINTS SCORED VS POINTS AGAINST  0.030399    0.005929   5.13  0.000
TAKEAWAYS VS. GIVEAWAYS DIFFERE -0.002684    0.002508  -1.07  0.295
TOTAL OFFENSE X4                -0.00003884  0.00004762 -0.82  0.422
TOTAL DEFENSE X5                 0.00003095  0.00004280  0.72  0.476
DID THE TEAM MAKE THE PLAYOFFS     -0.01736     0.04142  -0.42  0.679


S = 0.0836844   R-Sq = 84.9%   R-Sq(adj) = 81.3%


Analysis of Variance

Source          DF       SS       MS      F      P
Regression       6  0.98769  0.16462  23.51  0.000
Residual Error  25  0.17508  0.00700
Total           31  1.16277


Source                           DF   Seq SS
OVERALL QB RATING FOR ALL QUART   1  0.61916
POINTS SCORED VS POINTS AGAINST   1  0.35777
TAKEAWAYS VS. GIVEAWAYS DIFFERE   1  0.00258
TOTAL OFFENSE X4                  1  0.00278
TOTAL DEFENSE X5                  1  0.00418
DID THE TEAM MAKE THE PLAYOFFS    1  0.00123


Unusual Observations

     OVERALL QB
     RATING FOR           Win
Obs   ALL QUART  Percentage Y     Fit  SE Fit  Residual  St Resid
  7          89        0.7500  0.5961  0.0379    0.1539      2.06R
 26          76        0.6880  0.4725  0.0402    0.2155      2.94R

R denotes an observation with a large standardized residual.
```

After reviewing the regression output for the first order main effects model, R^2 and R^2 adjusted are .849 and .813 respectively. The R-squared term is a measure of how much of the variation in the dependent variable is explained by the model. The difference between the R^2 and R^2 adjusted is that the R^2 adjusted value takes into consideration both the sample size and the number of parameters in the model.

The least squares regression equation was Win Percentage Y = 0.432 + 0.00138

OVERALL QB RATING FOR ALL QUART + 0.0304 POINTS SCORED VS POINTS AGAINST - 0.00268 TAKEAWAYS VS. GIVEAWAYS DIFFERE- 0.000039 TOTAL OFFENSE X4 + 0.000031 TOTAL DEFENSE X5- 0.0174 DID THE TEAM MAKE THE PLAYOFFS.

The next step in the analysis was to interpret the β coefficients. β1 = .00138: We estimate the Win Percentage (y) to increase .00138 for every 1 unit increase in OVERALL QB RATING FOR ALL QUART (x1) when POINTS SCORED VS POINTS AGAINST (x2) , TAKEAWAYS VS. GIVEAWAYS DIFFERE (x3), TOTAL OFFENSE (x4), TOTAL DEFENSE (x5) and DID THE TEAM MAKE THE PLAYOFFS (x6) is held fixed. β2 = .0304: We estimate the Win Percentage (y) to increase .0304 for every 1 unit increase in POINTS SCORED VS POINTS AGAINST (x2) when the OVERALL QB RATING FOR ALL QUART (x1) , TAKEAWAYS VS. GIVEAWAYS DIFFERE (x3), TOTAL OFFENSE (x4), TOTAL DEFENSE (x5) and DID THE TEAM MAKE THE PLAYOFFS (x6) is held fixed.β3 = -.00268: We estimate the Win Percentage (y) to increase -.00268 for every 1 unit increase in TAKEAWAYS VS. GIVEAWAYS DIFFERE (x3) when the POINTS SCORED VS POINTS AGAINST (x2) , OVERALL QB RATING FOR ALL QUART (x1), TOTAL OFFENSE (x4), TOTAL DEFENSE (x5) and DID THE TEAM MAKE THE PLAYOFFS (x6) is held fixed.β4 = -.000039: We estimate the Win Percentage (y) to increase -.000039 for every 1 unit increase in TOTAL OFFENSE (x4) when the POINTS SCORED VS POINTS AGAINST (x2) , TAKEAWAYS VS. GIVEAWAYS DIFFERE (x3), OVERALL QB RATING FOR ALL QUART (x1), TOTAL DEFENSE (x5) and DID THE TEAM MAKE THE PLAYOFFS (x6) is held fixed.β5 =.000031: We estimate the Win Percentage (y) to increase .000031 for every 1 unit increase in TOTAL DEFENSE (x5) when the POINTS SCORED VS POINTS AGAINST (x2) ,

TAKEAWAYS VS. GIVEAWAYS DIFFERE (x3), TOTAL OFFENSE (x4), OVERALL QB RATING FOR ALL QUART (x1) and DID THE TEAM MAKE THE PLAYOFFS (x6) is held fixed.β6 = -.0174: We estimate the Win Percentage (y) to increase -.0174 for every 1 unit increase in DID THE TEAM MAKE THE PLAYOFFS (x6) when the POINTS SCORED VS POINTS AGAINST (x2) , TAKEAWAYS VS. GIVEAWAYS DIFFERE (x3), TOTAL OFFENSE (x4), TOTAL DEFENSE (x5) and OVERALL QB RATING FOR ALL QUART (x1) is held fixed. The value β0 = .432 is the estimated win percentage when all x's are zero.

Next a global overall utility test was needed to determine the model's usefulness. An α = .05 was used. Since α = .05 was greater than the observed significance level, p = .000, the data provides strong evidence that at least one of the model coefficients is nonzero. Therefore the overall model appears to be statistically useful for predicting Win Percentage. Local utility test were conducted at α = .05 and based off of each parameter's t-test statistic and p-values, the most significant predictor of Win Percentage was found to be average points scored vs. average points against differential.

The next step in the analysis was to determine if any trends possibly exist, if any terms interact, or if higher order terms would be needed.

The results for the interaction check are shown in the regression output below.

Regression Analysis: Win Percenta versus OVERALL QB R, POINTS SCORE, ...

```
The regression equation is
Win Percentage Y = 5.87 - 0.0720 OVERALL QB RATING FOR ALL QUART
                   + 0.051 POINTS SCORED VS POINTS AGAINST
                   + 0.0150 TAKEAWAYS VS. GIVEAWAYS DIFFERE
                   + 0.000629 TOTAL OFFENSE X4 - 0.00148 TOTAL DEFENSE X5
                   - 0.00040 QBR*Points Diff + 0.000700 QBR*Take
                   - 0.000006 QBR*Total Offense + 0.000018 QBR*Total Defense
                   - 0.000729 Points Scored*Takeaways
                   + 0.000020 Points Scored*Total Offense
                   - 0.000017 Points Scored*Total Defense
                   - 0.000014 Takeaways* Total Offense
                   - 0.000001 Takeaways* Total Defense
                   - 0.000000 Total Offense*Total Defense
                   + 0.934 DID THE TEAM MAKE THE PLAYOFFS
```

```
                  + 0.00471 DiD The Team*Overall QBR
                  - 0.0017 DiD The Team*Points Scored
                  + 0.0150 DiD The Team*Takeaways
                  - 0.000084 DiD The Team*Offense
                  - 0.000173 DiD The Team*Defense


Predictor                              Coef     SE Coef      T      P
Constant                              5.874       3.632   1.62  0.137
OVERALL QB RATING FOR ALL QUART     -0.07202     0.04220  -1.71  0.119
POINTS SCORED VS POINTS AGAINST       0.0511      0.1248   0.41  0.691
TAKEAWAYS VS. GIVEAWAYS DIFFERE      0.01498     0.04631   0.32  0.753
TOTAL OFFENSE X4                   0.0006286   0.0007421   0.85  0.417
TOTAL DEFENSE X5                    -0.001477    0.001014  -1.46  0.176
QBR*Points Diff                     -0.000396    0.001406  -0.28  0.784
QBR*Take                             0.0006999   0.0005187   1.35  0.207
QBR*Total Offense                 -0.00000614  0.00000703  -0.87  0.403
QBR*Total Defense                  0.00001843  0.00001126   1.64  0.133
Points Scored*Takeaways            -0.0007290   0.0004125  -1.77  0.108
Points Scored*Total Offense        0.00002033  0.00002016   1.01  0.337
Points Scored*Total Defense       -0.00001708  0.00001469  -1.16  0.272
Takeaways* Total Offense          -0.00001380  0.00000829  -1.66  0.127
Takeaways* Total Defense          -0.00000060  0.00000846  -0.07  0.945
Total Offense*Total Defense       -0.00000001  0.00000010  -0.10  0.923
DID THE TEAM MAKE THE PLAYOFFS         0.9343      0.9126   1.02  0.330
DiD The Team*Overall QBR             0.004706    0.008374   0.56  0.586
DiD The Team*Points Scored            -0.00173     0.02702  -0.06  0.950
DiD The Team*Takeaways                 0.01495     0.01313   1.14  0.281
DiD The Team*Offense                -0.0000843   0.0001692  -0.50  0.629
DiD The Team*Defense                -0.0001726   0.0001861  -0.93  0.375


S = 0.0722482   R-Sq = 95.5%   R-Sq(adj) = 86.1%


Analysis of Variance

Source          DF        SS        MS      F      P
Regression      21  1.110569  0.052884  10.13  0.000
Residual Error  10  0.052198  0.005220
Total           31  1.162767
```

After inspection and an α=.05, there were no terms whose T test statistic and P-value was significant. Therefore, according to the analysis no terms interact.

Higher order terms were then evaluated to see if they would be beneficial in the model. The results for the regression output are displayed below.

Regression Analysis: Win Percenta versus OVERALL QB R, POINTS SCORE, ...

```
The regression equation is
Win Percentage Y = - 2.03 - 0.0234 OVERALL QB RATING FOR ALL QUART
                   + 0.0305 POINTS SCORED VS POINTS AGAINST
```

```
                - 0.00169 TAKEAWAYS VS. GIVEAWAYS DIFFERE
                + 0.000831 TOTAL OFFENSE X4 + 0.000433 TOTAL DEFENSE X5
                + 0.000131 QBR^2 - 0.000188 Points Scored^2
                - 0.000050 Takeaways^2 - 0.000000 Total Offense^2
                - 0.000000 Total Defense^2 + 0.0017 Did the Team_Sqr
```

Predictor	Coef	SE Coef	T	P
Constant	-2.028	1.597	-1.27	0.219
OVERALL QB RATING FOR ALL QUART	-0.02343	0.02191	-1.07	0.298
POINTS SCORED VS POINTS AGAINST	0.030467	0.006005	5.07	0.000
TAKEAWAYS VS. GIVEAWAYS DIFFERE	-0.001686	0.002650	-0.64	0.532
TOTAL OFFENSE X4	0.0008315	0.0004680	1.78	0.091
TOTAL DEFENSE X5	0.0004326	0.0004492	0.96	0.347
QBR^2	0.0001306	0.0001205	1.08	0.291
Points Scored^2	-0.0001876	0.0003560	-0.53	0.604
Takeaways^2	-0.0000500	0.0001136	-0.44	0.665
Total Offense^2	-0.00000007	0.00000004	-1.84	0.080
Total Defense^2	-0.00000004	0.00000004	-0.90	0.377
Did the Team_Sqr	0.00166	0.04304	0.04	0.970

S = 0.0816566 R-Sq = 88.5% R-Sq(adj) = 82.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	11	1.029411	0.093583	14.04	0.000
Residual Error	20	0.133356	0.006668		
Total	31	1.162767			

From the values obtained from the regression output, a test was then conducted to see if higher order terms are beneficial.

The proposed higher order model is:

$y= \beta 0 + \beta 1x_1 + \beta 2x_2 + \beta 3x_3 + \beta 4\ x_4 + \beta 5x_5 + \beta 6x_6 + \beta 7x_1^2 + \beta 8x_2^2 + \beta 9x_3^2 + \beta 10\ x_4^2 + \beta 11x_5^2 + \beta 12x_6^2$

Where: y= Win Percentage, x_1= OVERALL QB RATING, x_2= POINTS SCORED VS. POINTS AGAINST DIFFERENTIAL, x_3= TAKEAWAYS VS. GIVEAWAYS DIFFERNTIAL, x_4= TOTAL OFFENSE, x_5= TOTAL DEFENSE, x_6= DID THE TEAM MAKE THE PLAYOFFS

H0: $\beta 7=\beta 8= \beta 9=\beta 10 =\beta 11=\beta 12=0$

H1: $\beta 7\neq 0$ or $\beta 8\neq 0$ $\beta 9\neq 0$ or $\beta 10\neq 0$ or $\beta 11\neq 0$ or $\beta 12\neq 0$

SSER= .17508 SSEC= .133356 MSEC= .006668

$F = \frac{(SSER - SSEC)/(k-g)}{SSEC/[n-(k+1)]}$

k= number of terms in the model

n= sample size

k − g = Number of β parameters specified in H0

$$F = \frac{(.17508 - .133356)/(6)}{.133356/[32 - 13]}$$

$$= .9908$$

The rejection region for the test of using α = .05.

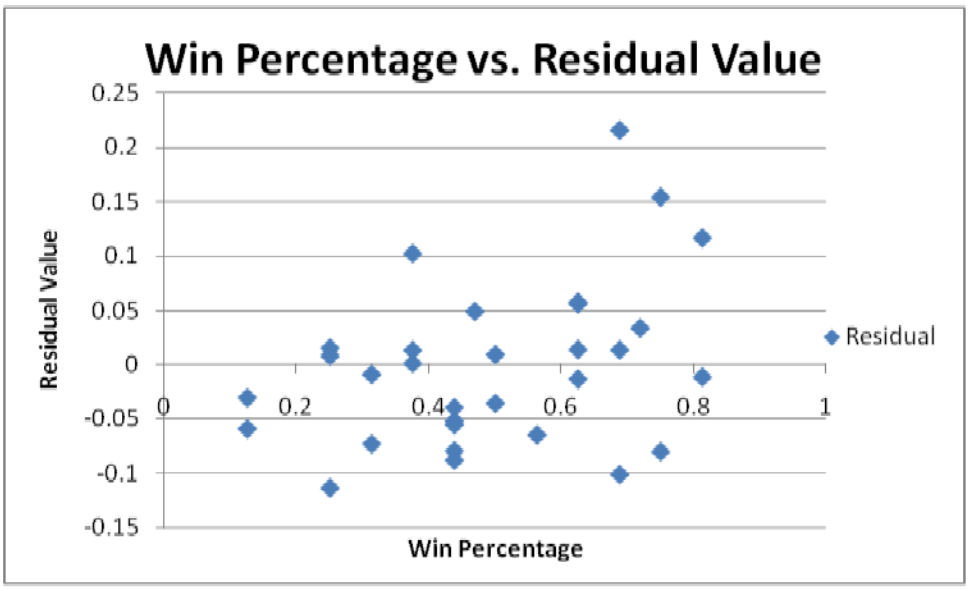
Fr= F at α, v1,v2

v1 = k − g = degrees of freedom for the numerator and v2 = n − (k + 1) = degrees of freedom for the denominator

The rejection region at α = .05, v1=6, v2= 19 is 2.6823

Since the calculated F test statistic .9908 is less than the critical 2.6823 F score, therefore the null hypothesis is not rejected and we conclude that the quadratic terms are not helpful in predicting values of y.

An analysis was done to validate the implementation of the error term €. A plot of residuals is show below for the proposed model.



After a quick inspection of the plot, it appears that a possible multiplicative trend may exist. A test for heteroscedasticity was conducted and the dataset was divided into two subsamples $X \geq .5$ and $x < .5$. The output for the datasets is displayed below.

Subset 1 Data Output

Regression Analysis: Win Percenta versus OVERALL QB R, POINTS SCORE, ...

```
The regression equation is
Win Percentage Y = 0.253 + 0.00258 OVERALL QB RATING FOR ALL QUART
                   + 0.0256 POINTS SCORED VS POINTS AGAINST
                   - 0.00270 TAKEAWAYS VS. GIVEAWAYS DIFFERE
                   - 0.000036 TOTAL OFFENSE X4 + 0.000036 TOTAL DEFENSE X5
                   - 0.100 Did The Team Make The Playoffs


Predictor                            Coef     SE Coef      T      P
Constant                           0.2534      0.2784   0.91  0.386
OVERALL QB RATING FOR ALL QUART  0.002576    0.003203   0.80  0.442
POINTS SCORED VS POINTS AGAINST  0.025555    0.005846   4.37  0.002
TAKEAWAYS VS. GIVEAWAYS DIFFERE -0.002702    0.002851  -0.95  0.368
TOTAL OFFENSE X4                -0.00003564  0.00005157  -0.69  0.507
TOTAL DEFENSE X5                 0.00003558  0.00004621   0.77  0.461
Did The Team Make The Playoffs     -0.10009     0.07101  -1.41  0.192


S = 0.0584368   R-Sq = 83.5%   R-Sq(adj) = 72.5%


Analysis of Variance

Source          DF        SS        MS     F      P
Regression       6  0.155754  0.025959  7.60  0.004
Residual Error   9  0.030734  0.003415
Total           15  0.186488
```

Subset 2 Data Output

Regression Analysis: Win Percenta versus OVERALL QB R, POINTS SCORE, ...

```
The regression equation is
Win Percentage Y = 0.302 - 0.00021 OVERALL QB RATING FOR ALL QUART
                   + 0.0152 POINTS SCORED VS POINTS AGAINST
                   - 0.00118 TAKEAWAYS VS. GIVEAWAYS DIFFERE
                   + 0.000036 TOTAL OFFENSE X4 + 0.000017 TOTAL DEFENSE X5
                   + 0.0060 Did The Team Make The Playoffs


Predictor                             Coef     SE Coef      T      P
Constant                            0.3017      0.3755   0.80  0.442
OVERALL QB RATING FOR ALL QUART   -0.000212    0.003343  -0.06  0.951
POINTS SCORED VS POINTS AGAINST     0.01517     0.01177   1.29  0.229
TAKEAWAYS VS. GIVEAWAYS DIFFERE   -0.001182    0.003751  -0.32  0.760
TOTAL OFFENSE X4                0.00003649  0.00008880   0.41  0.691
TOTAL DEFENSE X5                0.00001716  0.00007119   0.24  0.815
Did The Team Make The Playoffs      0.00604     0.05037   0.12  0.907


S = 0.0866011   R-Sq = 50.2%   R-Sq(adj) = 17.0%


Analysis of Variance

Source          DF        SS        MS     F      P
Regression       6  0.068002  0.011334  1.51  0.277
Residual Error   9  0.067498  0.007500
Total           15  0.135499
```

The test for heteroscedasticity is,

$H0 = \frac{\sigma1^2}{\sigma2^2} = 1$

$H1: = \frac{\sigma1^2}{\sigma2^2} \neq 1$

Where $\sigma1^2 = \sigma_1^2$, $\sigma2^2 = \sigma_2^2$ $\sigma1^2$

$F = \frac{MSE\ 1}{MSE\ 2} = \frac{.0075}{.0034} = 2.196$

With an α=.05

Degrees of freedom for subsample 1 and subsample 2 =7

From F table, @ v1=9, v2=9, α=.05. F= 3.787

Since the calculated F test statistic 2.196 is not greater than the critical 3.787 F score, the null hypothesis is not rejected and we conclude that the assumption of equal variance is accepted. Therefore the error term €, should be added to the complete model.

After all the analysis thus far, the completed model was:

$E(y)= \beta0 + \beta1x_1 + \beta2x_2 + \beta3x_3 + \beta4\ x_4 + \beta5x_5 + \beta6x_6 + €$

Where: y= Win Percentage, x_1= OVERALL QB RATING, x_2= POINTS SCORED VS. POINTS AGAINST DIFFERENTIAL, x_3= TAKEAWAYS VS. GIVEAWAYS DIFFERNTIAL, x_4= TOTAL OFFENSE, x_5= TOTAL DEFENSE, x_6= DID THE TEAM MAKE THE PLAYOFFS (0= no, 1=yes)

The least-squares regression was: Y = 0.432 + 0.00138 OVERALL QB RATING FOR ALL QUART + 0.0304 POINTS SCORED VS POINTS AGAINST - 0.00268 TAKEAWAYS VS. GIVEAWAYS DIFFERE- 0.000039 TOTAL OFFENSE X4 + 0.000031 TOTAL DEFENSE X5- 0.0174 DID THE TEAM MAKE THE PLAYOFFS.

The next step in the process was to remove any unnecessary terms for the final model. To arrive at the final model was an iterative process that methodically removes one insignificant term at a time and then refitting the model.

An α= .05 was used.

```
The regression equation is
Win Percentage Y = 0.432 + 0.00138 OVERALL QB RATING FOR ALL QUART
                   + 0.0304 POINTS SCORED VS POINTS AGAINST
```

```
                  - 0.00268 TAKEAWAYS VS. GIVEAWAYS DIFFERE
                  - 0.000039 TOTAL OFFENSE X4 + 0.000031 TOTAL DEFENSE X5
                  - 0.0174 DID THE TEAM MAKE THE PLAYOFFS


Predictor                            Coef     SE Coef      T      P
Constant                           0.4323      0.2391   1.81  0.083
OVERALL QB RATING FOR ALL QUART  0.001383    0.002491   0.56  0.584
POINTS SCORED VS POINTS AGAINST  0.030399    0.005929   5.13  0.000
TAKEAWAYS VS. GIVEAWAYS DIFFERE -0.002684    0.002508  -1.07  0.295
TOTAL OFFENSE X4                -0.00003884  0.00004762 -0.82  0.422
TOTAL DEFENSE X5                 0.00003095  0.00004280  0.72  0.476
DID THE TEAM MAKE THE PLAYOFFS     -0.01736     0.04142  -0.42  0.679


S = 0.0836844   R-Sq = 84.9%   R-Sq(adj) = 81.3%
```

Upon inspection of the regression output for this model, the most insignificant quantitative predictor in a team's Win Percentage was Overall QB Rating.

```
The regression equation is
Win Percentage Y = 0.504 + 0.0313 POINTS SCORED VS POINTS AGAINST
                  - 0.00248 TAKEAWAYS VS. GIVEAWAYS DIFFERE
                  - 0.000027 TOTAL OFFENSE X4 + 0.000027 TOTAL DEFENSE X5
                  - 0.0169 DID THE TEAM MAKE THE PLAYOFFS


Predictor                            Coef     SE Coef      T      P
Constant                           0.5043      0.1981   2.55  0.017
POINTS SCORED VS POINTS AGAINST  0.031269    0.005642   5.54  0.000
TAKEAWAYS VS. GIVEAWAYS DIFFERE -0.002477    0.002447  -1.01  0.321
TOTAL OFFENSE X4                -0.00002651  0.00004156 -0.64  0.529
TOTAL DEFENSE X5                 0.00002695  0.00004163  0.65  0.523
DID THE TEAM MAKE THE PLAYOFFS     -0.01692     0.04086  -0.41  0.682


S = 0.0825633   R-Sq = 84.8%   R-Sq(adj) = 81.8%
```

Upon inspection of the regression output for this model, the most insignificant quantitative predictor in a team's Win Percentage was Total Offense.

Regression Analysis: Win Percenta versus POINTS SCORE, TAKEAWAYS VS, ...

```
The regression equation is
Win Percentage Y = 0.452 + 0.0290 POINTS SCORED VS POINTS AGAINST
                  - 0.00167 TAKEAWAYS VS. GIVEAWAYS DIFFERE
                  + 0.000010 TOTAL DEFENSE X5
                  - 0.0255 DID THE TEAM MAKE THE PLAYOFFS
```

Predictor	Coef	SE Coef	T	P
Constant	0.4520	0.1784	2.53	0.017
POINTS SCORED VS POINTS AGAINST	0.029029	0.004366	6.65	0.000
TAKEAWAYS VS. GIVEAWAYS DIFFERE	-0.001675	0.002076	-0.81	0.427
TOTAL DEFENSE X5	0.00001043	0.00003223	0.32	0.749
DID THE TEAM MAKE THE PLAYOFFS	-0.02550	0.03816	-0.67	0.510

S = 0.0816516 R-Sq = 84.5% R-Sq(adj) = 82.2%

Upon inspection of the regression output for this model, the most insignificant quantitative predictor in a team's Win Percentage was Total Defense.

Regression Analysis: Win Percenta versus POINTS SCORE, TAKEAWAYS VS, ...

The regression equation is
Win Percentage Y = 0.509 + 0.0284 POINTS SCORED VS POINTS AGAINST
- 0.00140 TAKEAWAYS VS. GIVEAWAYS DIFFERE
- 0.0240 DID THE TEAM MAKE THE PLAYOFFS

Predictor	Coef	SE Coef	T	P
Constant	0.50938	0.01993	25.56	0.000
POINTS SCORED VS POINTS AGAINST	0.028368	0.003797	7.47	0.000
TAKEAWAYS VS. GIVEAWAYS DIFFERE	-0.001401	0.001865	-0.75	0.459
DID THE TEAM MAKE THE PLAYOFFS	-0.02395	0.03725	-0.64	0.525

S = 0.0803357 R-Sq = 84.5% R-Sq(adj) = 82.8%

Upon inspection of the regression output for this model, the most insignificant quantitative predictor in a team's Win Percentage was Takeaways vs. Giveaways Differential.

Regression Analysis: Win Percenta versus POINTS SCORE, DID THE TEAM

The regression equation is
Win Percentage Y = 0.507 + 0.0263 POINTS SCORED VS POINTS AGAINST
- 0.0189 DID THE TEAM MAKE THE PLAYOFFS

Predictor	Coef	SE Coef	T	P
Constant	0.50749	0.01962	25.87	0.000
POINTS SCORED VS POINTS AGAINST	0.026281	0.002568	10.23	0.000
DID THE TEAM MAKE THE PLAYOFFS	-0.01894	0.03637	-0.52	0.607

S = 0.0797297 R-Sq = 84.1% R-Sq(adj) = 83.1%

The current model has one mandatory qualitative predictor and one quantitative predictor. From the previously proposed model, ($E(y)= \beta0 + \beta1x_1 + \beta2x_2 + \beta3x_3 + \beta4\ x_4 + \beta5x_5 + \beta6x_6 + €$

Where: y= Win Percentage, x_1= OVERALL QB RATING, x_2= POINTS SCORED VS. POINTS AGAINST DIFFERENTIAL, x_3= TAKEAWAYS VS. GIVEAWAYS DIFFERNTIAL, x_4= TOTAL OFFENSE, x_5= TOTAL DEFENSE, x_6= DID THE TEAM MAKE THE PLAYOFFS), the R-squared and R-squared adjusted values were .849 and .813 respectively. The new current model had R-squared and R-squared adjusted values of .841 and .831. Because of the new current model's higher R-squared adjusted value and removed unnecessary quantitative terms, this model is the best model for the chosen objective of identifying what parameters are strong predictors of a team's win percentage.

The final model is:

$E(y) = \beta0 + \beta1x_1 + \beta2x_2 + €$

Where: y= Win Percentage, x_1= POINTS SCORED VS. POINTS AGAINST DIFFERENTIAL, x_2= DID THE TEAM MAKE THE PLAYOFFS (0= no, 1=yes)

Conclusion

After all the analysis was complete, the final model was:

$$E(y) = \beta 0 + \beta 1 x_1 + \beta 2 x_2 + €$$

Where: y= Win Percentage, x_1= POINTS SCORED VS. POINTS AGAINST DIFFERENTIAL, x_2= DID THE TEAM MAKE THE PLAYOFFS (0= no, 1=yes). It was determined that the most significant predictor of a team's win percentage was the average points scored versus points against differential. This result makes sense if one considers how the National Football League works. If a football team on average wins games by 12 points, then that team's win percentage will more than likely be higher than a team that on average wins games by 1 point. The point differential highlights this statistic. In the National Football League regardless of how many yards a team gains or how many turnovers the team forces, if the team does not score more points than the other team, then the team will lose. The analysis confirms this belief, and the results obtained are conclusive.

References

2012 NFL Regular Season Divisional Standings - National Football League - ESPN. (2012.). *ESPN: The Worldwide Leader In Sports*. Retrieved August 4, 2013, from http://espn.go.com/nfl/standings/_/year/2012/seasontype/2

2012 NFL Statistics - National Football League - ESPN. (2013). *ESPN: The Worldwide Leader In Sports*. Retrieved August 4, 2013, from http://espn.go.com/nfl/statistics