# Taming the Tiger with SAS®/STAT: Using the SAS® System to Analyze the 1997 Masters Golf Tournament

**Joseph Earley, Loyola Marymount University, Los Angeles**

## Abstract

This paper illustrates how the SAS System may be used to analyze a major sporting event, the 1997 Masters golf tournament. PROC PLOT, PROC CORR, PROC REG, PROC STEPWISE and PROC RSQUARE are used to develop an econometric model which points out the factors affecting the golfer's score. This paper may be of interest to golfers and those familiar with elementary regression analysis. The procedures were run using SAS for Windows version 6.12.

## Introduction

In April of each year the Augusta National Golf Club hosts the premiere golfing event of the year, the Masters Tournament. What singled out the 1997 tournament was the inclusion of a dynamic new 21 year old professional named Tiger Woods - who many watchers of the game would say is the best new member of the Professional Golfers Association to come along in a generation. In addition to being an excellent putter and having a fine short-game, what many golf fans enjoyed seeing was the tremendous length which Tiger Woods hit his tee shots. Since the Augusta course is known for favoring a golfer who can hit the tee shot a long way, the golfing community was very interested in how well Tiger would perform in his first Masters as a professional.

## Theoretical Model

A general knowledge of the game of golf would indicate that there are several measurable factors which account for a player's performance: the length and accuracy with which he hits his tee shots, the accuracy of his iron shots and finally, how well he putts. In order to determine whether or not these factors were important at the 1997 Masters, an econometric model was developed.

The theoretical model used to determine factors influencing scoring behavior in the tournament has the following specification:

SCORE = f( DD, GIR, FWAYS, PUTTS)

where:

| | |
|---|---|
| SCORE | Total score for the tournament |
| DD | Driving distance |
| GIR | Number of greens hit in regulation |
| FWAYS | Number of fairways hit in regulation |
| PUTTS | Total number of putts for entire tournament |

## Empirical Results

An ordinary least squares regression model of the following form was estimated using PROC REG as the statistical procedure:

$$\text{Score} = \beta_o + \beta_1 \text{ DD} + \beta_2 \text{ GIR} + \beta_3 \text{ FWAYS} + \beta_4 \text{ PUTTS} + \mu_i$$

with $\mu_i$ being the stochastic error term with the standard assumed statistical properties. Results derived from the estimation of this model are presented below in standard SAS format.

## Model: 1997 Masters

Dependent Variable: SCORE

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---|---|---|---|---|---|
| Model | 4 | 1598.99613 | 399.74903 | 57.106 | 0.0001 |
| Error | 41 | 287.00387 | 7.00009 | | |
| C Total | 45 | 1886.00000 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 2.64577 | R-square | 0.8478 |
| Dep Mean | 291.00000 | Adj R-sq | 0.8330 |
| C.V. | 0.90920 | | |

| Variable | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob > \|T\| |
|---|---|---|---|---|
| INTERCEP | 261.777331 | 16.34173505 | 16.019 | 0.0001 |
| DD | -0.048163 | 0.03055416 | -1.576 | 0.1226 |
| GIR | -1.306805 | 0.11216754 | -11.650 | 0.0001 |
| FWAYS | 0.002917 | 0.15212399 | 0.019 | 0.9848 |
| PUTTS | 0.851667 | 0.09598722 | 8.873 | 0.0001 |

For comparison purposes, the same model was run for the 1994 Masters tournament. Results for that model are presented below.

## Model: 1994 Masters

Dependent Variable: SCORE

Analysis of Variance

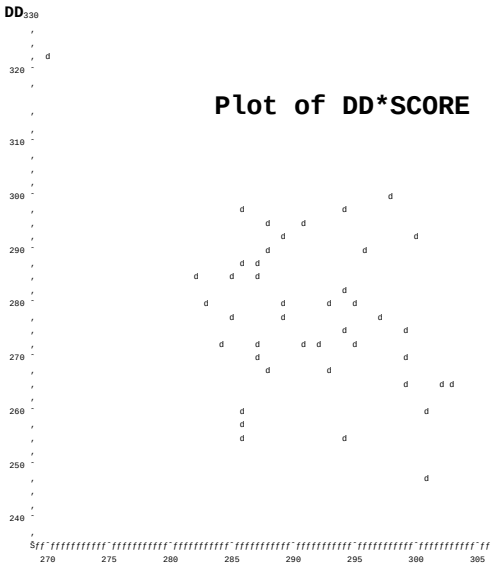
| Source | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---|---|---|---|---|---|
| Model | 4 | 1356.66465 | 339.16616 | 19.750 | 0.0001 |
| Error | 46 | 789.96280 | 17.17310 | | |
| C Total | 50 | 2146.62745 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 4.14404 | R-square | 0.6320 |
| Dep Mean | 293.45098 | Adj R-sq | 0.6000 |
| C.V. | 1.41218 | | |

| Variable | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob > \|T\| |
|---|---|---|---|---|
| INTERCEP | 229.529817 | 20.48366622 | 11.206 | 0.0001 |
| DD | 0.042684 | 0.05097906 | 0.837 | 0.4068 |
| GIR | -1.079873 | 0.15613453 | -6.916 | 0.0001 |
| FWAYS | -0.146346 | 0.18382187 | -0.796 | 0.4300 |
| PUTTS | 0.880718 | 0.15317316 | 5.750 | 0.0001 |

A noticeable occurrence between the 1994 and 1997 Masters tournament is summarized in the PROC PLOT procedure below. Tiger Woods averaged an astonishing 323 yards off the tee! This was 45 yards above the average of 274 for the rest of the

golfers in the tournament. This score was 3.15 standard deviations above average. The next longest driver in the 1997 tournament was professional Scott Hoch, whose average distance of 300 yards was 26 yards above average or 1.8 standard deviations above average. In contrast, the longest two drivers in the 1994 Masters were 300 and 299 yards respectively, which was 2.16 and 2.08 standard deviations above average. The enormous length off the tee is considered an advantage in achieving low scores. In the 1994 Masters, the DD variable was not statistically significant, the t-value being 0.837 with a p-value of 0.4068. In addition, the sign of the coefficient was incorrect. In 1997 however, the t-value for the DD variable was -1.576 with an associated p-value of 0.1226. This indicates that the old adage about "driving for show and putting for dough" may have changed. When the model is re-estimated without the Tiger Woods observation, the t-value of the DD variable drops to -0.932 with an associated p-value of 0.3567. This also illustrates the sensitivity of the regression equation to the deletion of one observation.




## Model: NoTiger

Dependent Variable: SCORE

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---|---|---|---|---|---|
| Model | 4 | 1163.88944 | 290.97236 | 42.899 | 0.0001 |
| Error | 40 | 271.31056 | 6.78276 | | |
| C Total | 44 | 1435.20000 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 2.60437 | R-square | 0.8110 |
| Dep Mean | 291.46667 | Adj R-sq | 0.7921 |
| C.V. | 0.89354 | | |

| Variable | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob > \|T\| |
|---|---|---|---|---|
| INTERCEP | 258.915086 | 16.19574082 | 15.987 | 0.0001 |
| DD | -0.030140 | 0.03232600 | -0.932 | 0.3567 |
| GIR | -1.254860 | 0.11557317 | -10.858 | 0.0001 |
| FWAYS | -0.011906 | 0.15006062 | -0.079 | 0.9372 |
| PUTTS | 0.820657 | 0.09665989 | 8.490 | 0.0001 |

## Conclusion

It appears that driving distance at the Master golf tournament may have been an important factor in the record setting victory of Tiger Woods. His ability to drive long distances places him in excellent position to attack the golf course and thus have shorter putts. Since the Masters tournament is always played at the Augusta National Golf Club, this may be a permanent advantage for long drivers such as Tiger Woods. Other major tournaments in golf are rotated among numerous courses and thus tournament results may be studied to determine whether or not driving distance is statistically important in general or just at Augusta.

## References

Freund, Rudolf and Littell, Ramon (1991), SAS System for Regression, Second edition, Cary, N.C.: SAS Institute, Inc.

Golfweek, April 19, 1997

Gujarati, Damodar (1995), Basic Econometrics, Third Edition, New York: McGraw-Hill, Inc.

Orzech, William and Earley, Joseph 1994), The Use of SAS Regression Procedures in Analyzing a Major Sporting Event: The Case of the 1994 Masters Tournament: Proceedings of the Second Annual Conference of the Western Users of SAS Software, pp. 164-168.

SAS Institute Inc.(1993), SAS/ETS User's Guide. Version 6. Second Edition, Cary, NC: SAS Institute Inc.

SAS Institute Inc.(1991), SAS/ETS Software: Applications Guide 1, Version 6, First Edition. Cary, N.C.: SAS Institute, Inc.

SAS Institute Inc.(1993), SAS/ETS Software: Applications Guide 2, Version

6, First Edition. Cary, N.C.: SAS Institute, Inc.

SAS Institute Inc.(1986), SAS System for Forecasting Time Series, Cary, N.C.: SAS Institute Inc.

Schlotzhauer, Sandra and Littell, Ramon(1987), SAS System for Elementary Statistical Analysis. Cary, N.C. SAS Institute Inc.

## Contact author at:

Joseph Earley
Department of Economics
Loyola Marymount University, Los Angeles
90045-8410
e-mail: jearley@popmail.lmu.edu