

PGA Tour Statistics

Data Munging

```
all <- read.csv("./data/PGA Stats.csv", header = FALSE, stringsAsFactors = F)
header <- all[c(8, 13, 18, 23, 33, 38, 104, 56, 61), 1]
header <- append(c("Name", "Year"), header)
```

Example of creating the player data from the csv

```
#### Tiger####
temp1 <- all[c(1, 2, 9, 14, 19, 24, 34, 39, 105, 57, 62), 1]
temp2 <- all[c(1, 2, 9, 14, 19, 24, 34, 39, 110, 57, 67), 2]
temp3 <- all[c(1, 2, 9, 14, 19, 24, 34, 39, 106, 53, 63), 3]
temp <- rbind(temp1, temp2, temp3)
temp4 <- rbind(header, temp)
temp4 <- as.data.frame(temp, stringsAsFactors = F)
colnames(temp4) <- header
# remove $ , % from data
for (i in grep("%", temp4)) temp4[, i] <- gsub("%", "", temp4[, i])
for (i in grep("\\$", temp4)) temp4[, i] <- gsub("\\$", "", temp4[, i])
for (i in grep(",", temp4)) temp4[, i] <- gsub(",", "", temp4[, i])
for (i in 2:11) temp4[, i] <- as.numeric(temp4[, i])
tiger <- temp4
```

##		Name	Year	Driving Distance	Driving Accuracy	Percentage
## temp1	Tiger	woods	2013	291.6		61.85
## temp2	Tiger	woods	2012	297.4		63.93
## temp3	Tiger	woods	2011	293.7		48.90
##						
##						
## temp1				67.55		0.835
## temp2				67.58		0.332
## temp3				67.74		0.258
##						
##						
## temp1				4.00	68.65 NA	3059
## temp2				3.97	68.90 NA	2269
## temp3				3.92	70.46 NA	318
##						
##						
## temp1						
## temp2						
## temp3						

Set Testing Data

```
allPlayers <- rbind(baddeley, bradely, clark, crane, duval, fowler, furyk, kuchar,  
  ohair, tiger, watson)  
testData <- allPlayers[, -c(1, 2, 8, 9, 10)]  
colnames(testData) <- c("drivingDistance", "drivingAccuracyPercentage",  
  "greensRegulationPercentage",  
  "Putting", "birdieAverage", "Money")
```

Test Regression Models

AIC

```
aicFormula <- step(lm1.1)
```

```
## Start:  AIC=748.8
## Money ~ drivingDistance + drivingAccuracyPercentage + greensRegulationPercentage +
##      Putting + birdieAverage
##
##              Df Sum of Sq      RSS AIC
## - greensRegulationPercentage  1  4.19e+10 5.28e+13 747
## <none>                                5.27e+13 749
## - drivingAccuracyPercentage    1  6.04e+12 5.88e+13 750
## - birdieAverage                1  6.72e+12 5.94e+13 750
## - drivingDistance              1  1.04e+13 6.32e+13 751
## - Putting                      1  1.29e+13 6.56e+13 752
##
## Step:  AIC=746.8
## Money ~ drivingDistance + drivingAccuracyPercentage + Putting +
##      birdieAverage
##
##              Df Sum of Sq      RSS AIC
## <none>                                5.28e+13 747
## - birdieAverage                1  7.24e+12 6.00e+13 748
## - drivingAccuracyPercentage    1  1.27e+13 6.55e+13 750
## - Putting                      1  1.29e+13 6.57e+13 751
## - drivingDistance              1  1.72e+13 7.00e+13 752
```

BIC

```
Posterior probabilities(%):
  drivingDistance drivingAccuracyPercentage greensRegulationPercentage      Putting
birdieAverage
59.0            84.6                75.0                28.3            94.7

Coefficient posterior expected values:
(Intercept) drivingDistance drivingAccuracyPercentage greensRegulationPercentage
Putting birdieAverage
-24235511      69202                85475                17339
1898624      -20312
```

Compare Different Inputs

Test model differences for different inputs

```
lm2.1 <- lm(Money ~ drivingDistance + drivingAccuracyPercentage +
greensRegulationPercentage +
  Putting + birdieAverage, data = testData)
lm3.1 <- lm(Money ~ drivingDistance + drivingAccuracyPercentage + Putting, data =
testData)
lm4.1 <- lm(Money ~ drivingDistance + drivingAccuracyPercentage +
greensRegulationPercentage +
  Putting, data = testData)
lm5.1 <- lm(Money ~ drivingDistance + drivingAccuracyPercentage + Putting +
birdieAverage, data = testData)
```

```
##
## Call:
## lm(formula = Money ~ drivingDistance + drivingAccuracyPercentage +
##     greensRegulationPercentage + Putting + birdieAverage, data = testData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1655175 -1038026  -325401   793270  3614976
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -28394819    11611967   -2.45   0.024 *
## drivingDistance     83687     42063     1.99   0.060 .
## drivingAccuracyPercentage    120142     79401     1.51   0.146
## greensRegulationPercentage   -13386    106187   -0.13   0.901
## Putting         1987725     897836     2.21   0.039 *
## birdieAverage     -30583     19161    -1.60   0.126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1620000 on 20 degrees of freedom
## (8 observations deleted due to missingness)
## Multiple R-squared:  0.497,    Adjusted R-squared:  0.371
## F-statistic: 3.95 on 5 and 20 DF,  p-value: 0.0118
```

```
##
## Call:
## lm(formula = Money ~ drivingDistance + drivingAccuracyPercentage +
##     Putting, data = testData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2558301 -1083434  -406456   775363  3885450
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -33035885    10178473   -3.25   0.0037 **
## drivingDistance     96693      30221    3.20   0.0041 **
## drivingAccuracyPercentage  116179      52111    2.23   0.0363 *
## Putting        1743837      896311    1.95   0.0646 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1650000 on 22 degrees of freedom
## (8 observations deleted due to missingness)
## Multiple R-squared:  0.427,    Adjusted R-squared:  0.349
## F-statistic: 5.47 on 3 and 22 DF,  p-value: 0.00578
```

```
##
## Call:
## lm(formula = Money ~ drivingDistance + drivingAccuracyPercentage +
##     greensRegulationPercentage + Putting, data = testData)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-2208768	-1163867	-346425	797109	3916308

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-34980793	11247008	-3.11	0.0053	**
drivingDistance	108449	40513	2.68	0.0141	*
drivingAccuracyPercentage	143417	80876	1.77	0.0907	.
greensRegulationPercentage	-48080	107701	-0.45	0.6599	
Putting	1802672	922543	1.95	0.0641	.

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1680000 on 21 degrees of freedom
## (8 observations deleted due to missingness)
## Multiple R-squared:  0.433,    Adjusted R-squared:  0.325
## F-statistic: 4.01 on 4 and 21 DF,  p-value: 0.0144
```



```
##
## Call:
## lm(formula = Money ~ drivingDistance + drivingAccuracyPercentage +
##     Putting + birdieAverage, data = testData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1642436 -1057258  -330514   850253  3601873
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -27769547    10250338   -2.71    0.013 *
## drivingDistance      80150      30600    2.62    0.016 *
## drivingAccuracyPercentage  112500      50064    2.25    0.036 *
## Putting        1975023      871008    2.27    0.034 *
## birdieAverage    -31078      18310   -1.70    0.104
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1590000 on 21 degrees of freedom
## (8 observations deleted due to missingness)
## Multiple R-squared:  0.496,    Adjusted R-squared:  0.401
## F-statistic: 5.18 on 4 and 21 DF,  p-value: 0.00462
```

Test Regression Models without Duval's Rows

Remove David Duval's results because they are poor and skew the models

AIC

```
aicFormula <- step(lm1.2)
```

```
## Start: AIC=652.3
## Money ~ drivingDistance + drivingAccuracyPercentage + greensRegulationPercentage +
## Putting + birdieAverage
##
##              Df Sum of Sq      RSS AIC
## <none>                2.83e+13 652
## - greensRegulationPercentage 1  8.52e+12 3.68e+13 656
## - drivingDistance            1  8.95e+12 3.73e+13 657
## - drivingAccuracyPercentage  1  9.62e+12 3.79e+13 657
## - birdieAverage              1  1.86e+13 4.70e+13 662
## - Putting                    1  2.69e+13 5.52e+13 666
```

```
##
## Call:
## lm(formula = Money ~ drivingDistance + drivingAccuracyPercentage +
## greensRegulationPercentage + Putting + birdieAverage, data = testNoDuval)
##
## Coefficients:
##              (Intercept)              drivingDistance
##              -51321525              88382
## drivingAccuracyPercentage greensRegulationPercentage
##              163659              269226
##              Putting              birdieAverage
##              4231188              -97641
```

BIC

Posterior probabilities(%):				
drivingDistance	drivingAccuracyPercentage	greensRegulationPercentage	Putting	
birdieAverage				
82.1	80.1	64.8	100.0	
95.2				
Coefficient posterior expected values:				
(Intercept)	drivingDistance	drivingAccuracyPercentage	greensRegulationPercentage	
Putting	birdieAverage			
-42365071	85704	140407	167625	
3350533	-73739			

Compare Different Inputs

Test model differences for different inputs

```
##
## Call:
## lm(formula = Money ~ drivingDistance + drivingAccuracyPercentage +
##     greensRegulationPercentage + Putting + birdieAverage, data = testNoDuval)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2173403  -831860    9165   597306  2230930
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -51321525   11735642  -4.37  0.00041 ***
## drivingDistance     88382     38133    2.32  0.03319 *
## drivingAccuracyPercentage  163659     68118    2.40  0.02798 *
## greensRegulationPercentage  269226     119074    2.26  0.03716 *
## Putting         4231188     1053920    4.01  0.00090 ***
## birdieAverage    -97641      29204   -3.34  0.00385 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1290000 on 17 degrees of freedom
## (8 observations deleted due to missingness)
## Multiple R-squared:  0.675,    Adjusted R-squared:  0.579
## F-statistic: 7.05 on 5 and 17 DF,  p-value: 0.000977
```

```
##
## Call:
## lm(formula = Money ~ drivingDistance + drivingAccuracyPercentage +
##     Putting, data = testNoDuval)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2732544 -1217783   -7088    854747   2986187
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -51818711   14289175   -3.63   0.0018 **
## drivingDistance    144416     38424     3.76   0.0013 **
## drivingAccuracyPercentage    186662     68413     2.73   0.0133 *
## Putting         3655916    1265920     2.89   0.0094 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1570000 on 19 degrees of freedom
## (8 observations deleted due to missingness)
## Multiple R-squared:  0.46,    Adjusted R-squared:  0.375
## F-statistic:  5.4 on 3 and 19 DF,  p-value: 0.00737
```

```
##
## Call:
## lm(formula = Money ~ drivingDistance + drivingAccuracyPercentage +
##     greensRegulationPercentage + Putting, data = testNoDuval)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2849266 -1185755  -108662   857575  2952935
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -51747742    14682696   -3.52   0.0024 **
## drivingDistance    142012     43285     3.28   0.0042 **
## drivingAccuracyPercentage    180197     85004     2.12   0.0482 *
## greensRegulationPercentage    15511     114812     0.14   0.8940
## Putting        3664471    1301492     2.82   0.0114 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1620000 on 18 degrees of freedom
## (8 observations deleted due to missingness)
## Multiple R-squared:  0.461,    Adjusted R-squared:  0.341
## F-statistic: 3.85 on 4 and 18 DF,  p-value: 0.0198
```

```
##
## Call:
## lm(formula = Money ~ drivingDistance + drivingAccuracyPercentage +
##     Putting + birdieAverage, data = testNoDuval)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1969698 -1067576  -19985    594207   2884846
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -52236759   12999514   -4.02  0.00081 ***
## drivingDistance    136278     35143    3.88  0.00110 **
## drivingAccuracyPercentage  237430     66274    3.58  0.00213 **
## Putting        3898766    1156694    3.37  0.00341 **
## birdieAverage   -55562      24944   -2.23  0.03891 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1430000 on 18 degrees of freedom
## (8 observations deleted due to missingness)
## Multiple R-squared:  0.577,    Adjusted R-squared:  0.483
## F-statistic: 6.14 on 4 and 18 DF,  p-value: 0.00269
```