First Name: Susheela

Last Name: Polepalli

M-ID: M10727836

**Homework 4**

**Generate a report to answer the following questions.**

**1. Read PGA data into R (PGA.csv). Below is the description of variables.**

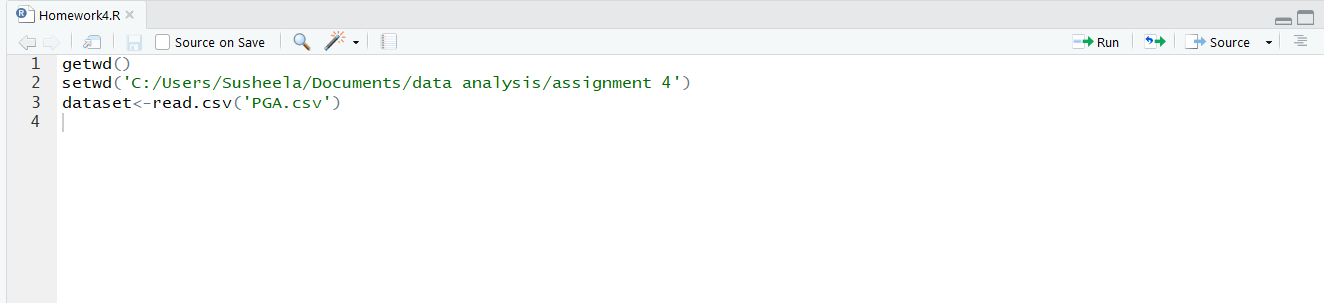
**Source: sportsillustrated.cnn.com**

**Description: Performance statistics and winnings for 196 PGA participants during, 2004 season.**

**Variable: Name, Age, Average Drive (Yards), Driving accuracy (percent), Greens on regulation (%), Average # of putts, Save Percent, Money Rank, # Events, Total Winnings ($), Average winnings ($).**

Solution:

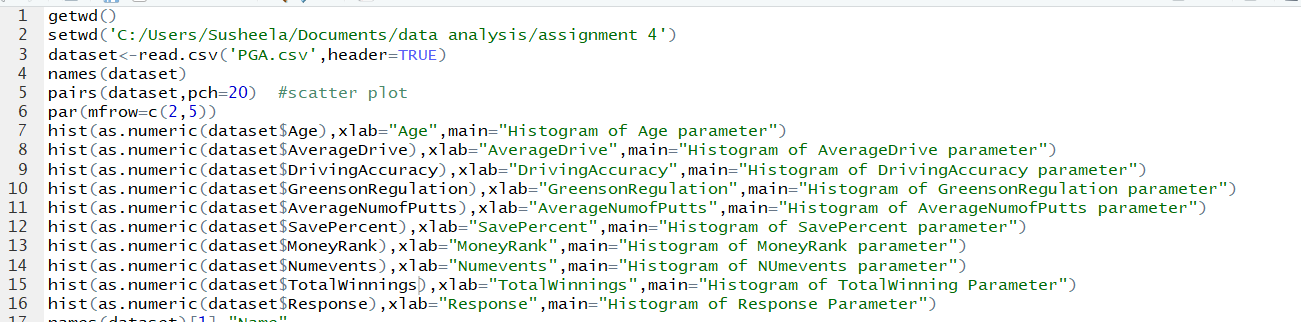
Code:



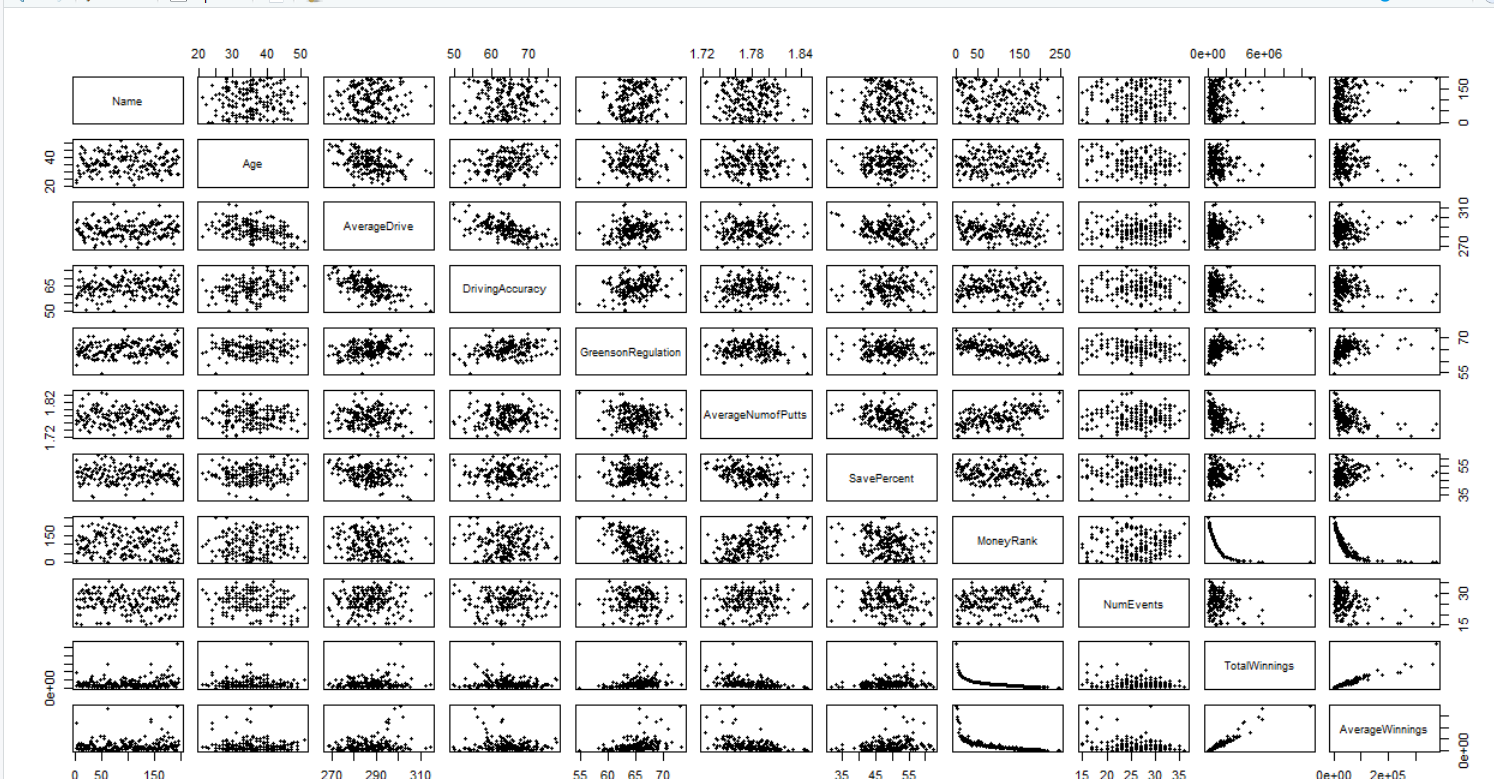
**2. Visualize the data using scatter plot and histogram.**

Solution:

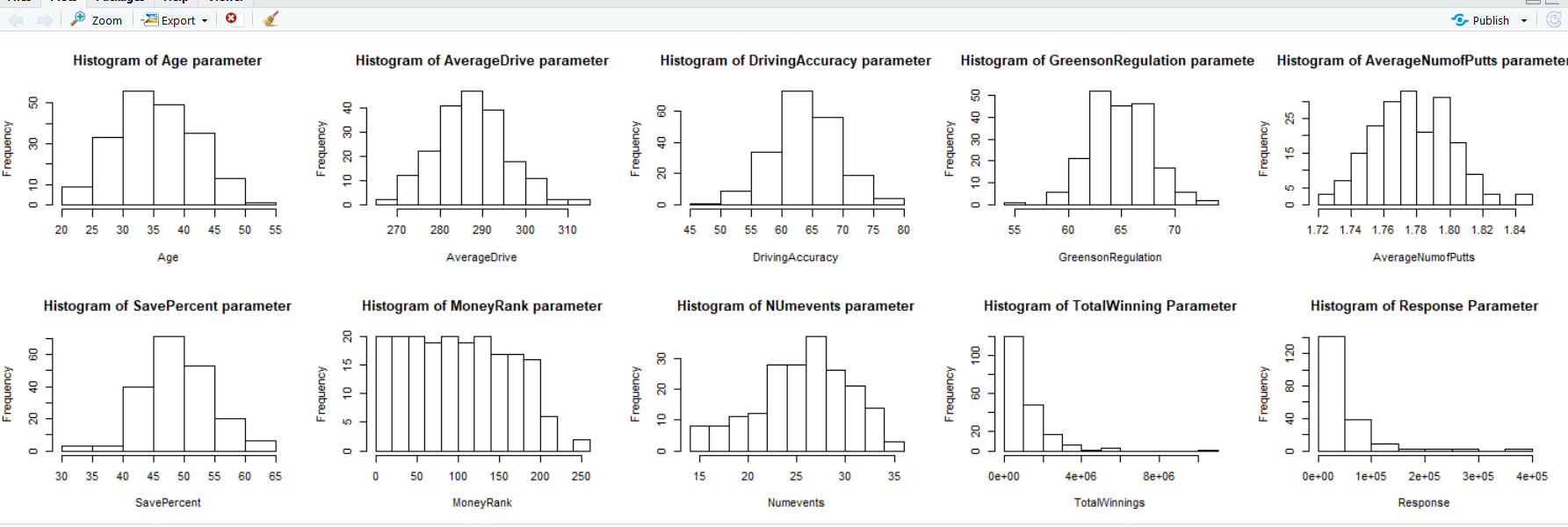
Code:



Output:

Scatter plot:

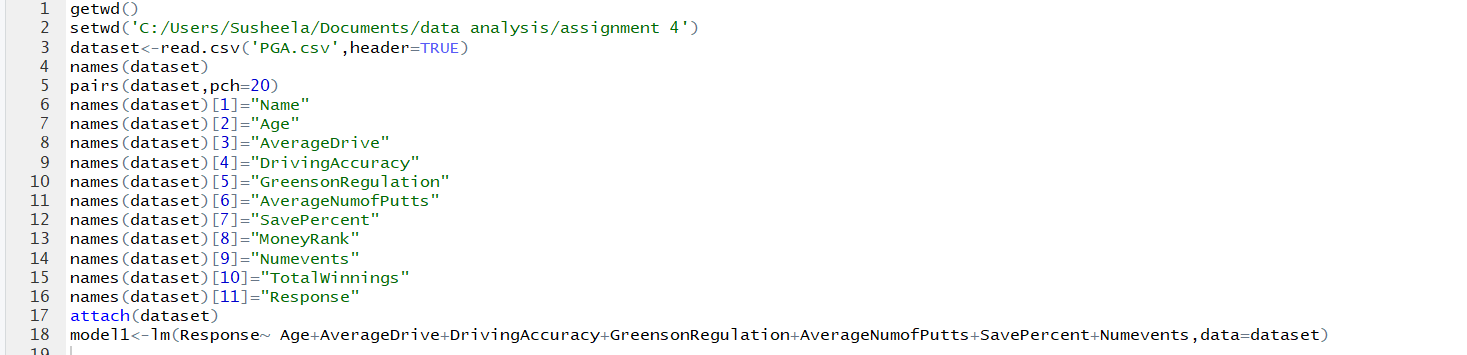
Histogram:



**3. Build a linear regression using Average winnings as response variable and using Age, Average Drive (Yards), Driving accuracy (percent), Greens on regulation (%), Average # of putts, Save Percent, and # Events as covariates.**

Solution:

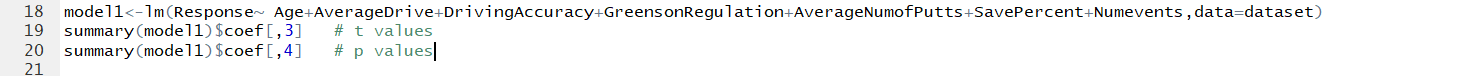
Code:



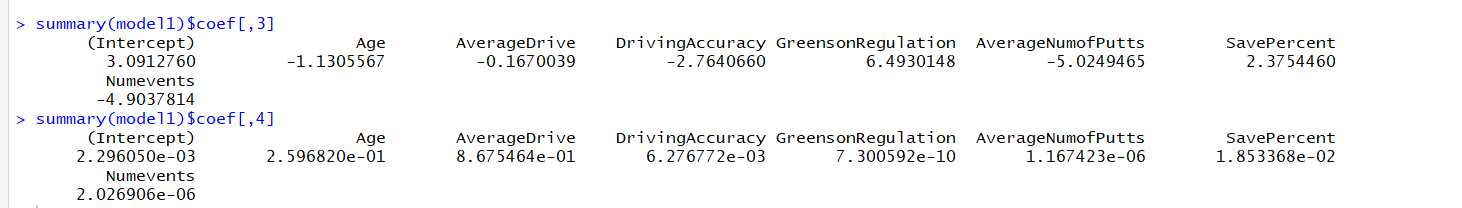
**4. Perform t tests for these coefficient estimates. Obtain t statistics and p values, interpret the results, make a conclusion (i.e. reject or not reject) and explain why. Note: please explain what the null hypothesis is.**

Solution:

Code:



Output:



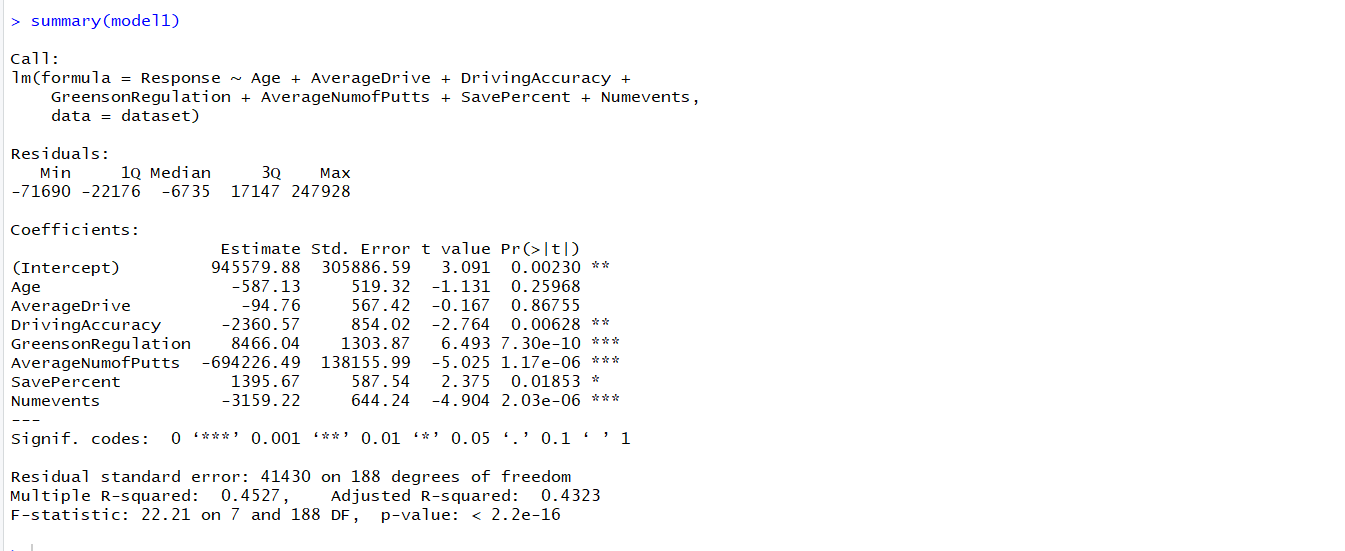
The t-value for intercept is 3.091 and p-value is 2.296050e-03. This means slope is different from 0. Thus, the null hypothesis H0: β1=0 is rejected.

**5. Use F test to test the significance of the regression. Obtain the F statistic and p value, interpret the results and make a conclusion.**

Code:



Output:

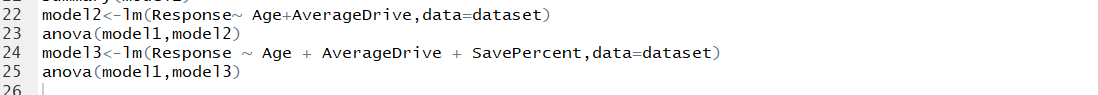


The F-value is 22.21383 and p value is < 2.2e – 16. The p value is less than 0.05. Thus, the null hypothesis is H0: β1=0 is rejected.

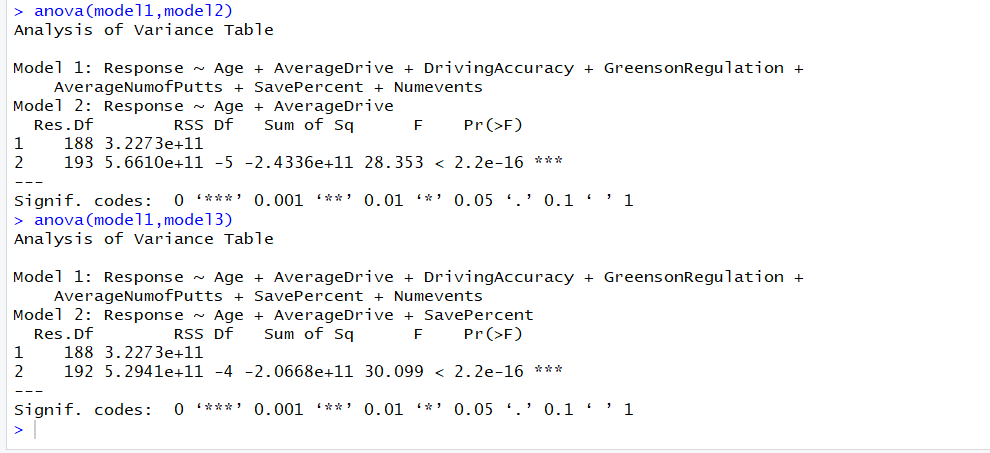
**6. Use a partial F test to test for two variables Age and Average Drive (Yards) together. According to your results, what do you conclude? Similarly, use the partial F test to test for three variables Age, Average Drive (Yards), and Save Percent together, what do you conclude?**

Solution:

Code:

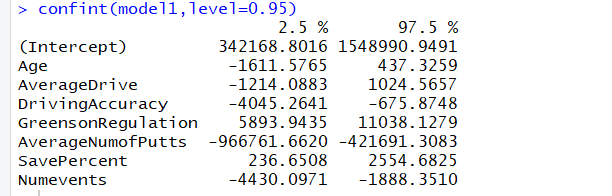


Output:



**7. Obtain the interval estimation for all the intercept and slope coefficients.**

**Solution:**



**8. Using the regression in question 3, make a prediction for the case of:**

**Age = 35,**

**AverageDrive = 287,**

**DrivingAccuracy = 64,**

**GreensonRegulation = 64.9,**

**AverageNumofPutts = 1.778,**

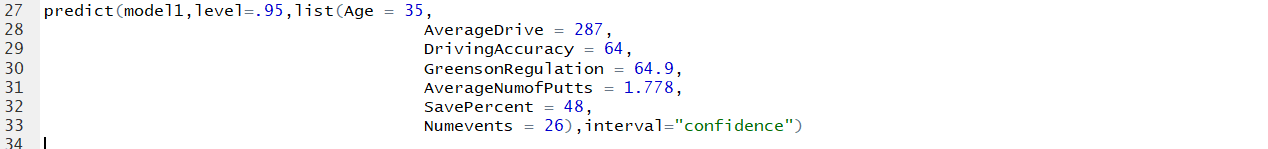
**SavePercent = 48,**

**NumEvents = 26,**

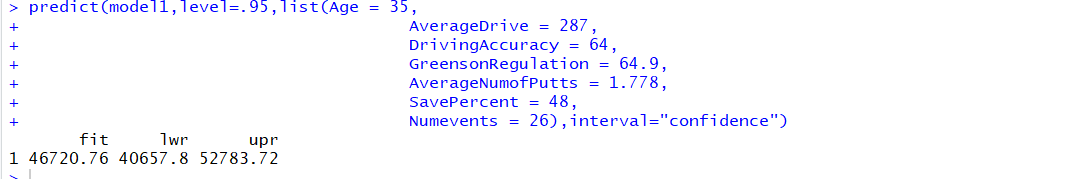
**The prediction should include fitted value and interval estimation.**

Solution:

Code:



Output:



**9. Similarly, make another prediction for the case of**

**Age = 42,**

**AverageDrive = 295,**

**DrivingAccuracy = 69,**

**GreensonRegulation = 67.7,**

**AverageNumofPutts = 1.80,**

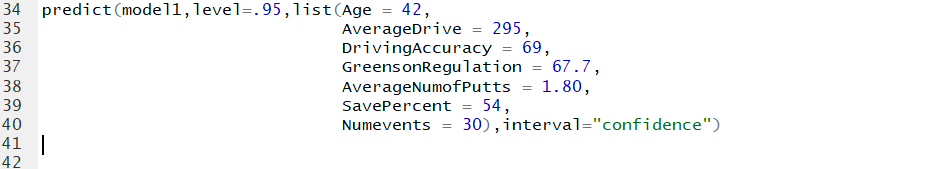
**SavePercent = 54,**

**NumEvents = 30,**

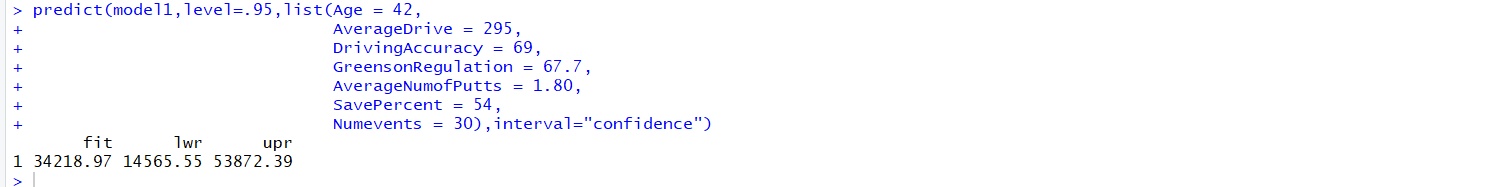
**The prediction should again include the fitted value and interval estimation. Compare the interval from question 8, what do you observe? For example, which interval is wider? And why?**

Solution:

Code:



Output:



It is observed that the interval has become wider than the interval obtained in question 8.

**10. Obtain the standardized regression coefficients and compare the influence of all variables.**

Solution:

Code:



Output:

