Andrew Goldstein
February 13th, 2012
Intro Data Analysis
Prof. Jay Emerson

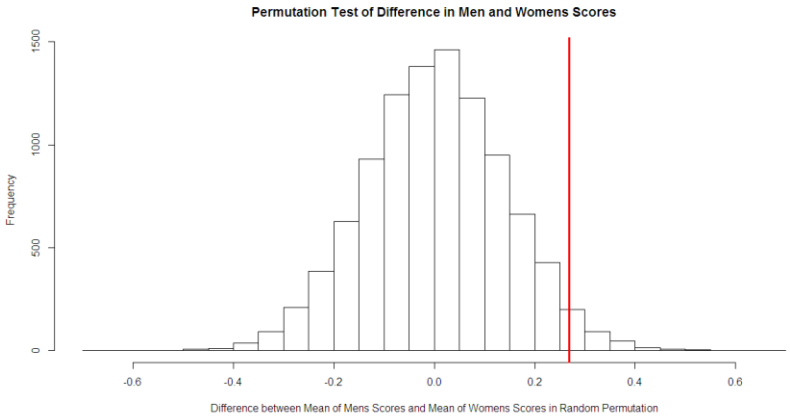
Homework Assignment #4

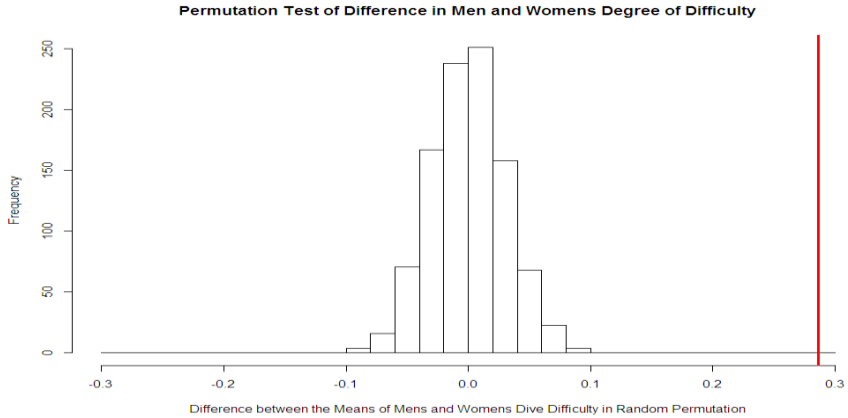
1) Degree of difficulty, revisited

Before examining the effect of gender on the data, it is valuable to do a quick analysis of the mean score and mean difficulty of the men's and women's dives:

```
#Scores
> mean(menmeans)
[1] 7.597222
> mean(womenmeans)
[1] 7.327976
```

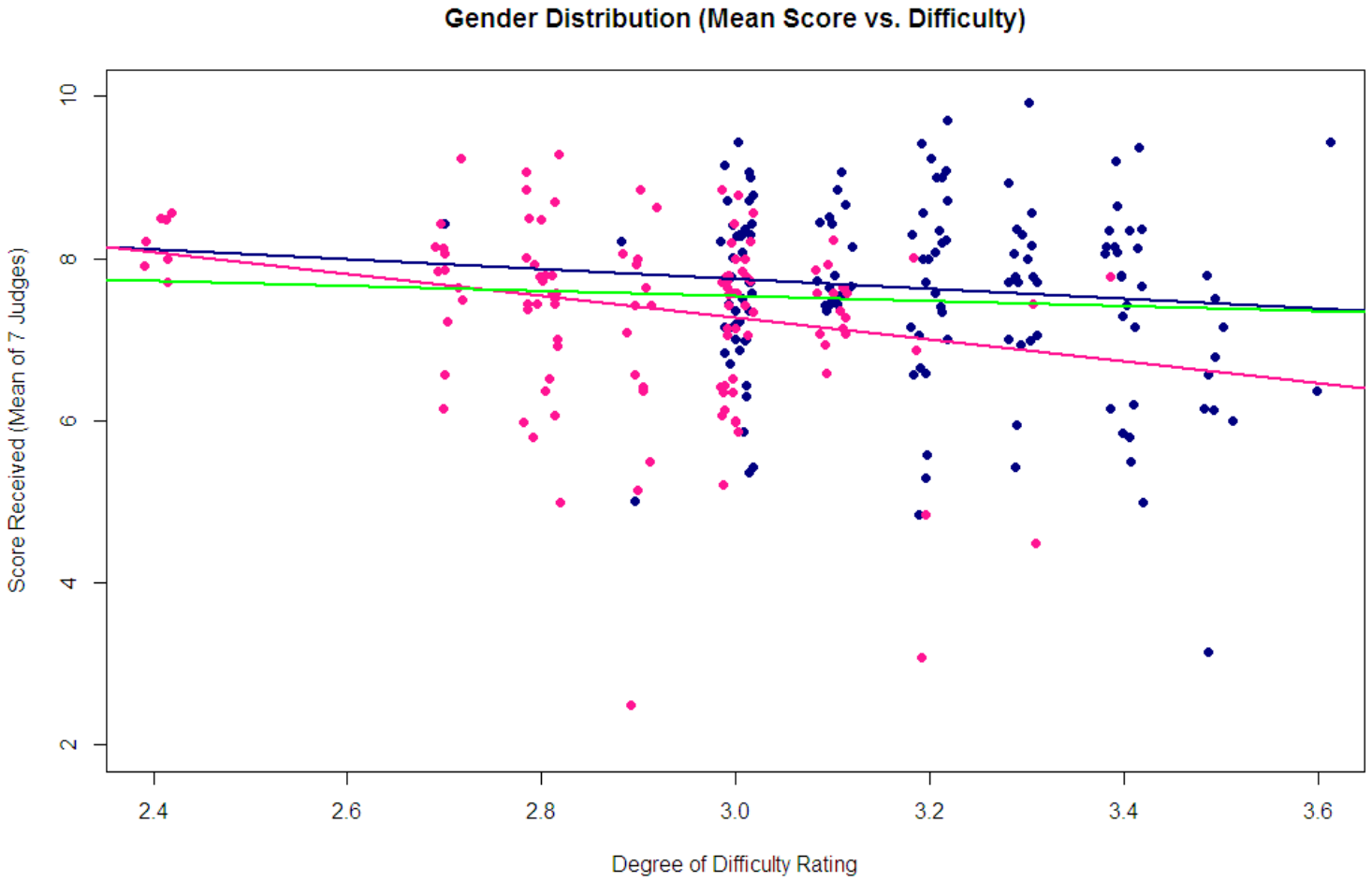
```
#Difficulty
> mean(men$Difficulty)
[1] 3.191667
> mean(women$Difficulty)
[1] 2.905
```





A permutation test was conducted for the differences between these two data sets and it appears that the split in this data was not arbitrary. The p-value for the plot on the left (difference in men and women's scores) is 0.05 and the p-value for the plot on the right (difference in men and women's degree of difficulty) is 0.00. Thus we can conclude that we are generally dealing with two distinct data sets, each with a different mean degree of difficulty and a different mean score. It is important to note that this conclusion means that the two-linear-model approach on the following page is appropriate.

The figure below shows the breakdown between men and women very clearly. Pink dots are women's dives and blue dots are men's dives. I have excluded three dives from the final round data as outliers given that they are more than 3.5 residual standard errors below the center of the data. I believe that these three low scores were not caused by degree of difficulty; they were simply large mistakes that would magnify a possible negative correlation (given what I know about how a regression line is calculated).



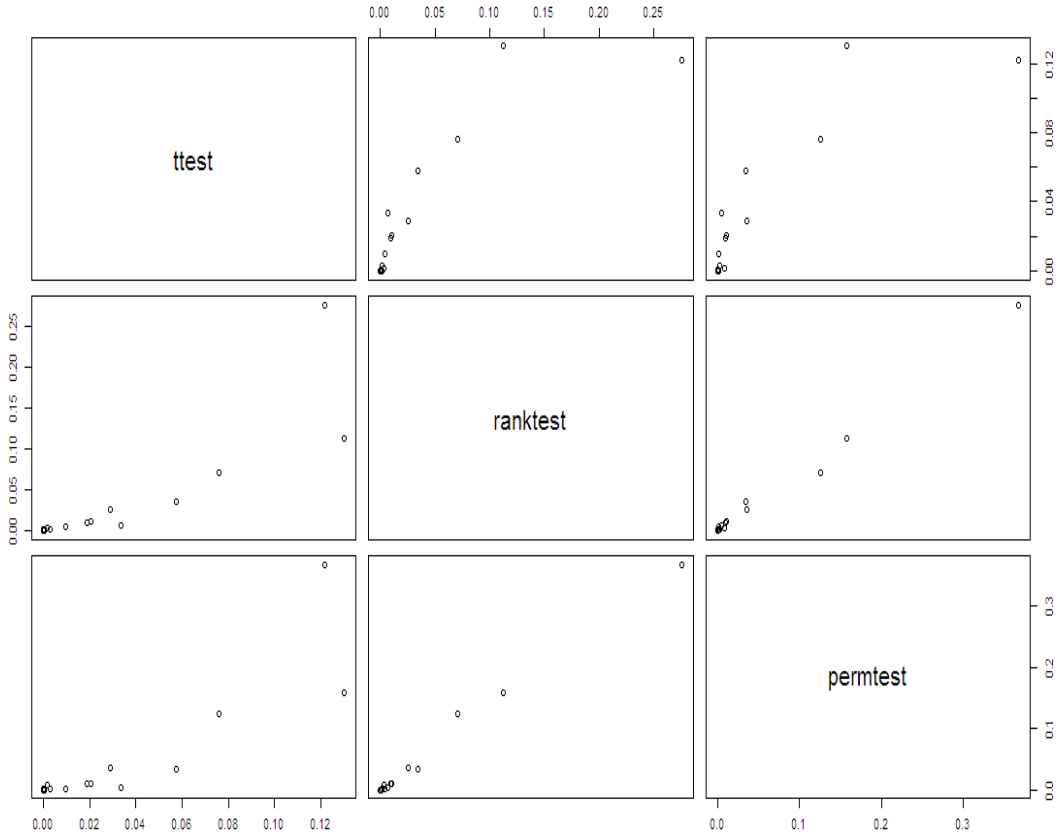

```
Combined: Slope of Green Line:  -0.3121     P-Value:  0.251
     Men:  Slope of Blue Line:  -0.6101     P-Value:  0.229
   Women:  Slope of Pink Line:  -1.354      P-Value:  0.00277**
```

To conclude, I do believe there is evidence that women have a strong negative association between difficulty of the dive and the score awarded. I do not extend this same conclusion to the men given the high p-value associated with the negative slope of the regression line. Additionally, these conclusions are restricted to the final round data of this particular data set.

2) Nationalistic bias, revisited

Using the code in the appendix, I generated the following matrix of p-values from the three tests (T-test, Rank sum test, and Permutation test) for each of the 16 judges who had the opportunity to judge divers from their own country. I excluded Facheng Wang from consideration because he was so honest that his data skewed the comparative "pairs" plot, which is shown to the right.

	ttest	ranktest	permtest
[1,]	2.937727e-06	5.119683e-06	0.0000
[2,]	3.811205e-04	5.394849e-04	0.0001
[3,]	2.049610e-02	1.044084e-02	0.0122
[4,]	1.216830e-01	2.754925e-01	0.3707
[5,]	2.892695e-02	2.522154e-02	0.0349
[6,]	9.533766e-03	4.066190e-03	0.0004
[7,]	7.599610e-02	7.099630e-02	0.1375
[8,]	1.793088e-03	3.281412e-03	0.0076
[9,]	3.374302e-02	6.541452e-03	0.0050
[10,]	5.753185e-02	3.462949e-02	0.0325
[11,]	2.195358e-04	1.566293e-03	0.0024
[12,]	2.627390e-04	2.646500e-04	0.0001
[13,]	1.912006e-02	9.732442e-03	0.0086
[14,]	1.300972e-01	1.120208e-01	0.1537
[15,]	3.042726e-03	9.427125e-04	0.0041
[16,]	5.904769e-06	1.415609e-05	0.0000



This look at the data shows that some level of nationalistic bias is common across most judges. Aside from Facheng Wang of China, the only other judges which did not show convincing levels of nationalistic bias were Judge 4 (Michel Boussard of France), Judge 7 (Felix Calderon), and Judge 14 (Kathy Seaman).

Another component to consider is that the number of times which a judge actually has the opportunity to score someone from their own country varies quite widely. Judge 9 (Michael Geissbuhler) only had 3 opportunities to score his own countrymen and so I think it would be hard to conclude that he is actively including a nationalistic bias. However, Judge 2 (Madeline Barnett of Australia), Judge 9 (Steve McFarland of the United States), and Judge 16 (Oleg Zaitsev of Russia) are all definitely incorporating elements of nationalistic bias into their scoring because they had p-values of essentially 0 and judged roughly 40 dives from their own countrymen.

Looking at the three different tests is a valuable tool in data analysis because it helps to ensure that the assumptions regarding the hypothesis are accurate. Finding similar results across multiple tests simply ensures that the assumptions are valid and make the tests' conclusions more credible. Just as the three p-values for any particular judge (across a row) are quite similar, the "pairs" scatterplots show that the p-values of the tests are very highly correlated. If the tests were to have given different results, then that would have been an indicator that an assumption was being violated or that the variables were not independent in some way. It would be a hint to the person analyzing the data to go back and explore the data further to avoid the making a false conclusion or misusing a certain statistical test.