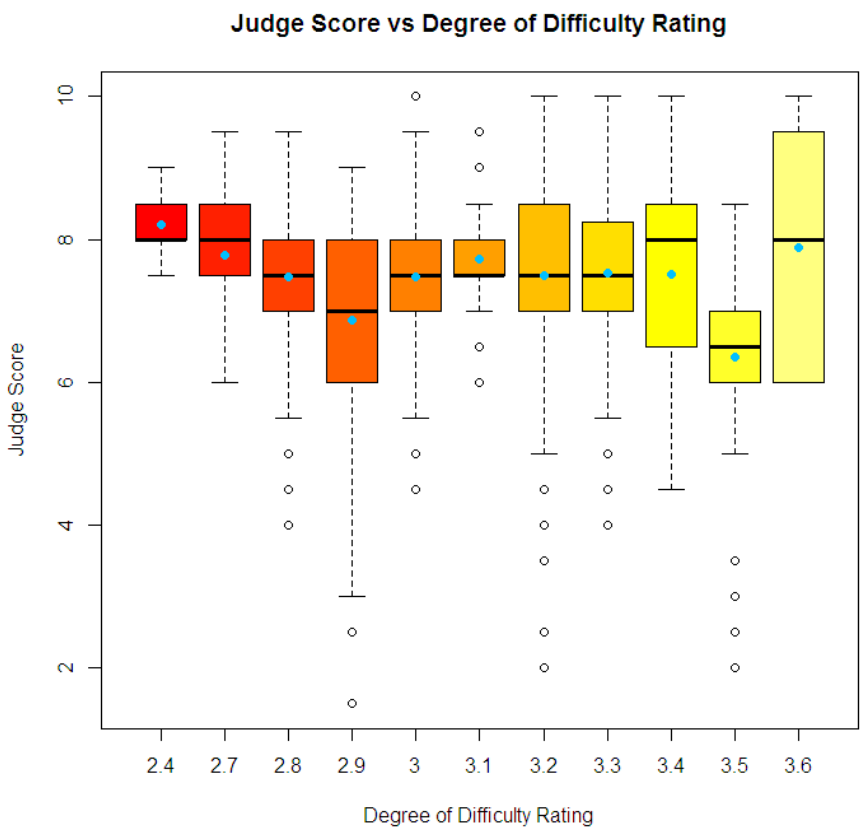


Slope of Green Line: -0.4280 P-Value: 0.000303

The side-by-side figures are expressing the same data in slightly different fashions. The boxplot on the right side has black lines indicating the median values and teal dots indicating the mean values for those dives conducted within each discrete difficulty level. The scatter plot on the left shows each score given by any of the judges in the final round as a function of the dive difficulty. The green regression line has a slight negative slope with a p-value of 0.000303, which means that there is greater than a 99.9% chance that – if and only if we agree with the assumptions and context of this plot – we can reject the null hypothesis and declare a negative association between degree of difficulty and scores.

However, I do not agree with the assumptions and context of this plot: I have three main concerns which I will address on the following pages.

My first concern is addressed in the figure below. There are seven judges who score each dive. These ratings are dependent on the quality of the dive and therefore the fact that 7 data points appear for a single event artificially increases the correlation due to

clustering. I plotted the mean score of these seven judges for each dive against its degree of difficulty. While this did not change the slope of the regression line, it did drastically increase the p-value for this set of data, calling into question the hypothesis of a negative association between degree of difficulty and scores.

A second concern that is worth noting under normal circumstances is that not all divers are created equal; the score is dependent on the quality of the *dive* (mentioned in previous paragraph) and the quality of the *diver*. Not all divers could execute a high-degree-of-difficulty dive at the same level. However, I chose to put aside this concern given that I was analyzing the top few divers in the world, whom I consider to be roughly equivalent in skill level.

Slope of Green Line: -0.4280 P-Value: 0.151

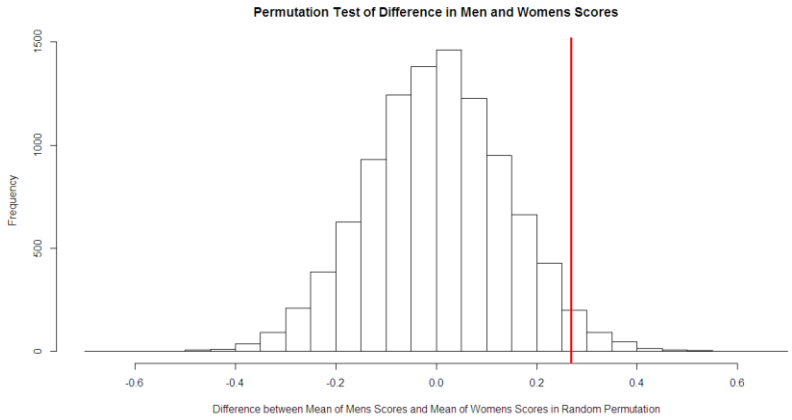
My third main concern is that this data set contains data from 4 distinct subgroups or events: Mens 3 Meter Springboard, Mens 10 meter Platform, Womens 3 Meter Springboard, and Womens 10 Meter Platform. I decided to take a closer look and determine if women and men have significantly different difficulty levels of their dives or significantly different mean scores.

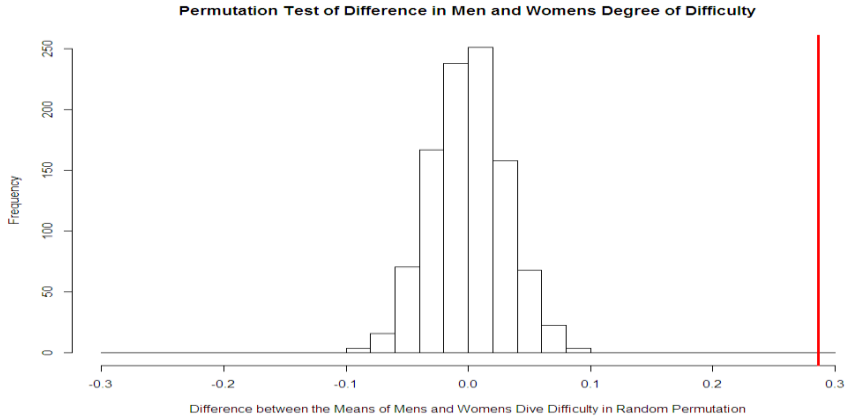
In fact, both of these concerns are well-founded:

Random sampling of the data could not naturally produce subgroups comparable to those shown above. The two values for the means above were consistently outliers on histograms of the means of hundreds of randomly generated subgroups. 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
```

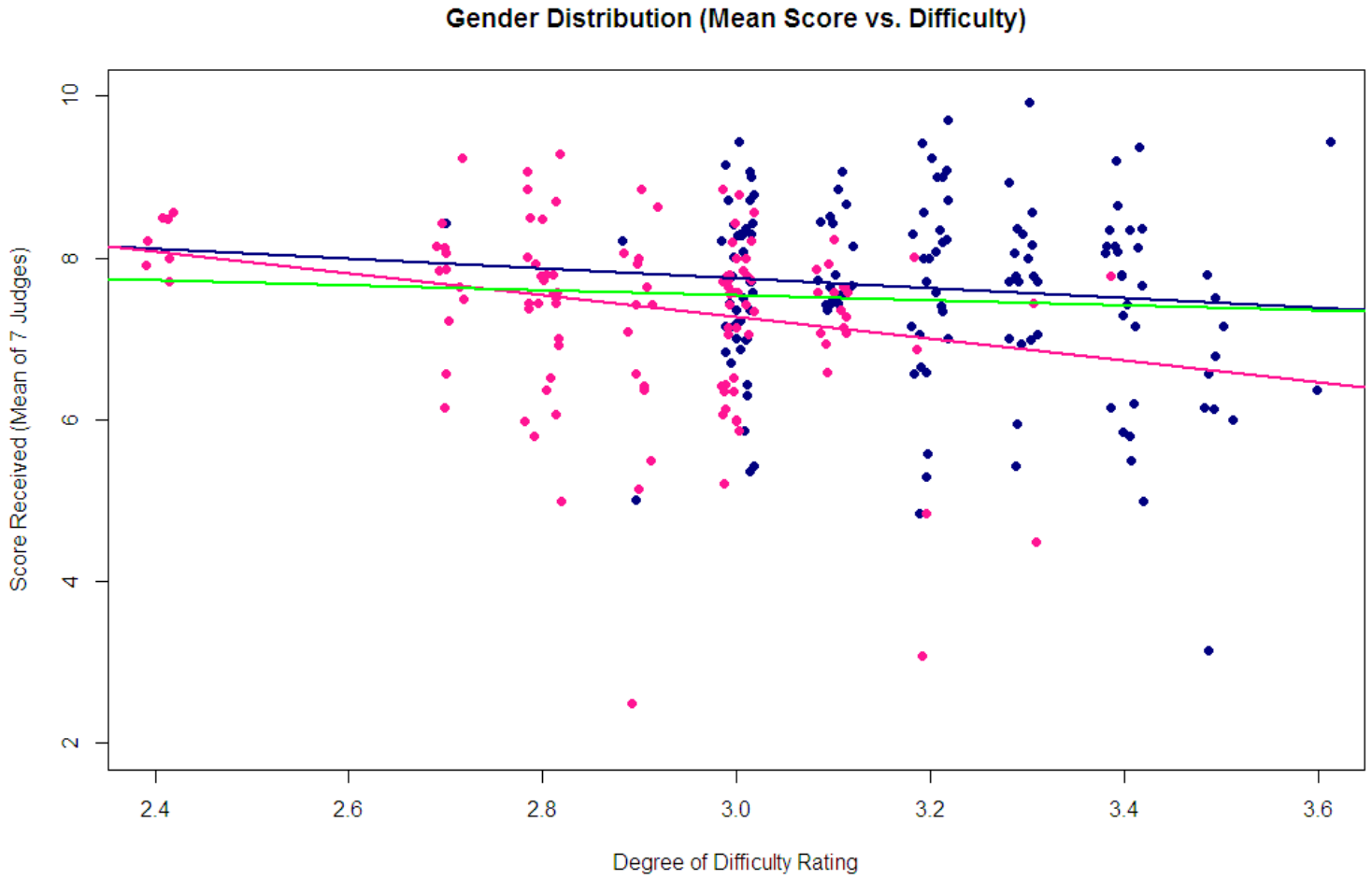




A permutation test was conducted for the differences between these two data sets and it appears that the split in this data was not arbitrary. The p-value for the plot on the left (difference in men and women's scores) is 0.05 and the p-value for the plot on the right (difference in men and women's degree 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.