Final Project Report

Zhenan Hong, Chih-Ming Yi, Wentao Wang

Project Hypothesis

This project is dedicated to investigate the relationship between tennis players’ performance and other factors, such as physical condition and nationality. From the current tennis development trend, we assume players’ physical condition, such as height and weight, have a different degree of impact on men and women players’ performance. To be more specific, men players are more likely to be balance to achieve high performance. On the other hand, it seems the women players are becoming more and more emphasize on power and strength. Our research is intended to explore the diverse requirement of physical condition for men and women players.

Data Collection

There are primarily two datasets for this project, one for men players, and one for women players. The men player dataset, we collected authorized data from the official website of ATP (Association of Tennis Professionals). The women dataset is collected from the official website of WTA (Women's Tennis Association). Both of these dataset are collected at the same time for consistency. For each dataset, it includes player name, ranking, nationality, height, weight, hand holding racket, and accumulated professional points. Each dataset has the sample population of the world top 128 players.

Data Analysis

Based on our hypothesis, we concentrate on the relationship between professional performance and physical condition (height and weight). Thus we adopted some data presentation methods, such as barplot, boxplot, and histogram to observe the data characteristics and data tendency first. Then other methods, including, correlation test, and CHI test were conducted to investigate the inner relationship.

Primary Finding

- For men tennis players

At first, we tried to examine the correlation between player’s performance and physical condition. We found that weight is more likely to affect player’s points.

Then, we draw two boxplots base on players’ weight regarding their points and ranking. We found that players with weight from 81 to 91 are more likely to have higher ranking. And players with weight of 81, 84, 85, 89, 91 earn average higher points.
We noticed that, as the weight approaches 85(kg), the points is rising.

To probe a bit more, we examine the correlation of weight and player’s performance using the following xxx:

```
> cor.test(abs(Weight-85),Points)

Pearson's product-moment correlation

data:  abs(Weight - 85) and Points
t = -2.4431, df = 126, p-value = 0.01595
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:      
-0.3724555  -0.0406397
sample estimates:                    
cor
-0.2126704
```

The p-value is the most significant (p-value=0.01595) when we test the correlation between abs(Weight-85) and Points. Therefore under the confidence level of 95%, men tennis players with weight of 85 (kg) are more likely to achieve brilliant professional achievement.

❖ For women tennis players

Similar methods as for men players were finished to test the relationship between height/weight and professional performance.

From the result, it reveals that both weight and height might be critical and individual factor affecting women tennis players’ performance. Hence, we plot the
Then t-test was implemented to test these two pairs of relationship. The result turned out that there was significant relationship between Height ~ Points, but no significant relationship between Weight ~ Points. We noticed that, as the height approaches 177(cm) and 182(cm), the points is rising. Details as the following:

```r
> cor.test(abs(Height-177),Points)

     Pearson's product-moment correlation

data:  abs(Height - 177) and Points
t = -1.9904, df = 126, p-value = 0.04971
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:  
-0.33789218 0.00198989
sample estimates:
cor
-0.1745971

> cor.test(abs(Height-182),Points)

     Pearson's product-moment correlation

data:  abs(Height - 182) and Points
t = -2.615, df = 126, p-value = 0.009927
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:  
-0.38547430 0.0580648
sample estimates:
cor
-0.2271374
```

**Result Analysis**

From the result, it is obvious that men and women tennis players have different primary influential factors. For men players, with weight close to 85(kg), on the contrary for women players, with height close to 177(cm) or 182(cm) increase their likelihood of better professional performance respectively.

**Potential Bias**

There is probability of bias in the analysis, because both women players and men players have a big Points gap between top players and other players in the back. For example, only the top two men players accumulate professional points above ten thousand. The player ranking at the fifth gets only less than five thousand points, only one third of points the first player got.

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Name</th>
<th>Nationality</th>
<th>Height</th>
<th>Weight</th>
<th>Plays</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Rafael Nadal</td>
<td>ESP</td>
<td>185</td>
<td>85</td>
<td>Left-handed</td>
<td>14360</td>
</tr>
<tr>
<td>2</td>
<td>Roger Federer</td>
<td>SUI</td>
<td>185</td>
<td>85</td>
<td>Right-handed</td>
<td>11000</td>
</tr>
<tr>
<td>3</td>
<td>Novak Djokovic</td>
<td>SRB</td>
<td>187</td>
<td>80</td>
<td>Right-handed</td>
<td>9420</td>
</tr>
<tr>
<td>4</td>
<td>Andy Murray</td>
<td>GBR</td>
<td>190</td>
<td>84</td>
<td>Right-handed</td>
<td>7400</td>
</tr>
<tr>
<td>5</td>
<td>Nikolay Davydenko</td>
<td>RUS</td>
<td>177</td>
<td>70</td>
<td>Right-handed</td>
<td>4955</td>
</tr>
<tr>
<td>6</td>
<td>Andy Roddick</td>
<td>USA</td>
<td>187</td>
<td>89</td>
<td>Right-handed</td>
<td>4440</td>
</tr>
<tr>
<td>7</td>
<td>Juan Martin Del Potro</td>
<td>ARG</td>
<td>198</td>
<td>83</td>
<td>Right-handed</td>
<td>4290</td>
</tr>
<tr>
<td>8</td>
<td>Gilles Simon</td>
<td>FRA</td>
<td>180</td>
<td>69</td>
<td>Right-handed</td>
<td>3925</td>
</tr>
<tr>
<td>9</td>
<td>Gael Monfils</td>
<td>FRA</td>
<td>193</td>
<td>80</td>
<td>Right-handed</td>
<td>3550</td>
</tr>
<tr>
<td>10</td>
<td>Fernando Verdasco</td>
<td>ESP</td>
<td>187</td>
<td>81</td>
<td>Left-handed</td>
<td>3330</td>
</tr>
</tbody>
</table>

This will cause the top four players set a phenomenon influence on the final testing result. This is pretty obvious in our findings on men tennis players, since the top two men player are both 85(kg) in weight. In dealing with this potential bias, we also tried to test relationship between log\(^\text{Points}\) and height/weight. Based on this method, log function can shrink the gap between top players and players in the back. The result is similarly significant, even though the significance is in different scale.

**About This Report**

Please feel free to use the data in future versions of the course and post final report on the public class webpage.