**Analysis on the Advantage of Height and Play Handedness in Tennis**

***Kah Hin Chin, Nicholas Wong Wai Hong, Judyanne Sharmini***

[*khchi17@student.monash.edu*](mailto:khchi17@student.monash.edu)*,* [*nicholas.wong1@monash.edu*](mailto:nicholas.wong1@monash.edu)*,* [*jsgil5@student.monash.edu*](mailto:jsgil5@student.monash.edu)

*Faculty of Information Technology, Monash University, Sunway Campus, Malaysia*

**INTRODUCTION**

Sports are competitive physical activities with a set of rules to provide participants a fair chance of winning. If a particular physique or play style can affect these chances of winning, then it would skew the fairness of the competition and defeat its purpose. This study analyzes advantages in terms of ranking and points gained from height and play hand of both male and female tennis players. This investigation sets out to test two hypotheses; height is an advantage and being left-handed is an advantage.

**METHODS**

**Data Extraction and Cleaning**

Data for this investigation was retrieved from two online databases belonging to the Association of Tennis Professionals (ATP) and Women’s Tennis Association (WTA), which are governing bodies for professional male and female players.

A Python web crawler was implemented to extract information on player details and ranking by crawling the official websites of these two professional bodies and the raw data that was extracted was then stored in Excel spreadsheets. The Python crawler generated two separate singles match play ranking spreadsheets which also included player details for men and women. The raw data was cleaned and consolidated into a single file for analysis. Data was cleaned to ensure that handedness entries only stated “left” or “right” without the style of play such as “two-handed backhand”, “two-handed both sides” etc. The consolidated spreadsheet contained the details of 3392 players, 2062 male and 1330 female.

**Data for Ranking and Height Hypothesis Testing**

Details on height were available for about two thirds, 65% (2208 records) of the players in the dataset. Height for more than half of the male players (53%) and a small number of the female players (7%) were not available. The statistics on complete and incomplete records are available in Table 1. A total of 1994 records will be excluded for the height analysis.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Gender** | **Height Information Available** | | **Height Information Not Available** | |
| **N** | **Percentage (%)** | **N** | **Percentage (%)** |
| **Male** | 968 | 47 | 1094 | 53 |
| **Female** | 1240 | 93 | 90 | 7 |

**Table 1**: Player records that do not contain height information

**Data for Ranking and Handedness Hypothesis Testing**

Details on handedness were only available for 1607 (47%) players in the dataset. This is less than half of the original number of records that were extracted. Handedness information for most of the female players (78%) was not available. The statistics on complete and incomplete records are available in Table 2. However, the analysis will be carried out on the remaining 1607 player records.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Gender** | **Handedness Information Available** | | **Handedness Information Not Available** | |
| **N** | **Percentage (%)** | **N** | **Percentage (%)** |
| **Male** | 1320 | 64 | 742 | 36 |
| **Female** | 287 | 22 | 1043 | 78 |

**Table 2**: Player records that do not contain handedness information

The height of players based on gender is tabulated in Table 3 below. The percentage values do not consider records that were excluded from the height analysis. The dataset used for analysis contains 968 male player records and 1240 female player records. It is interesting to note that there are exactly the same number of players in each gender group for the last three categories; 191-200, 201-210 and 211-220.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Gender** | **Height (cm)** | | | |  | | | |
| **151-160** | **161-170** | **171-180** | **181-190** | | **191-200** | **201-210** | **211-220** |
| **Male** | 1 | 22 | 278 | 433 | | 223 | 9 | 2 |
| **Female** | 5 | 127 | 405 | 469 | | 223 | 9 | 2 |

**Table 3**: Height of professional tennis players according to gender

The handedness of players based on gender is tabulated in Table 4 below. The percentage values do not consider records that were excluded from the handedness analysis. The dataset used for analysis contains 1320 male player records and 287 female player records. Only 13% of the players are left-handed.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Gender** | **Left-handed (LH)** | | **Right-handed (RH)** | |
| **N** | **Percentage (%)** | **N** | **Percentage (%)** |
| **Male** | 177 | 11 | 1143 | 71 |
| **Female** | 25 | 2 | 262 | 16 |

**Table 4**: Handedness of professional tennis players according to gender

Player ranking ranged from 1 (top player) to 1981. There were duplicate ranks in the table because more than one player could have the same ranking. Due to the large range of the ranking information, ranking was grouped into categories. Four quartiles of ranks were created; 1-500, 501-1000, 1001-1500 and 1501-2000.

**Statistical Analysis**

**Chi-square for Ranking and Height Hypothesis Testing**

The association between variables in this study; height and player ranking was tested using chi-square test. The first step of conducting a chi-square test is to create two extreme hypotheses. The hypotheses for this experiment are:

H0 : Ranking is influenced by height.

H1 : Ranking is **not** influenced by height.

H0 represents the null hypothesis which assumes that the two variables are not related while H1 which is the alternative hypothesis which assumes that a relationship between the two variables exist, thus contradicting the null hypothesis.

The next step of this test is to build a Rows x Columns contingency table. For this study, we would be using the 7 x 4 contingency table, creating a table with twenty-eight cells (please refer to highlighted section of Table 5). The values in the highlighted area indicate the ranking of players for each height category.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Height** | **Ranking Categories** | | | |
| **0-500** | **501-1000** | **1001-1500** | **1501-2000** |
| **151-160** | 6 | 0 | 0 | 0 |
| **161-170** | 149 | 0 | 0 | 0 |
| **171-180** | 345 | 338 | 0 | 0 |
| **181-190** | 568 | 334 | 0 | 0 |
| **191-200** | 0 | 198 | 166 | 82 |
| **201-210** | 0 | 0 | 0 | 18 |
| **211-220** | 0 | 0 | 0 | 4 |

**Table 5**: Player ranking based on height

After the contingency table is created, the chi-square (χ2) score needs to be computed. This score is calculated by summing squared values of the discrepancies between the observed and expected frequencies of each cell in the contingency table. The observed (O) frequencies are provided in the contingency table, but the expected (E) frequencies need to be calculated. This can be done using the formula provided below:



The following table (Table 4) provides the observed and expected frequencies for each cell for the contingency table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Height** | **Ranking Categories** | | | |
| **0-500** | **501-1000** | **1001-1500** | **1501-2000** |
| **151-160** | 6 **(O)** | 0 **(O)** | 0 **(O)** | 0 **(O)** |
| 2.902174 **(E)** | 2.36413 **(E)** | 0.451087 **(E)** | 0.282609 **(E)** |
| **161-170** | 149 **(O)** | 0 **(O)** | 0 **(O)** | 0 **(O)** |
| 72.07065 **(E)** | 58.70924 **(E)** | 11.20199 **(E)** | 7.018116 **(E)** |
| **171-180** | 345 **(O)** | 338 **(O)** | 0 **(O)** | 0 **(O)** |
| 330.3641 **(E)** | 269.1168 **(E)** | 51.34873 **(E)** | 32.17029 **(E)** |
| **181-190** | 568 **(O)** | 334 **(O)** | 0 **(O)** | 0 **(O)** |
| 436.2935 **(E)** | 355.4076 **(E)** | 67.81341 **(E)** | 42.48551 **(E)** |
| **191-200** | 0 **(O)** | 198 **(O)** | 166 **(O)** | 82 **(O)** |
| 215.7283 **(E)** | 175.7337 **(E)** | 33.5308 **(E)** | 21.00725 **(E)** |
| **201-210** | 0 **(O)** | 0 **(O)** | 0 **(O)** | 18 **(O)** |
| 8.706522 **(E)** | 7.092391 **(E)** | 1.353261 **(E)** | 0.847826 **(E)** |
| **211-220** | 0 **(O)** | 0 **(O)** | 0 **(O)** | 4 **(O)** |
| 1.934783 **(E)** | 1.576087 **(E)** | 0.300725 **(E)** | 0.188406 **(E)** |

**Table 6:** Expected **(E)** and Observed **(O)** frequencies for ranking based on hieght

Using the values in Table 6, the formula to calculate chi-square (χ2) score is as follows:

χ2 =

Applying the abovementioned formula to each observed and expected frequency pair would result in values as depicted in Table 7 below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Height** | **Ranking Categories** | | | |
| **0-500** | **501-1000** | **1001-1500** | **1501-2000** |
| **151-160** | 3.306668 | 2.36413 | 0.451087 | 0.282609 |
| **161-170** | 82.1156 | 58.70924 | 11.20199 | 7.018116 |
| **171-180** | 0.648402 | 17.63133 | 51.34873 | 32.17029 |
| **181-190** | 39.75904 | 1.289465 | 67.81341 | 42.48551 |
| **191-200** | 215.7283 | 2.821248 | 523.3425 | 177.0873 |
| **201-210** | 8.706522 | 7.092391 | 1.353261 | 347.0017 |
| **211-220** | 1.934783 | 1.576087 | 0.300725 | 77.11148 |

**Table 7:** Chi-square (χ2) values for ranking based on handedness

The next step of this method is to determine the degree of freedom (*df*) which is an indicator of how many values within the chi-square grid are independent. The formula to calculate the degree of freedom is:



For this chi-square grid, *df =* 18 and the chi-square (χ2) score can be computed by adding all the values in the chi-square grid in Table 7, resulting in χ2 = 1782.65. Providing both these values; *df =* 18 and χ2 = 1782.65 to the CHIDIST function in Microsoft Excel would return a value of 0. Microsoft Excel 2010 was used to conduct the chi-square test outlined in this section.

**Chi-square for Ranking and Handedness Hypothesis Testing**

The association between variables in this study; handedness and player ranking was tested using chi-square test. The first step of conducting a chi-square test is to create two extreme hypotheses. The hypotheses for this experiment are:

H0 : Ranking is influenced by handedness.

H1 : Ranking is **not** influenced by handedness.

H0 represents the null hypothesis which assumes that the two variables are not related while H1 which is the alternative hypothesis which assumes that a relationship between the two variables exist, thus contradicting the null hypothesis.

The next step of this test is to build a Rows x Columns contingency table. For this study, we would be using the 2 x 4 contingency table, creating a table with eight cells (please refer to highlighted section of Table 8). The values in the highlighted area indicate the number of players who are left-handed and right-handed for each ranking category.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Handedness** | **Ranking Categories** | | | |
| **0-500** | **501-1000** | **1001-1500** | **1501-2000** |
| **Left-handed (LH)** | 246 | 36 | 5 | 0 |
| **Right-handed (RH)** | 437 | 364 | 278 | 241 |

**Table 8**: Player ranking based on handedness

After the contingency table is created, the chi-square (χ2) score needs to be computed. This score is calculated by summing squared values of the discrepancies between the observed and expected frequencies of each cell in the contingency table. The observed (O) frequencies are provided in the contingency table, but the expected (E) frequencies need to be calculated. This can be done using the formula provided below:



The following table (Table 9) provides the observed and expected frequencies for each cell for the contingency table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Handedness** | **Ranking Categories** | | | |
| **0-500** | **501-1000** | **1001-1500** | **1501-2000** |
| **Left-handed (LH)** | 246 **(O)** | 36 **(O)** | 5 **(O)** | 0 **(O)** |
| 121.9795 **(E)** | 71.43746 **(E)** | 50.542 **(E)** | 43.04107 **(E)** |
| **Right-handed (RH)** | 437 **(O)** | 364 **(O)** | 278 **(O)** | 241 **(O)** |
| 561.0205 **(E)** | 328.5625 **(E)** | 232.458 **(E)** | 197.9589 **(E)** |

**Table 9:** Expected **(E)** and Observed **(O)** frequencies for ranking based on handedness

Using the values in Table 9, the formula to calculate chi-square (χ2) score is as follows:

χ2 =

Applying the abovementioned formula to each observed and expected frequency pair would result in values as depicted in Table 10 below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Handedness** | **Ranking Categories** | | | |
| **0-500** | **501-1000** | **1001-1500** | **1501-2000** |
| **Left-handed (LH)** | 126.0958 | 17.5792 | 41.03664 | 43.04107 |
| **Right-handed (RH)** | 27.41627 | 3.822145 | 8.922361 | 9.358172 |

**Table 10:** Chi-square (χ2) values for ranking based on handedness

The next step of this method is to determine the degree of freedom (*df*) which is an indicator of how many values within the chi-square grid are independent. The formula to calculate the degree of freedom is:



For this chi-square grid, *df =* 3 and the chi-square (χ2) score can be computed by adding all the values in the chi-square grid in Table 10, resulting in χ2 = 277.27. Providing both these values; *df =* 3 and χ2 = 277.27 to the CHIDIST function in Microsoft Excel would return a value of 8.24 x 10-60. Microsoft Excel 2010 was used to conduct the chi-square test outlined in this section.

**Scatterplot for Ranking and Height Hypothesis Testing**

The R software was used to create charts that could be used to visualize associations between height and ranking. Three scatterplots were created, one for Ranking versus Height (Both Gender) (Fig 1a), Ranking versus Height (Female) (Fig 1b) and Ranking versus Height (Male) (Fig 1c).

**Boxplot for Ranking and Handedness Hypothesis Testing**

The R software was used to create charts that could be used to visualize associations between handedness and ranking. Three boxplots were created, one for Ranking versus Play (Both Gender) (Fig 2a), Ranking versus Play (Female) (Fig 2b) and Ranking versus Play (Male) (Fig 2c).

**Pearson, Spearman and Kendall’s Correlation Test for Ranking and Handedness Hypothesis Testing**

Lastly, correlation testing was done on the investigated variables; (1) ranking and handedness and (2) ranking and height. Parametric and non-parametric correlation tests were carried out. Pearson’s correlation test is parametric while Kendall and Spearman are non-parametric.

**RESULTS AND DISCUSSION**

**Data**

Quite a number of the records from the dataset had to be excluded because the required information (eg. height or handedness) was missing. This reduced the dataset significantly. The effect of the missing data on findings is unclear.

**Statistical Analysis**

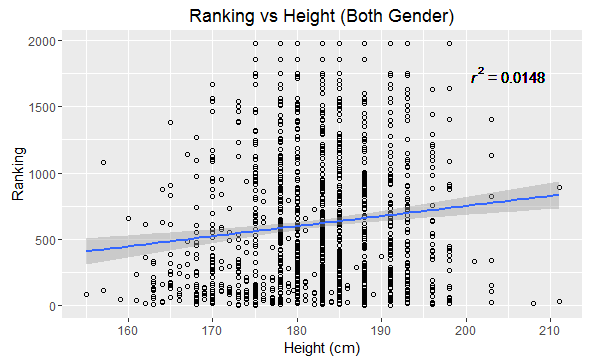
**Chi-square for Ranking and Handedness Hypothesis Testing**

The objective of this experiment was to determine if a correlation exists between the variables in this study; ranking and height. The results of the chi-square analysis returned a value of 0. This means that the null hypothesis, H0 can be rejected because the possibility of observing a result as extreme as this is very low, hence it can be concluded that ranking **is not** influenced by height.

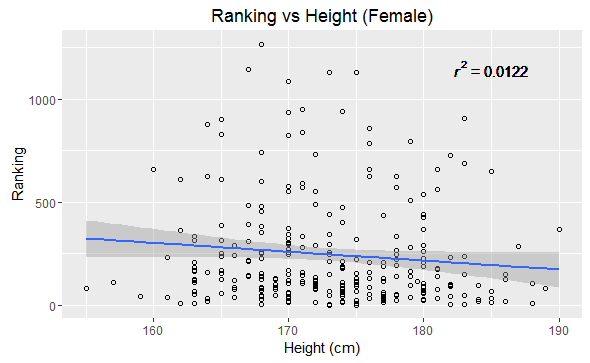
**Chi-square for Ranking and Handedness Hypothesis Testing**

The objective of this experiment was to determine if a correlation exists between the variables in this study; ranking and handedness. The results of the chi-square analysis returned a value of 8.24 x 10-60. This means that the null hypothesis, H0 can be rejected because the possibility of observing a result as extreme as this is very low, hence it can be concluded that ranking **is not** influenced by handedness.

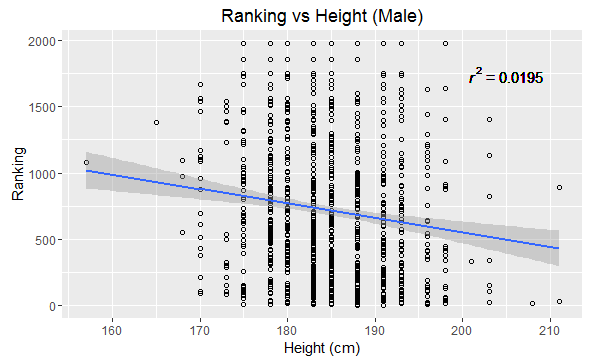
**Scatterplot for Ranking and Height Hypothesis Testing**



(a)



(b)



(c)

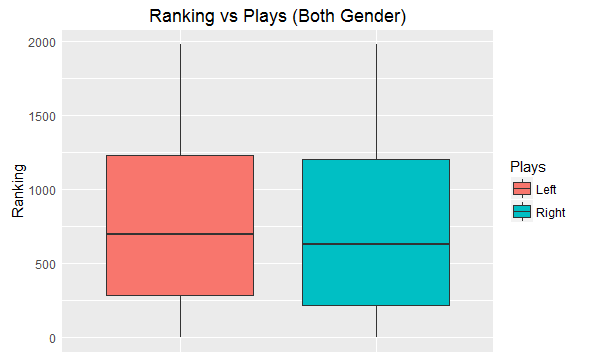
**Figure 1:** Ranking versus Height Scatterplots

(a) Ranking versus Height (Both Gender) (b) Ranking versus Height (Female)

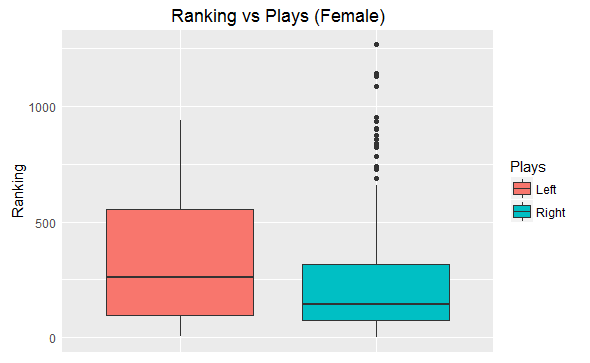
(c) Ranking versus Height (Male)

The scatterplots do not indicate clear correlations between height and ranking for all three plots. This further confirms the results of the chi-square analysis.

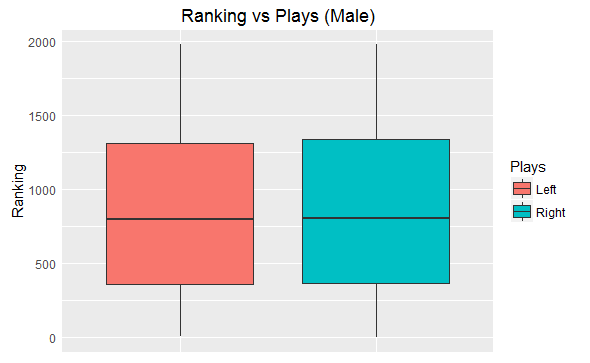
**Boxplot for Ranking and Handedness Hypothesis Testing**



(a)



(b)



(c)

**Figure 2:** Ranking versus Handedness Plots

(a) Ranking versus Play (Both Gender)

(b) Ranking versus Play (Female)

(c) Ranking versus Play (Male)

The boxplots do not indicate clear correlations between handedness and ranking for all three plots. This further confirms the results of the chi-square analysis.

**Pearson, Spearman and Kendall’s Correlation Test for Ranking and Handedness Hypothesis Testing**

The results of these three correlation tests are presented in Table 6 below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Correlation Test** | **ρ-value** | **Finding** | **Coefficient value** | **Finding** |
| **Pearson** | 0.1787 | Not statistically significant | cor = 0.1214988 | Low correlation |
| **Spearman** | 0.2389 | Not statistically significant | rho = 0.1196733 | Low correlation |
| **Kendall** | 0.3382 | Not statistically significant | tau = 0.08156836 | Low correlation |

**Table 6:** Results of parametric and non-parametric correlation tests

Pearson, Spearman and Kendall correlations all indicate that the observation **is not** statistically significant.

**Other Related Studies**

The data used in this study are on professional tennis players only. While left-handedness was no longer associated with higher probability of attaining high year-end world ranking position in male professionals, there may still be an advantage among female amateur players [3]. Although this study does not support height advantage, it is possible that there are both advantages and disadvantages to being short or tall, ironing out the average in this analysis. Another study on gender differences suggests that height and body mass index may affect performance [4].

**CONCLUSION**

A battery of statistical tests carried out in this investigation indicate that ranking and play handedness are not correlated, thus proving that being left-handed is not an advantage in tennis. Similarly, ranking and height were also found not to be correlated.

**REFERENCES**

1. Women's Tennis Association. *WTA Singles Rankings & Player Details* <http://www.wtatennis.com/singles-rankings> (Accessed 27/04/2016)
2. Association of Tennis Professionals. *ATP Single Rankings & Player Details* http://www.atpworldtour.com/en/rankings/singles (Accessed 27/04/2016)
3. Loffing, F., Hagemann, N., & Strauss, B. (2012). Left-handedness in professional and amateur tennis.*PLoS ONE, 7*(11) doi:10.1371/journal.pone.0049325
4. Krumer, A., Rosenboim, M., & Shapir, O. M. (2016). Gender, competitiveness, and physical characteristics: Evidence from professional tennis.*Journal of Sports Economics, 17*(3), 234-259. doi:10.1177/1527002514528516