

## **Relationship Between Finish Times and Starting Order for Athletes in the Vancouver Women's Downhill Skiing Olympics from 2010-2022**

**Table 1 : Summary Statistics and Relationship of Downhill Skiing Finish Times and Starting Orders**

<b>Year</b>	<b>N</b>	<b>Average Finish Times</b>	<b>Median Finish Times</b>	<b>Range of Finish Times</b>	<b><math>R^2</math></b>	<b>Regression Coefficient</b>	<b>Correlation</b>
2010	37	110.1 (5.2)	108.1	25.08	0.159	0.159	0.425
2014	35	103.8 (2.1)	103.2	9.4	0.42	0.109	0.648
2018	31	102.1 (2.1)	101.2	7.58	0.703	0.147	0.839
2022	31	95 (2.8)	94	12.66	0.405	0.164	0.647

*Note 1: N represents the number of observations for each year, where each unit of observation is an athlete. The Average, Median and Range of the Finish Times for the skiers are measured in terms of seconds. The values in the brackets represent the standard deviations from the mean for each of the years, in terms of seconds. The coefficient of determination ( $R^2$ ) estimates the amount of variation in finish times that is explained by the variation in starting order. The regression coefficient represents the slope of the regression line where the predictor variable is the starting order for athletes and the dependent variable is the finish time for the athletes. The correlation( $r$ ) measures the relationship between finish times and starting orders for each year.*

Our table displays data collected on the Women's Vancouver Olympics from 2010 to 2022 for downhill skiing. We can see that average finish times for skiers have been on the decline over the years, dropping from an average of 110.1 seconds in 2010 to 95 seconds in 2022. This indicates skiers in the Vancouver Women's Olympics have gotten 13.7% faster over the years. The standard deviation in finish times have also been declining over the years, suggesting reduced variance in skiers' recorded finish times, with respect to the average finish time for the respective years. Further, for every year, the average finish times are higher than the median finish times, insinuating a right skew in the data. That is, majority of the skiers have short finish times, with a few skiers trailing behind with comparatively longer finish times. Skiers also appear to be performing better over the years as the difference between the highest and lowest finish times have been gradually decreasing from 2010 to 2018. However, the range is greater in 2022 relative to previous years, due to there being a large difference between the fastest skier's time and the slowest skier's time.

Unsurprisingly, there is a positive correlation between the start order and finish times for each year of the Women's Vancouver Downhill Skiing Olympics. This positive relationship appears to be strongest in the years after 2010. Skiers that have a start order one standard deviation higher on average in 2022, have a finish time 0.65 standard deviations higher in the same year. Further, the regression coefficient is positive for every year, which means as the starting orders increase, the finish times for each skier gets consequently longer. That is, a starting position that is further back acts as a disadvantage, generally resulting in a longer time to finish the race. More precisely, in 2022, for every one unit increase in the start order, the finish times were on average, 0.16 percentage points higher. The respective  $R^2$  value for 2022, however, is quite low, indicating only 4.1% of the variation in skiers' finish times is explained by the variation in their respective starting orders.

### Replication Steps:

1. Open **skiing\_2022.xlsx**, **skiing\_2018.xlsx**, **skiing\_2014.xlsx** and **skiing\_2010.xlsx**.
2. Extract the following variables from each excel dataset: *fis\_code*, *lname*, *fname*, *country\_code*, *time\_seconds* and *start\_order*.
3. Create a new excel dataset and add all the selected variables for every year into a worksheet.
4. Create separate worksheets for the data for each year, naming each worksheet after the respective years (e.g., **2010 data**).
5. Copy the *start\_order* and *time\_seconds* variables for the year 2010 from the main worksheet in that specific order and paste it to the respective 2010 data worksheet.
6. Click Data Analysis on the Data tab. Select Descriptive Statistics.
7. Set the Input Range as the entire column for the *time\_seconds* variable (e.g., **\$B:\$B**). Check the boxes for Labels in first row and Summary Statistics. Set the Output Range as any column beside the two variable columns (e.g., **\$D\$1**).
8. This gives us a descriptive statistics table, from which we can extract the values for Mean (**Average Finish Times**), Median (**Median Finish Times**), Range (**Range of Finish Times**) and Count (N).
9. Next, in the Data tab, click on Data Analysis. Select Correlation.
10. Set the Input Range as the entire columns of *start\_order* and *time\_seconds* (e.g., **\$A:\$B**). Check the boxes for Grouped By: Columns and Labels. Set the Output Range as any column beside the two variable columns and the descriptive statistics table (e.g., **\$D\$20**).
11. This gives us the correlation matrix, from which we can extract the value for the Correlation between the starting orders and finish times (**Correlation**).
12. Next, in Data tab, click on Data Analysis. Select Regression.
13. Select the Input Y Range as the entire column of *time\_seconds*, including the variable name (e.g., **\$B\$1:\$B\$38**) and the Input X Range as the entire column of *start\_order*, including the variable name (e.g., **\$A\$1:\$A\$38**). Check the box for Labels. Select the Output Range as any column beside the two variable columns, the descriptive statistics table, and the correlation matrix (e.g., **\$D\$30**).
14. This gives us the summary output, from which we can extract the value for the correlation of determination ( $R^2$ ). We also get the ANOVA table, which gives us the slope of the regression line (**Regression Coefficient**).
15. Repeat steps 5 to 14 for the rest of the years, i.e., 2014, 2018 and 2022, to obtain all the necessary values for our summary statistics table.

### Tables for Inspiration:

1. April 2022 Exam, Questions (1) and (6): “Table 1 – Summary Statistics” from Aydin (2022) and “Table 3. Descriptive Statistics” from Bachmann et al. (2021).
2. April 2013 Exam, Questions (20) – (26) “Table 1: Descriptive statistics for variables used in analysis (World Values Survey, 1990, 2001, 2007)” from Steele and Lynch (2013).

