Data analysis

The merged dataset is the *times_shanghai_mergedRankings.xlsx* file created by merging data from shanghaiData.csv with the focal dataset, timesData.csv. The detailed step-by-step procedure of how this file was created can be found on the URL https://github.com/aditi-sharma/INFM-600---Information-Discovery-and-Analysis in the 'Processing Documentation.pdf ' file

This dataset can be used to answer following research question and hypothesis:

1. Determine if there is a correlation between the world rank of a university and its award and citations score

2. Analyze the hypothesis that higher the amount of international students in a university, higher would be the world rank of the university.

The results for the above question and hypothesis have been answered using the tool SPSS.

Answer to Question 1.

We perform correlation operation in SPSS for the three variables world_rank, award, citations. To do this, we opened the *times_shanghai_mergedRankings.xlsx* file in SPSS. We performed the correlation analysis of the three variables world_rank, award, and citations. For this go to the menu bar, Click Analyze -> Correlate ->Bivariate. Select the variables world_rank, award, citations, and click OK, as shown in the dialog box of Fig 1.0

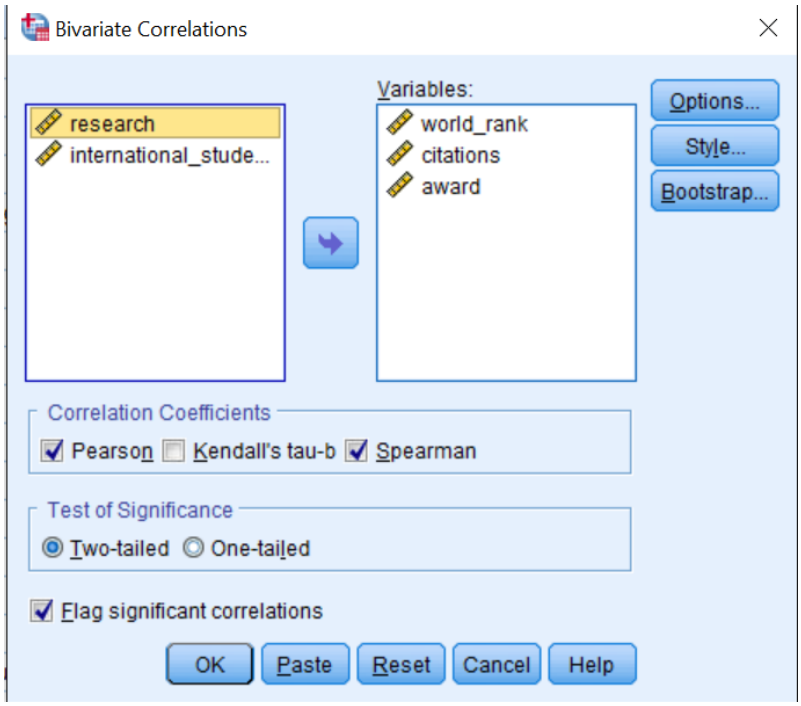

Fig 1.0

The output of the correlation analysis is shown in Table.1.0. The absolute value of the correlation co-efficient is > 0.5 between the three variables. (GraphPad Statistics Guide) Hence, we can conclude from the table that the world rank is highly correlated to the award and citation score of a university.

Correlations

			world_rank	citations	award
Spearman's rho	world_rank	Correlation Coefficient	1.000	-.559**	-1.000**
		Sig. (2-tailed)	.	.000	.000
		N	100	100	100
	citations	Correlation Coefficient	-.559**	1.000	.557**
		Sig. (2-tailed)	.000	.	.000
		N	100	100	100
	award	Correlation Coefficient	-1.000**	.557**	1.000
		Sig. (2-tailed)	.000	.000	.
		N	100	100	100

**. Correlation is significant at the 0.01 level (2-tailed).

Table.1.0

Although the correlation between ‘world rank and award’, ‘world rank and citations’ is negative; it is statistically significant. Awards and citations are highly positively correlated, meaning higher the number of citations, higher is the number of awards that the university bags.

The visual representation of the relation between world rank versus awards and citations can be seen in Fig.1.1 below

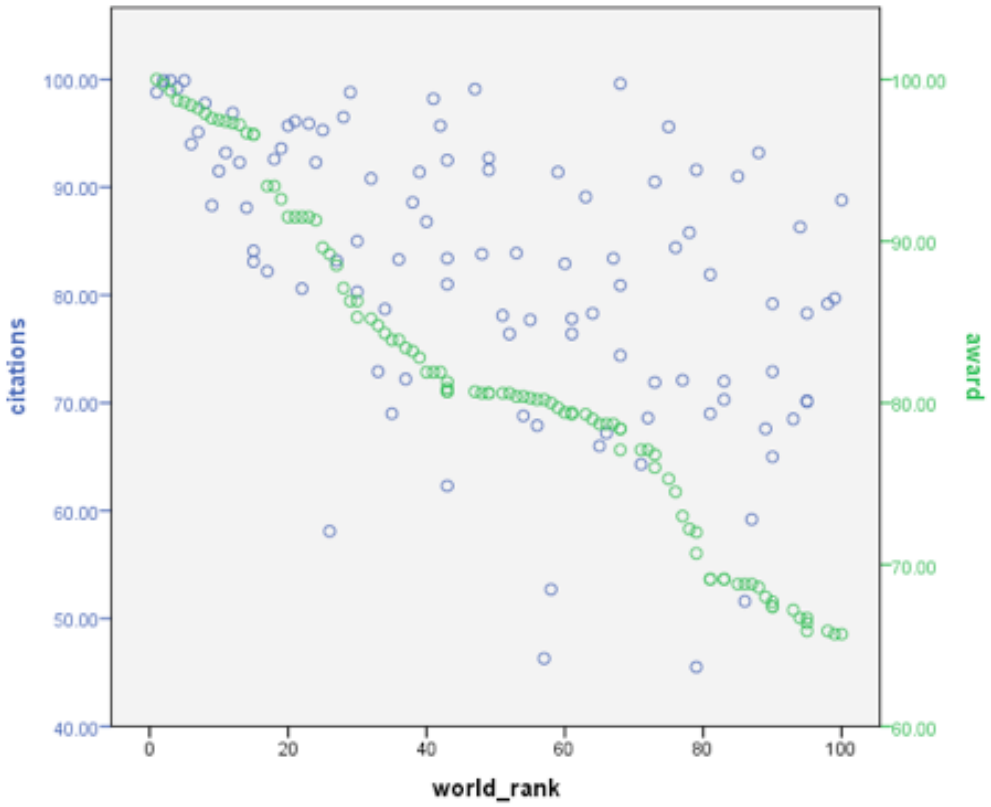

Fig.1.1

Answer to Question 2.

We perform correlation between world_rank and international students to determine if they are significantly correlated. If a positive correlation exists, this will mean the hypothesis will be proved. Alternatively, if a negative correlation exists, or no correlation exists at all; it will disprove the hypothesis.

Again, we use SPSS to test this hypothesis. Open the *times_shanghai_mergedRankings.xlsx* file in SPSS. Click on Analyze in the menu bar, go to Correlate -> select Bivariate. Select the two variables – world_rank, and international_students. Click ok.

From the Table 1.2, the absolute value of the correlation coefficient between the world rank and the international students is less than 0.2, we can establish that they are slightly correlated, and the results are not statistically significant.

Correlations

			world_rank	international_students
Spearman's rho	world_rank	Correlation Coefficient	1.000	-.256*
		Sig. (2-tailed)	.	.011
		N	100	97
	international_students	Correlation Coefficient	-.256*	1.000
		Sig. (2-tailed)	.011	.
		N	97	97

Table 1.2

The graph in Fig 1.2 below shows that the world rank is not a linear curve with respect to the percentage of international students. This disproves the hypothesis mentioned in question 2.

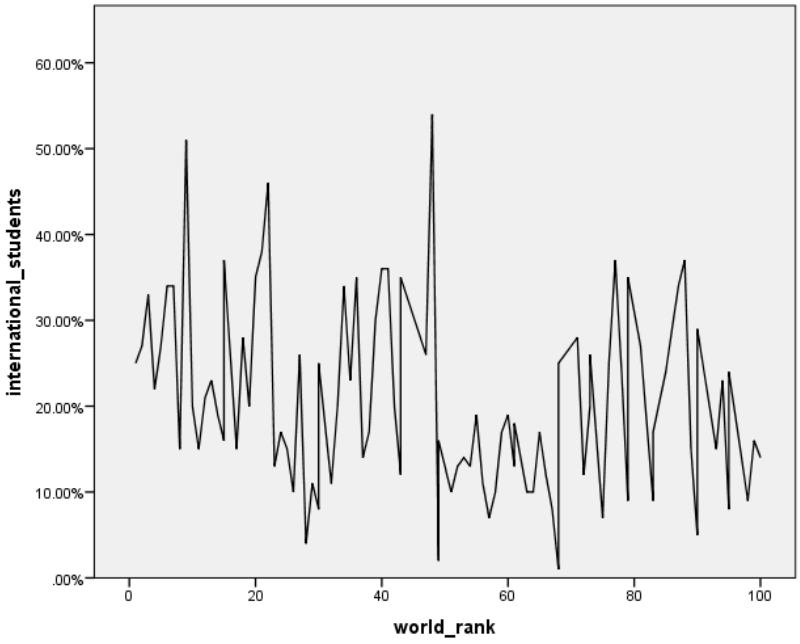

Fig 1.2