Olympinomics!

An Executive Summary by Viet Dang and Taylor Womack

Background: We procured Olympics cost and revenue data and performed statistical inference on it using multiple linear regression, Spearman's Rank Correlation test, Kendall Tau Rank Correlation test, and Pearson's Rank Correlation tests to determine which variables are best at predicting the percentage of tickets that will be sold at an olympic games (TP) and if any variables were correlated with TP. We then used multiple linear regression to create a model to predict TP and tried to simulate TP using the exponential distribution.

Methods: Using R, we fitted full models to predict TP; we also created box plots and other graphics and ran correlation tests to see if there were relationships. We also fitted models with interaction terms and used model selection techniques (looking at R-squared, viewing graphs for conditions, comparing AIC) to determine a final model. The other variables we looked at were Season (Winter/Summer), Year, Athletes (Number of athletes competing), Events (number of events hosted), Countries (number of countries participating), Media (number of accredited media organizations reporting/present), total revenue in USD (TRUSD), broadcast revenue (BRUSD), international sponsor revenue in USD (ISRUSD), domestic sponsor revenue in USD (DSRUSD), venue cost in USD (VCUSD), organization cost in USD (OCUSD), and change in ticket percentage from one year to another (TPX).

ARIMA time series predictions were used to predict ticketing and sponsorship revenues for the upcoming Olympics. Dickey-Fuller tests resulted in high p-values, suggesting non-stationary data. Removing outliers eliminated the issue and resulted in predictions of USD 335.8 million for ticketing sales and USD 752.0 million for sponsorship revenues.

Results: There is not a statistically significant correlation between TP and any variable but BRUSD, with a Pearson's p-value of 0.05929. The p-values for all the other correlation tests with the other variables were not significant at alpha=.1. Our multivariate model used Media (p=0.0638), TRUSD (p=0.0784), BRUSD (p=0.2185), Media:TRUSD (p=0.0604) Media:BRUSD (p=0.2857), TRUSD:BRUSD (p=0.2070), and Media:TRUSD:BRUSD (p=0.0972) and had an R-squared value of 0.9936.

Conclusion: Based on the results of the correlation tests, we conclude that there is not a monotonic relationship between TP and any variable and that TP is not strongly correlated with any one variable; however, our multivariate model suggests that Media, TRUSD, and BRUSD, when interacting, predict an incredible amount of the variability in TP. This makes sense, as media attention likely influences audience excitement.