Assessed Practical: Predicting the Olympic Games

Assignment due: 12pm Friday March 15

The Olympic Games have always mixed pure sporting spectacle with national competition. During the Cold War the USA and the Soviet Union competed fiercely to win the most medals in each games. On a somewhat milder level, in Britain we often compare our medal count with that of Australia, one of our traditional sporting rivals. If you were in the UK during the summers of 2012 and 2016 you cannot have missed the excitement caused by the UK's success relative to previous years.

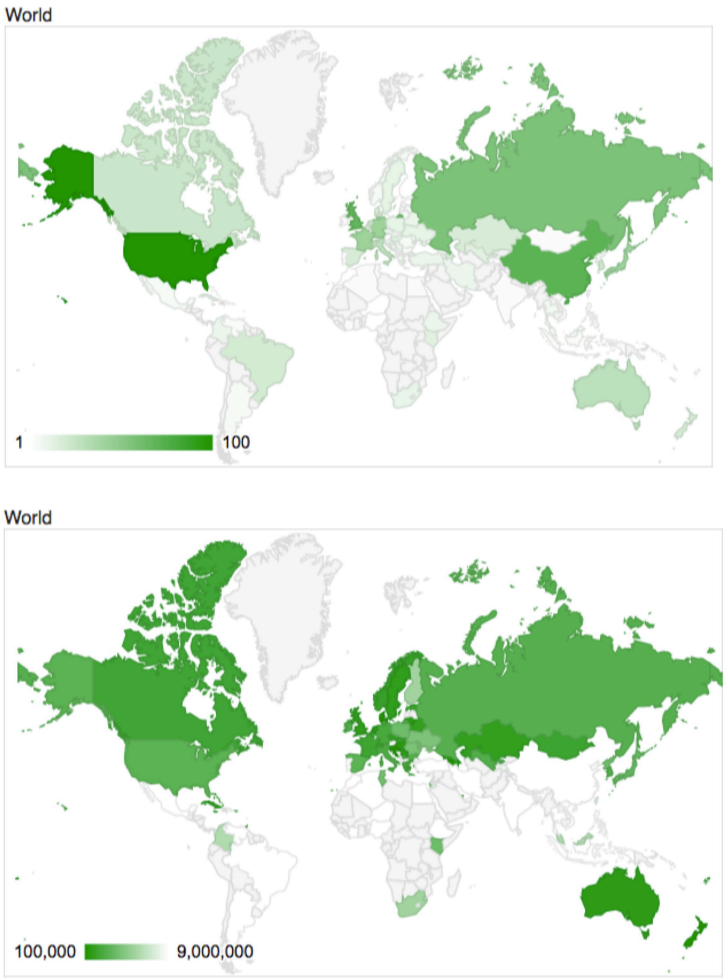

Figure 1: Total medals per country (top) and medals per capita (bottom) in the Rio 2016 Olympic Games (credit: http://www.medalspercapita.com/)

This competition is usually expressed in terms of the number of medals won by each country's athletes ((Figure 1 top panel). However, many interested watchers, especially those from smaller countries, have pointed out that the medal table is hardly a fair reflection of a country's sporting prowess. Some countries have a strong tradition of sporting excellence, but are simply too small to make an impact in terms of total medals. These commentators would rather look at the per capita medal count (Figure 1 bottom panel).

Looking at the per capita map above though, we see that large areas of the world are still very under-represented. Specifically, poorer countries do not win many medals per head of population. There are many reasons for this, including a lack of investment in sport and facilities, and fewer individuals who are wealthy enough to devote their life to training. As such, it has been suggested that we should compensate for wealth when measuring a country's Olympic performance.

In this practical you will investigate how the number of medals a country wins can be predicted from national population and GDP, and how consistent these relationships are. This practical will run over this week and the next.

Begin by downloading the data file medal_pop_gdp_data_statlearn.csv from MINERVA

This data file contains the following infomation for 71 countries (those that won at least one gold medal in each of the last three games):

- Country name (as recognised by the IOC)
- Population
- GDP (in billions of US dollars)
- Medals won in Beijing 2008, London 2012 and Rio 2016

Tasks (Regression):

1. Perform a linear regression to predict the medal count in 2008 and 2012 (separately, in two regressions) from Population and GDP and report your results.
2. How consistent are the effects of Population and GDP over time?
3. Using the regression for the 2012 medal count make a prediction for the results of 2016.
4. Plot your predictions against the actual results of 2016. If the results are hard to see, use a transformation of the axes to make it these clearer. How good are the predictions? Which countries are outliers from the trend?

Tasks (Model Selection):

1. Fit linear regressions models for the total medal count in 2012 using: (i) Population alone; (ii) GDP alone; (iii) Population and GDP. Select the model that minimises the Akaike Information Criterion.
2. Use cross-validation to perform a model selection between (i) Population alone; (ii) GDP alone; (iii) Population and GDP. Does your result agree with the model selected by AIC?
3. Using the three fitted models from Q1, predict the results of Rio 2016. Which predicts best? Compare this result with earlier answers.