**FURTHER MATHEMATICS (Unit 3)**

**APPLICATION TASK (SAC) 2022**

**CORE: Data Analysis**

***OLYMPIC MEDALS***

**SOLUTION**

**Start date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ End date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Time: *4 periods* (200 mins)**

**The SAC is designed to run in 2 sittings (3 parts)**

**Student Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Student Number: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**SAC Details:**

This task is divided into three components to be completed over 4 periods.

The SAC assesses the 3 outcomes, with the marking allocation as outlined below, and contributes to 40 % of the marks for Unit 3 and 4 SAC work combined.

**Marking allocation**

|  |  |
| --- | --- |
| **Outcomes** | **Marks allocated** |
| **Outcome 1** | 10 |
| **Outcome 2** | 20 |
| **Outcome 3** | 10 |
| **TOTAL MARKS** | **40** |

**Conditions and restrictions**

* Students are permitted to bring into the room for this task: pens, pencils, highlighters, erasers, sharpeners, and rulers. They may also bring in an approved bound reference.
* Students are NOT permitted to bring into the room for this task: blank sheets of paper and/or white out liquid/tape.
* An approved graphics or CAS calculator is permitted
* No copying of the task book is allowed in any way, shape or form.

**Materials supplied**

* Task book of 19 pages.

**Instructions**

* Print your name in the space provided on the bottom of the front page.
* All written responses must be in English.

**Students are NOT permitted to bring mobile phones and/or any other unauthorised electronic communication devices into the room for this task.**

**Data Collection**

The first Modern Summer Olympic Games were held in 1896 in Athens, Greece. Since then, the competition has moved all over the world and over 200 nations participate.

The Winter Olympic Games is a major international multi-sport event held once every four years for sports practiced on snow and ice. The first Winter Olympic Games, the 1924 Winder Olympics, were held in Chamonix, France.

The first Winter Olympics were held the same year as the Summer Olympics but since 1992 have been held on alternate even years*.*

Due to the Covid pandemic the Summer Olympics scheduled for Tokyo in 2020 were held in 2021 and this year the Winter Olympics are being held in Beijing. No spectators were allowed in the Summer Olympics.

**Part 1: Looking at your data over time**

This data relates to **the number of** **Olympic medals Australia and Great Britain have won** in total from the 1948 Olympics through to the 2021 Olympics (19 data points).

You are going to compare the **number of medals won by countries** and the **GDP\* of the countries**.

*\* The total value of all the goods and services produced by a country in a particular year, divided by the number of people living there*

*Table 1)*

|  |  |  |
| --- | --- | --- |
| **Host City and Year** | **Total Medals Won by Australia** | **Total Medals Won by Great Britain** |
| London 1948 | 13 | 23 |
| Helsinki 1952 | 11 | 11 |
| Melbourne 1956 | 35 | 24 |
| Rome 1960 | 22 | 20 |
| Tokyo 1964 | 18 | 18 |
| Mexico City 1968 | 17 | 13 |
| Munich 1972 | 17 | 18 |
| Montreal 1976 | 5 | 13 |
| Moscow 1980 | 9 | 21 |
| Los Angeles 1984 | 24 | 37 |
| Seoul 1988 | 14 | 24 |
| Barcelona 1992 | 27 | 20 |
| Atlanta 1996 | 41 | 15 |
| Sydney 2000 | 58 | 28 |
| Athens 2004 | 49 | 30 |
| Beijing 2008 | 44 | 51 |
| London 2012 | 35 | 65 |
| Rio 2016 | 29 | 67 |
| Tokyo 2021 | 46 | 65 |

a) Based on table 1, construct a time series plots showing the number of medals won by Australia and Great Britain from 1948 to 2021, per each Olympic year.

b) Describe the two time series plots commenting on any patterns, similarities or differences in the trends between the results for the two countries.

For both Great Britain and Australia there was an overall increasing trend with random variation.

Great Britain hosted the games in1948, and in 2012 and there are peaks at those two dates, with Great

Britain winning 23 and 65 medals respectively on those two dates, both were a substantial rise from the previous years.

Australia hosted the Olympics in 1956 and in 2000 and won 35 and 58 medals respectively.

**Part 2: Comparing number of medals**

a) Based on table 1, calculate the five-figure summary, mean, standard deviation, range and interquartile range for the number of medals won by Australia and Great Britain.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Min | Q1 | Q2 | Q3 | Max | IQR | Mean | *Sx* | *range* |
| Australia | 5 | 14 | 24 | 41 | 58 | 27 | 27.05 | 15.22 | 53 |
| Great Britain | 11 | 18 | 23 | 37 | 67 | 19 | 29.63 | 18.51 | 58 |

b) Construct an appropriate graphical display for total number of medals won by Australia and Great Britain. Discuss why you have chosen this particular display.

Chart

Description automatically generated

A box plot was chosen as it allows for data to be compared that has a categorical variable (country) and numerical variable (number of medals).

A stem and leaf plot could have also been used as there were only 2 groups with the categorical variable.

c) Compare and contrast the distributions by commenting on shape, centre, spread and presence of outlier(s). Discuss if there is an association.

The distribution of medals won for Australia is positively skewed with no outlier, Great Britain’s distribution is also positively skewed, but it has an outlier at 67.

The center, as measured by the median, for Australia is 23 medals, which is similar to the median for Great Britain with 24 medals.

The spread, as measured by the IQR, for Australia is 27 which is larger than the IQR of 19 for Great Britain.

There does not appear to be an association between the countries and the number of medals won as the median for Australia; 23 is like the median for Great Britain; 24.

**Part 3: Regression analysis**

This table demonstrates **the total number of Olympic medals won** by **31 countries** and compares them to see if **the countries with a greater GDP** (i.e. richer) have won more medals.

*Table 2)*

|  |  |  |
| --- | --- | --- |
| **Country** | **Total Medals Won** | **GDP (billion USD)** |
| Australia | 557 | 1380.207 |
| Belgium | 155 | 533.097 |
| China | 634 | 14342.933 |
| Finland | 305 | 269.296 |
| Canada | 326 | 1741.496 |
| Great Britain | 916 | 2826.441 |
| Greece | 121 | 205.326 |
| Hungary | 511 | 163.469 |
| Iceland | 4 | 24.188 |
| Japan | 497 | 5082.465 |
| Luxembourg | 2 | 71.104 |
| Malaysia | 13 | 364.684 |
| Netherlands | 321 | 907.05 |
| New Zealand | 139 | 206.936 |
| Paraguay | 1 | 38.086 |
| Peru | 4 | 226.85 |
| Ukraine | 139 | 153.781 |
| United Arab Emirates | 2 | 421.142 |
| United States | 2636 | 21433.226 |
| Belarus | 85 | 63.08 |
| Austria | 94 | 445.075 |
| Azerbaijan | 49 | 48.047 |
| Romania | 308 | 250.075 |
| Russia | 505 | 1692.93 |
| Serbia | 24 | 51.475 |
| Singapore | 5 | 372.073 |
| Slovakia | 32 | 105.079 |
| Guatemala | 1 | 76.71 |
| France | 764 | 2715.518 |
| Jamaica | 87 | 15.83 |
| Switzerland | 205 | 731.425 |

a) Based on table 2, construct a **scatterplot** of the GDP of the 31 countries (Explanatory variable) versus the number of medals won (Response variable). Describe the association using appropriate terminology.

From the scatterplot and Person’s product moment Correlation Coefficient, r = 0.8630, we can see that there is a strong, positive, linear relationship with two possible outliers at (14342, 634) and (21433, 2636).

b) Calculate the equation of the least-squares regression line, draw the regression line on the scatterplot and interpret the summary statistics using context.

The least squares regression line:

The vertical intercept of this line is 128.88 showing that you could win 129 medals even if your GDP is zero.

The slope of 0.0956 indicates that you can win an extra 0.1 (rounded) or 1 medal per one hundred million dollars in GDP increase.

The coefficient of determination indicates that 74.47% of the variation in medal counts can be explained by the variation in GDP the other 25.53% can be attributed to other factors.

c) Carry out a residual analysis and create a residual plot. Comment on what the residual plot indicates.

Chart, scatter chart

Description automatically generated

Although the residuals appear randomly scattered therefore indicating that the data is linear, there is a great deal of data at the bottom end of the scale so there could be two transformations that could be tried to linearize the data by spreading out the bottom end of the scale.

d) Choose and apply two appropriate transformations to either of the variables and select which model is best suited for making predictions.

log transformation of the GDP values

Chart, scatter chart

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Residual Graph of Log 10 GDP

Chart, scatter chart

Description automatically generated

(Number of medals )2 Transformation

Chart, line chart, scatter chart

Description automatically generated

Table

Description automatically generated with medium confidence

Residual Graph of (Number of medal)2 Transformation

Chart, scatter chart

Description automatically generated

|  |  |  |  |
| --- | --- | --- | --- |
| Data involved | Coefficient of determination  *r* | Shape of residual graph | Equation of the Least Squares regression line |
| Original data | 0. 8630 | Difficult to see pattern as there are a lot of smaller values grouped together at the lower end of the GDP axis | *Number of medals*  *= 128.88 + 0.0956 × GDP* |
| Log10 GDP  This was attempted so that the larger end of the GDP scale would be compressed to make all the values smaller and more closely grouped together | 0.7123 | There appears to be a trough pattern in the residuals indicating that the linearization has not been successful, and it also has the lowest coefficient of determination value | *Number of medals*  *= -864.65 +459.04× log 10 (GDP)* |
| (Number of Medals) 2  This was also attempted to make the medal values larger and extending that scale | 0.8438 | The residuals look randomly scattered but again there is a great deal of data at the smaller vales of the (number of medasl)2  so it is difficult to see if there is a pattern at that end of the data | *(Number of medals)2*  *= -91372.53 + 232.16 ×(GDP)* |

Discussion on the most suitable model.

Looking at the table above the Least Squares regression line generated from the original data is the best one to use as it has the largest Coefficient of determination value.

The Residual analysis is difficult to consider as there is such a lot of data at the smaller end of the scale, where you cannot distinguish clearly if there is a pattern in the residuals or not.

e) Use the most suitable model to predict the number of tests carried out and the GDP values in your actual data set.

Least Squares Lines are generally suited to interpolation rather than extrapolation as you are predicating for data inside the data set rather than outside the data set.

Predicting how many medals a country can win if it has a GDP of 1380 billion (AUSTALIA’S GDP)

This is interpolation so you would expect this to be reliable.

In fact, Australia has done much better than this and has won 559 medals to date.

The other GDP we will look at is 15.83 billion. This is a small GDP relative to our data set. (JAMAICA’S GDP)

this is also interpolation, and you would expect this to be reliable.

However, Jamaica has won 87 medals, therefore underperforming according to its GDP.