VCE General Mathematics Unit 3

Area of Study 1 sample application task: Data analysis
Live long and prosper

The application task is to be of 4–6 hours’ duration over a period of 1–2 weeks.

Introduction

Gapminder is an independent foundation based in Sweden that makes available a large amount of publicly funded data collected by governments and non-governmental organisations from around the world. This data can be accessed at [Gapminder](https://www.gapminder.org/) and [downloaded](https://www.gapminder.org/data/).

In this task we consider life expectancy for people living in countries around the world, and how life expectancy is related to economic conditions in those countries. These relationships can be explored within the Gapminder website itself through the Gapminder World tab on the site; alternatively, the data can be explored by downloading it and using suitable technology.

The data sets considered in this task are Life Expectancy (Years) and Income per person (GDP per capita) for the set of 260 countries, for which data may have been recorded over the years 1800 to 2016 (note that gaps in the spreadsheet reflect instances for which no data was available for the respective country and year combinations). The data set recording total population (Population, total) of each country over the years 1800 to 2015 may also be useful. To download these data sets, search for the terms ‘life expectancy’, ‘income’ and ‘population’ using the search box within the Data tab on the Gapminder website.

Component 1

Compare the distribution of life expectancies in 1900 with those in 2015 for countries for which relevant data for these years is available.

1. Construct a suitable display to compare the average life expectancies for these countries in 1900 and 2015.
2. Compare the distributions of life expectancies for these countries in terms of shape, centre, spread and outliers.
3. It is suggested that a ‘world life expectancy’ for 2015 can be calculated by averaging the life expectancies for each of the countries for which an average life expectancy figure is available. Is this suggestion reasonable? Explain how the calculation could be adjusted to refine this estimate of life expectancy.

Component 2

1. Consider Japan and the United States. For each of these countries construct a time series plot of the life expectancies from 1900 to 2015 on the same axes.
2. Describe the two time series plots, noting any similarities or differences in the trends between the two countries. Comment on any unusual features of the two-time series.

Component 3

In this component the relationship between how long a person lives and how wealthy they are is explored – do we ‘live long *and* prosper’?

Consider all countries for which data has been collected for the year 2000.

1. Construct a scatterplot of the life expectancies in 2000 against the income per capita in that year, and describe it in terms of direction, form and strength. Find the equation of the regression line for this scatterplot. Discuss whether this linear model is appropriate or not.
2. Apply a suitable data transformation and construct the corresponding scatterplot. Describe this relationship in terms of direction, form and strength. Discuss which of the two scatterplots is most useful for predicting longevity based on income.
3. For the relationship in b. above, find the equation of the regression line, draw the regression line on the scatterplot, and interpret the intercept and slope in terms of the how income related to life expectancy in 2000.
4. Use this model to predict the life expectancy in 2000 based on the income per capita reported for a randomly selected country. Compare the prediction from your model with the actual recorded life expectancy for that country, using percentage error to determine the accuracy of your prediction.
5. For both models find the correlation coefficient and the coefficient of determination between the life expectancy and the respective income variables in each case, and interpret these statistics. Explain whether the coefficient of determination is a reasonable measure of the strength of the relationship between life expectancy and the respective income-related variables for each of the cases described in a. and b. above.

Areas of study

The following content from the areas of study is addressed through this task.

|  |
| --- |
| **Unit 3** |
| **Area of study** | **Topics** | **Content dot points** |
| Data analysis, probability and statistics | Investigating data distributionsInvestigating associations between two variablesInvestigating and modelling linear associationsInvestigating and modelling time series data | 3, 4, 51, 3, 4, 5, 61, 2, 3, 4, 5, 6, 7, 81 |

Outcomes

The following outcomes, key knowledge and key skills are addressed through this task.

|  |
| --- |
| **Unit 3** |
| **Outcome** | **Key knowledge dot point** | **Key skills dot point** |
| 1 | 1, 2, 4, 7, 8, 9, 10, 11, 13 | 2, 3, 7, 8, 9, 10, 11, 12, 13, 14, 18 |
| 2 | 1, 2, 3, 4 | 1, 4 |
| 3 | 1, 3, 4, 5, 6 | 1, 3, 4, 5, 6, 7, 9, 10, 11, 12 |