MXN500: Problem Solving Task 2

Due: Friday 31-May-2024 11:59 pm

### Student Details

Family Name:

Given Name:

Student Number:

### Introduction

#### Background

Large scale climate drivers such as the El Niño Southern Oscillation (ENSO) are known to have an impact on Australian rainfall patterns. ENSO has three phases, El Niño, Neutral and La Niña. Generally, in the La Niña phase of ENSO conditions are cooler and wetter along the Eastern Australian coast. In comparison during the El Niño phase conditions are hotter and drier. In this problem solving task, you will be using your knowledge of regression to explore the relationship between ENSO and Australian rainfall.

Note that it is not possible to measure the strength of ENSO directly, so in regression models the Southern Oscillation Index (SOI) is commonly used to represent the strength of ENSO. The SOI is a climate index that measures the normalised pressure difference between Tahiti and Darwin. You do not need to provide units when displaying SOI on a plot axis as it is an index. ENSO is considered to be in the La Niña phase when there are sustained SOI values above 7, the El Niño phase when there are sustained SOI values below -7, and the Neutral phase otherwise. A csv file containing the monthly SOI values can be obtained from blackboard. More details about ENSO and SOI can be found on the Bureau of Meteorology website (<http://www.bom.gov.au/climate/enso>).

#### Datasets

Two datasets are provided for this assignment. Similarly to the data used in problem solving task 1, there is a precipitation dataset, total\_seasonal\_rainfall.csv. This dataset contains the seasonal rainfall totals recorded at BRISBANE AERO station and includes variables:

* The GHCN Daily station id (character)
* The Season of the observation (ordinal, categorical)
* The Year of the observation (numeric)
* The total\_seas\_prcp is the total amount of rainfall received during that Season in mm (numeric).
* The GHCN Daily station name (character)

In the dataset, seasonal\_soi\_data.csv, the variables included are:

* The Season of the observation (ordinal, categorical)
* The Year of the observation (numeric)
* The SeasonalSOI is the mean seasonal SOI value (numeric)
* The Phase of the ENSO (ordinal, categorical).

### Preprocessing

##### Question 1.1: (1 mark)

Download the the files total\_seasonal\_rainfall.csv and seasonal\_soi\_data.csv from Canvas. Combine all the variables from these two datasets into the single dataset total\_seasonal\_rainfall.

Print the first three rows to show the form of your new dataset.

##### Question 1.2: (3 marks)

Convert the appropriate variables to factors in your dataset. Be sure to set the factor levels appropriately for later analysis.

Show your code and show the factor levels.

### Exploratory Visualisaion

For each season, we are interested in whether a simple linear regression

could be used to model the relationship between the mean seasonal SOI value and total seasonal precipitation.

##### Question 2.1: (5 marks)

Create a visualisation that explores the relationship between SeasonalSOI and total\_seasonal\_prcp for each season.

Use geom\_smooth() to add both the null model and the linear model from equation (1).

##### Question 2.2: (3 marks)

Using your visualisation, for which seasons would you expect there to be a significant linear relationship between total seasonal precipitation and the mean seasonal SOI value? Give your reasoning.

### Simple linear Regression

For the season with the strongest relationship (as determined in the previous section) fit the regression model from equation (1) and answer the following questions.

##### Question 3.1: (2 marks)

Fill in the blanks in the following sentence so that it refers to the terms in the regression model.

*For the BRISBANE AERO Station and the season, a linear model was specified to model how the total seasonal precipitation, , is related to the mean seasonal SOI value, . The parameter describes the rate of change in the total seasonal precipitation with an increase in mean seasonal SOI value. The parameter represents the total seasonal precipitation when the mean seasonal SOI value is 0.*

##### Question 3.2: (2 marks)

Write down the equation for the linear model above, substituting the parameter values into the equation.

##### Question 3.3: (2 marks)

Provide a 95% confidence interval for the parameter estimates.

##### Question 3.4: (1 mark)

How much variability in the data is explained by this model?

##### Question 3.5: (4 marks)

Visualise the fitted values compared with the residuals, and visualise the standardised quantiles of the residuals compared with the theoretical quantiles.

Discuss the validity of the assumptions of this linear regression.

##### Question 3.6: (4 marks)

Print out a summary of your linear model and interpret the results. Make sure you discuss the physical meaning of the model, which parameters are significant and whether the linear model is significantly different compared to the null model.

##### Question 3.7: (2 marks)

Discuss whether your fitted model is a good model to use to predict seasonal rainfall totals using the mean seasonal SOI.

### Polynomial Lines of Best Fit

Climate experts think that in places where it rains a lot, it might rain even more. But in dry places, it might not get any drier. This means that in Australia, the amount of rain might change differently during La Niña and El Niño, which are two opposite weather patterns. Because of this, it could be useful to see if a polynomial regression fits our data better.

##### Question 4.1: (2 marks)

For BRISBANE AERO and the season you chose earlier, fit a linear regression using polynomial explanatory variables of up to order 2 of the SeasonalSOI.

Write down the equation with your estimated parameter values.

##### Question 4.2: (3 marks)

Print out a summary of your fitted model, interpret the significance of the results and explain the related physical meaning.

##### Question 4.3: (3 marks)

Create a prediction interval for a mean seasonal SOI value of 25 and a mean seasonal SOI value of -25. Discuss the appropriateness of extrapolating the model to these values.

##### Question 4.4: (3 marks)

Decide whether a linear or polynomial regression is preferred using a statistical test. Be sure to describe your statistical test in detail.

### Linear Regression with Categorical Explanatory Variables

Within the analysis so far the role of ENSO phases has been modelled solely using the SOI. For model simplicity and for physical understanding, it may be useful to consider only the phases.

##### Question 5.1: (3 marks)

Consider again BRISBANE AERO and your chosen season. To better understand the role of different Phases of ENSO, fit a categorical regression model of the form:

Print the model summary.

##### Question 5.2: (1 mark)

Write down the linear model substituting the parameter values into the equation.

##### Question 5.3: (3 marks)

Interpret the significance of the results of your fitted model and explain the related the physical meaning.

##### Question 5.4: (3 marks)

Do the results of this categorical linear model support your choice of linear or polynomial regression above?

### Code quality and readability

##### Readability and clarity of code (5 marks)

This assignment is primarily about your ability to perform a statistical data analysis, with additional marks based on how clear and readable your code is. To facilitate the markers’ evaluation, please ensure that you provide comments in your code corresponding to each question.