**Day 4 Individual Assignment**

|  |
| --- |
| Name: |

**Use the format “YourName\_Assignment4.docx” for assignment submission.**

**Day 4 Part I. Simple Linear Regression**

Q1. Sunflowers Apparel is looking to expand its business by opening a new retail outlet. Leasing agents from the Triangle Mall Management Corporation have suggested that Sunflowers Apparel consider several locations in some of Triangle’s newly renovated shopping malls that cater to shoppers with higher-than-mean disposable income.

Although the locations are smaller than the typical Sunflowers location, the leasing agents argue that higher-than-mean disposable income in the surrounding community is a better predictor than store size of higher sales. The leasing agents claims that sample data from 14 Sunflowers stores (as shown in Table 1 below) prove that this is true.

**Table 1**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Store | Average Disposable Income  ($000) | Annual Sales (Millions $) | Store | Average Disposable Income  ($000) | Annual Sales (Millions $) |
| 1 | 22.3 | 3.7 | 8 | 21.4 | 2.7 |
| 2 | 36.6 | 3.9 | 9 | 44.4 | 5.5 |
| 3 | 55.5 | 6.7 | 10 | 34.1 | 2.9 |
| 4 | 46.7 | 9.5 | 11 | 51.8 | 10.7 |
| 5 | 32.4 | 3.4 | 12 | 45.1 | 7.6 |
| 6 | 31.7 | 5.6 | 13 | 52.0 | 11.8 |
| 7 | 41.6 | 3.7 | 14 | 49.2 | 4.1 |

The dataset provided to you (“Triangle.csv”) consists of 14 observations and 3 variables as shown above. Answer the following questions by analysing the dataset in R:

Q1a. Construct a scatter plot and comment on the relationship between the Average Disposable Income and Annual Sales.

|  |
| --- |
|  |

Q1b. Compute and interpret the correlation coefficient.

|  |
| --- |
|  |

Q1c. At the 0.05 level of significance, is there evidence of a linear relationship between Average Disposable Income and Annual Sales? Thus, should the management of Sunflowers Apparel accept the claims of Triangle’s leasing agents?

|  |
| --- |
|  |

Q1d. Perform the necessary statistical tests to check for the following model assumptions:

* Independence of observations
* Error Terms with Constant Variance
* Normality of Residuals

|  |
| --- |
|  |

**Day 4 Part II. Multiple Regression**

Q2. Starbucks launched its prepaid Starbucks Card in November 2001. The card, which holds between $5 and $500, can be used at virtually any Starbucks location. The card was so popular when it first was released that many stores ran out. By mid-2002, Starbucks had activated more than 5 million of these cards. It is believed that the card accounted for a large portion of the company’s 7% same store increase in sales in early 2002 and that it is responsible attracting many new patrons to the store.

Starbucks has gone on to promote their Starbucks Card as a flexible marketing tool that can be used by individuals as a gift of thanks and appreciation for friendship or service and can be used by companies to reward loyal customers and as an incentive to employees.

The Starbucks management wants to study the reason why some people purchase Starbucks Card with higher prepaid values than do other people. 25 prepaid card purchases have been randomly selected in this study. The dataset provided to you (“StarbucksPrepaid.csv”) consists of 25 observations and 5 variables. Descriptions of the variables in this dataset are shown in the **Table 2** below:

**Table 2**

|  |  |
| --- | --- |
| **Variable(s)** | **Description** |
| Amount | Amount of Prepaid Card ($) |
| Age | Age as provided by the Cardholder |
| Days | Days per Month at Starbucks |
| Cups | Cups of Coffee per Day |
| Income | Income (in $100s) |

Answer the following questions by analysing the dataset in R:

Q2a. Using these data, develop a Multiple Linear Regression model to study how well the amount of the prepaid card can be predicted by the other variables and which variables seem to be more important in doing the prediction. Interpret the results from your model in the given context.

|  |
| --- |
|  |

Q2b. How good is your model fit? Explain your answer in the given context.

|  |
| --- |
|  |

Q2c. Perform the necessary statistical tests to check for the following model assumptions:

* Independence of observations
* Error Terms with Constant Variance
* Normality of Residuals
* No Multicollinearity between the independent variables

|  |
| --- |
|  |

Q2d. How does your analysis help Starbucks management in attempting to determine what drives sales revenues?

|  |
| --- |
|  |

**Day 4 Part I & II. Conceptual Understanding**

Q3. Answer all three questions below with **clear explanations**.

Q3a. In Simple Linear Regression, does the model fit affects the statistical significance of regression coefficients?

|  |
| --- |
|  |

Q3b. Why do we need to ensure that the error terms have constant variance in regression models (Assumption 7)?

|  |
| --- |
|  |

Q3c. In Multiple Linear Regression model, is it an issue if some of the independent variables have nonlinear relationships with each other?

|  |
| --- |
|  |