Mid-term Test

ST3131 Regression Analysis
Semester II, 2020/2021

Read all of the following information before starting the exam:

- Please show all work, clearly and in order. **Indicate your matrix number in every page of your submission.**
- Please leave your final answer to **at least 3 significant figures**.
- The duration of this test is **90** minutes.
- This test has **THREE** problems.
- All the questions are worth 50 points.

1. (*10 points*) Dataset `QFR.csv` is given on Luminus. This dataset includes a response variable **quality** of a type of fruit, and two possible regressors **flavor** and **region** (location where the fruits came from).

a. (*3 pts*) Read the dataset into R and plot a histogram of quality variable. Report the range of the quality and give your comment on this histogram.

b. (*3 pts*) Plot a scatter plot between flavor and quality. Derive the correlation coefficient of these two variables. Give your comment on the plot and the correlation coefficient.

c. (*2 pts*) Fit a linear model for the response quality with the two regressors given (called Model 4). Write down this fitted model.

d. (*2 pts*) Report R^2 of Model 4 and derive the total sum of squares (SST) of Model 4.

2. (*25 points*) A selling cars company would relate the amount of money spent for marketing every month (x) and the sales of that month (y). The data were collected for a year (12 months). A linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$ where ε is assumed to have mean zero and constant variance σ^2 is fitted to the data. Denote the fitted response as $\hat{y}$, the residuals are $e_i = y_i - \hat{y}_i, \quad i = 1, ..., n$. Figure 1 below is the output of the fitted model from R with some omitted information.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1383.4714  1255.2404       ?    0.296
x             10.6222           ?  65.378 1.71e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2313 on 10 degrees of freedom
Multiple R-squared:       ?     Adjusted R-squared:       ?
F-statistic:       ? on 1 and 10 DF,  p-value:          ?
```

Figure 1: Summary output of Model 1

a. (*10 pts*) Calculate the values that are replaced by the question mark in the figure above.

b. (*2 pts*) Estimate σ.

c. (*2 pts*) Write down the fitted model. (We name this fitted model as Model 1.)

d. (*4 pts*) Test the significance of variable x in Model 1.

e. (*2 pts*) Use Model 1 to predict the sale for a month for which the amount of money spent for marketing is \$10000.

f. (*5 pts*) Derive a 95% CI for the coefficient of variable x. If the intercept of Model 1 is fixed as given in the output above, derive a 95% CI for the mean sales of a month for which the amount of money spent for marketing is \$10000.

3. (*15 points*) The quality of a type of wine in France is thought to be related to the properties of aroma, flavor and the region (location, A, B and C) where the wine was produced. A linear model is fitted (called Model 2) where we assume the error terms are independent, identical normally distributed with constant variance. The output in R is given in Figure 2.

```
Call:
lm(formula = Quality ~ Aroma + Flavor + reg, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-1.97642 -0.58602  0.02108  0.51519  1.93049

Coefficients:
            Estimate Std. Error
(Intercept)  7.12079    0.94826
Aroma       -0.01262    0.24029
Flavor       1.12206    0.21572
regB        -1.53452    0.37484
regC         1.23557    0.46813
---
```

```
Analysis of Variance Table

Response: Quality
           Df Sum Sq
Aroma       1 77.442
Flavor      1 24.494
reg         2 25.641
Residuals  33 27.211
```

Figure 2: Summary output of Model 2

a. (*3 pts*) Write down the fitted model (Model 2).

b. (*3 pts*) Test the significance of variable aroma in Model 2.

c. (*3 pts*) Test the significance of variable region in Model 2.

d. (*4 pts*) If we fit another model for the same response (called Model 3) with only variable aroma as regressor, test the significance of Model 3.

e. (*2 pts*) Derive R^2 of Model 2.

END OF PAPER