SMM634 - Individual Assignment

(worth 50% of final grading)

Deadline 4 November 2022

1. For this assignment you need to use data frame `wine` which contains information on prices and growing characteristics of 25 Bordeaux wines from 1952 to 1998. The data frame contains 7 columns and 47 rows. The columns are: year of production `year`, average price of the wines as a percentage of the 1961 price (`price`), mm of rain in the harvest month (`h.rain`), average temperature (C) over the summer preceding harvest (`s.temp`), mm of rain in the winter preceding harvest (`w.rain`), average temperature (C) at harvest (`h.temp`), a rating of the wine quality (`parker`).
See `https://www.wine-searcher.com/critics-27-robert+parker+the+wine+advocate` for details on `parker`.

The aim of the analysis is to model the response variable `price` as a function of the variables described above.

Using these data, write a report addressing the following points:

(a) Justification of the chosen regression model specification. [5]

(b) Using the final model, provide a summary (e.g., using tables and figures) of the empirical findings as well as interpretation of the estimated model parameters. [5]

(c) Provide recommendations and limitations of your analysis. [5]

(d) What did you learn from the analysis? What is the answer, if any, to the questions you set out to address? How can the analysis be improved? [5]

Other 5 marks will be for overall report structure (e.g., report neatness, presentation style, extra effort).

The report must be in pdf format. It (excluding the title page) must not be longer than 5 pages (including graphs, tables, etc.) using font size 12pt with one and a half line spacing and at least 2.5 cm margin.

[Total marks: 25]

2. An economist is interested in trying to find simple models for predicting the amount of carbon dioxide that individuals are responsible for emitting into the atmosphere each year. For a sample of two hundred adults she draws up a detailed carbon budget to estimate each person's CO2 emissions in tonnes per year (`CO2`). As part of the study their household income is recorded (`income` in thousands of pounds), along with whether or not they own a 4 wheel drive car (`fwd`, a factor with levels 0, for no and 1 for yes) and their beliefs about climate change (`belief`, a factor with levels: `yes`, they believe the science; `no`, they do not; and `not.understand`, for claiming not to understand the science.)

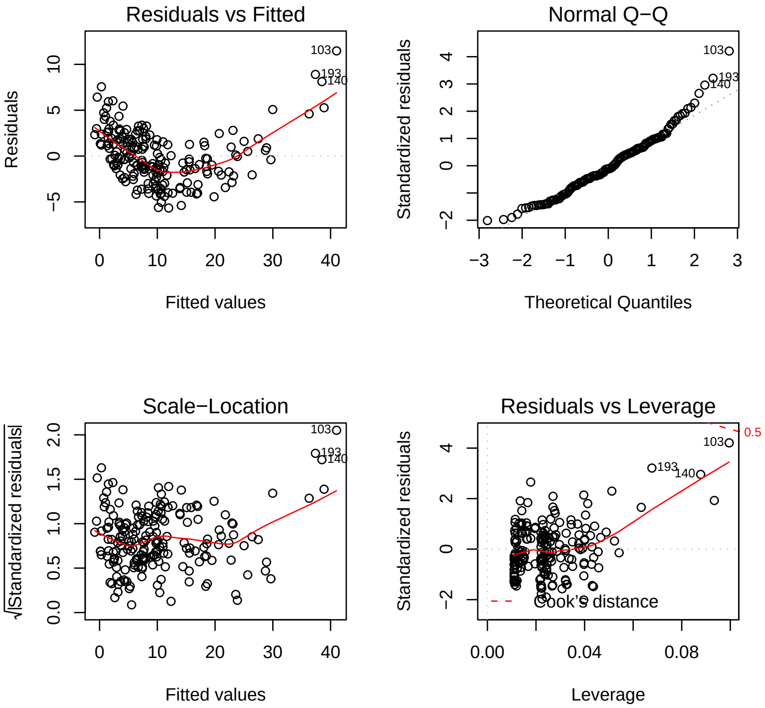
The following R session is an attempt to build a model for these data.

(a) By looking at the R code below write down the statistical model that has been fitted and the model assumptions. [2]

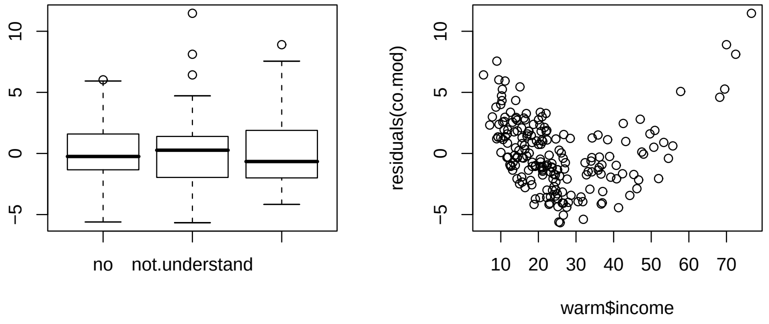
```
> co.mod <- lm(CO2~income+fwd+belief,data=warm)
```

(b) Comment on the following residual plots. [3]

```
> par(mfrow=c(2,2))
> plot(co.mod)
```



```
> par(mfrow=c(1,2))
> plot(warm$belief,residuals(co.mod))
> plot(warm$income,residuals(co.mod))
```

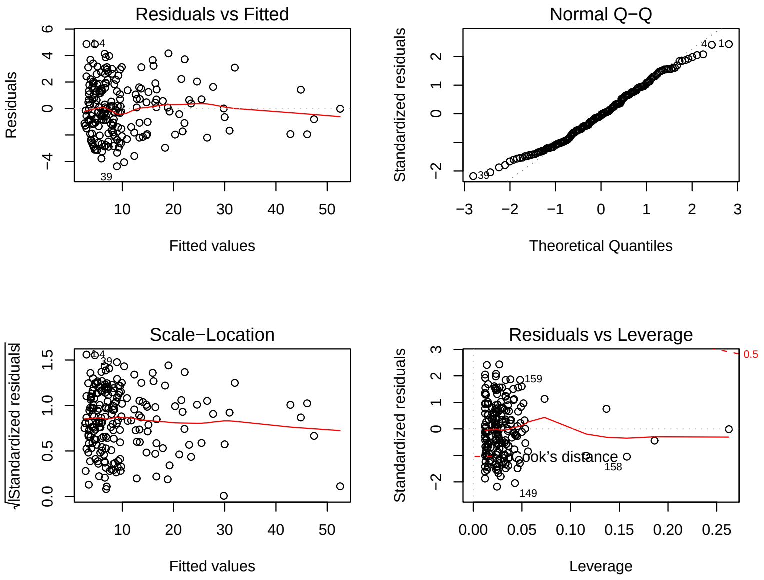


(c) Using the R code below, explain what model has been fitted. [1]

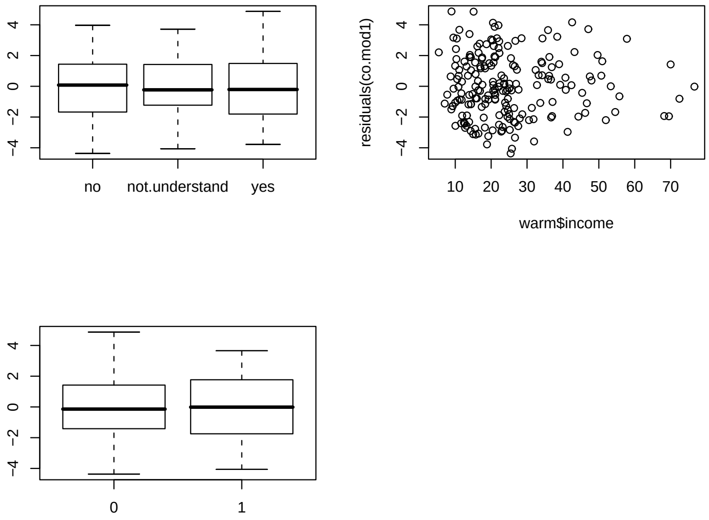
```
> co.mod1 <- lm(CO2~income+I(income^2)+fwd+belief,data=warm)
```

(d) Interpret the residual plots below. [4]

```
> par(mfrow=c(2,2))
> plot(co.mod1)
```



```
> plot(warm$belief,residuals(co.mod1))
> plot(warm$income,residuals(co.mod1))
> plot(warm$fwd,residuals(co.mod1))
```



(e) What conclusions can be drawn from the following analysis of variance? [2]

```
> anova(co.mod1)
Analysis of Variance Table
Response: CO2
             Df  Sum Sq   Mean Sq  F value     Pr(>F)
income        1 13513.8  13513.8  3282.6195  < 2.2e-16 ***
I(income^2)   1   793.7    793.7   192.8083  < 2.2e-16 ***
fwd           1   104.9    104.9    25.4765   1.027e-06 ***
belief        2     7.8      3.9     0.9475   0.3895
Residuals   194   798.7      4.1
```

(f) Explain the results of the following analysis of variance. Also comment on the two-way table below. [3]

```
> anova(lm(CO2~belief,data=warm))
Analysis of Variance Table
Response: CO2
             Df   Sum Sq  Mean Sq F value  Pr(>F)
belief        2    16.8    8.4     0.1089  0.8969
Residuals   197 15202.0   77.2

> table(warm$fwd,warm$belief)
   no not.understand yes
0  36    38           87
1  21    13            5
```

(g) Looking at the R code below, explain which model has been fitted and comment on the analysis of variance table. [2]

```
> co.mod2 <- lm(CO2~income+I(income^2)+fwd,data=warm)
> anova(co.mod2)
Analysis of Variance Table
Response: CO2
             Df  Sum Sq  Mean Sq  F value   Pr(>F)
income        1 13513.8  13513.8  3284.38  < 2.2e-16 ***
I(income^2)   1   793.7    793.7   192.91  < 2.2e-16 ***
fwd           1   104.9    104.9    25.49  1.013e-06 ***
Residuals   196   806.5      4.1
```

(h) By looking at the R code below which model has been fitted? What are the conclusions from the analysis of variance table? [3]

```
> co.mod3 <- lm(CO2~income+I(income^2)+fwd*income,data=warm)
> anova(co.mod3)
Analysis of Variance Table
Response: CO2
             Df  Sum Sq  Mean Sq   F value     Pr(>F)
income        1 13513.8 13513.8   3272.8146  < 2.2e-16 ***
I(income^2)   1   793.7   793.7    192.2324  < 2.2e-16 ***
fwd           1   104.9   104.9     25.4004  1.060e-06 ***
income:fwd    1     1.3     1.3      0.3098  0.5784
Residuals   195   805.2     4.1
```

(i) Do you think that the following model fitted in R is reasonable as compared to all models fitted above? Comment on the relationships between `CO2` and `income`, and `CO2` and `fwd`. [5]

```
> summary(co.mod2)

Call:
lm(formula = CO2 ~ income + I(income^2) + fwd, data = warm)

Residuals:
   Min      1Q    Median   3Q     Max
-4.3988 -1.5711 -0.1593 1.5466 4.8989

Coefficients:
             Estimate    Std. Error   t value   Pr(>|t|)
(Intercept)  1.6037309    0.5752177    2.788    0.00583 **
income       0.0860234    0.0387115    2.222    0.02742 *
I(income^2)  0.0075880    0.0005418   14.004    < 2e-16 ***
fwd          1.8310510    0.3626723    5.049    1.01e-06 ***

Residual standard error: 2.028 on 196 degrees of freedom
Multiple R-Squared: 0.947, Adjusted R-squared: 0.9462
F-statistic: 1168 on 3 and 196 DF, p-value: < 2.2e-16
```

[Total marks: 25]