MAST7866 Linear Regression

Computing Session 2

For this session, we will use the data sets that you used in Computing Session 1 (`production.txt`, `SaltBP.txt`, `diamonds.txt`) and also `AdRevenue.csv`.
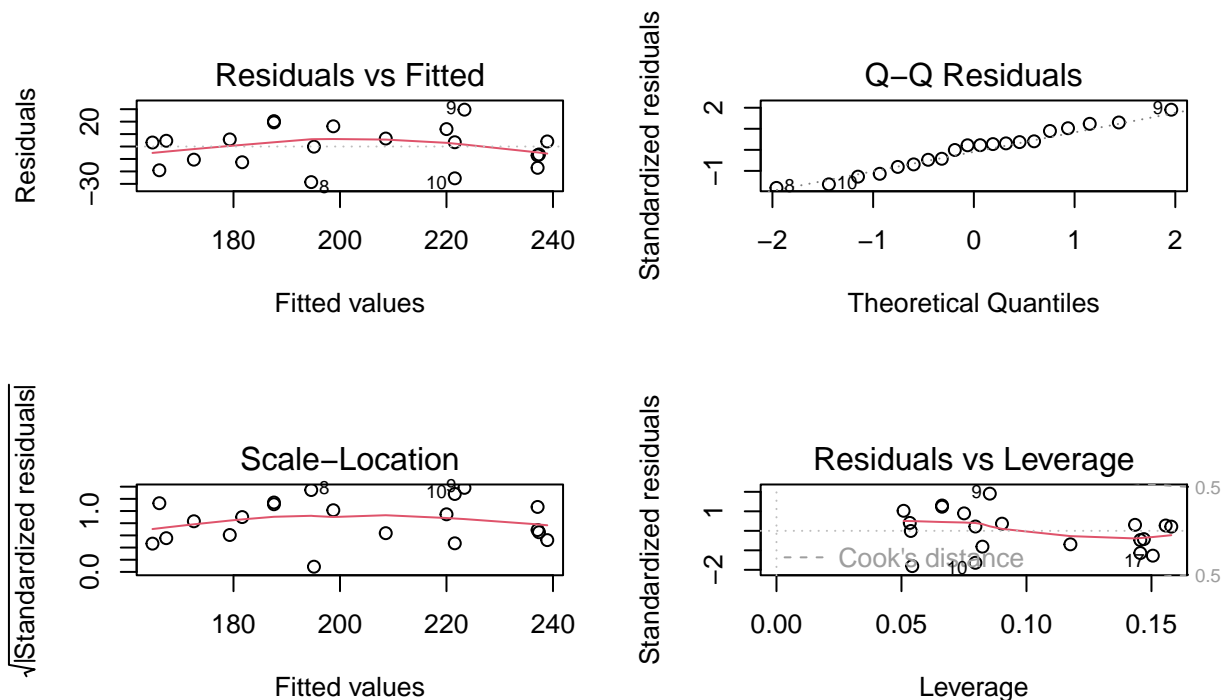
1 Production runs data

If you saved the workspace from Computing Session 1 then you should have the fitted regression model in an object called `fit_production`. If you did not save the workspace, the object can be re-created (after re-loading the data) using the command:

```
production <- read.table("production.txt", header = T)
fit_production <- lm(RunTime ~ RunSize, data = production)
```

In the previous sheet, you looked at estimation of the linear regression model. This sheet will look at model checking (using plots of the residuals) and transforming the data. Various graphs of the residuals can be plotted using the command

```
par(mfrow = c(2,2))
plot(fit_production)
```


As we saw in the previous session, more information about the fitted regression is provided if you type

```
summary(fit_production)  ##This produces the output below
```

```
##
## Call:
## lm(formula = RunTime ~ RunSize, data = production)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.597 -11.079   3.329   8.302  29.627
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 149.74770    8.32815   17.98 6.00e-13 ***
## RunSize       0.25924    0.03714    6.98 1.61e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.25 on 18 degrees of freedom
## Multiple R-squared:  0.7302, Adjusted R-squared:  0.7152
## F-statistic: 48.72 on 1 and 18 DF,  p-value: 1.615e-06
```

2 The effect of salt on blood pressure

The file saltBP.txt contains measurements on 25 elderly people. The variables measured are

- BP – denotes the systolic blood pressure
- salt – the average daily intake of salt in grams.

1. Produce the residuals plots.

   Do these indicate that the regression model is appropriate?

3 Pricing diamond rings

In the previous computing sheet, you looked at predicting the price of a diamond ring from the size of its diamond stone. After fitting a linear regression model, answer the following question.

1. Look at plots of the residuals.

   Do these indicate that the linear regression model is suitable?

4 Advertising revenue

The data is stored in the file `AdRevenue.csv`. After fitting a linear regression model to predict advertising revenue from circulation, answer the following questions:

1. Look at the residuals from your linear regression model.

   What do this indicate about the suitability of the linear regression model for these data?

2. Try fitting the regression model on transformed data by replacing Revenue by log(Revenue) and Circulation by log(Circulation) in the lm function.

   Look at the residual plots and comment on whether the linear regression model is appropriate.