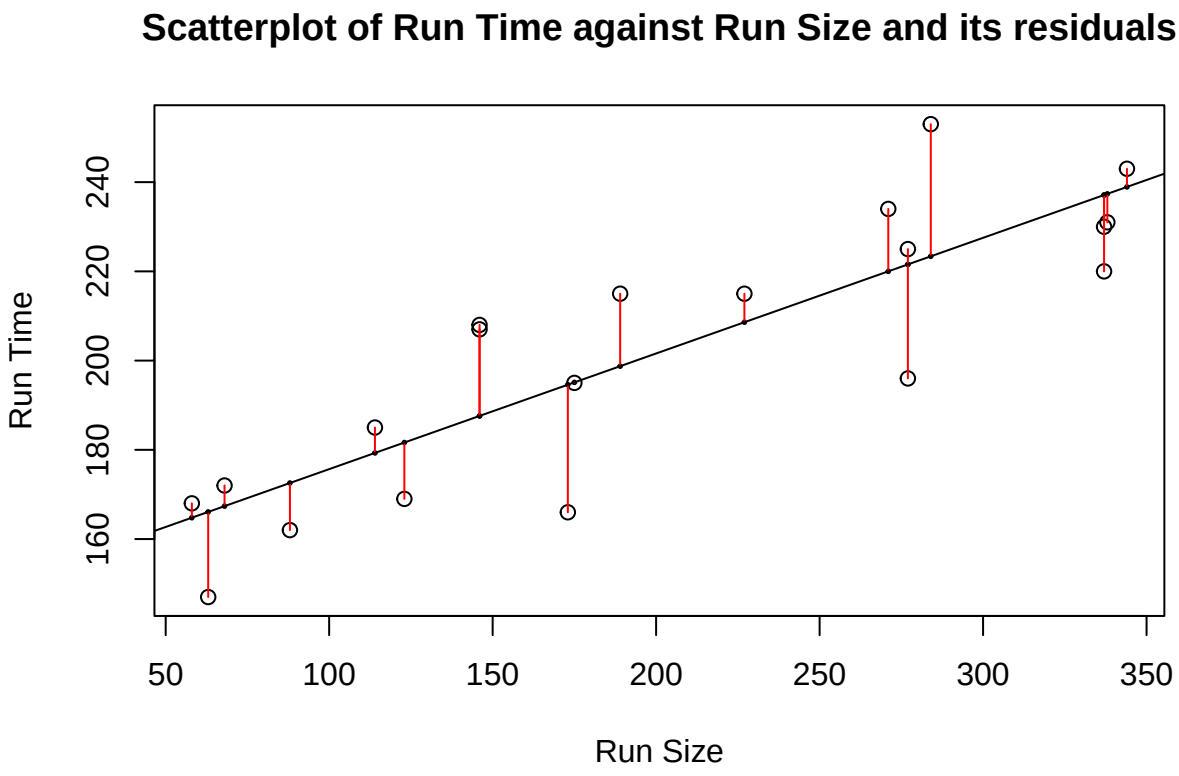
MAST7866 Linear Regression

Computing Session 3

For this session, we will use the data sets that you used in Computing Sessions 1 and 2 (`production.txt`, `SaltBP.txt`, `diamonds.txt`) and a new data set `glakes.txt`.

1 Production runs data

In the previous sheets, you looked at estimation of the linear regression model and model-checking using diagnostics. This sheet will look at other output for the `lm` function, confidence and prediction intervals. Try to reproduce the results for the production runs data from the slides, specifically the following plot. Check the function `segments` to produce the red segments.



2 The effect of salt on blood pressure

In the previous computer class sheet, you looked at the following data. 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. Use a fitted linear regression model where BP is the dependent variable and salt is the independent variable to answer the following questions:

(a) i. What are the values of the t-statistic and F-statistic to decide whether salt has an effect on blood pressure?

ii. What is the null hypothesis tested using the statistics in the previous part of this question?

iii. What is the result of the test?

iv. What conclusion can you draw from the test result?

(b) Find a 95% confidence interval for the effect of salt on blood pressure.

(c) Construct a confidence interval for the average blood pressure of people who have a salt intake of six grams per day.

(d) You are assisting a doctor. A patient has a salt intake of six grams per day.

i. Construct a 95% prediction interval for this patient's blood pressure.

ii. The patient has a blood pressure of 138. Is this an unusual blood pressure for a patient with this level of salt intake?

(e) What is the value of R^2? What does this indicate about the fit of the model?

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 questions:

(a) Use the summary function to decide whether there is evidence that the effect of the size on the price is different to zero.

(b) What is the value of R^2? What does this indicate about the fit of the model?

(c) The linear regression model predicts that the price of a 0.25 carat diamond ring is 670.63 and the price of a 0.15 carat diamond ring is 298.53. Find the prediction interval for 0.15 and 0.25 carat diamond rings.

(d) You see a 0.25 carat diamond ring available for a price of 580. Do you consider this ring to be good value?

(e) You can build a regression model with only α (the intercept) using the command

```
fit_diamonds <- lm(Price ~ 1, data = diamonds)
```

Use this model for prediction of the prices of a 0.15 carat ring and a 0.25 carat ring and compare to your answer to part (c).

4 Port management

This example involves the management at a Canadian port. The port management want to estimate the relationship between the volume of a ship's cargo loaded and unloaded at the port (Tonnage) and the time required (Time). Understanding this relationship will allow the management to better plan and to make comparisons to the time taken at other ports. The data are available in the file `glakes.txt`.

(a) Fit the linear regression model

$Time = \alpha + \beta * Tonnage + \varepsilon$

What are the estimates of α and β?

(b) Consider the residuals and a scatterplot of Time against Tonnage (with a fitted regression line). What do these graphs indicate about the suitability of the linear regression model?

(c) What is the value of R^2? What does this indicate about the fit of the model?

(d) If a ship has a tonnage of 8000, what does your model predict about the length of time taken to load and unload this cargo?