STAT2008/STAT2014/STAT6014

Tutorial 2

Question 1. The data file **Lubricant.csv** (available on Wattle) contains 53 measurements of the viscosity of a particular lubricating agent at various temperatures and pressures. The names of the three variables in the data are **viscos, pressure** and **tempC**. At the end of this question, remember to save the related R code as we will be using it again in next tutorial.

(a) Use **lm()** to perform a simple linear regression with viscosity as the response and pressure as the predictor variable. What are the least-squares estimates of the slope and intercept?

(b) Plot viscosity against pressure and use **abline()** to superimpose the estimated regression line. Use the estimated coefficients of the regression line to predict what the viscosity of the lubricant would be at a pressure of 1,000? Also predict what the viscosity of the lubricant would be at a pressure of 10,000? Locate these predictions on your plot and comment on whether or not they appear to be sensible predictions.

(c) Use R to find the means of both pressure and viscosity and check that together the two means form a point (called the centroid of the data) which is located on the estimated regression line.

Question 2. Show the following equations:

(a) $\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})(Y_i - \hat{Y}_i) = 0$.

(b) $E(b_1) = \beta_1$ and $Var(b_1) = \frac{\sigma^2}{S_{xx}}$.

(c) $E(b_0) = \beta_0$ and $Var(b_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{S_{xx}} \right]$.