DEPARTMENT OF STATISTICS

EKT 720 - Statistical learning

QUESTION 1

1. Study Draper and Smith, Chapter 26
2. Use the data as attached to determine/estimate the distribution of the estimated simple regression parameters by bootstrapping the
   (a) Pairs (X,Y)
   (b) Residuals
3. Determine a 95% confidence interval for the true value of R^2.

   For Questions 2 and 3 – Study the attached programmes, ensure that you have a full understanding of these programmes

QUESTION 2

Consider the dataset as attached in the file: cdata.sas7bdat (or cdata.xls). The following model needs to be fitted to the data:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 (X_i - X_1^*) D_{1i} + \beta_3 (X_i - X_2^*) D_{2i} + u_i$$

where
$$\begin{aligned} D_{1i} &= 1 \quad if \quad X_i > X_1^* \\ D_{1i} &= 0 \quad if \quad X_i \leq X_1^* \\ D_{2i} &= 1 \quad if \quad X_i > X_2^* \\ D_{2i} &= 0 \quad if \quad X_i \leq X_2^* \end{aligned}$$

(a) Make a graphical representation of the data.

(b) Use a value search to determine the structural break points, X_1^* and X_2^*, if they exist.

(c) Use a bootstrap approach to determine a $(1-\alpha)\%$ confidence interval for the values of X_1^* and X_2^*.

(d) Give estimates of the pdf of X_1^* and X_2^* respectively.