STAT 353 (Fall 2022)

Assignment 5

Note: Write up your solution carefully for each question. For data analysis questions, you also need to submit your R code and related output.

   Submit your assignment solution in Brightspace using one pdf file for each assignment.

Due: December 5, 8:30am

1. [10 marks] Suppose we wish to fit the piecewise quadratic polynomial with a knot at $x = t$,

$$E(y) = S(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 (x - t)^0_+ + \beta_4 (x - t)^1_+ + \beta_5 (x - t)^2_+,$$

where function $(x - t)_+ = max\{0, x - t\}$, and t is given.

   (a) Show how to test the hypothesis that this quadratic spline model fits the data significantly better than an ordinary quadratic polynomial.

   (b) The quadratic spline model is not continuous at the knot t. How can the model be modified so that continuity at $x = t$ is obtained?

   (c) Show how the model can be modified so that both $E(y)$ and $dE(y)/dx$ are continuous at $x = t$.

2. [5 marks] An operations research analyst is investigating the relationship between production lot size x in units and the average production cost per unit y. A study of recent operations provides the following data:

x	100	120	140	160	180	200	220	240	260	280	300
y (\$)	9.73	9.61	8.15	6.98	5.87	4.98	5.09	4.79	4.02	4.46	3.82

   The analyst suspects that a piecewise linear regression model should be fit to these data. Estimate the parameters in such a model assuming that the slope of the line changes at $x = 200$ units and $E(y)$ is continuous at $x = 200$. Do the data support the use of this model?

3. [10 marks] Consider the house price data in data-table-B4.csv in Brightspace, where there are 9 independent variables $x_1, \ldots, x_9$, and the dependent variable y is the house price. The variable definition is given on page 557 (in the 5th edition).

(a) Build an appropriate regression model for y using the stepwise regression in R with the AIC criterion. State the fitted model.

(b) Perform the residual analysis for the model in part (a), and comment on the model assumptions.

(c) Interpret the parameters in the model in part (a).