

Homework 3 – Due Feb 3rd @ 11pm

1. Exercise 3.4.1

The data file *airfares.txt* on the book web site gives the one-way airfare (in US dollars) and distance (in miles) from city A to 17 other cities in the US. Interest centers on modeling airfare as a function of distance. The first model fit to the data was

$$\text{Fare} = \beta_0 + \beta_1 x + e \quad (3.7)$$

(a) Based on the output for model (3.7) a business analyst concluded the following:

The regression coefficient of the predictor variable, Distance is highly statistically significant and the model explains 99.4% of the variability in the Y-variable, Fare. Thus model (1) is a highly effective model for both understanding the effects of Distance on Fare and for predicting future values of Fare given the value of the predictor variable, Distance.

Provide a detailed critique of this conclusion.

Note: Focus on the model assessment. Use both numerical and graphical output for the critique.

(b) Does the ordinary straight line regression model (3.7) seem to fit the data well? If not, carefully describe how the model can be improved. Given below and in Figure 3.41 is some output from fitting model (3.7).

Note: Focus on improving the current model. If you have found the way to improve the model, try to fit the alternative model and show how it improves the fit. You may try several alternative models and choose the best fitted one for your model of choice.

2. Exercise 3.4.5

3. Exercise 3.4.8