StuDocu.com

Linear Statistical Models Assignment 2

Linear Statistical Models (University of Melbourne)

StuDocu is not sponsored or endorsed by any college or university
Downloaded by Michael Le (mikele0217@gmail.com)

MAST30025: Linear Statistical Models

Assignment 2, 2020

Due: 5pm Friday May 15 (week 8)

- This assignment is worth 7% of your total mark.
- You may use R for this assignment, including the `lm` function unless specified. If you do, include your R commands and output.
- Your assignment must be submitted on Canvas LMS as a single PDF document only (no other formats allowed). You may choose to either typeset your assignment or handwrite and scan it to produce an electronic version. The LMS will not accept late submissions. It is your responsibility to ensure that your assignments are submitted correctly and on time, and problems with online submissions are not a valid excuse for submitting a late or incorrect version of an assignment.
- Your assignment must clearly show your name and student ID number, your tutor's name and the time and day of your tutorial class. Your assignment must be submitted in the correct format and the correct orientation. Your answers must be clearly numbered and in the same order as the assignment questions.

1. Consider a general full rank linear model $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with $p > 2$ parameters. Derive an expression for a joint $100(1-\alpha)\%$ confidence region for parameters β_i and β_j, where i and j are arbitrary.

2. An experiment is conducted to estimate the annual demand for cars, based on their cost, the current unemployment rate, and the current interest rate. A survey is conducted and the following measurements obtained:

Cars sold ($\times 10^3$)	Cost ($\$k$)	Unemployment rate (%)	Interest rate (%)
5.5	7.2	8.7	5.5
5.9	10.0	9.4	4.4
6.5	9.0	10.0	4.0
5.9	5.5	9.0	7.0
8.0	9.0	12.0	5.0
9.0	9.8	11.0	6.2
10.0	14.5	12.0	5.8
10.8	8.0	13.7	3.9

For this question, you may not use the `lm` function in R.

(a) Fit a linear model to the data, and estimate the parameters and error variance.

(b) Calculate 95% confidence intervals for the model parameters.

(c) In a year with 8% unemployment rate and 3.5% interest rate, we price a car at $\$12,000$ and observe that 7,000 cars are sold. Is this an atypical year (according to your model)?

(d) Using your answer from question 1, find and draw a joint 95% confidence region for the parameters corresponding to unemployment rate and interest rate. Superimpose a rectangle corresponding to the confidence intervals found in (b).

(e) Do you expect the confidence region to be larger or smaller than the rectangle? Justify your answer.

This document is available free of charge on StuDocu.com

Downloaded by Michael Le (mikele0217@gmail.com)

(f) (Bonus) What is the probability that the true parameters for unemployment rate and interest rate (jointly) lie in the rectangle you drew in (d)?

3. Show that for a full rank linear model with p parameters, the Akaike's information criterion, defined as $-2\log(\text{Likelihood}) + 2p$, can be written as

$$n \log\left(\frac{SS_{Res}}{n}\right) + 2p + \text{const.}$$

4. For this question we use the data set `UCD.csv` (available on the LMS). This data set, collected on 158 UC Davis students (self-reported), includes the following variables:

ID = the ID for that student

alcohol = average number of alcoholic drinks consumed per week

exercise = average hours per week the student exercises

height = the student's height (in inches)

male = indicator variable, 1 if male and 0 if female

dadht = the student's father's height

momht = the student's mother's height

We seek to predict a person's height, based on the given data.

(a) Fit a linear model using all of the variables (except ID).

(b) Test for model relevance, using a corrected sum of squares.

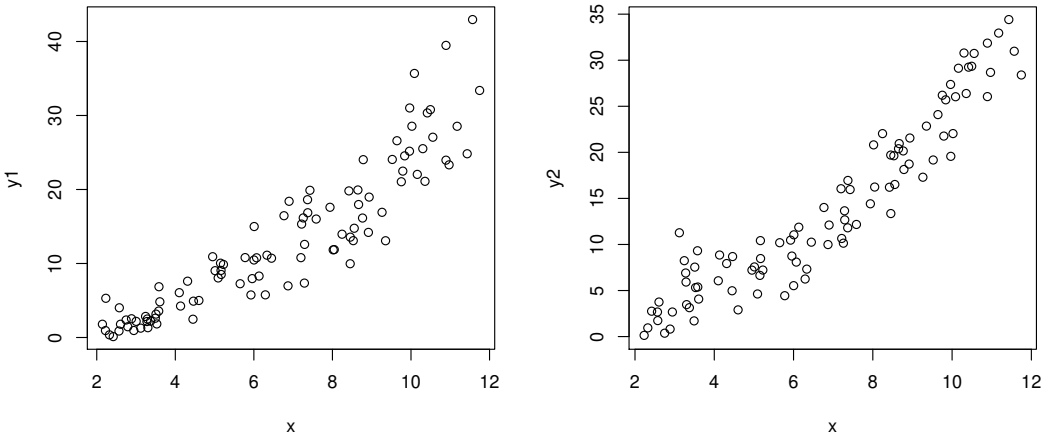
(c) Use forward selection with F tests to select variables for your model.

(d) Starting from a full model, use stepwise selection with AIC to select variables for your model. Use this as your final model; comment briefly on the variables included.

(e) Using your final model, test whether the parameters corresponding to father's and mother's heights are equal.

(f) Comment on the suitability of your final model, using diagnostic plots.

5. Suppose that we have a response variable y which is known to have a quadratic relationship with a predictor variable x. Explain *all* of the differences between fitting a linear model of y against x and x^2, versus a linear model of $\sqrt{y}$ against x. Which would you use for the two datasets shown below?



Downloaded by Michael Le (mikele0217@gmail.com)