Final Exams Papers of 2023-2022
(Statistical Modelling I)

Lubna Shaheen

Week 12, Revision

Question 1

Question 1 [50 marks]. The table below shows the amount of $CO2$ in the atmosphere (in parts per million) measured at the Mauna Loa Observatory in Hawaii in January of each year from 2011 to 2022 (x_i) and the average global surface temperature in the following year (y_i) where temperatures are expressed as a percentage of the value in 2001. A climate scientist fits a simple linear regression model to this data.

Year	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
x_i (ppm)	391	393	396	398	400	403	406	408	411	414	416	418
y_i (%)	113	120	126	139	167	189	170	157	181	189	157	167

The model to be fitted is

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where $i = 1, ..., 12$.

(a) What assumptions are usually made about the ϵ_i? **[3]**

You are given that for this data, $\sum x_i = 4{,}854$, $\sum y_i = 1{,}875$, $\sum x_i y_i = 760{,}359$ and $\sum x_i^2 = 1{,}964{,}356$.

(b) Find the least squares estimates of β_0 and β_1. **[6]**

Queen Mary University of London

Question 1

When the least squares estimates are used the total sum of squares is found to be 7,576.3 and the residual sum of squares is 3,532.3.

(c) Calculate the Coefficient of Determination. **[2]**

(d) Complete the Analysis of Variance Table for this model. **[9]**

(e) Use your Table in (d) above to estimate $var(\epsilon_i)$. **[2]**

(f) What hypothesis can be tested using the Table in (d) above? **[1]**

Question 1

d.f.1	1	1	2	2	10
d.f.2	10	11	11	12	1
$F(0.05)$	4.965	4.844	3.982	3.885	241.882

(g) Complete the test of hypothesis in (f) above at a 95% significance level. [**4**]

Final Exams Paper 2023

Question 1

In January 2023 the amount of CO_2 in the atmosphere was measured as 419 parts per million. You are given that $S_{xx} = 913$ and $t_{0.025;10} = 2.228139$.

(h) Calculate a point estimate for the global surface temperature in 2023 as a percentage of the 2001 temperature using the regression parameters. **[2]**

(i) Calculate a 95% confidence interval for the global surface temperature when atmospheric CO_2 measures 419 parts per million. **[4]**

Queen Mary University of London

Final Exams Paper 2023

Question 1

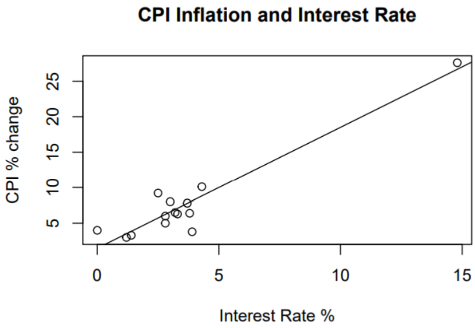
(j) Calculate a 95% prediction interval for the global surface temperature in 2023. [**4**]

(k) Comment on your answers to (h), (i) and (j) above. [**4**]

(l) Explain how residual plots can be used to check the assumptions referred to in (a) above. You should include what is to be plotted, what assumption each plot checks and what type of output to look for. [**9**]

Question 2

Question 2 [18 marks]. An economist wishes to analyse the relationship between interest rates and CPI inflation in different countries. They plot observations from 14 countries and fit a simple linear regression model as shown by the plot below.



(a) Based on the evidence of this plot, comment on the suitability of this data set for a simple linear regression model. [4]

Final Exams Paper 2023

Question 2

(b) What is the average leverage of an observation in this data set? **[1]**

(c) Explain what each of the following lines of R code are doing. **[6]**

```
econ <- lm(y~ x)
di <- rstandard(econ)
vi <- hatvalues(econ)
Di <- cooks.distance(econ)
i <- 1:n
plot(i, vi)
```

The observation with the highest interest rate is Pakistan which has leverage of 0.89 and a Cook's Statistic of 7.98.

Queen Mary University of London

Question 2

(d) Explain carefully how you would evaluate whether this observation had significant influence on the linear regression results. Include all of the steps you would take. [**7**]

Final Exams Paper 2023

Question 4

Question 4 [18 marks]. The crop yield of oranges (y_i) harvested from a group of trees in an orange grove is measured in different years. A multiple linear regression model is fitted with three explanatory variables:

x_1: number of nights in the year when the temperature fell below 5 degrees
x_2: the longest drought period of the year in days
x_3: a score out of 20 given for the mineral content of the soil

The full model fitted is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i$$

(a) If there are just 6 years of observations so $i = 1, 2, ..., 6$, write out the full model in terms of vectors and matrices. [4]

With 15 years of observation data the model is fitted using R. The full model R^2 is 98.79% and the total sum of squares is 7,675.5.

Queen Mary University of London

Question 4

(b) Estimate the variance of the residuals in the full model. **[3]**

Because the full model R^2 is so large, the analyst wishes to consider a simpler two explanatory variable model. They fit each of the three possible combinations of variables and compute the following sums of squares of residuals for these three reduced models.

Model	Residual sum of squares
x1 + x2	863.8
x1 + x3	6,451.3
x2 + x3	700.7

Queen Mary University of London

Question 4

(c) Using the data from the full and reduced models and the critical values of the F distribution in the table below, determine whether any of the three explanatory variables could be deleted by a F-test at 95% significance showing all your working. **[11]**

s	1	2	1	2	1	2	11	11	12	12
t	11	11	12	12	13	13	1	2	1	2
$F_t^s(0.05)$	4.84	3.98	4.75	3.89	4.67	3.81	242.98	19.40	243.91	19.41

Question 1

Question 1 [16 marks]. Ten observations for an explanatory variable (x_i) and a response variable (y_i) where $i = 1, 2, ..., 10$ are given in the table below.

x_i	3.6	3.5	2.9	1.8	3.4	2.3	3.6	3.7	1.6	2.0
y_i	9.6	9.4	8.0	7.3	9.2	6.5	9.6	6.9	7.8	6.8

(a) State the three usual assumptions made in a simple linear regression model giving each in terms of the response variable. **[3]**

You are given that $\sum x_i = 28.4$, $\sum y_i = 81.1$, $\sum x_i y_i = 236.2$ and $\sum x_i^2 = 86.92$.

(b) Find the least squares estimates for the intercept (β_0) and slope (β_1) parameters. **[6]**

You are given that the standard error of the slope parameter estimator is 0.4074 and that $t_{0.025;n-2} = 1.8595$.

(c) Calculate a 95% confidence interval for β_1. **[4]**

(d) What does your answer in (c) above say about the hypothesis H_0: $\beta_1 = 0$? **[2]**

(e) State another way in which the same hypothesis could be tested. **[1]**

Queen Mary
University of London

LSR Exams Paper 2023

Question 2

Question 2 [20 marks]. A simple linear regression model is constructed from 32 observations where the mean of the explanatory variable observations is 3.7 and the mean of the response variable observations is 4.75. The following model is fitted

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

The least squares estimates of the parameters β_0 and β_1 are 9.2928 and -1.2269 respectively. The model R^2 is 63.07%.

(a) If the Residual Sum of Squares (SS_E) is 60.977 complete the Analysis of Variance Table for this model. **[9]**

(b) Use your table in (a) to estimate the variance of the residuals. **[1]**

A new x_i observation is taken with a value of 5. The value of the response that corresponds to this new x_i is not known.

(c) Calculate a 95% prediction interval for the new value of the response variable if S_{xx}= 69.18 and $t_{0.025;n-2} = 1.812$. **[7]**

(d) If another 16 (x_i, y_i) observations were obtained and the regression model re-run with these in addition to the initial 32 observations, what would be the impact on the width of prediction intervals such as those in (c) above? **[3]**

LSR Exams Paper 2023