STAT 40001/ MA 59800 **Statistical Computing** **Fall 2017**

Homework 5

Name:

Due : November 14, 2017 PUID:

Instruction: Please submit your R code along with a brief write-up of the solutions. Some of the questions below can be answered with very little or no programming. However, write code that outputs the final answer and does not require any additional paper calculations.

Q.N. 1) The mammals data set in the *MASS* package records brain size and body size of 62 different mammals.
a) Fit a regression model to describe the relation between brain size and body size.
b) Calculate the **95%** confidence interval for the slope parameter of the model.
c) Calculate the **90%** confidence interval for the slope parameter of the model.
d) Display a residual plot using the plot method for the results of the lm function.
e) Which observation(mammal) has the largest residual in your fitted model?

Q.N. 2) The data set *cars* is one of the data sets installed with R and is available in base package. The data set contains 50 observations of speed(mph) and dist(stopping distance in feet).
a) Display the data using scatter plot.
b) Fit a simple regression model using speed as a predictor variable.
c) Add the fitted line to the scatter plot.
d) Calculate the residuals and fitted values and print only first five observations of the residuals and fitted values.
e) Create a scatter plot of the residuals and fitted values.
f) Assuming that no intercept model is appropriate fit a simple linear regression model.
g) Calculate and compare the coefficient of determination for both with intercept and no-intercept models.

Q.N. 3) In the manufacture of commercial wood products, it is important to estimate the relationship between the density of a wood product and its stiffness. A relatively new type of particleboard is being considered that can be formed with considerably more ease than the accepted commercial product. It is necessary to know at what density the stiffness is comparable to that of the well-known, well-documented commercial product. A study was done by Terrance E. Conners, Investigation of Certain Mechanical Properties of a Wood-Foam Composite (M.S. Thesis, Department of Forestry and Wildlife Management, University of Massachusetts). Thirty particleboards were produced at densities ranging from roughly 8 to 26 pounds per cubic foot, and the stiffness was measured in pounds per square inch. Table below shows the data.

Density: 9.50, 9.80, 8.30, 8.60, 7.00, 17.40, 15.20, 16.70, 15.00, 14.80, 25.60, 24.40, 19.50, 22.80, 19.80, 8.40, 11.00, 9.90, 6.40, 8.20, 15.00, 16.40, 15.40, 14.50, 13.60, 23.40, 23.30, 21.20, 21.70, 21.30

Stiffness:14814, 14007, 7573, 9714, 5304, 43243, 28028, 49499, 26222, 26751, 96305, 72594, 32207, 70453, 38138, 17502, 19443, 14191, 8076, 10728, 25319, 41792, 25312, 22148, 18036, 104170, 49512, 48218, 47661, 53045

a) Import the data in R and display it graphically.
b) Fit a simple linear regression model by choosing appropriate response variable and regressor variable.

c) Perform the residual analysis and comment on the appropriateness of the model.
d) Use Box-Cox transformation and check whether transformation would improve the model.

Q.N. 4) The electric power consumed each month by a chemical plant is thought to be related to the average ambient temperature $\boldsymbol{x_1}$, the number of days in the month $\boldsymbol{x_2}$, the average product purity $\boldsymbol{x_3}$, and the tons of product produced $\boldsymbol{x_4}$. The past years historical data are available and are presented in the following table.

y	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	$\boldsymbol{x_3}$	$\boldsymbol{x_4}$
240	25	24	91	100
236	31	21	90	95
290	45	24	88	110
274	60	25	87	88
301	65	25	91	94
316	72	26	94	99
300	80	25	87	97
296	84	25	86	96
267	75	24	88	110
276	60	25	91	105
288	50	25	90	100
261	38	23	89	98

a) Fit a multiple linear regression model using these data set.
b) Predict the power consumption for a month with $\boldsymbol{x_1 = 75, x_2 = 24, x_3 = 90}$ and $\boldsymbol{x_4 = 98}$.
c) Construct a 95% confidence interval and 95% prediction interval for a month with $\boldsymbol{x_1 = 75, x_2 = 24, x_3 = 90}$ and $\boldsymbol{x_4 = 98}$.

Q.N. 5) An author maintains a website on a particular book and using Google Analytics, records the number of visits on this particular website on each day of the year. As expected there are more hits during weekdays then on weekends. Since the book is used as a textbook for a statistics course there are more hits during the time when the classes are in session. Table below provides the data for 35 weeks from April through November 2009. To explore the week by week visit patterns of these
a) Display the data using a scatterplot.
b) Calculate the rank correlation coefficient to measure the association between the week and the number of hits on the website.
c) Test for the significance of the correlation at **0.05** level.

Week	Hits
1	148
2	148
3	157
4	112
5	125
6	155
7	154
8	135
9	140
10	164
11	154
12	138
13	129
14	131
15	113
16	124
17	119
18	110
19	166
20	105
21	132
22	132
23	144
24	152
25	152
26	166
27	161
28	168
29	170
30	179
31	154
32	136
33	147
34	151
35	188