STAT 40001/MA 59800   Statistical Computing/ Computational Statistics   Fall 2013
Homework 6

Name:

Due : November 7, 2013   PUID:

Instruction: Please submit your R code along with a brief write-up of the solutions (do not submit raw R codes with Errors!). 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) 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

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

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.

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 scatteplot.

b) Fit a simple regression model using speed as a predictor variable.

c) Add the fitted line to the scatterable.

d) Calculate the residuals and fitted values and print only first five observations of the residuals and fitted values.

e) Create a scatterable 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 the with intercept and no-intercept models.

h) Using your fitted model predict the stopping diatance for a car with an speed of 21 mph.

Q.N. 3) The mammals dataset 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. 4) The mammals dataset in the MASS package records brain size and body size of 62 different mammals.

a) Display a scatterplot of the log(brain) vs. log(body).

b) Fit a simple linear regression model to the transformed data.

c) What is the equation of the fitted model.

Q.N. 5) A marketing researcher studied annual sales of a product that had been introduced 10 years ago. The data are as follows, where x is the year coded and y is the sales in thousands of units:

x_i:	0	1	2	3	4	5	6	7	8	9
y_i :	98	135	162	178	221	232	283	300	374	395

a) Prepare a scatter plot of the data. Does a linear relation appear adequate?

b) Use Box-Cox procedure to find an appropriate transformation of **y**.

c) Plot the estimated regression line for the transformed data.

d) Obtain the residuals and plot them against the fitted values. Also prepare a normal probability plot. What do your plot show?

e) Express the estimated regression function in the original units.