Flipped Assignments

Flipped Assignment 1 (1/18)

Do the following problem by hand (upload a scanned pdf file). Consider the following data:

x	y
1	4
2	2
3	3

a. Plot the points on a graph by hand. Eyeball a straight line that you believe fits the points "best"
b. Fit a straight line to the data, f(x)=mx+b, using least squares. Sketch the fitted curve on the same graph created in part a. Comment on any similarities or dissimilarities between these lines.
c. Calculate SST
d. Calculate SSE
e. Calculate SSR
f. Sketch the errors (total, error, regression) for each of the data points on the graph.

Flipped Assignment 2 (1/25)

Consider the following data and assume that the response variable Y is Normally distributed. Upload a pdf of your handwritten work in response to this assignment

x	y
1	1.5
2	2
3	3
3	4

4	4.5
4.5	6

a. Fit a simple linear regression model to this data using least squares
b. What is your estimate of E[Y|X]?
c. What is your estimate of Var[Y|X] = Var[Y]?
d. Test the hypothesis that $\beta_1=0$ versus the alternative that it is not at the .05 level of significance
e. Interpret your findings, is the regression significant?
f. Calculate a 95% confidence interval on the mean response when x=3
g. Calculate a 95% prediction interval on a new observation when x=3

Flipped Assignment 3 (1/27)

Attached Files: Problem2_13.docx (557.919 KB)

Consider problem 2.13 from your book.

- Regress Days on Index using simple linear regression, what are the estimates of your fitted regression line?
- What is the value of R^2?
- Test for the signifiance of the regression at a 0.05 level of signficance assuming the response is Normally distributed, what is your conclusion?
- Regardless of whether you conclude that the regression is signficant above, make a scatterplot of the data showing the fitted regression line, confidence interval, and prediction interval
- Calculate a 95% confidence interval on the mean number of days the ozone level exceeds 20ppm when the meterological index is 17.0. Comment on the meaning of this interval?
- Calculate a 95% prediction interval on the mean number of days the ozone level exceeds 20ppm when the meterological index is 17.0. Comment on the meaning of this intervall? Compare the width of the prediction interval to that of the confidence interval and comment

Flipped Assignment 4 (2/8)

Attached Files: Problem2_13.docx (95.074 KB)

Consider problem 2.13 in your text book. You may use R to do your calculations, but you need to additionally generate a report in pdf/Word using the output from R with appropriate comments. Upload your RScript+Word Document or RDM+Word Document or RMD+pdf

a. Make a scatterplot of the data, Title the plot and label the axes appropriately
b. What are the Least Squares estimates of the parameters of a simple linear regression?
c. Add the least squares line to the scatterplot.
d. Check for model adequacy using diagnostic plots. Show the plots and comment on the assumptions of Constant Variance, Normality, and whether there appear to be outliers
e. Does the regression appear to be signficant? Why or why not?

Flipped Assignment 5 (2/10)

Consider problem 5.3 in your text book, in which the Number of Bacteria is regressed on Minutes of Exposure. The data is reported as follows.

Number of Bacteria	Minutes of Exposure
175	1
108	2
95	3
82	4
71	5
50	6
49	7
31	8
28	9
17	10
16	11
11	12

Upload a commented R file that answers the following questions.

a. Fit a simple linear regression model to the data. What Is the value of R^2?
b. Check for model adequacy (comment)
c. Use Box-Cox to perform a power transformation, transform the data as appropriate
d. Fit a simple linear regression model to the transformed data. What Is the value of R^2?
e. Check for model adequacy using the transformed data (comment)
f. Estimate the number of bacteria at 10 minutes of exposure, how does this compare with the observed value? (note, the transformed estimate from the fitted regression line must be transformed back)
g. Provide a 95% prediction interval on the number of bacteria at 10 minutes

of exposure.

Flipped Assignment 6 (2/15)

Attached Files: data-table-B8.csv (453 B)

Do problem 3.12 in the book **without using the LM function in R**. The corresponding data file is data-table-B8.csv is attached. Specifically, the problem is regressing clathrate formation (y - mass%) on the amount of surfactant (x1 - mass %) and time (x2 - min).

a. Write the model equation of clathrate formation regressed on both surfactant and time.
b. Fit the model in R without using the LM function. What is the dimensionality of Y? What is the dimensionality of X?
c. What are the least squares estimates of the regression parameters in the model?

Flipped Assignment 7 (2/17)

Attached Files: data-table-B8.csv (453 B)

Do problem 3.12 in the book **using the LM function in R**. The corresponding data file is data-table-B8.csv is attached. Specifically, the problem is regressing clathrate formation (y - mass%) on the amount of surfactant (x1 - mass %) and time (x2 - min).

a. Fit the model in R using the LM function.
b. What are the least squares estimates of the regression parameters in the model?

Flipped Assignment 8 (2/22)

Attached Files: data-table-B9.csv (1.548 KB)

Do problem 3.13 in the book **using the LM function in R**. Specifically, the problem studies the effect of four factors (x1 = superficial fluid velocity of the gas (cs/s), x2= kinematic velocity, x3= mesh opening (cm), and x4= dimensionless number relating the superficial velocity of the gas to the fluid) on the pressure drop in a screen plate bubble column (y). The corresponding data file data-table-B9.csv is attached.

a) Consider a first order multiple regression model with two-factor interactions.

Check for model adequacy and make any corrective actions if deemed necessary. Test for the signifcance of the full regression model, what do you conclude?
b) Test for the signifiance of all 2 factor interactions using a partial F-test. What are your findings?
c) Determine the best fitting model using partial F and/or t-tests. What is the final model?
d) Using the model from part c), calculate a 95% confidence interval on the mean response at the following points of interest (note, if a variable in the points of interest are not in the model, then it is omitted).
e) Using the model from part c), calculate a 95% prediction interval on the mean response at the points of interest (note, if a variable in the points of interest are not in the model, then it is omitted).

Points of Interest			
X1	**X2**	**X3**	**X4**
5.0	10.0	0.5	0.75
10.0	3.0	0.25	0.85

Flipped Assignment 9 (2/24)

Attached Files: data-table-B8.csv (453 B)

Do problem 3.12 in the book **using the LM function in R**. The corresponding data file is data-table-B8.csv is attached. Specifically, the problem is regressing clathrate formation on the amount of surfactant and time. Answer the following ...
a) Fit a first order multiple linear regression model with interactions relating calthrate formation to sufactant and time
b) Check for model adequacy, peform any transformation and repeat part a) if deemed necessary.
c) Test for the significance of the regression (using ANOVA). What do you conclude?
d) Test for the signifcance of the regression parameters, eliminating those that are deemed not signficant. What is your final model?

Flipped Assignment 10 (3/1)

Attached Files: data-SoftDrinkDeliveryTime.csv (428 B)

The file data-SoftDrinkDeliveryTime.csv contains data on the time that it takes a driver to deliver soft drinks to different locations. Regress delivery time on the

number of cases and the distance walked, including their second order interaction. Please answer all questions in a commented R file.

1. What are your estimates of the regression parameters and what is the associated value of R^2?
2. What do you notice in the diagnostic plots (all of them)
3. Remove the observation(s) that appears to be the most influential
4. What are your estimates of the regression parameters what is the associated value of R^2?
5. What do you notice in the diagnostic plots (all of them) after removing this point?
6. Is there now another point that might be influential or that has leverage?
7. Do you believe that this point should have be removed? (explain)

Flipped Assignment 11 (3/24)

Attached Files: data-table-B21.csv (276 B)

Consider the data file *data-table-B21*, which contains the response y and four predictor variables. We are interested in determining the best first order model for this data. (**Use the build in R commands to do this problem**).

a. When you regress y on all four predictors, what do you notice about the p-value for the f-statistic and the t-tests for the individual regression coefficients?
b. What are the VIFs for the predictors in this model?
c. Which first order model do you think describes the response with interpretable regression parameters the "best", why?

Flipped Assignment 12 (3/29)

Attached Files: data-table-B8.csv (453 B)

Conside the datafile data-table-B8.csv. Specifically, the problem is regressing clathrate formation on the amount of surfactant and time. Answer the following ...
a) Fit a first order model and check the VIF for the fitted regression parameters
b) Fit a first order model with second order interactions and check the VIF for the fitted regression parameters
c) Fit a first order model with standardized predictor variables and check the VIF for the fitted regression parameters
d) Fit a first order model with second order interactions with standardized predictor variables and check the VIF for the fitted regression parameters
e) Comment on VIFs for parts a,b,c,d in the context on standardized and non-standardized variables.

Flipped Assignment 13 (3/31)

Attached Files: data-heartbeat-3.csv (1.774 KB)

Consider the attached dataset collected on heartbeats from IE 5344 students over the years.

Predictors

X1 = Continent of Birth (NAm-North America, SA-South America, AF-Africa, AS-Asia, EU-Europe, AUS-Australia)

X2 = Frequency of Exercise (1-Never, 2-Not often, 3 – Moderate, 4 – Fairly Regularly, 5 – Everyday)

X3 = Gender (M or F)

X4 = Hours of sleep each night

X5 = Early-Bird (0) or Night-Owl (1)

X6 = Sleep Schedule: Regular (reg) or Irregular (irr)

Response

Y = heartbeat (in beats per minute, bpm)

1. Create a model that regresses heartbeat on continent.
    a. Is continent a significant indicator?
    b. Interpret the regression parameters
2. Create a model that regresses heartbeat on exercise frequency, assuming the marginal response is fixed
    a. Is exercise frequency significant?
    b. Interpret the regression parameters
3. Create a model that regresses heartbeat on exercise frequency, assuming the marginal response varies
    a. Is exercise frequency significant?
    b. Interpret the regression parameters
4. Create a model considering all predictors as candidates. In this full model, consider only first-order effects. Working backwards, determine the model that "best" fits the data.
    a. Did you consider effects on the response to be marginal or fixed for X2? Why or why not?
    b. How much variation is described by the model? Interpret.
    c. Which factors are significant, is this surprising? Interpret those that are significant.

** As discussed in Lecture, please include the variable for Coffee in your analysis, thanks**

Flipped Assignment 14 (4/5)

Attached Files: data-table-B9.csv (1.548 KB)

The problem corresponding to data-table-B9.csv studies the effect of four factors (x1 = superficial fluid velocity of the gas (cs/s), x2= kinematic velocity, x3= mesh opening (cm), and x4= dimensionless number relating the superficial velocity of the gas to the fluid) on the pressure drop in a screen plate bubble column (y).

Fit all possible first order models to this dataset (note there will 2^k where k=4 possible models). Select the best model using available criterion (AIC, AdjR^2, etc.) and further analysis of candidate models. Justify your selection in a commented RScript or RMarkdown file.

Flipped Assignment 15 (4/7)

Attached Files: data-table-B9.csv (1.548 KB)

The problem corresponding to data-table-B9.csv studies the effect of four factors (x1 = superficial fluid velocity of the gas (cs/s), x2= kinematic velocity, x3= mesh opening (cm), and x4= dimensionless number relating the superficial velocity of the gas to the fluid) on the pressure drop in a screen plate bubble column (y).

Find the “best” fitting model using stepwise regression. Consider a linear function of the predictor variables and their two-factor interactions as candidates for inclusion.

a. Prepare the data prior to analysis. Why would centering/scaling the predictor variables make sense?
b. Use forward stepwise regression, which model is selected?
c. Use stepwise regression in both directions starting with a first order model, which model is selected?
d. Use backwards stepwise regression, which model is selected?
e. Do you find the same model in parts b, c, and d? Explain

Flipped Assignment 16 (4/12)

Attached Files: data-likelihood-exponential.csv (649 B)

The data file *data-likelihood-exponential.csv* contains 50 observations that were drawn from an exponential distribution. You may use R to support your calcuations (Hint: use the command dexp(x,rate=) in your calcuations). Upload a RMarkdown or commented R file.

Based on visual inspection of the data alone, pick three possible values for the rate lambda. Remember, rate=lambda=1/E[X]. These should just be reasonable guesses ... by inspection of the data, guess (don't calculate) three possible values for the expectation/average and convert these to lambda.

a. Considering the first 25 observations, which of the three values of your guessed values for lambda is the most likely? (explain)
b. Considering all 50 observations, which of the three values of your guessed values for is the most likely? (explain)
c. Were your guesses at lambda the same when considering 25 versus 50 observations?

Flipped Assignment 17 (4/21)

Attached Files: LogisticRegressionAssignment.docx (3.103 MB)

Please see the attached flipped assignment

Flipped Assignment 18 (4/26)

Attached Files: Problem_13_5.pdf (352.838 KB)

Please work the attached problem

Flipped Assignment 19 (4/28)

Attached Files: PoissonRegression.docx (134.322 KB)

Consider the attached assignment