MCP261 IE Lab I: March 10, 2021

Exercise 5: Analytics I: Linear and Logistic Regression

Due date: March 13, 11:59 PM

Solve problems 1-3 using Excel, and problem 4 using Python's *sklearn* package.

1. (2 marks) Dataset for this problem is available here: https://bit.ly/2FDvKG6

Copier maintenance. The Tri-City Office Equipment Corporation sells an imported copier on a franchise basis and performs preventive maintenance and repair service on this copier. The data below have been collected from 45 recent calls on users to perform routine preventive maintenance service; for each call, X is the number of copiers serviced and Y is the total number of minutes spent by the service person. Assume that first-order regression model (1.1) is appropriate.

i:	1	2	3	...	43	44	45
X_i:	2	4	3	...	2	4	5
Y_i:	20	60	46	...	27	61	77

a. Obtain the estimated regression function.

b. Plot the estimated regression function and the data. How well does the estimated regression function fit the data?

c. Interpret b_0 in your estimated regression function. Does b_0 provide any relevant information here? Explain.

d. Obtain a point estimate of the mean service time when $X = 5$ copiers are serviced.

2. (2 marks) Solve the following problem, with respect to the data from the previous problem.

a. Estimate the change in the mean service time when the number of copiers serviced increases by one. Use a 90 percent confidence interval. Interpret your confidence interval.

b. Conduct a t test to determine whether or not there is a linear association between X and Y here; control the α risk at .10. State the alternatives, decision rule, and conclusion. What is the P-value of your test?

c. Are your results in parts (a) and (b) consistent? Explain.

3. (3 marks)

*6.15. **Patient satisfaction.** A hospital administrator wished to study the relation between patient satisfaction (Y) and patient's age (X_1, in years), severity of illness (X_2, an index), and anxiety

level (X_3, an index). The administrator randomly selected 46 patients and collected the data presented below, where larger values of Y, X_2, and X_3 are, respectively, associated with more satisfaction, increased severity of illness, and more anxiety.

i:	1	2	3	...	44	45	46
X_{i1}:	50	36	40	...	45	37	28
X_{i2}:	51	46	48	...	51	53	46
X_{i3}:	2.3	2.3	2.2	...	2.2	2.1	1.8
Y_i:	48	57	66	...	68	59	92

The dataset is available here: https://goo.gl/K69YZC

(i) Fit a linear regression model for three predictor variables to the data and state the estimated regression function. How does the adjusted R-squared value compare to the R-squared value?

(ii) Test whether there is a regression relation between patient's age and patient satisfaction.

(iii) Test whether there is a regression relation at all between the three predictor variables and the dependent variable.

4. (3 marks) Dataset for this problem is available here: https://bit.ly/3qAT4cS

Toxicity experiment. In an experiment testing the effect of a toxic substance, 1,500 experimental insects were divided at random into six groups of 250 each. The insects in each group were exposed to a fixed dose of the toxic substance. A day later, each insect was observed. Death from exposure was scored 1, and survival was scored 0. The results are shown below; X_j denotes the dose level (on a logarithmic scale) administered to the insects in group j and $Y_{\cdot j}$ denotes the number of insects that died out of the 250 (n_j) in the group.

j:	1	2	3	4	5	6
X_j:	1	2	3	4	5	6
n_j:	250	250	250	250	250	250
$Y_{\cdot j}$:	28	53	93	126	172	197

Logistic regression model (14.20) is assumed to be appropriate.

a. Plot the estimated proportions $p_j = Y_{\cdot j}/n_j$ against X_j. Does the plot support the analyst's belief that the logistic response function is appropriate?

b. Find the maximum likelihood estimates of β_0 and β_1. State the fitted response function.

c. Obtain a scatter plot of the data with the estimated proportions from part (a), and superimpose the fitted logistic response function from part (b). Does the fitted logistic response function appear to fit well?

d. Obtain $\exp(b_1)$ and interpret this number.