Homework No.2 (MSDS 954:567) Spring 2022

Due date: 3/10/2022

Problem 1. We want to test the effect of light level and amount of water on the yield of tomato plants. Each potted plant receives one of three levels of light (1 = 5 hours, 2 = 10 hours, 3 = 15 hours) and one of two levels of water (1 = 1 quart, 2 = 2 quarts). The yield, in pounds, is recorded. The results are as follows:

Yield	Light	Water	Yield	Light	Water
12	1	1	20	2	2
9	1	1	16	2	2
8	1	1	16	2	2
13	1	2	18	3	1
15	1	2	25	3	1
16	1	2	20	3	1
16	2	1	25	3	2
14	2	1	27	3	2
12	2	1	29	3	2

Perform a multiple regression.

[*Remark: Use a computer for you calculation; explain your analysis and results carefully*]

Problem 2. The following data responses y are generated from a regular Poisson model with a single covariate variable x:

x	y
0.4	1
0.6	0
0.8	6
1.0	6
1.2	7
1.4	9

(a) Please write down the Poisson model for this data set, stating all requirements.

(b) Calculate the maximum likelihood estimates (MLE) $\hat{\beta}_0$ and $\hat{\beta}_1$ for β_1 and β_2, and then provide the variance estimates for $\hat{\beta}_0$ and $\hat{\beta}_1$.

(c) The values of the linear predictor are $0.37, 0.77, 1.16, 1.56, 1.95, 2.34$ for the 6 observations. Please compute the deviance residuals and draw the index plot of deviance residuals.

(d) Draw a partial residual plot to study the linearity of the covariate variable x (show your calculation).

(e) The following is an incomplete "Analysis of Deviance Table" for testing $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$. Please fill in the question marks. What is your conclusion of the test? Use the level of significance $\alpha = 0.05$.

Model	Degrees of Freedom (DF)	Deviance	Difference of DFs	Difference of Deviance
without x	5	?		
with x	4	?	1	?

(The probability (density) function for a Poisson distribution $y \sim \text{Poisson}(\mu)$ is $f(y|\mu) = e^{-\mu}\mu^y/y!$ for $y = 0, 1, 2, \ldots$.)

[*Remark: Solve this problem by paper, pencil and calculator.*]

Problem 3. Knight & Skagen (1988) collected the data shown in the table (and in data frame eagles) during a field study on the foraging behavior of wintering Bald Eagles in Washington State, USA. The data concern 160 attempts by one (pirating) Bald Eagle to steal a chum salmon from another (feeding) Bald Eagle. The abbreviations used are

L = large S = small A = adult I = immature

Number of successful attempts	Total number of attempts	Size of pirating eagle	Age of pirating eagle	Size of feeding eagle
17	24	L	A	L
29	29	L	A	S
17	27	L	I	L
20	20	L	I	S
1	12	S	A	L
15	16	S	A	S
0	28	S	I	L
1	4	S	I	S

Report on factors that explain the success of the pirating attempt and give a prediction formula for the probability of success.

[*Remark: Use a computer for you calculation; explain your analysis and results carefully*]

Problem 4. A marketing research firm was engaged by an automobile manufacturer to conduct a pilot study to examine the feasibility of using logistic regression for ascertaining the likelihood that a family will purchase a new car during the next year. A random sample of 33 suburban families was selected. Data on annual family income (X_1, in thousand dollars) and the current age of the oldest family automobile (X_2, in years) were obtained. A follow-up interview conducted 12 months later was used to determine whether the family actually purchased a new car ($Y = 1$) or did not purchase a new car ($Y = 0$) during the year.

i:	1	2	3	$\cdots$	31	32	33
X_{i1}:	32	45	60	$\cdots$	21	32	17
X_{i2}:	3	2	2	$\cdots$	3	5	1
Y_i:	0	0	1	$\cdots$	0	1	0

Complete dataset is provided in Stat567_hw2_problem4.txt

(a) Find the maximum likelihood estimates of β_0, β_1, and β_2. State the fitted response function.
(b) Obtain $\exp(\beta_1)$ and $\exp(\beta_2)$ and interpret these numbers.
(c) What is the estimated probability that a family with annual income of \$50 thousand and an oldest car of 3 years will purchase a new car next year?
(d) Obtain the deviance residuals and present them in an index plot. Do there appear to be any outlying cases?

(e) Construct a half-normal probability plot of the absolute deviance residuals. Do any cases here appear to be outlying?

[*Remark: Use a computer for you calculation; explain your analysis and results carefully*]

Problem 5. (This is a follow up question for homework 1, problem 3) Given the following data points:

-1.43 -0.95 -0.19 0.02 0.14 0.83 1.35 1.46 2.62

Compute the kernel density estimate $\hat{f}(x)$ at point $x = 0.05$. Use the rectangular kernel $K(t)$ with binwidth $h = 0.22$. Here, $K(t) = 1/2$ if $|t| \leq 1$, and it equals 0 if $|t| > 1$.

[*Remark: Solve this problem by paper and pencil.*]