**STAT34700 HW1**

**Problem 4**

**(a)**

**Model: Logistic Regression Model (Binomial)**

**Output:**

> sumary(oringsfit)

Estimate Std. Error z value Pr(>|z|)

(Intercept) 11.662990 3.296263 3.5382 0.0004028

temp -0.216234 0.053177 -4.0663 4.777e-05

n = 23 p = 2

Deviance = 16.91228 Null Deviance = 38.89766 (Difference = 21.98538)

**Fitted Model:**

**Comments:**

The summary coincides with the result shown in Faraway page 52 but the estimates here ( and ) are different from the estimates in the paper ( and ). The estimates of equation 3.2 are obtained using the data in table 1 columns Joint-Temperature and Erosion-or-blowby (Field).

In table 1, when the temperature is 53, the number of damage is recorded as 2 but in the orings data in R, the number of damage is 5. In addition, the orings data does not contain the data when temperature is 31. The number of damage when temperature is 75 are also different in table 1 and orings data in R. As a result, the difference of data used in the main reason of difference estimates.

**(b)**

**Model: Logistic Regression Model (Binary)**

**Summary:** The result of binary analysis using the data in R is the same as the result got from the paper.

The binary analysis: convert the binomial response to binary response, Y=1 if there is damage and Y=0 if there is no damage.

**Output:**

> sumary(fit\_binary)

Estimate Std. Error z value Pr(>|z|)

(Intercept) 15.04290 7.37863 2.0387 0.04148

temp -0.23216 0.10824 -2.1450 0.03196

n = 23 p = 2

Deviance = 20.31519 Null Deviance = 28.26715 (Difference = 7.95196)

**Fitted Model:**

**Comments:**

In the paper, and .

The estimated parameters of binary analysis using the data in R is the same as the estimates got from the paper.

**(c)**

Sparsity means that there are too many 0’s in the observed response values. Looking at the original binomial response *damage,* there are 44 zero’s and only 7 non-zero values out of total 51 response values. Therefore, there also existing sparsity problem in the original data.

**(d)**

We can use the retrospective sampling to remedy the sparsity problem in the binary analysis. In the original binary data, take all 7 data points where damage occur (orings$binary=1). Then

take a random sample of 7 points where damage does not occur (orings$binary=0). Use these 14 data points to fit the regression model. in this way, the sparsity problem can be remedied because there is same amount of 0’s and 1’s in the binary response.

**(e)**

**From the binary model:**

The fitted probability is the probability that there is event (damage) occurring in the six O-rings so the probability that there is no event (damage) in the six O-rings is .

**From the binomial model:**

The fitted probability is the probability that each one of the six O-rings has damage. Because each O-ring is independent of other five O-rings, the probability that there is no damage (event) in the six O-rings is .

As a result, we can get .

From the binary model we get in part (b),

Hence, the expected number of events (damages) is

, where x = temperature

**R Code**

**# Problem 3 ---------------------------------------------------------------**

install.packages("faraway")

library(faraway)

**#a**

data(orings)

oringsfit=glm(cbind(damage, 6-damage) ~ temp, family = binomial, orings)

sumary(oringsfit)

**#b**

orings$binary = orings$damage

orings[1,3]=1

fit\_binary = glm(binary ~ temp, family = binomial, data=orings)

sumary(fit\_binary)