Logistic Regression

STAT/BIOS 823

Homework 11

Directions

Using `RMarkdown` in `RStudio`, complete the following questions. Launch RStudio and open a new RMarkdown file or use the class RMarkdown template provided and save it on your working directory as a `.Rmd` file. At the end of the activity, save your **pdf** generated from `RMarkdown+Knitr` and submit your homework on the Blackboard.

Only question 1 is required. Question 2 is optional.

If you have questions, please post them on the **lesson discussion board**.

Some **R-codes** and **output** from the code have been provided for you. R codes and output must be clearly shown.

Homework submitted after the due date will attract a penalty of 10 points per day after the due date.

---

1 Logistic Regression Analyses

1. Use R to complete the following:
   (a) Plot the mean response function of a logistic regression model,

$$Pr(Y_i = 1) = \frac{exp(\beta_0 + \beta_1 X_i)}{1 + exp(\beta_0 + \beta_1 X_i)}$$

   when $\beta_0 = -25$ and $\beta_1 = 0.2$. **Hint:** generate values of X over the range $\sim 90 \leq X \leq \sim 160$, then plug X into the mean response function and use the `plot(x,y,type = "l")` command.

```
x <- seq(90, 160, 1)
b0 <- -25
b1 <- 0.2
y <- exp(b0 + b1 * x)/(1 + exp(b0 + b1 * x))
# ilogit is a function that does: exp(x)/(1+exp(x))
```

   (b) For what value of X is $Pr(Y) = 0.5$?

   (c) Find the odds:

$$\frac{Pr(Y = 1)}{1 - Pr(Y = 1)}$$

when $X = 150$ and when $X = 151$, and the ratio of the odds when $X = 151$ (numerator) to the odds when $X = 150$ (denominator). Is this odds ratio equal to $\exp(\beta_1)$ as it should be?

Optional 2. A marketing research firm was engaged by an automobile manufacturer to conduct a pilot study to examine the feasibility of using logistic regression for predicting whether 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. Use the attached dataset `Q2` to answer the following. Assume that a multiple logistic regression model with two predictor variables is appropriate:

(a) Fit the model and find the estimates of β_0, β_1, and β_2. Plot the estimated function over the data.

```
rm(list = ls(all = TRUE)[!grepl("global.var.A", ls(all = TRUE))])
q2 <- read.table("data/Q2.txt", quote = "\"", comment.char = "")
str(q2)
colnames(q2) <- c("y", "income", "age")
require(epiDisplay)
summ(q2)
mod2 <- glm(y ~ income + age, family = binomial(link = "logit"),
    data = q2)
summary(mod2)
library(epiDisplay)
```

(b) Find and interpret estimates of $\exp(\beta_1)$ and $\exp(\beta_2)$.

(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? Compute a 95% interval estimate for this probability.

```
# predicted probability for income = 50, age = 3
library(faraway)
ilogit(b0 + b1 * 50 + b2 * 3)  # Approach 1
# A 60.9% predicted chance of buying a car x0 <-
# c(1, 50, 3) eta0 <- sum(x0*coef(mod2))
# ilogit(eta0)
predict(mod2, newdata = data.frame(income = 50, age = 3),
    type = "response")
(pr1 <- predict(mod2, newdata = data.frame(income = 50,
    age = 3), se = T))
# predicted 95% interval for this probability
ilogit(c(pr1$fit - 1.96 * pr1$se.fit, pr1$fit + 1.96 *
    pr1$se.fit))
```

(d) Calculate the confidence intervals for $\exp(\beta_1)$ and $\exp(\beta_2)$ and interpret.

(e) Assess model goodness of fit. Explain your results.