Topics to be covered this week

Statistics 138
Fall Quarter, 2018

Monday, Nov 5 Logistic regression (Handout 10, chaps 4.1-4.3 (except Section 4.3.4), 5.2 in the text).

Wednesday, Nov 7 Multiple logistic regression (Handout 11, chaps 4.4 and 5.1 in the text).

Friday, Nov 9 Multiple logistic regression (Handout 11, chaps 5.1 and 6.1 in the text).

Homework 5 (Due on Monday, Nov 12)
You may form a group of three registered in this course and submit one completed homework for the group. The front page should be blank. Please write down the names of the students in the group on the submitted work.

1. For the dataset flu.xls, the columns of interest are shot (1 indicates flu shot, 0 indicates no flu shot) and age (age of the subject). We ignore the other columns at this time. It is desired to model the logit of π, the probability of getting the flu shot, as a linear function of age.
(a) Obtain the maximum likelihood estimates of β_0 and β_1. Write down the estimated logistic regression function.
(b) Plot the estimated probabilities (obtained from the fitted logistic model) against age. Does it seem that the probability of a flu shot depend on age?
(c) Using part (a), estimate the probability that a randomly chosen 55 year old will get the flu shot, along with a 95% confidence interval. [Obtain the confidence interval in the logit scale first using the R function 'predict', and then convert it.]
(d) Interpret the value of $\exp(\hat{\beta}_1)$.

2. Refer to Problem 1 above.
(a) Obtain an estimate of the age at which the probability of a flu shot is 0.5.
(b) Would it make sense to predict the probability of a flu shot of a 15 year old using the fitted model based on this data set? Does $\hat{\beta}_0$ have a practical meaning in this case? Explain your answers.
(c) Obtain a 95% confidence interval β_1. What does your interval say about testing if $\beta_1 = 0$?

3. A carefully controlled experiment was conducted to study the effect of the size of the deposit level on the likelihood that a returnable one-liter soft-drink bottle will be returned. A bottle return was scored 1, and no return was scored 0. The file BottleReturn.xls contains the number of bottles returned ($Y_{i\cdot}$) out of 100 sold (n_i) at each of six deposit levels (X_i, in cents).

An analyst believes that a logistic model is appropriate in modeling the probability of return as a function of deposit level.

(a) Plot the logit of the sample proportions $p_i = Y_{i\cdot}/n_i$ against X_i. Does the plot seem to support the analyst's belief?
(b) Find the maximum likelihood estimates of β_0 and β_1. Write down the estimated logistic function.
(c) Obtain a scatterplot of the data with the sample proportions from part (a) against deposit levels, and superimpose the fitted logistic probabilities from part (b). Does the fitted logistic model appear to fit the data well?
(d) Obtain the Pearson as well as the deviance residuals. Also obtain the standardized Pearson and standardized deviance residuals. Summarize your findings.

4. Refer to Problem 3 above.
(a) Estimate the probability that a bottle will be returned when the deposit is 15 cents.
(b) Estimate the amount of deposit for which 75% of the bottles are expected to be returned.
(c) Obtain a 95% confidence interval for β_1 and interpret it. Convert this interval into one for the odds ratio. Interpret this interval.
(d) Conduct a likelihood ratio test to determine whether deposit level X is related to the probability is returned at level $\alpha = 0.05$. Write down the null and the alternative hypotheses, the decision rule, and state your conclusion.

5. Refer to Problem 3 above.
(a) It is desired to carry out a likelihood ratio test to determine if a logistic linear model is appropriate for the data at level $\alpha = 0.05$. Write down the null and the alternative hypotheses, the decision rule, and state your conclusion.
(b) The analyst wanted to carry out a Pearson's chi-square test for the testing problem in part (c). Obtain the (estimated) expected number of bottle returns under the null hypothesis.
(c) Use part (b) to calculate the Pearson's chi-square statistic, carry out the test at level $\alpha = 0.05$, find the p-value of your test, and state your conclusion.