LAB 7: LOGISTIC REGRESSION

Problem 1. Bottle return. 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 data to follow show the number of bottles that were returned (Y.j) out of 500 sold (n j) at each of six deposit levels (X j, in cents):

j:	1	2	3	4	5	6
Deposit level X_j:	2	5	10	20	25	30
Number sold n_j:	500	500	500	500	500	500
Number returned $Y_{\cdot j}$:	72	103	170	296	406	449

An analyst believes that logistic regression model is appropriate for studying the relation between size of deposit and the probability a bottle will be returned.

a. Plot the estimated proportions $p_j = Y_j/n_j$ 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.

e. What is the estimated probability that a bottle will be returned when the deposit is 15 cents?

f. Estimate the amount of deposit for which 75 percent of the bottles are expected to be returned.

Problem 2. Flu shots. A local health clinic sent fliers to its clients to encourage everyone, but especially older persons at high risk of complications, to get a flu shot in time for protection against an expected flu epidemic. In a pilot follow-up study, 159 clients were randomly selected and asked whether they actually received a flu shot. A client who received a flu shot was coded Y = 1, and a client who did not receive a flu shot was coded Y = 0. In addition, data were collected on their age (X1) and their health awareness. The latter data were combined into a health awareness index (X2), for which higher values indicate greater awareness. Also included in the data was client gender, where males were coded X3 = 1 and females were coded X3 = 0.

i:	1	2	3	...	157	158	159
X_{i1}:	59	61	82	...	76	68	73
X_{i2}:	52	55	51	...	22	32	56
X_{i3}:	0	1	0	...	1	0	1
Y_i:	0	0	1	...	1	1	1

Multiple logistic regression model (14.41) with three predictor variables in first-order terms is assumed to be appropriate.

a. Find the maximum likelihood estimates of β0, β1, β2, and β3. State the fitted response function.

b. Obtain exp(b1), exp(b2), and exp(b3). Interpret these numbers.

c. What is the estimated probability that male clients aged 55 with a health awareness index of 60 will receive a flu shot?

d. Use the Wald test to determine whether X3, client gender, can be dropped from the regression model; use α = .05. State the alternatives, decision rule, and conclusion. What is the approximate P-value of the test?

-- END --