2019Fall STAT5107 Assignment 4

Department of Statistics, The Chinese University of Hong Kong

Due 9:30pm, Thursday, November 14, 2019

1. For a binary response variable Y and an explanatory variable X, let $\pi(x) = P(Y = 1|X = x) = 1 - P(Y = 0|X = x)$. The logistic regression model is

$$\pi(x) = \frac{exp(\alpha + \beta x)}{1 + exp(\alpha + \beta x)}.$$

When $\pi(x)$ is small, explain why one can interpret $exp(\beta)$ approximately as $\pi(x+1)/\pi(x)$.

2. A study for several professional sports used the model $logit(\pi) = \alpha + \beta logd$ to study the effect of a player's draft position d ($d = 1, 2, 3...$) (of selection from the pool of potential players in a given year) on the probability π of eventually being named an all star.

   a. Show that $\pi/(1 - \pi) = e^{\alpha}d^{\beta}$. Show that e^{α}=odds for the first draft pick.

   b. In the United States, Berry reported $\hat{\alpha} = 2.3$ and $\hat{\beta} = -1.1$ for probasketball and $\hat{\alpha} = 0.7$ and $\hat{\beta} = -0.6$ for pro baseball. This suggests that in basketball a first draft pick is more crucial and picks with high d are relatively less likely to be all-stars. Try to give some explaination.

3. For a study using logistic regression to determine characteristics associated with remission in cancer patients, the first Table below shows the most important explanatory variable, a labeling index (LI). This index measures proliferative activity of cells after a patient receives an injection of tritiated thymidine, representing the percentage of cells that are "labeled." The response Y measured whether the patient achieved remission (1=yes). Software reports the second Table below for a logistic regression model using LI to predict the probability of remission, where the model is:

$$logit(E[Y = 1|LI = x]) = logit(\pi(x)) = \alpha + \beta x$$

TABLE

LI	Number of Cases	Number of Remissions	LI	Number of Cases	Number of Remissions	LI	Number of Cases	Number of Remissions
8	2	0	18	1	1	28	1	1
10	2	0	20	3	2	32	1	0
12	3	0	22	2	1	34	1	1
14	3	0	24	1	0	38	3	2
16	3	0	26	1	1			

Source: Data reprinted with permission from E. T. Lee, *Comput. Prog. Biomed.* **4**: 80–92 (1974).

TABLE **Computer Output**

Criterion	Intercept Only	Intercept and Covariates
−2 Log L	34.372	26.073

Testing Global Null Hypothesis: BETA = 0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	8.2988	1	0.0040
Score	7.9311	1	0.0049
Wald	5.9594	1	0.0146

Parameter	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	−3.7771	1.3786	7.5064	0.0061
li	0.1449	0.0593	5.9594	0.0146

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
li	1.156	1.029	1.298

Estimated Covariance Matrix

Variable	Intercept	li
Intercept	1.900616	−0.07653
li	−0.07653	0.003521

Obs	li	remiss	n	pi_hat	lower	upper
1	8	0	2	0.06797	0.01121	0.31925
2	10	0	2	0.08879	0.01809	0.34010

(The fitted parameters intercept $\hat{\alpha}$, li $\hat{\beta}$, their standard error and other results are showed in above.)

a. Show how software obtained $\hat{\pi}$=0.068 when $LI = 8$.

b. Show that $\hat{\pi} = 0.5$ when $LI = 26.0$.

c. .Show that the rate of change in $\hat{\pi}$ is 0.009 when $LI = 8$ and 0.036 when $LI = 26$.

d. The lower quartile and upper quartile for LI are 14 and 28. Show that $\hat{\pi}$ increases by 0.42, from 0.15 to 0.57, between those values.

e. For a unit change in LI, show that the estimated odds of remission multiply by 1.16.

f. Please explain how to obtain the confidence interval reported for the odds ratio.

g. Construct a Wald test for the effect.And try to interpret it.

h. Conduct a likelihood-ratio test for the effect, showing how to construct the test statistic using the $-2logL$ values reported.

4. Table below is a $2 \times 2 \times 2$ contingency table-two rows, two columns, and two layers-from an article that studied effects of racial characteristics on whether persons convicted of homicide received the death penalty. The 674 subjects classified in the table were the defendants in indictments involving cases with muliple murders in Florida between 1976 and 1987. The variables are Y=death penalty verdict, having the categories (yes, no), X=race of defendant, Z=race of victims, each having the categories (white, black).

TABLE **Death Penalty Verdict by Defendant's Race and Victims' Race**

Victims' Race	Defendant's Race	Death Penalty Yes	Death Penalty No	Percent Yes
White	White	53	414	11.3
	Black	11	37	22.9
Black	White	0	16	0.0
	Black	4	139	2.8
Total	White	53	430	11.0
	Black	15	176	7.9

Source: M. L. Radelet and G. L. Pierce, *Florida Law Rev.* **43**: 1–34 (1991). Reprinted with permission from the *Florida Law Review*.

Below another Table shows the results of fitting a logit model, treating death penalty as the response (1=yes), treating defendant's race (1=white) and victims' race (1=white) as dummy predictors. The model is:

$$logit(E[Y = 1|X = x, Z = z]) = logit[\pi(x, z)] = \alpha + \beta_1 x + \beta_2 z$$

where α is intercept, β_1 and β_2 are parameters of defendant's race and victim's race respectively.

TABLE **Computer Output**

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value
Deviance	1	0.3798
Pearson Chi-Square	1	0.1978
Log Likelihood		−209.4783

Parameter	Estimate	Standard Error	Likelihood Ratio 95% Conf Limits		Chi-Square
Intercept	−3.5961	0.5069	−4.7754	−2.7349	50.33
def	−0.8678	0.3671	−1.5633	−0.1140	5.59
vic	2.4044	0.6006	1.3068	3.7175	16.03

LR Statistics

Source	DF	Chi-Square	Pr > ChiSq
def	1	5.01	0.0251
vic	1	20.35	<.0001

a. Please interpret parameter estimates. Which group is most likely to have the yes response? Find the estimated probability in that case.

b. Interpret 95% confidence intervals for conditional odds ratios (in terms of victim's race) given defendant's race.

c. Test the effect of defendant's race, controlling for victims' race, using a (i) Wald test, and (ii) likelihood-ratio test.