Stat 421: Applied Categorical Data Analysis, Spring 2020
Lec. 002
Derek Bean
Homework 9
Due: Friday 4/24/20 by 11:59pm CST through Canvas

Each question worth 10 points. TOTAL: 50 points.

Suggested Readings in Agresti (2nd Ed.)

1. Chapter 4, Section 4.3: §4.3.1–4.3.3
2. Chapter 4, Section 4.4

Problems

1. A study in Florida found that the death penalty was given in 19 out of 151 cases in which a white killed a white, in 0 out of 9 cases in which white killed a black, in 11 out of 63 cases in which a black killed a white, and in 6 out of 103 cases in which a black killed a black. The below table shows software output when logic model is fit for death penalty as a response (1 = yes), with defendant's race (1 = white) and victim's race (1 = white) as indicator predictors.

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) Based on the parameter estimates, which group is most likely to have the "yes" response? Estimate the probability in that case.

(b) Interpret the parameter estimate for victim's race.

(c) Using the information shown, construct and interpret a 95% likelihood-ratio confidence interval for the conditional odds ratio between death penalty verdict and victim's race.

(d) Test the effect of victim's race, controlling for defendant's race, using a Wald test or likelihood-ratio test. Interpret.

2. The below table shows results of an eight-center clinical trial to compare a drug to placebo for curing an infection. At each center, subjects were randomly assigned to groups.

Center	Treatment	Response: Success	Response: Failure	Sample Odds Ratio
1	Drug	11	25	1.19
	Control	10	27	
2	Drug	16	4	1.82
	Control	22	10	
3	Drug	14	5	4.80
	Control	7	12	
4	Drug	2	14	2.29
	Control	1	16	
5	Drug	6	11	∞
	Control	0	12	
6	Drug	1	10	∞
	Control	0	10	
7	Drug	1	4	2.0
	Control	1	8	
8	Drug	4	2	0.33
	Control	6	1	

Source: P. J. Beitler and J. R. Landis, *Biometrics*, **41**: 991–1000, 1985.

(a) Analyze these data using logistic regression, describing and making inferences about the group effect.

(b) Give possible reasons for why we would want to control for the center variable in our model.

3. A sample of subjects were asked their opinion about current laws legalizing abortion (support, oppose). For the explanatory variables gender G (female, male), religious affiliation R (Protestant, Catholic, Jewish), and political party affiliation P (Democrat, Republican, Independent), the model for the probability π of supporting legalized abortion,

$$\text{logit}(\pi) = \alpha + \beta_h^G + \beta_i^R + \beta_j^P,$$

has reported parameter estimates (setting the parameter for the last listed category of a variable equal to 0) $\hat{\alpha} = -0.11$, $\hat{\beta}_1^G = 0.16$, $\hat{\beta}_2^G = 0$, $\hat{\beta}_1^R = -0.57$, $\hat{\beta}_2^R = -0.66$, $\hat{\beta}_3^R = 0$, $\hat{\beta}_1^P = 0.84$, $\hat{\beta}_2^P = -1.67$, $\hat{\beta}_3^P = 0$.

(a) Interpret how the odds of supporting legalized abortion depend on gender.

(b) Find the estimated probability of supporting legalized abortion for (i) male Catholic Republicans and (ii) female Jewish Democrats.

(c) If we defined parameters such that the *first* category of a variable has value 0, then what would $\hat{\beta}^G$ equal? Show then how to obtain the odds ratio that describes the conditional effect of gender.

(d) If we defined parameters such that they sum to 0 across the categories of a variable, then what would $\hat{\beta}_1^G$ and $\hat{\beta}_2^G$ equal? Show then how to obtain the odds ratio that describes the conditional effect of gender.

4. The table below shows estimated effects for a fitted logistic regression model with squamous cell esophageal cancer (1 = yes, 0 = no) as the response variable Y. Smoking status (S) equals 1 for at least one pack per day and 0 otherwise, alcohol consumption (A) equals the average number of alcoholic drinks consumed per day, and race (R) equals 1 for blacks and 0 for whites.

Variable	Effect	P-value
Intercept	−7.00	<0.01
Alcohol use	0.10	0.03
Smoking	1.20	<0.01
Race	0.30	0.02
Race × smoking	0.20	0.04

(a) To describe the race-by-smoking interaction, construct the prediction equation when $R = 1$ and again when $R = 0$. Find the fitted YS conditional odds ratio for each case. Similarly, construct the prediction equation when $S = 1$ and again when $S = 0$, and find the fitted YR conditional odds ratio for each case. Note that, for each association, the coefficient of the cross-product term is the difference between the log odds ratios at the two fixed levels for the other variable.

(b) Explain what the coefficients of R and S represent for the coding as given above. What hypotheses do the P-values refer to for these variables?

(c) Suppose the model also contained an $A \times R$ interaction term, with coefficient 0.04. In the prediction equation, show that this represents the difference between the effect of A for blacks and for whites.

5. The table below shows results of a study about Y = whether a patient having surgery with general anesthesia experienced a sore throat on waking (1 = yes) as a function of D = duration of the surgery (minutes) and T = type of device used to secure the airway (0 = laryngeal mask airway, 1 = tracheal tube).

Patient	D	T	Y	Patient	D	T	Y	Patient	D	T	Y
1	45	0	0	13	50	1	0	25	20	1	0
2	15	0	0	14	75	1	1	26	45	0	1
3	40	0	1	15	30	0	0	27	15	1	0
4	83	1	1	16	25	0	1	28	25	0	1
5	90	1	1	17	20	1	0	29	15	1	0
6	25	1	1	18	60	1	1	30	30	0	1
7	35	0	1	19	70	1	1	31	40	0	1
8	65	0	1	20	30	0	1	32	15	1	0
9	95	0	1	21	60	0	1	33	135	1	1
10	35	0	1	22	61	0	0	34	20	1	0
11	75	0	1	23	65	0	1	35	40	1	0
12	45	1	1	24	15	1	0				

Source: Data from D. Collett, in *Encyclopedia of Biostatistics*, Wiley, New York, 1998, pp. 350–358. Predictors are D = duration of surgery, T = type of device.

(a) Fit a main effects model using these predictors. Interpret parameter estimates.

(b) Conduct inference about the D effect in the model in part (a).

(c) Fit a model permitting interaction. Report the prediction equation for the effect of D when (i) $T = 1$, (ii) $T = 0$. Interpret.

(d) Conduct inference about whether you need the interaction term in the model in (c).