Categorical Data Analysis
Homework III

04/06/2021

1. The table below shows the results of a study about $Y =$ whether a patient having surgery with general anesthesia experienced a sore throat on waking ($0 =$ no, $1 =$ yes) as a function of the $D =$ duration of the surgery (in minutes) and the $T =$ 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 "Binary Data" by D. Collett, in *Encyclopedia of Biostatistics*, 2nd ed. Hoboken, NJ: Wiley, 2005, pp. 439–446.

   (a) Use a model-building strategy to select a logistic model for these predictors. For your model, interpret parameter estimates, and conduct inference about the effects.

   (b) Refer to (a). For your preferred model, summarize predictive power using classification tables with $p_0 = 0.5$ and $p_0 = \bar{y}$. In each case, report and interpret the sensitivity and specificity.

   (c) Summarize the predictive power of your preferred model using a ROC curve. Report and interpret the concordance index.

2. The table below refers to the effectiveness of immediately injected or $1\frac{1}{2}$-hour-delayed penicillin in protecting rabbits against lethal injection with β-hemolytic streptococci.

Penicillin Level	Delay	Response	
		Cured	Died
$\frac{1}{8}$	None	0	6
	$1\frac{1}{2}$ h	0	5
$\frac{1}{4}$	None	3	3
	$1\frac{1}{2}$ h	0	6
$\frac{1}{2}$	None	6	0
	$1\frac{1}{2}$ h	2	4
1	None	5	1
	$1\frac{1}{2}$ h	6	0
4	None	2	0
	$1\frac{1}{2}$ h	5	0

Source: Reprinted with permission from Mantel (1963).

(a) Let $X =$ delay, $Y =$ whether cured, and $Z =$ penicillin level. Fit the logistic model

$$\text{logit}(p_{ik}) = \alpha + \beta x_i + \beta_k^Z, i = 1,2, k = 1,\ldots,K.$$

Argue that the pattern of 0 cell counts suggests that (with no intercept) $\hat{\beta}_1^Z = -\infty$ and $\hat{\beta}_5^Z = \infty$. What does your software report?

(b) Using the logistic model, conduct the likelihood-ratio test of XY conditional independence. Interpret.

(c) Test XY conditional independence using the Cochran-Mantel-Haenszel test. Interpret.

3. For the table below on maternal alcohol consumption and child's congenital malformations, use the following two priors respectively to perform Bayesian analysis for the linear **logit** model with scores $(0, 0.5, 1.5, 4.0, 7.0)$, report the posterior mean and standard deviation and a 95% posterior interval (equal-tail or HPD) for β, and compare results to those obtained with ML.

Alcohol Consumption	Malformation	
	Present	Absent
0	48	17,066
<1	38	14,464
1–2	5	788
3–5	1	126
≥6	1	37

(a) $\mathrm{N}(0, 1000^2)$ prior,

(b) $\mathrm{N}(0, 1)$ prior.

4. Refer to Exercise 3. For the same data, conduct Bayesian analyses with the **probit** link, using two prior distributions respectively which (on the probit scale) are comparable to the two priors used for the logit link in Exercise 3. Compare results to those obtained with ML probit and with the Bayesian logistic analysis.