CatData Final

Matthew Vanaman

05-18-2019

5.19

(a)

$\text{logit}(\pi) = \alpha + \beta_1 M_1 + \beta_2 M_2 + \beta_3 M_3 + \beta_4 M_4 + \beta_5 M_5 + \beta_6 M_6$, with M = major, once for each level. When you want to assess the effect of a particular level of major, code that major (e.g. M_3) to 1 and the rest to zero.

(b)

The deviance goodness-of-fit test approximates the χ^2 for GLMs with binomial responses and large numbers of successes and failures. Our model meets these criteria. The p-value for the deviance test is 0.001, indicating that other possible predictors excluded from this model may not equal zero (i.e. bad fit).

(c)

The rule of thumb for standardized residuals is that residuals with aboslute values above 2 or 3 are suspect. Most of the standardized residuals given are pretty small, with the exception of the first Major. This indicates that the model in (a) with Major as the only predictor may not have good fit because it excludes another important predictor (gender).

(d)

Because the model in (a) excludes a predictor for gender, it assumes an equal probability of admission for males and females across majors. If this is true, the standardized residuals should be close to zero (they are normally distributed when the model holds). Because gender is a binary predictor in this case, we can expect the standardized residual for males to be identical to female, but with an opposite sign. In other words, whatever deviation from the expected probability there is for females, there is probability of equal magnitude in the opposite direction for males. The residual of -4.15 for males indicates that males have a lower-than-expected probability of admission, likely violating the assumption of equal probability for males and females within that major.

(e)

First of all, this illustrates Simpson's paradox: failing to account for differences across majors leads to an effect in the opposite direction. When you condition on major females have 10% lower odds of admission, yet collapsing over major, females have 84% greater odds of admission. This *could* mean that females apply to the more competitive majors relatively more often, so are accepted less often than males after conditioning on major. So when you control for major, females fare worse because the majors they are applying to are more competitive. Males on the other hand, are applying to less competitive majors, meaning that when you average across majors of varying difficulty, it appears that males are doing worse. When you allow for the conditionality of major, males fare better because they apply in relatively higher numbers to the easier majors. This shows why it's important to keep track of whom applies to where when making group comparisons.

7.9

9 Table7.23referstoapplicantstograduateschoolattheUniversityofCalifornia, Berkeley for the fall 1973 session. Admissions decisions are presented by gender of applicant, for the six largest graduate departments. Denote the three variables by A = whether admitted, F = female, and M = major Fit loglinear model (AM, AF, MF).

a. Report the estimated AF conditional odds ratio, and compare it with the AF marginal odds ratio. Why are they so different?

b. Report G^2 and df values, and comment on the quality of fit. Conduct a residual analysis. Describe the lack of fit.

c. Deleting the data for major 1, re-fit the model. Interpret.

d. Deleting the data for Department 1 and treating A as the response variable, fit an equivalent logistic model for model (AM, AF, MF) in (c). Show how to use each model to obtain an odds ratio estimate of the effect of F on A, controlling for D.

(a)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	5.5936	0.0548	102.10	0.0000
admit	1.1287	0.0669	16.88	0.0000
female	-1.6557	0.0874	-18.94	0.0000
major	-0.0222	0.0141	-1.58	0.1149
admit:major	-0.4848	0.0199	-24.40	0.0000
admit:female	-0.1084	0.0711	-1.52	0.1275
female:major	0.3780	0.0192	19.66	0.0000

Major	Myes	Mno	Fyes	Fno
1.00	262.82	500.33	73.24	125.11
2.00	257.06	301.35	104.54	109.96
3.00	251.43	181.50	149.22	96.66
4.00	245.92	109.32	213.00	84.96
5.00	240.53	65.84	304.03	74.68
6.00	235.26	39.66	433.97	65.64

a. Estimated odds ratios are 0.9 for conditional and 1.8 for marginal. Men apply in greater numbers to departments (1, 2) having relatively high admissions rates and women apply in greater numbers to departments (3, 4, 5, 6) having relatively low admissions rates.

b. Deviance G 2 = 20.2 (df = 5), poor fit. Standardized residuals show lack of fit only for Department 1.

c. G 2 = 2.56, df = 4, good fit.

d. Logit model with main effects for department and gender has estimated conditional odds ratio = 1.03 between gender and admissions. Model deleting gender term fits essentially as well, with G 2 = 2.68 (df = 5); plausible that admissions and gender are conditionally independent for these departments.

4.37