STAC51 TUT02

Week 13: Apr 8, 2021

To study the effects of racial characteristics on whether individual convicted of homicide receive the death penalty, 674 subjects were classified in the following table.

Victim's Race	Defendent's Race	Death Penalty Yes	Death Penalty No
White	White	53	414
	Black	11	37
Black	White	0	16
	Black	4	139

Let D = defendants race, V = victims race, and P = death penalty verdict.

(a) Fit model (DV, DP, PV).

(b) Using the fitted values, calculate the odds ratio between D and P at each level of V. Note the common odds ratio property of this model, and interpret the value.

(c) Calculate the marginal odds ratio between D and P, (i) using the fitted values, (ii) using the sample data. Contrast the fitted odds ratio with that in (b), and remark on Simpsons paradox.

(d) Test the goodness of fit of this model. Interpret.

(e) Fit the simpler model (DV, PV). Interpret associations, test the fit and conduct a residual analysis.

(f) Test the $D - P$ partial association by comparing the fits of models (DV, PV) and (DV, DP, PV). Interpret.

(b) The fitted values are:

Victims	Defendants	Death Penalty	
Race	Race	Yes	No
White	White	52.82	414.18
	Black	11.18	36.82
Black	White	0.18	15.82
	Black	3.82	139.18

From the fitted values, estimated odds ratio between D and P when V = White is 0.42 and when V = Black, it is 0.42. Since this is a homogeneous association model, the odds ratios are same at each level of V . Note that, this odds ratio can also be obtained as

$$\exp\{\hat{\lambda}_{11}^{DP} + \hat{\lambda}_{22}^{DP} - \hat{\lambda}_{12}^{DP} - \hat{\lambda}_{21}^{DP}\} = \exp(-0.8678) = 0.42.$$

The value of this common odds ratio is smaller than 1 indicating that the odds of getting death penalty for a white defendant is smaller than that of a black defendant controlling for the victims race.

The fitted marginal table between Death Penalty verdict and defendant's race is

Defendants	Death Penalty	
Race	Yes	No
White	53	430
Black	15	176

2. Suppose that $\mu_{ij} = n\pi_{ij}$ satisfy the independence model in an $I \times J$ contingency table.

(a) Show that $\lambda_a{}^Y$ - $\lambda_b{}^Y = \log(\pi_{+a}/\pi_{+b})$.

(b) Show that {all $\lambda_j^Y = 0$} is equivalent to $\pi_{+j} = 1/J$ for all j.

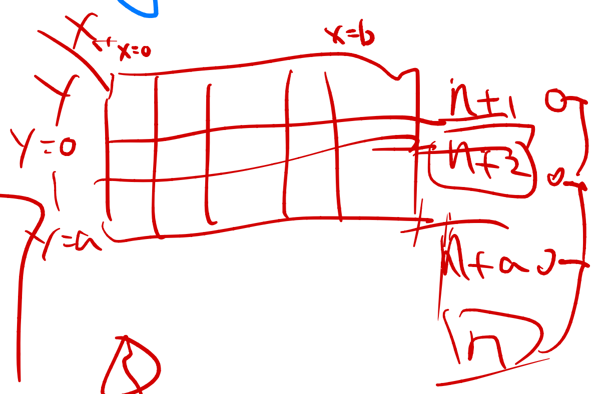
X, Y

- $\log(\mu_{ij}) = \log \pi_{i+} + \log \pi_{+j} + \log(n)$

(a) $\log(\mu_{ia}) - \log(\mu_{ib}) = (\log \pi_{i+} + \log \pi_{+a} + \log n) - (\log \pi_{i+} + \log \pi_{+b} + \log n)$

$= \log(\pi_{+a}/\pi_{+b})$

$\Rightarrow = (\lambda + \lambda_i^X + \lambda_a^Y) - (\lambda + \lambda_i^X + \lambda_b^Y) = \lambda_a^Y - \lambda_b^Y$



(b) If All $\lambda_j^Y = 0 \Rightarrow \pi_{+a} = \pi_{+b}$.

Let $\pi_{+j} = \pi$.

$\sum_{j=1}^{J} \pi_{+j} = 1$.

$\Rightarrow \pi_{+j} = 1/J$.

$\pi_{ij} = \pi_{+j}\pi_{i+}$.

$\log(\pi_{ij}) = \lambda + \underset{(I-1)}{\lambda_i^X} + \underset{(J-1)}{\lambda_j^Y}$

1 + +

0

$= 0$

4. Below, please find the ouputs of a logistic regression, where the outcome was Y = low birth weight of children. Y is a binary outcome, where, $Y = 1$ indicates that the children had "low birth weight" and $Y = 0$ indicates "not low birthweight". The two covariates included in the model were X = "smoking status" of mothers (1 = smoker & 0 = non-smoker) and Z = "Whether the birth was classified as premature or full-term" (1 = premature & 0 = full term).

	Estimate	SE	z-value	p-value
=0 Intercept	-3.58	0.25	-14.55	0.00
Smoker=Yes (X = 1)	0.49	0.64	0.76	0.45
Premature=Yes (Z=1)	3.55	0.32	11.00	0.00
Smoker = Yes & Premature = Yes (X=1 & Z=1)	-0.06	0.85	-0.07	0.95

(a) Write the regression model that was fitted. State and interpret the parameters.

X=1 premature, Z=1

(b) For a smoker mother and ~~a full term~~ baby calculate the predictive probability of low birth weight (i.e., $P(Y = 1 \mid X = 1, Z = 1)$) ~~and deviance residual~~ from the model output. ASSUME THE BABY HAD LOW BIRTH WEIGHT (y=1).

(a) $\log\left(\frac{E(y|X,Z)}{1-E(y|X,Z)}\right) = \beta_0 + \beta_x X + \beta_z Z + \boxed{\beta_{xz}} XZ$.

β_0: log-odds low birthweight when mother is non-smoker.

β_x: log-OR.

β_z: $\rightarrow$ main effect

(b) $\hat{\pi} = P(y=1 \mid x=1, z=1) = \frac{\exp(-3.58+0.49+3.55-0.06)}{1+\Delta} = \boxed{0.5986877}$

$\Delta = \exp(-3.58+0.49+3.55-0.06)$

$\Delta_i = 2\left(y_i \log\left(\frac{y_i}{\hat{\pi}_i}\right) + (1-y_i)\log\left(\frac{1-y_i}{1-\hat{\pi}_i}\right)\right)$ $\quad y_i = 1, \quad \hat{\pi}_i = 0.5987$

$= 1.02603$.

$y_i > \hat{\pi}_i$. $\quad \sqrt{\Delta_i} = \sqrt{1.02603} = 1.01293$

Let's assume we are interested in the relationship between job satisfaction and income among some employees of a company. Job satisfaction is a categorical response variable with four categories and they are, "0 = very dissatisfied", "1 = dissatisfied", "2 = satisfied" and "3 = very satisfied". Income is a continuous explanatory variable. First we fit a multinomial logistic regression. Below is the output,

```
### The multinomial data ###
mod.fit<-multinom(formula = satis ~ score, data=satisData)
summary(mod.fit)
Call:
multinom(formula = satis ~ score, data = satisData)

Coefficients:
(Intercept)     score
dis        0.08129887 0.1282872
satis      1.25461877 0.1518843
v_satis -0.56384582 0.1987923

Std. Errors:
(Intercept)     score
dis         0.9434572 0.1032131
satis       0.8573106 0.1000908
v_satis     0.9601394 0.1021073

   Residual Deviance: 205.4909
   AIC: 217.4909
```

(a) Write the model in terms of the parameters of the j^{th} category ($j \in \{1, 2, 3\}$). Specify the reference group. [5 Marks]

(b) Find the odds ratio of "dissatisfied" and "satisfied" (i.e., $\frac{\pi_1(x)}{\pi_2(x)}$) for one unit change in the income score. [8 Marks]

Table below contains results of a study comparing radiation therapy with surgery in treating cancer of the larynx. Use Fishers exact test to test $H_0 : \theta = 1$ vs $H_a : \theta > 1$. Here θ is the odds ratio. [12 Marks]

	Cancer controlled	Cancer Not Controlled	Total
Surgery	17	2	19
Radiation Therapy	12	3	15
Total	29	5	34