Loglinear Modeling of Berkeley Admissions

Greg Johnson

Table of Contents

Introduction

Our data are counts for the admittance of Berkeley students, distinguished by gender and department. The result is a $2 \times 2 \times 6$ three-way table of counts.

	Dept. A		Dept. B		Dept. C		Dept. D		Dept. E		Dept. F	
	M	F	M	F	M	F	M	F	M	F	M	F
Admit	512	89	353	17	120	202	138	131	53	94	22	24
No Admit	313	19	207	8	205	391	279	244	138	299	351	317

Loglinear Models

Let's run all possible loglinear models given our three categorical variables: Admit (A), Sex (S), and Department (D). There are $2^{\binom{3}{2}} + 1 = 9$ different models: one saturated model, seven conditional association models and one mutual independence model. Their formulas are given below.

		Loglinear Formula	GLM Formula
Model 1	Mutual Independence	(A, S, D)	$Y \sim A + S + D$
Model 2	Admit-Sex Interaction	(AS, D)	$Y \sim A + S + D + AS$
Model 3	Admit-Dept	(AD, S)	$Y \sim A + S + D + AD$
Model 4	Sex-Dept	(SD, A)	$Y \sim A + S + D + SD$
Model 5	Admit-Dept, Admit-Sex	(AD, AS)	$Y \sim A + S + D + AD + AS$
Model 6	Admit-Sex, Dept-Sex	(AS, DS)	$Y \sim A + S + D + AS + DS$
Model 7	Admit-Dept, Dept-Sex	(AD, DS)	$Y \sim A + S + D + AD + DS$
Model 8	Homogeneous Association	(AS, AD, DS)	$Y \sim A + S + D + AS + AD + DS$
Model 9	Saturated	(ASD)	$Y \sim A + S + D + AS + SD + AD + ASD$

In terms of overall fit, every model except the saturated model (obviously) fits terribly. Both the Likelihood Ratio Test (Deviance) and Pearson's Chi-Squared Test reject the null hypothesis of good fit for every model.

	Model	G^2 (Deviance)	df	p-value	X^2 (Chi-squared)	p-value
1	(A, S, D)	2097.7	16	≈0	2000.0	≈0
2	(AS, D)	2004.2	15	≈0	1748.2	≈0
3	(AD, S)	1242.4	11	≈0	1078.1	≈0
4	(SD, A)	877.1	11	≈0	797.7	≈0
5	(AD, AS)	1148.9	10	≈0	1015.7	≈0
6	(AS, DS)	783.6	10	≈0	715.3	≈0
7	(AD, DS)	21.8	6	0.003	19.9	0.003
8	(AS, AD, DS)	20.2	5	0.002	18.8	0.002

9	(ASD)	0	0	1	0	1

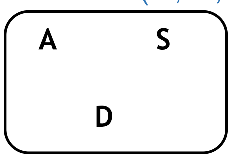
Admission vs. Sex in Dept. A vs. Dept. F

To look at the relationship between admission (X) and sex (Y) conditional on department (Z), we look at the conditional odds ratio of admission and sex for department k:

$$\log\theta = \log\left[\frac{\mu_{11k}\mu_{22k}}{\mu_{12k}\mu_{21k}}\right] = \log\mu_{11k} + \log\mu_{22k} - \log\mu_{12k} - \log\mu_{21k}$$

Specifically we want to compare the odds ratio conditional on Department A $(Z = 1)$ and Department F $(Z = 6)$ for the 9 models.

Model 1 (A, S, D)



For the conditional odds ratio of admission and sex in department A and F:

$$\log\theta_{Z=1} = \log\mu_{111} + \log\mu_{221} - \log\mu_{121} - \log\mu_{211}$$

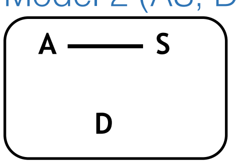
$$= \left(\lambda + \lambda_1^x + \lambda_1^y + \lambda_1^z\right) + \left(\lambda + \lambda_2^x + \lambda_2^y + \lambda_1^z\right) - \left(\lambda + \lambda_1^x + \lambda_2^y + \lambda_1^z\right) - \left(\lambda + \lambda_2^x + \lambda_1^y + \lambda_1^z\right) = 0$$

$$\log\theta_{z=1} = 0 \;\rightarrow\; \theta_{z=1} = \theta_{z=6} = 1$$

This tells us that under Model 1, Admission and Sex are conditionally independent in department A; we can tell from the above formula that this is also true for department F. This is to be expected under the mutual independence model. In fact, Admission and Sex are also marginally independent.

So, under the model of mutual independence, the odds of admission is the same for both sexes, for every department.

Model 2 (AS, D)



For the conditional odds ratio of admission and sex in department A and F:

$$\log\theta_{Z=1} = \log\mu_{111} + \log\mu_{221} - \log\mu_{121} - \log\mu_{211}$$

$$= \left(\lambda + \lambda_1^x + \lambda_1^y + \lambda_1^z + \lambda_{11}^{xy}\right) + \left(\lambda + \lambda_2^x + \lambda_2^y + \lambda_1^z + \lambda_{22}^{xy}\right)$$

$$-\left(\lambda+\lambda_1^x+\lambda_2^y+\lambda_1^z+\lambda_{12}^{xy}\right)-\left(\lambda+\lambda_2^x+\lambda_1^y+\lambda_1^z+\lambda_{21}^{xy}\right)$$

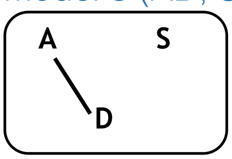
$$=\lambda_{11}^{xy}+\lambda_{22}^{xy}-\lambda_{12}^{xy}-\lambda_{21}^{xy}=4\lambda_{11}^{xy}\text{ (under contrast parameterization)}$$

$$\log\theta_{z=1}=4\cdot 0.153=0.61$$

$$\theta_{z=1}=\theta_{z=6}=1.84$$

Notice that under this model, the conditional odds ratio is not zero, so admission and sex have a conditional association. Furthermore, the magnitude of this association isn't determined by the department conditioned on. So, the odds of being admitted is 84% greater for males than it is for females in all departments.

Model 3 (AD, S)



For the conditional odds ratio of admission and sex in department A and F:

$$\log\theta_{Z=1}=\log\mu_{111}+\log\mu_{221}-\log\mu_{121}-\log\mu_{211}$$

$$=\left(\lambda+\lambda_1^x+\lambda_1^y+\lambda_1^z+\lambda_{11}^{xz}\right)+\left(\lambda+\lambda_2^x+\lambda_2^y+\lambda_1^z+\lambda_{21}^{xz}\right)$$

$$-\left(\lambda+\lambda_1^x+\lambda_2^y+\lambda_1^z+\lambda_{11}^{xz}\right)-\left(\lambda+\lambda_2^x+\lambda_1^y+\lambda_1^z+\lambda_{21}^{xz}\right)$$

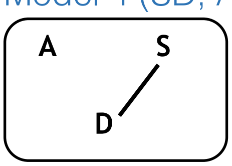
$$=\lambda_{11}^{xz}+\lambda_{21}^{xz}-\lambda_{11}^{xz}-\lambda_{21}^{xz}=0$$

$$\log\theta_{z=1}=0$$

$$\theta_{z=1}=\theta_{z=6}=1$$

Under this model, admission and sex are conditionally independent.

Model 4 (SD, A)



For the conditional odds ratio of admission and sex in department A and F:

$$\log\theta_{Z=1}=\log\mu_{111}+\log\mu_{221}-\log\mu_{121}-\log\mu_{211}$$

$$=\left(\lambda+\lambda_1^x+\lambda_1^y+\lambda_1^z+\lambda_{11}^{yz}\right)+\left(\lambda+\lambda_2^x+\lambda_2^y+\lambda_1^z+\lambda_{21}^{yz}\right)$$

$$-\left(\lambda+\lambda_1^x+\lambda_2^y+\lambda_1^z+\lambda_{21}^{yz}\right)-\left(\lambda+\lambda_2^x+\lambda_1^y+\lambda_1^z+\lambda_{11}^{yz}\right)$$

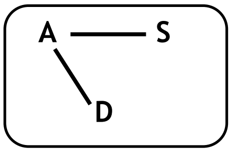
$$= \lambda_{11}^{yz} + \lambda_{21}^{yz} - \lambda_{21}^{yz} - \lambda_{11}^{yz} = 0$$

$$\log\theta_{z=1} = 0$$

$$\theta_{z=1} = \theta_{z=6} = 1$$

Under this model, admission and sex are conditionally independent.

Model 5 (AD, AS)



For the conditional odds ratio of admission and sex in department A and F:

$$\log\theta_{Z=1} = \log\mu_{111} + \log\mu_{221} - \log\mu_{121} - \log\mu_{211}$$

$$= \left(\lambda + \lambda_1^x + \lambda_1^y + \lambda_1^z + \lambda_{11}^{xy} + \lambda_{11}^{xz}\right) + \left(\lambda + \lambda_2^x + \lambda_2^y + \lambda_1^z + \lambda_{22}^{xy} + \lambda_{21}^{xz}\right)$$

$$-\left(\lambda + \lambda_1^x + \lambda_2^y + \lambda_1^z + \lambda_{12}^{xy} + \lambda_{11}^{xz}\right) - \left(\lambda + \lambda_2^x + \lambda_1^y + \lambda_1^z + \lambda_{21}^{xy} + \lambda_{21}^{xz}\right)$$

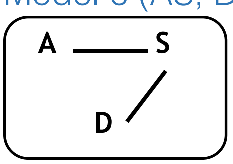
$$= \left(\lambda_{11}^{xy} + \lambda_{22}^{xy} - \lambda_{12}^{xy} - \lambda_{21}^{xy}\right) + \left(\lambda_{21}^{xz} + \lambda_{11}^{xz} - \lambda_{21}^{xz} - \lambda_{11}^{xz}\right) = 4\lambda_{11}^{xy}$$

$$\log\theta_{z=1} = \log\theta_{z=6} = 0.61$$

$$\theta_{z=1} = \theta_{z=6} = 1.84$$

Under this model, admission and sex are conditionally associated. The nature of this association isn't determined by the department conditioned on. So, the odds of being admitted is 84% greater for males than it is for females in all departments.

Model 6 (AS, DS)



For the conditional odds ratio of admission and sex in department A and F:

$$\log\theta_{Z=1} = \log\mu_{111} + \log\mu_{221} - \log\mu_{121} - \log\mu_{211}$$

$$= \left(\lambda + \lambda_1^x + \lambda_1^y + \lambda_1^z + \lambda_{11}^{xy} + \lambda_{11}^{yz}\right) + \left(\lambda + \lambda_2^x + \lambda_2^y + \lambda_1^z + \lambda_{22}^{xy} + \lambda_{21}^{yz}\right)$$

$$-\left(\lambda + \lambda_1^x + \lambda_2^y + \lambda_1^z + \lambda_{12}^{xy} + \lambda_{21}^{yz}\right) - \left(\lambda + \lambda_2^x + \lambda_1^y + \lambda_1^z + \lambda_{21}^{xy} + \lambda_{11}^{yz}\right)$$

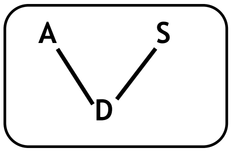
$$= \left(\lambda_{11}^{xy} + \lambda_{22}^{xy} - \lambda_{12}^{xy} - \lambda_{21}^{xy}\right) + \left(\lambda_{11}^{yz} + \lambda_{21}^{yz} - \lambda_{21}^{yz} - \lambda_{11}^{yz}\right) = 4\lambda_{11}^{xy}$$

$$\log\theta_{z=1} = \log\theta_{z=6} = 0.61$$

$$\theta_{z=1} = \theta_{z=6} = 1.84$$

Under this model, admission and sex are conditionally associated. The nature of this association isn't determined by the department conditioned upon. So, the odds of being admitted is 84% greater for males than it is for females in all departments.

Model 7 (AD, DS)



For the conditional odds ratio of admission and sex in department A and F:

$$\log\theta_{Z=1} = \log\mu_{111} + \log\mu_{221} - \log\mu_{121} - \log\mu_{211}$$

$$= \left(\lambda + \lambda_1^x + \lambda_1^y + \lambda_1^z + \lambda_{11}^{xz} + \lambda_{11}^{yz}\right) + \left(\lambda + \lambda_2^x + \lambda_2^y + \lambda_1^z + \lambda_{21}^{xz} + \lambda_{21}^{yz}\right)$$

$$-\left(\lambda + \lambda_1^x + \lambda_2^y + \lambda_1^z + \lambda_{11}^{xz} + \lambda_{21}^{yz}\right) - \left(\lambda + \lambda_2^x + \lambda_1^y + \lambda_1^z + \lambda_{21}^{xz} + \lambda_{11}^{yz}\right)$$

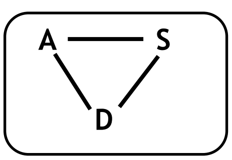
$$= \left(\lambda_{11}^{xz} + \lambda_{21}^{xz} - \lambda_{11}^{xz} - \lambda_{21}^{xz}\right) + \left(\lambda_{11}^{yz} + \lambda_{21}^{yz} - \lambda_{21}^{yz} - \lambda_{11}^{yz}\right) = 0$$

$$\log\theta_{z=1} = \log\theta_{z=6} = 0$$

$$\theta_{z=1} = \theta_{z=6} = 1$$

Under this model, admission and sex are conditionally associated. The nature of this association isn't determined by the department conditioned upon. So, the odds of being admitted is 84% greater for males than it is for females in all departments.

Model 8 (AS, AD, DS)



For the conditional odds ratio of admission and sex in department A and F:

$$\log\theta_{Z=1} = \log\mu_{111} + \log\mu_{221} - \log\mu_{121} - \log\mu_{211}$$

$$= \left(\lambda + \lambda_1^x + \lambda_1^y + \lambda_1^z + \lambda_{11}^{xy} + \lambda_{11}^{yz} + \lambda_{11}^{xz}\right) + \left(\lambda + \lambda_2^x + \lambda_2^y + \lambda_1^z + \lambda_{22}^{xy} + \lambda_{21}^{yz} + \lambda_{21}^{xz}\right)$$

$$-\left(\lambda + \lambda_1^x + \lambda_2^y + \lambda_1^z + \lambda_{12}^{xy} + \lambda_{21}^{yz} + \lambda_{11}^{xz}\right) - \left(\lambda + \lambda_2^x + \lambda_1^y + \lambda_1^z + \lambda_{21}^{xy} + \lambda_{11}^{yz} + \lambda_{21}^{xz}\right)$$

$$= \left(\lambda_{11}^{xy} + \lambda_{22}^{xy} - \lambda_{12}^{xy} - \lambda_{21}^{xy}\right) + \left(\lambda_{11}^{yz} + \lambda_{21}^{yz} - \lambda_{21}^{yz} - \lambda_{11}^{yz}\right) + \left(\lambda_{11}^{xz} + \lambda_{21}^{xz} - \lambda_{11}^{xz} - \lambda_{21}^{xz}\right)$$

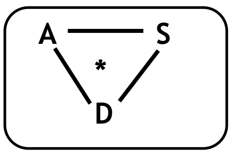
$$= 4\lambda_{11}^{xy}$$

$$\log\theta_{z=1} = \log\theta_{z=6} = -0.09$$

$$\theta_{z=1} = \theta_{z=6} = 0.91$$

Under this model, admission and sex are conditionally associated. The nature of this association isn't determined by the department conditioned upon. So, the odds of being admitted is 11% greater for females than it is for males in all departments (reciprocal of odds ratio).

Model 9 (ASD)



For the conditional odds ratio of admission and sex in department A:

$$\log\theta_{Z=1} = \log\mu_{111} + \log\mu_{221} - \log\mu_{121} - \log\mu_{211}$$

$$= \left(\lambda + \lambda_1^x + \lambda_1^y + \lambda_1^z + \lambda_{11}^{xy} + \lambda_{11}^{yz} + \lambda_{11}^{xz} + \lambda_{111}^{xyz}\right) + \left(\lambda + \lambda_2^x + \lambda_2^y + \lambda_1^z + \lambda_{22}^{xy} + \lambda_{21}^{yz} + \lambda_{21}^{xz} + \lambda_{221}^{xyz}\right)$$

$$-\left(\lambda + \lambda_1^x + \lambda_2^y + \lambda_1^z + \lambda_{12}^{xy} + \lambda_{21}^{yz} + \lambda_{11}^{xz} + \lambda_{121}^{xyz}\right) - \left(\lambda + \lambda_2^x + \lambda_1^y + \lambda_1^z + \lambda_{21}^{xy} + \lambda_{11}^{yz} + \lambda_{21}^{xz} + \lambda_{211}^{xyz}\right)$$

$$= \left(\lambda_{11}^{xy} + \lambda_{22}^{xy} - \lambda_{12}^{xy} - \lambda_{21}^{xy}\right) + \left(\lambda_{11}^{yz} + \lambda_{21}^{yz} - \lambda_{21}^{yz} - \lambda_{11}^{yz}\right) + \left(\lambda_{11}^{xz} + \lambda_{21}^{xz} - \lambda_{11}^{xz} - \lambda_{21}^{xz}\right)$$

$$+(\lambda_{111}^{xyz} + \lambda_{221}^{xyz} + \lambda_{121}^{xyz} + \lambda_{211}^{xyz})$$

$$= 4\lambda_{11}^{xy} + 4\lambda_{111}^{xyz} = 4(-0.05 - 0.212)$$

$$\log\theta_{z=1} = -1.048$$

$$\theta_{z=1} = 0.35$$

Under this model, the odds of a female applicant being accepted to department A are 2.86 times higher than the odds of a male applicant.

For the conditional odds ratio of admission and sex in department F:

$$\log\theta_{Z=1} = \log\mu_{116} + \log\mu_{226} - \log\mu_{126} - \log\mu_{216}$$

$$= \left(\lambda + \lambda_1^x + \lambda_1^y + \lambda_6^z + \lambda_{11}^{xy} + \lambda_{16}^{yz} + \lambda_{16}^{xz} + \lambda_{116}^{xyz}\right) + \left(\lambda + \lambda_2^x + \lambda_2^y + \lambda_6^z + \lambda_{22}^{xy} + \lambda_{26}^{yz} + \lambda_{26}^{xz} + \lambda_{226}^{xyz}\right)$$

$$-\left(\lambda + \lambda_1^x + \lambda_2^y + \lambda_6^z + \lambda_{12}^{xy} + \lambda_{26}^{yz} + \lambda_{16}^{xz} + \lambda_{126}^{xyz}\right) - \left(\lambda + \lambda_2^x + \lambda_1^y + \lambda_6^z + \lambda_{21}^{xy} + \lambda_{16}^{yz} + \lambda_{26}^{xz} + \lambda_{216}^{xyz}\right)$$

$$= \left(\lambda_{11}^{xy} + \lambda_{22}^{xy} - \lambda_{12}^{xy} - \lambda_{21}^{xy}\right) + \left(\lambda_{16}^{yz} + \lambda_{26}^{yz} - \lambda_{26}^{yz} - \lambda_{16}^{yz}\right) + \left(\lambda_{16}^{xz} + \lambda_{26}^{xz} - \lambda_{16}^{xz} - \lambda_{26}^{xz}\right)$$

$$+(\lambda_{116}^{xyz} + \lambda_{226}^{xyz} + \lambda_{126}^{xyz} + \lambda_{216}^{xyz})$$

$$= 4\lambda_{11}^{xy} + 4\lambda_{116}^{xyz} = 4(-0.05 + .003)$$

$$\log\theta_{z=1} = -0.19$$

$$\theta_{z=1} = 0.83$$

Under this model, the odds of a female applicant being accepted to department F are 1.21 times higher than the odds of a male applicant.

Under this model, admission and sex are conditionally associated. Further, the nature of this association is determined by the department being conditioned upon.

Model Selection

In terms of absolute fit to the data, Model 8 (AS, AD, DS) is the best choice $(G^2(6) = 20.2,\ X^2(6) = 18.8,\ p < 0.05)$. However it still fits the data terribly. In addition, it does not provide a significantly better fit $(\Delta G^2(1) = 1.6,\ p > 0.05)$ than the nested Model 7 (AD, DS) which also happens to be the second-best choice. According to G^2 and X^2, Model 7 is the best choice.

Performing a stepwise subset selection based on AIC also yielded Model 7. Therefore based on absolute fit and relative fit, the model that imposes conditional independence between admission and sex fits the data best.