36-720: Log-Linear Models: Three-Way Tables

Brian Junker

September 5, 2007

- Aggregation and Association
- Modelling the Three-Way Table
- The Model of Complete Independence
- Models with One Factor Independent of the Others
- Conditional Independence Models
- The Model of No Three-Way Interaction

Aggregation and Association

Consider two questions on a math test:

Y_1: "What is the slope of $y = 2x + 3$ at $x = 0$?"
  ($Y_1 = 0$ for wrong, $Y_1 = 1$ for right).

Y_2: "What is the slope of $y = (x + 1)^2 + 5$ at $x = 0$?"
  ($Y_2 = 0$ for wrong, $Y_2 = 1$ for right).

For 500 students a table cross-classifying responses to these two questions might be as follows:

OBS	$Y_2 = 0$	$Y_2 = 1$	Total
$Y_1 = 0$	88	52	140
$Y_1 = 1$	192	168	360
Total	280	220	500

RESIDS	$Y_2 = 0$	$Y_2 = 1$
$Y_1 = 0$	1.08	−1.22
$Y_1 = 1$	−0.68	0.76

The odds ratio is 1.48, and the Wald test of $H_0 : OR = 1$ yields $z = 1.92$ ($p \approx 0.05$).

For example, $\hat{P}[Y_2 = 1|Y_1 = 1] = \frac{168}{360} = 0.47 > 0.44 = \frac{220}{500} = \hat{P}[Y_2 = 1]$.

On further investigation the students could be separated into a group of 200 students who had taken an algebra class (only) and 300 students who had taken a calculus class as well. Looking at the data separately for these two groups of students, we see

Algebra Students

	$Y_2 = 0$	$Y_2 = 1$	Total
$Y_1 = 0$	64	16	80
$Y_1 = 1$	96	24	120
Total	160	40	200

Calculus Students

RESIDS	$Y_2 = 0$	$Y_2 = 1$	Total
$Y_1 = 0$	24	36	60
$Y_1 = 1$	96	144	240
Total	120	180	300

The odds ratios here are $\frac{(64)(24)}{(16)(96)} = 1$ and $\frac{(24)(144)}{(36)(96)} = 1$. *So, Y_1 and Y_2 are conditionally independent, given class:* $(Y_1 \perp\!\!\!\perp Y_2)|(\text{class})$.

Note that

- $\hat{P}[Y_1 = 1 \mid Algebra] = 0.6 < \hat{P}[Y_1 = 1 \mid Calculus] = 0.8$;
- $\hat{P}[Y_2 = 1 \mid Algebra] = 0.2 < \hat{P}[Y_2 = 1 \mid Calculus] = 0.6$;

This co-monotonicity in probabilities is what increases the association in the combined table. See Esary, Proschan and Walkup (1967, *Ann. Math. Stat.*).

Simpson's (Yule's) Paradox

Hospital A

	Recovered	Died
Treatment	n_{111}	n_{121}
Placebo	n_{211}	n_{221}

Hospital B

	Recovered	Died
Treatment	n_{112}	n_{122}
Placebo	n_{212}	n_{222}

Simpson's paradox occurs, e.g., when

$$
\begin{aligned}
OR_A &= \frac{n_{111}n_{221}}{n_{211}n_{121}} \geq 1 \\
OR_B &= \frac{n_{112}n_{222}}{n_{212}n_{122}} \geq 1 \\
OR_{A+B} &= \frac{(n_{111}+n_{112})(n_{221}+n_{222})}{(n_{211}+n_{212})(n_{121}+n_{122})} < 1
\end{aligned}
$$

(or reverse all inequalities, as in the math test example).

For example:

Hospital A	Recovered	Died
Treatment	81	6
Placebo	234	36

Hospital B	Recovered	Died
Treatment	192	71
Placebo	55	25

Here, $OR_A = \frac{(81)(36)}{(6)(234)} = 2.08$; $OR_B = \frac{(192)(25)}{(71)(55)} = 1.23$; and $OR_{A+B} = \frac{(273)(61)}{(77)(289)} = 0.74$.

Note that

- $\hat{P}[Recov. \mid Hosp.\ A] = 0.88 > \hat{P}[Recov. \mid Hosp.\ B] = 0.72$
- $\hat{P}[Treat. \mid Hosp.\ A] = 0.24 < \hat{P}[Treat. \mid Hosp.\ B] = 0.77$

This reversal of probababilities is what *reduces* the association in the combined table.

If the probabilities both increased (or both decreased) it would tend to *increase* the association in the combined table (as in the math test example).

Holland & Rosenbaum (1986, *Ann. Stat.*); Kadane, Meyer & Tukey (1999, *JASA*).

Modelling the Three-Way Table

Phenomena like Simpson's Paradox make it clear that we cannot learn all there is to know about a three- (or higher-) way table by just looking at 2-way subtables or 2-way aggregates.

We now turn directly to models for m_{ijk} in a three-way table. We will develop, simultaneously,

- Direct models for m_{ijk}; these show what (in-)dependence and conditional (in-)dependence assumptions are being made by the model;
- The corresponding log-linear models; these make a link between statistical (in-)dependence and terms in an ANOVA-like model.

In general we are looking at an $I \times J \times K$ table. Here is a $3 \times 2 \times 2$ example:

$k = 1$

	$j = 1$	$j = 2$
$i = 1$	n_{111}	n_{121}
$i = 2$	n_{211}	n_{221}
$i = 3$	n_{311}	n_{321}

$k = 2$

	$j = 1$	$j = 2$
$i = 1$	n_{112}	n_{122}
$i = 2$	n_{212}	n_{222}
$i = 3$	n_{312}	n_{322}

We consider mainly <u>multinomial</u> sampling, although the formalisms go through for the <u>product multinomial</u> model as well.

You can compute MLE's for most of these models by hand.

We will also show how to compute MLE's using R's `glm()` function.

The Model of Complete Independence

The *model of complete independence* is

$$M^{(0)} : \; p_{ijk} = p_{i++}p_{+j+}p_{++k} \quad or \;\; \perp\!\!\!\perp(\text{rows,columns,layers})$$

The MLE's are

$$\begin{aligned}
\hat{p}^{(0)}_{ijk} &= \hat{p}_{i++}\hat{p}_{+j+}\hat{p}_{++k} \\
&= (n_{i++}/n_{+++})(n_{+j+}/n_{+++})(n_{++k}/n_{+++}) \; ; \\
\hat{m}^{(0)}_{ijk} &= n_{+++}\hat{p}^{(0)}_{ijk} \\
&= (n_{i++}n_{+j+}n_{++k})/n^2_{+++} \\
&= (\hat{m}_{i++}\hat{m}_{+j+}\hat{m}_{++k})/\hat{m}^2_{+++}
\end{aligned}$$

where $\hat{m}_{ijk}$ satisfy[a] the "marginal constraints" $\hat{m}_{i++} = n_{i++}$, $\hat{m}_{+j+} = n_{+j+}$, $\hat{m}_{++k} = n_{++k}$, and $\hat{m}_{+++} = n_{+++}$.

[a]This is so because n_{i++}, etc. are the sufficient statistics for the p's when the multinomial model $M^{(0)}$ is viewed as an exponential family model: The MLE in an exponential family model has to equate observed (n_{i++}, etc.) and expected ($\hat{m}_{i++}$, etc.) sufficient statistics.

The two test-statistics are

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \frac{(n_{ijk} - \hat{m}_{ijk}^{(0)})^2}{\hat{m}_{ijk}^{(0)}}$$

$$G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} n_{ijk} \log n_{ijk}/\hat{m}_{ijk}^{(0)}$$

with df. for the χ^2 tests above are
$(IJK - 1) - (I - 1) - (J - 1) - (K - 1) = IJK - I - J - K + 2.$

The corresponding log-linear model is

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)}$$

which can be seen by log'ing $m_{ijk}^{(0)} = n_{+++}p_{i++}p_{+j+}p_{++k}$ under M_0.

Example

We consider the $2 \times 2 \times 3$ table

		Adversity of school (k)						
		Low		Med		High		
	Family Risk (j)	N	R	N	R	N	R	Total
Classrooom	Nondeviant	16	7	15	34	5	3	80
Behavior (i)	Deviant	1	1	3	8	1	3	17
	Total	17	8	18	42	6	6	97

The model can be fitted directly or by using `glm()` in R. For example,

```
n <- scan(sep="&")
16 & 7 & 15 & 34 & 5 & 3
 1 & 1 & 3 & 8 & 1 & 3
Beh <- c(rep("N",6),rep("D",6))
Risk <- rep(c("N","R"),3)
Schl <- rep(c(rep("L",2),rep("M",2),rep("H",2)),2)
summary(fit0 <- glm(n ~ Beh + Risk + Schl,family=poisson))
mhat <- fitted(fit0)
```

The table of fitted values is

	Low		Med		High		
	N	R	N	R	N	R	Total
Nondeviant	8.72	11.90	20.92	28.57	4.18	5.71	80
Deviant	1.85	2.53	4.44	6.07	0.89	1.21	17
Total	17	8	18	42	6	6	97

and it is easy to verify that this table has the same marginal totals as the original (as it must).

$X^2 = 17.30$ on $df = (IJK - 1) - (I - 1) - (J - 1) - (K - 1) = 2 \cdot 2 \cdot 3 - 1 - 2 - 1 - 1 = 7$; with a p-value of 0.015.

$G^2 = 16.42$, $df = 7$, p-value=0.02. Note that this is the same as the "residual deviance" in the `glm()` output.

The Pearson residuals for $M^{(0)}$ are

	Low		Med		High	
	N	R	N	R	N	R
Nondeviant	2.47	−1.42	−1.29	1.02	0.40	−1.14
Deviant	−0.63	−0.96	−0.69	0.78	0.12	1.62

Models with One Factor Independent of the Others

There are three such models:

$$M^{(1)}: \quad p_{ijk} = p_{i++}p_{+jk} \quad \textit{or} \quad \text{(rows)}\perp\!\!\!\perp\text{(columns,layers)}$$

$$M^{(2)}: \quad p_{ijk} = p_{+j+}p_{i+k} \quad \textit{or} \quad \text{(columns)}\perp\!\!\!\perp\text{(rows,layers)}$$

$$M^{(3)}: \quad p_{ijk} = p_{++k}p_{ij+} \quad \textit{or} \quad \text{(layers)}\perp\!\!\!\perp\text{(rows,columns)}$$

The MLE's are straightforward again. For example under $M^{(1)}$,

$$\hat{p}^{(1)}_{ijk} = \hat{p}_{i++}\hat{p}_{+jk} = n_{i++}n_{+jk}/n^2_{+++}$$

and

$$\hat{m}^{(1)}_{ijk} = \hat{m}_{i++}\hat{m}_{+jk}/\hat{m}_{+++}$$

subject to the same sorts of constraints, e.g. $m_{+jk} = n_{+jk}$, again due to equating expected and observed sufficient statistics for MLE's in the exponential family model.

The test-statistics are again

$$
\begin{aligned}
X^2 &= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\frac{(n_{ijk}-\hat{m}^{(s)}_{ijk})^2}{\hat{m}^{(s)}_{ijk}} \\
G^2 &= 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} n_{ijk}\log n_{ijk}/\hat{m}^{(s)}_{ijk} ,
\end{aligned}
$$

depending on which model $s = 1, 2$, or 3 is in use; the df. for the χ^2 tests above are

$$
\begin{aligned}
(IJK-1)-(I-1)-(JK-1) &= (I-1)(JK-1) \text{ for } M^{(1)}; \\
(IJK-1)-(J-1)-(IK-1) &= (J-1)(IK-1) \text{ for } M^{(2)}; \\
(IJK-1)-(K-1)-(IJ-1) &= (K-1)(IJ-1) \text{ for } M^{(3)}.
\end{aligned}
$$

The corresponding log-linear models are

$$
\begin{aligned}
\log m_{ijk} &= u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{23(jk)} \\
\log m_{ijk} &= u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)} \\
\log m_{ijk} &= u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)}
\end{aligned}
$$

which can be seen by log'ing the corresponding m_{ijk} under each $M^{(s)}$.

Example, Continued

Returning to the school behavior data, we can fit all three models in R as

```
summary(fit1 <- glm(n ~ Beh + Risk*Schl,family=poisson))
summary(fit2 <- glm(n ~ Risk + Beh*Schl,family=poisson))
summary(fit3 <- glm(n ~ Schl + Beh*Risk,family=poisson))
```

and we discover

Resid Deviance	df	p-value	AIC
$G^2_{(1)} = 5.56$	5	0.35	60.39
$G^2_{(2)} = 12.76$	5	0.03	67.51
$G^2_{(3)} = 52.52$	8	0.00	101.27

It seems that $M^{(1)}$ provides the best fit relative to the saturated model: (classroom behavior)$\perp\!\!\!\perp$(school adversity, family risk)

The Pearson residuals for $M^{(1)}$ are

	Low		Med		High	
	N	R	N	R	N	R
Nondeviant	0.53	0.16	0.04	−0.11	0.02	−0.88
Deviant	−1.15	−0.34	−0.09	0.24	−0.05	1.90

Conditional Independence Models

For a model asserting (rows)$\perp\!\!\!\perp$(columns) | (layers), we would have

$$P[ijk] = P[ij|k]P[k] = P[i|k]P[j|k]P[k] = \frac{p_{i+k}}{p_{++k}}\frac{p_{+jk}}{p_{++k}}p_{++k} = \frac{p_{i+k}p_{+jk}}{p_{++k}}$$

There are clearly three such models:

$$M^{(4)}: \qquad p_{ijk} = p_{i+k}p_{+jk}/p_{++k} \quad \textit{or} \;\; \text{(rows)}\perp\!\!\!\perp\text{(columns) | (layers)}$$

$$M^{(5)}: \qquad p_{ijk} = p_{ij+}p_{+jk}/p_{+j+} \quad \textit{or} \;\; \text{(rows)}\perp\!\!\!\perp\text{(layers) | (columns)}$$

$$M^{(6)}: \qquad p_{ijk} = p_{ij+}p_{i+k}/p_{i++} \quad \textit{or} \;\; \text{(columns)}\perp\!\!\!\perp\text{(layers) | (rows)}$$

The MLE's are straightforward again. For example under $M^{(4)}$,

$$\hat{p}^{(4)}_{ijk} = \hat{p}_{i+k}\hat{p}_{+jk}/\hat{p}_{++k} = (n_{i+k}n_{+jk})/(n_{++k}n_{+++})$$

and

$$\hat{m}^{(4)}_{ijk} = \hat{m}_{i+k}\hat{m}_{+jk}/\hat{m}_{++k}$$

subject to the same sorts of constraints, e.g. $m_{+jk} = n_{+jk}$, again due to equating expected and observed sufficient statistics for MLE's in the exponential family model.

The test-statistics are again

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \frac{(n_{ijk} - \hat{m}^{(s)}_{ijk})^2}{\hat{m}^{(s)}_{ijk}}$$

$$G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} n_{ijk} \log n_{ijk}/\hat{m}^{(s)}_{ijk} ,$$

depending on which model $s = 4, 5$, or 6 is in use; the df. for the χ^2 tests above are

$$\begin{array}{ll} (I-1)(J-1)K & \text{for } M^{(4)}; \\ (I-1)(K-1)J & \text{for } M^{(5)}; \\ (J-1)(K-1)I & \text{for } M^{(6)}. \end{array}$$

For example, the test for $M^{(6)}$ clearly pools I tests for independence of the conditional $J \times K$ subtable at each row, and each of these tests has $(J-1)(K-1)$ df, so the total df is $(J-1)(K-1)I$.

The corresponding log-linear models are

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)} + u_{23(jk)}$$

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{23(jk)}$$

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)}$$

For example, when the in the first model we fix k, we get an overparametrized version of the independence model $\log m_{ij} = u + u_{1(i)} + u_{2(j)}$.

Example, Continued

Returning to the school behavior data, we can fit all three models in R as

```
summary(fit4 <- glm(n ~ Beh*Schl + Risk*Schl,family=poisson))
summary(fit5 <- glm(n ~ Schl*Risk + Beh*Risk,family=poisson))
summary(fit6 <- glm(n ~ Beh*Risk + Beh*Schl,family=poisson))
```

and we discover

Resid Deviance	df	p-value	AIC
$G^2_{(4)} = 1.90$	3	0.60	60.65
$G^2_{(5)} = 4.12$	4	0.39	60.87
$G^2_{(6)} = 11.32$	4	0.02	68.07

Although two of these models fit well

$$\text{(classroom behavior)} \perp\!\!\!\perp \text{(family risk)} \mid \text{(school adversity)}$$
$$\text{(classroom behavior)} \perp\!\!\!\perp \text{(school adversity)} \mid \text{(family risk)}$$

the best fitting of the previous three models is simpler, with a similar AIC (60.39):

$$\text{(classroom behavior)} \perp\!\!\!\perp \text{(school adversity, family risk)}.$$

The Model of No Three-Way Interaction

Each of the models we have considered so far have involved the absence of one or more interaction in the log-linear model

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)}$$

or equivalently setting the corresponding odds-ratios equal to 1.

- The independence model sets $u_{12(ij)} = u_{23(jk)} = u_{13(ik)} = u_{123(ijk)} = 0$ $\forall i, j, k$.
- Independence of one factor from the other two sets the three-way interaction $u_{123(ijk)} = 0$ and two of the three two-way interactions to zero (which two depends on the model).
- Conditional independence sets the thre-way interaction $u_{123(ijk)} = 0$ and sets one of the two-way interactions to zero (again, depending on the model).

There is one more model commonly considered, that sets only the three-way interaction $u_{123(ijk)} = 0, \forall i, j, k$. From this model,

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} \qquad (*)$$

generically taking $a = e^u$, we see that the model for the cell means is

$$m_{ijk} = a \cdot a_{1(i)} \cdot a_{2(j)} \cdot a_{3(k)} \cdot a_{12(ij)} \cdot a_{13(ik)} \cdot a_{23(jk)}$$

Now consider an odds ratio in the $I \times J$ table in any of the K layers:

$$\begin{aligned} OR(i, j, i', j'|k) &= \frac{p_{ijk}p_{i'j'k}}{p_{i'jk}p_{ij'k}} = \frac{m_{ijk}m_{i'j'k}}{m_{i'jk}m_{ij'k}} \\ &= \frac{[a\, a_{1(i)}a_{2(j)}a_{3(k)}a_{12(ij)}a_{13(ik)}a_{23(jk)}][a\, a_{1(i')}a_{2(j')}a_{3(k)}a_{12(i'j')}a_{13(i'k)}a_{23(j'k)}]}{[a\, a_{1(i')}a_{2(j)}a_{3(k)}a_{12(i'j)}a_{13(i'k)}a_{23(jk)}][a\, a_{1(i)}a_{2(j')}a_{3(k)}a_{12(ij')}a_{13(ik)}a_{23(j'k)}]} \\ &= \frac{[a_{12(ij)}a_{13(ik)}a_{23(jk)}][a_{12(i'j')}a_{13(i'k)}a_{23(j'k)}]}{[a_{12(i'j)}a_{13(i'k)}a_{23(jk)}][a_{12(ij')}a_{13(ik)}a_{23(j'k)}]} = \frac{a_{12(ij)}a_{12(i'j')}}{a_{12(i'j)}a_{12(ij')}}, \end{aligned}$$

which does not depend on k!

Therefore, the model of no three-way interaction

$$M^{(7)} : \ (*) \ holds$$

is equivalent to the model in which the odds ratio is the same in every layer of the table.

For this model, there are not closed-form MLE's; they must be computed iteratively. However, the test statistics are as usual

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{(n_{ijk} - \hat{m}_{ijk}^{(7)})^2}{\hat{m}_{ijk}^{(7)}}$$

$$G^2 = 2 \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} n_{ijk} \log n_{ijk} / \hat{m}_{ijk}^{(7)}$$

with df. for the χ^2 tests above are $(I-1)(J-1)(K-1)$.

The model is easy to fit with R

```
summary(fit7<-glm(n~Beh*Risk*Schl-Beh:Risk:Schl,family=poisson))
```

This yields a fit of $G^2 = 0.94$ on 2 d.f. Clearly there is overfit here.

However, given that the model fits we can now estimate a common odds-ratio for (say) the interaction of classroom behavior and family risk, across all levels of school adversity:

```
n.mtx <- matrix(n,nrow=2,byrow=T)
mhat.mtx <- matrix(fitted(fit7),nrow=2,byrow=T)

OR(n.mtx[,(1:2)+0])$OR    # 2.29; CI is (0.12, 41.98)
OR(n.mtx[,(1:2)+2])$OR    # 1.18; CI is (0.27, 5.06)
OR(n.mtx[,(1:2)+4])$OR    # 5.00; CI is (0.34, 72.77)

OR(mhat.mtx[,(1:2)+0])$OR  # 1.80
OR(mhat.mtx[,(1:2)+2])$OR  # 1.80
OR(mhat.mtx[,(1:2)+4])$OR  # 1.80
```

When $M^{(7)}$ holds, the common odds ratio is more stably estimated since it uses all the data in the table, instead of one layer of data at a time (see next slide).

Following our work above we know that under the model of no three-factor interaction, for all k,

$$OR(i, j, i', j'|k) = \frac{a_{12(ij)}a_{12(i'j')}}{a_{12(i'j)}a_{12(ij')}} = \exp\left[u_{12(ij)} + u_{12(i'j')} - u_{12(i'j)} - u_{12(ij')}\right]$$

Once we have fitted the model above we can access the estimated u-terms and their variance-covariance matrix as

```
print(U <- summary(fit7)$coefficients[,1])
# (Intercept)        BehN       RiskR       SchlL       SchlM
#   0.4765646   1.0026434   0.3945061  -0.3748634   0.3486696
#  BehN:RiskR  BehN:SchlL  BehN:SchlM RiskR:SchlL RiskR:SchlM
#  -0.5898590   1.6615308   0.9269025  -0.6094350   0.9456019
V <- summary(fit7)$cov.unscaled
```

Under the model of no three-way interaction,

$$\log(OR[Beh, Risk|Schl]) = BehN{:}RiskN + BehD{:}RiskR - BehN{:}RiskR - BehD{:}RiskN$$

but we can see above that R set all but one of these equal to zero to identify the model. For R's reduced model, then, we have that the log common odds ratio is

$$\log(OR[Beh, Risk]) = \log(OR[Beh, Risk|Schl]) = -BehN{:}RiskR = -U[6] = 0.59$$

with standard error $\sqrt{V[6]} = 0.39$.

Note also that $\exp(-U[6]) = 1.80$, the same as we calculated from the fitted subtables on the slide above.

An approximate 95% CI for the log common odds ratio is

$$-U[6] + (-2, 2)\sqrt{V[6])} = (-0.19, 1.37)$$

which we can exponentiate to get an approximate 95% interval for the common odds ratio itself

$$\exp(-0.19, 1.37) = (0.83, 3.93)$$

Note that

- The final CI contains 1, suggesting that we cannot reject either
  - $\log(OR[Beh, Risk]) = 1$, i.e. (behavior) $\perp\!\!\!\perp$ (family risk)
  - $\log(OR[Beh, Risk|S\,chl]) = 1$, i.e. (behavior) $\perp\!\!\!\perp$ (family risk) | (school);
- The CI here is much shorter than any of the CI's for OR's calculated from the 2×2 subtables above.

  Using all of the data really does sharpen the inference!