

Multilevel generalized linear models

As with linear and logistic regressions, generalized linear models can be fit to multilevel structures by including coefficients for group indicators and then adding group-level models. We illustrate in this chapter with three examples from our recent applied research: an overdispersed Poisson model for police stops, a multinomial logistic model for storable voting, and an overdispersed Poisson model for social networks.

15.1 Overdispersed Poisson regression: police stops and ethnicity

We return to the New York City police example introduced in Sections 1.2 and 6.2, where we formulated the problem as an overdispersed Poisson regression, and here we generalize to a multilevel model. In order to compare ethnic groups while controlling for precinct-level variation, we perform multilevel analyses using the city's 75 precincts. Allowing precinct-level effects is consistent with theories of policing such as the “broken windows” model that emphasize local, neighborhood-level strategies. Because it is possible that the patterns are systematically different in neighborhoods with different ethnic compositions, we divide the precincts into three categories in terms of their black population: precincts that were less than 10% black, 10%–40% black, and more than 40% black. We also account for variation in stop rates between the precincts within each group. Each of the three categories represents roughly one-third of the precincts in the city, and we perform separate analyses for each set.

Overdispersion as a variance component

As discussed in Chapter 6, data that are fit by a generalized linear model are *overdispersed* if the data-level variance is higher than would be predicted by the model. Binomial and Poisson regression models are subject to overdispersion because these models do not have variance parameters to capture the variation in the data. This was illustrated at the end of Section 6.2 for the Poisson model of police stops.

As discussed at the end of Section 14.4, overdispersion can be directly modeled using a data-level variance component in a multilevel model. For example, we can extend the classical Poisson regression of Section 6.2,

$$\text{Poisson regression: } y_i \sim \text{Poisson}(u_i e^{X_i \beta}),$$

to the multilevel model,

$$\begin{aligned} \text{overdispersed Poisson regression: } y_i &\sim \text{Poisson}(u_i e^{X_i \beta + \epsilon_i}) \\ \epsilon_i &\sim N(0, \sigma_\epsilon^2). \end{aligned}$$

The new hyperparameter σ_ϵ measures the amount of overdispersion, with $\sigma_\epsilon = 0$ corresponding to classical Poisson regression. We shall use this model for the police stops.

Section 15.3 illustrates the use of the negative binomial model, which is a different overdispersed extension of the Poisson.

Multilevel Poisson regression model

For each ethnic group $e = 1, 2, 3$ and precinct p , we model the number of stops y_{ep} using an overdispersed Poisson regression with indicators for ethnic groups, a multilevel model for precincts, and using n_{ep} , the number of arrests recorded by the Department of Criminal Justice Services (DCJS) for that ethnic group in that precinct in the previous year (multiplied by 15/12 to scale to a 15-month period), as a baseline, so that $\log(\frac{15}{12}n_{ep})$ is an offset:

$$\begin{aligned} y_{ep} &\sim \text{Poisson}\left(\frac{15}{12}n_{ep}e^{\mu+\alpha_e+\beta_p+\epsilon_{ep}}\right) \\ \alpha_e &\sim N(0, \sigma_\alpha^2) \\ \beta_p &\sim N(0, \sigma_\beta^2) \\ \epsilon_{ep} &\sim N(0, \sigma_\epsilon^2), \end{aligned} \tag{15.1}$$

where the coefficients α_e control for ethnic groups, the β_p 's adjust for variation among precincts, and the ϵ_{ep} 's allow for overdispersion (see Chapter 6). The parameter σ_β represents variation in the rates of stops among precincts, and σ_ϵ represents variation in the data beyond that explained by the Poisson model. We are not particularly interested in the other variance parameter, σ_α ; instead we work with the individual coefficients, α_e .

Constraining a batch of coefficients to sum to 0. When comparing ethnic groups, we can look at the ethnicity coefficients relative to their mean:

$$\alpha_e^{\text{adj}} = \alpha_e - \bar{\alpha}, \text{ for } e = 1, 2, 3. \tag{15.2}$$

We examine the exponentiated coefficients $\exp(\alpha_e^{\text{adj}})$, which represent relative rates of stops compared to arrests, after controlling for precinct. Having done this, we also adjust the intercept of the model accordingly:

$$\mu^{\text{adj}} = \mu + \bar{\alpha}. \tag{15.3}$$

Now $\mu^{\text{adj}} + \alpha_e^{\text{adj}} = \mu + \alpha_e$ for each ethnic group e , and so we can use μ^{adj} and α_e^{adj} in place of μ and α without changing the model for the data.

In multilevel modeling, it makes sense to fit the full model (15.1) and then define the constrained parameters of interest as in (15.2) and (15.3), rather than trying to fit a model in which the original α parameters are constrained. We discuss this issue further in Section 19.4.

Separately fitting the model to different subsets of the data. By comparing to arrest rates, we can also separately analyze stops associated with different sorts of crimes. We do a separate comparison for each of four types of offenses ("suspected charges" as characterized on the official form): violent crimes, weapons offenses, property crimes, and drug crimes. For each, we model the number of stops y_{ep} by ethnic group e and precinct p for that crime type, using as a baseline the DCJS arrest rates n_{ep} for that crime type.

We thus estimate model (15.1) for twelve separate subsets of the data, corresponding to the four crime types and the three categories of precincts (less than 10% black population, 10–40% black, and more than 40% black). An alternative approach would be to fit a single model to all the data with interactions between crime types and precinct categories, and the other predictors in the model. It is simpler, however, to just fit the model separately to each of the twelve subsets.

Proportion black in precinct	Parameter	Crime type			
		Violent	Weapons	Property	Drug
< 10%	intercept, μ^{adj}	-0.85 (0.07)	0.13 (0.07)	-0.58 (0.21)	-1.62 (0.16)
	α_1^{adj} [blacks]	0.40 (0.06)	0.16 (0.05)	-0.32 (0.06)	-0.08 (0.09)
	α_2^{adj} [hispanics]	0.13 (0.06)	0.12 (0.04)	0.32 (0.06)	0.17 (0.10)
	α_3^{adj} [whites]	-0.53 (0.06)	-0.28 (0.05)	0.00 (0.06)	-0.08 (0.09)
	σ_β	0.33 (0.08)	0.38 (0.08)	1.19 (0.20)	0.87 (0.16)
	σ_ϵ	0.30 (0.04)	0.23 (0.04)	0.32 (0.04)	0.50 (0.07)
10–40%	intercept, μ^{adj}	-0.97 (0.07)	0.42 (0.07)	-0.89 (0.16)	-1.87 (0.13)
	α_1^{adj} [blacks]	0.38 (0.04)	0.24 (0.04)	-0.16 (0.06)	-0.05 (0.05)
	α_2^{adj} [hispanics]	0.08 (0.04)	0.13 (0.04)	0.25 (0.06)	0.12 (0.06)
	α_3^{adj} [whites]	-0.46 (0.04)	-0.36 (0.04)	-0.08 (0.06)	-0.07 (0.05)
	σ_β	0.49 (0.07)	0.47 (0.07)	1.21 (0.17)	0.90 (0.13)
	σ_ϵ	0.24 (0.03)	0.24 (0.03)	0.38 (0.04)	0.32 (0.04)
> 40%	intercept, μ^{adj}	-1.58 (0.10)	0.29 (0.11)	-1.15 (0.19)	-2.62 (0.12)
	α_1^{adj} [blacks]	0.44 (0.06)	0.30 (0.07)	-0.03 (0.07)	0.09 (0.06)
	α_2^{adj} [hispanics]	0.11 (0.06)	0.14 (0.07)	0.04 (0.07)	0.09 (0.07)
	α_3^{adj} [whites]	-0.55 (0.08)	-0.44 (0.08)	-0.01 (0.07)	-0.18 (0.09)
	σ_β	0.48 (0.10)	0.47 (0.11)	0.96 (0.18)	0.54 (0.11)
	σ_ϵ	0.24 (0.05)	0.37 (0.05)	0.42 (0.07)	0.28 (0.06)

Figure 15.1 *Estimates and standard errors for the intercept μ^{adj} , ethnicity parameters α_e^{adj} , and the precinct-level and precinct-by-ethnicity-level variance parameters σ_β and σ_ϵ , for the multilevel Poisson regression model (15.1), fit separately to three categories of precinct and four crime types. The estimates of $e^{\mu+\alpha_e}$ are displayed graphically in Figure 15.2, and alternative model specifications are shown in Figure 15.5. It would be preferable to display these results graphically. We show them in tabular form here to give a sense of the inferences that result from the 12 multilevel models that were fit to these data.*

Figure 15.1 shows the estimates from model (15.1) fit to each of four crime types in each of three categories of precinct. The standard-deviation parameters σ_β and σ_ϵ are substantial,¹ indicating the relevance of multilevel modeling for these data.

The parameters of most interest are the rates of stop (compared to previous year’s arrests) for each ethnic group, $e^{\mu+\alpha_e}$, for $e = 1, 2, 3$. We display these graphically in Figure 15.2. Stops for violent crimes and weapons offenses were the most controversial aspect of the stop-and-frisk policy (and represent more than two-thirds of the stops), but for completeness we display all four categories of crime here.

Figure 15.2 shows that, for the most frequent categories of stops—those associated with violent crimes and weapons offenses—blacks and hispanics were much more likely to be stopped than whites, in all categories of precincts. For violent crimes, blacks and hispanics were stopped 2.5 times and 1.9 times as often as whites, respectively, and for weapons crimes, blacks and hispanics were stopped 1.8 times and 1.6 times as often as whites. In the less common categories of stop, whites were slightly more often stopped for property crimes and more often stopped for drug crimes, in proportion to their previous year’s arrests in any given precinct.

¹ Recall that these effects are all on the logarithmic scale, so that an effect of 0.3, for example, corresponds to a multiplicative effect of $\exp(0.3) = 1.35$, or a 35% increase in the probability of being stopped.

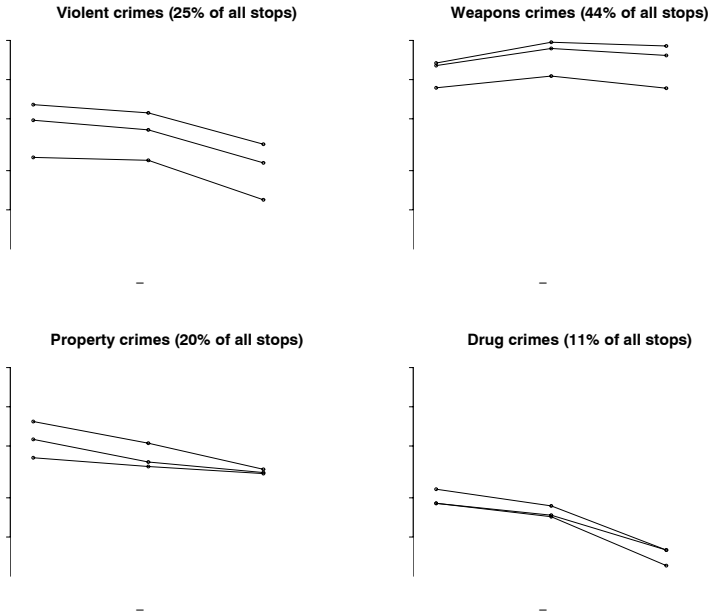


Figure 15.2 Estimated rates $e^{\mu+\alpha_e}$ at which people of different ethnic groups were stopped for different categories of crime, as estimated from multilevel regressions (15.1) using previous year's arrests as a baseline and controlling for differences between precincts. Separate analyses were done for the precincts that had less than 10%, 10%–40%, and more than 40% black population. For the most common stops—violent crimes and weapons offenses—blacks and hispanics were stopped about twice as often as whites. Rates are plotted on a logarithmic scale. Numerical estimates and standard errors appear in Figure 15.1.

Alternative model specifications

In addition to fitting model (15.1) as described above, we consider two forms of alternative specifications: first, fitting the same model but changing the batching of precincts; and, second, altering the role played in the model by the previous year's arrests. We compare the fits under these alternative models to assess the sensitivity of our findings to the details of model specification.

Modeling variability across precincts

The batching of precincts into three categories is convenient and makes sense—neighborhoods with different levels of minority populations differ in many ways, and fitting the model separately to each group of precincts is a way to include contextual effects. However, there is an arbitrariness to the division. We explore this by portioning the precincts into different numbers of categories and seeing how the model estimates change.

Including precinct-level predictors. Another approach to controlling for systematic variation among precincts is to include precinct-level predictors, which can be included along with the individual precinct-level effects in the multilevel model.

Parameter	Crime type			
	Violent	Weapons	Property	Drug
intercept, μ^{adj}	-0.66 (0.08)	0.08 (0.11)	-0.14 (0.24)	-0.98 (0.17)
α_1^{adj} [blacks]	0.41 (0.03)	0.24 (0.03)	-0.19 (0.04)	-0.02 (0.04)
α_2^{adj} [hispanics]	0.10 (0.03)	0.12 (0.03)	0.23 (0.04)	0.15 (0.04)
α_3^{adj} [whites]	-0.51 (0.03)	-0.36 (0.03)	-0.05 (0.04)	-0.13 (0.04)
ζ_1 [coeff for prop. black]	-1.22 (0.18)	0.10 (0.19)	-1.11 (0.45)	-1.71 (0.31)
ζ_2 [coef for prop. hispanic]	-0.33 (0.23)	0.71 (0.27)	-1.50 (0.57)	-1.89 (0.41)
σ_β	0.40 (0.04)	0.43 (0.04)	1.04 (0.09)	0.68 (0.06)
σ_ϵ	0.25 (0.02)	0.27 (0.02)	0.37 (0.03)	0.37 (0.03)

Figure 15.3 *Estimates and standard errors for the parameters of model (15.4) that includes proportion black and hispanic as precinct-level predictors, fit to all 75 precincts. The results for the parameters of interest, α_e^{adj} , are similar to those obtained by fitting the basic model separately to each of three categories of precincts, as displayed in Figures 15.1 and 15.2. As before, the model is fit separately to the data from four different crime types.*

As discussed earlier, the precinct-level information that is of greatest interest, and also that has greatest potential to affect our results, is the ethnic breakdown of the population. Thus we consider as regression predictors the proportion black and hispanic in the precinct, replacing model (15.1) by

$$y_{ep} \sim \text{Poisson}\left(\frac{15}{12}n_{ep}e^{\mu+\alpha_e+\zeta_1z_{1p}+\zeta_2z_{2p}+\beta_p+\epsilon_{ep}}\right), \tag{15.4}$$

where z_{1p} and z_{2p} represent the proportion of the people in precinct p who are black and hispanic, respectively. We also considered variants of model (15.4) including the quadratic terms, z_{1p}^2 , z_{2p}^2 , and $z_{1p}z_{2p}$, to examine sensitivity to nonlinearity.

Figure 15.3 shows the results from model (15.4), which is fit to all 75 precincts but controls for the proportion black and proportion hispanic in precincts. The inferences are similar to those obtained from the main analysis presented earlier. Including quadratic terms and interactions in the precinct-level model (15.4), and including the precinct-level predictors in the models fit to each of the three subsets of the data, similarly had little effect on the parameters of interest, α_e^{adj} .

Changing the number of precinct categories. Figure 15.4 displays the estimated rates of stops for violent crimes, compared to the previous year’s arrests, for each of the three ethnic groups, for analyses dividing the precincts into 5, 10, and 15 categories ordered by percent black population in precinct. For simplicity, we only give results for violent crimes; these are typical of the alternative analyses for all four crime types. For each of the three graphs in Figure 15.4, the model was separately estimated for each batch of precincts, and these estimates are connected in a line for each ethnic group. Compared to the upper-left plot in Figure 15.2, which shows the results from dividing the precincts into three categories, we see that dividing into more groups adds noise to the estimation but does not change the overall pattern of differences between the groups.

Modeling the relation of stops to previous year’s arrests

We also consider different ways of using the number of DCJS arrests n_{ep} in the previous year, which plays the role of a baseline (or offset, in generalized linear models terminology) in model (15.1).

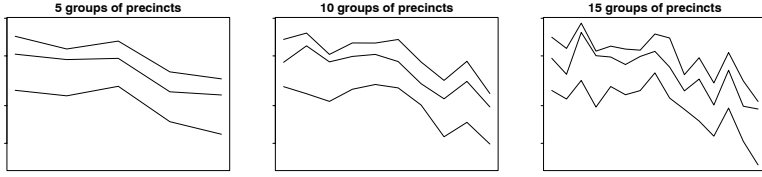


Figure 15.4 Estimated rates $e^{\mu+\alpha_e}$ at which people of different ethnic groups were stopped for violent crimes, as estimated from models dividing precincts into 5, 10, and 15 categories. For each graph, the top, middle, and lower lines correspond to blacks, hispanics, and whites, respectively. These plots show the same general patterns as the model with 3 categories (the upper-left graph in Figure 15.2) but with increasing levels of noise.

Using a linear predictor instead of an offset. Including the past arrest rate as an offset makes sense because we are interested in the rate of stops per crime, and we are using past arrests as a proxy for crime rate and for police expectations about the demographics of perpetrators. However, another option is to include the logarithm of the number of past arrests as a linear predictor instead:

$$y_{ep} \sim \text{Poisson} \left(\frac{15}{12} e^{\gamma \log n_{ep} + \mu + \alpha_e + \beta_p + \epsilon_{ep}} \right). \quad (15.5)$$

Model (15.5) reduces to the offset model (15.1) if $\gamma = 1$. We can thus fit (15.5) and see if the inferences for α_e^{adj} change compared to the earlier model that implicitly fixes γ to 1.

Two-stage models of arrest and stop rates. We can take this idea further by modeling past arrests as a proxy rather than the actual crime rate. We try this in two ways, for each labeling the true crime rate for each ethnicity in each precinct as θ_{ep} , with separate multilevel Poisson regressions for this year's stops and last year's arrests (as always, including the factor $\frac{15}{12}$ to account for our 15 months of stop data). In the first formulation, we model last year's arrests as Poisson distributed with mean θ :

$$\begin{aligned} y_{ep} &\sim \text{Poisson} \left(\frac{15}{12} \theta_{ep} e^{\mu + \alpha_e + \beta_p + \epsilon_{ep}} \right) \\ n_{ep} &\sim \text{Poisson}(\theta_{ep}) \\ \log \theta_{ep} &= \log N_{ep} + \tilde{\alpha}_e + \tilde{\beta}_p + \tilde{\epsilon}_{ep}. \end{aligned} \quad (15.6)$$

Here we are using N_{ep} , the population of ethnic group e in precinct p , as a baseline for the model of crime frequencies. The second-level error terms $\tilde{\beta}$ and $\tilde{\epsilon}$ are given normal hyperprior distributions as with model (15.1).

Our second two-stage model is similar to (15.6) but moving the new error term $\tilde{\epsilon}$ to the model for n_{ep} :

$$\begin{aligned} y_{ep} &\sim \text{Poisson} \left(\frac{15}{12} \theta_{ep} e^{\mu + \alpha_e + \beta_p + \epsilon_{ep}} \right) \\ n_{ep} &\sim \text{Poisson}(\theta_{ep} e^{\tilde{\epsilon}_{ep}}) \\ \log \theta_{ep} &= \log N_{ep} + \tilde{\alpha}_e + \tilde{\beta}_p. \end{aligned} \quad (15.7)$$

Under this model, arrest rates n_{ep} are equal to the underlying crime rates, θ_{ep} , on average, but with overdispersion compared to the Poisson error distribution.

Figure 15.5 displays parameter estimates from the models that differently incorporate the previous year's arrest rates n_{ep} . For conciseness we display results for violent crimes only, for simplicity including all 75 precincts in the models. (Similar results are obtained when fitting the model separately in each of three categories

Parameter	offset (15.1)	Model for previous year's arrests		
		regression (15.5)	2-stage (15.6)	2-stage (15.7)
intercept, μ^{adj}	-1.08 (0.06)	-0.94 (0.16)	-1.07 (0.06)	-1.13 (0.07)
α_1^{adj} [blacks]	0.40 (0.03)	0.41 (0.03)	0.40 (0.03)	0.42 (0.08)
α_2^{adj} [hispanics]	0.10 (0.03)	0.10 (0.03)	0.10 (0.03)	0.14 (0.09)
α_3^{adj} [whites]	-0.50 (0.03)	-0.51 (0.03)	-0.50 (0.03)	-0.56 (0.09)
γ [coef for $\log n_{ep}$]		0.97 (0.03)		
σ_β	0.51 (0.05)	0.51 (0.05)	0.51 (0.05)	0.27 (0.12)
σ_ϵ	0.26 (0.02)	0.26 (0.02)	0.24 (0.02)	0.67 (0.04)

Figure 15.5 *Estimates and standard errors for parameters under model (15.1) and three alternative specifications for the previous year's arrests n_{ep} : treating $\log(n_{ep})$ as a predictor in the Poisson regression model (15.5), and the two-stage models (15.6) and (15.7). For simplicity, results are displayed for violent crimes only, for the model fit to all 75 precincts. The three α_e^{adj} parameters are nearly identical under all four models, with the specification affecting only the intercept.*

of precincts, and for the other crime types.) The first two columns of Figure 15.5 show the result from our main model (15.1) and the alternative model (15.5), which includes $\log n_{ep}$ as a regression predictor. The two models differ only in that the first restricts γ to be 1, but as we can see, γ is estimated very close to 1 in the regression formulation, and the coefficients α_e^{adj} are essentially unchanged. (The intercept changes a bit because $\log n_{ep}$ does not have a mean of 0.)

The last two columns in Figure 15.5 show the estimates from the two-stage regression models (15.6) and (15.7). The models differ in their estimates of the variance parameters σ_β and σ_ϵ , but the estimates of the key parameters α_e^{adj} are essentially the same as in the original model.

We also performed analyses including indicators for the month of arrest. Rates of stops were roughly constant over the 15-month period and did not add anything informative to the comparison of ethnic groups.

15.2 Ordered categorical regression: storable votes

At the end of Section 6.5, we described data on “storable votes” from a study in experimental economics. Each student in the experiment played a voting game 30 times, in each play receiving an input x between 1 and 100 and then giving a response of 1, 2, or 3. As described in Section 6.5, we fit an ordered multinomial logistic regression to these outcomes, fitting the model separately for each student.

We now return to these data and fit a multilevel model. With only 30 data points on each student, there is some uncertainty in the estimate for the parameters for each student, and a multilevel model should allow more precise inferences—at least if the students are similar to each other. (As discussed in Section 12.7, multilevel modeling is most effective when the parameters in a batch are similar to each other.)

In our multilevel model for the storable votes study, we simply expand the model (6.11) allowing each of the parameters to vary by student j :

$$\begin{aligned} c_{j\ 1.5} &\sim N(\mu_{1.5}, \sigma_{1.5}^2) \\ c_{j\ 2.5} &\sim N(\mu_{2.5}, \sigma_{2.5}^2) \\ \log \sigma_j &\sim N(\mu_{\log \sigma}, \sigma_{\log \sigma}^2), \end{aligned}$$

with the hyperparameters $\sigma_{1.5}, \sigma_{2.5}, \mu_{\log \sigma}, \sigma_{\log \sigma}$ estimated from the data. The un-

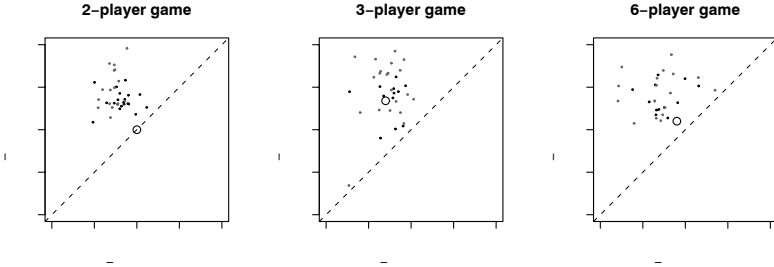


Figure 15.6 Estimated cutpoints $c_{j\,1.5}, c_{j\,2.5}$ for students j in three experiments on storable voting. Each set of parameters was estimated using a multilevel ordered logistic regression model. The dots all fall above the diagonal line because the 2–3 cutpoint is restricted to be at least as high as the 1–2 cutpoint. The large circle on each graph corresponds to a theoretically optimal cutpoint for each game. Deviations of actual strategies from the theoretical optimum are of interest.

pooled model, estimating the parameters separately for each student, corresponds to hyperparameters set to $\sigma_{1.5} = \infty, \sigma_{1.5} = \infty, \sigma_{\log \sigma} = \infty$. (The log transformation is used for σ_j so that its distribution is restricted to be positive.)

Having fit the multilevel model, we can now examine how the parameters vary among the students in the experiment. The parameters $c_{1.5}$ and $c_{2.5}$ correspond to the cutpoints for voting 1, 2, 3; variation in these two parameters thus corresponds to variation in the monotonic strategies used by the students in the voting game. The parameter σ represents the variation in the model, with higher values indicating a less purely monotonic strategy.

Figure 15.6 shows estimated cutpoints $c_{j\,1.5}, c_{j\,2.5}$ for students j in three different storable voting experiments. The general distribution of the cutpoints is similar in the three games, despite the different positions of the theoretically optimal cutpoints (indicated by a solid circle in each plot). This suggests that students are following a range of approximately monotonic strategies that are determined more by the general structure of the game than by detailed strategic reasoning. For the practical implementation of storable votes, this finding is somewhat encouraging, in that the “players” are using storable votes to express their preferences on individual issues without feeling the need to manipulate the voting game.

15.3 Non-nested negative-binomial model of structure in social networks

Understanding the structure of social networks, and the social processes that form them, is a central concern of sociology for both theoretical and practical reasons. Networks have been found to have important implications for social mobility, getting a job, the dynamics of fads and fashion, attitude formation, and the spread of infectious disease.

This section discusses how we used an overdispersed Poisson regression model to learn about social structure. We fit the model to a random-sample survey of Americans who were asked,² “How many X’s do you know?” for a variety of characteristics X, defined by name (Michael, Christina, Nicole, ...), occupation (postal worker,

² The respondents were told, “For the purposes of this study, the definition of knowing someone is that you know them and they know you by sight or by name, that you could contact them,

pilot, gun dealer, ...), ethnicity (Native American), or experience (prisoner, auto accident victim, ...). For a complete list of the groups, see Figure 15.10 on page 338.

Background

The original goals of the survey were (1) to estimate the distribution of individuals' network size, defined to be the number of acquaintances, in U.S. population and (2) to estimate the sizes of certain subpopulations, especially those that are hard to count using regular survey results.

Before describing our regression modeling, we explain how these two estimates can be roughly computed from the data. First, to estimate the social network size of a single individual, one can use his or her total response for a set of subpopulations with known sizes, and then scale up using the sizes of these groups in the population.

To illustrate, suppose you know 2 persons named Nicole. At the time of the survey, there were 358,000 Nicoles out of 280 million Americans. Thus, your 2 Nicoles represent a fraction $\frac{2}{358,000}$ of all the Nicoles. Extrapolating to the entire country yields an estimate of $\frac{2}{358,000} \cdot (280 \text{ million}) = 1560$ people known by you. A more precise estimate can be obtained by averaging these estimates using a range of different groups. This is only a crude inference since it assumes that everyone has equal propensity to know someone from each group. However, as an estimation procedure, it has the advantage of not requiring a respondent to recall his or her entire network, which typically numbers in the hundreds.

The second use for which this survey was designed is to estimate the size of certain hard-to-count populations. To do this, one can combine the estimated network size information with the responses to the questions about how many people the respondents know in the hard-to-count population.

For example, the survey respondents know, on average, 0.63 homeless people. If it is estimated that the average network size is 750, then homeless people represent a fraction of $\frac{0.63}{750}$ of an average person's social network. The total number of homeless people in the country can then be estimated as $\frac{0.63}{750} \cdot (280 \text{ million}) = 0.24 \text{ million}$. This estimate relies on idealized assumptions (most notably, that homeless persons have the same social network size, on average, as Americans as a whole) but can be used as a starting point for estimating the sizes of groups that are difficult to measure directly.

Our regression model performs more precise versions of these estimates but, more interestingly, uses overdispersion in the data to reveal information about social structure in the acquaintanceship network. We use the variation in response data to study the heterogeneity of relative propensities for people to form ties to people in specific groups.

that they live within the United States, and that there has been some contact (either in person, by telephone or mail) in the past two years."

In addition, the data have some minor problems. For the fewer than 0.4% of responses that were missing, we followed the usual practice with this sort of unbalanced data of assuming an ignorable model (that is, constructing the likelihood using the observed data). Sometimes responses were categorized, and then we use the central value in the bin (for example, imputing 7.5 for the response "5-10"). In addition, to correct for some responses that were suspiciously large (for example, a person claiming to know more than 50 Michaels), we truncate all responses at 30. (As a sensitivity analysis, we tried changing the truncation point to 50; this had essentially no effect on our results.) We also inspected the data using scatterplots of responses, which revealed a respondent who was coded as knowing 7 persons of every category. We removed this case from the dataset.

Modeling the data

For respondent $i = 1, \dots, 1370$ and subpopulations $k = 1, \dots, 32$ (see Figure 15.10 for the list of groups asked about in the survey), we use the notation y_{ik} for the number of persons in group k known by person i . We consider three increasingly complicated models for these data.

Most simply, the *Erdos-Renyi* model is derived from the assumption that any person in the population is equally likely to know any other person, so that social links occur completely at random. A consequence of this assumption is that y_{ik} should have a Poisson distribution with mean proportional to the size of subpopulation k .

More generally, one could consider a *null model* in which individuals i have varying levels of gregariousness or popularity, so that the expected number of persons in group k known by person i will be proportional to this gregariousness parameter, which we label a_i . Departure from this model—patterns not simply explained by differing group sizes or individual popularities—can be viewed as evidence of structured social acquaintance networks.

As we shall see, the null model (and its special case, the Erdos-Renyi model) fails to account for much of social reality, including the “How many X’s do you know?” survey data. In fact, some individuals are much more likely to know people of some groups. To capture this aspect of social structure, we set up a new model in which, for any individual, the relative propensities to know people from different groups vary. We call this the *overdispersed model* since this variation results in overdispersion in the count data y_{ik} .

Figure 15.7 shows some of the data—the distributions of responses to the questions, “How many people named Nicole do you know?” and “How many Jaycees do you know?” along with the expected distributions under the Erdos-Renyi model, our null model, and our overdispersed model. We chose these two groups to plot because they are close in average number known (0.9 Nicoles, 1.2 Jaycees) but have much different distributions. The distribution for Jaycees has much more variation, with more zero responses and more responses in the upper tail.³

Comparing the models, the Erdos-Renyi model implies a Poisson distribution for the responses to each question, whereas the other models allow for more dispersion. The distributions under the null model are more dispersed to reflect that social network sizes vary greatly among individuals. The distributions under the overdispersed model are even more spread out—especially for the Jaycees—reflecting estimated variation in relative propensities for people to know members of the Jaycees. As we shall see, both these sources of variation—variation in social network sizes and variations in relative propensities to form ties to specific groups—can be estimated from the data in the survey.

The three models can be written as follows in statistical notation as $y_{ik} \sim \text{Poisson}(\lambda_{ik})$, with increasingly general forms for λ_{ik} :

$$\begin{aligned} \text{Erdos-Renyi model:} \quad & \lambda_{ik} = ab_k \\ \text{our null model:} \quad & \lambda_{ik} = a_i b_k \\ \text{our overdispersed model:} \quad & \lambda_{ik} = a_i b_k g_{ik}. \end{aligned}$$

The null model goes beyond the Erdos-Renyi model by allowing the gregariousness parameters to differ between individuals (a_i) and prevalence parameters

³ “Jaycees” are members of the Junior Chamber of Commerce, a community organization of people between the ages of 21 and 39. Because the Jaycees are a social organization, it makes sense that not everyone has the same propensity to know one—people who are in the social circle of one Jaycee are particularly likely to know others.

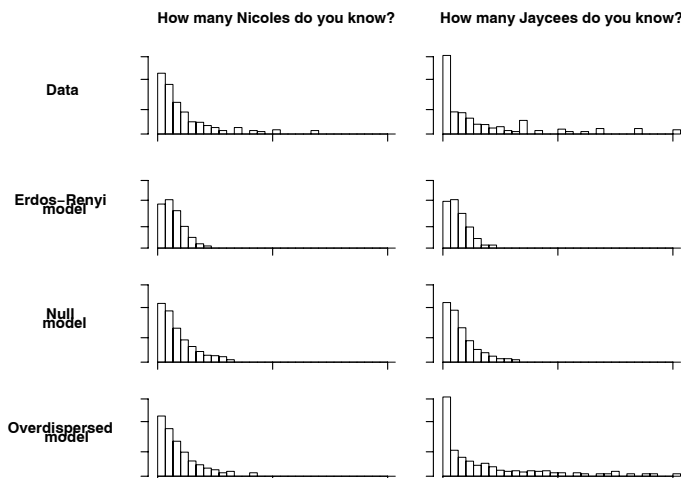


Figure 15.7 Distributions of responses to “How many persons do you know named Nicole?” and “How any Jaycees do you know?” from survey data and from random simulations under three fitted models: the Erdos-Renyi model (completely random links), our null model (some people more gregarious than others, but uniform relative propensities for people to form ties to all groups), and our overdispersed model (variation in gregariousness and variation in propensities to form ties to different groups). The four models are listed in order of increasing dispersion than the one above, with the overdispersed model fitting the data reasonably well. The propensities to form ties to Jaycees show much more variation than the propensities to form ties to Nicoles, and hence the Jaycees counts are much more overdispersed. (The data also show minor idiosyncrasies such as small peaks at the responses 10, 15, 20, and 25. All values greater than 30 have been truncated at 30.)

between groups (b_k). The overdispersed model generalized further by allowing different individuals to differ in their relative propensities to form ties to people in specific groups (g_{ik}). When fitting the overdispersed model, we will not attempt to estimate all the individual g_{ik} ’s; rather, we estimate certain properties of their distributions.

The overdispersed model

Overdispersion in these data can arise if the relative propensity for knowing someone in prison, for example, varies from respondent to respondent. We can write this in the generalized linear model framework as

$$y_{ik} \sim \text{Poisson}(e^{\alpha_i + \beta_k + \gamma_{ik}}), \quad (15.8)$$

where each $\gamma_{ik} = \log(g_{ik}) \equiv 0$ in the null model. For each subpopulation k , we let the multiplicative factors $g_{ik} = e^{\gamma_{ik}}$ follow a gamma distribution with a value of 1 for the mean and a value of $1/(\omega_k - 1)$ for the shape parameter.⁴ This distribution

⁴ If we wanted, we could allow the mean of the gamma distribution to vary also; however, this would be redundant with a location shift in β_k ; see model (15.8). The mean of the gamma distribution for the $e^{\gamma_{ik}}$ ’s cannot be identified separately from β_k , which we are already estimating from data.

is convenient because then the γ 's can be integrated out of model (15.8) to yield

$$y_{ik} \sim \text{negative-binomial}(\text{mean} = e^{\alpha_i + \beta_k}, \text{overdispersion} = \omega_k). \quad (15.9)$$

(The usual parametrization of this distribution is $y \sim \text{Negative-binomial}(A, B)$, but for this example it is more convenient to express in terms of the mean $\lambda = A/B$ and overdispersion $\omega = 1 + 1/B$.) Setting $\omega_k = 1$ corresponds to setting the shape parameter in the gamma distribution to ∞ , which in turn implies that the g_{ik} 's have zero variance, reducing to the null model with no overdispersion. Higher values of ω_k correspond to overdispersion—that is, more variation in the distribution of connections involving group k than would be expected under the Poisson regression, as would be expected if there is variation among respondents in the relative propensity to know someone in group k .

Our primary goal in fitting model (15.9) is to estimate the overdispersions ω_k and thus learn about biases that exist in the formation of social networks. As a byproduct, we also estimate the gregariousness parameters α_i and the group prevalence parameters β_k . As we have discussed, $a_i = e^{\alpha_i}$ represents the expected number of persons known by respondent i , and $b_k = e^{\beta_k}$ is the proportion of subgroup k in the social network, counting each link in the network in proportion to the probability that it will be recalled by a random respondent in the survey.

We estimate the α 's, β 's, and ω 's with a multilevel model. The respondent parameters α_i are assumed to follow a normal distribution with unknown mean μ_α and standard deviation σ_α . We similarly fit the subgroup parameters β_k with a normal distribution $N(\mu_\beta, \sigma_\beta^2)$, with these hyperparameters also estimated from the data. For simplicity, we assign independent Uniform(0,1) prior distributions to the overdispersion parameters on the inverse scale: $p(1/\omega_k) \propto 1$. (The overdispersions ω_k are constrained to the range $(1, \infty)$, and so it is convenient to put a model on the inverses $1/\omega_k$, which fall in the interval $(0, 1)$.)

Nonidentifiability

The model as given is not fully identified. Any constant C can be added to all the α_i 's and subtracted from all the β_k 's, and the likelihood will remain unchanged (since it depends on these parameters only through sums of the form $\alpha_i + \beta_k$). If we also add C to μ_α and subtract C from μ_β , then the prior density also is unchanged as well. It would be possible to identify the model by anchoring it at some arbitrary point—for example, setting μ_α to zero—but we prefer to let all the parameters float, since including this redundancy can allow the iterative computations to converge more quickly (a point we discuss more generally in Section 19.4).

We choose a constant C so that the parameters e^{β_k} correspond to the proportion of the entire social network associated with subpopulation k . We perform this renormalization using the known population sizes of the named subgroups (the number of Michaels, Nicoles, and so forth, in the population), which can be obtained from the U.S. Census. The choice of C is somewhat elaborate, including adjustments for the rare and common male and female names, and we do not present the details here.

Computation

We would like to fit the model in Bugs, but with more than 1400 parameters (an α_j for each of the 1370 survey respondents, a β_k for each of the 32 subgroups, and some hyperparameters), it runs too slowly. Instead, we use a more efficient program called

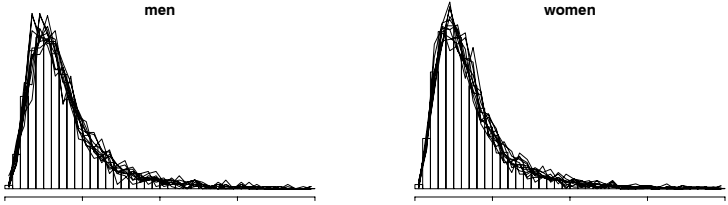


Figure 15.8 *Estimated distributions of “gregariousness” or expected number of acquaintances, $a_i = e^{\alpha_i}$ from the fitted model. Men and women have similar distributions (with medians of about 610 and means about 750), with a great deal of variation among persons. The overlain lines are posterior simulation draws indicating inferential uncertainty in the histograms.*

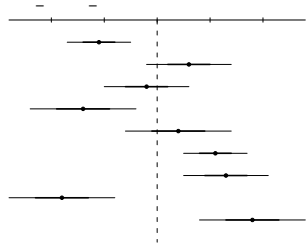


Figure 15.9 *Coefficients of the regression of estimated log gregariousness parameters α_i on personal characteristics. Because the regression is on the logarithmic scale, the coefficients (with the exception of the constant term) can be interpreted as proportional differences: thus, with all else held constant women have social network sizes 11% smaller than men, persons over 65 have social network sizes 14% lower than others, and so forth. The R^2 of the model is only 10%, indicating that these predictors explain very little of the variation in gregariousness in the population.*

Umucs (Universal Markov Chain sampler) which requires the posterior distribution to be specified as an R function. Details appear in Section 18.7.

We fit the model first using all the data and then separately for the male and female respondents (582 men and 784 women, with 4 individuals excluded due to missing gender information). Fitting the models separately for men and women makes sense since many of the subpopulations under study are single-sex groups. As we shall see, men tend to know more men and women tend to know more women, and more subtle sex-linked patterns also occur.

The distribution of social network sizes a_i

Figure 15.8 displays estimated distributions of the gregariousness parameters $a_i = e^{\alpha_i}$ for the survey respondents, showing separate histograms of the posterior simulations from the model estimated separately to the men and the women.

The spread in each of the histograms of Figure 15.8 almost entirely represents population variability. The model allows us to estimate the individual a_i ’s to within a coefficient of variation of about $\pm 25\%$. When taken together this allows us to

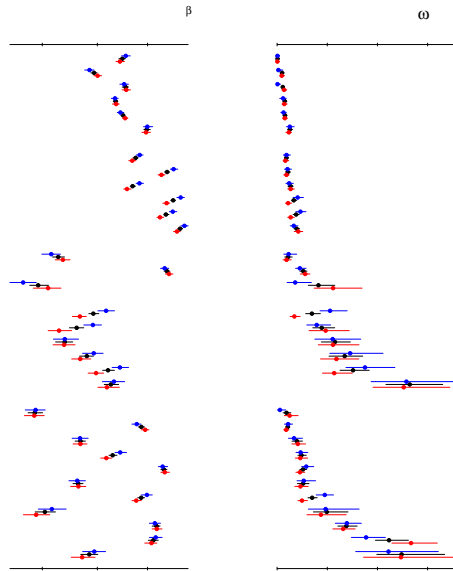


Figure 15.10 For each group k in the “How many X’s do you know?” survey, we plot the estimate (and 95% interval) of b_k and ω_k . The estimates and uncertainty lines are clustered in groups of three; for each group, the top, middle, and bottom dots/lines correspond to men, all respondents, and women, respectively. The groups are listed in categories—female names, male names, female, male (or primarily male), and mixed-sex groups—and in increasing average overdispersion within each category.

estimate the distribution very precisely. This precision can be seen in the solid lines that are overlaid on Figure 15.8 and represent inferential uncertainty.

Figure 15.9 presents a simple regression analysis estimating some of the factors predictive of $\alpha_i = \log(a_i)$, using questions asked of the respondents in the survey. These explanatory factors are relatively unimportant in explaining social network size: the regression summarized in Figure 15.9 has an R^2 of only 10%. The largest effects are that persons with a college education, a job outside the home, and high incomes know more people, and persons over 65 and those having low incomes know fewer people. These factors all have effects in the range of 10%–20%.

Relative sizes b_k of subpopulations

We now consider the parameters describing the 32 subpopulations. The left panels of Figure 15.10 show the 32 subpopulations k and the estimates of $b_k = e^{\beta_k}$, the proportion of links in the network that go to a member of group k . The right panel displays the estimated overdispersions ω_k . The sample size is large enough that the 95% error bars are tiny for the β_k ’s and reasonably small for the ω_k ’s as well. (It is a general property of statistical estimation that mean parameters, such as the β ’s in this example, are easier to estimate than dispersion parameters such as the ω ’s.) The figure also displays the separate estimates from the men and women.

Considering the β ’s first, the clearest pattern in Figure 15.10 is that respondents

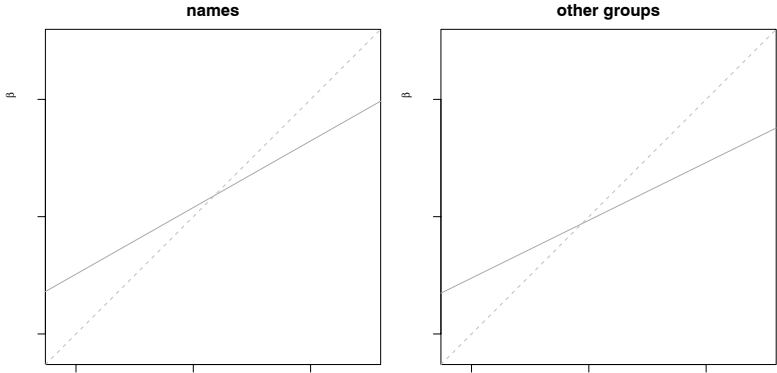


Figure 15.11 *Log-log plot of estimated prevalence of groups in the population (as estimated from the “How many X’s do you know?” survey) plotted versus actual group size (as determined from public sources). Names and other groups are plotted separately, on a common scale, with fitted regression lines shown. The solid lines have slopes 0.53 and 0.42, compared to a theoretical slope of 1 (as indicated by the dotted lines) that would be expected if all groups were equally popular, on average, and equally recalled by respondents.*

of each sex tend to know more people in groups of their own sex. We can also see that the 95% intervals are wider for groups with lower β ’s, which makes sense because the data are discrete and, for these groups, the counts y_{ik} are smaller and provide less information.

Another pattern in the estimated b_k ’s is the way that they scale with the size of group k . One would expect an approximate linear relation between the number of people in group k and our estimate for b_k : that is, on a graph of $\log b_k$ versus $\log(\text{group size})$, we would expect the groups to fall roughly along a line with slope 1. As can be seen in Figure 15.11, however, this is not the case. Rather, the estimated prevalence increases approximately with square root of population size, a pattern that is particularly clean for the names. This relation has also been observed by earlier researchers.

Discrepancies from the linear relation can be explained by difference in average degrees (for example, as members of a social organization, Jaycees would be expected to know more people than average, so their b_k should be larger than another group of equal numbers), inconsistency in definitions (for example, what is the definition of an American Indian?), and ease or difficulty of recall (for example, a friend might be a twin without you knowing it, whereas you would probably know whether she gave birth in the past year).

This still leaves unanswered the question of why square root (that is, a slope of $1/2$ in the log-log plot) rather than linear (a slope of 1). It is easier to recall rare persons and events, whereas more people in more common categories are easily forgotten. You will probably remember every Ulysses you ever met, but it can be difficult to recall all the Michaels and Roberts you know even now. The recall process for rarer names reaches deeper into one’s personal network of acquaintances.

Another pattern in Figure 15.11 is that the line for the names is higher than for the other groups. We suppose that is because, for a given group size, it is easier

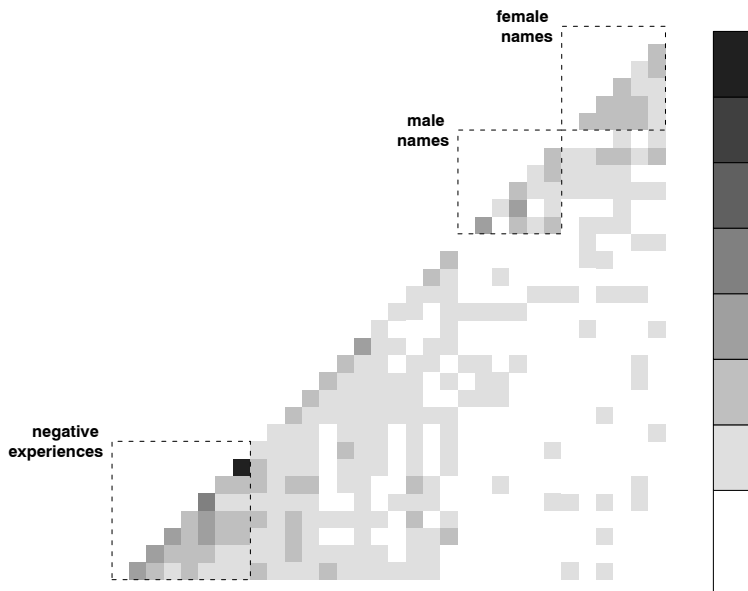


Figure 15.12 *Correlations of the residuals r_{ik} among the survey respondents: people who know more HIV-positive persons know more AIDS patients, etc. The groups other than the names are listed based on a clustering algorithm that maximizes correlations between nearby groups.*

to recall names than characteristics. After all, you know the name of almost all your acquaintances, but you could easily be unaware that a friend has diabetes, for example.

Overdispersion parameters ω_k for subpopulations

Recall that we introduced the overdispersed model to attempt to estimate the variability in respondents' relative propensities to form ties to members of different groups. For groups where $\omega_k = 1$, we can conclude that there is no variation in these relative propensities, so that persons in group k appear to be randomly distributed in the social network. However, for groups where ω_k is much greater than 1, the null model is a poor fit to the data, and persons in group k do not appear to be uniformly distributed in the social network. Rather, overdispersion implies that the relative propensity to know persons of group k varies in the general population.

The right panel of Figure 15.10 displays the estimated overdispersions ω_k , and they are striking. First, we observe that the names have overdispersions of between 1 and 2—that is, indicating very little variation in relative propensities. In contrast, the other groups have a wide range of overdispersions, with the highest values for Jaycees and American Indians (two groups with dense internal networks) and homeless persons, who are both geographically and socially localized.

These results are consistent with our general understanding and also potentially reveal patterns that would not be apparent without this analysis. For example, it

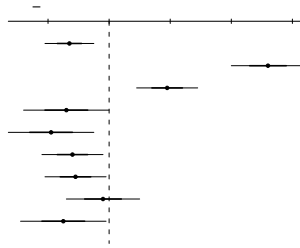


Figure 15.13 *Coefficients of the regression of residuals for the “How many males in federal prison do you know?” question on personal characteristics. Being male, nonwhite, young, unmarried, etc., are associated with knowing more people than expected in federal prison. However, the R^2 of the regression is only 11%, indicating that most of the variation in the data is not captured by these predictors.*

is no surprise that there is high variation in the propensity to know someone who is homeless, but it is perhaps surprising that AIDS patients are less overdispersed than HIV-positive persons, or that new business owners are no more overdispersed than new mothers.

Analysis of residuals

Further features of these data can be studied using residuals from the overdispersed model. A natural object of study is correlation: for example, do people who know more Anthonys tend to know more gun dealers (after controlling for the fact that social network sizes differ, so that anyone who knows more X’s will tend to know more Y’s)? For each survey response y_{ik} , we can define the standardized residual as

$$\text{residual: } r_{ik} = \sqrt{y_{ik}} - \sqrt{a_i b_k}, \quad (15.10)$$

the excess people known after accounting for individual and group parameters. (It is standard to compute residuals of count data on the square root scale to stabilize the variance.)

For each pair of groups k_1, k_2 , we can compute the correlation of their vectors of residuals; Figure 15.12 displays the matrix of these correlations. Care must be taken when interrupting the figure. At first, it may appear that the correlations are quite small. However, this is in some sense a natural result of our model. That is, if the correlations were all positive for a specific group, then the popularity b of that group would increase.

Several patterns can be seen in Figure 15.12. First, there is a slight positive correlation within male and female names. Second, perhaps more interesting sociologically, there is a positive correlation between the categories that can be considered negative experiences—homicide, suicide, rape, died in a car accident, homelessness, and being in prison. That is, someone with a higher relative propensity to know someone with one bad experience is also likely to have a higher propensity to know someone who had a different bad experience.

Instead of correlating the residuals, we could have examined the correlations of the raw data. However, these would be more difficult to interpret because we would find positive correlations everywhere, for the uninteresting reason that some

respondents know many more people than others, so that if you know more of any category of person, you are likely to know more in just about any other category.

One can also model the residuals based on individual-level predictors. For example, Figure 15.13 shows the estimated coefficients of a regression model fit to the residuals of the null model for the “How many males do you know in state or federal prison?” question. It is no surprise that being male, nonwhite, young, unmarried, less educated, unemployed, and so forth are associated with knowing more men than expected in state or federal prison. However, the R^2 of the regression model is only 11%.

As with the correlation analysis, by performing this regression on the residuals and not the raw data, we are able to focus on the relative number of prisoners known, without being distracted by the total network size of each respondent (which we have separately analyzed in Figure 15.9).

15.4 Bibliographic note

Examples of hierarchical generalized linear modeling in the literature include Novick, Lewis, and Jackson (1973), Clayton and Kaldor (1987), Karim and Zeger (1992), Barry et al. (2003), and many others. Johnson (1996, 1997) applies multilevel ordered discrete regression models to student grades.

The New York City police analysis comes from Gelman, Fagan, and Kiss (2006). Some references on neighborhoods and crime include Wilson and Kelling (1982), Skogan (1990), and Sampson, Raudenbush, and Earls (1997). The storable votes experiment and model are described in Casella, Gelman, and Palfrey (2006). The social network example comes from Zheng, Salganik, and Gelman (2006), with the analysis based on survey data of Killworth et al. (1998) and McCarty et al. (2000). For more on statistical models for network data, see Newman (2003), Hoff, Raftery, and Handcock (2002), and Hoff (2003, 2005).

15.5 Exercises

1. Multilevel ordered logit: using the National Election Study data from the year 2000 (data available in the folder `nes`), set up an ordered logistic regression predicting the response to the question on vote intention (0 = Gore, 1 = no opinion or other, 2 = Bush), given the predictors shown in Figure 5.4 on page 84, and with varying intercepts for states. (You will fit the model using Bugs in Exercise 17.10.)
2. Using the same data as the previous exercise:
 - (a) Formulate a model to predict party identification (which is on a five-point scale) using ideology and demographics with a multilevel ordered categorical model allowing both the intercept and the coefficient on ideology to vary over state.
 - (b) Fit the model using `lmer()` and discuss your results.
3. Multinomial choice models: fit and evaluate a multilevel model to the Academy Awards data from Exercise 6.11.

Part 2B: Fitting multilevel models

We next explain how to fit multilevel models in Bugs, as called from R. We illustrate with several examples and discuss some general issues in model fitting and tricks that can help us estimate multilevel models using less computer time. We also present the basics of Bayesian inference (as a generalization of the least squares and maximum likelihood methods used for classical regression), which is the approach used in problems such as multilevel models with potentially large numbers of parameters.

Appendix C discusses some software that is available to quickly and approximately fit multilevel models. We recommend using Bugs for its flexibility in modeling; however, these simpler approaches can be useful to get started, explore models quickly, and check results.