16.4 Example: hierarchical Poisson regression for police stops

There have been complaints in New York City and elsewhere that the police harass members of ethnic minority groups. In 1999, the New York State Attorney General's Office instigated a study of the New York City police department's 'stop and frisk' policy: the lawful practice of 'temporarily detaining, questioning, and, at times, searching civilians on the street.' The police have a policy of keeping records on stops, and we obtained all these records (about 175,000 in total) for a fifteen-month period in 1998–1999. We analyzed these data to see to what extent different ethnic groups were stopped by the police. We focused on blacks (African Americans), hispanics (Latinos), and whites (European Americans). The ethnic categories were as recorded by the police making the stops. We excluded members of other ethnic groups (about 4% of the stops) because of the likelihood of ambiguities in classifications. (With such a low frequency of 'other,' even a small rate of misclassifications could cause large distortions in the estimates for that group. For example, if only 4% of blacks, hispanics, and whites were mistakenly labeled as 'other,' then this would nearly double the estimates for the 'other' category while having little effect on the three major groups.)

Aggregate data

Blacks and hispanics represented 51% and 33% of the stops, respectively, despite comprising only 26% and 24%, respectively, of the population of the city. Perhaps a more relevant comparison, however, is to the number of crimes committed by members of each ethnic group.

Data on actual crimes are not available, so as a proxy we used the number of arrests within New York City in 1997 as recorded by the Division of Criminal Justice Services (DCJS) of New York State. These were deemed to be the best available measure of local crime rates categorized by ethnicity. We used these numbers to represent the frequency of crimes that the police might suspect were committed by members of each group. When compared in that way, the ratio of stops to DCJS arrests was 1.24 for whites, 1.54 for blacks, and 1.72 for hispanics: based on this comparison, blacks are stopped 23% and hispanics 39% more often than whites.

Regression analysis to control for precincts

The analysis so far looks at average rates for the whole city. Suppose the police make more stops in high-crime areas but treat the different ethnic groups equally within any locality. Then the citywide ratios could show strong differences between ethnic groups even if stops are entirely determined by location rather than ethnicity. In order to separate these two kinds of predictors, we performed hierarchical analyses using the city's 75 precincts. Because it is possible that the patterns are systematically different in neighborhoods with different ethnic compositions, we divided the precincts into three categories in terms of their black population: precincts that were less than 10% black, 10%–40% black, and over 40% black. Each of these represented roughly 1/3 of the precincts in the city, and we performed separate analyses for each set.

For each ethnic group $e = 1, 2, 3$ and precinct p, we modeled the number of stops y_{ep} using an overdispersed Poisson regression with indicators for ethnic groups, a hierarchical model for precincts, and using n_{ep}, the number of DCJS arrests for that ethnic group in that precinct (multiplied by 15/12 to scale to a fifteen-month period), as an offset:

$$\begin{aligned} y_{ep} &\sim \text{Poisson}\left(n_{ep} e^{\alpha_e + \beta_p + \epsilon_{ep}}\right) \\ \beta_p &\sim \text{N}(0, \sigma_\beta^2) \end{aligned}$$

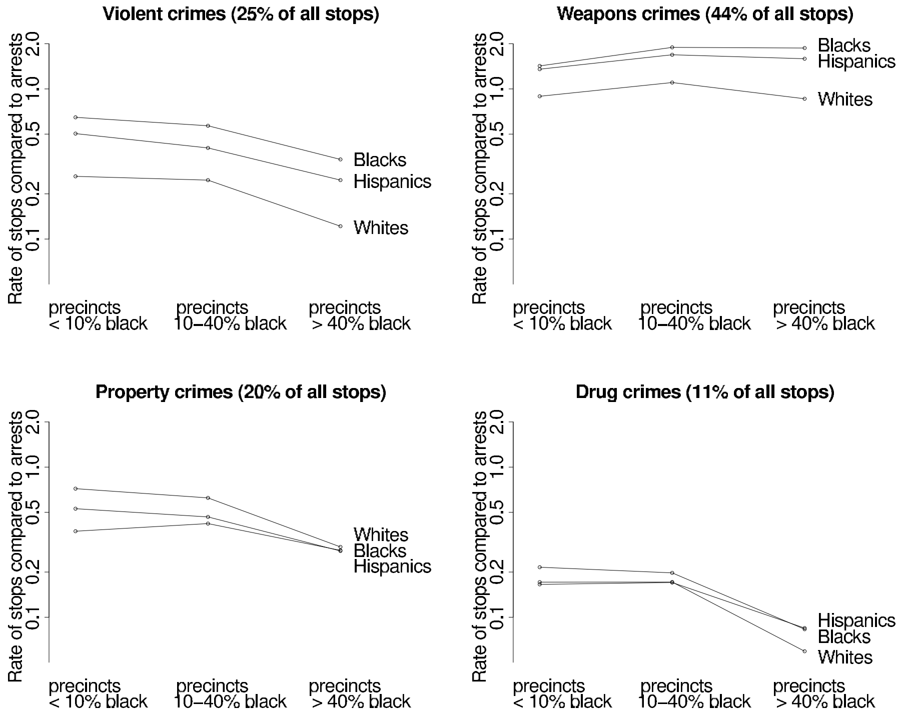

Figure 16.5 *Estimated rates* $\exp(\alpha_e)$ *at which people of different ethnic groups were stopped for different categories of crime, as estimated from hierarchical regressions (16.12) using previous year's arrests as a baseline and controlling for differences between precincts. Separate analyses 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.*

$$\epsilon_{ep} \sim \mathrm{N}(0, \sigma_\epsilon^2), \tag{16.12}$$

where the coefficients α_e's control for ethnic groups, the β_p's adjust for variation among precincts, and the ϵ_{ep}'s allow for overdispersion. Of most interest are the exponentiated coefficients $\exp(\alpha_e)$, which represent relative rates of stops compared to arrests, after controlling for precinct.

By comparing to arrest rates, we can also separately analyze stops associated with different sorts of crimes. We did a separate comparison for each of four types of offenses: violent crimes, weapons offenses, property crimes, and drug crimes. For each, we modeled the number of stops y_{ep} by ethnic group e and precinct p for that crime type, using as a baseline the previous year's DCJS arrest rates n_{ep} for that crime type.

We thus estimated model (16.12) for twelve separate subsets of the data, corresponding to the four crime types and the three categories of precincts. We performed the computations using `Bugs` (a predecessor to the program Stan described in Appendix C). Figure 16.5 displays the estimated rates $\exp(\hat{\alpha}_e)$. For each type of crime, the relative frequencies of stops for the different ethnic groups, are in the same order for each set of precincts. We also performed an analysis including 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.

Figure 16.5 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 are 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.

Does the overall pattern of disproportionate stops of minorities imply that the NYPD was acting in an unfair or racist manner? Not at all. It is reasonable to suppose that effective policing requires many people to be stopped and questioned in order to gather information about any given crime. In the context of some difficult relations between the police and ethnic minority communities in New York City, it is useful to have some quantitative sense of the issues under dispute. Given that there have been complaints about the frequency with which the police have been stopping blacks and hispanics, it is relevant to know that this is indeed a statistical pattern. The police department then has the opportunity to explain its policies to the affected communities.

16.5 Example: hierarchical logistic regression for political opinions

We illustrate the application of the analysis of variance (see Section 15.6) to hierarchical generalized linear models with a model of public opinion.

Dozens of national opinion polls are conducted by media organizations before every election, and it is desirable to estimate opinions at the levels of individual states as well as for the entire country. These polls are generally based on national random-digit dialing with corrections for nonresponse based on demographic factors such as sex, ethnicity, age, and education. We estimated state-level opinions from these polls, while simultaneously correcting for nonresponse, in two steps. For any survey response of interest:

1. We fit a regression model for the individual response y given demographics and state. This model thus estimates an average response θ_j for each cross-classification j of demographics and state. In our example, we have sex (male or female), ethnicity (African American or other), age (4 categories), education (4 categories), and 50 states; thus 3200 categories.
2. From the U.S. Census, we get the adult population N_j for each category j. The estimated population average of the response y in any state s is then $\theta_s = \sum_{j\in s} N_j\theta_j / \sum_{j\in s} N_j$, with each summation over the 64 demographic categories in the state.

We need a large number of categories because (a) we are interested in separating out the responses by state, and (b) nonresponse adjustments force us to include the demographics. As a result, any given survey will have few or no data in many categories. This is not a problem, however, if a multilevel model is fitted. Each factor or set of interactions in the model corresponds to a row in the Anova plot and is automatically given a variance component.

As discussed in the survey sampling literature, this inferential procedure works well and outperforms standard survey estimates when estimating state-level outcomes. For this example, we choose a single outcome—the probability that a respondent prefers the Republican candidate for President—as estimated by a logistic regression model from a set of seven CBS News polls conducted during the week before the 1988 presidential election. We focus here on the first stage of the estimation procedure—the inference for the logistic regression model—and use our Anova tools to display the relative importance of each factor in the model.

We label the survey responses y_i as 1 for supporters of the Republican candidate and 0 for supporters of the Democrat (with undecideds excluded) and model them as independent, with $\Pr(y_i = 1) = \text{logit}^{-1}(X_i\beta)$. The design matrix X is all 0's and 1's with indicators