Homework 4, due 15 Nov

1. Cox and Snell (1980) present data on satisfaction with housing. Variables in the data set include a record number, satisfaction, the amount of influence the tenant has over management of the housing, the type of housing, the amount of contact the tenant has with other residents, and the number of respondents matching this pattern. The data set may be found at at http://stat.rutgers.edu/~kolassa/Data/housing.dat, although the question below is intended to be conceptual and not data analytic, and so you don’t need to use the data.

   a. Suppose that you believe a Poisson regression model, containing main effects for satisfaction, influence, type, and contact, fits the data. Suppose further that you want to learn about the interactions between satisfaction and contact, and furthermore, you believe that contact and type of housing are related (ie, the Poisson regression needs the interaction between these variables to be valid.) Which of the above variables may be collapsed over, and why?

   b. Again consider the model with main effects and interactions between satisfaction and contact, and between contact and residence type. Suppose that no other interactions are necessary to fit these data. What can you deduce about the relationship between satisfaction and residence type, taking contact into account?

2. Investigators plan a study of the effectiveness of seat belt usage in avoiding fatalities in serious traffic accidents. These accidents are single-vehicle accidents severe enough that the insurance company decided to scrap the vehicle. Investigators plan to collect data from outcomes for the drivers in 20 accidents involving each of six vehicle types; these vehicle types vary according to their implicit safety, and are expected to have fatality probabilities of .1, .1, .2, .2, .3, and .3 in such accidents. Furthermore, investigators expect that half of all passengers will use seatbelts. Calculate the power that a one-sided test of size $\alpha = .025$ rejects the null hypothesis of no association between belt use and fatality, when the alternative is that lack of belt use and fatality are associated by a log odds ratio of .2.

3. Caroll, Gail, and Lubin (1993) describe a case–control study in which 39 women with cervical cancer and 76 women without cervical cancer were screened for exposure to Herpes simplex virus. Investigators expect that roughly 30% of the controls will have Herpes simplex virus exposure, and expect under their alternative hypothesis to see an elevated proportion of the cases to have Herpes simplex virus exposure, with an odds ratio of 1.4.

   a. Calculate the power for detecting an effect this large for a one-sided test of size .025.

   b. Suppose that the investigators wanted 80% power under the above alternative, and planned to use the same number of cases and controls. How many women would be necessary in each group?
c. Suppose that the investigators wanted 80% power under the above alternative, and planned to use ten times as many controls as cases. How many women would be necessary in each group? Compare your answer to part (b).