Statistical Models for Multi-level Data
2004 Summer Epi-Biostat Institute

Participant Homework

Module I: Statistical Background on Multi-level Models

1. The data from the "Alcohol Dependence" crossover-trial example given in class actually pertains to a 2x2 crossover trial of Cerebrovascular disease examined in the text: *Analysis of Longitudinal Data,* (Diggle, Heagerty, Liang & Zeger, 2002. pg. 148-150 & pg. 180-181) and reproduced below.

Data from a 2 x 2 crossover trial on cerebrovscular deficiency adapted from Jones and Kenward (1989, p. 90), where treatments A and B are active drug and placebo, respectively; the outcome indicates whether an electrocardiogram was judged abnormal (0) or normal (1).

	Responses					Period	
Group	(1,1)	(0,1)	(1,0)	(0,0)	Total	1	2
AB	22	0	6	6	34	28	22
BA	18	4	2	9	33	20	22

In SAS's program editor, open and run the program: `"Problem 1 xover trial.sas"` found on the class website to reproduce the results in DHLZ (2002). Interpret the regression coefficients and their confidence intervals for the period and treatment effects for both the marginal (GEE) and conditional (Random Effect) analyses. (Note: the results are the same as in the Module 1 class notes so you can use them and look over the program later if you're short on time.)

2. In slides 50 and 53 of the module 1 presentation, results of a conditional and marginal logistic multilevel model are presented. Focus on Model 1 in the tables (the models with the simplest association structures). Interpret the regression coefficients for gender and for %protestant in each of these two models.

3. Download the winbugs software on your laptop or visit the School computer lab to visit the BUGS website and watch the movie demo:

http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml
http://www.statslab.cam.ac.uk/~krice/winbugsthemovie.html

Module II: A two-stage model example: The DZAPS study

Below find a table of maximum likelihood estimates of the log relative risk (percent increase per 10 micrograms per cubic meter) and their statistical standard errors for 6 cities from the hypothetical DZAPS study (note: the data are not real).

Calculations for Empirical Bayes Estimates

City	Log RR $\hat{\beta}_c$	Statistical std error (vc)	Total Var (TVc)	1/TVc	wc	θ_c	RR.EB $\tilde{\beta}_c$	se RR.EB
LA	**0.30**	**0.10**						
NYC	**0.50**	**0.12**						
Chi	**0.4**	**0.15**						
Dal	**0.0**	**0.30**						
Hou	**1.0**	**0.40**						
SD	**-0.10**	**0.50**						
Over-all								

1. Use the estimates above and their standard errors to estimate the natural variance in the true log relative risks across these 6 cities. Follow the calculations made in the lecture for module 2.

2. Calculate the overall estimate of the log relative risk weighting the individual city estimates by the inverse of their total variances

3. .Calculate the standard error for the overall estimate and make a 95% confidence interval for the true population mean.

4. Now complete the table above producing the empirical-Bayes estimate and standard error for each city

5. Compare the empirical-Bayes and maximum likelihood estimates for San Diego (SD). Which estimate do you prefer and why? Comment on whether you think air pollution saves lives?

6. Fit the two-stage normal-normal model below in Winbugs to re-analyze the NMMAPS 6 cities data using MCMC.

```
model
        {

        for( i in 1 : N)
        {
        p.hat[i] <-(1/se[i])*(1/se[i])
        beta.hat[i] ~dnorm(b[i],p.hat[i])
                        b[i] <-alpha + u[i]
                        u[i] ~dnorm(0,tau)
                }


        tau ~dgamma(0.001,0.001)
        sigma <- 1 / sqrt(tau)
        alpha ~ dnorm(0.0,1.0E-6)

}

Inits <- list(alpha =0,tau=1) <-
Data <- list(N=6,beta.hat  = c(0.3,0.5,0.4,0.0,1.0,-0.1), se =c(0.10,0.12,0.15,0.30,0.40,0.50)) <-
```

Module III: Applications of Multilevel Models to Profiling of Health Care Providers

1. From the Winbugs help menu, copy the "Institutional ranking" example (look under the help menu under Vol I examples)
2. Reproduce the example discussed in class on "Institutional ranking"
3. Discuss whether a marginal model may or may not be appropriate for the analysis of the institutional ranking data.
4. Write an abstract for a scientific journal that summarizes the results of the "Institutional ranking" example. Report statistical uncertainty associated with ranking.

5. From the case study by Normand and et al JASA 1997, which are the three most “aberrant hospitals”? How is the uncertainty in ranking reported?

Module IV: Applications of Multilevel Models to Spatial Epidemiology

1. From the Winbugs help menu, copy the “Scottish Lip Cancer” example (look under the map menu under examples)
2. Reproduce the statistical analyses performed in class
3. For the area with the largest observed SMR, estimate the posterior probability of having the largest relative risk of lip cancer
4. For the area with the lowest observed SMR, estimate the posterior probability of having the largest relative risk of lip cancer . Why do you think the rank of the observed SMR and the rank of the smoothed SMR might be different?