

Stat 771 Homework 5, due Monday, April 25

We will consider data from a longitudinal study at Harvard of effects of air pollution on respiratory illness in children. The children were examined annually at ages 7 through 10 and classified according to the presence of absence of respiratory illness. The only predictor is whether a kid's mom smoked at the start of the study: $s_i = 1$ for smoking regularly and $s_i = 0$ otherwise.

1. Marginal model Fit the following marginal model in PROC GENMOD:

$$\text{logit } \pi_{ij} = \beta_0 + \beta_1 s_i + \beta_2 t_j + \beta_3 t_j^2 + \beta_4 s_i t_j.$$

Here,

$$\mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4} \end{bmatrix} \quad \text{and } \mathbf{t} = \begin{bmatrix} 7 \\ 8 \\ 9 \\ 10 \end{bmatrix}.$$

Try AR(1), compound symmetry, and unstructured for the correlation matrices; choose one via QIC. If you can drop the smoking by time interaction, do so. Interpret your final model.

2. Conditional random coefficient models Now let's assume each child has his/her own probability trajectory that's linear on the log-odds scale:

$$\text{logit } \pi_{ij} = \beta_0 + \beta_1 s_i + \beta_2 t_j + b_{i0} + b_{i1} t_j, \quad \mathbf{b}_i \stackrel{iid}{\sim} N_2(\mathbf{0}, \boldsymbol{\Sigma}).$$

Fit this model in PROC GLIMMIX; use `method=laplace`. Interpret the model.

3. Assume a quadratic mean effect, but only child-specific random intercepts:

$$\text{logit } \pi_{ij} = \beta_0 + \beta_1 s_i + \beta_2 t_j + \beta_3 t_j^2 + b_{i0} \quad b_{i0} \stackrel{iid}{\sim} N(0, \sigma^2).$$

Fit this model in PROC GLIMMIX; use `method=laplace`. Interpret the model. Which model (in 2 or 3) has lower AIC?