This final is assigned May 6th, 2009.

Your solution write-up and appendix is due on May 13th, 2009 at 5PM sharp.

To turn in exams:

A hard copy of the write-up and appendix must be placed in either Haley Hedlin's, Bruce Swihart's or Howard Chang’s departmental mailbox on the 3rd floor outside of the Biostatistics Departmental office.

Late submissions are not allowed.
Electronic (email) submissions are not allowed.

Final Rules:

- This exam is to be considered a take-home test. Please do not collaborate or consult with others.
- You may use any reading material (class notes, books, etc) you wish.
- Please keep your write-ups short and do not include reams of analyses output. Please attach an appendix of the (well documented) code you used.
- For the final exam, the BBS will be used only for clarification of the exam questions, wording or expectations.
- You may also carry out additional analyses that you feel are appropriate.

Final: Demographics and contraceptive use in Bangladesh.

Goal: Quantify how factors such as urban vs. rural residence, age, and number of children affect the use of contraceptives in a developing nation.


Variables (i – indexes district, j – indexes woman)

**use** \((y_{ij})\): use of contraception (1: using contraception; 0: otherwise)

**district** \((d_{ij})\): district in which the subject lives

**age** \((a_{ij})\): age (in years) at time of survey (centered at the sample mean age)

**urban** \((r_{ij})\): residence location (1: urban; 0: rural)

**nchild**: number of (living) children at time of survey (1: none; 2: one, 3: two; 4: three or more) (let \(c_{1ij}, c_{2ij}, c_{3ij}, c_{4ij}\) be indicator variables for each nchild category)

1 Exploratory Data Analysis

a) How many women participated in the survey? How many districts are included in the fertility survey? What are the maximum and the minimum number of women participating in the survey in a district?

b) Overall, what is the proportion of contraceptive use by women in the survey?

c) What is the proportion of contraceptive use by women in urban areas? What is the
proportion in rural areas? How many of the women surveyed are in urban areas? How many are in rural areas?

d) Stratify the women by parity (number of children) and report the proportion of contraceptive use in each stratum. Of the four nchild categories, which is the most common among this group of women?

e) Create a binary variable ageb (0: age ≤0 ; 1: age > 0). Calculate the proportion of women who use contraceptives, stratifying by ageb.

2 Ignore clustering by district

a) Run and report your results from the following model for the binary outcome \( y_{ij} \), ignoring potential clustering by district of women’s contraceptive use:

\[
\text{logit}(P(y_{ij} = 1)) = \beta_0 + \beta_1 r_{ij} + \beta_2 a_{ij} + \beta_3 c_{2ij} + \beta_4 c_{3ij} + \beta_5 c_{4ij}
\]

b) Report and interpret your estimate of \( \beta_1 \), translating your result to a more interpretable scale (include a 95% confidence interval on this interpretable scale).

c) Without using a Stata command that specifies the clustering variable, run a model that acknowledges the fact that the independence assumption is probably violated. (Hint: you can do this by adding a single option to a Stata command that you could have used for part (a). This command calculates a part of the regression results differently to make up for the violation in the independence assumption.) Write a sentence comparing the results from part (c) to those in part (a). Explain very briefly why the two results are different.

3 Account for clustering by district

a) Use gllamm to fit a model analogous to that in 2(a), but that accounts for clustering by district. Make sure you are satisfied with the parameter estimates (re-run gllamm changing the necessary options until you arrive at adequately accurate results). Report your results and comment on any differences from the results in 2(c). Which set of results do you prefer and why?

b) Write down the model that you fit in 3(a) using the same mathematical notation from 2(a) where appropriate (i.e., use \( \beta_1 \) as the coefficient for \( r_{ij} \), even though this may have a different value than the \( \beta_1 \) from the model in 2(a)). Call the variance of the random intercept \( \psi \).

c) Report and interpret your estimate of \( \beta_1 \), translating your result to a more interpretable scale (include a 95% confidence interval on this scale). Report and interpret \( \psi \).

4 Random slope on age

a) Write down a model that builds on the model in 3 (c) but includes a random slope on age as an additional random effect.

b) Fit this model using gllamm, report the results and comment on whether you believe the random slope on age is needed in our model.

c) Interpret the correlation between these two random effects.
5 Random slope on children
Idea: Replace the `nchild` categorical variable in our model with the binary `children` variable, so we fit a total of 2 random effects instead of 4 random effects.

a) Create a binary variable `children (m_{ij})` that is 0 if the woman has no children, and 1 if the woman has 1 or more children.

b) Write down a model that builds on

\[
\logit(P(y_{ij} = 1)) = \beta_0 + \beta_1 r_{ij} + \beta_2 a_{ij} + \beta_3 m_{ij}
\]

by including a random intercept for district and random slope on the indicator of having children.

c) Fit this model using `gllamm`, report the results and comment on whether you believe this random slope on `children` is needed in our model.

6 Final steps
a) Choose a final model (include at least one random effect). Explain why you made your choice.

b) Use your model to predict the probability that each woman uses contraception. *(Hint: use empirical Bayes methods to predict the random effect(s).)*

c) Create three 2x2 tables:

**Table 1**: Report the mean predicted probability of using contraception from your final model in a table that stratifies by urban vs. rural and no children vs. children.

**Table 2**: Report the mean predicted probability of using contraception from the model in 2(a) that ignored clustering in a table that stratifies by urban vs. rural and no children vs. children.

**Table 3**: Report the observed proportion of contraception usage in a table that stratifies by urban vs. rural and no children vs. children.

Comment on the three tables.

d) **Summarize**: Write a 2-3 paragraph summary of the final analysis results, including a few introductory sentences summarizing your exploratory data analysis.