Longitudinal Data Analysis (BIOS 767)
Final Exam, April 29, 2019

Name:..
I pledge to observe the UNC Honor Code in full. I didn't receive or give help from or to anyone on this exam. Signature:....................................(4/29/2019)
Don't discuss these problems with any person other than the course instructor.

1. Low birth weight is an important public health problem because it is associated with perinatal mortality and morbidity, stunted growth and chronic diseases in adult life. Hence it is important to identify important risk factors. A study was conducted to investigate the risk of low birth weight and how it relates to birth order and to mother's age. Data were collected from hospital records on all women with exactly five children (on the date the study started) in a certain geographic area in the United States. There were $K = 198$ women who qualified for the study and the data are available in files `babies1.dat` and `babies2.dat`. The two files contain the same data, but the first is formatted with five records per subject, while the second has one record per subject.

   Let the random variable Y_{ij} represent the birth weight in grams of the jth child for the ith woman, and let age_{ij} denote her age in years (truncated to an integer value) when she delivered her jth child. Define the vector Y_i to be $Y_i := (Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4}, Y_{i5})^\top$. In what follows, when fitting models for Y_i we will assume that Y_i is distributed as multivariate normal and specify various models for its mean and covariance. Subjects are assumed mutually independent throughout.

   Two common cutpoints used to define "low birth weight" are 2500 grams and 3000 grams. In this problem we will adopt the latter; a birth weight under 3000 grams will be considered a low birth weight. Hence, define Y_{ij}^* as follows: $Y_{ij}^* = 1$ if $Y_{ij} < 3000$ and $Y_{ij}^* = 0$ if $Y_{ij} \geq 3000$. Define the vector Y_i^* to be $Y_i^* := (Y_{i1}^*, \ldots, Y_{i5}^*)^\top$.

   In what follows, for hypothesis testing, use likelihood ratio tests whenever possible/feasible. Otherwise, use Wald-type tests. In either case, you should describe your methods, present test statistics, degrees of freedom and p-values. Use common statistical language with no reference to any computer code, statements or options. Do not assume that the reader knows anything about the software you used, and do not submit any computer code.

   (a) Present one or two useful summaries, numerical and/or graphical.

   (b) Model M_1: Fit the model

   $$\mathrm{E}[Y_{ij}] = \alpha_j + \beta_j(age_{ij} - 20),$$

   $i = 1, \ldots, K, j = 1, \ldots, 5$, assuming that $\mathrm{cov}(Y_i)$ is the same for all subjects, but otherwise has no special structure. Fit this model, present parameter and standard error estimates and comment on them. Interpret the parameters that appear in the mean structure.

   (c) In the context of M_1, test the hypothesis $H_0 : \beta_1 = \beta_2 = \ldots = \beta_5$ against its complement. Report and interpret your findings.

(d) In the context of M_1, test the hypothesis $H_0 : \beta_j = 0, j = 1, \ldots, 5$ against its complement. Report and interpret your findings.

(e) In the context of M_1 and assuming that $\beta_1 = \ldots = \beta_5$, test the hypothesis $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$. Report and intepret your findings.

(f) In the context of M_1 and assuming that $\beta_1 = \ldots = \beta_5$, test the hypothesis $H_0 : \alpha_1 = ... = \alpha_5$ against its complement. Report and interpret your findings.

(g) Model M_2: Fit the model

$$\mathrm{E}[Y_{ij}] = \alpha + \beta(age_{ij} - 20),$$

$i = 1, \ldots, K, j = 1, \ldots, 5$, assuming that $\mathrm{cov}(Y_i)$ is the same for all subjects, but otherwise has no special structure. Fit this model, then calculate estimates of and 95% confidence intervals for the probability of a low-birth weigth for the first birth if a woman's age at that birth was 15, 20, 25, 30, 35 or 40 years. Describe your methods and present the results in a table. Comment.

(h) Fit the model M_3:

$$\Phi^{-1}(\mathrm{E}[Y^*_{ij}]) = \theta + \psi(age_{ij} - 20),$$

$i = 1, \ldots, K, j = 1, \ldots, 5$, assuming that $\mathrm{cov}(Y^*_i)$ is the same for all subjects, but otherwise has no special structure. Fit this model, then explore possible theoretical connections between the mean structures in models M_2 and M_3, and how far these connections are borne out in the analysis results (the estimates).

(i) Based on the estimates from M_3, calculate and present estimates and confidence intervals as in part (g).

(j) Fit the model M_4:

$$\Phi^{-1}(\mathrm{E}[Y^*_{ij}|U_i]) = \gamma + \delta(age_{ij} - 20) + U_i,$$

$i = 1, \ldots, K, j = 1, \ldots, 5$, where $U_1, \ldots, U_K$ are unobserved random variables distributed as iid normal with mean 0 and variance σ^2_u. Further, assume that $Y^*_{i1}, \ldots, Y^*_{i5}$ are conditionally independent given U_i. Explore possible theoretical connections between the mean structures in models M_3 and M_4, and how far these connections are borne out in the analysis results.

(k) Based on the estimates from M_4, present a "between and within" decomposition of the variance of Y^*_{i3} assuming $age_{i3} = 25$, and compute the intra-class correlation based on this decomposition.

(l) Write one or two paragraphs discussing the study design and your main findings for a medical journal.

Points: a 8, b 8, c-f 4 each, g 16, h 12, i 8, j 12, k 8, l 12.