MS WRITTEN EXAMINATION IN BIOSTATISTICS, PART II

Thursday, August 3, 2017, 9am-3pm.

1. The gestational age (GA) of an embryo or newborn infant is approximately the time since the mother's last menstrual period. GA is used to determine whether the pregnancy has reached full term (40 weeks). An infant born prior to 37 weeks GA is regarded as being premature or preterm. There had been some epidemiologic evidence for an association between a pregnant woman having periodontal disease and giving birth prematurely or for the infant's birthweight to be below normal.

   The MOTOR Study was a randomized clinical trial to investigate whether treating periodontal disease in pregnant women reduces the risk of preterm delivery and/or increases average birthweight. About 1,800 pregnant women with periodontal disease were randomized into two treatment groups. The "prenatal treatment" group received periodontal therapy early in pregnancy. The "post-partum treatment" group received periodontal therapy a few weeks after delivery. (The reason for giving the post-partum treatment group any periodontal therapy was so that the women would receive the same benefit in terms of their own periodontal health as those in the prenatal treatment group. By providing the treatment after delivery it could not affect birth outcomes in this group.)

   In MOTOR, two measures of GA at birth were available. The more accurate one was made using an ultrasound examination early in the pregnancy. This is given in weeks, calculated from days. So, for instance, 38.1429 corresponds to a GA of 38 weeks and 1 day. The other measure of GA was an estimate made when the infant was born and is in whole weeks.

   Various periodontal measurements were made around each tooth at baseline (early in the pregnancy, before randomization) and were repeated shortly after giving birth. For each woman, the measurements on the teeth were averaged to give a summary score for each type of measurement. One such measurement is probing depth, in millimeters, with larger values indicating more periodontal disease.

   The file `motor.dat` contains data from a subset of the live births in the trial. The columns in the file are, respectively, ID (participant identifier), group (treatment group; 1 = prenatal, 2 = post-partum), GA_ultra (GA estimated by ultrasound), GA_est (GA estimated at birth), bweight (birthweight, in grams), sex (of the infant; 1 = male, 2 = female), ppnum (number of previous pregnancies), PD_pre (average pocket depth at the time of randomization), PD_post (average pocket depth after delivery). Note that the data file was printed from SAS, so a missing value for a numeric variable appears as a period (".").

   If you detect any obvious data errors, make a note of the errors and set the corresponding values to missing.

Because of the limited time available, state any assumptions you need to make for any statistical tests, but you do not need to check these assumptions.

Submit your computer code for this problem and **only the relevant parts** of the computer output.

(a) Determine the mean number of previous pregnancies per woman. (Report the value rounded to 2 decimal places.)

(b) In other studies the distribution of the number of previous pregnancies is approximated reasonably well by the Poisson distribution. Assuming a Poisson distribution having the mean calculated in part (a), how many women in this dataset would one expect to have had at least 4 previous pregnancies? Test whether the observed number differs significantly from the expected number.

(c) Test whether the normal distribution is a good model for the distribution of the ultrasound version of GA.

(d) Classify both versions of gestational age into 3 intervals, $(0, 37)$, $[37, 40)$, and $[40, \infty)$. Determine how well the two versions agree and provide a 95% confidence interval for the true agreement.

(e) Does the risk of preterm delivery vary monotonically with the number of previous pregnancies?

(f) Based on this study, is treating periodontal disease in pregnant women effective in terms of reducing the risk of prematurity?

The effect of the periodontal therapy on mean birthweight was smaller than the investigators had expected – there was not a statistically significant difference between the mean birthweights in the two treatment groups. One potential explanation for the lack of effect is that the periodontal therapy provided may not have been intensive enough to yield a substantial and sustained reduction in the amount of periodontal disease.

(g) For the prenatal treatment group, did the mean average pocket depth change from baseline to after delivery?

(h) Did the mean change in average pocket depth differ between the two treatment groups?

(i) Based on the data on average pocket depth, discuss the effectiveness of the periodontal therapy and the consequences for the potential to affect birthweight.

Points: (a) 2, (b) 4, (c) 2, (d) 3, (e) 3, (f) 3, (g) 3, (h) 3, (i) 2.

2. Treatment selection in breast cancer has traditionally been based upon a patient's clinical tumor subtype. This subtype is usually determined through a pathology review of a sample from the patient's tumor after diagnosis. Based upon historical clinical trial data, an "optimal" treatment is suggested for patients belonging to a particular subtype. This treatment is assumed to have the best response among the set of treatments available to a patient, however for various reasons not all physicians choose this treatment.

   Among patients of the same subtype, significant variation in tumor response to optimal treatment is usually observed. It is hypothesized that an alternate subtyping system, molecular subtyping, may provide better criteria for treatment selection. Based upon molecular subtype, a different set of treatments may be suggested to be optimal for an individual. This approach (PAM50) measures the expression of 50 genes in a patient's tumor sample. The expression levels are then used to assign patients to one of four molecular subtypes.

   To address the question of whether clinical versus molecular subtyping should be used to direct therapy, researchers at the North Carolina Cancer Hospital have reviewed data pertaining to 300 breast cancer patients treated in the past year at UNC. The data are in file `subtype.dat`, and the variables are as follows:

   - `id`: An assigned id number, from 1 to 300.
   - `clin_sub`: Clinical subtype; HR+/HER2+ (1) or HR-/HER2+ (2)
   - `treatment`: Treatment given; Hormone Therapy (0) or Chemotherapy (1)
   - `molecular_sub`: Molecular subtype from PAM50; Luminal A (1), Luminal B (2), Basal (3) or Claudin-low (4)
   - `response`: Tumor response; the percentage of baseline tumor volume remaining 30 days after treatment, $100 \times \frac{V_{30}}{V_b}$, where V_b is the baseline (pre-treatment) tumor volume and V_{30} is the tumor volume 30 days post-treatment. Of course, lower values indicate better response.

   (a) The researchers would like to first verify whether the mean response in patients **receiving chemotherapy** differs across clinical subtypes. Please use cell means coding for this section.

      i. Write a linear model in matrix notation (using cell means coding) to evaluate this question. Specify the dimension of each component in your model

      ii. Fit this linear model and interpret the coefficients. Then, evaluate whether mean response in patients receiving chemotherapy differs across clinical subtypes.

      iii. The researchers hypothesize that the variance in chemotherapy response may be larger in the HR+/HER2+ group compared to the HR-/HER2+ group. Evaluate whether

the variance in chemotherapy response is homogeneous across clinical subtypes, reporting the test, test statistic, df, p-value for the test, and your interpretation of the result. What additional steps should be taken given your test result, if any?

iv. Traditionally, the mean response to chemotherapy among HR+/HER2+ patients has been shown to be twice that of HR-/HER2+ patients. Evaluate whether this relation holds in this study. Write the H_0, H_A, C, β and θ_0 for carrying out this test in the GLH testing framework. Carry out this test and report the test statistic, its df and p-value. Does this result agree with or contradict previous results?

v. Calculate the estimated covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$. Explain how one may arrive at this result without access to computer software.

(b) Next, the researchers hypothesize that the relative effectiveness of chemotherapy versus hormone therapy may vary across subtypes. That is, the difference in mean response between patients receiving chemotherapy and those receiving hormone therapy may not be the same across subtypes.

i. With respect to clinical subtypes, evaluate this claim by first choosing an appropriate model. Then write out the H_0, H_A, C, β and θ_0 for carrying out this test in the GLH testing framework with your chosen model. Carry out this test and interpret the result. Based on the fitted model, suggest which therapy may be better for each clinical subtype and why.

ii. Next, the researchers would like to evaluate whether molecular subtype is better at explaining response to treatment than clinical subtype. Fit the same model as the prior sub-part, except this time using molecular subtype instead of clinical subtype. Report the R^2 from each model. Can we use R^2 to help choose between the two models? Why or why not? If no, propose an alternative measure and use it to help suggest which model may be more appropriate.

iii. Researchers would like to directly test whether the fit of the previous model using clinical subtype is much better than the fit of the model using molecular subtype. Carry out this test. If you cannot carry out this test, explain why.

Points: (a) 12, (b) 13.

3. A study was conducted to investigate the risk of low borth weight (LBW) in first-born and second-born children in both normal and hypertensive women.

   The data were from a random sample of 100 non-hypertensive women and 100 hypertensive women who had two or more children. Low birth weight was defined as a binary outcome; 0 if birth weight was at least 3000 grams, 1 if the birth weight was below 3000 grams. The data, obtained from hospital records, are presented at the end.

   For all hypothesis tests: state your methods, show your computations, give the test statistic, degrees of freedom (if applicable) and p-value (either the actual value or a range, e.g. $0.6 < p < 0.7$).

   For full credit, show *all computations*, not just the final answer. This problem is to be done using **only a calculator**. A computer can be used but only as a simple calculator. Specifically, use of any SAS procedures or the analogus R functions is not allowed.

   (a) Test the null hypothesis that, in non-hypertensive women, low birth weight of the first child is not associated with low birth weight of the second child. Apply the same test to the hypertensive group.

   (b) Suggest a measure of association between low birth weight of the first child and low birth weight of the second child. Compute a point estimate and a 95% confidence interval for this measure, separately for each group of women.

   (c) Test the null hypothesis that, in non-hypertensive women, the risk of low birth weight for the first born child is the same as the risk of low birth weight for the second born child. Apply the same test to the hypertensive group.

   (d) Test the null hypothesis that the *joint distribution* of the first and second low birth weight outcomes is the same in the two groups of women.

   (e) The data from *only* the first child (200 births) were used to fit a logistic regression model (using maximum-likelihood) with low birth weight as response and an indicator of hypertension ($x = 0$ if non-hypertensive, $x = 1$ if hypertensive) as a covariate,

   $$\text{logit}P(\text{low birth weight}) = \beta_1 + \beta_2 x.$$

   To the extent possible, present a table of parameter and standard error estimates of β_1 and β_2. (Remember: only a calculator is allowed).

   (f) Consider the following model:

   $$\text{logit}P(\text{low birth weight}) = \gamma_1 - \gamma_2 x + \gamma_3 z - \gamma_4 z * x,$$

   where z is an indicator for the first child ($z = 1$ for the first child, $z = 0$ for the second child). Interpret γ_3 and γ_4. (Notice the negative signs in the model equation.)

(g) An investigator suggested using all the data (400 births) in PROC LOGISTIC in SAS to fit the above model to obtain estimates, confidence intervals, p-values, etc. Comment on this suggestion.

DATA:

The non-hypertensive group:

	Second born not LBW	Second born LBW
First born not LBW	60	10
First born LBW	20	10

The hypertensive group:

	Second born not LBW	Second born LBW
First born not LBW	48	12
First born LBW	22	18

Points: Equal weight, 25 points total.