Biostatistics 140.656, 2018-19
QUIZ 1

Quiz Guidelines:

Please read the following quiz guidelines carefully:

- For this quiz, you are to work ALONE. You may use your course notes and lab materials to help answer the questions.
- Submit your answers to Courseplus by 5pm Friday April 5~~th~~.
- DO NOT discuss this quiz or your solution to this quiz with other students from the course on Wednesday ~~April 3rd~~ through Friday ~~April 5th~~. The solution to the quiz will be available Saturday, ~~April 6th~~.
- By submitting your answers to Courseplus, you are acknowledging that you have read the guidelines carefully and will adhere to these guidelines.

Please check course plus for due dates.

Scientific Background:

Consider a two-stage normal-normal model defined for the High School and Beyond data as follows:

$$y_{ij} = \beta_{oi} + \varepsilon_{ij} = \beta_0 + b_i + \varepsilon_{ij}$$

$$\beta_{oi} \sim N(\beta_0, \tau^2), b_i \sim N(0, \tau^2), \varepsilon_{ij} \sim N(0, \sigma^2)$$

where i denotes the schools, i = 1, …, 156, and j denotes the students within schools, j = 1, …, n_i, and the measure y is the mathematics achievement score.

The model results are below.

```
mixed mathach || newid: , stddev

Mixed-effects ML regression                     Number of obs     =      7,042
Group variable: newid                           Number of groups  =        156

                                                Obs per group:
                                                              min =         14
                                                              avg =       45.1
                                                              max =         67

                                                Wald chi2(0)      =          .
Log likelihood = -23094.799                     Prob > chi2       =          .
------------------------------------------------------------------------------
     mathach |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   12.76768   .2396075    53.29   0.000     12.29806     13.2373
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
newid: Identity              |
                   sd(_cons) |   2.831786    .180498      2.499222    3.208603
-----------------------------+------------------------------------------------
                sd(Residual) |    6.26665   .0534091       6.16284    6.372209
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 889.26        Prob >= chibar2 = 0.0000
```

1. In the above model, which parameter describes the **variation** in observed student's MA scores relative to the school's mean MA score?
    a. β_{oi}
    b. b_i
    c. ε_{ij}
    d. τ^2
    e. σ^2

2. Based on the results above, you would expect that roughly 95% of schools will have mean MA scores ranging from roughly
    a. 12.30 to 13.24
    b. 7.22 to 18.32
    c. 2.50 to 3.21
    d. 6.16 to 6.37

3. Based on the results above, the 95% confidence interval for the population mean MA score is
    a. 12.30 to 13.24
    b. 7.22 to 18.32
    c. 2.50 to 3.21
    d. 6.16 to 6.37

4. What is the estimated intra-class correlation (ICC) from the model fit?
    a. 2.83 / (6.27+2.83)
    b. 6.27 / (6.27+2.83)
    c. $2.83^2 / (6.27^2+2.83^2)$
    d. $6.27^2 / (6.27^2+2.83^2)$

5. Which phrase is a correct interpretation of the intra-class correlation (ICC)?
    a. The proportion of total variation in MA scores that is attributable to differences in MA scores within schools
    b. The proportion of total variation in MA scores that is attributable to differences in the sample average MA scores between schools
    c. The proportion of total variation in MA scores that is attributable to differences in the mean MA scores between schools
    d. The proportion of total variation in MA scores that is attributable to measurement error (i.e. natural fluctuations in MA scores)

6. What parameter in the two-stage normal normal model represents the difference between a schools' mean MA score and the population mean MA score?
    a. β_{oi}
    b. b_i
    c. ε_{ij}
    d. τ^2
    e. σ^2

7. Suppose there are two schools in the data, call them A and B. Both schools have a sample mean MA score of 14.5 and the estimated value of b_i is greater than 0 for both schools; call these values b_A and b_B, respectively. If $b_A > b_B$, then
    a. $n_A > n_B$
    b. $n_A < n_B$
    c. $n_A = n_B$
    d. we don't know anything about the relationship between n_A and n_B

8. Suppose we recruited another 50 schools for the study and refit the model. The estimated value of τ^2 increased where the estimated values of β_0 and σ^2 remained the same. In this larger study (the original 156 schools plus 50 additional schools), the estimates of the school mean MA score for the original 156 schools would be
    a. "shrunk" closer to the population mean MA score compared to the estimates based on the model fit to only the original sample of 156 schools.
    b. closer to the school sample mean MA score compared to the estimates based on the model fit to only the original sample of 156 schools.
    c. the same as the estimates based on the model fit to only the original sample of 156 schools.
    d. it is not possible to tell what will happen to the estimates of the school mean MA score for the original 156 schools.

9. The two-stage normal normal model assumes that $Corr(b_i, \varepsilon_{ij}) = 0$ for all i and j. In fact, you can think of the two-stage normal normal model as

$$y_{ij} = \beta_{oi} + \varepsilon_{ij} = \beta_0 + b_i + \varepsilon_{ij} = \beta_0 + {\varepsilon^*}_{ij},$$

where ${\varepsilon^*}_{ij}$ is the total residual and the two-stage normal normal model decomposes this total residual into two independent parts; one corresponding to differences between the school means with respect to the population mean (i.e. b_i) and one corresponding to differences between MA scores within a school with respect to the given schools mean (i.e. ε_{ij}).

If we wanted to add variables to the model to explain some of the variation we see between student's MA scores from the same school, we could consider adding:

    a. student gender
    b. school-average SES
    c. proportion of students in the school on the academic track
    d. none of the above