

**INFOSYS 750, Assignment 2**

**Before you start:**

1. Your assignment should be in the form of a report that responds to the parts of this assignment.

2. Sections from your R output should be embedded in appropriate places in your report. PLEASE NOTE THAT INCLUDING NON-RELEVANT OUTPUT MIGHT BE PENALIZED.

3. Please mention the additional R packages you are using for producing the outputs of analysis.

4. You will need different datasets for completing this assignment provided in Canvas:

a. gpa.csv.csv

b. mathscore.csv

# Question 1

Use Data Set "***gpa.csv***", and answer the following questions

|  |  |
| --- | --- |
| **GPA** | GPA measures on 6 consecutive occasions |
| **Occas** | Variable representing time, 6consecutive occasions when students’ GPA was recorded |
| **Sex** | Dummy for gender (1 male, 0 female) |
| **Highgpa** | Students’ GPA in high school |

Employ a Multilevel longitudinal data analysis approach to answer the following questions:

1. Visualize the growth plots for the first ten students. Does GPA increase over time? Describe the change in GPA over time based on growth plots.

2. Generate the unconditional means and unconditional growth models and write down the corresponding equations? Discuss the fixed and random effects of the model predict the rate of change in students' GPA scores over time.

3. Examine if gender affects the intercept (initial GPA score) and slope (change over time). Interpret the estimates of fixed effects, variance components, and Pseudo R2 statistics.

4. How do GPA in high school and gender affect changes in GPA over time? Examine the effect on initial status and rate of change.

5. Discuss the best model to explain the change in GPAs over time.

# Question 2

Please use the dataset: "***mathscore.csv"***, using the variables listed below to fit your model to the data and answer the following questions:

|  |  |
| --- | --- |
| **test** | The dependent variable representing each |
| students' individual scores on the math tests over time | |
| **time** | Variable representing three linear occasions in time |
|  | measuring students' math achievement. |
| **effective** | Two-category predictor variable representing |
|  | teachers' effectiveness in teaching math. |
| **ses** | Predictor interval variable (z score) measuring student |
|  | socioeconomic status composition within the schools. |

Employ a Multilevel longitudinal data analysis approach to answer the following questions:

1. Using visualization techniques, compare the trend of changes in test results across two categories of the effective variable over time. Discuss your findings of the observed pattern.
2. What is your "intercept only" model? Interpret the fixed effect; variance components (within-person variance and between-person variance); and ICC (Intra-class Correlation Coefficient)?
3. Discuss the model to predict students' math scores from the intercept and time. Interpret the rate of change?
4. Whether the intercept (initial math score) and slope (change over time) affected by effective teachers? Interpret the estimates of fixed effects, variance components?
5. How do socioeconomic status (characteristics – level 2) and effective teachers (characteristics – level 2) influence students' math scores?
6. Whether the intercept and slope (for time) are affected by an effective teacher and the student's socioeconomic status (at the start of the study)? Interpret the estimates of fixed effects; variance components?
7. Is there a change in math achievement over time? What model is the most appropriate to use at the end?