Homework 3
Biostatistics 755
Due February 21st, 2025

1. The dataset `MITgrowth.csv` on the course website data are from a prospective study on body fat accretion in a cohort of 162 girls from the MIT Growth and Development Study. The study was designed to look at changes in percent body fat in girls before and after menarche. All subjects had to be pre-menarche and non-obese to enter the study. Observations were taken annually until 4 years after menarche. At each observation percent body fat was measured.

   Two time-scales are included: age, and time since menarche (which can be negative). Time since menarche is the more biologically relevant time scale to use. The variables (in order) are: *Subject ID, Current Age (years), Age at Menarche (years), time relative to Menarche (years), Percent Body Fat.*

   (a) **(10 points)** Produce a spaghetti plot with the overall mean of the outcome using (i) age and (ii) time relative to menarche as the x-axis.

   (b) **(10 points)** Which time scale from (b) appears to have a stronger relationship with percent body fat? Which time scale is best to answer the study question? Comment on the possible parametric methods that could be used to include time in the model.

   (c) **(10 points)** Fit a model with a random intercept and time effect. Use the time effect you found fit the data best in (a).

      i. What is the estimated variance of the random intercepts?

      ii. What is the estimated variance of the random slope(s)?

      iii. What is the estimated correlation between the random intercepts and slopes?

   (d) **(10 points)** Print the `v` and `vcorr` matrix from the model in (c). Comment on the heterogeneity and pattern in the correlations over time.

   (e) **(10 points)** Fit a model with random intercepts only. What do you think about this model versus the previous model? Should this be used? Why or why not?

   (f) **(20 points)** Recall that the study was designed to look at changes in percent body fat in girls before and after menarche. One question of interest may be "is the effect of time the same pre- and post-menarch?" Does the model you fitted in (c) address that question? If so, what β coefficient(s) and hypothesis test(s) answer that question? If not, design a model that does and run it using at most two random effects.

   (g) **(30 points)** Using the model you fit in (f), give a full description of your findings. Include interpretations and results of the hypothesis test(s) as necessary. All descriptions should be in the study's context and the study's goals.