Question 1 [18 marks]

CD4 cells, also known as T cells, play an essential role in your immune system for fighting infection. A CD4 percentage represents the percentage of total lymphocytes that are CD4 cells. For HIV negative people, this is usually around 40%, for people with HIV, it will generally be lower than this.

The CD4 data has CD4 percentages for a set of young children with HIV who had their CD4 % measured several times over a period of two years. These children were assigned to either a control group (treatment=1) or a zinc treatment group (treatment =2) to see if zinc improved CD4 counts.

Here is a description of the non-obvious variables:

- newpid= child identifier
- baseage = age at initial doctor visit
- visage = age at particular visit
- time = visage-baseage

a) [2 marks] Plot the distribution of the outcome (CD4PCT). Based on what you see, propose a reasonable transformation for the outcome.
b) [2 marks] Plot the transformed CD4 percents at each time point for any 10 children in your dataset. You can do this in one plot.
c) [5 marks] Write a model predicting CD4 percents as a function of time, with intercepts varying by child. Use latex. Define any notation you introduce. Fit this model using lmer(), neatly present your model code and estimated model parameters.
d) [3 marks] Extend your model in c) to include treatment and baseline age as predictors. What is different about these predictors than the predictors we used in the radon example? What effect did including these predictors have on your model parameters? Comment.
e) [1 mark] Keeping your transformation in mind, interpret the effect of treatment. Is the treatment effective?
f) [2 marks] For the child with newpid=9 in the dataset, predict their CD4 percentage at 1.4 years, include a prediction interval (use simulations).
g) [3 marks] For a child with average baseline age and treatment=1, predict their CD4 percentage after 1 year, include a prediction interval. If you had used your model from c) instead, would the prediction interval be larger or smaller? Why?