BIOS 6612 Homework 5: Intro to Longitudinal Analysis

This homework focuses on the simplest longitudinal analysis (2 time points). The data `hw5.txt` contains pre- and post-bronchodilator FEV1 measurements, and the difference between them, for the first 50 COPDGene subjects. In questions asking you to write out a model equation, write the equations in subject form and use Y_{i1}, Y_{i2} for the values of FEV1 at baseline and post-bronchodilator measurements respectively; include numeric estimates for each parameter represented in the model equation.

1. **Model 1**: change-score model. Fit an intercept-only linear regression model for the difference (`delta_FEV1`).
   (a) Write out the model equation, including the distribution of the error term.
   (b) Is there a significant difference between post- and pre-bronchodilator FEV1? Give a p-value to support your decision.
   (c) What simple test other than a linear regression yields the same results?
2. **Model 2**: baseline-as-covariate model. Perform a linear regression for the `post_FEV1` value using the baseline variable (`pre_FEV1`) as a covariate.
   (a) Write out the model equation, including the distribution of the error term.
   (b) Is there a significant association between post- and pre-bronchodilator FEV1? Give a p-value to support your decision.
3. **Model 3**: hybrid model. Fit the change-score model (outcome `delta_FEV1`) using pre-bronchodilator FEV1 (`pre_FEV1`) as a covariate.
   (a) Write out the model equation, including the distribution of the error term.
   (b) Is the change in FEV1 significantly associated with the baseline pre-bronchodilator FEV1 value? Give a p-value to support your decision.
   (c) Is there a significant **difference** between pre- and post-bronchodilator FEV1, controlling for baseline FEV1 values?
4. Show algebraically that Model 1 is nested within Model 2. What parameter values are fixed in Model 1 that are allowed to vary in Model 2? What does this say about the baseline-as-covariate model versus the change-score model in general?
5. Show algebraically that Models 2 and 3 are equivalent.

6. **Model 4**: ignore within-subject aspect of the data. If we convert this data into "long" format (i.e., one row per subject per time point), then we can write the model as $Y_{ij} = \beta_0 + \beta_1 \mathbb{1}(j = 2) + \epsilon_i, j = 1, 2$. The estimates and standard errors from this model fit are given in the table below:

	Estimate	Std. Error	t value	Pr(>|t|)
$\hat{\beta}_0$	1.6480	0.1256	13.1176	0.0000
$\hat{\beta}_1$	0.0758	0.1777	0.4268	0.6705

   (a) Interpret the intercept estimate in this model.

   (b) Compare the results from Model 4 with the results from Model 1. What parameters are equivalent?

   (c) Explain why the standard error is higher for $\hat{\beta}_1$ in Model 4 than for $\hat{\beta}_0$ in Model 1.

   (d) What simple test other than a linear regression yields the same results?