1. Intercepts and slopes: Consider two groups A and B, in both of which the relationship of an outcome y is well described as a linear function of an explanatory variable x. Let the data in group A follow the model $y_{i,A} = \alpha_A + \beta_A x_{i,A} + \epsilon_{i,A}$, and the data in group B follow the model $y_{i,B} = \alpha_B + \beta_B x_{i,B} + \epsilon_{i,B}$. Let $\bar{y}_A$ and $\bar{y}_B$ be the sample means for groups A and B respectively.

   (a) Compute the expected value of the sample mean $\bar{y}_A$ as a function of α_A, β_A and $\bar{x}_A$.

   (b) Describe a situation, either graphically or mathematically, where $\mathrm{E}[\bar{y}_A] > \mathrm{E}[\bar{y}_B]$ but the regression lines are the same ($\alpha_A = \alpha_B$ and $\beta_A = \beta_B$).

   (c) Describe a situation in which $\mathrm{E}[\bar{y}_A] > \mathrm{E}[\bar{y}_B]$ but $\alpha_A < \alpha_B$ and the distributions of x-values in the two groups are the same.

   (d) Describe a situation in which $\mathrm{E}[\bar{y}_A] > \mathrm{E}[\bar{y}_B]$ but $\beta_A < \beta_B$ and the distributions of x-values in the two groups are the same.

   (e) Is it possible to have $\mathrm{E}[\bar{y}_A] > \mathrm{E}[\bar{y}_B]$ but $\alpha_A < \alpha_B$ and $\beta_A < \beta_B$?

2. CD4 counts: The dataset `cd4.dat` contains data on CD4 cell percentage (`cd4`) among a set of HIV patients under two different treatments (`trt`). The variable `time` indicates the time in years during the two-year study that the measurement took place.

   (a) Fit a model `fit0` that represents `cd4` as a function of `time`. Fit a model `fit1` that represents `cd4` as a function of `time`, but has a separate slope and intercept for each treatment. Evaluate the residuals of `fit1` for a mean-variance relationship and for normality. Compare `fit0` to `fit1` via an F-test and discuss evidence for a treatment effect.

   (b) Fit a model `fit2` that allows for a separate slope and intercept for each person `pid`, but doesn't have `trt` in the model. Fit a model `fit2b` that has the same terms as `fit2`, in addition to a separate intercept and slope for each treatment group. Compare `fit2` and `fit2b` via an F-test and explain your results. Does this approach adequately evaluate the effects of `trt` while accounting for across-group heterogeneity?

(c) Now obtain OLS regression coefficients $(\hat{\beta}_{0,j}, \hat{\beta}_{1,j})$ for each subject j separately. Make a histogram of both coefficients and describe their heterogeneity across subjects. Using a simple paired t-test, evaluate evidence that treatment has an effect on intercepts (the $\hat{\beta}_{0,j}$'s) or the slopes (the $\hat{\beta}_{1,j}$'s).

(d) Fit a HLM `fit.lme1` to estimate the effects of `trt` while accounting for across-subject heterogeneity. Specifically, fit an HLM that includes fixed effects for `trt`, `time` and `trt`×`time`, and a random intercept and random effect for `time`. Obtain an approximate 95% CI for the fixed effects for `trt` and `trt`×`time`, and discuss the evidence for a treatment effect.

(e) Fit an HLM `fit.lme0` as in the previous part but without the `trt` and `trt`×`time` terms. Compare `fit.lme0` to `fit.lme1` in terms of AIC and BIC and discuss the evidence for a treatment effect. Also, obtain shrinkage regression coefficients for each subject and compare them graphically to the OLS estimates.

(f) Discuss the different approaches used here to evaluate the effects of treatment. What are your conclusions about a treatment effect?