

## Problem Set 9

**Problem 1 (tomato plants).** The file `tplant.dat` contains data on the heights of ten tomato plants, grown under a variety of soil pH conditions. Each plant was measured twice. During the first measurement, each plant's height was recorded and a reading of soil pH was taken. During the second measurement only plant height was measured, although it is assumed that pH levels did not vary much from measurement to measurement.

- (a) Using ordinary least squares, fit a linear regression to the data, modeling plant height as a function of time (measurement period) and pH level. Interpret your model parameters.
- (b) Perform model diagnostics. In particular, carefully analyze the residuals and comment on possible violations of assumptions. In particular, assess (graphically or otherwise) whether or not the residuals within a plant are independent. What parts of your ordinary linear regression model do you think are sensitive to any violations of assumptions you may have detected?
- (c) Hypothesize a new model for your data which allows for observations within a plant to be correlated. Fit the model using a MCMC approximation to the posterior distribution. Check the convergence of your MCMC sampler and present diagnostics for your approximation.
- (d) Discuss the results of your data analysis. In particular, discuss similarities and differences between ordinary linear regression and the model fit with correlated responses. Are the conclusions different?

Solutions will be discussed in class on October 31.