Christian Conroy

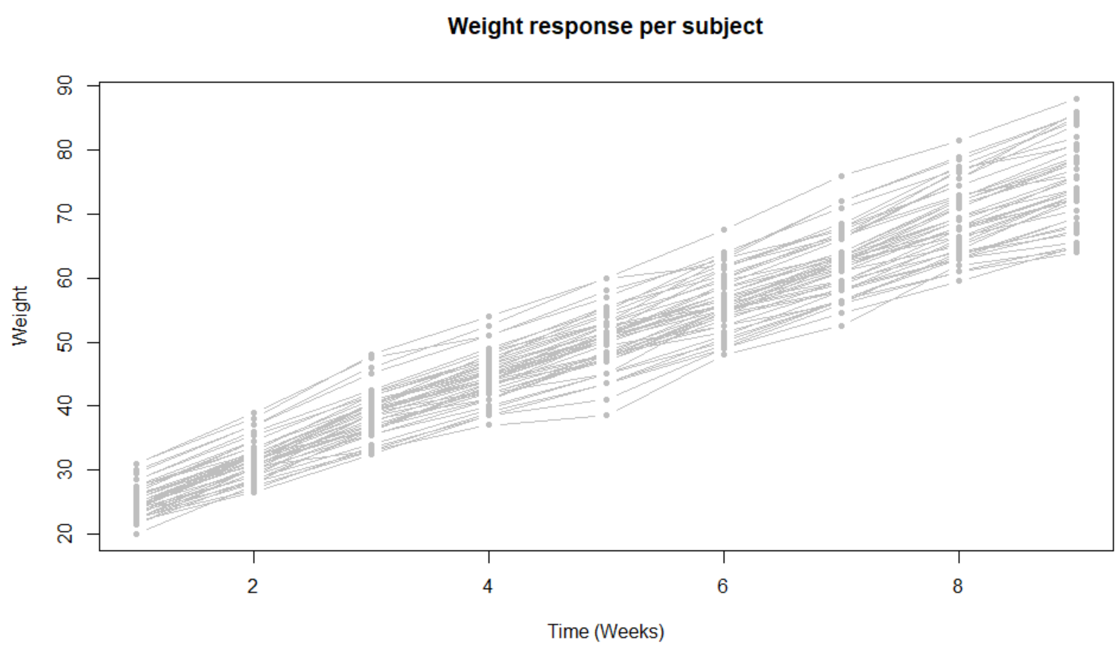
MATH 426: Applied Longitudinal Analysis
Homework 6, due Thursday, November 30

Please submit a PDF or .doc version of your homework to Blackboard by 11:59pm on the due date. Please type all responses. You are encouraged to use R for all calculations. Please include a commented code appendix at the end of the assignment.

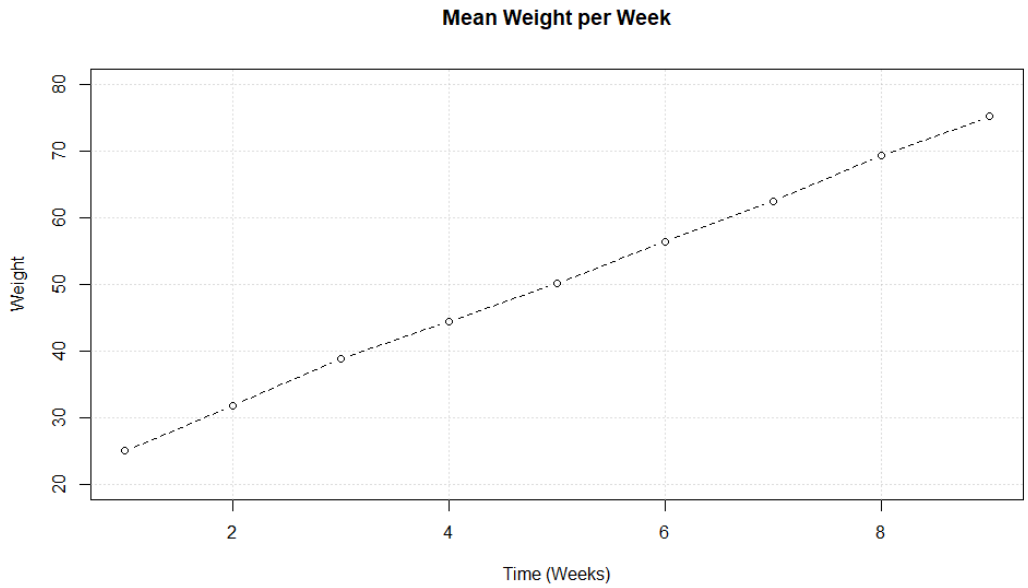
Linear Mixed Effects Models

In the R package SemiPar is a dataset called pig.weights that includes data on the bodyweights of 48 pigs measured in nine successive weeks (so spanning an eight week period between the first and last measurement). Each row of the dataset includes a pig ID number and then the nine successive weights.

1. Descriptive Analysis: Create two plots, one showing the trajectories of each pig's weight over time and the other showing the trajectory of the mean weight over time. Comment on the notable features in these plots.



Christian Conroy



In looking at the mean response per pig, it is clear that there is generally an upward trajectory in weight across the weekly measurement periods for all pigs. However, there are occasional departures from the general upward trajectory, with some pigs experiencing declines in weight at certain points only to experience greater weight gains in the future in line with the general upward trajectory. We are therefore not surprised that the mean response profile showing the trajectory of the average weight for the 48 pigs at each weekly measurement period exhibits a consistent upward trend.

2. Obtaining and Interpreting a Linear Mixed Effects Model for Bodyweight: Using lme(), fit a model for weight over (continuous) time which includes subject-specific intercepts and slopes as random effects.

(a) Obtain and interpret parameter estimates and associated 95% confidence intervals for the fixed effects in the model for the mean trajectory.

The estimated slope for the pigs is 6.21, representing a 288.67% increase in weight for each additional week. We are 95% confident that the true parameter for the mean trajectory of pig weight over time falls between 6.03 and 6.39.

(b) Obtain and interpret parameter estimates for the variances and correlation in the model.

The within-subject "error" variance is 1.6 and the between subject variance of the intercepts is 6.99. We would conclude then that there appears to be more variance between subject intercepts than there is for within subject "error" variance. The variance for the slope or change over time among subjects is 0.379 and the correlation of the intercept and the slope among subjects that describes the relationship between the subject-specific distributions over time and the subject-specific distribution at baseline is -0.063.

(c) Obtain and interpret the 90% normal range for trends over time in body weight among pigs in the population sampled (i.e. for the pig-specific random effects for trend over time).

Christian Conroy

The 90% normal range for trends over time in body weight among pigs in the population sampled is (5.59, 6.83). Under the assumption of normality, we expect 90% of the pig-specific random effects for trend over time to fall between this range.

3. The investigator who provided the data is interested in designing a randomized clinical trial to evaluate an additive to the standard pig feed which might increase the rate of growth over time in bodyweight in pigs. He feels that an increase in bodyweight of 0.2 kg per week above that observed in the study for which the data are provided would be important, and would like to design a study to have 90% power to detect this increase using a two-sided 0.05 level of significance.

(a) What sample size would be needed in the randomized trial if the growth rate in the control group (without the additive) was the same as observed in the study for which the data are provided?

A sample size of 213 would be needed in the randomized trial if the growth rate in the control group was the same as observed in the study for which the data are provided.

(b) Using the same number of pigs per group as in the pilot study and 80% power, find the smallest detectable effect size.

The smallest detectable effect size would be 0.364.

GLM Review

For each of the following, summarize the **<u>estimated mean model</u>** in a table along with 95% confidence intervals and provide meaningful interpretations of the estimated coefficients (i.e. not on the β scale).

1. The Asthma Coalition on Community, Environment, and Social Stress (ACCESS) project, a prospective pregnancy cohort, was designed to study effects of prenatal maternal and early life stress and other environmental risk factors on childhood asthma risk. Using the dataset access.txt, build a model to predict a child's wheeze status (defined as two or more wheezing events in given the mother's typical bedtime cortisol level (cort5) and BMI (bmi) during the second trimester.

Holding the bedtime cortisol level constant in the model, we might say that the odds of a child having a wheeze for someone considered obese are e to the 1.034 times the odds (2.812) of someone considered non-obese having a wheeze. Holding BMI status constant in the model, the odds of having a wheeze for individuals with a bedtime cortisol level of x + 1 are e^0.864 = 2.373 times the odds of having a wheeze for individuals with a bedtime cortisol level of x. Both effects are statistically significant at the 95% level of confidence. We are 95% confident that the true parameter for the mean trajectory of bedtime cortisol level falls between 1.11 and 1.21. We are 95% confident that the true odds of someone whose mother was considered obese at the time of the second trimester having a wheeze is between 1.13 and 7.15 times greater than the odds of someone whose mother was not considered obese at the time of the second trimester having a wheeze.

2. A health insurance company collected information on 788 of its subscribers who had made claims resulting from ischemic (coronary) heart disease. Data were obtained on total costs of services provided for these subjects and the nature of the various services for the period of January 1, 1998 through December 31, 1999. Using the dataset ihd.txt, build a model to predict the number of interventions performed per subject during the two year period using subject's sex and presence of at least one comorbidity as predictors.

Christian Conroy

Holding the presence of at least one comorbidity constant in the model, the rate of interventions performed per subject over the two year period is exp(0.1129) = 1.20 greater for males than that of females over the same time period. Similarly, holding sex constant in the model, the rate of interventions performed per subject over the two year period is exp(0.4639) = 1.59 greater for those experiencing the presence of at least one comorbidity than that of those experiencing no comorbidity over the same period. The effect of sex is statistically significant at the .01 level of confidence and the effect of comorbidity is statistically significant at the .001 level of confidence. We are 95% confident that the true rate ratio for the incidence of interventions performed per subject over the two years for males versus females falls between 1.04 and 1.21. We are 95% confident that the true rate ratio for the incidence of interventions performed per subject over the two years for those experiencing at least one comorbidity versus those experiencing no comorbidity falls between 1.48 and 1.71.