case1final

Nathaniel Brown, Annie Tang, William Yang

September 13, 2017

Introduction

The purpose of this project is to analyze rat bioassays. is there variation in results between labs? there shouldn't be. Here are a few plots showing the four protocols and other important things.

Model-Fitting

$$y_{ij} \sim \beta_{0,i}+\beta_{1,i}x_{ij,d_1}+\beta_{2,i}x_{ij,d_2}+\beta_{3,i}x_{ij,p_B}+\beta_{4,i}x_{ij,p_C}+\beta_{5,i}x_{ij,p_D}+\beta_{6,i}x_{ij,log(w)}+\epsilon\beta_{0:5,i} \sim N(\mu_{0:5,i}, \sigma^2_{0:5,i})\epsilon \sim N(0,\sigma^2)$$

To fit the data, we used a mixed model, with fixed and random effects. Let y_{ij} be the observed log(blotted uterus weight) for subject x_{ij}, the jth individual in lab i. x_{ij,d_1} and x_{ij,d_2} are the values of dose1 and dose2 for subject x_{ij}. x_{ij,p_B}, x_{ij,p_C}, and x_{ij,p_D} are dummy variables indicating which protocol x_{ij} was subjected to. $x_{ij,log(w)}$ is the log(body weight) for x_{ij}. Body weight and uterus weight are log-transformed to account for the right skew in the data (Figure 1). We make the Gaussian assumption that the coefficients, β, are normally distributed according to some μ_i and σ_i. We add a random effect on all $\beta_{0:5,i}$ to account for the lab-to-lab variability in the intercepts and slopes of the blotted weight (Figure 2). The summary statistics of the model described above can be found on Table 1.

We start at a reduced form of this model and augment it to its full form after initial analyses.

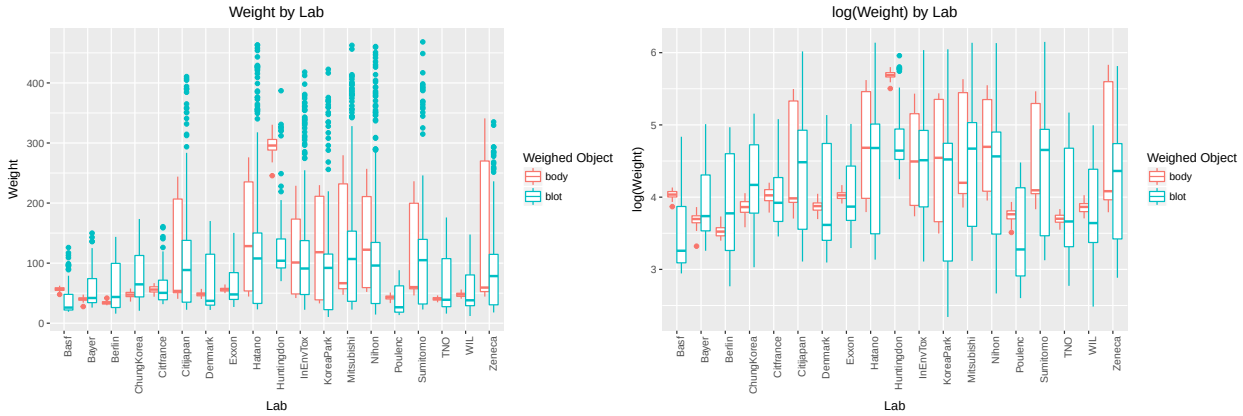


Figure 1: Log Transformations of Weights

Mean Fixed Eff	ects Variance Fixed	Effects
(Intercept)	4.9844	0.1421
protoB	0.0397	0.0007
protoC	1.2966	0.0193
protoD	1.2949	0.0204
log(body)	-0.0264	0.0092

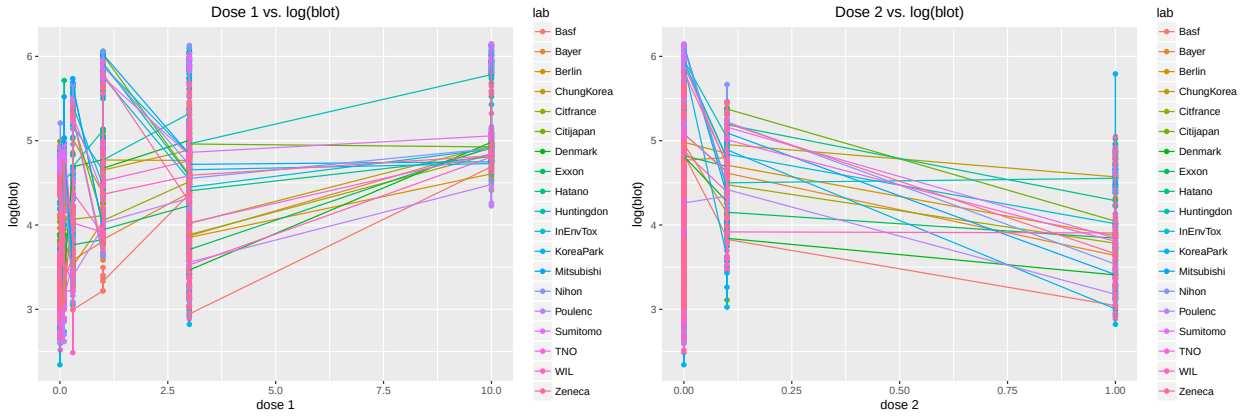


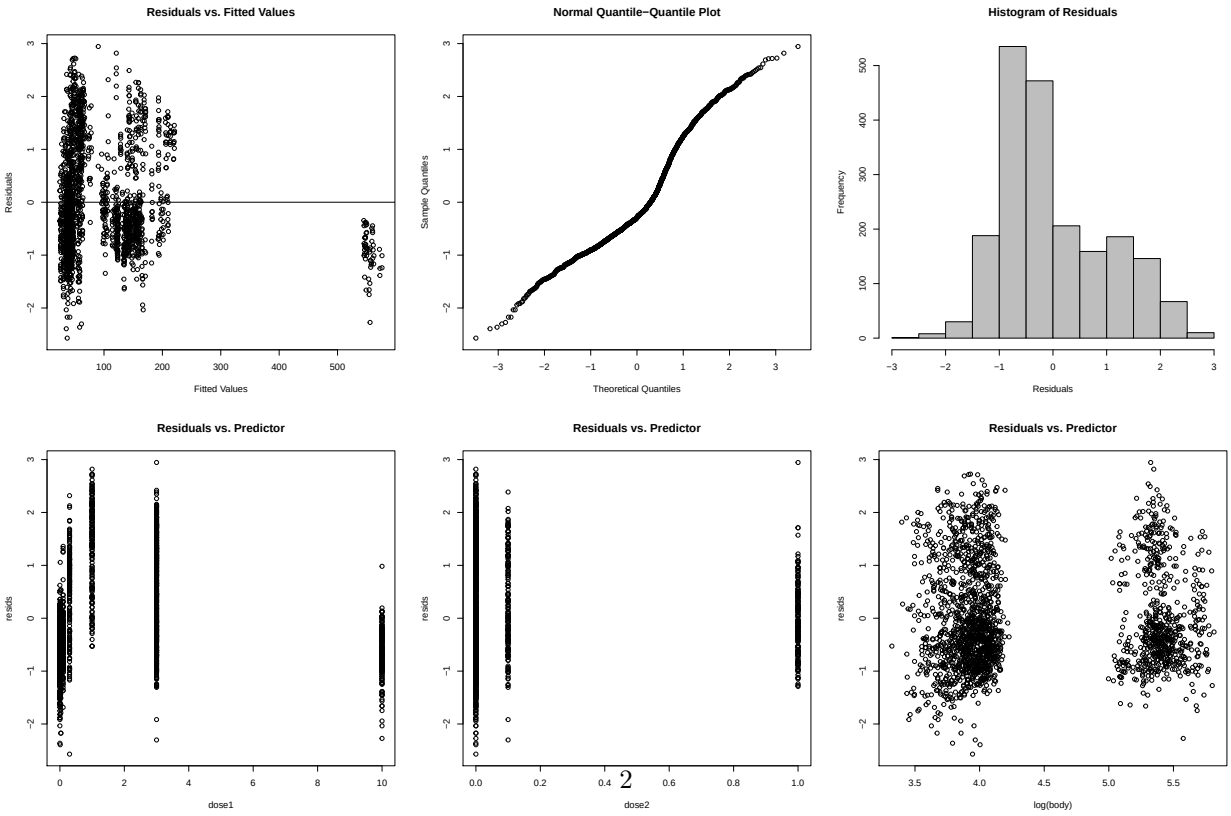
Figure 2: Lab-to-Lab Variability in Intercepts and Slopes

Table 2: Summary Statistics of Initial Model

	Mean Random Effects	Variance Random Effects
(Intercept)	2.9018	1.6181
dose1	0.0000	0.0211
dose2	0.2938	0.2691

	Residual Variance
residual	0.1922

The diagnostic plots to assess model fit are below:

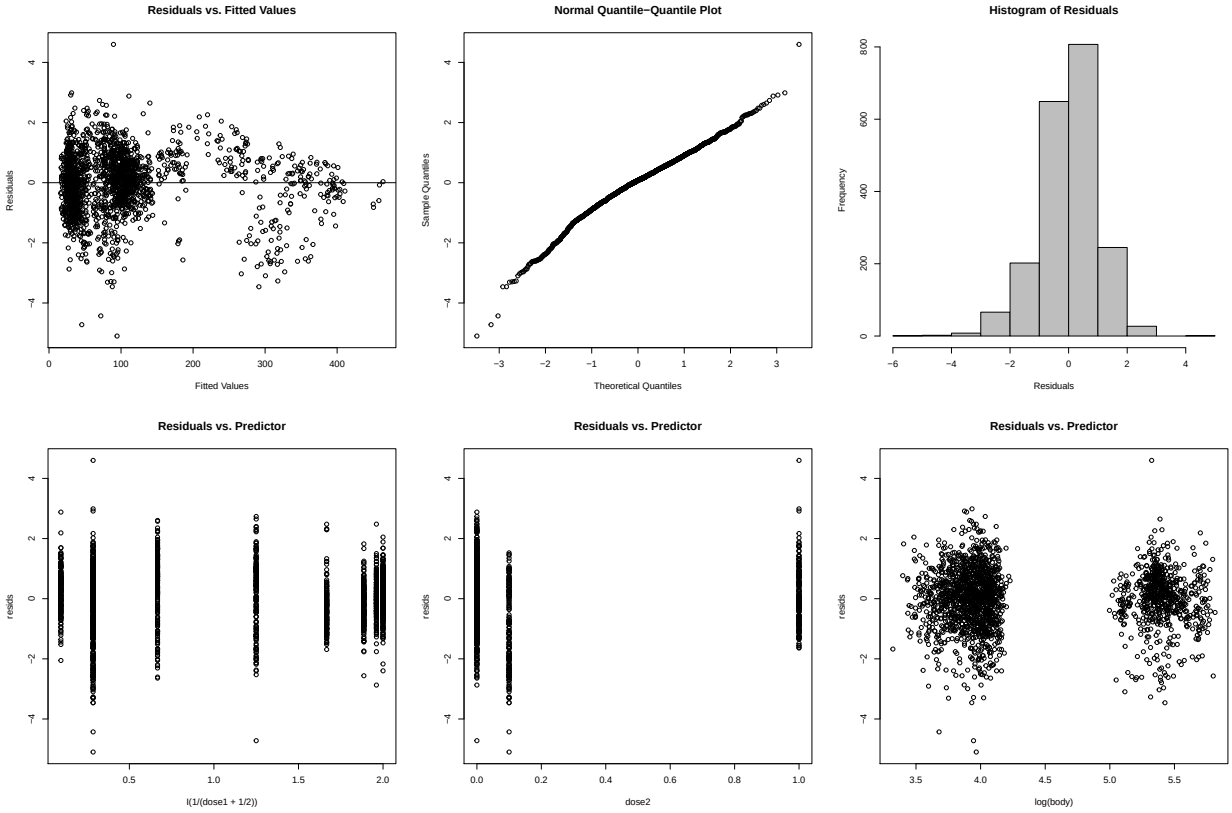


Among the problems in these plots is the nonlinearity in the Residuals vs. Predictor plot for dose1. To counter this, we transformed dose 1 by the reciprocal of (dose1 + 1/2) (we add a small number to dose because we

	Mean Fixed Effects	Variance Fixed Effects
protoD	0.8524	0.0100
log(body)	0.2946	0.0046

	Mean Random Effects	Variance Random Effects
(Intercept)	3.4691	0.9593
I(1/(dose1 + 1/2))	0.4969	0.5371
dose2	0.7532	1.0481

	Residual Variance
residual	0.0797



	MAE	RMSE
Random Dose + Transformations	21.7968	37.7587
Fixed Dose	36.4387	61.0759

We see that the variability due to sources other than the random effects is 0.0797 (decreased from 0.193). Also, the diagnostic plots improved, although it is notable that the residual variance decreases for larger fitted values. The fixed dose model is also much worse at predicting than the random dose model with transformations. We still see problems with this model, including the non-randomness in the body weight

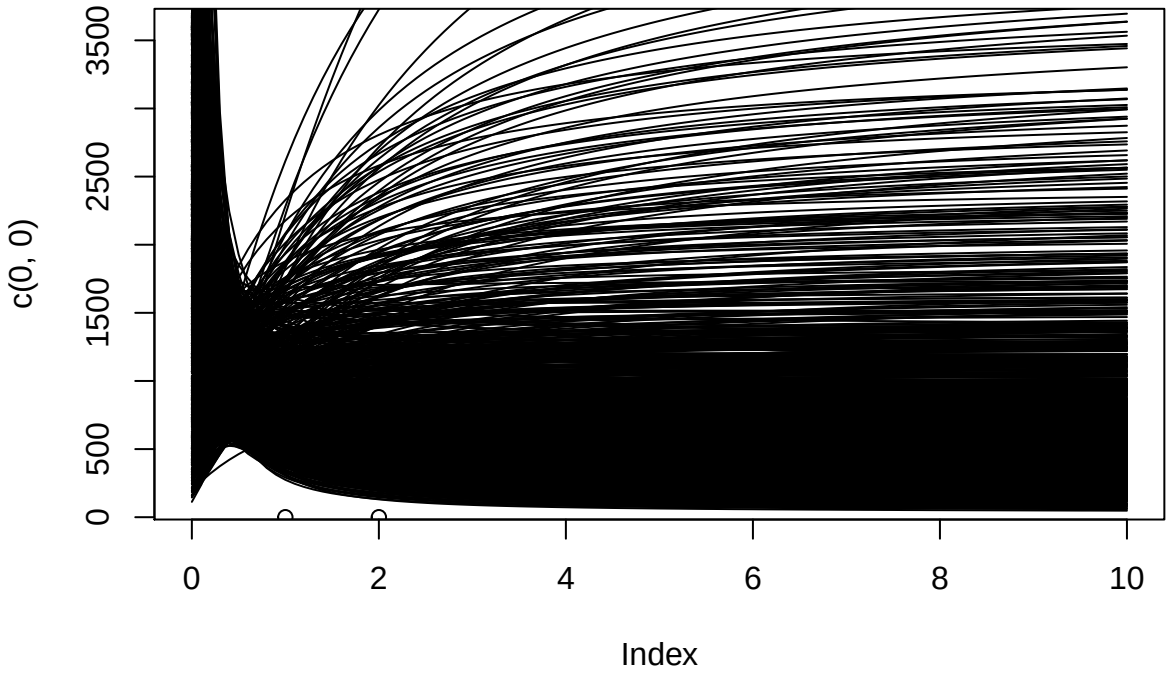
residuals. To counteract this, we added interactions between log body weight and protocol, and an anova test ($p = 0.2458$, χ^2=17.8594) confirmed that the interactions improve the model.

m5 stuff

Lab Variation

INCOMPLETE

The plot below samples random dose1 effects (slopes) and lab effects (intercepts) from the mixed model, holding all other predictors constant, and plots the results. The variance in the dose1 effects by lab is noticeably high, so we say that the bioassay does depend on lab and thus this study fails miserably.



Conclusion

Because the dose 1 effect has such a high variance relative to the mean, our model shows that the bioassay being studied does not measure a consistent response to dosage across labs.

Contributions

Nathaniel Brown made the visualizations for this report. He also organized the relevant files in a Github repository for the group to access and edit. Annie Tang compiled the group work done on EDA into a .rmd and wrote the accompanying explanations for the EDA and approaches to analysis. William Yang helped

pair on EDA analysis and identify approaches to handle the data. Approaches to analysis were a joint effort by all members of the group. Nathanial implemented analysis for the univariate normal and multivariate normal approaches. Implementation and analysis of the mixed effects model was a joint effort by all members of the group.