STA 723 HW05 Report **Zach White**

Executive Summary

We will test a new approach, the rat uterotrphic bioassay, for screening chemicals for endocrine disrupting effects. We use a Bayesian hierarchical model to analyze the effectiveness of this bioassay across multiple labs and protocols in identifying the effects of estrogen agonists and antagonists. We use random effects for both lab and protocol, and we use the covariates of a known estrogen agonist (EE), antagonist (ZM), and EE^2. Through this model, we obtain a posterior distribution and perform two-sample t-tests to compare the differences between control and experimental groups using the posterior predictive.

Introduction

In today's world, there are a significant number of substances that surround us that behave like estrogens, which are a group of hormones produced by both males and females. Estrogens are influential during puberty, menstruation, and pregnancy, but they also help regulate the growth of bones, skin, and other organs and tissues. Environmental estrogens are synthetic compounds and plant products present in the environment. These chemicals can cause endocrine disruption. We test the effectiveness of the rat uterotrophic bioassay, which is designed to test the effect of estrogen agonists and antagonists on a particular hormonal response. The main goal of this international multi-laboratory study we analyze is to assess whether the results of the bioassay are consistent across laboratories.

In addition, we seek to answer three additional questions. The first is whether the bioassay is successful at identifying estrogenic effects of EE and anti-estrogenic effects of ZM and whether or not there are differences between labs. The second is whether or not the dose response varies across labs, and the third is whether or not the different protocols differ in their sensitivity to detect estrogenic and anti-estrogenic effects. The response variable is uterus weight because hypothetically estrogenic effects will lead to an increased uterus weight, while anti-estrogenic effects will have the opposite effects.

Analysis

Exploratory Data Analysis

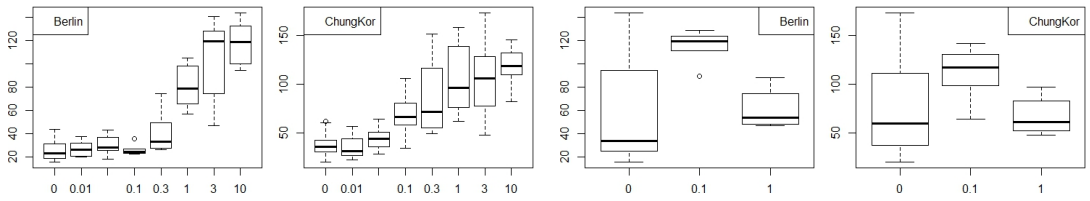
We must first understand the structure of the study and the data to answer these questions. Uterus weight is the response variable. We have 6 possible explanatory variables: weight of rat (g), EE (mg/kg/day), ZM (mg/kg/day), Lab, Group, and protocol. This study was carried out across 19 different labs using four different protocols described by the following.

Protocol	Description
A	immature female rats dosed by oral gavage (3 days)
B	immature female rats dosed by injection (3 days)
C	adult ovariectomized female rats dosed by injection (3 days)
D	adult ovariectomized female rats dosed by injection (7 days)

Analyzing a scatterplot of uterus weight by overall weight, the data are bunched in two groups, which are clearly associated with protocol and weight. The group associated with lower weights represents protocols A and B, while the heavier group represents protocols C and D. It is clear that protocol is associated with both weight and uterus weight, which is intuitive, since protocol is linked to age, which will influence those characteristics. Thus, we include protocol rather than weight since it aligns better with the goals of our analysis, and they seem to share much of the same information. The variable group seems more

important in the design of the experiment, but it does not seem associated with uterus weight. We will not include it in our model.

The following plot shows first uterus weight by EE dose for two arbitrary labs, and second uterus weight by ZM dose for the same two labs. There is a positive relationship



with EE and uterus weight. Adjusting the scale of the x-axis to appropriately match the dose scale makes it clear that this relationship may not be completely linear under EE. We propose a model that includes quadratic of EE. For this reason, we will treat EE as a continuous variable, even though under the design of the study, they seem like factors. We do this because of the flexibility and interpretability of a continuous predictor and its quadratic term. As far as ZM is concerned, the control group has much higher variance than the experimental groups, and the control is actually associated with a lower mean, which isn't expected since ZM is known to have anti-estrogenic effects, which lowers uterus weight. We will include ZM to see if this relationship holds. We will also treat ZM as a continuous predictor.

Model Choice

A clear, intuitive hierarchical structure for the purposes of our analysis is some hierarchy around lab and protocol. These are two facets of interest from the study that fit well in some hierarchy. From three potential models, we choose a Bayesian hierarchical normal regression model. The first model we explored was a hierarchical ANOVA model with varying intercepts and coefficients, but under our hierarchical structure, there were too many parameters. This model wasn't identifiable, and interpretation of the parameters was too ridged and unintuitive. The second model was nested-hierarchical normal regression model, but there were similar issues as the previous model. Our model of choice is a Bayesian hierarchical normal regression model with random effects on both lab and protocol. We use the covariates of EE, ZM, and EE^2.

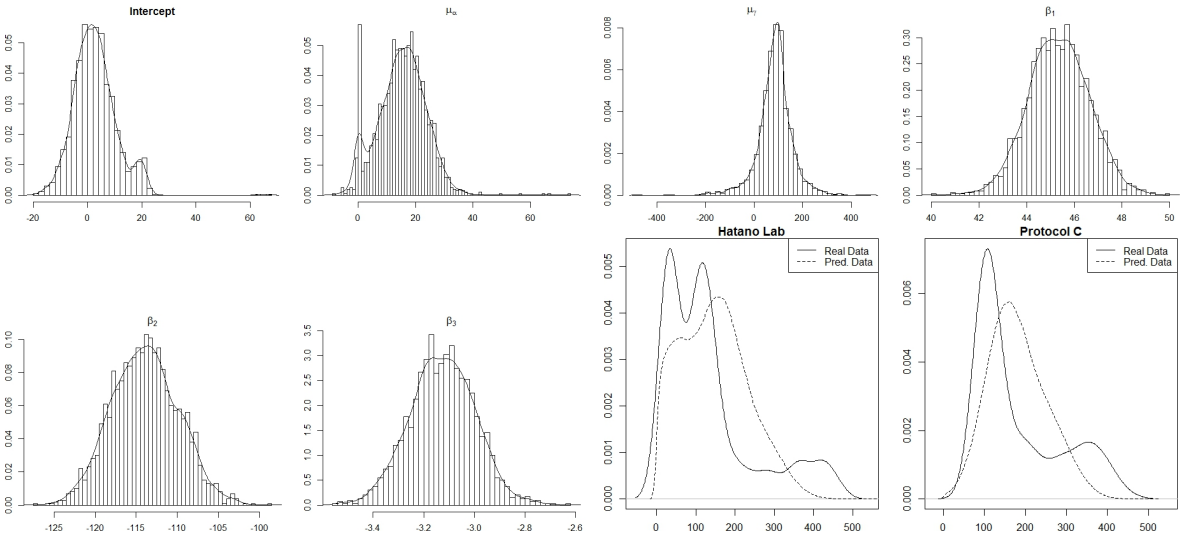
$$Y_{ijt} \sim N(\mu_{ijt}, \sigma^2) \qquad \alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2)$$
$$\mu_{ijt} = \mu + \alpha_i + \gamma_j + \beta_1 \text{EE}_{ijt} + \beta_2 \text{ZM}_{ijt} + \beta_3 \text{EE}^2_{ijt} \qquad \gamma_j \sim N(\mu_\gamma, \sigma_\gamma^2)$$

Under this model, we also place diffuse priors on μ, μ_α, σ_α^2, μ_γ, σ_γ^2, β_1, β_2, β_3. We set both $\alpha_1 = \gamma_1 = 0$ for identifiability. Under this model, we propose the data follow a normal distribution with μ_{ijt}. We propose random effects on lab(α_i) and protocol (γ_j). This is not a nested hierarchical model, and thus we propose shared coefficients across labs and protocols for EE, ZM, and EE^2. We generate draws from the posterior distribution, and we then use these draws to find posterior predictive distribution using the values of the given data. We perform t-tests on the generated datasets to perform lab and protocol specific comparisons of the control and experimental groups for both EE and ZM.

Results

Analyzing posterior draws from the figure below, we can test the significance of the parameters of our proposed model. Under our model, the intercept does not seem to be significant. For sake of space, we just include μ_α, which is the super mean of the labs.

We can see that uterus weights vary between labs, but the super mean is clearly positive. The super mean μ_γ is less clearly non-negative, the majority of the draws are greater than zero. However, the magnitude is much larger, which indicates that protocol will be associated with larger changes of uterus weight. Each of the coefficients β_1, β_2, β_3 are significant, and β_1 is positive while the other two are negative, which at least for β_1 and β_2 is to be expected and proves to some degree the efficacy of the bioassay.



The last two plots show a comparison for two labs between simulated uterus weights and their corresponding actual uterus weights. It is clear that our model does a good job at achieving the overall shape, but it misses some features. For examples, the distribution for uterus weights in the Hetano lab has three modes, and our predictive distribution only has 2 modes. However, under our model, we do model the means quite effectively. Thus although our predictive distribution does not capture all the features, it does capture the mean.

With our simulated data, we perform two-sample t-tests to compare experimental and control groups for each lab and protocol. By doing this, we can test if certain labs or protocols can detect the effects of EE or ZM on uterus. By so doing, we can conclude that each of the labs can detect the effect of EE on rats with equal efficacy according to a Bonferroni adjusted p-value. however, it seems that there are some labs that don't detect the effect of ZM effectively. Both labs Basf and Hatano did not detect effects of ZM well. As far as protocol is concerned, it seems that protocol A is not as effective at the other as the other protocols at detecting the effects of EE and ZM. We can conclude that there is a difference between labs and protocols. However, for the most part, the labs detect changes in EE more effectively that they do ZM, which also may be due to the difference in sample size.

Conclusion

We use a Bayesian non-nested hierarchical model to analyze the effects of ZM and EE on uterus weights across labs and protocols. We use random effects on both lab and protocol because a goal of the study is to analyze the differences between both labs and protocols. We find that there are differences across labs and protocols in the ability of the bioassay to detect effects of EE and ZM. A clear limitation of our method is that we use t-tests to analyze the difference between control and experimental groups of ZM and EE, but we used our generative model to simulate this data. Thus, if our generative model doesn't represent the data well, we are overstating our confidence int the results.