Bioassay 808

Team Multicollinearity

11/3/2019

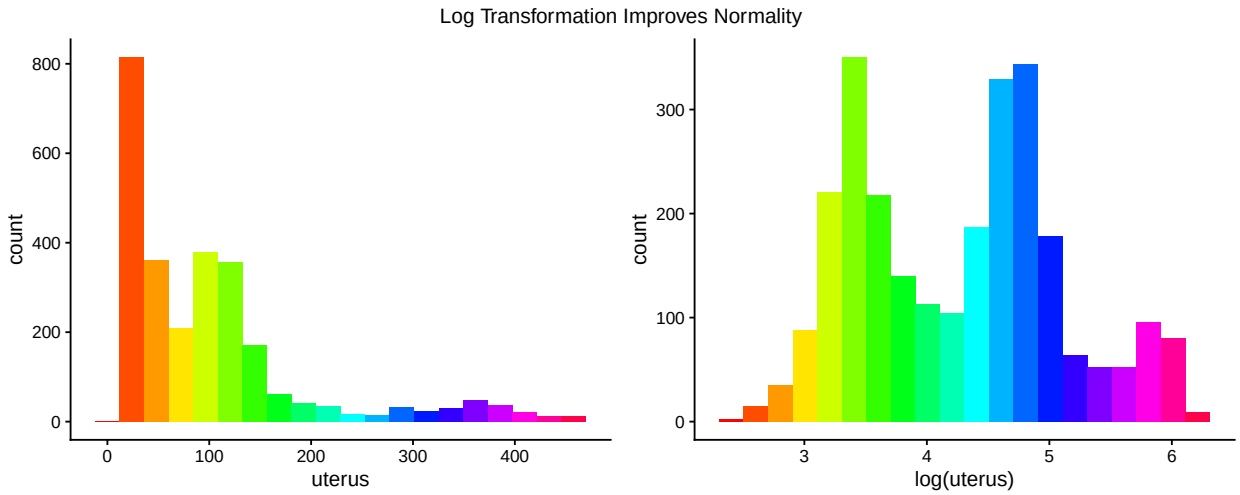
Summary:

Throughout the paper, a hierarchical model was used to understand effects of agonist, EE, and antagonist, ZM, on uterine weight, with the use of 4 different dosing protocols. It was found that EE was correlated with an increase in uterine weight per increase in EE dose and at EE = 3, ZM was associated with a decrease in uterine weight for all protocols except protocol A.

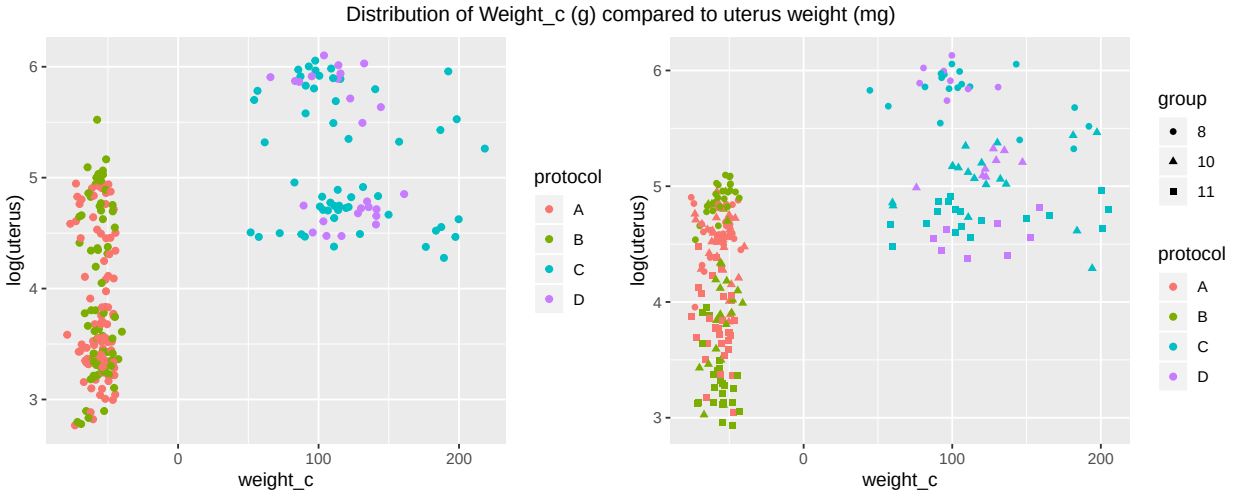
Intro:

The purpose of this paper is to understand the experiments which were conducted by multiple labs on rats receiving an estrogen agonist EE and antagonist ZM to see if it had an impact on the weight of the uterus where the rats either were immature or had their ovaries removed. The design of the experiment randomized female rats to different groups where several groups had increases doses of the agent and the remaining were a control group.

EDA:



When exploring the data, the first issue that stood out was that there were a few instances where there was missing data. Because only 4 rows had missing data, and were in separate groups, we felt that that this would have little to no impact on the overall analysis and thus removed them from the data set. From there a series of density plots were made to understand the data, when looking at the continuous variables, we see that their distributions were mostly negative/right skewed. When applying a log transformation to the uterus variable the distribution did improve though not ideally.

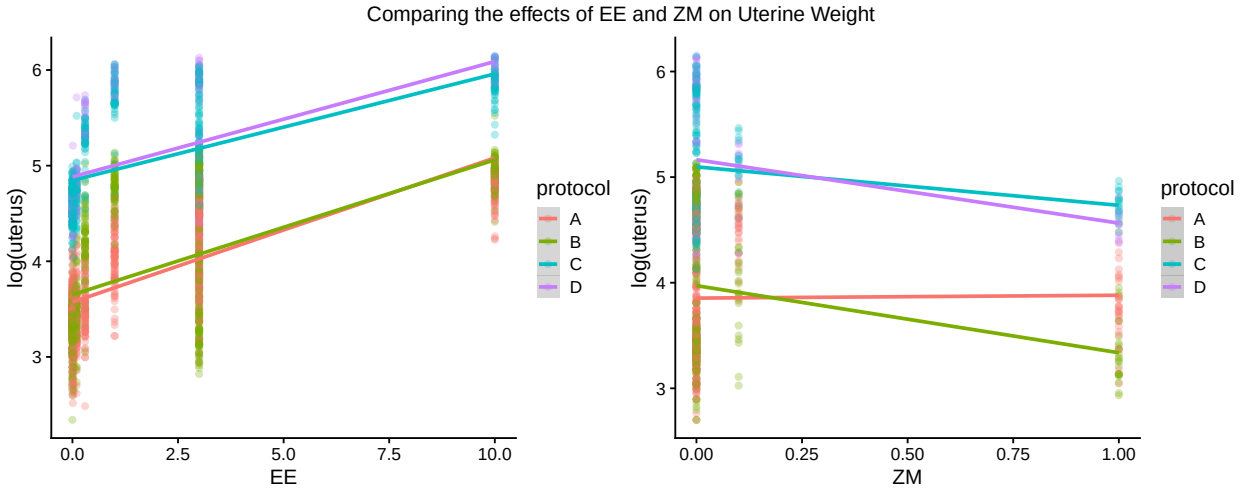


When looking at the first graph on the left where we compare the weight of the mice to the log of uterus, color by protocol. From this graph we can see that the distribution of A and B is quite similar only varying in uterus weight and not mouse weight, while C and D are also quite similar. However, Group C varies a lot more in weight that group D but the variance in uterine weight appears to be similar. The key finding from A and B is that the method of dosing the mice does not have any major difference based on how well overlapped. The key finding from C and D is that even when dosing for an additional 4 days, there is no major effect in uterus weight.

The second graph on the right we compare group number to log uterus, what we find is that from groups 1 to 9 as the dosage of EE increases the uterus weight also increases, however, past a certain dosage we see that there are diminishing returns. In addition, we see that the trends for all protocols are similar, but A and B have different intercepts compared to C and D. Lastly, there are slightly different intercepts between the labs.

Where does this go?

Focusing now on the third graph, we plot both control groups (1,2) and group 8 which has an EE of 3, and group 10 and 11 which have EE of 3 and ZM of 0.1 and 1 respectively. Group 8 overall has the uterine weight in all protocols, which is to be expected because it has and EE of 3 and no ZM, followed by group 10. This can be explained since ZM was added to that group. In group 11 which has an EE of 3 and ZM of 1, which is substantially more ZM compared to group 10 is near the same range as the control group 1 and 2.



On the left using groups 1 through 9 we compare EE to log uterus, what we find is that protocol B is the most sensitive to EE treatment as we increase the EE dosage. Groups C and D are near parallel, suggesting that injection to adult rats leads to the lowest EE sensitivity. One the right using groups 1,2,8,10,11 we compare ZM to log uterus, similar to the previous graph, protocol B is the most sensitive based of the steep decreasing slope.

Where does this go??

For the fifth graph, we selected groups 1,2,8,10,11 where we are looking at the antagonist effects of ZM. Looking at the results, we see that group 8 with EE of 3 is leads to the highest uterine weight. This is followed by groups 10,11,2,1. We see that adding the ZM does lead to a lower uterine weight and is quite close to the control levels.?? Where does this go?

Modeling:

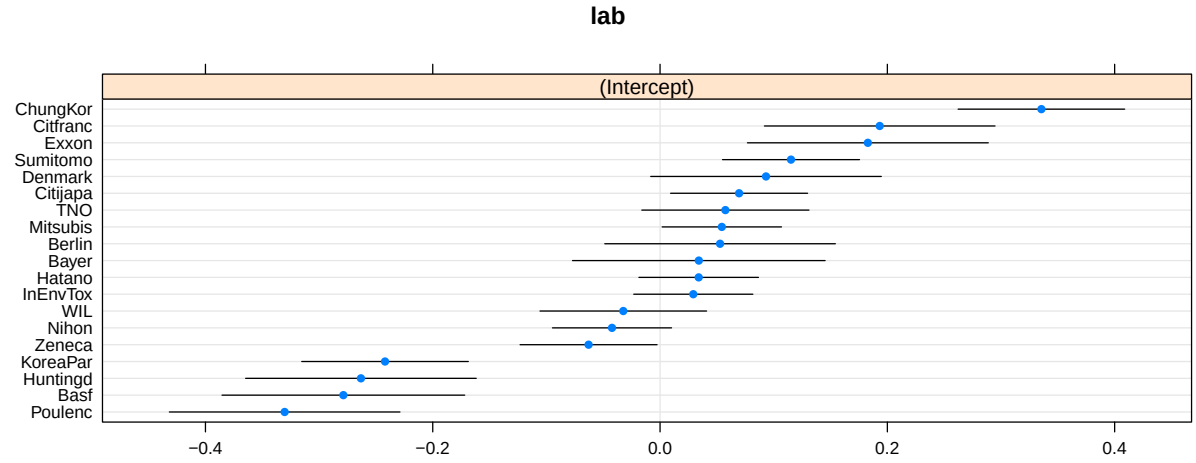
Starting with the linear model, we considered putting all the variables in the model, however it was realized that group was comprised of different combinations of EE's and ZM's, so it was discarded. Looking at the summary, we see that the remaining predictors were significant. We then performed a VIF test and the results showed weight_c and protocol had a VIF score around 15, this can be explained as protocol A&B were the low weight immature groups and C&D would be the heavier adult rats.

$$lmer(log(uterus)) = \beta_0 + \beta_{EE} + \beta_{ZM} + (1|lab)$$

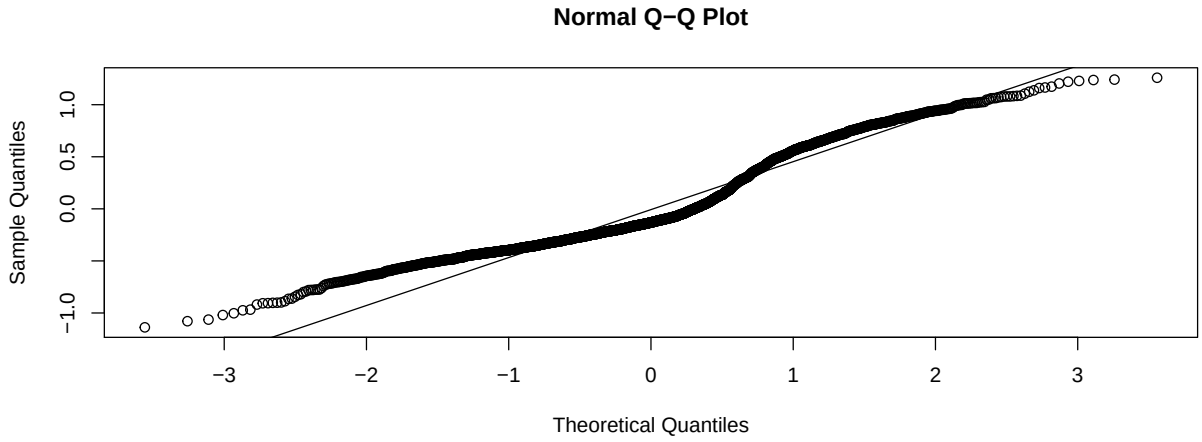
To create a model to answer whether the results from the laboratories are consistent, we decided to use a lmer model and level by lab. When taking the summary of the lmer model, weight was not found to be significant, so we decided to remove it from the model. This led to the final model.

	Coef.Estimate	Coef.Std..Error	Coef.df	Coef.t.value	Coef.Pr...t..
(Intercept)	3.61	0.04	20.31	82.07	0.00
ZM	-0.51	0.03	2652.47	-17.04	0.00
EE	0.14	0.00	2652.66	47.97	0.00
protocolB	0.05	0.02	2541.05	2.14	0.03
protocolC	1.26	0.03	2478.06	47.04	0.00
protocolD	1.26	0.03	2652.83	36.74	0.00

To see if having weight_c made the model significantly better, we performed and ANOVA test which concluded that the simpler model without weight. Looking at the results of the final model, we see that all variables are of some level of significance, in addition we see that a one unit increase in EE would lead to a 0.14 increase of log uterus. Furthermore, when EE is equal to 3, a unit increase in ZM leads to a 0.51 decrease in the log uterus. For protocols B, C, and D are similar to what we found in the EDA process based on the coefficients of these variables.



When looking at the random effects (dot plot), what we can conclude is that ChungKor and the bottom four labs have more extreme magnitude which is a result of their random effects. The variance between the rest of the labs is not too different from one another, suggesting that the model of the random effects is reasonably accurate.



Looking at the assumptions of normality based of the non-linear QQ plot it does not meet normality. Lastly, when we look at the fitted vs. residuals plot, it seems that the errors are concentrated in certain areas, which can be concerning.

Conclusion:

In conclusion, by using a linear hierarchical model where we level be lab, we find that the uterotphic bioassay was indeed successful at identifying the estrogenic effects of EE. Overall, uterus weight exhibits an increasing dose response trend for EE for all labs. For the uterus weight ZM does exhibit a decreasing uterus weight while increasing in dose for protocols B, C, and D but not for protocol A when analyzing all the labs. When faceted by lab, we see that certain labs have a positive slope for uterus weight when increasing ZM dose, given that EE = 3, while other labs display a 0 slope or slightly negative slope. This is likely due to a lack of data points for ZM = 0.1 and ZM = 1.0 across all labs. Lastly, we have also identified that group B is the most sensitive to EE and ZM as it has the highest slope when comparing it to log uterus.