1.Summary

I constructed a linear regression model to fit the data from international mutilaboratory studies. The model includes all variables except `group` and interaction between estrogen chemicals and different labs along with protocols. Apart from `weight` and `uterus`, all other variables are considered as factor. I use MLE/OLS to estimate the parameters and use t-test and F-test to assess the effects of `EE` and `ZM`, and whether the results were consistent across the laboratories. Overall, the effects are significant. And the changing dose point for `EE` is 3. Protocol A,B would be recommended. However, these results are not consistant among labs. These results can also be presented by 3 simple graph.

2.Introduction

Uterotropic bioassay is one approach for screening chemicals for endocrine disrupting effect. An international mutilaboratory study was conducted with this method using known estrogen agonist(`EE`) and antagonist(`ZM`). The main goal is to assess the effects of `EE` and `ZM`, and whether the results were consistent across the laboratories.

The dataset including: response variable `uterus`(Uterus weight (mg)); key explanatory variables: `EE`(Dose of estrogen agonist,mg/kg/day),`ZM`(Dose of estrogen antagonist, mg/kg/day);other explanatory variables: `Lab`(19 Laboratory at which assay was conducted), `Protocol`(4 levels factor, representing immature female rats dosed by oral gavage (3 days), immature female rats dosed by injection (3 days), adult ovariectomized female rats dosed by injection (3 days), adult ovariectomized female rats dosed by injection (7 days) seperately.); `Weight`(Body weight of rat (g)),`Group` (Lab replicate group (6 rats were used per group)).

3.Methods and Model

3.1 EDA

From the plot, we get main information on the dataset.1)There are significant interaction among variables. Because each labs adopted differet treatment. So we have to include the iteraction into our regression model. 2)Many discrete variables. We need to keep the model simple.

I chose boxplot to illustrate the relationship between continuous variable (`uterus`) and discrete variable(`EE` and `ZM`).Different colors represent different `labs`, and I use facet to represent different `protocol`. From the plot1, we can see that: 1)Overall, the uterus decrease as the zm increase; The variance under different level of `ZM` is different. The variance of `ZM`=0 is much larger than the other two groups; 2)Overall, the boxplots look similar. However, the different labs act different, especially when `ZM`=0. Some labs, like TNO in `protocol="B"`,`ZM=0.1` acts like an "outlier". There is interaction between `lab` and `ZM`. 3)Each labs adopted different treatment. So for some case, limited labs conducted the study. Similar for the case `EE`.

I chose scatterplots with regression line to illustrate the relationship between two continuous variable (`uterus` and `weight`). Because the significant differece on `weight` between immature rats and adult rats, I use facet to represent different `protocol`. Notice the range for y-axis is different. We can see that: 1)For immature rats, `weight` seems do not effect the `uterus`. While for adult rats, the `uterus` decrease as the `weight` increase. There is significant interaction.

3.2 Model

I build a linear regression model with all variables except `group`. Apart from `weight` and `uterus`, I consider all other variables as factor. Also, I include the interaction between estrogen chemicals and different labs along with protocols. Notice for the factors need to use a set of dummy variables to represent, so there are 218 predictors including the intercept.

$Y = intercept + \beta_1 EE + \beta_2 ZM + \beta_3 lab + \beta_4 protocol + \beta_5 weight + \beta_6 EE : protocol + \beta_7 ZM : protocol + \beta_8 EE : lab + \beta_9 ZM : lab + \epsilon$

where $\epsilon \sim N(0, \sigma^2)$,Adjusted R-squared: 0.9538.

Notes:

1. The value of `EE` and `ZM` is discrete. It is resonable to set it as factor. And compared to the model with numeric `EE` and `ZM`, the model I chose performed better in fitting the model(Ajusted R-square 0.9538; the model include numeric `EE` and `ZM` and quadradic form can only achieve 0.8286) and get the results as I expected from the boxplot.(`ZM:labs` is significant in F-test)
2. The model does not include `group` for it does not pass the F-test. This make sense, which also prove that the study is replicable.
3. Notice that each labs adopted different schemes for the study.(See the jitter plot) So interaction would be significant among all the variables (except for `group`). However, there are 19 labs and 8 different levels for `EE`. We need to be careful when include the interaction. I adopted the rule "keep the simplest model" and make sure that each variable and interaction I include in is super significant.(with p-value for F-test less than 2.2e-16).

3.3 Methods

I use MLE/OLS to estimate the parameters. I use F-test, t-test for a single coefficient and t-test for linear combination of coefficients to answer the ques we care about. Actually, all the questions can be answer by only two plots.(See the Page 3 for details.)

a.1 Is the uterotrophic bioassay successful overal at identifying effects of EE and ZM

Using F-test.

And check all the sign for the coefficients `EE` (positive) and `ZM` (negative).

Hypothesis: $H_0 : \beta_{11} = \beta_{12} = \beta_{13} = 0$(failed to detect) v.s. $H_1 : \exists \beta_{1i} \neq 0$(detect successfully)

statistics: F-stat(using anova() in R)

$F = \frac{\|(P_k - P_{k-1})Y\|^2/(r(P_k)-r(P_{k-1}))}{\hat{\sigma}^2} \sim F(r(P_k) - r(P_{k-1}), n - p)$

Note: here the $\beta 1i$ means all the coefficients before `EE`. Similar F-test for `ZM`

a.2 Do some labs fail to detect such effects?

Constructing t-test for linear combination of coefficients.Take lab Bayer for example. The sum of coefficients before `EE` (`EE` here including `EE0.01`, `EE0.03`,...) and `EE:labBayer` is t-distrbution. All also, we need to check the sign for significant coefficients. Or, instead ,using t-test for "EE:lab_i" just pick out the labs with significant p value at "EE:lab_i" but opposite t value. ("Huntingd""Bayer""Zeneca")

Hypothesis: $H_0 : \Sigma_i(\beta_{1i} + \beta_{1i:labBayer}) = 0$(lab Bayer failed to detect) v.s. $H_1 : \Sigma_i(\beta_{1i} + \beta_{1i:labBayer}) \neq 0$(lab Bayer detect successfully)

statistics: t-stat(need construct function in R)

$t = \frac{\lambda^T \hat{\beta} - \lambda^T b_0}{\hat{\sigma}\sqrt{\lambda^T (X^T X)^{-1} \lambda}} \sim t(n - p, 0, 1)$

where $\lambda_i = 1$ for the position `EE` or `EE:labBayer`,else $\lambda_i = 0$.

a.3 The change level of dose for the effect? Does this vary across labs?

Using t-test for single coefficient and find the change point from insignificant to significant. For the first question, we concerntrate on coefficients of `EE`, `ZM`. For the second question, for each lab_i we concerntrate on coefficients of `EE:lab_i`, `ZM:lab_i`.

Hypothesis: $H_0 : \beta_{1i} = 0$ v.s. $H_1 : \beta_{1i} \neq 0$

statistics: t-stat

$t = \frac{\hat{\beta} - b_0}{\hat{\sigma}\sqrt{(X^T X)^{-1}}} \sim t(n-p, 0, 1)$

b Does the dose response vary across labs? If so, which one stands out as being different?

Using F-test to see whether the dose response vary across labs. If F-test for `EE:labs`,`ZM:labs` are significant, then vary across labs. Using t-test for single coefficient.For each lab_i we concerntrate on coefficients of `EE:lab_i`, `ZM:lab_i`. For certain $dose_j$, if some of the pvalue are especially small, those labs may stand out as being different.

c Do the protocols differ in their sensitivity to detect the effects? Is there one can be recommend?

Using F-test for the interaction `EE:protocol`, `ZM:Protocol`.If they are significant, we may say the protocols differ.

A good protocol should have following properties: 1)Robust. The variance of coefficients of `EE:protocol_i` and `ZM:protocol_i` should be small. 2)Consistent. Under the protocol, different labs should get similar result.For 1) compare variance for coefficients `protocol` and `EE:protocol`. For 2), we need to include `labs:protocol` inside the model. Or we can use boxplot to get the result.

Notes:

1. The main problem of this dataset is too much factors and interactions leading to too much predictor. If adopted Bayesian methods, it would be time consuming.(There are some improvements–adding selection process in Appendix) So I adopted MLE/OLS for it is simple and fast.
2. For overall test for discrete variable (factors), I adopted F-test. And for specific problem, I use t-test.

3.4 Outliers and Influential points

According the residuals, there are three possible "outliers"(1426,926,1586). However, I text these outliers in the boxplots. I find that points 1586,926 are outliers with $ZM = 0, EE = 0.1$. In the boxplot, they are much higher compared to all other points. So I consider it as an outlier. However, the points 1426 is with $ZM = 1, EE = 3$. It maybe the result of the interaction of `EE` and `ZM`.And not much labs conduct studies under this treatment. Only a few studies including both `EE` and `ZM`. So I keep the points 1426.

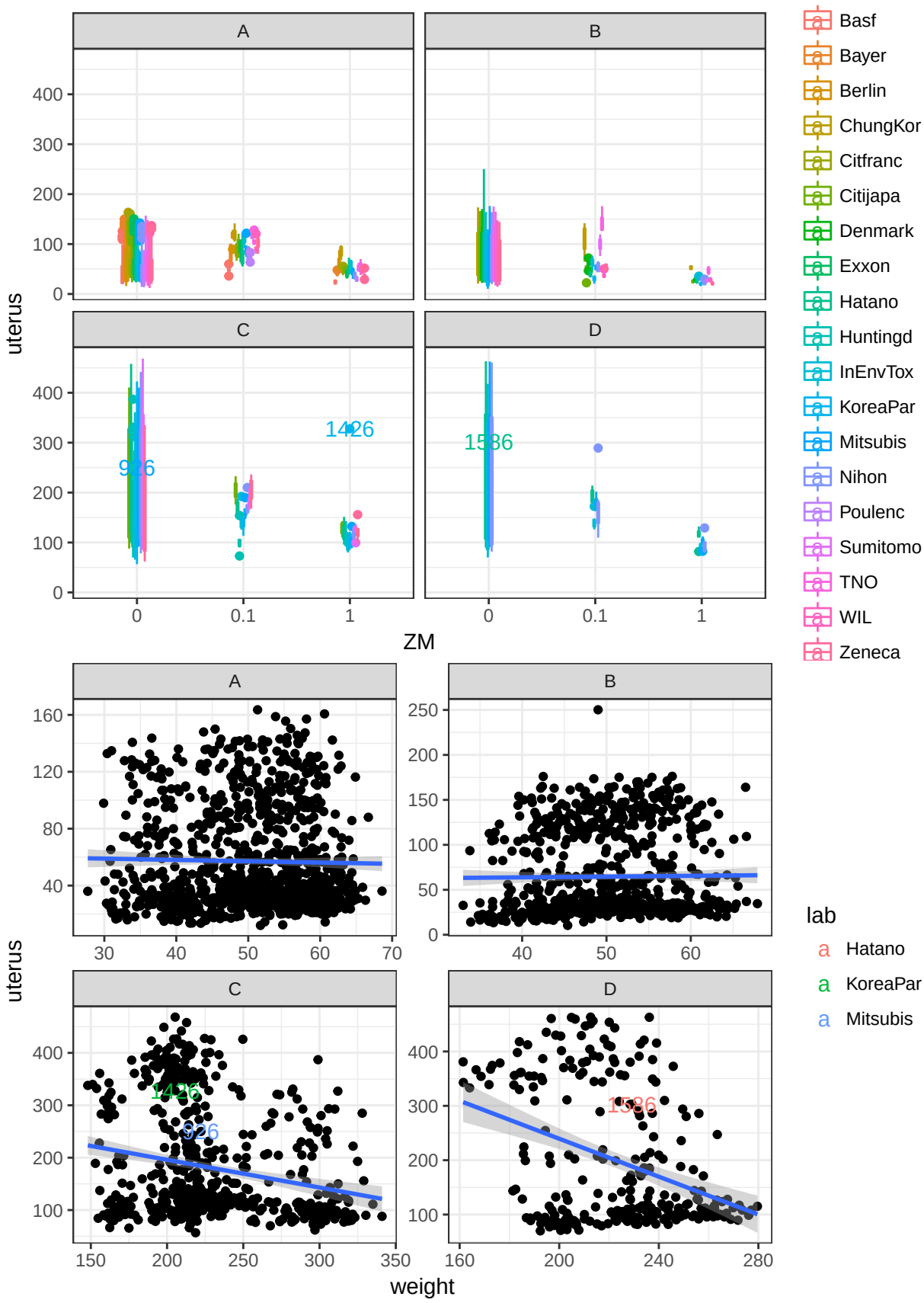
Notices: all the outliers appears in protocol C,D. This suggest us these procotols are not recommended.

4.Conclusion

Overall, uterotrophic bioassay successful overal at identifying effects of EE and ZM.The change dose for EE is 3. Dose larger than this have significant effect while less than this is not significant.The protocols differ in their sensitivity to detecting the effect. Protocol B would be recommended. Futher improvement is to reparametrize the level for protocol to compare protocol A and protocol B.

The result is not consistent among labs. First, there are some labs fail to detect such effects:Huntingd, Bayer, ChungKor, TNO, Zeneca. Second, the changing value for dose varies across labs. One of the reasons is for different labs, the dose changing points is different. For example, Sumitomo. EE0.3 may be the changing dose point. For Huntingd, EE0.1 may be the changing dose point. Third, dose reponse vary across labs. The labs Berlin and Sumitomo stands out as being different.(with pvalue<0.001 for EE:labs). The labs Bayer,Poulenc,Zeneca stands out as being different.(with pvalue<0.005 for ZM:labs)

EDA can help us answer part of the questions. The dose response vary across labs because the different shape and location of boxplot. The certain "box" behave different may be the "outlier" we want to find in b. Protocol differs in sensitivity for the length of "box" is different in 4 facets. Protocol A may be recommeded for it keep the consistence among labs. And the variance for each lab in A is smaller compared to other groups. Worth noting, I use EDA to help me deal with outliers.


A
B
C
D
uterus
ZM
weight
0
0.1
1
Basf
Bayer
Berlin
ChungKor
Citfranc
Citijapa
Denmark
Exxon
Hatano
Huntingd
InEnvTox
KoreaPar
Mitsubis
Nihon
Poulenc
Sumitomo
TNO
WIL
Zeneca
lab
Hatano
KoreaPar
Mitsubis