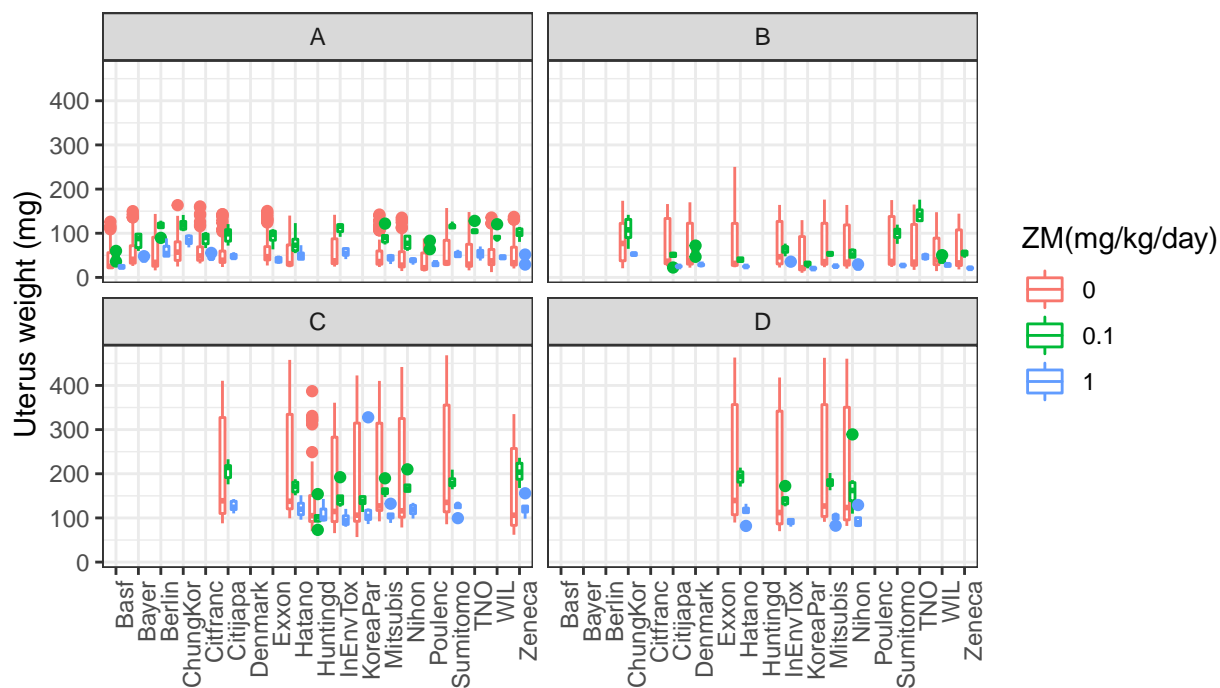# STA721 Final Project

Shuangjie Zhang, Xiyang Hu

12/8/2018
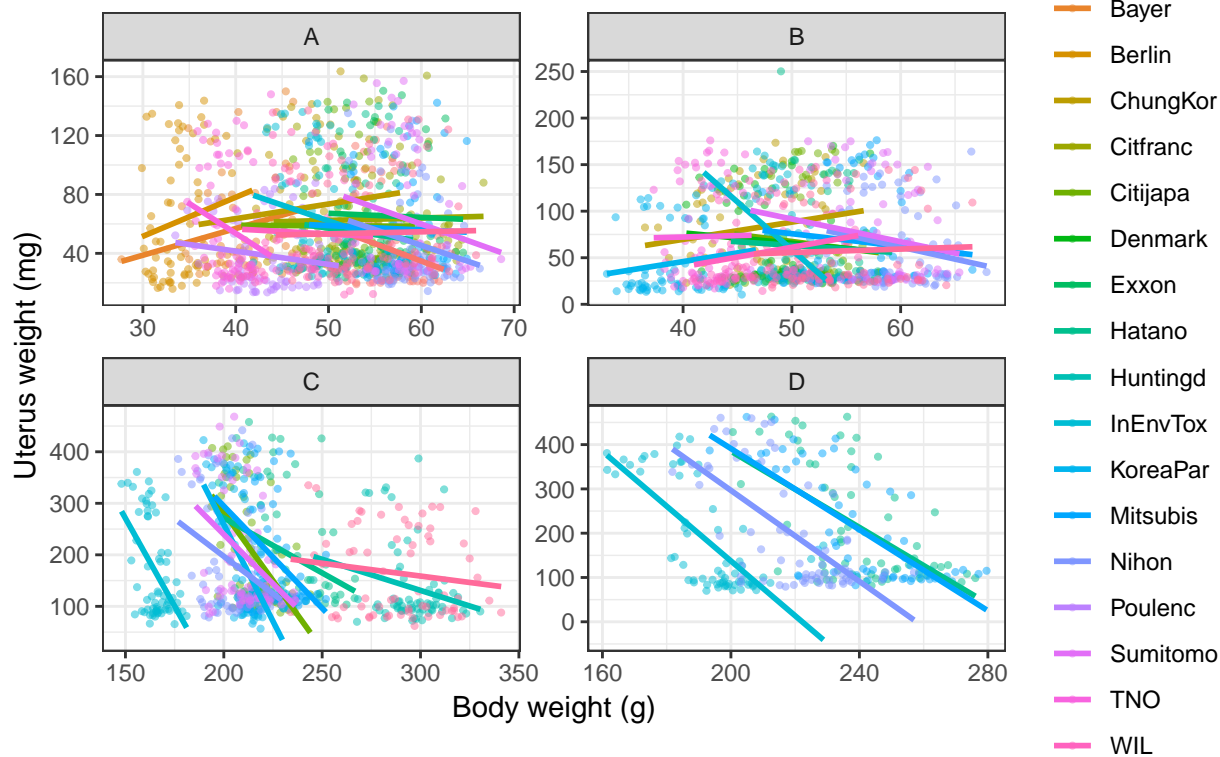
1. Summary

2. Introductions

The side−by−side boxplot of uterus weight to estrogen antagonist(ZM), facet by protocol

The side−by−side scatterplots of Uterus weight to Body weight (g), facet by protocol

Apeendix

EDA

```r
bioassay_lm = bioassay[,-7]
str(bioassay_lm)
```

```
## 'data.frame':    2677 obs. of  6 variables:
##  $ uterus  : num  21 22 21 26 24 25 22 26 24 22 ...
##  $ weight  : num  61.9 55.9 59.1 54.8 57.5 57.6 60.3 59 59.1 61.4 ...
##  $ protocol: Factor w/ 4 levels "A","B","C","D": 1 1 1 1 1 1 1 1 1 1 ...
##  $ EE      : Factor w/ 8 levels "0","0.01","0.03",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ ZM      : Factor w/ 3 levels "0","0.1","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ lab     : Factor w/ 19 levels "Basf","Bayer",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```r
table(bioassay_lm$EE, bioassay_lm$ZM)
```

```
## 
##          0 0.1   1
##   0    484   0   0
##   0.01 234   0   0
##   0.03 239   0   0
##   0.1  246   0   0
##   0.3  246   0   0
##   1    246   0   0
##   3    246 245 246
##   10   245   0   0
```

```r
ggplot(data=bioassay,mapping = aes(y = uterus,x = lab,color=EE))+
  geom_boxplot()+theme_bw()+facet_wrap(~ protocol) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(x = "Lab", y="Uterus weight (mg)", title="The side-by-side boxplot of uterus weight for different
       different dose of estrogen agonist(EE), facet by protocol", caption="", colour="EE(mg/kg/day)")
```

The side—by—side boxplot of uterus weight for different labs and different dose of estrogen agonist(EE), facet by protocol
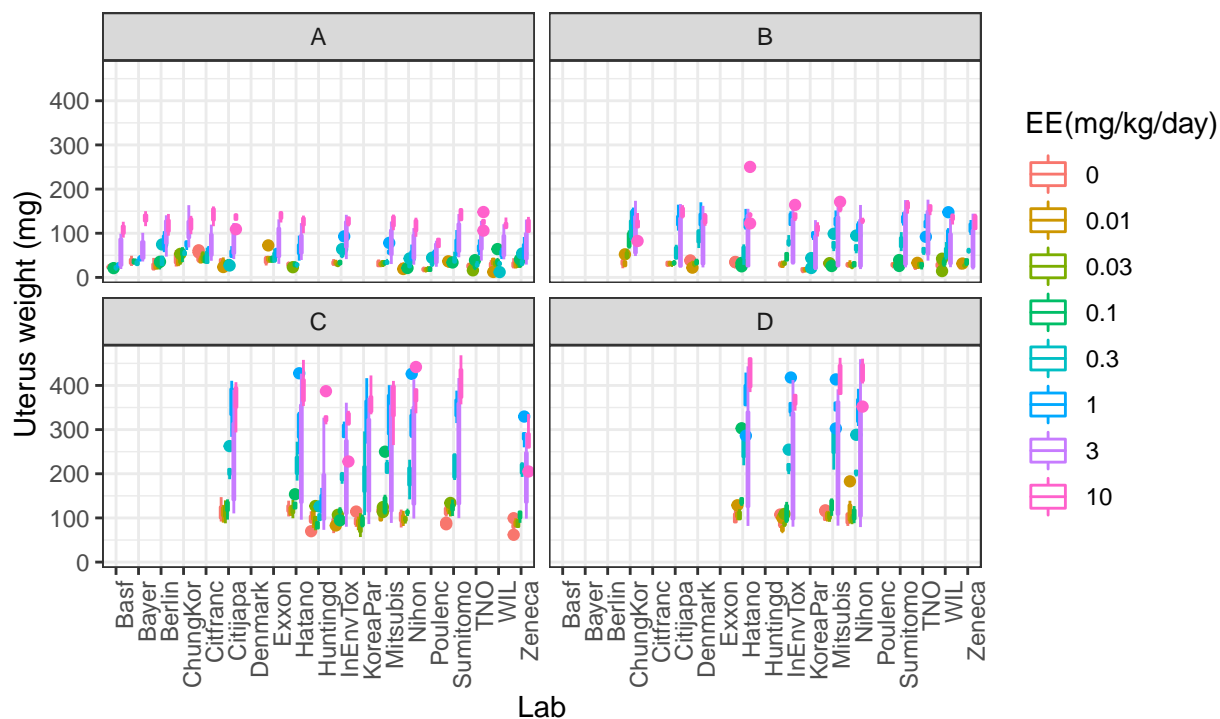
Model Part I

```
lm1 = lm(uterus~., data = bioassay_lm)
summary(lm1)
```

```
##
## Call:
## lm(formula = uterus ~ ., data = bioassay_lm)
##
## Residuals:
##       Min       1Q   Median       3Q      Max
## -107.625  -30.150    2.595   21.979  190.872
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  15.82251    6.55993   2.412 0.015933 *
## weight       -0.45365    0.05508  -8.237 2.75e-16 ***
## protocolB     7.84315    2.26416   3.464 0.000541 ***
## protocolC   207.53588    9.48173  21.888  < 2e-16 ***
## protocolD   221.22623   10.07610  21.956  < 2e-16 ***
## EE0.01       -0.60177    3.31535  -0.182 0.855982
## EE0.03        0.26008    3.28953   0.079 0.936989
## EE0.1         8.01257    3.25677   2.460 0.013946 *
## EE0.3        47.94479    3.25716  14.720  < 2e-16 ***
## EE1         106.35605    3.26542  32.570  < 2e-16 ***
## EE3         136.45891    3.27333  41.688  < 2e-16 ***
## EE10        150.55730    3.28955  45.768  < 2e-16 ***
```

4

```
## ZM0.1          -80.51563     3.75770 -21.427  < 2e-16 ***
## ZM1           -127.18576     3.75162 -33.902  < 2e-16 ***
## labBayer         2.60266     7.88343   0.330 0.741318
## labBerlin       14.84134     7.55929   1.963 0.049713 *
## labChungKor     32.46041     6.63073   4.895 1.04e-06 ***
## labCitfranc     26.21060     7.45111   3.518 0.000443 ***
## labCitijapa     21.52689     6.33273   3.399 0.000686 ***
## labDenmark      18.95727     7.80137   2.430 0.015165 *
## labExxon        23.72114     7.62673   3.110 0.001889 **
## labHatano       26.83352     6.19974   4.328 1.56e-05 ***
## labHuntingd      0.09856     8.84741   0.011 0.991112
## labInEnvTox      0.58445     6.38094   0.092 0.927028
## labKoreaPar     -2.51500     6.88744  -0.365 0.715023
## labMitsubis     24.63683     6.19749   3.975 7.22e-05 ***
## labNihon        13.18893     6.20345   2.126 0.033590 *
## labPoulenc      -4.14169     7.49225  -0.553 0.580450
## labSumitomo     28.52520     6.32781   4.508 6.83e-06 ***
## labTNO          16.56429     6.68045   2.480 0.013218 *
## labWIL          10.05022     6.63237   1.515 0.129809
## labZeneca       17.93047     6.42998   2.789 0.005332 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 41.56 on 2645 degrees of freedom
## Multiple R-squared:  0.802,  Adjusted R-squared:  0.7997
## F-statistic: 345.7 on 31 and 2645 DF,  p-value: < 2.2e-16
```

```
step(lm1, k=log(2677))
```

```
## Start:  AIC=20175.57
## uterus ~ weight + protocol + EE + ZM + lab
##
##            Df Sum of Sq      RSS   AIC
## <none>                   4568714 20176
## - lab      18    304839  4873553 20206
## - weight    1    117187  4685901 20236
## - protocol  3    855660  5424374 20612
## - ZM        2   2030817  6599531 21144
## - EE        7   7683826 12252540 22761
##
##
## Call:
## lm(formula = uterus ~ weight + protocol + EE + ZM + lab, data = bioassay_lm)
##
## Coefficients:
## (Intercept)       weight     protocolB     protocolC     protocolD
##    15.82251     -0.45365       7.84315     207.53588     221.22623
##       EE0.01        EE0.03         EE0.1         EE0.3           EE1
##     -0.60177       0.26008       8.01257      47.94479     106.35605
##          EE3          EE10         ZM0.1           ZM1       labBayer
##    136.45891     150.55730     -80.51563    -127.18576       2.60266
##     labBerlin    labChungKor    labCitfranc   labCitijapa     labDenmark
##      14.84134      32.46041      26.21060      21.52689      18.95727
##       labExxon     labHatano    labHuntingd   labInEnvTox    labKoreaPar
##      23.72114      26.83352       0.09856       0.58445      -2.51500
```

```
## labMitsubis      labNihon   labPoulenc  labSumitomo        labTNO
##     24.63683      13.18893    -4.14169     28.52520     16.56429
##       labWIL     labZeneca
##     10.05022      17.93047
```

```r
library(MASS)
```

```
##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select
```

```r
box =boxcox(lm1)
```
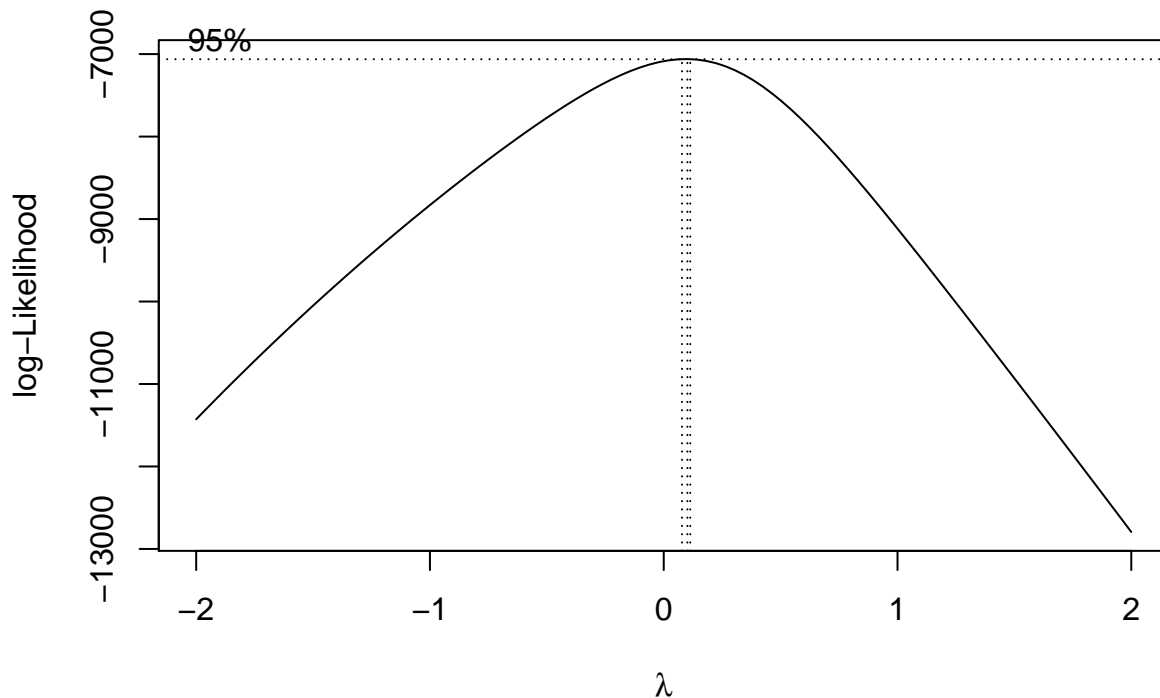


```r
lm2 = lm(formula = log(uterus) ~ log(weight) + protocol + EE + ZM + lab, data = bioassay_lm)
summary(lm2)
```

```
##
## Call:
## lm(formula = log(uterus) ~ log(weight) + protocol + EE + ZM +
##     lab, data = bioassay_lm)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.38682 -0.16223  0.01173  0.16583  1.21149
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.321144   0.239602   5.514 3.85e-08 ***
## log(weight) 0.410375   0.058372   7.030 2.61e-12 ***
## protocolB   0.055004   0.014890   3.694 0.000225 ***
## protocolC   0.689389   0.083618   8.244 2.58e-16 ***
```

```
## protocolD     0.675012    0.085999    7.849 6.03e-15 ***
## EE0.01         0.002747    0.021776    0.126 0.899614
## EE0.03         0.012017    0.021606    0.556 0.578144
## EE0.1          0.131728    0.021388    6.159 8.44e-10 ***
## EE0.3          0.551201    0.021396   25.762  < 2e-16 ***
## EE1            1.061739    0.021438   49.527  < 2e-16 ***
## EE3            1.352063    0.021477   62.954  < 2e-16 ***
## EE10           1.451617    0.021571   67.294  < 2e-16 ***
## ZM0.1         -0.531428    0.024670  -21.541  < 2e-16 ***
## ZM1           -1.145005    0.024622  -46.504  < 2e-16 ***
## labBayer       0.457789    0.055245    8.286  < 2e-16 ***
## labBerlin      0.610866    0.057515   10.621  < 2e-16 ***
## labChungKor    0.762522    0.044567   17.109  < 2e-16 ***
## labCitfranc    0.559241    0.048950   11.425  < 2e-16 ***
## labCitijapa    0.445367    0.041829   10.647  < 2e-16 ***
## labDenmark     0.506016    0.051920    9.746  < 2e-16 ***
## labExxon       0.517897    0.050097   10.338  < 2e-16 ***
## labHatano      0.383445    0.040728    9.415  < 2e-16 ***
## labHuntingd   -0.053061    0.053724   -0.988 0.323416
## labInEnvTox    0.455932    0.042501   10.728  < 2e-16 ***
## labKoreaPar    0.178793    0.046846    3.817 0.000138 ***
## labMitsubis    0.392970    0.040688    9.658  < 2e-16 ***
## labNihon       0.307153    0.040710    7.545 6.19e-14 ***
## labPoulenc     0.097765    0.051708    1.891 0.058768 .
## labSumitomo    0.464850    0.041551   11.187  < 2e-16 ***
## labTNO         0.533952    0.047525   11.235  < 2e-16 ***
## labWIL         0.380265    0.044646    8.517  < 2e-16 ***
## labZeneca      0.251125    0.041670    6.027 1.91e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.273 on 2645 degrees of freedom
## Multiple R-squared:  0.8937, Adjusted R-squared:  0.8925
## F-statistic: 717.6 on 31 and 2645 DF,  p-value: < 2.2e-16
```

```r
par(mfrow=c(2,2))
plot(lm1)
```
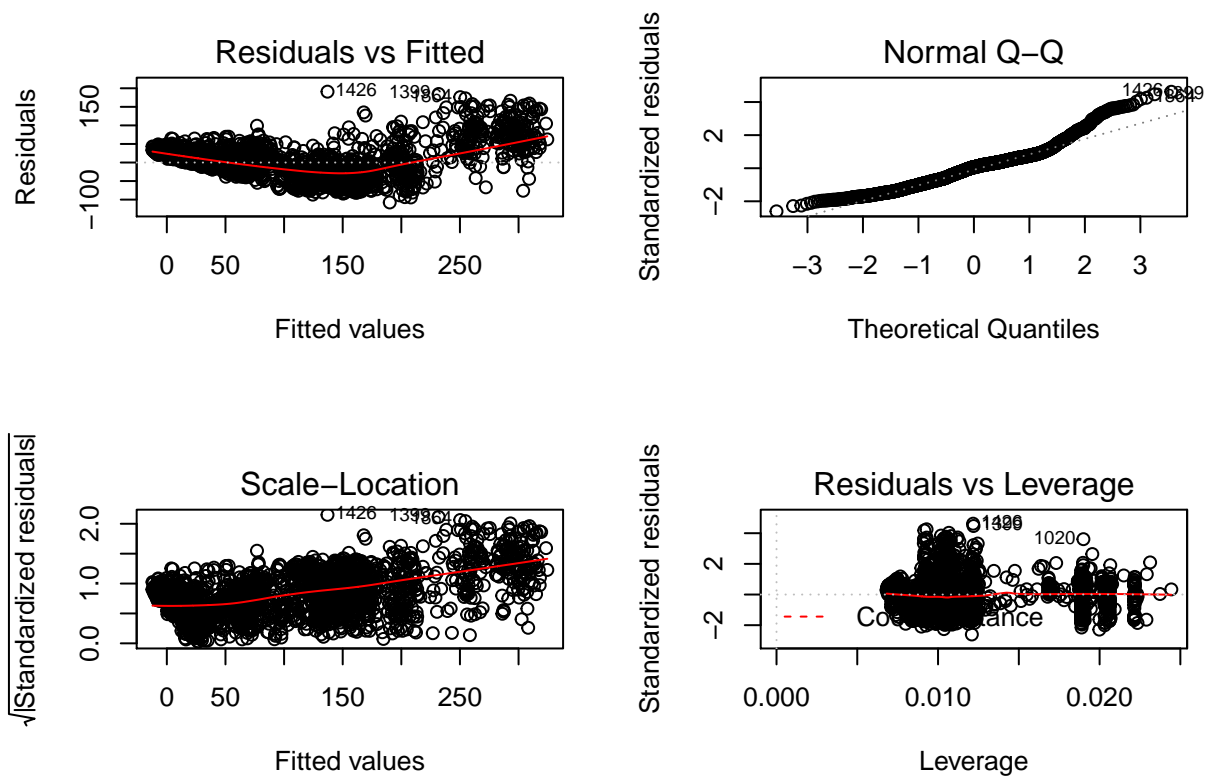
Frequentist Random Effect Model:

```r
library(lme4)
```

```
## Loading required package: Matrix
```

```r
randomeffect = lmer(log(uterus) ~ log(weight) + protocol + EE + ZM + (1+EE+ZM|lab), data = bioassay_lm)
```

```
## Warning in commonArgs(par, fn, control, environment()): maxfun < 10 *
## length(par)^2 is not recommended.
```

```
## Warning in optwrap(optimizer, devfun, getStart(start, rho$lower, rho$pp), :
## convergence code 1 from bobyqa: bobyqa -- maximum number of function
## evaluations exceeded
```

```
## singular fit
```

```r
summary(randomeffect)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: log(uterus) ~ log(weight) + protocol + EE + ZM + (1 + EE + ZM |
##     lab)
##    Data: bioassay_lm
##
## REML criterion at convergence: 74.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.4143 -0.5782  0.0019  0.5717  5.4534
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
```

8

```
##  lab        (Intercept) 0.038505 0.19623
##             EE0.01      0.002410 0.04909  -0.13
##             EE0.03      0.003451 0.05874   0.52 -0.11
##             EE0.1       0.015031 0.12260   0.39 -0.89  0.55
##             EE0.3       0.094878 0.30802  -0.21 -0.76 -0.02  0.62
##             EE1         0.170539 0.41296  -0.34 -0.58 -0.11  0.41  0.94
##             EE3         0.045229 0.21267  -0.45 -0.26 -0.12  0.14  0.78
##             EE10        0.015601 0.12491  -0.62  0.27 -0.33 -0.41  0.35
##             ZM0.1       0.108467 0.32934   0.34  0.34  0.88  0.11 -0.28
##             ZM1         0.058896 0.24268   0.35  0.25  0.80  0.15 -0.52
##  Residual              0.053099 0.23043
##
##
##
##
##
##
##
##   0.94
##   0.62  0.84
##  -0.30 -0.16 -0.13
##  -0.53 -0.44 -0.40  0.80
##
## Number of obs: 2677, groups:  lab, 19
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  1.62766    0.19575   8.315
## log(weight)  0.43491    0.04834   8.998
## protocolB    0.05597    0.01250   4.478
## protocolC    0.65373    0.06946   9.412
## protocolD    0.64037    0.07140   8.969
## EE0.01       0.02789    0.02172   1.284
## EE0.03       0.02528    0.02277   1.110
## EE0.1        0.11530    0.03376   3.416
## EE0.3        0.46623    0.07330   6.361
## EE1          0.95558    0.09684   9.868
## EE3          1.30993    0.05233  25.033
## EE10         1.43968    0.03423  42.062
## ZM0.1       -0.45946    0.07888  -5.825
## ZM1         -1.07974    0.05999 -17.999
##
## Correlation matrix not shown by default, as p = 14 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
##
## convergence code: 1
## singular fit
## maxfun < 10 * length(par)^2 is not recommended.
```

Model Part II

Model Part III