

# Final Project

Ryan Chiang

2024-05-07

select for model correctness BIC

#they use seconds moving and bodily contact time to measure activity and affiliation

```
#install.packages("Matrix")
#install.packages("MCMCglmm")
library(Matrix)
#remove.packages("Matrix")
#install.packages("/Users/apple/Downloads/Matrix_1.2-7.1.tar", repos = "NULL", type = "source")
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
library(rptR)
```

```
#library(lme4)
```

```
library(MCMCglmm)
```

```
## Loading required package: coda
```

```
## Loading required package: ape
```

```
##
```

```
## Attaching package: 'ape'
```

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

```
##
```

```
##      where
```

```
library(MPV)
```

```
## Loading required package: lattice

## Loading required package: KernSmooth

## KernSmooth 2.23 loaded
## Copyright M. P. Wand 1997-2009

## Loading required package: randomForest

## randomForest 4.7-1.1

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':
##
##     margin

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

```
library(leaps)
library(MASS)
```

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

## The following object is masked from 'package:MPV':
##
##     cement

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

```
wasp <- read.csv("../final_project_STA101/wasp.csv")
```

<https://cran.r-project.org/web/packages/rptR/vignettes/rptR.html>

Repeatability is an important measurement when conducting and analyzing research. Finish doing the repeatability values later.

```
#summary(wasp)
```

```
###trying to use lmer function
```

```
library(lme4)
```

```
lmer_fxn_trial <- lmer(seconds_moving ~ Trial + weight + (1|Wasp.ID), data = wasp)
```

```
summary(lmer_fxn_trial)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: seconds_moving ~ Trial + weight + (1 | Wasp.ID)
## Data: wasp
##
## REML criterion at convergence: 3811.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.1901 -0.5627 -0.1258  0.6035  2.4349
##
## Random effects:
## Groups Name Variance Std.Dev.
## Wasp.ID (Intercept) 12179 110.4
## Residual 18399 135.6
## Number of obs: 296, groups: Wasp.ID, 74
##
## Fixed effects:
## Estimate Std. Error t value
## (Intercept) 155.021 115.789 1.339
## Trial 24.520 7.052 3.477
## weight 95.618 749.541 0.128
##
## Correlation of Fixed Effects:
## (Intr) Trial
## Trial -0.152
## weight -0.980 0.000
```

model selection

article and important quote: <https://stats.stackexchange.com/questions/231080/can-mcmcglmm-in-r-be-used-similarly-to-lmer>

“MCMCglmm and lmer are both functions that can be used for fitting linear mixed models. MCMCglmm takes a Bayesian approach where priors must be specified for fixed and random effects, enabling inference via Markov Chain Monte Carlo sampling, whereas lmer takes a likelihood approach within the frequentist paradigm.”

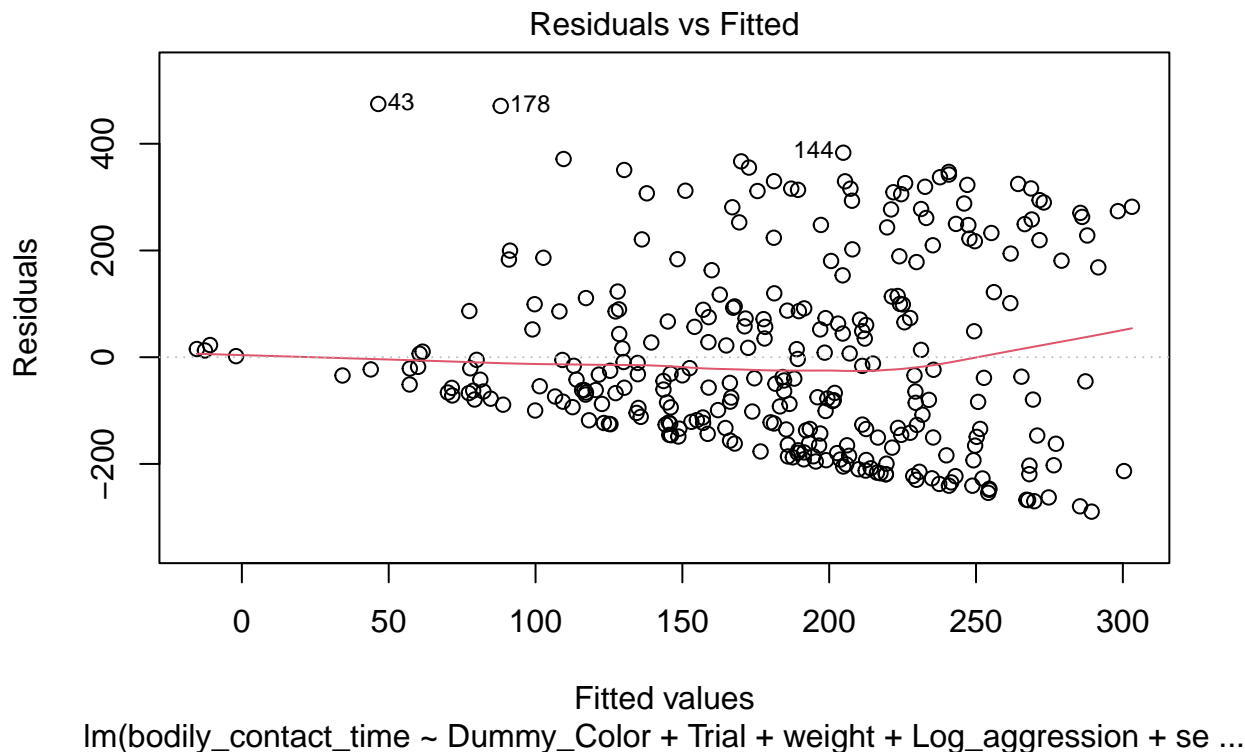
include other predictor variables??? potentially take out + Log\_aggression + scale(seconds\_moving) + anntenation... ask.

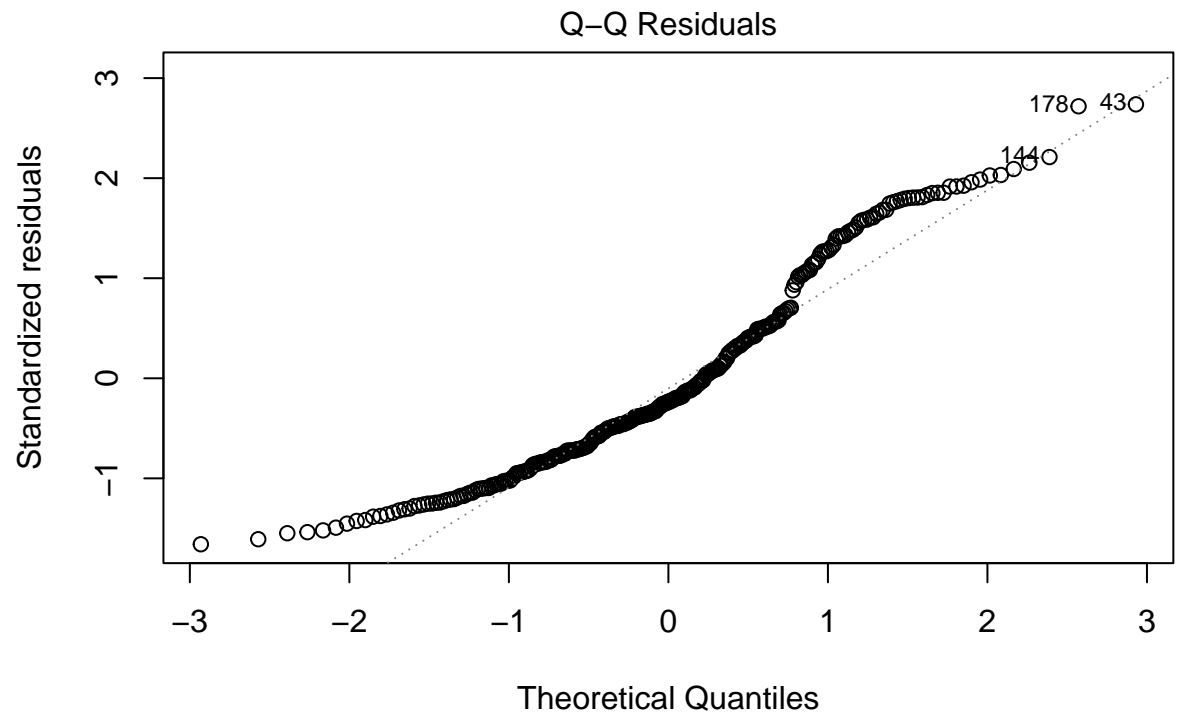
```
lm_model_affiliation <- lm(bodily_contact_time ~ Dummy_Color + Trial + weight + Log_aggression + seconds_moving + chambers_entered, data = wasp)
lm_model_affiliation
```

```
##
## Call:
## lm(formula = bodily_contact_time ~ Dummy_Color + Trial + weight +
##     Log_aggression + seconds_moving + annatenation + chambers_entered,
##     data = wasp)
##
## Coefficients:
##      (Intercept)  Dummy_Colorgoldsilver  Dummy_Colorgreen
##             155.2345             52.2334             28.2654
##      Dummy_Colorred  Dummy_Colorsilver  Dummy_Colorwhite
##             -107.7908             6.1998             37.4223
##      Dummy_Coloryellow  Trial  weight
##             -26.5071      21.3410     291.4209
##      Log_aggression  seconds_moving  annatenation
##             -12.7849      -0.2476      -3.7196
##      chambers_entered
##             2.3193
```

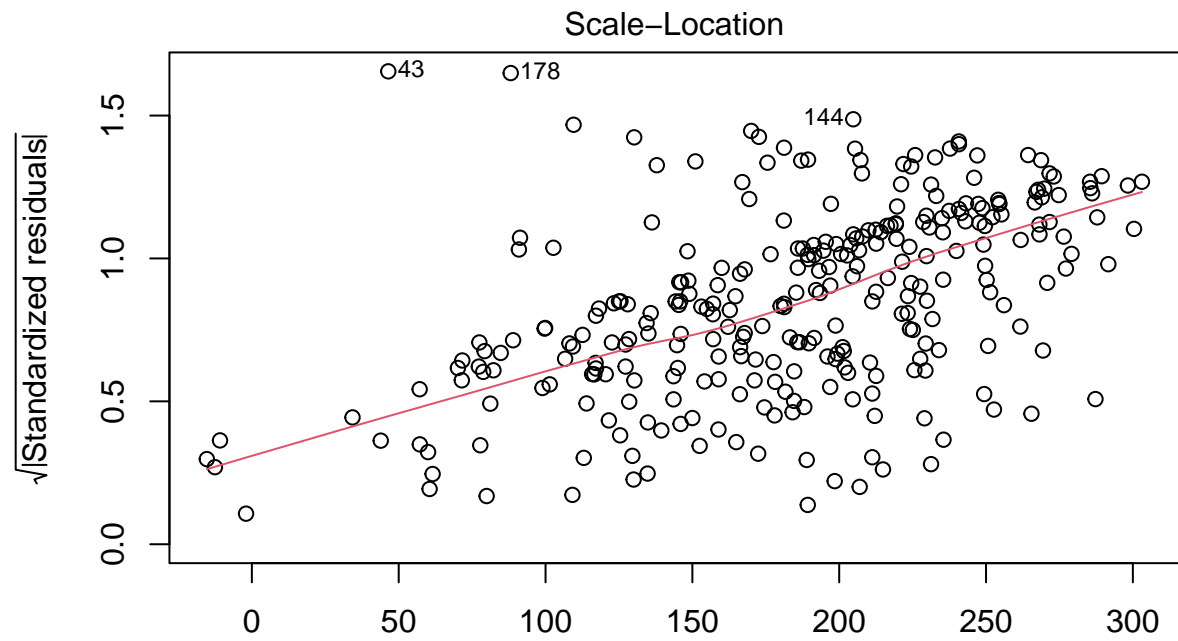
```
lm_empty_affiliation_model <- lm(bodily_contact_time ~ 1, data = wasp)
```

```
plot(lm_model_affiliation)
```

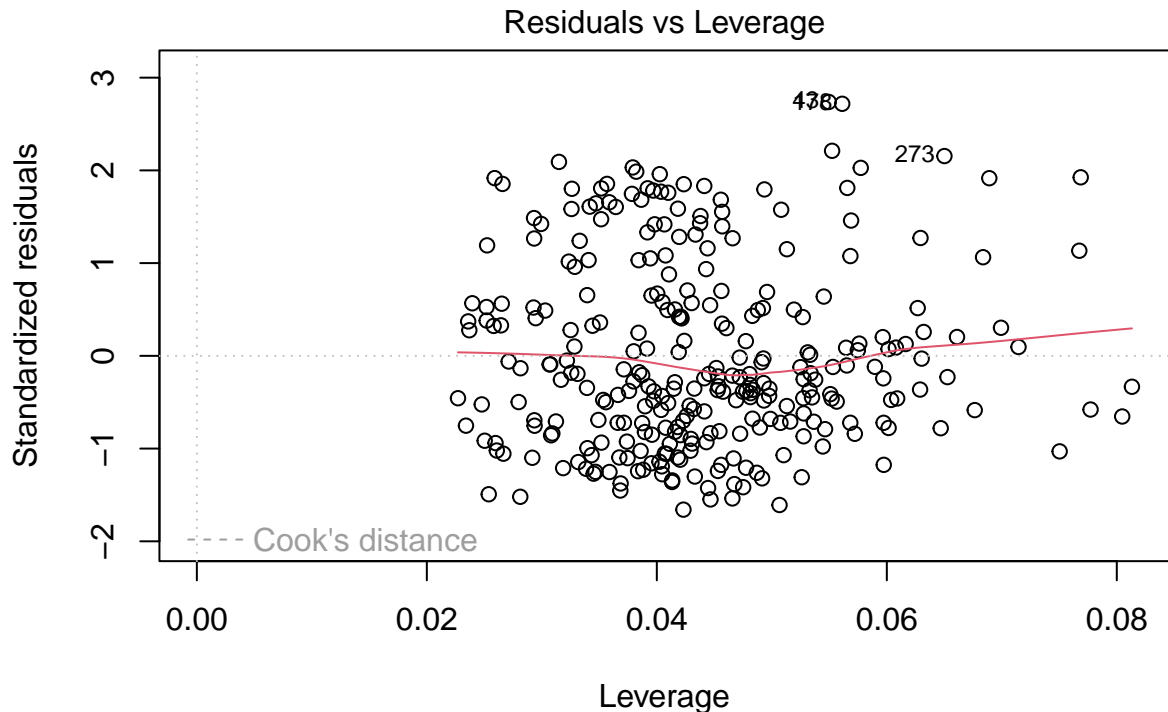




lm(bodily\_contact\_time ~ Dummy\_Color + Trial + weight + Log\_aggression + se ...



Fitted values  
`lm(bodily_contact_time ~ Dummy_Color + Trial + weight + Log_aggression + se ...`



lm(bodily\_contact\_time ~ Dummy\_Color + Trial + weight + Log\_aggression + se ...)

```
n = nrow(wasp)
```

```
aff_backward.model.BIC = stepAIC(lm_model_affiliation, scope = list(lower = lm_empty_affiliation_model
aff_forward.model.BIC = stepAIC(lm_empty_affiliation_model, scope = list(lower = lm_empty_affiliation_model
```

```
## Start: AIC=3096.71
## bodily_contact_time ~ 1
##
##           Df Sum of Sq    RSS   AIC
## + seconds_moving  1    291403  9858807 3093.8
## <none>                        10150210 3096.7
## + chambers_entered  1    183378  9966832 3097.0
## + Trial              1    177237  9972973 3097.2
## + anntenation       1     43558 10106652 3101.1
## + Log_aggression    1     21167 10129043 3101.8
## + weight            1      8887 10141323 3102.1
## + Dummy_Color       6    556433  9593776 3114.2
##
## Step: AIC=3093.78
## bodily_contact_time ~ seconds_moving
##
##           Df Sum of Sq    RSS   AIC
## + Trial      1    261705  9597101 3091.5
## <none>                        9858807 3093.8
## + anntenation  1     36697  9822110 3098.4
```

```
## + weight          1      10036 9848770 3099.2
## + Log_aggression  1       5211 9853596 3099.3
## + chambers_entered 1       4165 9854642 3099.3
## + Dummy_Color     6      638176 9220630 3108.1
##
## Step: AIC=3091.51
## bodily_contact_time ~ seconds_moving + Trial
##
##              Df Sum of Sq    RSS    AIC
## <none>                9597101 3091.5
## + antennation      1     58302 9538799 3095.4
## + Log_aggression    1     14556 9582545 3096.8
## + weight            1     10214 9586887 3096.9
## + chambers_entered  1       4212 9592889 3097.1
## + Dummy_Color       6     500220 9096881 3109.8
```

```
aff_FB.model.BIC = stepAIC(lm_empty_affiliation_model, scope = list(lower = lm_empty_affiliation_model, upper = lm_model_affiliation))
aff_BF.model.BIC = stepAIC(lm_model_affiliation, scope = list(lower = lm_empty_affiliation_model, upper = lm_model_affiliation))
BIC(aff_backward.model.BIC)
```

```
## [1] 3937.21
```

```
BIC(aff_forward.model.BIC)
```

```
## [1] 3937.21
```

```
BIC(aff_FB.model.BIC)
```

```
## [1] 3937.21
```

```
BIC(aff_BF.model.BIC)
```

```
## [1] 3937.21
```

```
aff_BF.model.BIC
```

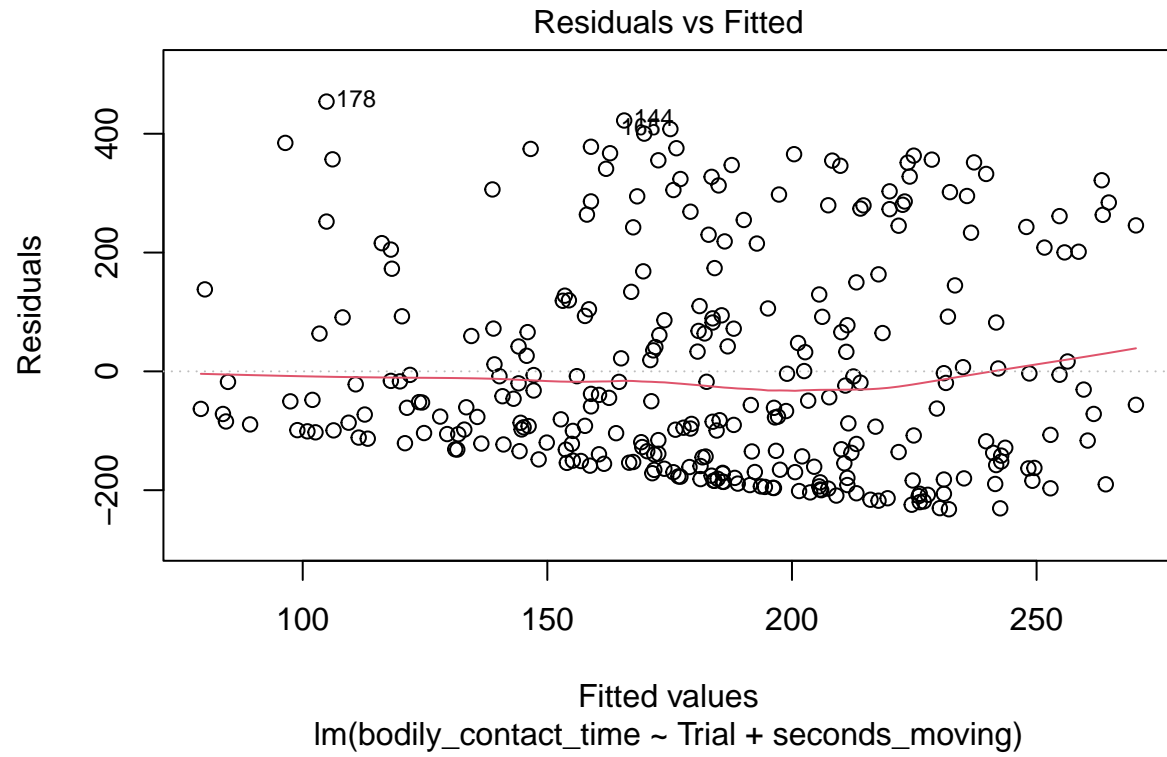
```
##
## Call:
## lm(formula = bodily_contact_time ~ Trial + seconds_moving, data = wasp)
##
## Coefficients:
##      (Intercept)          Trial  seconds_moving
##      163.6369         26.9257         -0.2055
```

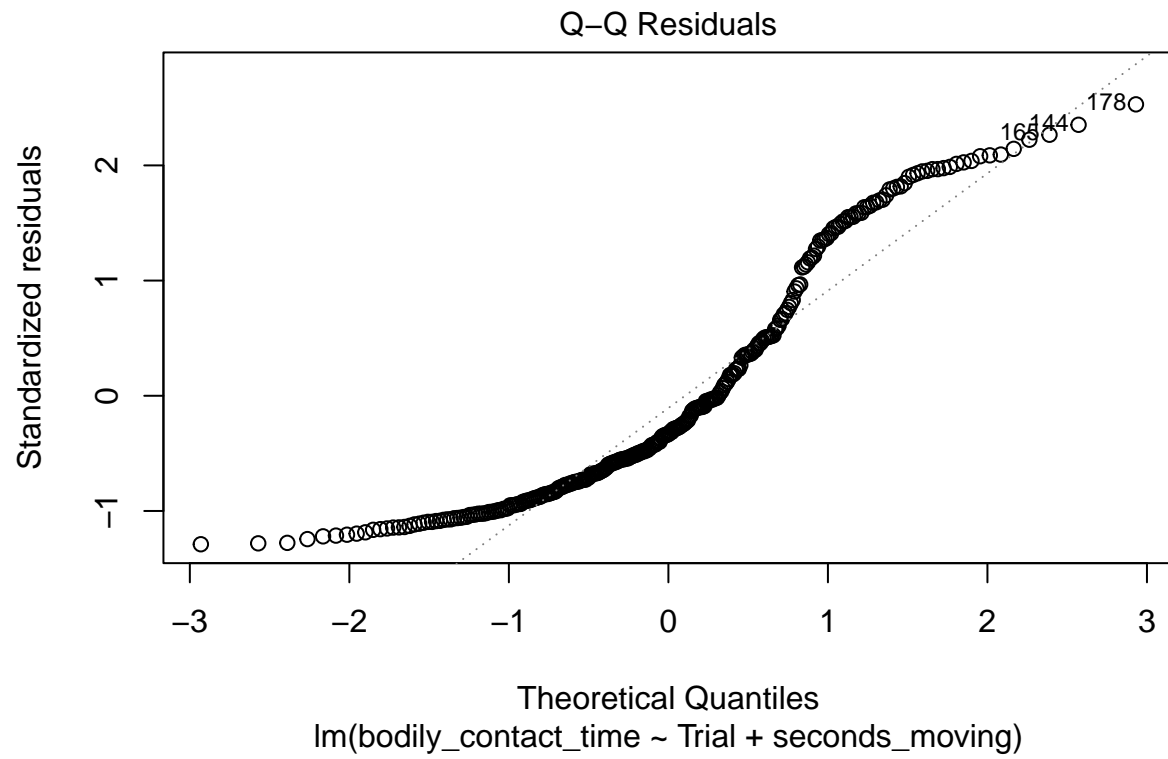
```
confint(aff_BF.model.BIC)
```

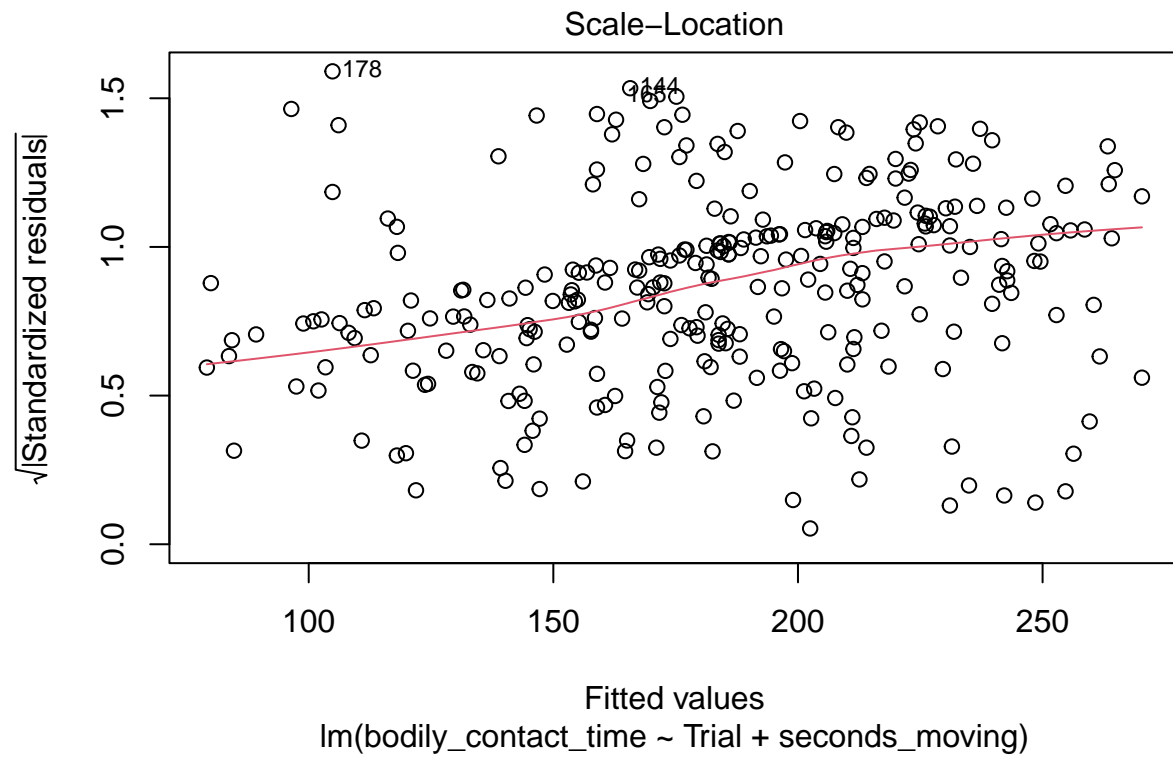
```
##              2.5 %      97.5 %
## (Intercept) 109.0359340 218.23782494
## Trial        8.1782302  45.67314211
## seconds_moving -0.3249099 -0.08611324
```

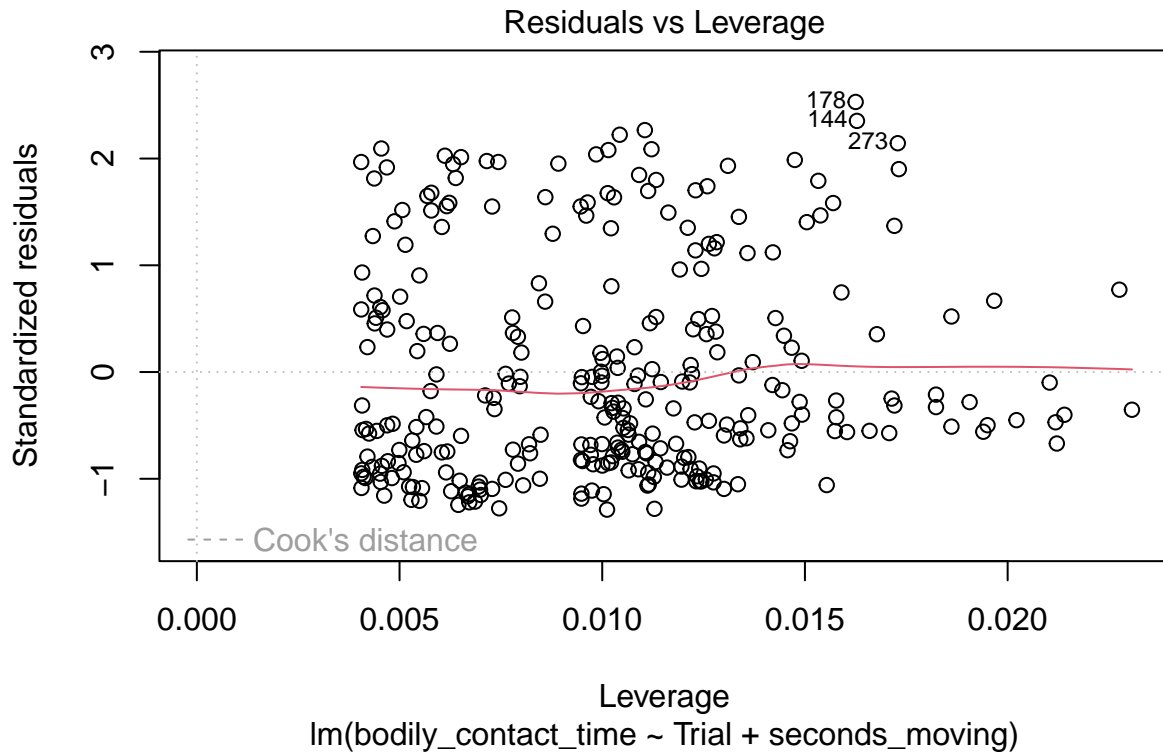


```
plot(aff_BF.model.BIC)
```









```
lmer_model_affilitation <- lmer(bodily_contact_time ~ Trial + seconds_moving + (1|Wasp.ID), data = wasp)
BIC(lmer_model_affilitation)
```

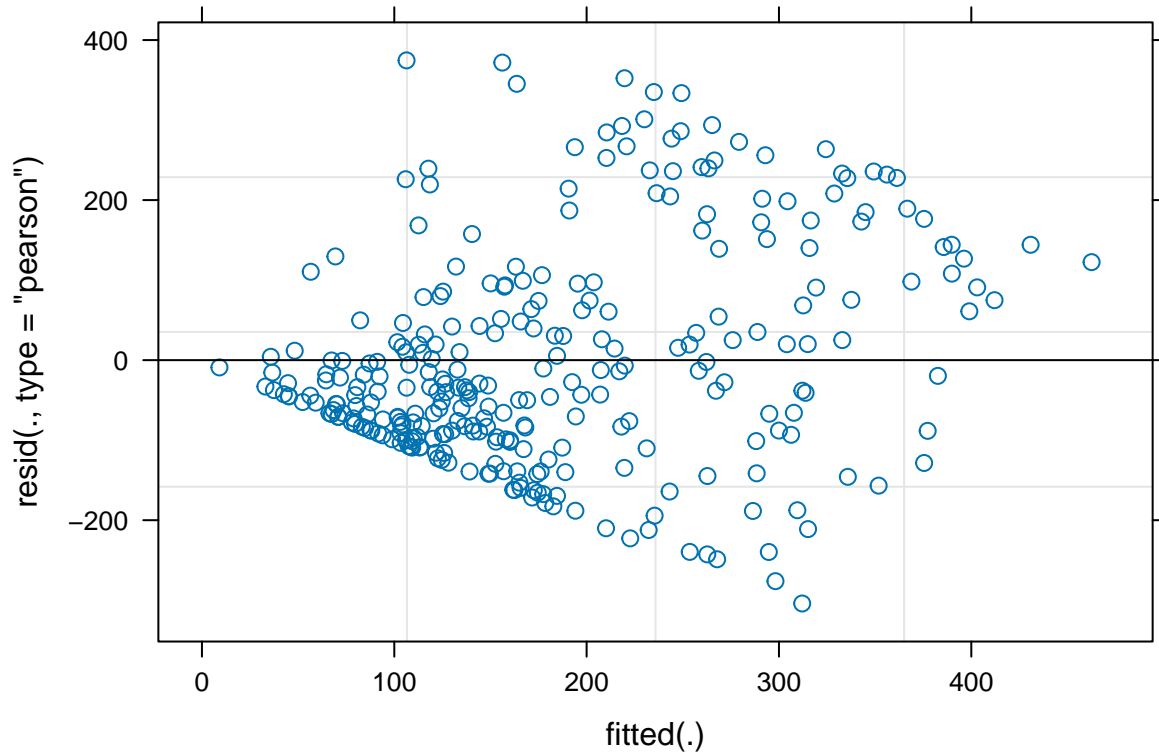
```
## [1] 3898.763
```

```
shapiro.test(resid(lmer_model_affilitation))
```

```
##
## Shapiro-Wilk normality test
##
## data: resid(lmer_model_affilitation)
## W = 0.95022, p-value = 1.777e-08
```

```
#Group.aff = rep("Lower", nrow(wasp)) #Creates a vector that repeats "Lower" n times
#Group.aff[wasp$bodily_contact_time < median(wasp$bodily_contact_time)] = "Upper" #Changing the appropriate
#Group.aff = as.factor(Group.aff) #Changes it to a factor, which R recognizes as a grouping variable.
#fligner.test(lmer_model_affilitation$residuals, Group.aff)
```

```
plot(lmer_model_affilitation)
```



```
lm_model_ant <- lm(anntenation ~ Dummy_Color + Trial + weight + Log_aggression + seconds_moving + bodily_contact_time + chambers_entered, data = wasp)
lm_model_ant
```

```
##
## Call:
## lm(formula = anntenation ~ Dummy_Color + Trial + weight + Log_aggression +
##     seconds_moving + bodily_contact_time + chambers_entered,
##     data = wasp)
##
## Coefficients:
##      (Intercept) Dummy_Colorgoldsilver Dummy_Colorgreen
##           3.842039          -0.213593           0.546385
## Dummy_Colorred   Dummy_Colorsilver   Dummy_Colorwhite
##           0.499150           0.257893           0.920762
## Dummy_Coloryellow      Trial      weight
##          -0.909314      0.340920      6.853397
## Log_aggression      seconds_moving      bodily_contact_time
##           0.681158      0.002366      -0.001398
## chambers_entered
##          -0.188734
```

```
lm_empty_ant <- lm(anntenation ~ 1, data = wasp)
```

```
ant_backward.model.BIC = stepAIC(lm_model_ant, scope = list(lower = lm_empty_ant, upper= lm_model_ant))
ant_forward.model.BIC = stepAIC(lm_empty_ant, scope = list(lower = lm_empty_ant, upper= lm_model_ant),
```

```
## Start: AIC=748.14
## anntenation ~ 1
##
##           Df Sum of Sq  RSS   AIC
## + Log_aggression    1   103.378 3532.7 745.30
## + <none>                3636.0 748.14
## + Trial              1    35.743 3600.3 750.91
## + bodily_contact_time 1    15.603 3620.4 752.56
## + weight             1     6.094 3629.9 753.34
## + seconds_moving     1     3.708 3632.3 753.53
## + chambers_entered   1     1.808 3634.2 753.69
## + Dummy_Color        6    91.455 3544.6 774.75
##
## Step: AIC=745.3
## anntenation ~ Log_aggression
##
##           Df Sum of Sq  RSS   AIC
## + <none>                3532.7 745.30
## + Trial              1    23.620 3509.0 749.00
## + bodily_contact_time 1    12.176 3520.5 749.97
## + weight             1     6.480 3526.2 750.44
## + chambers_entered   1     5.970 3526.7 750.49
## + seconds_moving     1     0.289 3532.4 750.96
## + Dummy_Color        6    69.089 3463.6 773.59
```

```
ant_FB.model.BIC = stepAIC(lm_empty_ant, scope = list(lower = lm_empty_ant, upper= lm_model_ant), k = 1)
ant_BF.model.BIC = stepAIC(lm_model_ant, scope = list(lower = lm_empty_ant, upper= lm_model_ant), k = 1)
```

```
BIC(ant_backward.model.BIC)
```

```
## [1] 1590.999
```

```
BIC(ant_forward.model.BIC)
```

```
## [1] 1590.999
```

```
BIC(ant_BF.model.BIC)
```

```
## [1] 1590.999
```

```
BIC(ant_FB.model.BIC)
```

```
## [1] 1590.999
```

```
ant_BF.model.BIC
```

```
##  
## Call:  
## lm(formula = anntenation ~ Log_aggression, data = wasp)  
##  
## Coefficients:  
##      (Intercept)  Log_aggression  
##          5.1270          0.7365
```

```
lmer_model_ant <- lmer(anntenation ~ Log_aggression + (1|Wasp.ID), data = wasp)
```

```
shapiro.test(resid(lmer_model_ant))
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  resid(lmer_model_ant)  
## W = 0.9772, p-value = 0.0001166
```

try fligner test for this

###mixed effect model better.

```
BIC(lmer_model_ant)
```

```
## [1] 1555.338
```

```
lm_model_agg <- lm(Log_aggression ~ Dummy_Color + Trial + weight + bodily_contact_time + seconds_moving  
lm_model_agg
```

```
##  
## Call:  
## lm(formula = Log_aggression ~ Dummy_Color + Trial + weight +  
##      bodily_contact_time + seconds_moving + anntenation + chambers_entered,  
##      data = wasp)  
##  
## Coefficients:  
##      (Intercept)  Dummy_Colorgolddsilver  Dummy_Colorgreen  
##          0.3166691          0.5499208          -0.0585080  
##      Dummy_Colorred  Dummy_Colorsilver  Dummy_Colorwhite  
##          -0.0641304          0.5887456          -0.1366981  
##      Dummy_Coloryellow  Trial  weight  
##          -0.1920834          -0.0390310          0.5698611  
##      bodily_contact_time  seconds_moving  anntenation  
##          -0.0002190          0.0007041          0.0310535  
##      chambers_entered  
##          -0.0144865
```

```
lm_empty_agg <- lm(Log_aggression ~ 1, data = wasp)
```

```
agg_backward.model.BIC = stepAIC(lm_model_agg, scope = list(lower = lm_empty_agg, upper= lm_model_agg))
agg_forward.model.BIC = stepAIC(lm_empty_agg, scope = list(lower = lm_empty_agg, upper= lm_model_agg),
```

```
## Start: AIC=-124.63
## Log_aggression ~ 1
##
##               Df Sum of Sq   RSS   AIC
## + Dummy_Color    6   29.3719 161.21 -140.03
## + anntenation    1    5.4185 185.16 -127.48
## <none>                                190.58 -124.63
## + seconds_moving  1    3.5797 187.00 -124.55
## + Trial            1    2.4368 188.15 -122.75
## + chambers_entered 1    2.1703 188.41 -122.33
## + bodily_contact_time 1    0.3974 190.18 -119.56
## + weight          1    0.0109 190.57 -118.96
##
## Step: AIC=-140.03
## Log_aggression ~ Dummy_Color
##
##               Df Sum of Sq   RSS   AIC
## + anntenation    1    3.6845 157.53 -141.19
## <none>                                161.21 -140.03
## + seconds_moving  1    2.3637 158.85 -138.71
## + bodily_contact_time 1    1.3528 159.86 -136.84
## + chambers_entered 1    1.0123 160.20 -136.21
## + Trial            1    0.1200 161.09 -134.56
## + weight          1    0.0778 161.13 -134.49
##
## Step: AIC=-141.19
## Log_aggression ~ Dummy_Color + anntenation
##
##               Df Sum of Sq   RSS   AIC
## <none>                                157.53 -141.19
## + seconds_moving  1    2.24455 155.28 -139.74
## + chambers_entered 1    1.15309 156.37 -137.67
## + bodily_contact_time 1    1.04936 156.48 -137.47
## + Trial            1    0.26916 157.26 -136.00
## + weight          1    0.03459 157.49 -135.56
```

```
agg_FB.model.BIC = stepAIC(lm_empty_agg, scope = list(lower = lm_empty_agg, upper= lm_model_agg), k = 1)
agg_BF.model.BIC = stepAIC(lm_model_agg, scope = list(lower = lm_empty_agg, upper= lm_model_agg), k = 1)
```

```
BIC(agg_backward.model.BIC)
```

```
## [1] 704.5164
```

```
BIC(agg_forward.model.BIC)
```

```
## [1] 704.5164
```



```
BIC(agg_BF.model.BIC)
```

```
## [1] 704.5164
```

```
BIC(agg_FB.model.BIC)
```

```
## [1] 704.5164
```

```
agg_BF.model.BIC
```

```
##  
## Call:  
## lm(formula = Log_aggression ~ Dummy_Color + anntenation, data = wasp)  
##  
## Coefficients:  
##      (Intercept) Dummy_Colorgoldsilver Dummy_Colorgreen  
##           0.33147           0.53597           -0.02882  
##      Dummy_Colorred      Dummy_Colorsilver      Dummy_Colorwhite  
##          -0.02245           0.59191           -0.09271  
##      Dummy_Coloryellow      anntenation  
##          -0.15931           0.03224
```

```
lmer_model_agg <- lmer(Log_aggression ~ Dummy_Color + anntenation + (1|Wasp.ID), data = wasp)  
lmer_model_agg
```

```
## Linear mixed model fit by REML ['lmerMod']  
## Formula: Log_aggression ~ Dummy_Color + anntenation + (1 | Wasp.ID)  
## Data: wasp  
## REML criterion at convergence: 668.824  
## Random effects:  
## Groups Name Std.Dev.  
## Wasp.ID (Intercept) 0.2920  
## Residual 0.6798  
## Number of obs: 296, groups: Wasp.ID, 74  
## Fixed Effects:  
##      (Intercept) Dummy_Colorgoldsilver Dummy_Colorgreen  
##           0.318735           0.527531           -0.009458  
##      Dummy_Colorred      Dummy_Colorsilver      Dummy_Colorwhite  
##           0.027020           0.584500           -0.046199  
##      Dummy_Coloryellow      anntenation  
##          -0.125362           0.032218
```

```
#Mixed effect model not better
```

```
BIC(lmer_model_agg)
```

```
## [1] 725.7275
```

```
lm_model_exp <- lm(chambers_entered ~ Dummy_Color + Trial + weight + Log_aggression + seconds_moving +
lm_model_exp
```

```
##
## Call:
## lm(formula = chambers_entered ~ Dummy_Color + Trial + weight +
##     Log_aggression + seconds_moving + bodily_contact_time + anntenation,
##     data = wasp)
##
## Coefficients:
##             (Intercept) Dummy_Colorgoldsilver      Dummy_Colorgreen
##                   2.587045              0.054953              0.127885
##      Dummy_Colorred      Dummy_Colorsilver      Dummy_Colorwhite
##                   0.112001              0.222856             -0.474804
##      Dummy_Coloryellow              Trial              weight
##                   -0.240149             -0.041701             -3.678774
##      Log_aggression      seconds_moving      bodily_contact_time
##                   -0.070734              0.015302              0.000194
##      anntenation
##                   -0.042012
```

```
lm_empty_exp <- lm(chambers_entered ~ 1, data = wasp)
```

```
exp_backward.model.BIC = stepAIC(lm_model_exp, scope = list(lower = lm_empty_exp, upper= lm_model_exp)
exp_forward.model.BIC = stepAIC(lm_empty_exp, scope = list(lower = lm_empty_exp, upper= lm_model_exp),
```

```
## Start:  AIC=679.19
## chambers_entered ~ 1
##
##              Df Sum of Sq    RSS    AIC
## + seconds_moving      1    2106.82  773.61 295.75
## <none>                                2880.43 679.19
## + bodily_contact_time  1      52.04 2828.39 679.48
## + Trial                 1      51.10 2829.33 679.58
## + Log_aggression       1      32.80 2847.63 681.49
## + anntenation          1       1.43 2879.00 684.73
## + weight               1       0.93 2879.50 684.78
## + Dummy_Color          6      45.26 2835.17 708.64
##
## Step:  AIC=295.75
## chambers_entered ~ seconds_moving
##
##              Df Sum of Sq    RSS    AIC
## <none>                                773.61 295.75
## + anntenation          1      7.0960 766.51 298.71
## + weight               1      2.1437 771.46 300.62
## + bodily_contact_time  1      0.3268 773.28 301.31
## + Log_aggression       1      0.3235 773.28 301.32
## + Trial                 1      0.0004 773.60 301.44
## + Dummy_Color          6     10.1387 763.47 325.99
```

```
exp_FB.model.BIC = stepAIC(lm_empty_exp, scope = list(lower = lm_empty_exp, upper= lm_model_exp), k = 1)
exp_BF.model.BIC = stepAIC(lm_model_exp, scope = list(lower = lm_empty_exp, upper= lm_model_exp), k = 1)
```

```
BIC(exp_backward.model.BIC)
```

```
## [1] 1141.451
```

```
BIC(exp_forward.model.BIC)
```

```
## [1] 1141.451
```

```
BIC(exp_BF.model.BIC)
```

```
## [1] 1141.451
```

```
BIC(exp_FB.model.BIC)
```

```
## [1] 1141.451
```

```
exp_BF.model.BIC
```

```
##
## Call:
## lm(formula = chambers_entered ~ seconds_moving, data = wasp)
##
## Coefficients:
##      (Intercept)  seconds_moving
##           1.6986           0.0152
```

```
shapiro.test(resid(exp_BF.model.BIC))
```

```
##
## Shapiro-Wilk normality test
##
## data:  resid(exp_BF.model.BIC)
## W = 0.98843, p-value = 0.01837
```

```
Group.exp = rep("Lower",nrow(wasp)) #Creates a vector that repeats "Lower" n times
Group.exp[wasp$chambers_entered < median(wasp$chambers_entered)] = "Upper" #Changing the appropriate va
Group.exp = as.factor(Group.exp) #Changes it to a factor, which R recognizes as a grouping variable.
fligner.test(exp_BF.model.BIC$residuals, Group.exp)
```

```
##
## Fligner-Killeen test of homogeneity of variances
##
## data:  exp_BF.model.BIC$residuals and Group.exp
## Fligner-Killeen:med chi-squared = 9.9773, df = 1, p-value = 0.001585
```

```
##LMER MODEL NOT BETTER.
```

```
lmer_model_exp <- lmer(chambers_entered ~ seconds_moving + (1|Wasp.ID), data = wasp)
BIC(lmer_model_exp)
```

```
## [1] 1142.717
```

```
#install.packages("MPV")
#install.packages("leaps")
library(MPV)
library(leaps)
library(MASS)
#install.packages("lmerTest")
```

```
#lmerTest::step(lmer_model_affilitation_full)
```