

Linear Mixed effect model-day2

Chandan Kumar Pandey

2022-10-20

Mixed effect models

possum Morphometric dataset.

```
possum <- read.csv("./data/possum.csv",header = T)
## Basic data data exploration
summary(possum)
```

```
##      case      site      Pop      sex
## Min.   : 1.00   Min.   :1.000   Length:104   Length:104
## 1st Qu.: 26.75   1st Qu.:1.000   Class :character   Class :character
## Median : 52.50   Median :3.000   Mode  :character   Mode  :character
## Mean   : 52.50   Mean    :3.625
## 3rd Qu.: 78.25   3rd Qu.:6.000
## Max.   :104.00   Max.    :7.000
##
##      age      hdlngth      skullw      totlngth
## Min.   :1.000   Min.   : 82.50   Min.   :50.00   Min.   :75.00
## 1st Qu.:2.250   1st Qu.: 90.67   1st Qu.:54.98   1st Qu.:84.00
## Median :3.000   Median : 92.80   Median :56.35   Median :88.00
## Mean   :3.833   Mean    : 92.60   Mean    :56.88   Mean    :87.09
## 3rd Qu.:5.000   3rd Qu.: 94.72   3rd Qu.:58.10   3rd Qu.:90.00
## Max.   :9.000   Max.    :103.10   Max.    :68.60   Max.    :96.50
## NA's      :2
##      taill      footlngth      earconch      eye      chest
## Min.   :32.00   Min.   :60.30   Min.   :40.30   Min.   :12.80   Min.   :22.0
## 1st Qu.:35.88   1st Qu.:64.60   1st Qu.:44.80   1st Qu.:14.40   1st Qu.:25.5
## Median :37.00   Median :68.00   Median :46.80   Median :14.90   Median :27.0
## Mean   :37.01   Mean    :68.46   Mean    :48.13   Mean    :15.05   Mean    :27.0
## 3rd Qu.:38.00   3rd Qu.:72.50   3rd Qu.:52.00   3rd Qu.:15.72   3rd Qu.:28.0
## Max.   :43.00   Max.    :77.90   Max.    :56.20   Max.    :17.80   Max.    :32.0
## NA's      :1
##      belly
## Min.   :25.00
## 1st Qu.:31.00
## Median :32.50
## Mean   :32.59
## 3rd Qu.:34.12
## Max.   :40.00
```

```
##
```

```
## convert the population,site and gender into factors
```

```
possum$site <-as.factor(possum$site)
```

```
possum$sex <- as.factor(possum$sex)
```

```
possum$Pop <-as.factor(possum$Pop)
```

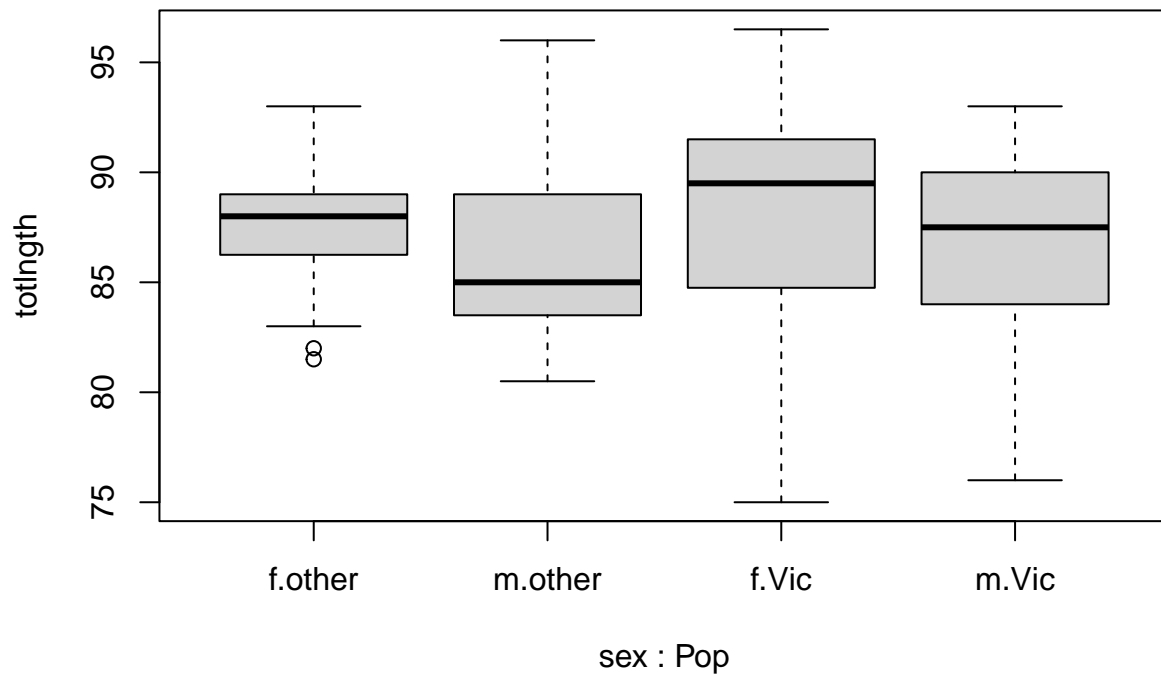
```
summary(possum)
```

```
##      case      site      Pop      sex      age      hdlngth
## Min.   : 1.00    1:33    other:58    f:43    Min.   :1.000    Min.   : 82.50
## 1st Qu.: 26.75    2:13    Vic  :46    m:61    1st Qu.:2.250    1st Qu.: 90.67
## Median : 52.50    3: 7                                Median :3.000    Median : 92.80
## Mean   : 52.50    4: 7                                Mean   :3.833    Mean   : 92.60
## 3rd Qu.: 78.25    5:13                                3rd Qu.:5.000    3rd Qu.: 94.72
## Max.   :104.00    6:13                                Max.   :9.000    Max.   :103.10
##                                     7:18                                NA's    :2
##      skullw      totlngth      taill      footlngth
## Min.   :50.00    Min.   :75.00    Min.   :32.00    Min.   :60.30
## 1st Qu.:54.98    1st Qu.:84.00    1st Qu.:35.88    1st Qu.:64.60
## Median :56.35    Median :88.00    Median :37.00    Median :68.00
## Mean   :56.88    Mean   :87.09    Mean   :37.01    Mean   :68.46
## 3rd Qu.:58.10    3rd Qu.:90.00    3rd Qu.:38.00    3rd Qu.:72.50
## Max.   :68.60    Max.   :96.50    Max.   :43.00    Max.   :77.90
##                                     NA's    :1
##      earconch      eye      chest      belly
## Min.   :40.30    Min.   :12.80    Min.   :22.0    Min.   :25.00
## 1st Qu.:44.80    1st Qu.:14.40    1st Qu.:25.5    1st Qu.:31.00
## Median :46.80    Median :14.90    Median :27.0    Median :32.50
## Mean   :48.13    Mean   :15.05    Mean   :27.0    Mean   :32.59
## 3rd Qu.:52.00    3rd Qu.:15.72    3rd Qu.:28.0    3rd Qu.:34.12
## Max.   :56.20    Max.   :17.80    Max.   :32.0    Max.   :40.00
##
```

```
## you can see the difference in the way population,site and gender columns
```

```
## box plot of lenght vs gender and pop
```

```
boxplot(totlngth~sex+Pop, data = possum)
```



```
## pair plot to see if there is co - linearity
ggpairs(possum[,c(8,7,9,10)])
```

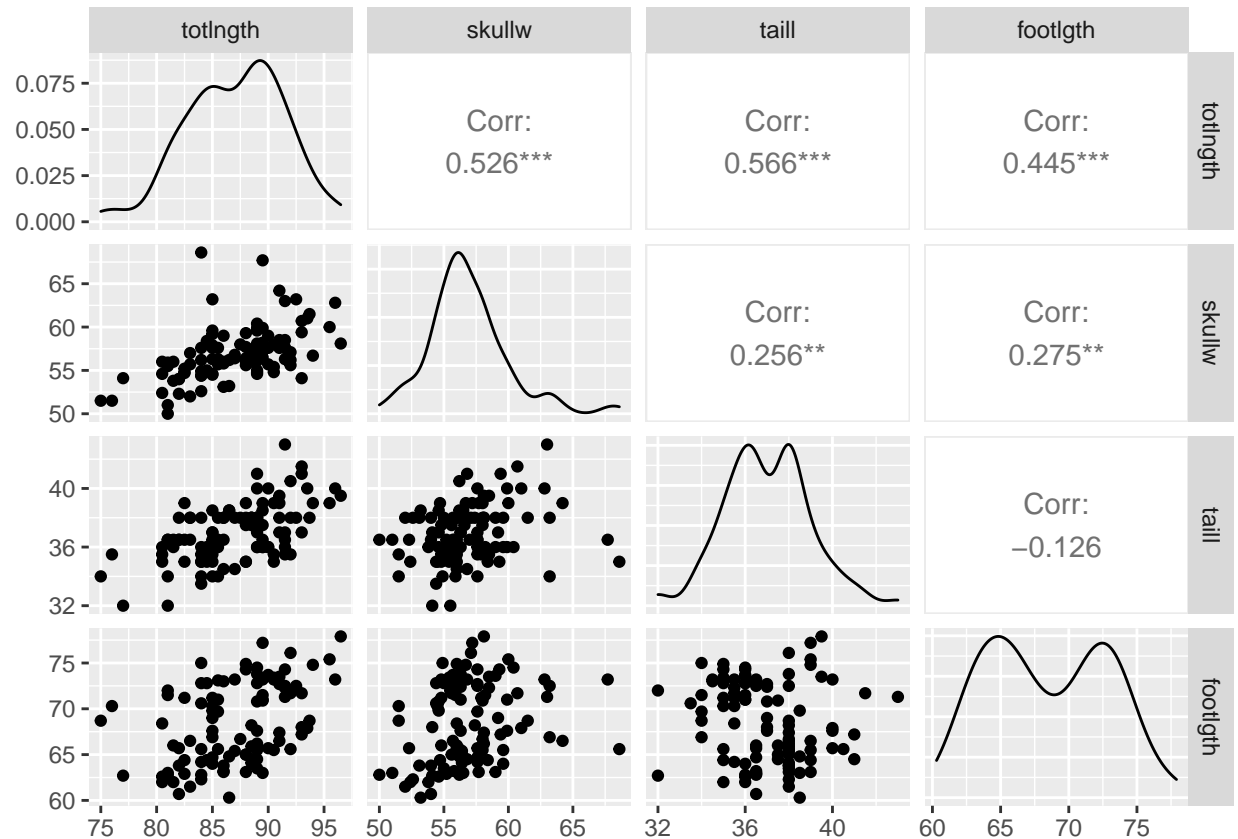
```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
## Removed 1 rows containing missing values (geom_point).
## Removed 1 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing non-finite values (stat_density).
```

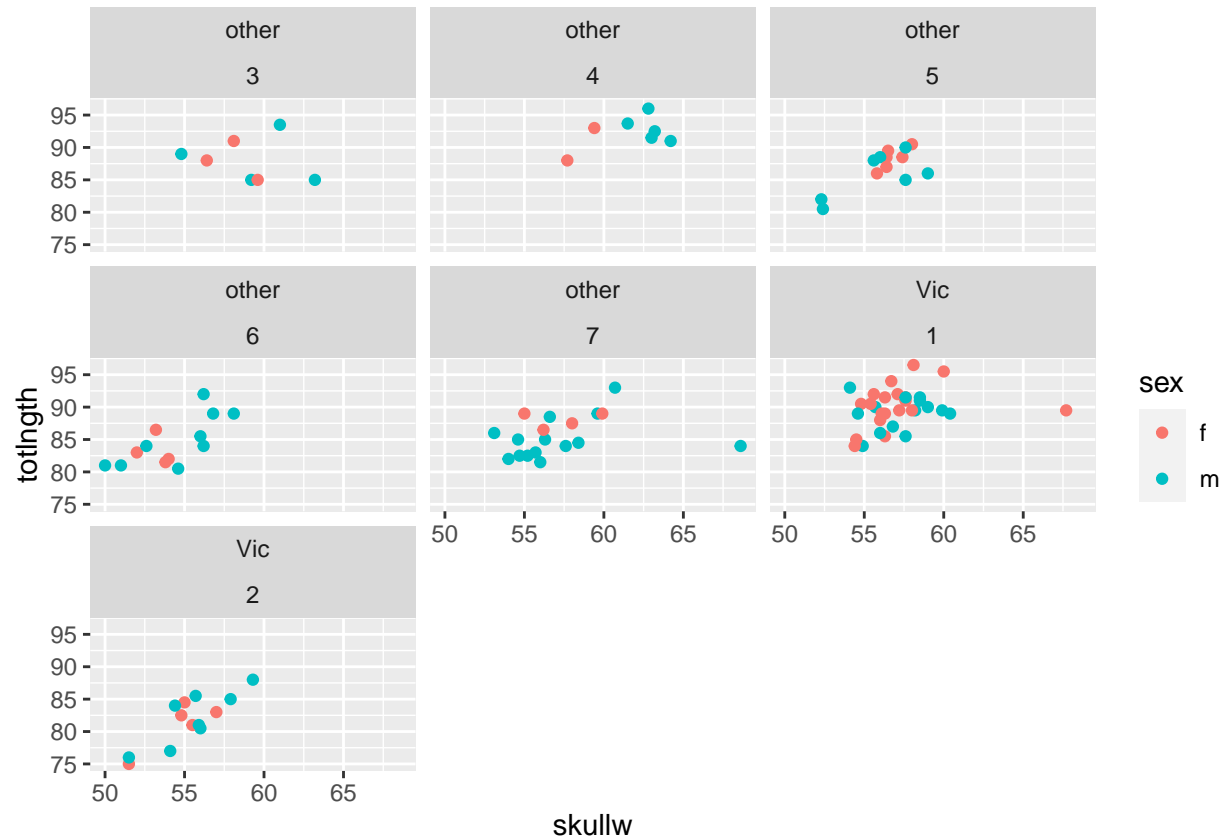


#

Dropping the co-linear variable.

There is small amount of correlation between tail and skull, foot len vs skull. However, since we want to know of skull width can predict the possum length.

```
#scatter plot
scat_plot <- ggplot(data = possum, aes(x=skullw,y=totlngth,col=sex))
scat_plot + geom_point()+facet_wrap(~Pop+site)
```



- What you take out of this graphs on the population trend of the possum

Let fit simple linear model

$$Lenght = \alpha + \beta_1 * Skullwidth + \epsilon$$

Where

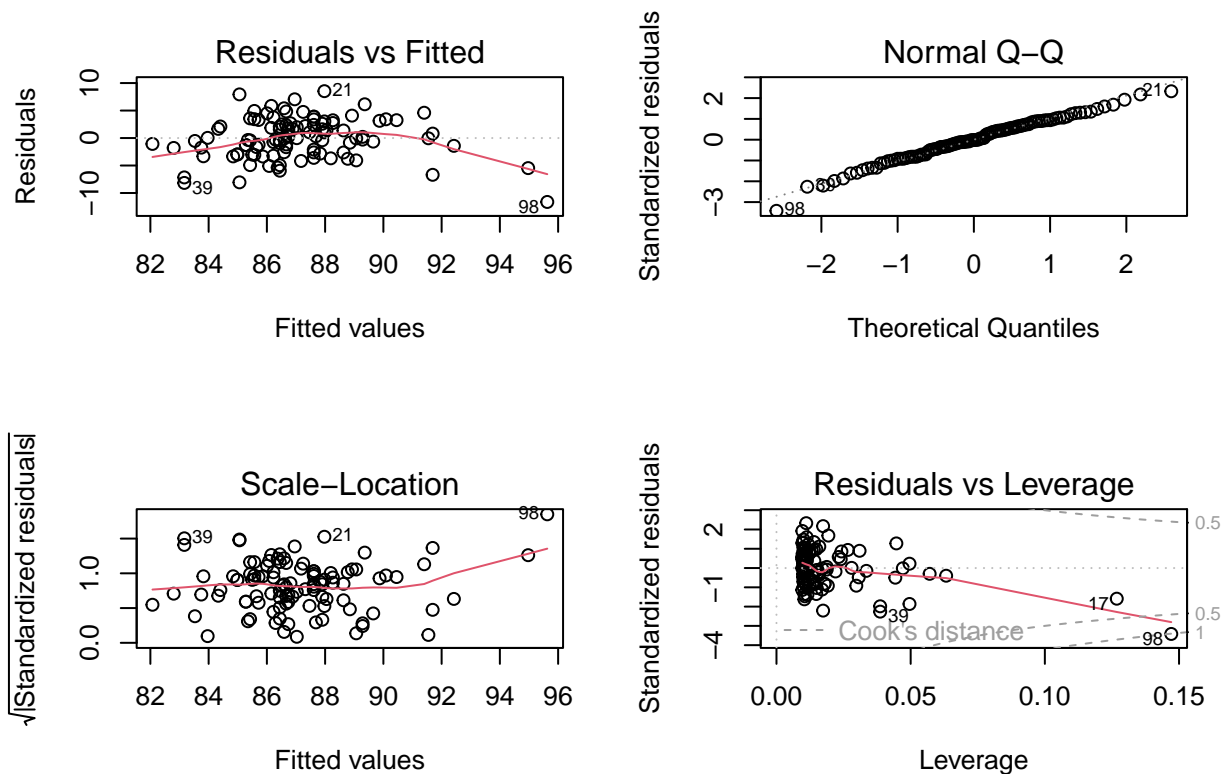
$$\epsilon \sim N(0, \sigma^2)$$

```
mod_lm <- lm(totlngth~skullw,data = possum)
summary(mod_lm)
```

```
##
## Call:
## lm(formula = totlngth ~ skullw, data = possum)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.6276  -2.6156  -0.0572   2.6212   8.5250
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  45.6305     6.6399   6.872 5.13e-10 ***
## skullw       0.7288     0.1166   6.253 9.50e-09 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.683 on 102 degrees of freedom
## Multiple R-squared:  0.2771, Adjusted R-squared:  0.27
## F-statistic: 39.1 on 1 and 102 DF, p-value: 9.498e-09
```

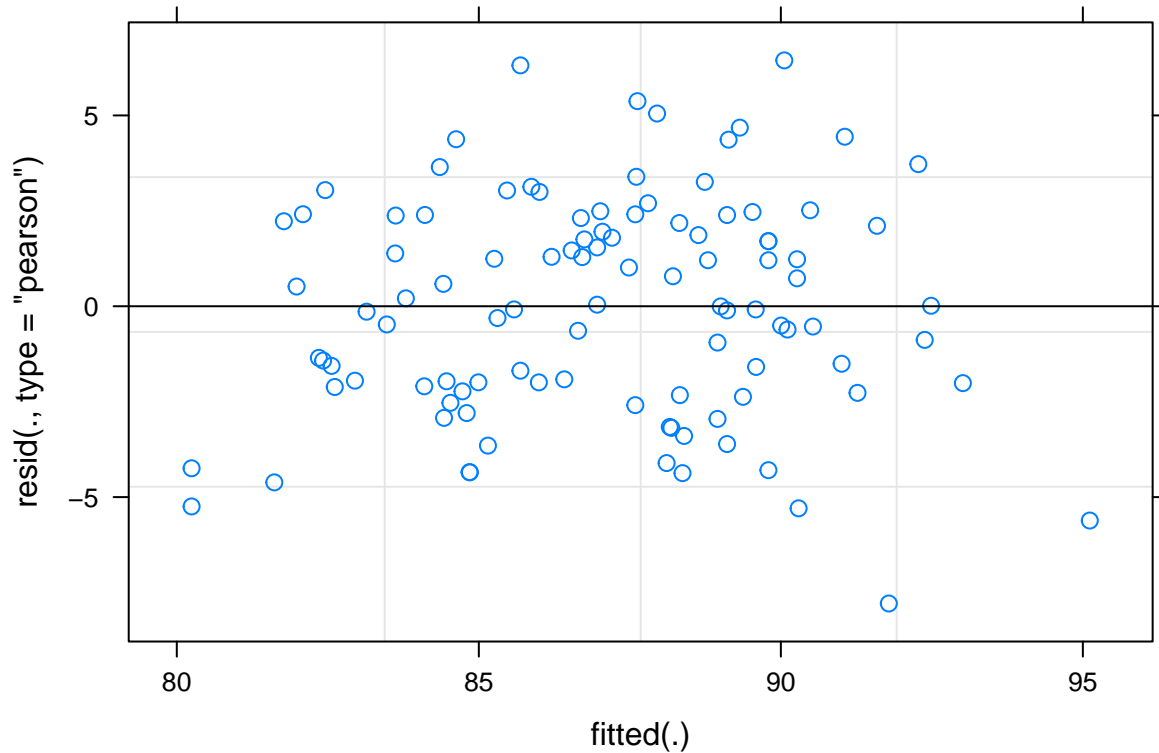
```
par(mfrow=c(2,2))
plot(mod_lm)
```



The data seems to be heterogeneous due to fact from the first graph

Let take population as random effect

```
model_LMM <- lmer(totlngth~skullw + (1|Pop:site),data = possum)
plot(model_LMM)
```



```
## clearly intercept are different.
dotplot(ranef(model_LMM, which = "Pop", condVar = TRUE))
```

```
## named list()
```

```
## let us see the model
print(model_LMM, corr = FALSE)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: totlngh ~ skullw + (1 | Pop:site)
## Data: possum
## REML criterion at convergence: 536.663
## Random effects:
## Groups Name Std.Dev.
## Pop:site (Intercept) 2.334
## Residual 2.982
## Number of obs: 104, groups: Pop:site, 7
## Fixed Effects:
## (Intercept) skullw
## 56.8234 0.5267
```

```
## ##getting SE and tabular format for fixed effect.
se <- sqrt(diag(vcov(model_LMM)))
# table of estimates with 95% CI
```

```
(tab <- cbind(Est = fixef(model_LMM), LL = fixef(model_LMM) - 1.96 * se, UL = fixef(model_LMM) + 1.96 *  
se))
```

```
##           Est           LL           UL  
## (Intercept) 56.8233765 44.2382563 69.4084967  
## skullw      0.5267382  0.3089698  0.7445067
```