Assignment 4 STAT 315-463: Multivariable Statistical Methods and Applications

```
library(here)
library(lattice)
library(lme4)
library(ggplot2)

library(tidyverse)
```

```
# Read in data file
tern14 <- read.table("Terns2014.csv", header = TRUE, sep = ',', na.strings = "na")
```

```
# Only using simple Linear Regression
Model <- lm(Age ~ Wing * ID, data = tern14)
summary(Model)
```

```
##
## Call:
## lm(formula = Age ~ Wing * ID, data = tern14)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6826 -1.6299 -0.5009  1.5109  6.3868
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.6194915  0.4952646   5.289 2.06e-07 ***
## Wing         0.1530470  0.0050336  30.405  < 2e-16 ***
## ID           0.0101452  0.0159284   0.637    0.525
## Wing:ID     -0.0002035  0.0001494  -1.363    0.174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.094 on 387 degrees of freedom
## Multiple R-squared:  0.8963, Adjusted R-squared:  0.8955
## F-statistic:  1115 on 3 and 387 DF,  p-value: < 2.2e-16
```

```
AIC(Model)
```

```
## [1] 1693.598
```

Model 1: Random intercepts

```
Model.1 <- lmer(Age ~ Wing + (1|ID), data = tern14)
AIC(Model.1)
```

```
## [1] 1143.745
```

In Model 1, we are looking at the model with the random intercepts, which assume that some terns have more and some have less wing length.

The AIC value of this mixed linear model indicates that it fits better than the model that ignores individual effects (The AIC score of random intercepts model is smaller than the simple linear regression model).

```
summary(Model.1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Age ~ Wing + (1 | ID)
##    Data: tern14
##
## REML criterion at convergence: 1135.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.4005 -0.4709 -0.0712  0.4367  5.6895
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 4.2721   2.0669
##  Residual             0.5619   0.7496
## Number of obs: 391, groups:  ID, 65
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 2.976157   0.283128   10.51
## Wing        0.147138   0.001146  128.34
##
## Correlation of Fixed Effects:
##      (Intr)
## Wing -0.395
```

Random Effects: Around 88.4% of variability is explained by the random intercepts. The variance of residual shows that there are still some factors from the simple body measurements of terns which affect age and cannot be explained by the model. This value is pretty close to zero, which

Fixed Effects: The intercept in the summary above shows that the average tern's wing length when they are born is 2.97 mm and the slope suggests that in average, each increase in wing length by 1 mm takes 0.14 days.

```
model1_coefs <- coef(Model.1)$ID %>%
  rename(Intercept = '(Intercept)', Slope = Wing) %>%
  rownames_to_column("ID")

# See coefficients
summary(model1_coefs$Intercept)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.4788  1.3941  2.6807  2.9762  4.6634  7.3462
```

```
hist(model1_coefs$Intercept)
```

Model 2: Random slopes

In this case, we are expecting all terns to start off at around the same point, but the effects of wing length on age differs across each tern.

```
Model.2 <- lmer(Age ~ Wing + (0 + Wing|ID), data = tern14)
AIC(Model.2)
```

```
## [1] 1174.457
```

The AIC value suggests this model outperforms the simple linear regression model, which does not take the different growth rate into consideration.

```
summary(Model.2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Age ~ Wing + (0 + Wing | ID)
##    Data: tern14
```

```
##
## REML criterion at convergence: 1166.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.0715 -0.4559 -0.0779  0.4010  4.2163
##
## Random effects:
##  Groups   Name Variance  Std.Dev.
##  ID       Wing 0.0005676 0.02382
##  Residual      0.5962503 0.77217
## Number of obs: 391, groups:  ID, 65
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 2.527025   0.122949   20.55
## Wing        0.153688   0.003315   46.35
##
## Correlation of Fixed Effects:
##      (Intr)
## Wing -0.405
```

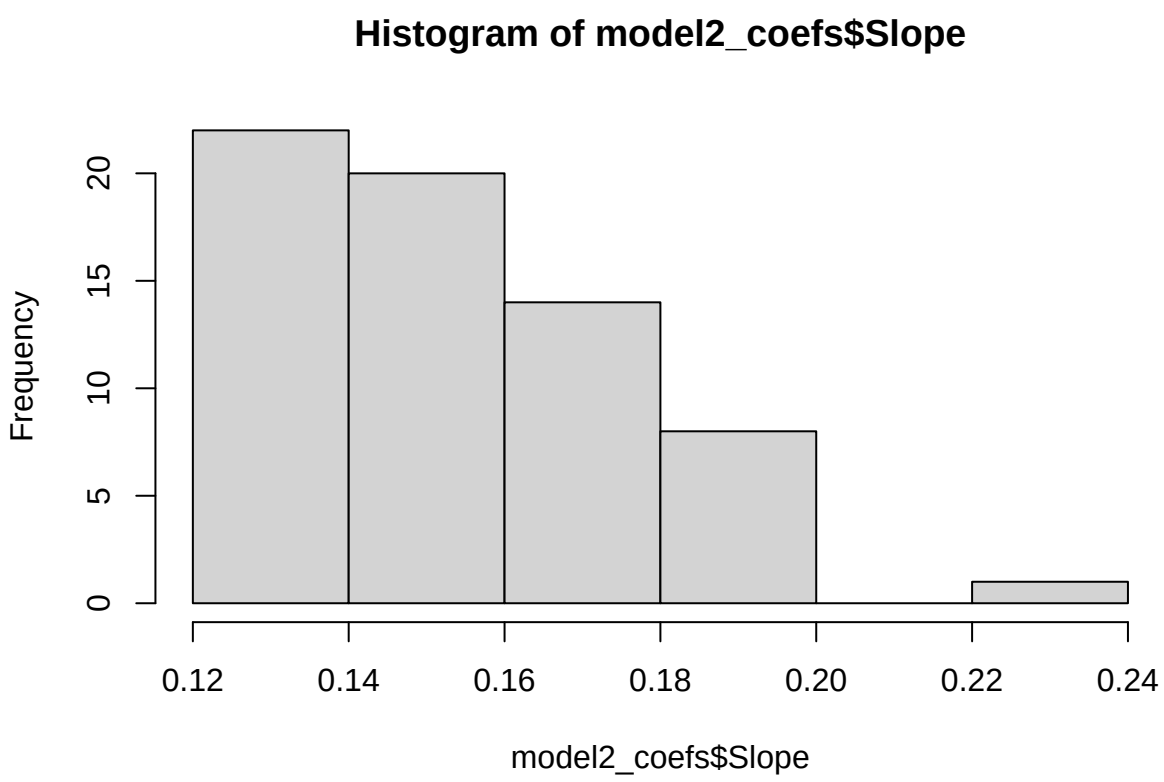
Random Effects:

Fixed Effects:

```
model2_coefs <- coef(Model.2)$ID %>%
  rename(Intercept = '(Intercept)', Slope=Wing) %>%
  rownames_to_column("ID")

summary(model2_coefs)
```

```
##       ID              Intercept         Slope
##  Length:65          Min.   :2.527   Min.   :0.1208
##  Class :character   1st Qu.:2.527   1st Qu.:0.1370
##  Mode  :character   Median :2.527   Median :0.1486
##                     Mean   :2.527   Mean   :0.1537
##                     3rd Qu.:2.527   3rd Qu.:0.1710
##                     Max.   :2.527   Max.   :0.2388
```

```
hist(model2_coefs$Slope)
```




```
coef(Model.2)
```

```
## $ID
##    (Intercept)      Wing
## 1     2.527025 0.1296118
## 2     2.527025 0.1266603
## 3     2.527025 0.1441504
## 4     2.527025 0.1283708
## 5     2.527025 0.1664200
## 6     2.527025 0.1352117
## 7     2.527025 0.1397402
## 8     2.527025 0.1387303
## 9     2.527025 0.1816841
## 10    2.527025 0.1647057
## 11    2.527025 0.1910185
## 12    2.527025 0.1718231
## 13    2.527025 0.1725178
## 14    2.527025 0.1304463
## 15    2.527025 0.1707286
## 16    2.527025 0.1621836
## 17    2.527025 0.1238346
## 18    2.527025 0.1445179
## 19    2.527025 0.1577769
## 20    2.527025 0.1469140
## 21    2.527025 0.1873774
## 22    2.527025 0.1885235
```

```
## 23    2.527025 0.1388217
## 24    2.527025 0.1799855
## 25    2.527025 0.1739090
## 26    2.527025 0.2388141
## 27    2.527025 0.1556988
## 28    2.527025 0.1982779
## 29    2.527025 0.1408648
## 30    2.527025 0.1738029
## 31    2.527025 0.1462256
## 32    2.527025 0.1276246
## 33    2.527025 0.1529132
## 34    2.527025 0.1370442
## 35    2.527025 0.1508789
## 36    2.527025 0.1676289
## 37    2.527025 0.1709695
## 38    2.527025 0.1383305
## 39    2.527025 0.1391249
## 40    2.527025 0.1517927
## 41    2.527025 0.1485587
## 42    2.527025 0.1720056
## 43    2.527025 0.1464499
## 44    2.527025 0.1618591
## 45    2.527025 0.1821830
## 46    2.527025 0.1417453
## 47    2.527025 0.1576291
## 48    2.527025 0.1300611
## 49    2.527025 0.1767589
## 50    2.527025 0.1207979
## 51    2.527025 0.1833605
## 52    2.527025 0.1438838
## 53    2.527025 0.1307264
## 54    2.527025 0.1551169
## 55    2.527025 0.1451454
## 56    2.527025 0.1361845
## 57    2.527025 0.1235031
## 58    2.527025 0.1246739
## 59    2.527025 0.1955957
## 60    2.527025 0.1401833
## 61    2.527025 0.1303502
## 62    2.527025 0.1509885
## 63    2.527025 0.1273769
## 64    2.527025 0.1520855
## 65    2.527025 0.1268382
##
## attr(,"class")
## [1] "coef.mer"
```

Model 3: Random slopes and intercepts

```
Model.3 <- lmer(Age ~ Wing + (1 + Wing|ID), data = tern14)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
```

```
## Model failed to converge with max|grad| = 0.439389 (tol = 0.002, component 1)
```

```
summary(Model.3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Age ~ Wing + (1 + Wing | ID)
##    Data: tern14
##
## REML criterion at convergence: 932.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2764 -0.4639 -0.0748  0.4586  4.9026
##
## Random effects:
##  Groups   Name        Variance  Std.Dev. Corr
##  ID       (Intercept) 3.4346907 1.85329
##           Wing        0.0003303 0.01817  -0.38
##  Residual             0.2256019 0.47498
## Number of obs: 391, groups:  ID, 65
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 2.697838   0.262709   10.27
## Wing        0.150908   0.002669   56.54
##
## Correlation of Fixed Effects:
##      (Intr)
## Wing -0.510
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.439389 (tol = 0.002, component 1)
```

```
AIC(Model.3)
```

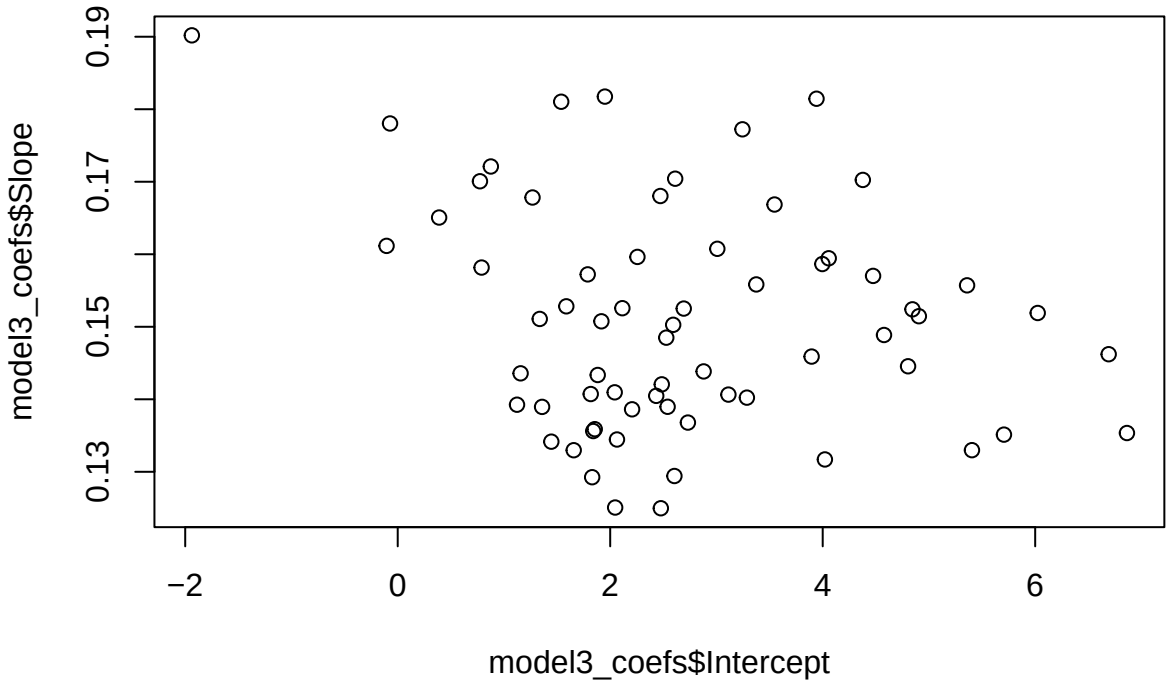
```
## [1] 944.8728
```

From the AIC result above, we can see that this model with random slopes and random intercepts has the best fit (the smallest AIC value among all four models). It does make sense because each fern has different growth rate and their wing lengths also differ when they were just born.

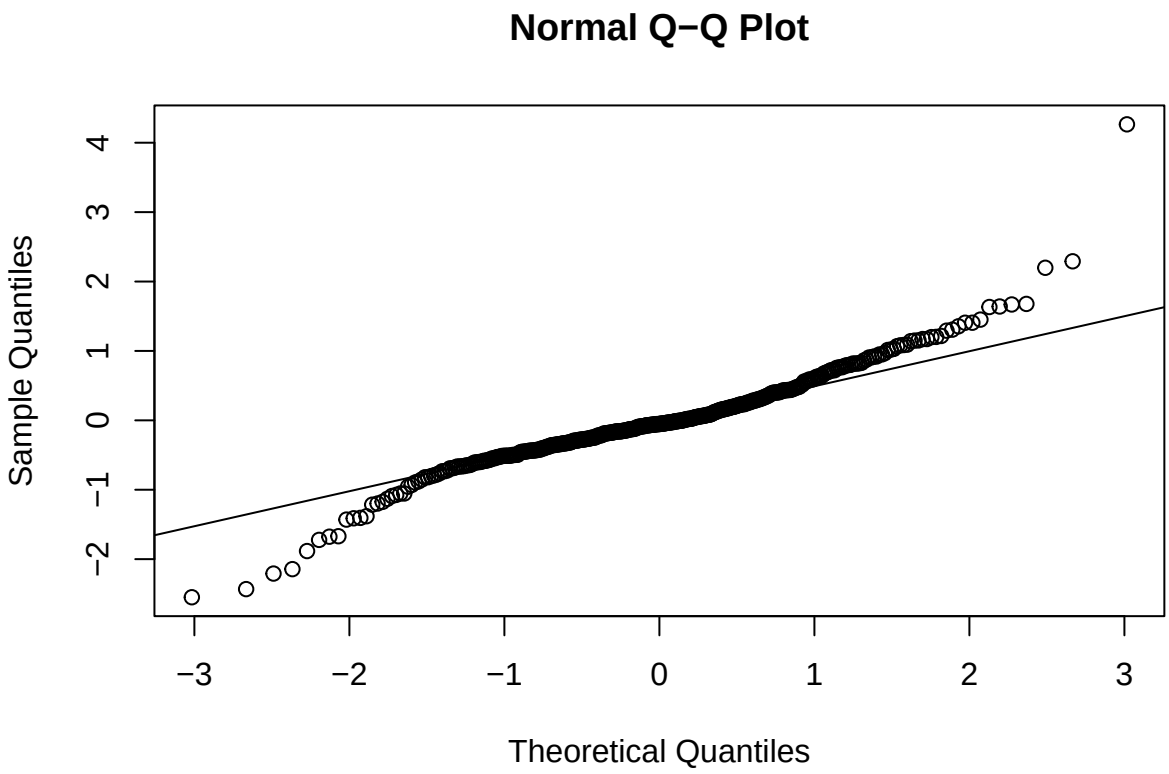
Random Effects:

Fixed Effects:

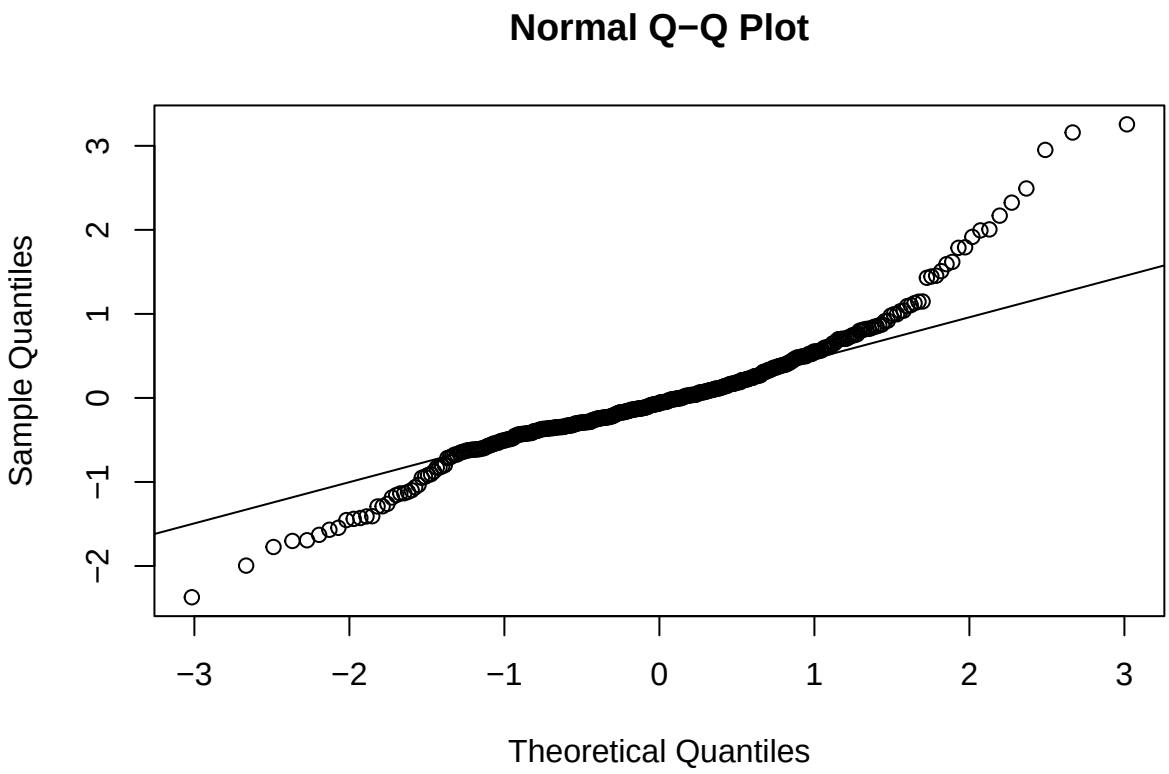
```
model3_coefs <- coef(Model.3)$ID %>%
  rename(Intercept = '(Intercept)', Slope=Wing) %>%
  rownames_to_column("ID")

plot(model3_coefs$Intercept, model3_coefs$Slope)
```



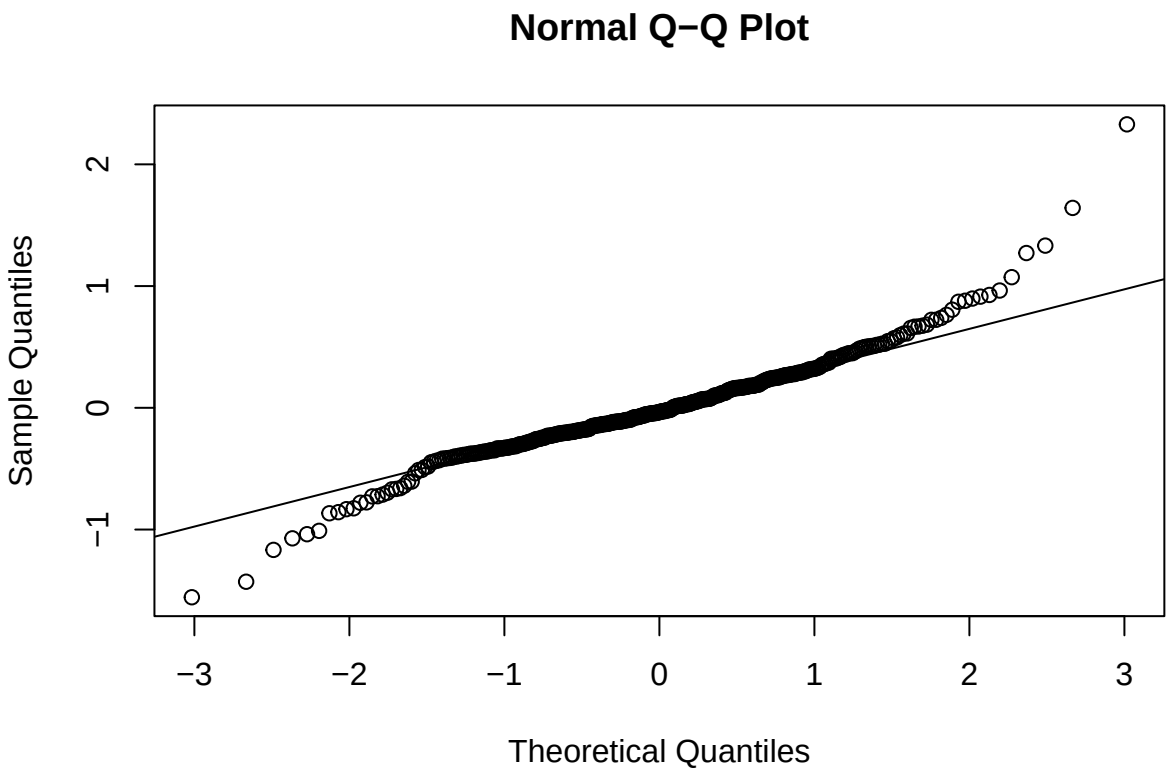

```
qqnorm(resid(Model.1))
qqline(resid(Model.1))
```




```
qqnorm(resid(Model.2))
qqline(resid(Model.2))
```




```
qqnorm(resid(Model.3))
qqline(resid(Model.3))
```


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−3
−2
−1
0
1
2
3
−1
0
1
2