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

Lisa Lu 31088272

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

```
# 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: By looking at the proportion of the random intercept variance compared to the total variance, we can see that 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.

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_coef <- coef(Model.1)$ID
summary(model1_coef)
```

```
##    (Intercept)              Wing
##  Min.    :-0.4788   Min.    :0.1471
##  1st Qu.: 1.3941    1st Qu.:0.1471
##  Median : 2.6807    Median :0.1471
##  Mean   : 2.9762    Mean    :0.1471
##  3rd Qu.: 4.6634    3rd Qu.:0.1471
##  Max.   : 7.3462    Max.    :0.1471
```

The intercepts vary from -0.48 to 7.35 with the slope being the same. This suggests that if we assume all the terns have the same grow rate of wing length, some may start off a really small initial wing length, whereas some may need a larger initial starting point.

Model 2: Random slopes

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

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

In this case, we are expecting all terns to start off at around the same wing length, but the effects of wing length on age differs across each tern. 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: From the proportion of random slope variance compared to the total variance, we can see that this value is close to 0, which indicates a low group level information, i.e., the variability here would not benefit much from this random effect.

Fixed Effects: The coefficient estimates suggest a similar model compared to other models. The t values here suggest that the wing length of a tern has a substantial effect on predicting the age of the tern because the t values are significantly different from zero.

```
model2_coef <- coef(Model.2)$ID
summary(model2_coef)
```

```
##   (Intercept)         Wing
##  Min.   :2.527   Min.   :0.1208
##  1st Qu.:2.527   1st Qu.:0.1370
##  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
```

When the intercept remains 2.5, namely, when the initial wing length of each tern is 2.53 mm, if the wing length increments by 1 mm, it takes from 0.12 to 0.23 days.

Model 3: Random slopes and intercepts

```
Model.3 <- lmer(Age ~ Wing + (1 + Wing|ID), data = tern14)
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 tern has different growth rate and their wing lengths also differ when they were just born.

```
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)
```

Random Effects: The variance of residual here shows that only a few factors from the wing length measurements that affect the prediction of age cannot be explained by the model. This is pretty good as this value is closer to zero than the other two models. The negative corr value here implies that the estimated values of the random effects are negatively related to each other. In the case, it means that as the random intercept (the initial wing length) increases, the random slope (the growth rate) decreases, and vice versa.

Fixed Effects: Here again, the t values suggest the highly significant effect that wing length has on predicting the tern's age.

All three mixed linear models suggest a negative correlation between slop and intercept. This means that a tern born with a longer wing tends to have a slower growth of wing length, whereas a tern's wing length will grow faster if it is born with a shorter wing length. This last model shows the strongest correlation between the wing lengths and the ages of terns.

```r
pardef <- par()
par(mfrow = c(2,2))

qqnorm(resid(Model), main = "Simple Linear Regression")
qqline(resid(Model))

qqnorm(resid(Model.1), main = "LMM with random intercepts")
qqline(resid(Model.1))

qqnorm(resid(Model.2), main = "LMM with random slopes")
qqline(resid(Model.2))

qqnorm(resid(Model.3), main = "LMM with random intercepts and slopes")
qqline(resid(Model.3))
```

Simple Linear Regression

LMM with random intercepts

LMM with random slopes

LMM with random intercepts and slope
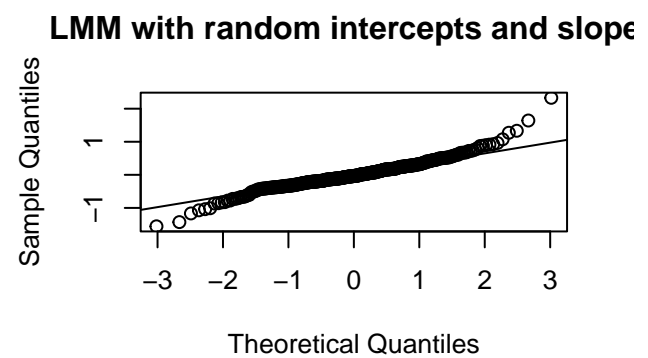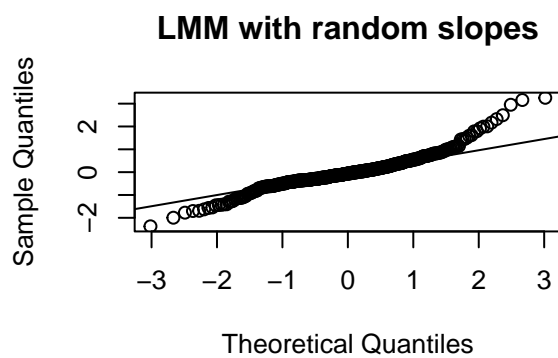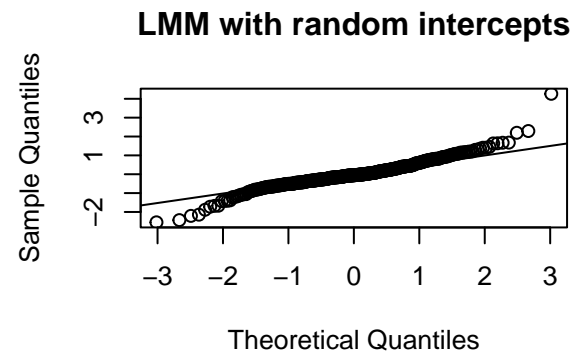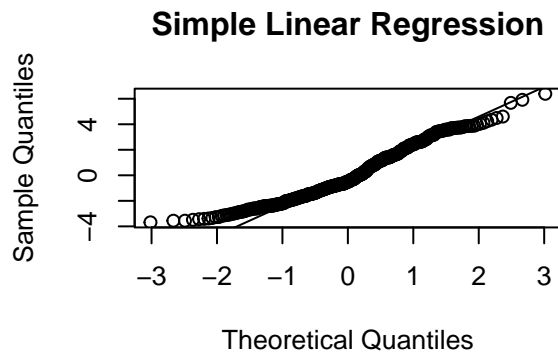
If we look at the normal Q-Q plots from all these four models, we can notice that with the simple linear regression, the data are skewed. With the other three linear mixed models, we can see the tailed distribution happened, which were probably caused by the skewed response variable (Age in this case). Another reason might be caused by the outliers in the data. From the plots below, we can see that there are few outliers in the original dataset.

```
pardef <- par()
par(mfrow = c(2,2))
plot(Model)
```

Residuals vs Fitted

Normal Q–Q

Scale–Location

Residuals vs Leverage