

# HW 5

Mohammad Imtiaz Nur

5/3/2021

**ID: 1878074**

## Task 1

```
require(lme4)
```

```
## Loading required package: lme4
```

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

```
require(lmerTest)
```

```
## Loading required package: lmerTest
```

```
##
```

```
## Attaching package: 'lmerTest'
```

```
## The following object is masked from 'package:lme4':
```

```
##
```

```
##      lmer
```

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

```
##
```

```
##      step
```

```
library(FSA)
```

```
## ## FSA v0.8.32. See citation('FSA') if used in publication.
```

```
## ## Run fishR() for related website and fishR('IFAR') for related book.
```

```
library(rcompanion)
```

```
time <- c(1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2,  
         4, 4, 4, 4, 4, 4, 4, 4, 8, 8, 8, 8, 8, 8, 8, 8,  
         13, 13, 13, 13, 13, 13, 13, 13)
```

```

temperature <- c(2, 2, 4.5, 4.5, 21, 21, 38, 38, 2, 2, 4.5, 4.5,
                21, 21, 38, 38, 2, 2, 4.5, 4.5, 21, 21, 38, 38,
                2, 2, 4.5, 4.5, 21, 21, 38, 38, 2, 2, 4.5, 4.5,
                21, 21, 38, 38 )

rating <- c(2.38, 2.19, 2.67, 2.39, 2.93, 2.72, 3.81, 3.07, 2.74,
           2.50, 2.81, 2.64, 2.97, 2.88, 4.14, 3.14, 2.75, 2.74,
           3.00, 2.79, 3.05, 3.21, 4.78, 3.45, 3.28, 2.83, 3.58,
           3.23, 3.68, 3.25, 5.78, 5.28, 3.81, 3.05, 3.67, 3.61,
           4.04, 4.23, 6.05, 7.17)

time <- as.factor(time)
temperature <- as.factor(temperature)

data <- data.frame(time,temperature,rating)

model = lmer(rating ~ temperature + (1|time), data=data, REML=TRUE)

model.null = lmer(rating ~ 1 + (1|time), data = data, REML = TRUE)

anova(model, model.null)

## refitting model(s) with ML (instead of REML)

## Data: data
## Models:
## model.null: rating ~ 1 + (1 | time)
## model: rating ~ temperature + (1 | time)
##          npar      AIC      BIC logLik deviance Chisq Df Pr(>Chisq)
## model.null    3 118.165 123.232 -56.083  112.165
## model         6  79.801  89.934 -33.900   67.801 44.364  3 1.263e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

From the anova test between the model and the null model, we found that the significance is 0.001 for the model. Hence, the model is better than the null model. Our the null hypothesis is the temperature has no effect on rating.

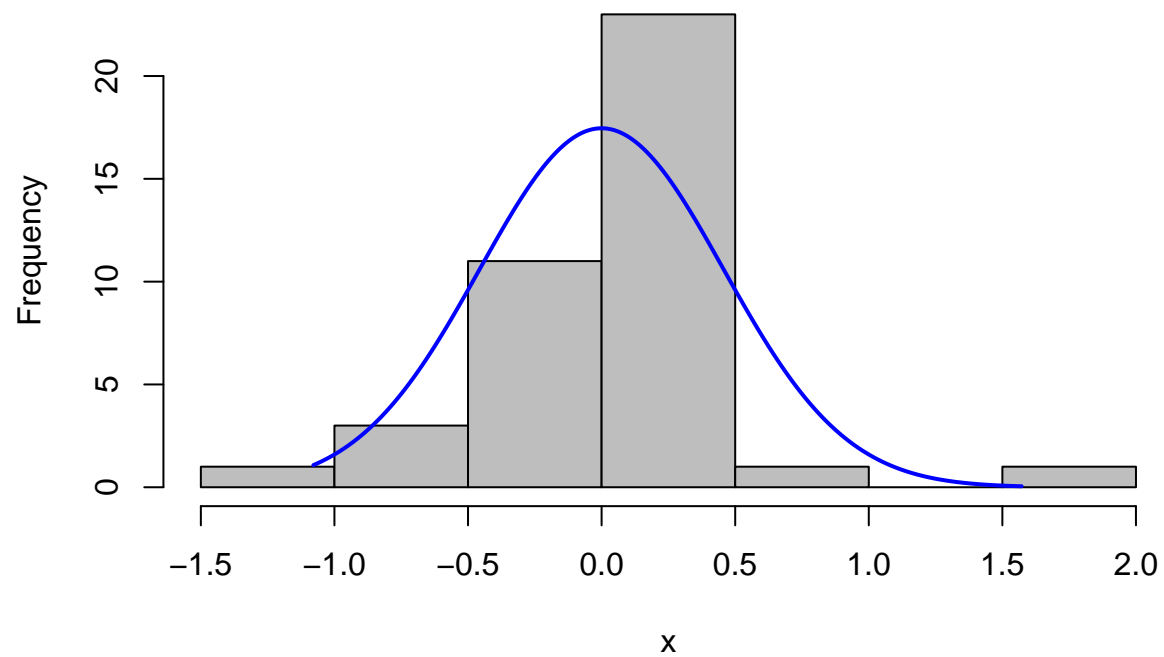
## Histogram

```

res = residuals(model)

plotNormalHistogram(res)

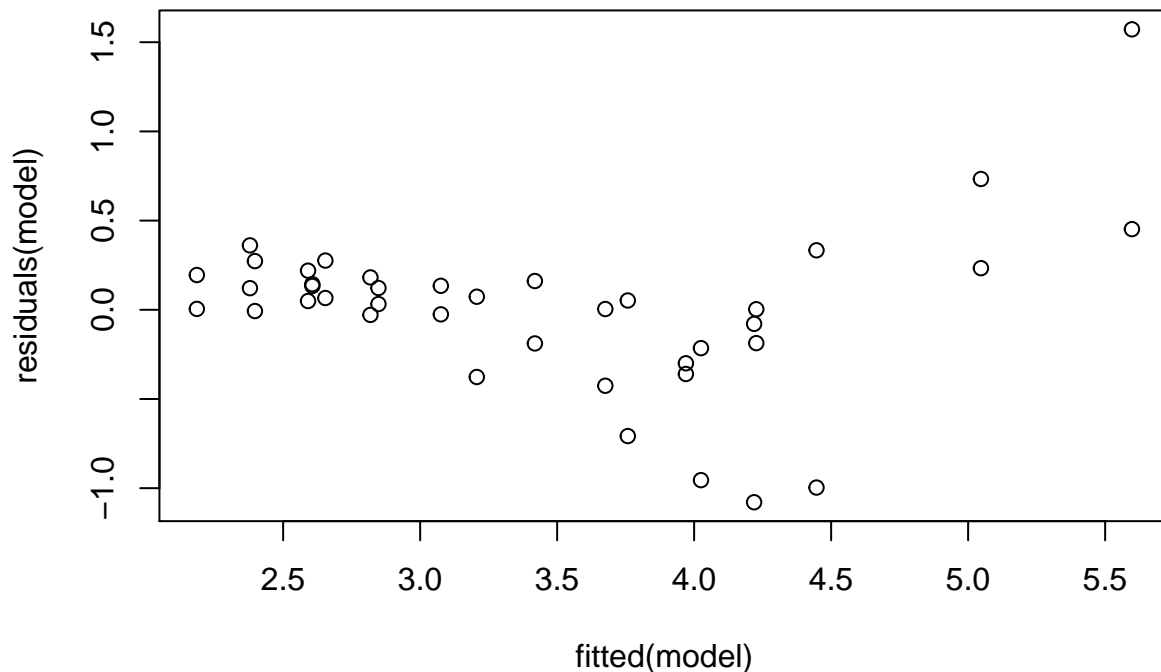
```



From the histogram of the residuals, it follows normal distribution.

### Fitted vs Residuals

```
plot(fitted(model),residuals(model))
```



From the residual vs fitted plot, we can see there is no consistent pattern. So the residuals are not reasonably homoscedastic.

## Comparing the LS mean for each temperature

```
require("lsmeans")
```

```
## Loading required package: lsmeans
```

```
## Loading required package: emmeans
```

```
## The 'lsmeans' package is now basically a front end for 'emmeans'.
## Users are encouraged to switch the rest of the way.
## See help('transition') for more information, including how to
## convert old 'lsmeans' objects and scripts to work with 'emmeans'.
```

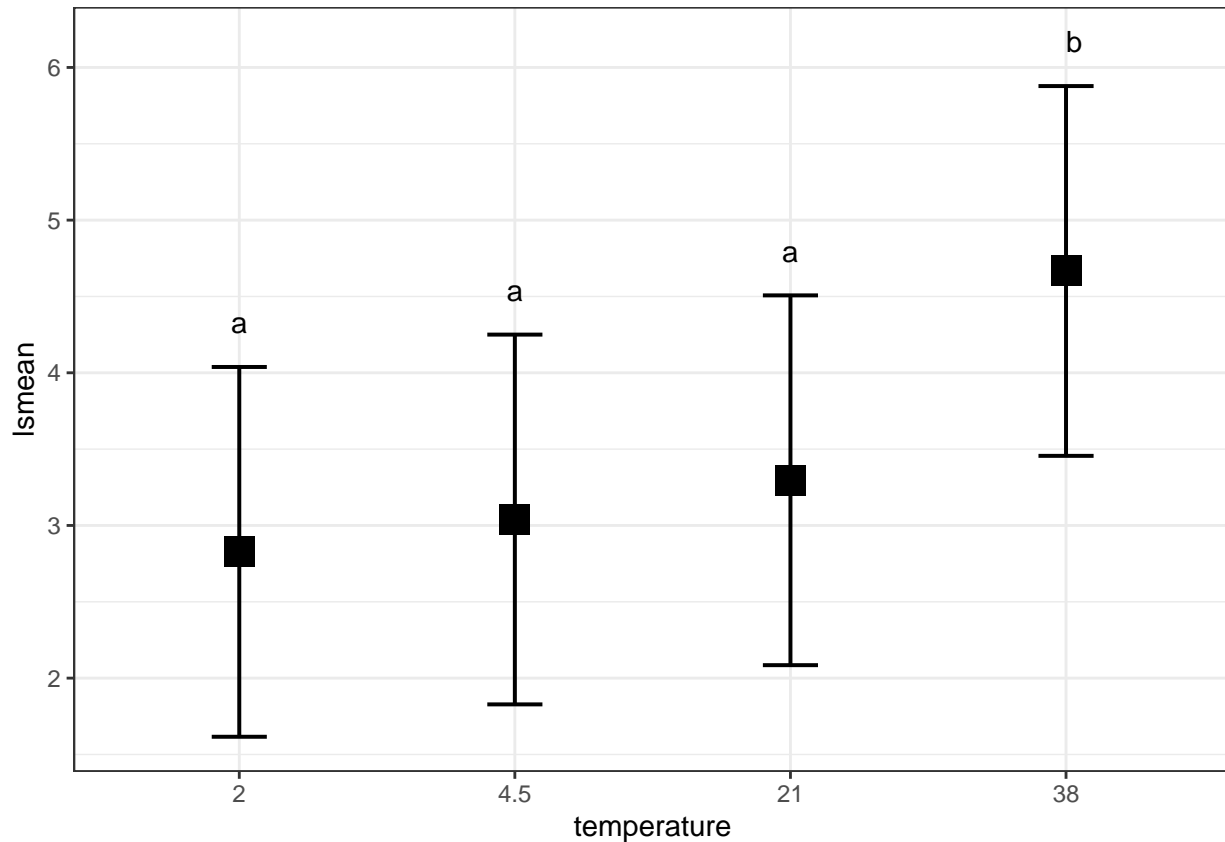
```
marginal <- lsmeans(model, "temperature")
```

```
CLD = multcomp::cld(marginal, alpha=0.05, Letters=letters, adjust="tukey")
```

```
## Note: adjust = "tukey" was changed to "sidak"
```

```
## because "tukey" is only appropriate for one set of pairwise comparisons
```

```
library(ggplot2)
ggplot(CLD, aes(x = temperature, y = lsmean, label = .group)) +
  geom_point(shape = 15, size = 5) +
  geom_errorbar(aes(ymin = lower.CL, ymax = upper.CL), width = 0.2, size = 0.7) +
  theme_bw() +
  geom_text(nudge_x = c(0, 0, 0), nudge_y = c(1.5, 1.5, 1.5))
```



From the above plot, it is clear that the temperature value 38 has higher value than all other temperatures.

## Significance of the Temperature and Time

```
print("Temperature")
```

```
## [1] "Temperature"
```

```
anova(model)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
## temperature 20.616  6.8721     3    32  27.223 6.163e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
print("Time")
```

```
## [1] "Time"
```

```
rand(model)
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## rating ~ temperature + (1 | time)
##      npar logLik    AIC    LRT Df Pr(>Chisq)
## <none>      6 -36.349  84.698
## (1 | time)   5 -47.940 105.880 23.182  1  1.473e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here, the p-value for both temperature and time are very much less than 0.001. Hence, both time and temperature have quite significant effect.

## Pairwise difference of temperature

```
pairs(marginal, adjust="tukey")
```

```
## contrast estimate    SE df t.ratio p.value
## 2 - 4.5      -0.212 0.225 32 -0.943  0.7818
## 2 - 21       -0.469 0.225 32 -2.087  0.1790
## 2 - 38       -1.840 0.225 32 -8.189  <.0001
## 4.5 - 21     -0.257 0.225 32 -1.144  0.6657
## 4.5 - 38     -1.628 0.225 32 -7.245  <.0001
## 21 - 38      -1.371 0.225 32 -6.102  <.0001
##
## Degrees-of-freedom method: kenward-roger
## P value adjustment: tukey method for comparing a family of 4 estimates
```

From the pairwise difference, we can see that the temperature 38 is significantly different than all other temperature group as all pairs with 38 has p\_value below 0.001 significance level. So, we can say it rejects the null hypothesis that all pairs are equal.

## Task 2

```
variety <- c("Variety A", "Variety B", "Variety C", "Variety D",
            "Variety A", "Variety B", "Variety C", "Variety D",
            "Variety A", "Variety B", "Variety C", "Variety D",
            "Variety A", "Variety B", "Variety C", "Variety D",
            "Variety A", "Variety B", "Variety C", "Variety D",
            "Variety A", "Variety B", "Variety C", "Variety D",
            "Variety A", "Variety B", "Variety C", "Variety D")
```

```

yield <- c(8.3, 9.1, 10.1, 7.8, 9.4, 9.0, 10.0, 8.2, 9.1, 8.1,
          9.6, 8.1, 9.1, 8.2, 9.3, 7.9, 9.0, 8.8, 9.8, 7.7,
          8.9, 8.4, 9.5, 8.0, 8.9, 8.3, 9.4, 8.1)

df <- data.frame(variety, yield)

df$variety <- ordered(df$variety,
                     levels = c("Variety A", "Variety B", "Variety C", "Variety D"))
levels(df$variety)

```

```
## [1] "Variety A" "Variety B" "Variety C" "Variety D"
```

```
kruskal.test(yield ~ variety, data = df)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  yield by variety
## Kruskal-Wallis chi-squared = 22.683, df = 3, p-value = 4.701e-05
```

Here, p-value (4.701e-05) is less than the significance level 0.01. Hence, there is significant evidence that the distribution is different for at least one variety.