# STAT302 Assignment 4 Solution

## Daisy Yu

## 8/4/2023

## Q5.16 (4 points)

**a. (2 points)**

The explanatory variable is type of dog food, and there are three different types of dog food being used. The response variable is the average number of hours of sleep per 24 hours, which is being used as a measure of energy.

**b. (1 point)**

This was a randomized experiment because the treatments (type of dog food) were randomly assigned to the subjects (Border Collies).

**c. (1 point)**

The subjects were randomly assigned to treatments, which makes the observations independent of each other.

## Q5.18 (4 points)

**a. (1 point)**

The 45 Border Collies available for the experiment.

**b. (1 point)**

The 3 different dog foods.

**c. (2 points)**

A design is balanced if all treatments are assigned to the same number of units. Since we have 45 units and 3 treatments, randomly assign 15 Border Collies to each type of dog food.

## Q5.20 (3 points)

2, 42, and 44.

## Q5.24 (4 points)

**a. (2 points)**

$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$ or the size of turtles are the same in three different states    $H_a$ : at least one $\alpha_i \neq$ 0 or at least one state has different size of turtles than the others.

**b. (1 point)**

We need to know the actual values of response variable, sample size and standard error of these three samples in order to conduct ANOVA.

**c. (1 point)**

We need residuals to get normal QQ plot, residuals plot and histogram to check assumptions for ANOVA.

## Q5.34 (9 points)

**a. (3 points)**

DF = 51 - 46 = 5
SSGroups = 64.77 - 39.87 = 24.90
F = 4.9807/0.8667 = 5.75

**b. (2 points)**

Since the degrees of freedom for aphid/plant combination is 5 and the degrees of freedom for treatments is the number of groups minus 1, the number of different occupations considered is 6.

**c. (4 points)**

$H_0$ : mean of all groups are same
$H_1$ : at least one combination of means are different
Test statistic: F = 5.75
P value $< 0.05$
We reject the null hypothesis. We have strong evidence that mean amount of honeydew produced by aphids is different for at least one aphid/plant combination.

## Q5.38 (5 points)

**a. (1 point)**

The hypotheses are $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$ vs $H_1$ : at least one $\alpha_i \neq 0$.

**b. (2 points)**

```r
library(Stat2Data)
library(mosaic)
```

```
## Registered S3 method overwritten by 'mosaic':
##   method                              from
##   fortify.SpatialPolygonsDataFrame ggplot2

##
## The 'mosaic' package masks several functions from core packages in order to add
## additional features.  The original behavior of these functions should not be affected by this.

##
## Attaching package: 'mosaic'

## The following objects are masked from 'package:dplyr':
##
##     count, do, tally

## The following object is masked from 'package:Matrix':
##
##     mean

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

## The following objects are masked from 'package:stats':
##
##     binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,
##     quantile, sd, t.test, var

## The following objects are masked from 'package:base':
##
##     max, mean, min, prod, range, sample, sum
```

```r
data(Meniscus)
Meniscus$Method <- as.factor(Meniscus$Method)
model <- aov(Stiffness~Method, data=Meniscus)
par(mfrow=c(2,2))
qqnorm(model$residuals,ylab="Residuals",xlab="Normal Quantiles",main="")
qqline(model$residuals,col='red')
boxplot(model$residuals~Method,data=Meniscus,ylab="Residual",xlab="Treatment")
hist(model$residuals,main='Histogram of residuals',xlab='Residuals')
plot(fitted(model),resid(model));abline(h=0,col='red')
```

### Histogram of residuals
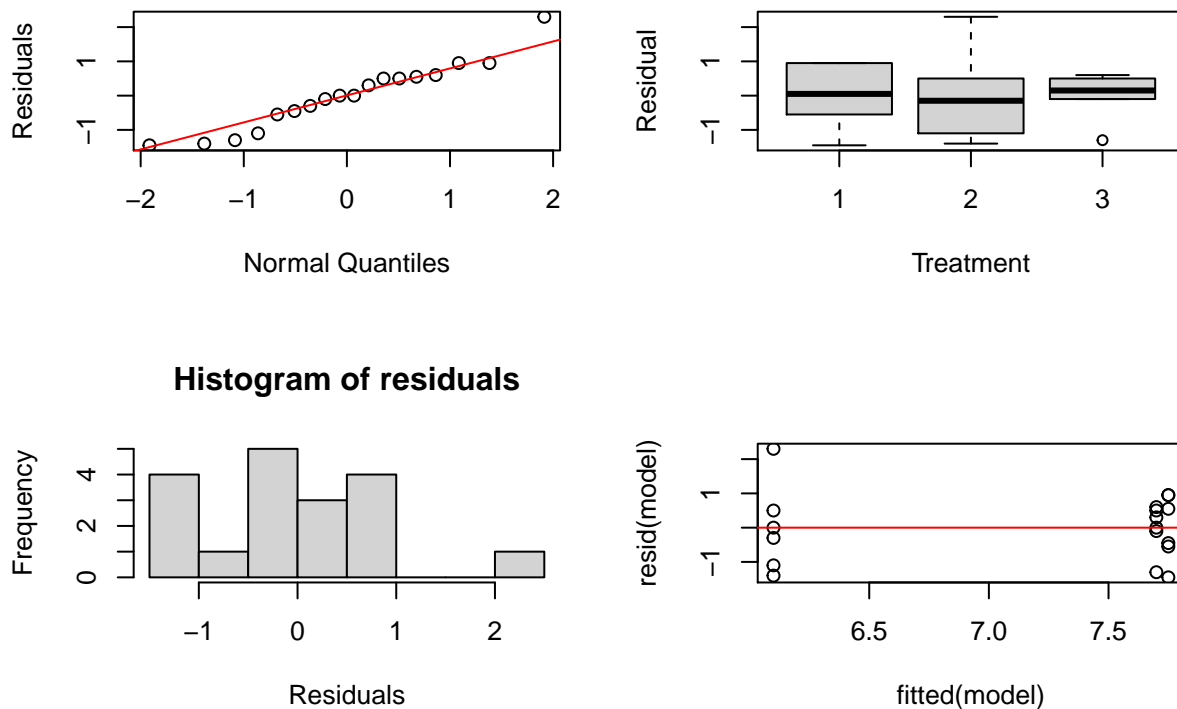




```
favstats(Stiffness~Method, data=Meniscus)[c("Method","n","mean","sd")]
```

```
##   Method n mean        sd
## 1      1 6 7.75 0.9710819
## 2      2 6 6.10 1.3266499
## 3      3 6 7.70 0.6928203
```

The graphs of the residuals are given below. The residuals are reasonably normally distributed and show similar variability. The actual standard deviations range from 0.693 to 1.327. These are close enough that an ANOVA procedure is acceptable.

### c. (2 points)

```
summary(model)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Method        2  10.57   5.285   4.981 0.0219 *
## Residuals    15  15.91   1.061
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value $= 0.022 < 0.05$, we reject the null hypothesis and conclude that at least one method of repairing has a different amount of stiffness in comparison to the other two methods.

## Q5.40 (5 points)

### a. (1 point)

The hypotheses are $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$ vs $H_1$ : at least one $\alpha_i \neq 0$.

**b. (2 points)**

```
model <- aov(Displacement~Method, data=Meniscus)
par(mfrow=c(2,2))
qqnorm(model$residuals,ylab="Residuals",xlab="Normal Quantiles",main="")
qqline(model$residuals,col='red')
boxplot(model$residuals~Method,data=Meniscus,ylab="Residual",xlab="Treatment")
hist(model$residuals,main='Histogram of residuals',xlab='Residuals')
plot(fitted(model),resid(model));abline(h=0,col='red')
```



### Histogram of residuals
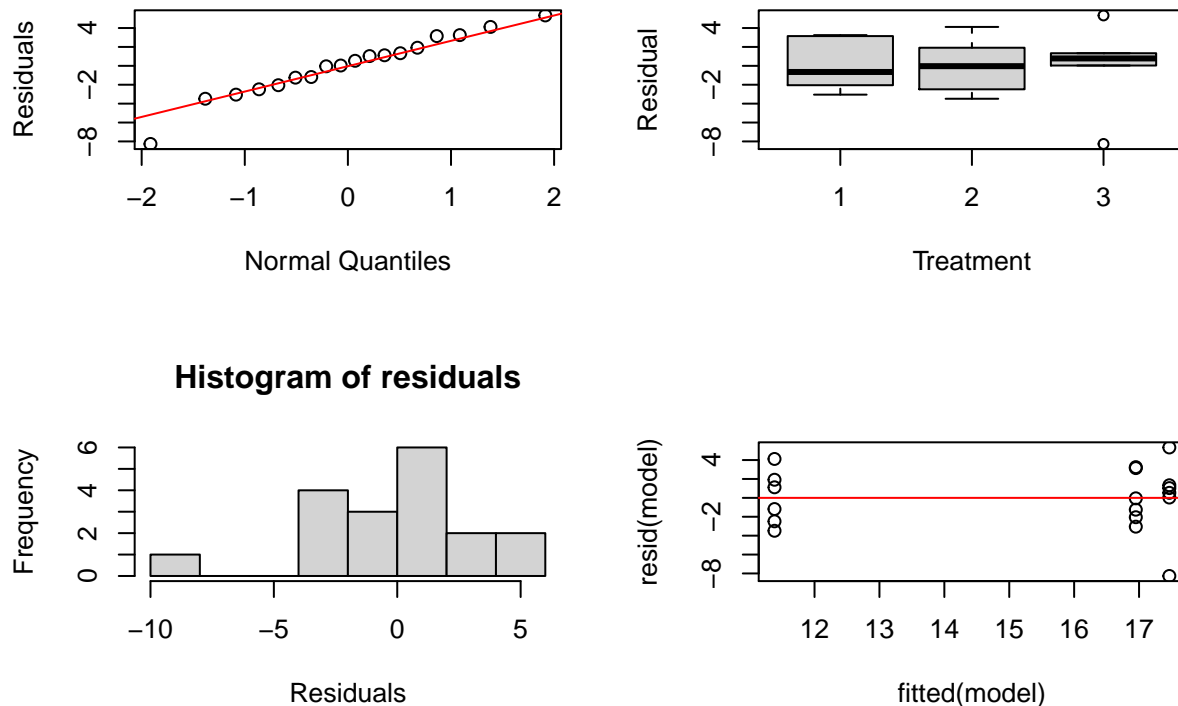


```
favstats(Displacement~Method, data=Meniscus)[c("Method","n","mean","sd")]
```

```
##   Method n      mean       sd
## 1      1 6 16.95000 2.666646
## 2      2 6 11.38333 2.883343
## 3      3 6 17.46667 4.470198
```

Normal probability plot and histogram of residuals indicate that residuals are normally distributed.Residual vs fitted value indicate deviation in constant variance, but reasonable and centered around zero. Thus, These are close enough that an ANOVA procedure is acceptable.

**c. (2 points)**

```
summary(model)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Method       2  136.5   68.26   5.784 0.0137 *
## Residuals   15  177.0   11.80
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value $= 0.014 < 0.05$, we reject the null hypothesis and conclude that there is at least one method of repairing that has a different amount of displacement in comparison to the other two methods.
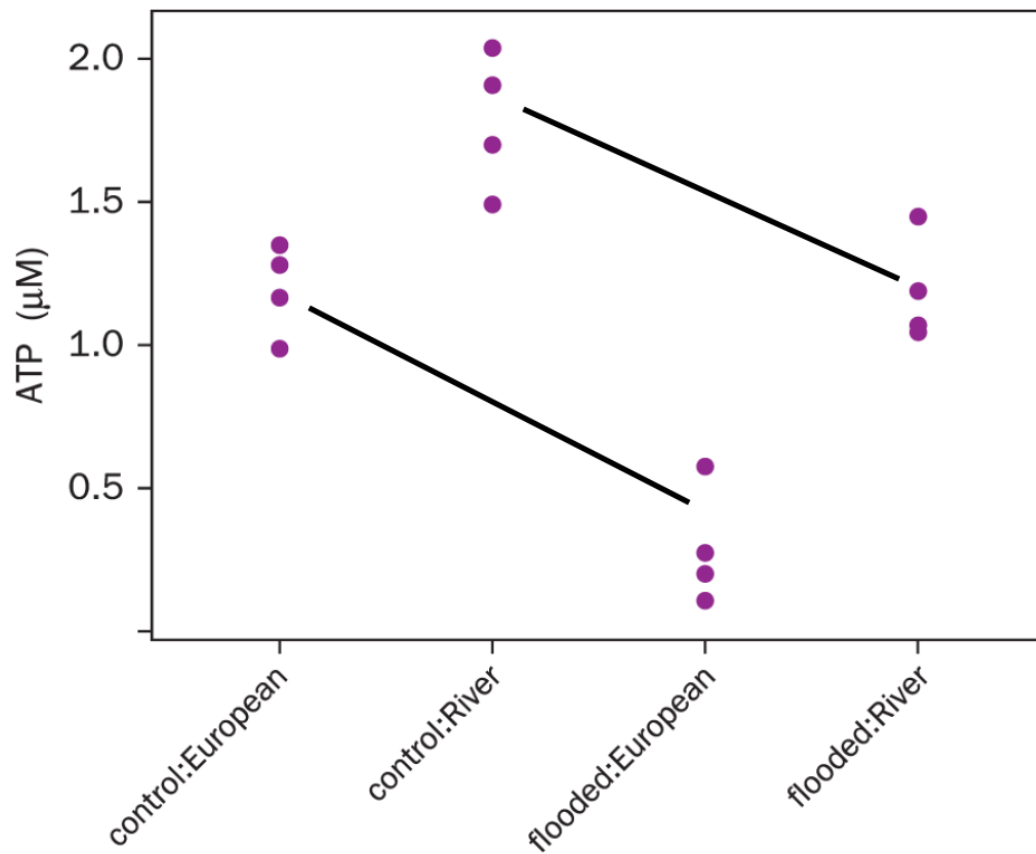
## Q7.2 (4 points)

a. (1 point) **2**

b. (1 point) **2**

c. (1 point) **4**

d. (1 point) **18**

## Q7.20 (1 point)



Cannon, et. al., *STAT2: Modeling with Regression and ANOVA,* 2e, © 2019
W. H. Freeman and Company

Since the two lines are parallel, there is no interaction effect between the conditions and the species on ATP concentration.

# Q7.22 (1 point)

Since the two lines are parallel, there is no interaction effect between the conditions and the species on ATP concentration.