HW#4 Solution (week5 HW)

10/3/2019

HW 18.17 Winding speeds

```
# HW 18.17

#HW18  <- read.table("CH18PR17.txt")
HW18  <- read.table(
   url("https://raw.githubusercontent.com/npmldabook/Stat3119/master/Week5/CH18PR17.txt"))
# rename the variables
names(HW18)<- c("response", "speed", "units")

HW18$speed <- as.factor(HW18$speed)
head(HW18)
```

```
##   response speed units
## 1        4     1     1
## 2        3     1     2
## 3        2     1     3
## 4        3     1     4
## 5        4     1     5
## 6        4     1     6
```

```
str(HW18)
```

```
## 'data.frame':    64 obs. of  3 variables:
##  $ response: num  4 3 2 3 4 4 3 6 5 4 ...
##  $ speed   : Factor w/ 4 levels "1","2","3","4": 1 1 1 1 1 1 1 1 1 1 ...
##  $ units   : int  1 2 3 4 5 6 7 8 9 10 ...
```

a. fit one-way ANOVA analysis, get fitted value and residulas

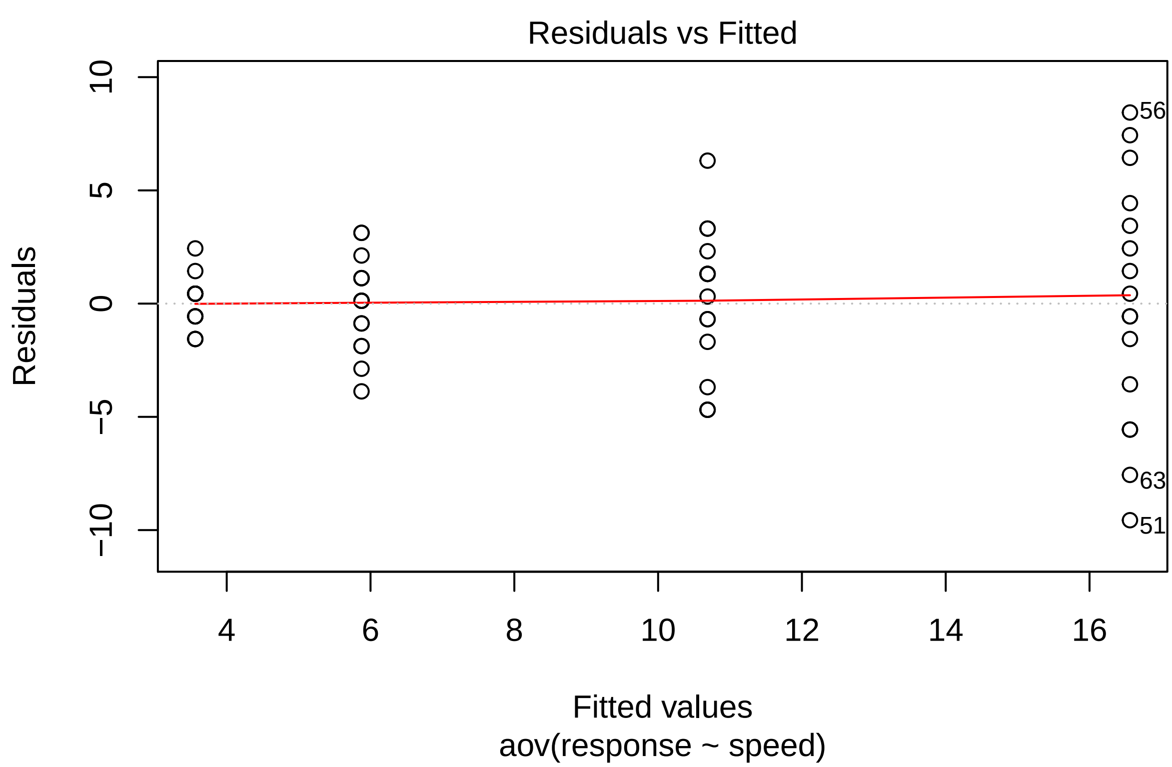
```
fit <- aov(response ~ speed, data = HW18)
list(fitted= predict(fit, newdata = data.frame(speed = factor(1:4))),
     Residuals=fit$residuals)
```

```
## $fitted
##       1       2       3       4
##  3.5625  5.8750 10.6875 16.5625
##
## $Residuals
##       1       2       3       4       5       6       7       8       9
##  0.4375 -0.5625 -1.5625 -0.5625  0.4375  0.4375 -0.5625  2.4375  1.4375
##      10      11      12      13      14      15      16      17      18
##  0.4375 -1.5625  0.4375  0.4375 -1.5625 -0.5625  0.4375  1.1250  0.1250
##      19      20      21      22      23      24      25      26      27
```

```
## -1.8750  0.1250  1.1250 -3.8750  3.1250 -0.8750 -0.8750  3.1250 -2.8750
##      28      29      30      31      32      33      34      35      36
##  2.1250  0.1250 -1.8750  1.1250  0.1250  1.3125 -4.6875  3.3125  1.3125
##      37      38      39      40      41      42      43      44      45
## -0.6875 -1.6875  1.3125  6.3125 -3.6875 -4.6875  1.3125  0.3125 -4.6875
##      46      47      48      49      50      51      52      53      54
##  2.3125 -0.6875  3.3125  0.4375 -1.5625 -9.5625  3.4375 -3.5625 -5.5625
##      55      56      57      58      59      60      61      62      63
## -0.5625  8.4375 -5.5625  7.4375  1.4375  4.4375 -0.5625  2.4375 -7.5625
##      64
##  6.4375
```

b. Residual plot to study whether or not the error variances are equal for the four winding speeds.

```r
plot(fit,1)
```



Results: The variances seem different for different speeds. The factor level for fater speeds had larger variances.

c) Use Brown-Forsythe test

```r
(mediani = with(HW18, by( response, speed, median)))
```

```
## speed: 1
## [1] 4
## ------------------------------------------------------------
## speed: 2
## [1] 6
## ------------------------------------------------------------
## speed: 3
## [1] 11.5
## ------------------------------------------------------------
## speed: 4
## [1] 16.5
```

```r
(Factor.median = rep(as.numeric(mediani), rep(16,4)))
```

```
##  [1]  4.0  4.0  4.0  4.0  4.0  4.0  4.0  4.0  4.0  4.0  4.0  4.0  4.0  4.0
## [15]  4.0  4.0  6.0  6.0  6.0  6.0  6.0  6.0  6.0  6.0  6.0  6.0  6.0  6.0
## [29]  6.0  6.0  6.0  6.0 11.5 11.5 11.5 11.5 11.5 11.5 11.5 11.5 11.5 11.5
## [43] 11.5 11.5 11.5 11.5 11.5 11.5 16.5 16.5 16.5 16.5 16.5 16.5 16.5 16.5
## [57] 16.5 16.5 16.5 16.5 16.5 16.5 16.5 16.5
```

```r
(dij= abs(HW18$response -Factor.median))
```

```
##  [1] 0.0 1.0 2.0 1.0 0.0 0.0 1.0 2.0 1.0 0.0 2.0 0.0 0.0 2.0 1.0 0.0 1.0
## [18] 0.0 2.0 0.0 1.0 4.0 3.0 1.0 1.0 3.0 3.0 2.0 0.0 2.0 1.0 0.0 0.5 5.5
## [35] 2.5 0.5 1.5 2.5 0.5 5.5 4.5 5.5 0.5 0.5 5.5 1.5 1.5 2.5 0.5 1.5 9.5
## [52] 3.5 3.5 5.5 0.5 8.5 5.5 7.5 1.5 4.5 0.5 2.5 7.5 6.5
```

```r
summary(aov(dij~speed, data = HW18))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## speed        3  111.5   37.18   9.542 3.04e-05 ***
## Residuals   60  233.8    3.90
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results: The Brown-Forsythe F test statistic= 9.54 with a p-value<0.001, therefore we reject the null hypothesis and concluded the variances are significantly different among different factor levels.

d) simple guide to decide the type of transformation

```r
# factor level mean
meani =  as.numeric(with(HW18, by( response, speed, mean)))
VARi = as.numeric( with(HW18, by( response, speed, var)))
list(mean=meani, sd= sqrt(VARi))
```

```
## $mean
## [1]  3.5625  5.8750 10.6875 16.5625
##
## $sd
## [1] 1.093542 1.995829 3.239727 5.378584
```

```r
Guide = data.frame(factor=1:4, Var.div.mean= VARi/meani, sd.div.mean= sqrt(VARi)/meani,
                    sd.div.meansq = sqrt(VARi)/meani^2)
round(Guide, 3)
```

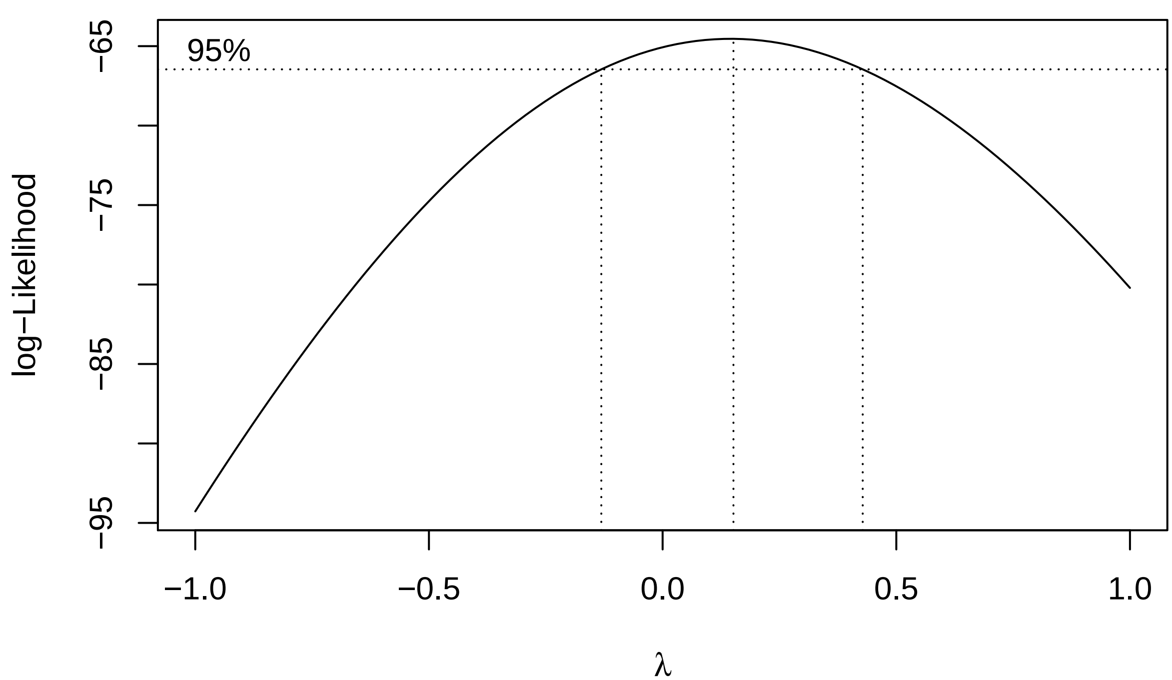
```
##   factor Var.div.mean sd.div.mean sd.div.meansq
## 1      1        0.336       0.307         0.086
## 2      2        0.678       0.340         0.058
## 3      3        0.982       0.303         0.028
## 4      4        1.747       0.325         0.020
```

Results: The sd/mean seems most stable, therefore, the simple guide suggests a log-transformation is a better choice.

e) Box-Cox

```r
library(MASS)

#  we can call boxcox function and use anova fitted model object
boxcox(fit,  lambda=seq(-1,1, by=0.1))
```



Results: The results suggest that the log-transformation appears to be reasonable, as $\lambda = 0$ is within the 95% CI of the optimal *lambda* that maximize the likelihood function,

HW 18.18 with additional problem (d) and (e)

(a) use log10 transformation then fit ANOVA model

```
fit2  = aov(log10(response)~ speed, data= HW18)
summary(fit2)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## speed        3  4.091   1.364   56.78 <2e-16 ***
## Residuals   60  1.441   0.024
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

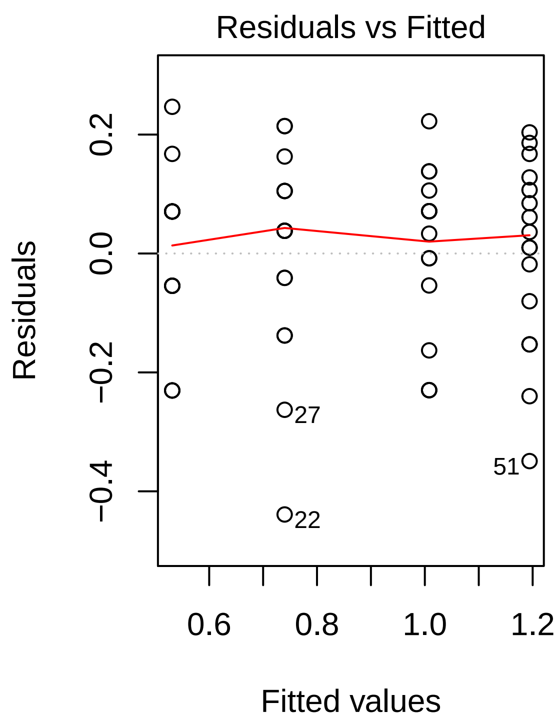
```
fit2$residuals
```

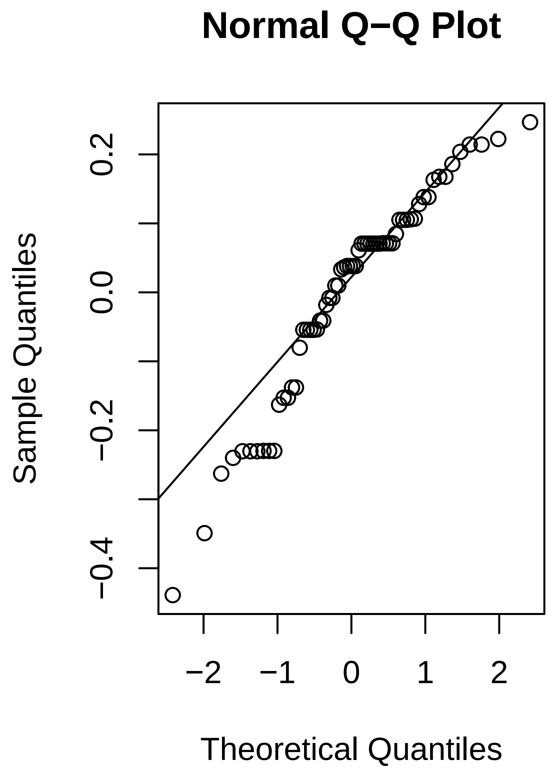
```
##            1            2            3            4            5
##  0.070615229 -0.054323508 -0.230414767 -0.054323508  0.070615229
##            6            7            8            9           10
##  0.070615229 -0.054323508  0.246706488  0.167525242  0.070615229
##           11           12           13           14           15
## -0.230414767  0.070615229  0.070615229 -0.230414767 -0.054323508
##           16           17           18           19           20
##  0.070615229  0.105117704  0.038170915 -0.137920344  0.038170915
##           21           22           23           24           25
##  0.105117704 -0.438950340  0.214262174 -0.041010331 -0.041010331
##           26           27           28           29           30
##  0.214262174 -0.262859081  0.163109651  0.038170915 -0.137920344
##           31           32           33           34           35
##  0.105117704  0.038170915  0.071146226 -0.229883769  0.138093016
##           36           37           38           39           40
##  0.071146226 -0.008035020 -0.053792510  0.071146226  0.222413902
##           41           42           43           44           45
## -0.162936980 -0.229883769  0.071146226  0.033357665 -0.229883769
##           46           47           48           49           50
##  0.105908333 -0.008035020  0.138093016  0.036198678 -0.018158985
##           51           52           53           54           55
## -0.349152204  0.106779752 -0.080306891 -0.152857559  0.009869739
##           56           57           58           59           60
##  0.203689765 -0.152857559  0.185960998  0.061022261  0.127969051
##           61           62           63           64
##  0.009869739  0.084503357 -0.240007734  0.167477592
```

Results: The ANOVA analysis of the log10 transformed responses shows that mean responses are significantly different among different speed levels.

(b) residual plots

```r
par(mfrow=c(1,2))
plot(fit2,1)
QQstat<- qqnorm(fit2$residuals) # same as plot(fit2, 2)
qqline(fit2$residuals)
```






```r
# correlation coefficient: first print the ordered residuals and the expected value under nomality
QQstat
```

```
## $x
##  [1]  0.13751340 -0.65010407 -1.27269864 -0.60244945  0.17716982
##  [6]  0.21710695 -0.55612559  2.41755902  1.27269864  0.25739353
## [11] -1.47346758  0.29810241  0.33931161 -1.36620382 -0.51096581
## [16]  0.38110545  0.65010407 -0.05878294 -0.80317257 -0.01958429
## [21]  0.69928330 -2.41755902  1.60100866 -0.42357608 -0.38110545
## [26]  1.76167041 -1.76167041  1.11319428  0.01958429 -0.75021538
## [31]  0.75021538  0.05878294  0.42357608 -1.18916435  0.97789754
## [36]  0.46682512 -0.29810241 -0.46682512  0.51096581  1.98742789
## [41] -0.97789754 -1.11319428  0.55612559 -0.13751340 -1.04315826
## [46]  0.80317257 -0.25739353  1.04315826 -0.09807215 -0.33931161
## [51] -1.98742789  0.85848447 -0.69928330 -0.91655667 -0.21710695
## [56]  1.47346758 -0.85848447  1.36620382  0.09807215  0.91655667
## [61] -0.17716982  0.60244945 -1.60100866  1.18916435
```

```
##
## $y
##           1           2           3           4           5
##  0.070615229 -0.054323508 -0.230414767 -0.054323508  0.070615229
##           6           7           8           9          10
##  0.070615229 -0.054323508  0.246706488  0.167525242  0.070615229
##          11          12          13          14          15
## -0.230414767  0.070615229  0.070615229 -0.230414767 -0.054323508
##          16          17          18          19          20
##  0.070615229  0.105117704  0.038170915 -0.137920344  0.038170915
##          21          22          23          24          25
##  0.105117704 -0.438950340  0.214262174 -0.041010331 -0.041010331
##          26          27          28          29          30
##  0.214262174 -0.262859081  0.163109651  0.038170915 -0.137920344
##          31          32          33          34          35
##  0.105117704  0.038170915  0.071146226 -0.229883769  0.138093016
##          36          37          38          39          40
##  0.071146226 -0.008035020 -0.053792510  0.071146226  0.222413902
##          41          42          43          44          45
## -0.162936980 -0.229883769  0.071146226  0.033357665 -0.229883769
##          46          47          48          49          50
##  0.105908333 -0.008035020  0.138093016  0.036198678 -0.018158985
##          51          52          53          54          55
## -0.349152204  0.106779752 -0.080306891 -0.152857559  0.009869739
##          56          57          58          59          60
##  0.203689765 -0.152857559  0.185960998  0.061022261  0.127969051
##          61          62          63          64
##  0.009869739  0.084503357 -0.240007734  0.167477592
```

```
cor(QQstat$x, QQstat$y)
```

```
## [1] 0.9704474
```

results: The log10 transformation has stablized the variance and now the variances are much similar for different levels of speed. For the residual normal QQ plot, it seems there are still some deviations, especially at the tails.

c) Use Brown-Forsythe test for transformed data

```
(mediani = with(HW18, by( log10(response), speed, median)))
```

```
## speed: 1
## [1] 0.60206
## --------------------------------------------------------
## speed: 2
## [1] 0.7781513
## --------------------------------------------------------
## speed: 3
## [1] 1.060287
## --------------------------------------------------------
## speed: 4
## [1] 1.217284
```

```r
(Factor.median = rep(as.numeric(mediani), rep(16,4)))
```

```
##  [1] 0.6020600 0.6020600 0.6020600 0.6020600 0.6020600 0.6020600 0.6020600
##  [8] 0.6020600 0.6020600 0.6020600 0.6020600 0.6020600 0.6020600 0.6020600
## [15] 0.6020600 0.6020600 0.7781513 0.7781513 0.7781513 0.7781513 0.7781513
## [22] 0.7781513 0.7781513 0.7781513 0.7781513 0.7781513 0.7781513 0.7781513
## [29] 0.7781513 0.7781513 0.7781513 0.7781513 1.0602870 1.0602870 1.0602870
## [36] 1.0602870 1.0602870 1.0602870 1.0602870 1.0602870 1.0602870 1.0602870
## [43] 1.0602870 1.0602870 1.0602870 1.0602870 1.0602870 1.0602870 1.2172845
## [50] 1.2172845 1.2172845 1.2172845 1.2172845 1.2172845 1.2172845 1.2172845
## [57] 1.2172845 1.2172845 1.2172845 1.2172845 1.2172845 1.2172845 1.2172845
## [64] 1.2172845
```

```r
(dij= abs(log10(HW18$response) -Factor.median))
```

```
##  [1] 0.00000000 0.12493874 0.30103000 0.12493874 0.00000000 0.00000000
##  [7] 0.12493874 0.17609126 0.09691001 0.00000000 0.30103000 0.00000000
## [13] 0.00000000 0.30103000 0.12493874 0.00000000 0.06694679 0.00000000
## [19] 0.17609126 0.00000000 0.06694679 0.47712125 0.17609126 0.07918125
## [25] 0.07918125 0.17609126 0.30103000 0.12493874 0.00000000 0.17609126
## [31] 0.06694679 0.00000000 0.01889428 0.28213572 0.08584107 0.01889428
## [37] 0.06028697 0.10604446 0.01889428 0.17016196 0.21518893 0.28213572
## [43] 0.01889428 0.01889428 0.28213572 0.05365639 0.06028697 0.08584107
## [49] 0.01316447 0.04119319 0.37218641 0.08374554 0.10334110 0.17589177
## [55] 0.01316447 0.18065556 0.17589177 0.16292679 0.03798805 0.10493484
## [61] 0.01316447 0.06146915 0.26304194 0.14444338
```

```r
summary(aov(dij~speed, data = HW18))
```

```
##             Df Sum Sq  Mean Sq F value Pr(>F)
## speed        3 0.0036 0.001214   0.098  0.961
## Residuals   60 0.7445 0.012408
```

Results: The F-stat= 0.098 with a p-value=0.96, we can't reject the hypothesis that the variances for different factor levels are the same.

(d) Use Hartley Test and Bartlett Test

```r
library(PMCMRplus)
hartleyTest(log10(response) ~ speed, data = HW18)
```

```
##
##  Hartley's maximum F-ratio test of homogeneity of variances
##
## data:  log10(response) by speed
## F Max = 1.5618, df = 15, k = 4, p-value = 0.8309
```

```r
bartlett.test(log10(response) ~ speed, data = HW18)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  log10(response) by speed
## Bartlett's K-squared = 0.93655, df = 3, p-value = 0.8166
```

Results: Both tests suggest similar results as those using Use Brown-Forsythe test in (c)

e) apply a nonparametric test

Solution 1:

```r
kruskal.test(response ~ speed, data = HW18)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  response by speed
## Kruskal-Wallis chi-squared = 48.025, df = 3, p-value = 2.104e-10
```

Solution 2: rank based F-test

```r
HW18$rank <- rank(HW18$response)
fit3 = aov( rank~ speed, data= HW18 )
summary(fit3)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## speed        3  16539    5513   64.14 <2e-16 ***
## Residuals   60   5157      86
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results: Using with Kruskal-Wallis test or the rank-based F-test is appropriate here. Both tests reject H0, suggesting the mean responses are different a mong d ifferent wi nding sp eed le vels. Th is is al so si milar to those results by applying the ANOVA analysis on the log-transformed variable.