# STAT4355HW8

Code

Ramesh Kanakala

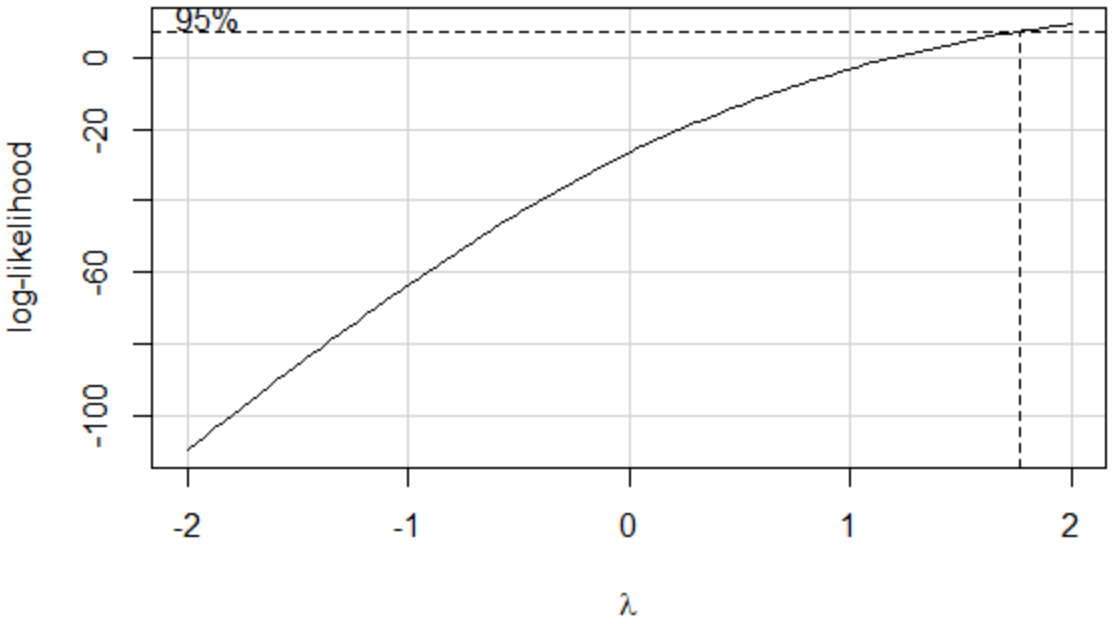
## Problem 1

### (a)

Hide

```
#load the data
wind <- read.csv(file = 'windmill.csv')

#Box-Cox transformation
library(car)
boxcox.wind <- boxCox(wind$y ~ wind$x,
                      lambda=seq(-2,2,1/10))
```




Hide

```
boxcox.wind # y=log-Likelihood, x=lambda
```

```
$x
  [1] -2.00000000 -1.95959596 -1.91919192 -1.87878788 -1.83838384 -1.79797980 -1.75757576
  [8] -1.71717172 -1.67676768 -1.63636364 -1.59595960 -1.55555556 -1.51515152 -1.47474747
 [15] -1.43434343 -1.39393939 -1.35353535 -1.31313131 -1.27272727 -1.23232323 -1.19191919
 [22] -1.15151515 -1.11111111 -1.07070707 -1.03030303 -0.98989899 -0.94949495 -0.90909091
 [29] -0.86868687 -0.82828283 -0.78787879 -0.74747475 -0.70707071 -0.66666667 -0.62626263
 [36] -0.58585859 -0.54545455 -0.50505051 -0.46464646 -0.42424242 -0.38383838 -0.34343434
 [43] -0.30303030 -0.26262626 -0.22222222 -0.18181818 -0.14141414 -0.10101010 -0.06060606
 [50] -0.02020202  0.02020202  0.06060606  0.10101010  0.14141414  0.18181818  0.22222222
 [57]  0.26262626  0.30303030  0.34343434  0.38383838  0.42424242  0.46464646  0.50505051
 [64]  0.54545455  0.58585859  0.62626263  0.66666667  0.70707071  0.74747475  0.78787879
 [71]  0.82828283  0.86868687  0.90909091  0.94949495  0.98989899  1.03030303  1.07070707
 [78]  1.11111111  1.15151515  1.19191919  1.23232323  1.27272727  1.31313131  1.35353535
 [85]  1.39393939  1.43434343  1.47474747  1.51515152  1.55555556  1.59595960  1.63636364
 [92]  1.67676768  1.71717172  1.75757576  1.79797980  1.83838384  1.87878788  1.91919192
 [99]  1.95959596  2.00000000

$y
  [1] -109.8312236 -107.8483826 -105.8729994 -103.9052962 -101.9454926  -99.9938055  -98.0504572
  [8]  -96.1156885  -94.1897450  -92.2728837  -90.3653709  -88.4674789  -86.5794976  -84.7017207
 [15]  -82.8344574  -80.9780278  -79.1327597  -77.2989995  -75.4770984  -73.6674246  -71.8703576
 [22]  -70.0862852  -68.3156138  -66.5587542  -64.8161334  -63.0881880  -61.3753615  -59.6781120
 [29]  -57.9969001  -56.3321966  -54.6844772  -53.0542196  -51.4419049  -49.8480142  -48.2730234
 [36]  -46.7174065  -45.1816283  -43.6661400  -42.1713869  -40.6977863  -39.2457450  -37.8156434
 [43]  -36.4078284  -35.0226340  -33.6603397  -32.3212071  -31.0054539  -29.7132478  -28.4447373
 [50]  -27.1999983  -25.9790850  -24.7819975  -23.6086819  -22.4590622  -21.3329900  -20.2302974
 [57]  -19.1507665  -18.0941361  -17.0601268  -16.0484072  -15.0586332  -14.0904308  -13.1434029
 [64]  -12.2171450  -11.3112337  -10.4252422   -9.5587431   -8.7113085   -7.8825174   -7.0719656
 [71]   -6.2792542   -5.5040146   -4.7458964   -4.0045719   -3.2797598   -2.5711910   -1.8786549
 [78]   -1.2019709   -0.5409962    0.1043449    0.7341102    1.3482782    1.9467913    2.5295425
 [85]    3.0963473    3.6469999    4.1812063    4.6986365    5.1989082    5.6815660    6.1461354
 [92]    6.5920658    7.0187801    7.4256639    7.8120570    8.1772911    8.5206716    8.8414961
 [99]    9.1390661    9.4126870
```

Hide

```
boxcox.wind$x[which.max(boxcox.wind$y)]
```

```
[1] 2
```

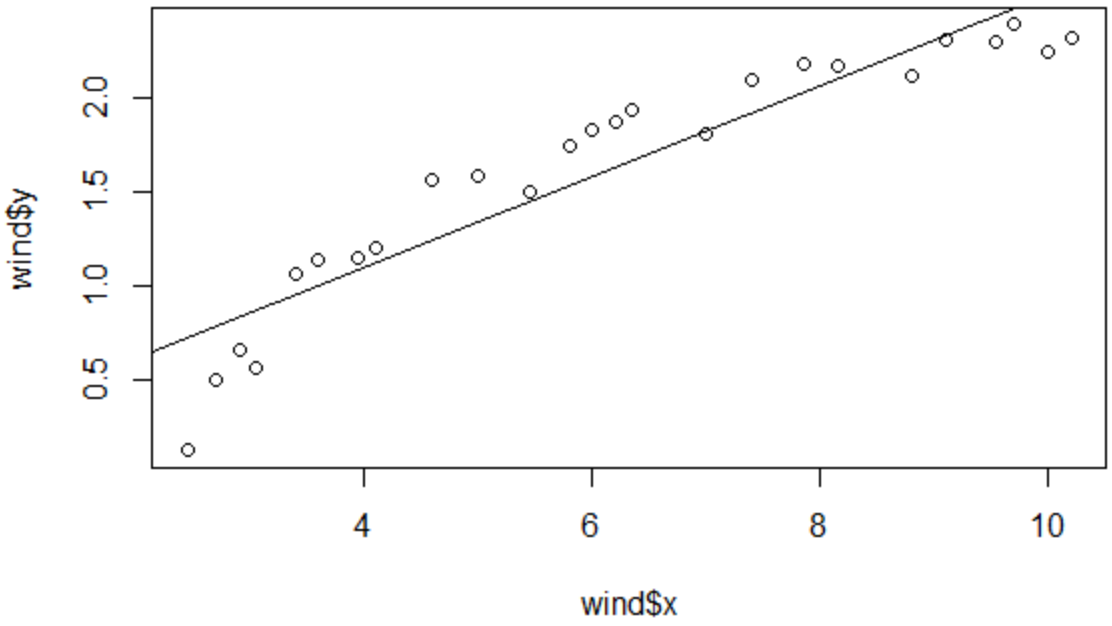
# (b)

Hide

```
#build linear model
lm1 <- lm(y~x, wind)

#scatterplot with regression line
plot(wind$x, wind$y) + abline(lm1)
```

```
integer(0)
```



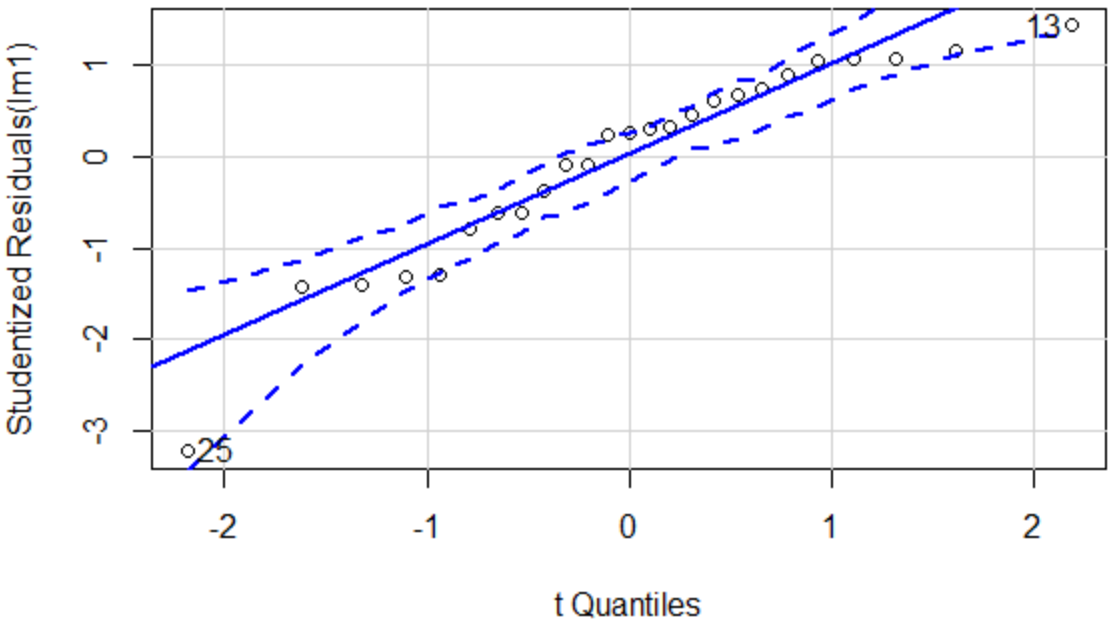

Hide

```
#normal probability plot
qqPlot(lm1)
```

```
[1] 13 25
```

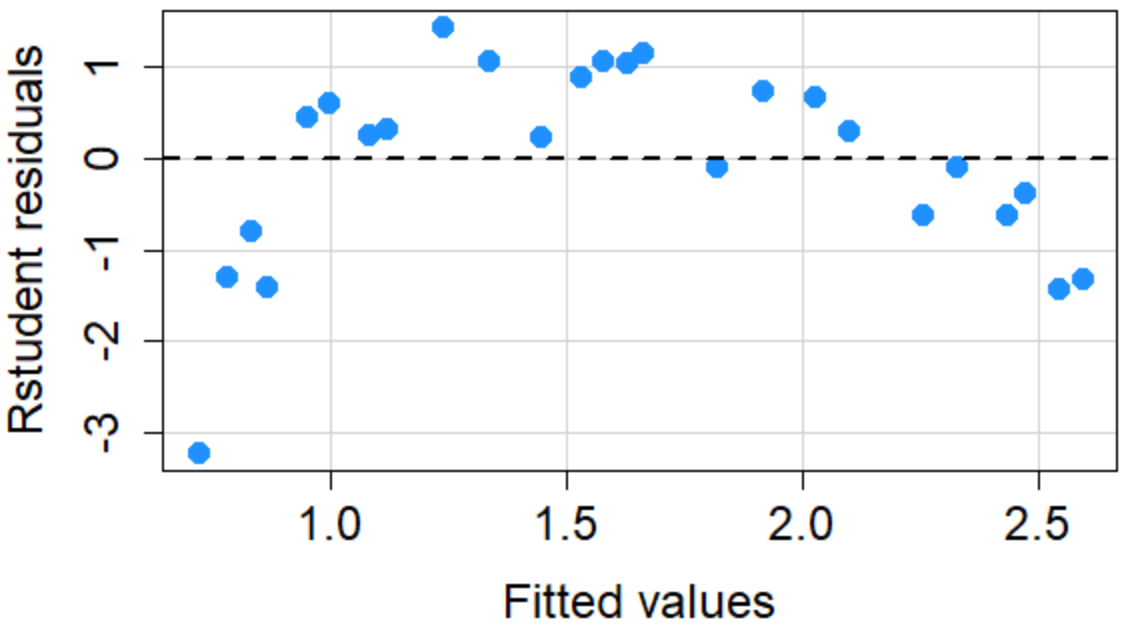
Hide

```
#studentized residuals versus the fitted values plot
par(mfrow=c(1,1))
```




Hide

```
residualPlot(lm1, type="rstudent", quadratic=F, col = "dodgerblue",
             pch=16, cex=1.5, cex.axis=1.5, cex.lab=1.5)
```




Hide

```
#influential analysis
myInf <- influence.measures(lm1)
myInf
```

```
Influence measures of
     lm(formula = y ~ x, data = wind) :
```

| | dfb.1_ <dbl> | dfb.x <dbl> | dffit <dbl> | cov.r <dbl> | cook.d <dbl> | hat <dbl> | inf <chr> |
|---|---|---|---|---|---|---|---|
| 1 | 0.1749174670 | -0.100094360 | 0.24091867 | 1.0378975 | 0.0288421683 | 0.04834508 | |
| 2 | 0.0916665937 | -0.011512178 | 0.21645112 | 1.0309042 | 0.0233028359 | 0.04011347 | |
| 3 | 0.1355473962 | -0.106977460 | 0.14443645 | 1.1761799 | 0.0107997414 | 0.08860703 | |
| 4 | -0.4550287954 | 0.379908437 | -0.46860871 | 1.0696651 | 0.1067252379 | 0.11670652 | |
| 5 | 0.3299085156 | -0.480183601 | -0.57029443 | 1.0610827 | 0.1555747283 | 0.13743403 | |
| 6 | 0.0764483382 | -0.114639354 | -0.13958124 | 1.2305302 | 0.0101202968 | 0.12290628 | |
| 7 | 0.1195456645 | -0.182390775 | -0.22529127 | 1.1941854 | 0.0260735760 | 0.11608199 | |
| 8 | -0.4538043800 | 0.369408522 | -0.47402926 | 1.0241531 | 0.1077543940 | 0.10185902 | |
| 9 | -0.0235933316 | 0.050527326 | 0.08002279 | 1.1613545 | 0.0033337440 | 0.06652040 | |
| 10 | 0.0741159018 | 0.005822197 | 0.21227680 | 1.0344342 | 0.0224520142 | 0.04003011 | |

Hide

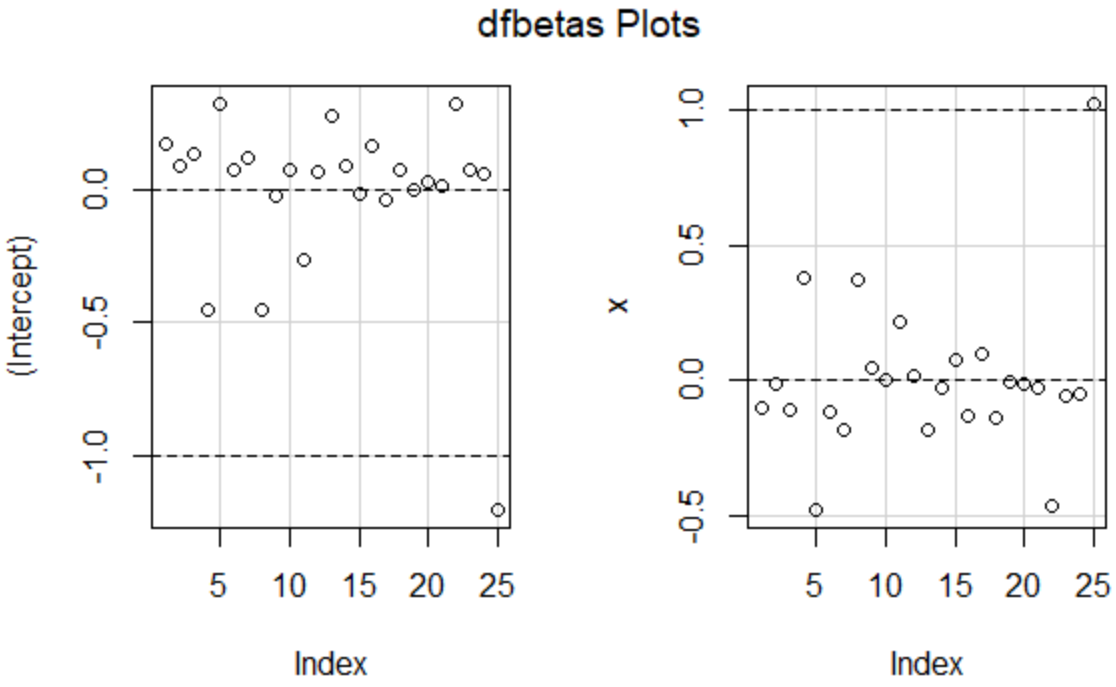
```
summary(myInf)
```

```
Potentially influential observations of
     lm(formula = y ~ x, data = wind) :

   dfb.1_  dfb.x   dffit   cov.r   cook.d hat
25 -1.21_*  1.03_* -1.24_*  0.58_*  0.54   0.13
```

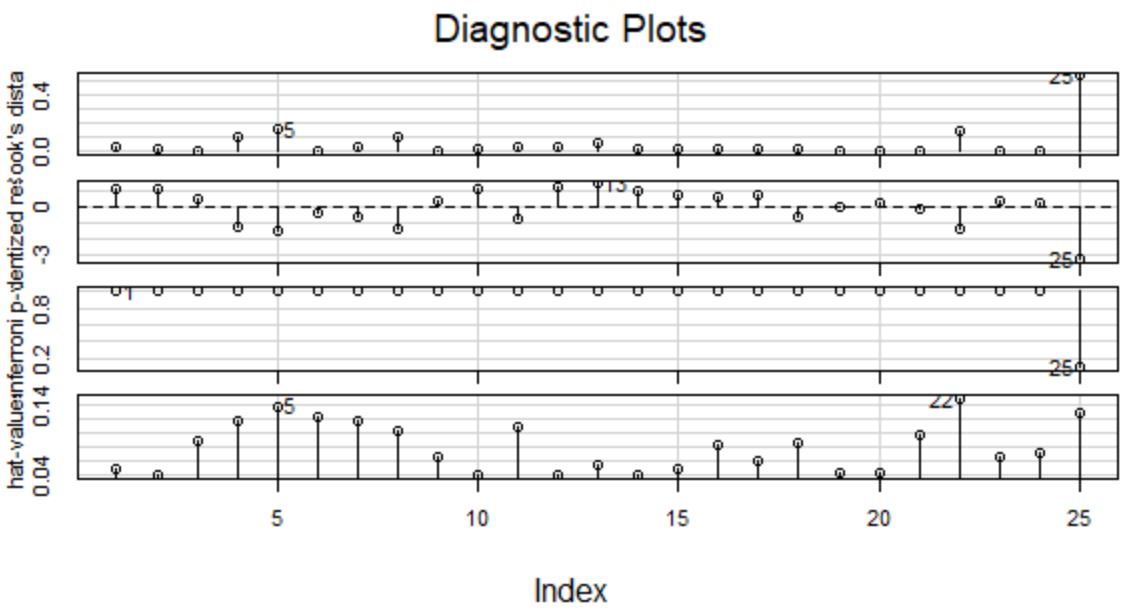
Hide

```
dfbetasPlots(lm1,intercept=T)
```




Hide

```
influenceIndexPlot(lm1)
```




# Problem 2

## (a)

Hide

```
#load the data
patient <- read.table('patient.txt', header=T)
x1 <- patient[,2] # age
x2 <- patient[,3] # severity
x3 <- patient[,4] # surgical-medical
x4 <- patient[,5] # anxiety
y <- patient[,6]  # satisfaction

x11=x12=x13=x14=array(0,length(y)) # n by 1 vector of 0's
x11[x1<30] <- 1          # if age group 1
x12[30<=x1 & x1<40] <- 1 # if age group 2
x13[40<=x1 & x1<50] <- 1 # if age group 3
x14[50<=x1 & x1<60] <- 1 # if age group 4

fit <- lm(y~x11+x12+x13+x14+x2+x3+x4)

#test for a general linear hypothesis
library(multcomp)
D <- matrix(0,3,8)
D[1,2]=D[2,3]=D[3,4]=1
D
```

```
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    0    1    0    0    0    0    0    0
[2,]    0    0    1    0    0    0    0    0
[3,]    0    0    0    1    0    0    0    0
```

Hide

```
d <- c(0,0,0)
mytest <- glht(fit, linfct=D, rhs=d)  # from page 15
summary(mytest,test=Ftest())
```

```
     General Linear Hypotheses

Linear Hypotheses:
       Estimate
1 == 0    48.71
2 == 0    49.14
3 == 0    42.14

Global Test:
```

| F | DF1 | DF2 | Pr(>F) |
|---|---|---|---|
| <dbl> | <int> | <int> | <dbl> |
| 11.78488 | 3 | 17 | 0.0002035494 |

1 row

i.

H0: β11 = β12 = β13 = 0

H1: β11 or β12 or β13 /= 0

ii.

D = 01000000 00100000 00010000

d = 000

iii.

As F = 11.78488 and p = 0.0002035494 < 0.05, we reject the null hypothesis and confirm either β11 or β12 or β13 does not equal zero.

# (b)

Hide

```
x11=x12=array(0,length(y))
x11[x1<50] <- 1         # if age group 1
x12[50<=x1 & x1<60] <- 1 # if age group 2

fit <- lm(y~x11+x12+x2+x3+x4)
summary(fit)
```

```
Call:
lm(formula = y ~ x11 + x12 + x2 + x3 + x4)

Residuals:
    Min      1Q  Median      3Q     Max
-13.157  -7.407  -1.525   6.006  16.002

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  38.8757    15.5052   2.507 0.021410 *
x11          48.8871     8.1199   6.021 8.59e-06 ***
x12          32.6479     8.3719   3.900 0.000963 ***
x2           -0.4358     0.2006  -2.172 0.042695 *
x3           -1.7310     4.0850  -0.424 0.676501
x4            4.9538     2.0958   2.364 0.028907 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.902 on 19 degrees of freedom
Multiple R-squared:  0.828, Adjusted R-squared:  0.7827
F-statistic: 18.29 on 5 and 19 DF,  p-value: 1.123e-06
```

x11: With all other variables remaining the same, if a patient is in the age group under 50, satisfaction is associated with a 48.8871 increase on average.

x12: With all other variables remaining the same, if a patient is in the age group between 50 and 60, satisfaction is associated with a 32.6479 increase on average.

x2: With all other variables remaining the same, for a one unit increase in severity, satisfaction is associated with a 0.4358 decrease on average.

x3: With all other variables remaining the same, if a patient is a medical patient, satisfaction is associated with a 1.7310 decrease on average.

x4: With all other variables remaining the same, for a one unit increase in anxiety, satisfaction is associated with a 4.9538 increase on average.

# R code is within this notebook