Stats 101B Code and Outputs

Loading data and creating a new variable that is the mean of the two repetitions. I created two dataframes for clarity between all three analysis.

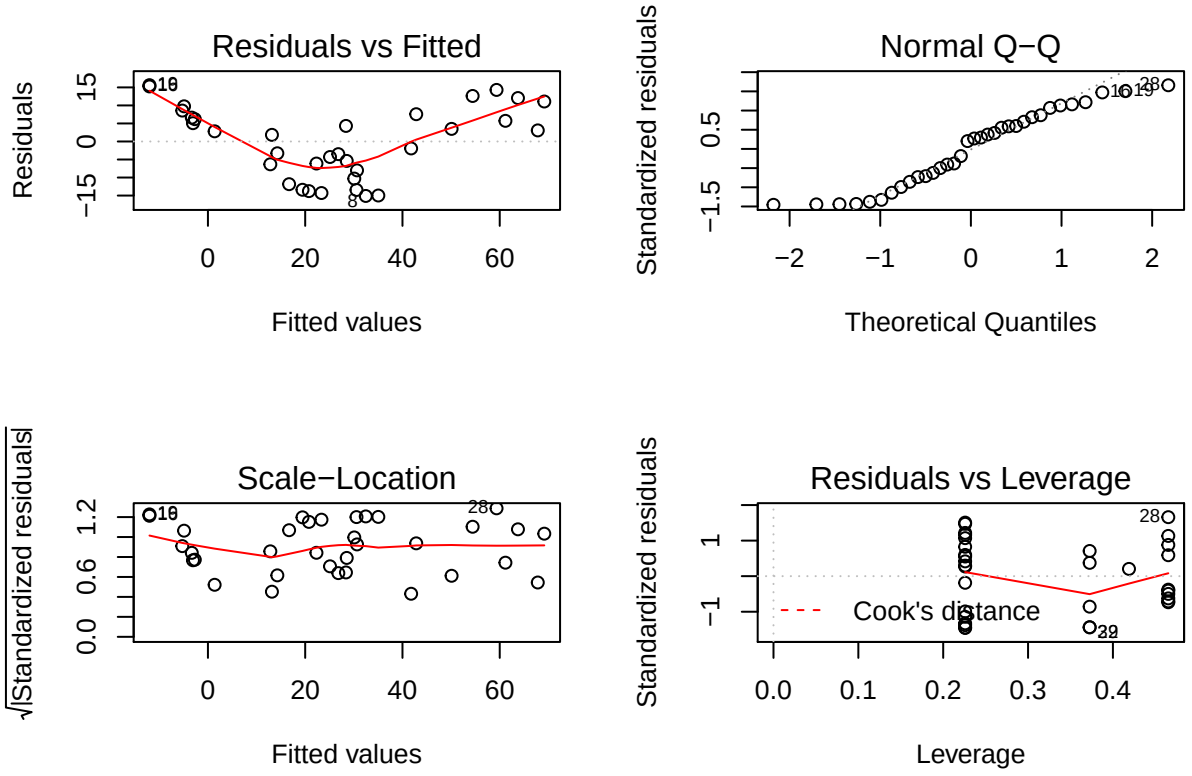
```
library(leaps)
library(car)
setwd("~/vmshare/ucla/classes/stats101b")
herpes <- read.table("hsv34.txt")
herpes$rep.avg <- (herpes$rep1 + herpes$rep2) / 2
cols <- c("A" ,"B", "C", "D", "E")
herpes[cols] <- lapply(herpes[cols], factor)

herpes16 <- herpes[1:16,]
herpes18 <- herpes[-(1:16),]
```

All 36 trials

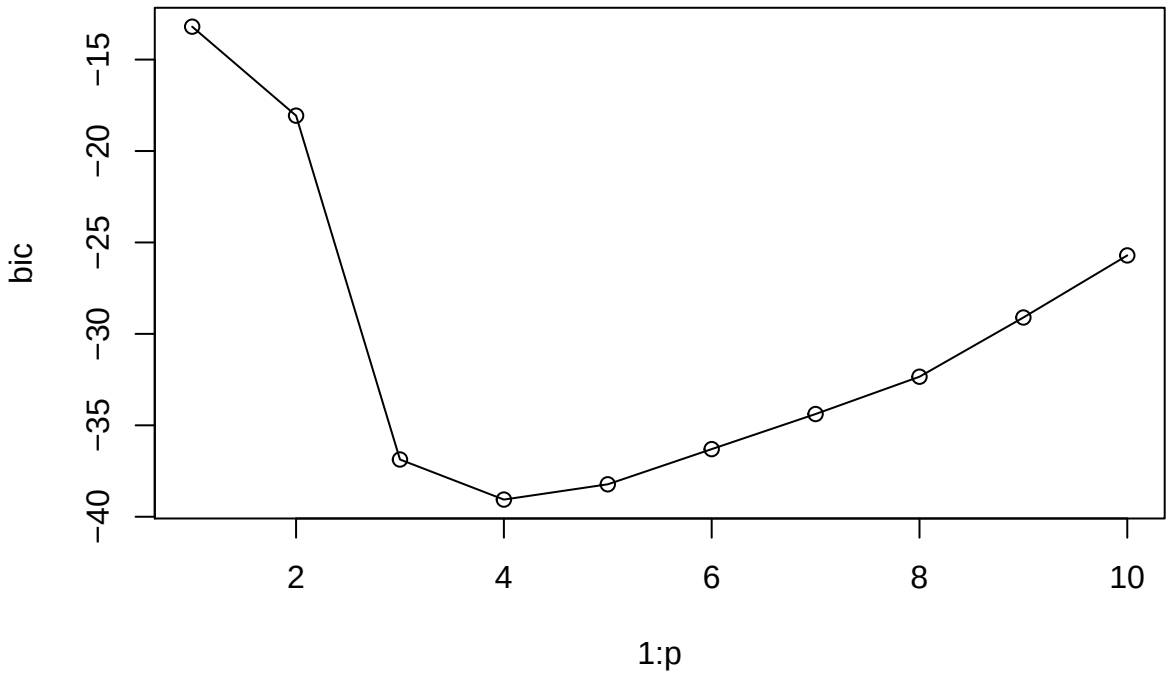
```
m1 <- lm(rep.avg ~ A + B + C + D + E, herpes)
summary(m1)
```

```
##
## Call:
## lm(formula = rep.avg ~ A + B + C + D + E, data = herpes)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.108  -7.589   2.329   7.332  15.620
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   69.162      5.603  12.343 1.26e-11 ***
## A0            -6.032      5.997  -1.006   0.3249
## A1            -1.336      4.456  -0.300   0.7671
## B0            -5.821      5.997  -0.971   0.3418
## B1            -7.989      4.456  -1.793   0.0862 .
## C0            -2.657      5.997  -0.443   0.6618
## C1            -5.411      4.456  -1.214   0.2370
## D0           -12.311      5.997  -2.053   0.0516 .
## D1           -39.068      4.456  -8.767 8.61e-09 ***
## E0           -29.178      5.997  -4.866 6.51e-05 ***
## E1           -27.379      4.456  -6.144 2.88e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.79 on 23 degrees of freedom
## Multiple R-squared:   0.85,  Adjusted R-squared:  0.7848
## F-statistic: 13.03 on 10 and 23 DF,  p-value: 2.928e-07
```

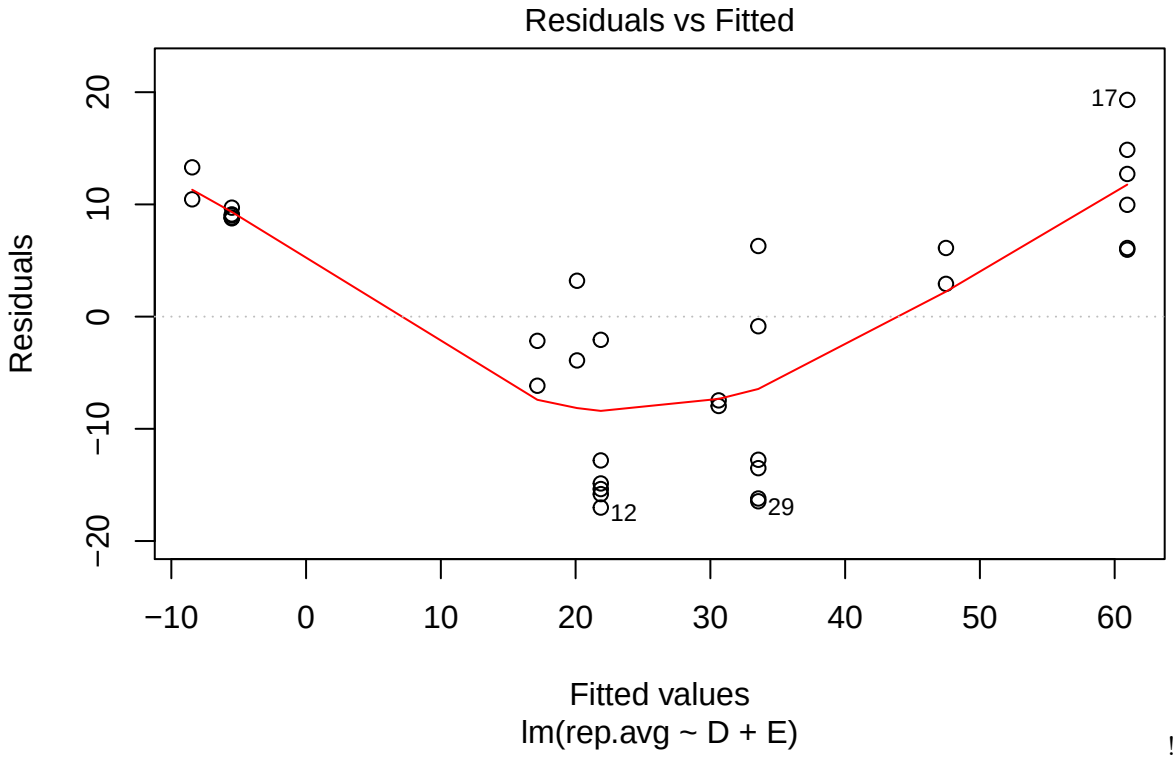
```
par(mfrow=c(2,2))
plot(m1)
```




```
herpes <- herpes[, -c(6,7)]
backwards <- regsubsets(rep.avg ~.,data=herpes,method="backward",nvmax=10)
backwards <- summary(backwards)
bic <- backwards$bic
p <- length(bic)
par(mfrow=c(1,1))
plot(1:p, bic)
lines(1:p, bic)
```




```
m2 <- lm(rep.avg ~ D + E, herpes)
plot(m2)
```



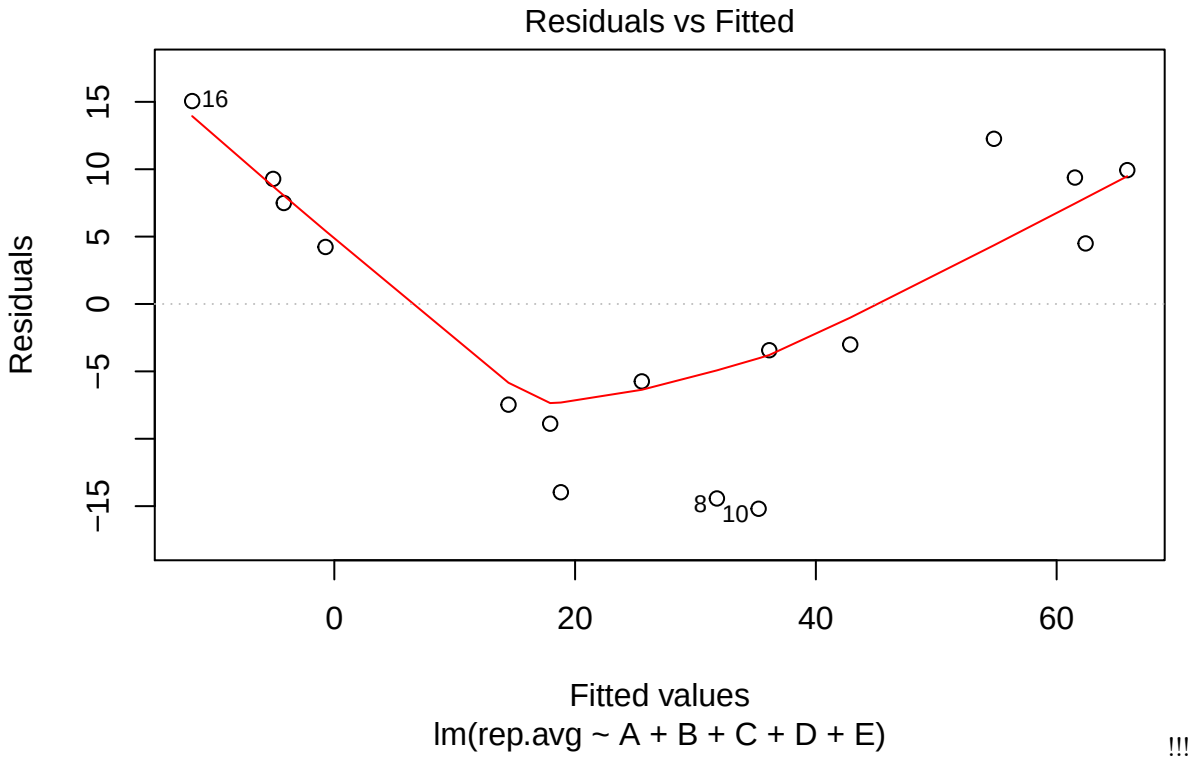
!!!

First 16 trials

```r
herpes16 <- herpes[1:16,]
m3 <-  lm(rep.avg ~ A + B + C + D + E, herpes16)
summary(m3)
```

```
##
## Call:
## lm(formula = rep.avg ~ A + B + C + D + E, data = herpes16)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.1937  -7.8219   0.6125   9.3062  15.0562
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   67.500      7.665   8.807 5.03e-06 ***
## A1            -5.981      6.258  -0.956  0.36174
## B1            -5.094      6.258  -0.814  0.43464
## C1            -1.631      6.258  -0.261  0.79965
## D1           -41.956      6.258  -6.704 5.34e-05 ***
## E1           -24.644      6.258  -3.938  0.00278 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.52 on 10 degrees of freedom
## Multiple R-squared:  0.8613, Adjusted R-squared:  0.7919
## F-statistic: 12.42 on 5 and 10 DF,  p-value: 0.000501
```

```r
plot(m3)
```



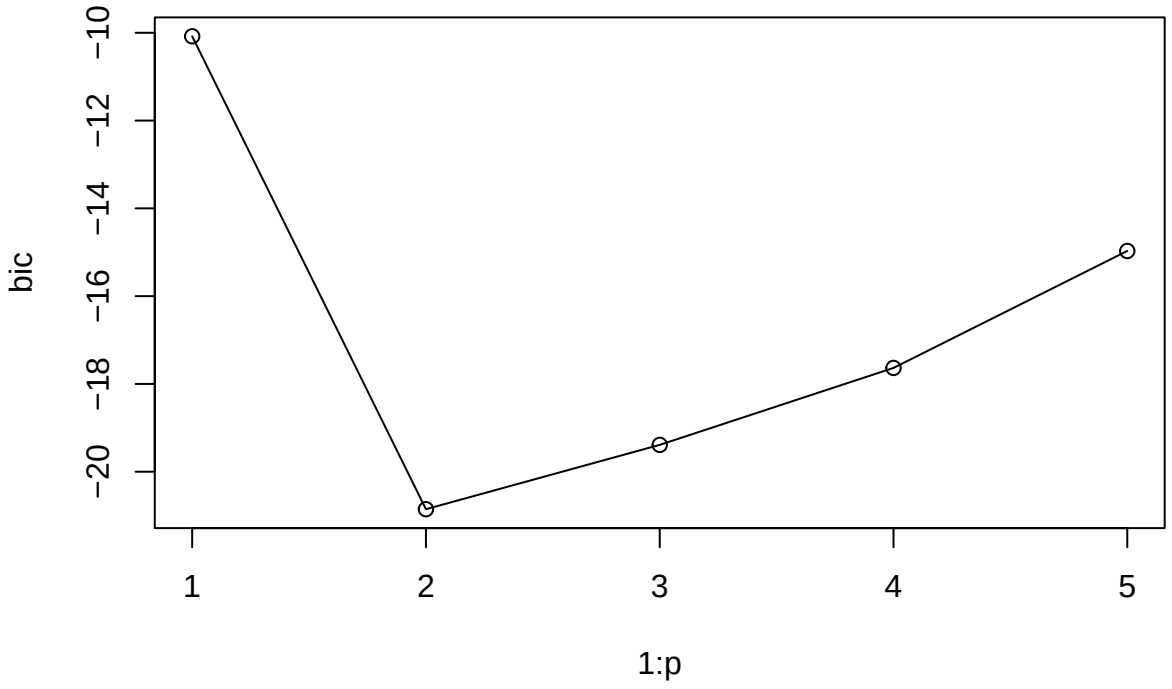
!!!

```
backwards <- regsubsets(rep.avg ~. ,data=herpes16 ,method="backward" ,nvmax=10)
```

```
## Warning in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax,
## force.in = force.in, : 5 linear dependencies found
```

```
## Reordering variables and trying again:
```

```
backwards <- summary(backwards)
bic <- backwards$bic
p <- length(bic)
par(mfrow=c(1,1))
plot(1:p, bic)
lines(1:p, bic)
```

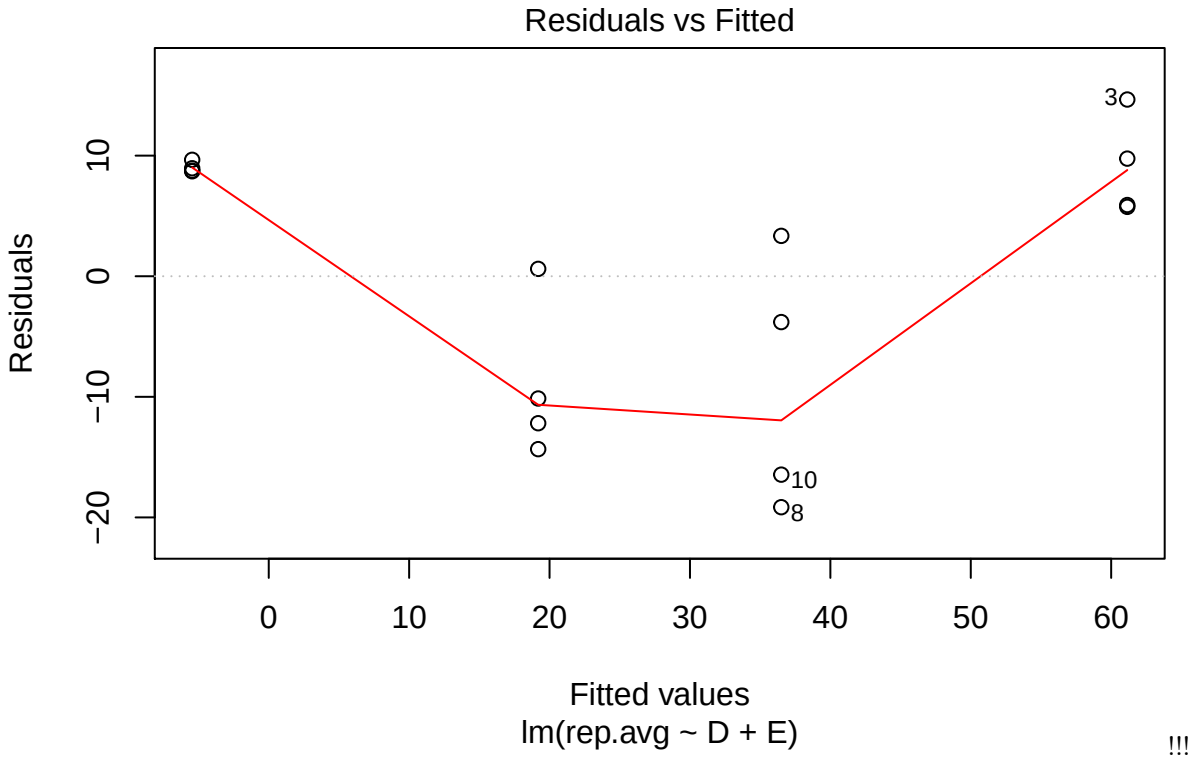



```
m4 <- lm(rep.avg ~ D + E, herpes16)
summary(m4)
```

```
##
## Call:
## lm(formula = rep.avg ~ D + E, data = herpes16)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.153 -10.653   4.550   8.803  14.653
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   61.147      5.129  11.921 2.26e-08 ***
```

```
## D1           -41.956      5.923  -7.084 8.25e-06 ***
## E1           -24.644      5.923  -4.161  0.00112 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.85 on 13 degrees of freedom
## Multiple R-squared:  0.8385, Adjusted R-squared:  0.8136
## F-statistic: 33.75 on 2 and 13 DF,  p-value: 7.132e-06
```

```
plot(m4)
```



!!!

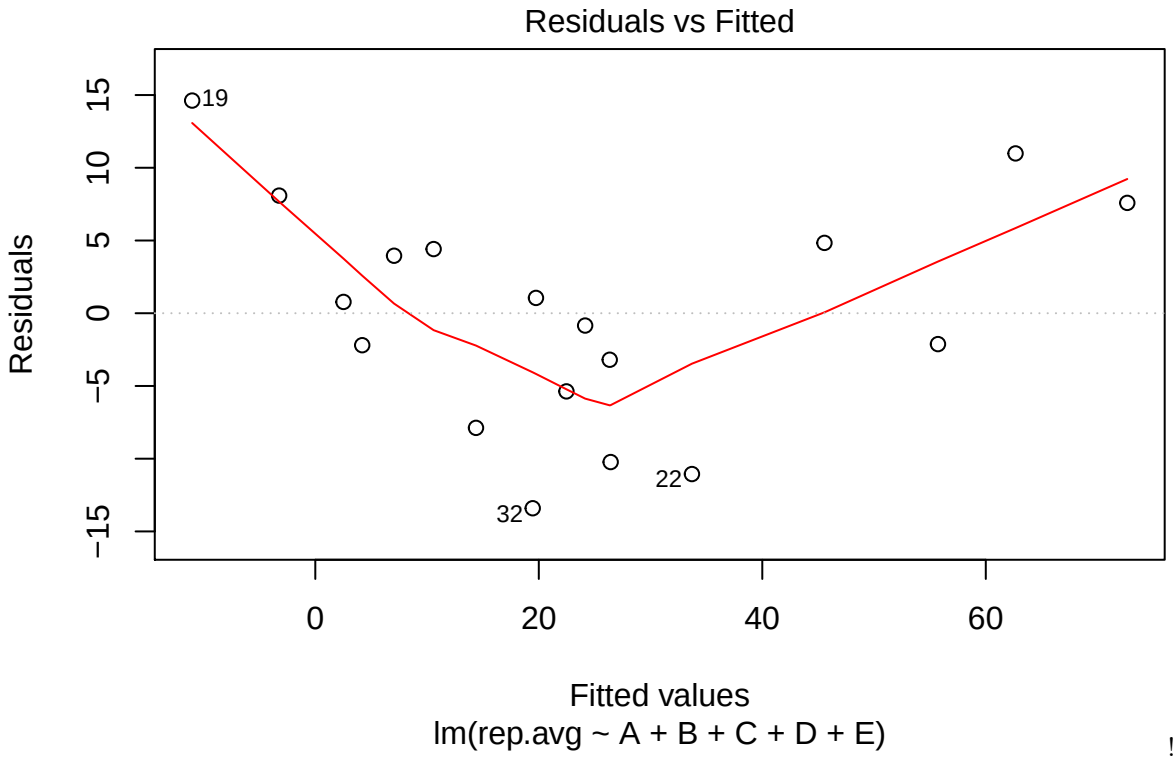
Last 18 trials

```
herpes18 <- herpes[-(1:16),]

m5 <-  lm(rep.avg ~ A + B + C + D + E, herpes18)
summary(m5)
```

```
##
## Call:
## lm(formula = rep.avg ~ A + B + C + D + E, data = herpes18)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.4056  -4.8285  -0.0347   4.7299  14.6194
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   72.664      9.507   7.643 0.000122 ***
```

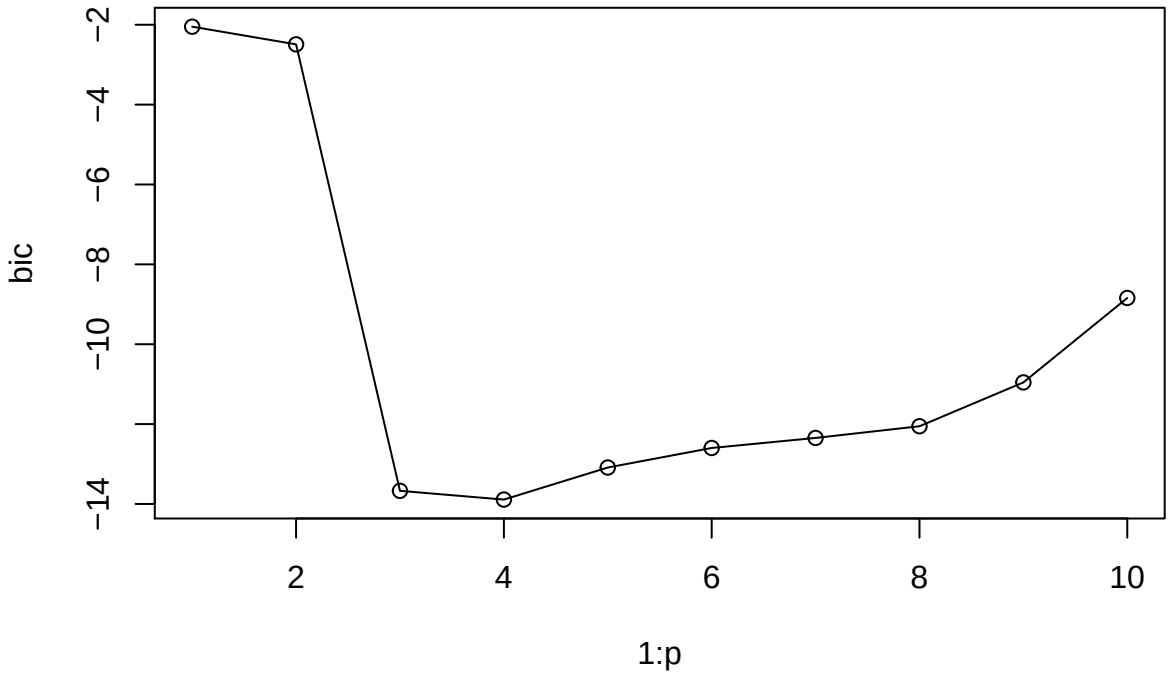
```
## A0            -3.900      7.022  -0.555 0.595900
## A1             4.858      7.022   0.692 0.511285
## B0            -8.717      7.022  -1.241 0.254441
## B1           -11.850      7.022  -1.688 0.135338
## C0            -6.142      7.022  -0.875 0.410754
## C1           -10.450      7.022  -1.488 0.180287
## D0           -11.350      7.022  -1.616 0.150035
## D1           -35.217      7.022  -5.015 0.001538 **
## E0           -31.967      7.022  -4.553 0.002628 **
## E1           -31.025      7.022  -4.418 0.003087 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.16 on 7 degrees of freedom
## Multiple R-squared:  0.8954, Adjusted R-squared:  0.746
## F-statistic: 5.992 on 10 and 7 DF,  p-value: 0.01329
```

```
plot(m5)
```



!!!

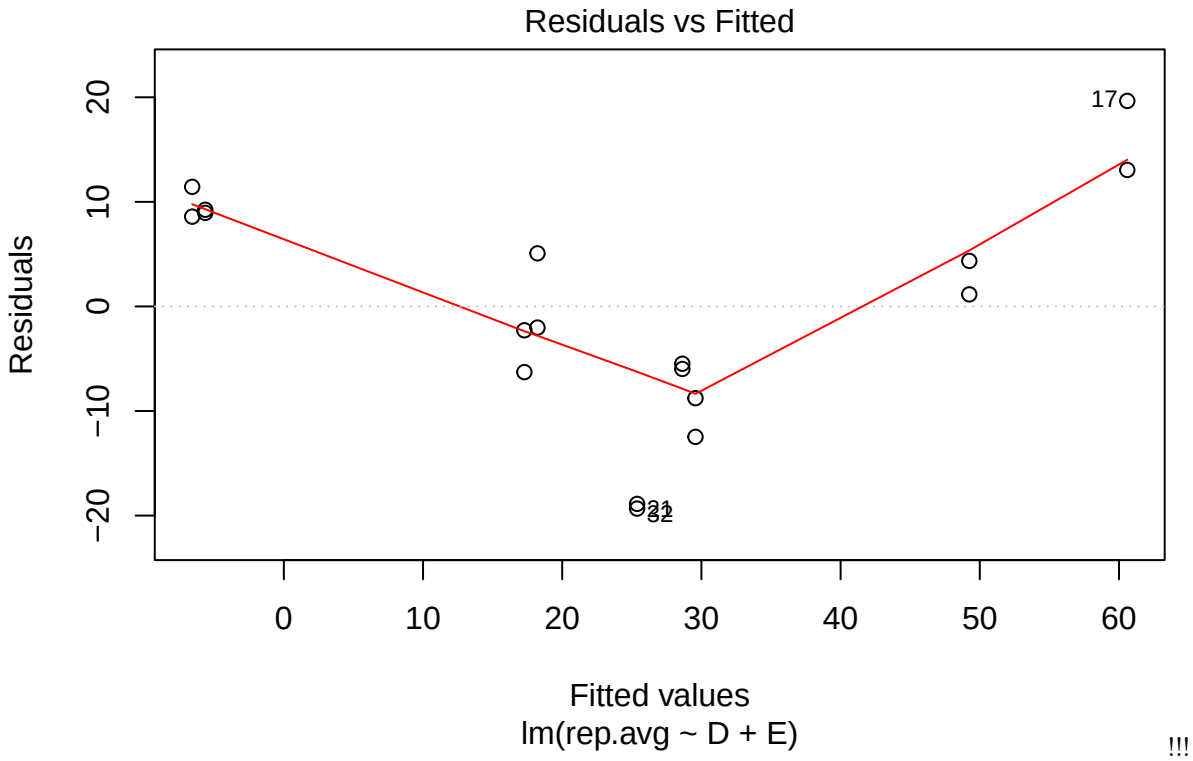
```
backwards <- regsubsets(rep.avg ~. ,data=herpes18 ,method="backward" ,nvmax=10)
backwards <- summary(backwards)
bic <- backwards$bic
p <- length(bic)
par(mfrow=c(1,1))
plot(1:p, bic)
lines(1:p, bic)
```




```
m6 <- lm(rep.avg ~ D + E, herpes18)
summary(m6)
```

```
##
## Call:
## lm(formula = rep.avg ~ D + E, data = herpes18)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.3306  -6.2056  -0.4347   8.8549  19.6528
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   60.597      6.618   9.156 4.95e-07 ***
## D0           -11.350      7.250  -1.566 0.141461
## D1           -35.217      7.250  -4.858 0.000313 ***
## E0           -31.967      7.250  -4.409 0.000705 ***
## E1           -31.025      7.250  -4.279 0.000897 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.56 on 13 degrees of freedom
## Multiple R-squared:  0.7929, Adjusted R-squared:  0.7292
## F-statistic: 12.44 on 4 and 13 DF,  p-value: 0.000221
```

```
plot(m6)
```


Residuals vs Fitted
Residuals
Fitted values
lm(rep.avg ~ D + E)
17
21
32
20
10
0
−10
−20
0
10
20
30
40
50
60

!!!