STA 108 Final Project: An analysis of crime prevalence using 5 socioeconomic variables for 4 geographic regions of the United States.

Import and format data:

```
library(MASS)
data_demo <- read.table("Demographic.txt")
names(data_demo) = c('id','county','state','land_area','total_population','percent_population_18to34','
```

Question 1:

Generate Regression Models for each geographic region:

```
model_demo <- lm(crime ~ per_cap_income + unemployment + below_poverty + bachelor + highschool, data=da
```

```
regional_model_generator = function(data_demo,region){
  loader <- data_demo[data_demo$geo_region == region,]
  y <- as.numeric(loader$crime)
  x1 <- as.numeric(loader$per_cap_income)
  x2 <- as.numeric(loader$unemployment)
  x3 <- as.numeric(loader$below_poverty)
  x4 <- as.numeric(loader$bachelor)
  x5 <- as.numeric(loader$highschool)
  output = lm(y ~ x1+x2+x3+x4+x5, data=data_demo)
  return(output)
}

loader1 = regional_model_generator(data_demo= data_demo,region =1)
loader2 = regional_model_generator(data_demo= data_demo,region =2)
loader3 = regional_model_generator(data_demo= data_demo,region =3)
loader4 = regional_model_generator(data_demo= data_demo,region =4)
```

```
loader1_data = subset(data_demo, geo_region ==1)
loader2_data = subset(data_demo, geo_region ==2)
loader3_data = subset(data_demo, geo_region ==3)
loader4_data = subset(data_demo, geo_region ==4)
```

Exploratory Data Analysis:

The initial regression model for the demographic dataset has the coefficients Y = -28798.161 + 4.729 - 815.696 + 3982.371 + 254.546 - 858.541. These negative and positive values for each coefficient responds to the real world relationship between these variables, with the 4.729 coefficient for per capita income indicating a positive relationship between serious crimes and amount of per capita income. Likewise, the negative coefficients for unemployment and prevalence of high school degrees indicate a negative relationship between serious crimes and these factors. A negative relationship in this situation indicates that a increase in the unemployment rate and decrease in the prevalence of high school degrees would lead to an increase in serious crimes.

The residual plots for each region all display fairly similar results, with the median centered around 0 and all are quite right skewed due to a few extreme outliers.

```
summary(loader1)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5, data = data_demo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -127848  -16268   -3728   13978  479383
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  47040.612 153904.343   0.306    0.761
## x1               3.670      2.336   1.571    0.119
## x2           -2934.758   4115.950  -0.713    0.478
## x3           13356.157   2742.171   4.871 4.32e-06 ***
## x4            3101.042   1887.071   1.643    0.104
## x5           -3000.954   1862.062  -1.612    0.110
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 56930 on 97 degrees of freedom
## Multiple R-squared:  0.3355, Adjusted R-squared:  0.3012
## F-statistic: 9.794 on 5 and 97 DF,  p-value: 1.341e-07
```

```
summary(loader2)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5, data = data_demo)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -67551 -14420  -4908   7774 347455
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  55890.278 146968.941   0.380 0.704524
## x1               8.493      2.236   3.798 0.000248 ***
## x2           -1110.576   2736.906  -0.406 0.685756
## x3            5350.915   2043.762   2.618 0.010186 *
## x4            1229.761   1320.021   0.932 0.353732
## x5           -3138.775   1742.387  -1.801 0.074591 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 40770 on 102 degrees of freedom
## Multiple R-squared:  0.2934, Adjusted R-squared:  0.2588
## F-statistic: 8.472 on 5 and 102 DF,  p-value: 9.81e-07
```

```
summary(loader3)
```

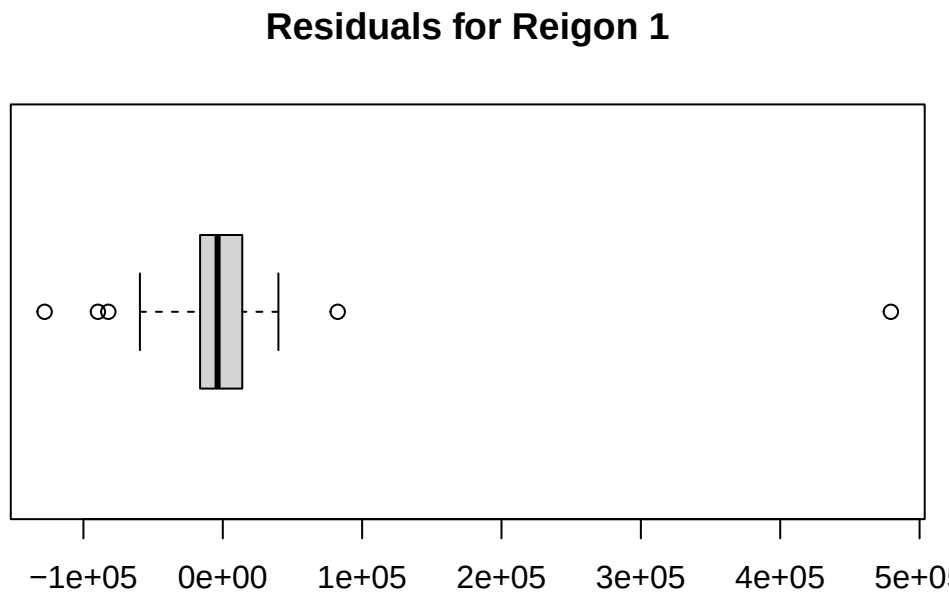
```
##
```

```
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5, data = data_demo)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -61813 -14846  -7166   2314 215373
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -77889.315  63697.010  -1.223  0.22337
## x1               4.047      1.352   2.993  0.00324 **
## x2            1721.363   2080.442   0.827  0.40936
## x3            1983.880    924.544   2.146  0.03354 *
## x4              -7.352    755.049  -0.010  0.99224
## x5              35.081    763.520   0.046  0.96342
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36800 on 146 degrees of freedom
## Multiple R-squared:  0.1132, Adjusted R-squared:  0.08287
## F-statistic: 3.729 on 5 and 146 DF,  p-value: 0.003313
```

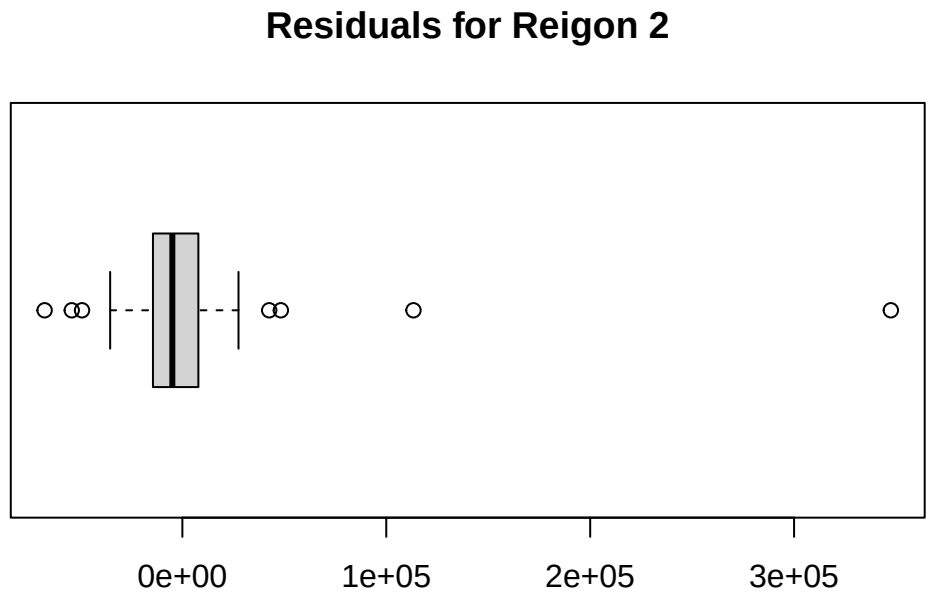
```
summary(loader4)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5, data = data_demo)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -91456 -35694  -6947  12592 580752
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 735587.395 366561.289   2.007   0.0486 *
## x1              1.170      4.464   0.262   0.7939
## x2          -9009.014   4879.278  -1.846   0.0690 .
## x3          -4101.237   5517.308  -0.743   0.4597
## x4           3144.073   2918.537   1.077   0.2850
## x5          -8602.818   3722.965  -2.311   0.0238 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 81060 on 71 degrees of freedom
## Multiple R-squared:  0.128,  Adjusted R-squared:  0.06661
## F-statistic: 2.085 on 5 and 71 DF,  p-value: 0.07732
```

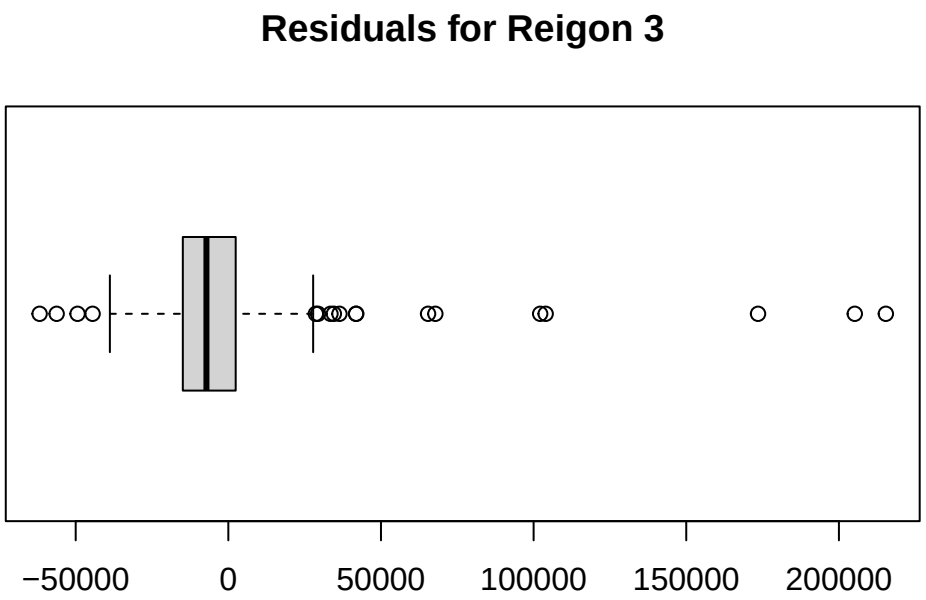
```
boxplot(loader1$residuals, horizontal = T, main = "Residuals for Reigon 1")
```



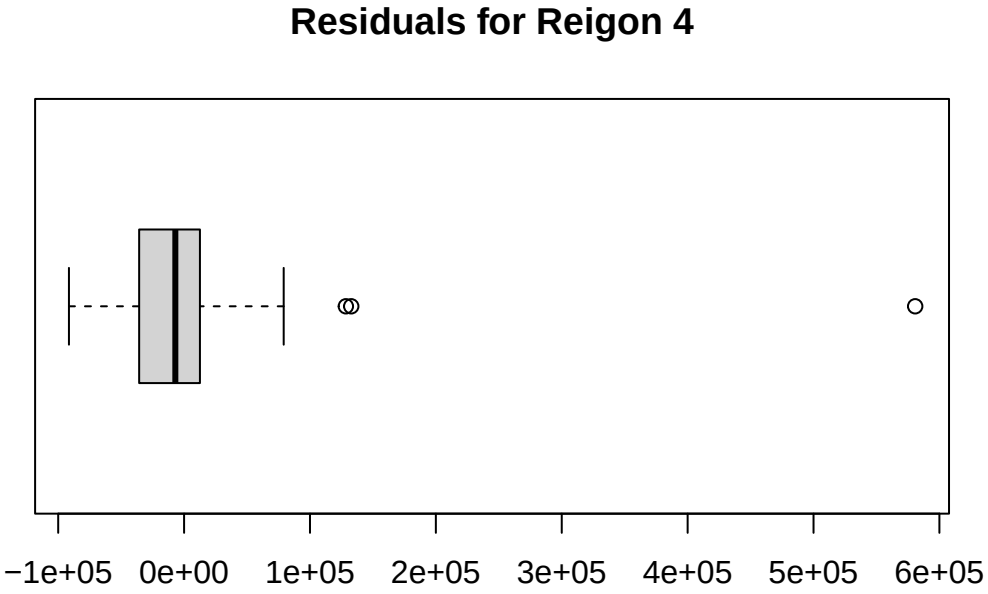

```
boxplot(loader2$residuals, horizontal = T, main = "Residuals for Reigon 2")
```




```
boxplot(loader3$residuals, horizontal = T, main = "Residuals for Reigon 3")
```




```
boxplot(loader4$residuals, horizontal = T, main = "Residuals for Reigon 4")
```



Question 2:

Functions to re-initialize variables and construct AIC and BIC models:

```
AICfunction <- function(data_demo, region){
  loader <- data_demo[data_demo$geo_region == region,]
  library(MASS)
  y <- as.numeric(loader$crime)
  x1 <- as.numeric(loader$per_cap_income)
  x2 <- as.numeric(loader$unemployment)
  x3 <- as.numeric(loader$below_poverty)
  x4 <- as.numeric(loader$bachelor)
  x5 <- as.numeric(loader$highschool)
  loader_mod <- lm(y~x1+x2+x3+x4+x5, data = loader)
```

```
  AIC_loader <- stepAIC(loader_mod, k =2)
  return(AIC_loader)
}
```

```
BICfunction <- function(data_demo, region){
  loader <- data_demo[data_demo$geo_region == region,]
  library(MASS)
  y <- as.numeric(loader$crime)
  x1 <- as.numeric(loader$per_cap_income)
  x2 <- as.numeric(loader$unemployment)
  x3 <- as.numeric(loader$below_poverty)
  x4 <- as.numeric(loader$bachelor)
  x5 <- as.numeric(loader$highschool)
  loader_mod <- lm(y~x1+x2+x3+x4+x5, data = loader)
  BIC_loader <- stepAIC(loader_mod, k =log(nrow(loader)))
  return(BIC_loader)
}
```

Construct AIC and BIC for each model per geographic region:

```
AIC_loader1 = AICfunction(data_demo, 1)
```

```
## Start:  AIC=2261.42
## y ~ x1 + x2 + x3 + x4 + x5
##
##        Df  Sum of Sq        RSS    AIC
## - x2    1 1.6475e+09 3.1598e+11 2259.9
## <none>               3.1433e+11 2261.4
## - x1    1 7.9952e+09 3.2233e+11 2262.0
## - x5    1 8.4169e+09 3.2275e+11 2262.1
## - x4    1 8.7510e+09 3.2309e+11 2262.2
## - x3    1 7.6877e+10 3.9121e+11 2281.9
##
## Step:  AIC=2259.95
## y ~ x1 + x3 + x4 + x5
##
##        Df  Sum of Sq        RSS    AIC
## <none>               3.1598e+11 2259.9
## - x1    1 7.8106e+09 3.2379e+11 2260.5
## - x5    1 8.3993e+09 3.2438e+11 2260.7
## - x4    1 1.0755e+10 3.2674e+11 2261.4
## - x3    1 7.6837e+10 3.9282e+11 2280.4
```

```
AIC_loader1$coefficients
```

```
##  (Intercept)           x1           x3           x4           x5
## 24461.535401     3.625788 12798.923610  3369.023096 -2997.812291
```

```
AIC_loader2 = AICfunction(data_demo, 2)
```

```
## Start:  AIC=2298.8
## y ~ x1 + x2 + x3 + x4 + x5
##
##        Df  Sum of Sq        RSS    AIC
## - x2    1 2.7364e+08 1.6979e+11 2297.0
```

```
## - x4    1 1.4424e+09 1.7095e+11 2297.7
## <none>               1.6951e+11 2298.8
## - x5    1 5.3930e+09 1.7491e+11 2300.2
## - x3    1 1.1392e+10 1.8090e+11 2303.8
## - x1    1 2.3976e+10 1.9349e+11 2311.1
##
## Step:  AIC=2296.97
## y ~ x1 + x3 + x4 + x5
##
##        Df  Sum of Sq        RSS    AIC
## - x4    1 2.7082e+09 1.7249e+11 2296.7
## <none>               1.6979e+11 2297.0
## - x5    1 5.4201e+09 1.7521e+11 2298.4
## - x3    1 1.1280e+10 1.8107e+11 2301.9
## - x1    1 2.5189e+10 1.9498e+11 2309.9
##
## Step:  AIC=2296.68
## y ~ x1 + x3 + x5
##
##        Df  Sum of Sq        RSS    AIC
## - x5    1 2.8323e+09 1.7533e+11 2296.4
## <none>               1.7249e+11 2296.7
## - x3    1 2.3323e+10 1.9582e+11 2308.4
## - x1    1 4.9408e+10 2.2190e+11 2321.9
##
## Step:  AIC=2296.44
## y ~ x1 + x3
##
##        Df  Sum of Sq        RSS    AIC
## <none>               1.7533e+11 2296.4
## - x1    1 4.6810e+10 2.2214e+11 2320.0
## - x3    1 5.1372e+10 2.2670e+11 2322.2
```

```
AIC_loader2$coefficients
```

```
##   (Intercept)            x1            x3
## -2.024345e+05  8.932537e+00  7.685012e+03
```

```
AIC_loader3 = AICfunction(data_demo, 3)
```

```
## Start:  AIC=3201.9
## y ~ x1 + x2 + x3 + x4 + x5
##
##        Df  Sum of Sq        RSS    AIC
## - x4    1 1.2838e+05 1.9771e+11 3199.9
## - x5    1 2.8587e+06 1.9771e+11 3199.9
## - x2    1 9.2706e+08 1.9864e+11 3200.6
## <none>               1.9771e+11 3201.9
## - x3    1 6.2352e+09 2.0395e+11 3204.6
## - x1    1 1.2133e+10 2.0984e+11 3209.0
##
## Step:  AIC=3199.9
## y ~ x1 + x2 + x3 + x5
##
##        Df  Sum of Sq        RSS    AIC
```

```
## - x5    1 3.4997e+06 1.9771e+11 3197.9
## - x2    1 1.0447e+09 1.9876e+11 3198.7
## <none>               1.9771e+11 3199.9
## - x3    1 9.1501e+09 2.0686e+11 3204.8
## - x1    1 2.0546e+10 2.1826e+11 3212.9
##
## Step:  AIC=3197.9
## y ~ x1 + x2 + x3
##
##        Df  Sum of Sq        RSS    AIC
## - x2    1 1.1429e+09 1.9886e+11 3196.8
## <none>               1.9771e+11 3197.9
## - x3    1 1.0401e+10 2.0811e+11 3203.7
## - x1    1 2.3238e+10 2.2095e+11 3212.8
##
## Step:  AIC=3196.78
## y ~ x1 + x3
##
##        Df  Sum of Sq        RSS    AIC
## <none>               1.9886e+11 3196.8
## - x3    1 1.5230e+10 2.1409e+11 3206.0
## - x1    1 2.2194e+10 2.2105e+11 3210.9
```

```
AIC_loader3$coefficients
```

```
##  (Intercept)           x1           x3
## -64787.42054      3.91437   2206.90440
```

```
AIC_loader4 = AICfunction(data_demo, 4)
```

```
## Start:  AIC=1746.42
## y ~ x1 + x2 + x3 + x4 + x5
##
##        Df  Sum of Sq        RSS    AIC
## - x1    1 4.5169e+08 4.6703e+11 1744.5
## - x3    1 3.6311e+09 4.7020e+11 1745.0
## - x4    1 7.6263e+09 4.7420e+11 1745.7
## <none>               4.6657e+11 1746.4
## - x2    1 2.2403e+10 4.8898e+11 1748.0
## - x5    1 3.5089e+10 5.0166e+11 1750.0
##
## Step:  AIC=1744.49
## y ~ x2 + x3 + x4 + x5
##
##        Df  Sum of Sq        RSS    AIC
## - x3    1 9.8347e+09 4.7686e+11 1744.1
## <none>               4.6703e+11 1744.5
## - x2    1 2.3284e+10 4.9031e+11 1746.2
## - x4    1 2.6717e+10 4.9374e+11 1746.8
## - x5    1 6.0540e+10 5.2757e+11 1751.9
##
## Step:  AIC=1744.09
## y ~ x2 + x4 + x5
##
##        Df  Sum of Sq        RSS    AIC
```

```
## <none>               4.7686e+11 1744.1
## - x2    1 2.3032e+10 4.9989e+11 1745.7
## - x4    1 2.8423e+10 5.0528e+11 1746.5
## - x5    1 5.3254e+10 5.3011e+11 1750.2
```

```
AIC_loader4$coefficients
```

```
## (Intercept)          x2          x4          x5
##  581617.594   -9087.879    3847.990   -7060.223
```

```
BIC_loader1 = BICfunction(data_demo, 1)
```

```
## Start:  AIC=2277.22
## y ~ x1 + x2 + x3 + x4 + x5
##
##        Df  Sum of Sq        RSS    AIC
## - x2    1 1.6475e+09 3.1598e+11 2273.1
## - x1    1 7.9952e+09 3.2233e+11 2275.2
## - x5    1 8.4169e+09 3.2275e+11 2275.3
## - x4    1 8.7510e+09 3.2309e+11 2275.4
## <none>               3.1433e+11 2277.2
## - x3    1 7.6877e+10 3.9121e+11 2295.1
##
## Step:  AIC=2273.13
## y ~ x1 + x3 + x4 + x5
##
##        Df  Sum of Sq        RSS    AIC
## - x1    1 7.8106e+09 3.2379e+11 2271.0
## - x5    1 8.3993e+09 3.2438e+11 2271.2
## - x4    1 1.0755e+10 3.2674e+11 2271.9
## <none>               3.1598e+11 2273.1
## - x3    1 7.6837e+10 3.9282e+11 2290.9
##
## Step:  AIC=2271.01
## y ~ x3 + x4 + x5
##
##        Df  Sum of Sq        RSS    AIC
## <none>               3.2379e+11 2271.0
## - x5    1 1.8073e+10 3.4187e+11 2272.0
## - x4    1 6.1535e+10 3.8533e+11 2284.3
## - x3    1 6.9756e+10 3.9355e+11 2286.5
```

```
BIC_loader1$coefficients
```

```
## (Intercept)          x3          x4          x5
##  147968.881   11134.647    5475.472   -4078.540
```

```
BIC_loader2 = BICfunction(data_demo, 2)
```

```
## Start:  AIC=2314.89
## y ~ x1 + x2 + x3 + x4 + x5
##
##        Df  Sum of Sq        RSS    AIC
## - x2    1 2.7364e+08 1.6979e+11 2310.4
## - x4    1 1.4424e+09 1.7095e+11 2311.1
## - x5    1 5.3930e+09 1.7491e+11 2313.6
## <none>               1.6951e+11 2314.9
```

```
## - x3    1 1.1392e+10 1.8090e+11 2317.2
## - x1    1 2.3976e+10 1.9349e+11 2324.5
##
## Step:  AIC=2310.38
## y ~ x1 + x3 + x4 + x5
##
##        Df  Sum of Sq        RSS    AIC
## - x4    1 2.7082e+09 1.7249e+11 2307.4
## - x5    1 5.4201e+09 1.7521e+11 2309.1
## <none>               1.6979e+11 2310.4
## - x3    1 1.1280e+10 1.8107e+11 2312.7
## - x1    1 2.5189e+10 1.9498e+11 2320.6
##
## Step:  AIC=2307.41
## y ~ x1 + x3 + x5
##
##        Df  Sum of Sq        RSS    AIC
## - x5    1 2.8323e+09 1.7533e+11 2304.5
## <none>               1.7249e+11 2307.4
## - x3    1 2.3323e+10 1.9582e+11 2316.4
## - x1    1 4.9408e+10 2.2190e+11 2329.9
##
## Step:  AIC=2304.49
## y ~ x1 + x3
##
##        Df  Sum of Sq        RSS    AIC
## <none>               1.7533e+11 2304.5
## - x1    1 4.6810e+10 2.2214e+11 2325.4
## - x3    1 5.1372e+10 2.2670e+11 2327.6
```

```
BIC_loader2$coefficients
```

```
##   (Intercept)            x1            x3
## -2.024345e+05  8.932537e+00  7.685012e+03
```

```
BIC_loader3 = BICfunction(data_demo, 3)
```

```
## Start:  AIC=3220.04
## y ~ x1 + x2 + x3 + x4 + x5
##
##        Df  Sum of Sq        RSS    AIC
## - x4    1 1.2838e+05 1.9771e+11 3215.0
## - x5    1 2.8587e+06 1.9771e+11 3215.0
## - x2    1 9.2706e+08 1.9864e+11 3215.7
## - x3    1 6.2352e+09 2.0395e+11 3219.7
## <none>               1.9771e+11 3220.0
## - x1    1 1.2133e+10 2.0984e+11 3224.1
##
## Step:  AIC=3215.02
## y ~ x1 + x2 + x3 + x5
##
##        Df  Sum of Sq        RSS    AIC
## - x5    1 3.4997e+06 1.9771e+11 3210.0
## - x2    1 1.0447e+09 1.9876e+11 3210.8
## <none>               1.9771e+11 3215.0
```

```
## - x3    1 9.1501e+09 2.0686e+11 3216.9
## - x1    1 2.0546e+10 2.1826e+11 3225.0
##
## Step:  AIC=3210
## y ~ x1 + x2 + x3
##
##        Df  Sum of Sq        RSS    AIC
## - x2    1 1.1429e+09 1.9886e+11 3205.9
## <none>               1.9771e+11 3210.0
## - x3    1 1.0401e+10 2.0811e+11 3212.8
## - x1    1 2.3238e+10 2.2095e+11 3221.9
##
## Step:  AIC=3205.85
## y ~ x1 + x3
##
##        Df  Sum of Sq        RSS    AIC
## <none>               1.9886e+11 3205.9
## - x3    1 1.5230e+10 2.1409e+11 3212.0
## - x1    1 2.2194e+10 2.2105e+11 3216.9
```

```
BIC_loader3$coefficients
```

```
##  (Intercept)           x1           x3
## -64787.42054      3.91437   2206.90440
```

```
BIC_loader4 = BICfunction(data_demo, 4)
```

```
## Start:  AIC=1760.48
## y ~ x1 + x2 + x3 + x4 + x5
##
##        Df  Sum of Sq        RSS    AIC
## - x1    1 4.5169e+08 4.6703e+11 1756.2
## - x3    1 3.6311e+09 4.7020e+11 1756.7
## - x4    1 7.6263e+09 4.7420e+11 1757.4
## - x2    1 2.2403e+10 4.8898e+11 1759.8
## <none>               4.6657e+11 1760.5
## - x5    1 3.5089e+10 5.0166e+11 1761.7
##
## Step:  AIC=1756.21
## y ~ x2 + x3 + x4 + x5
##
##        Df  Sum of Sq        RSS    AIC
## - x3    1 9.8347e+09 4.7686e+11 1753.5
## - x2    1 2.3284e+10 4.9031e+11 1755.6
## - x4    1 2.6717e+10 4.9374e+11 1756.2
## <none>               4.6703e+11 1756.2
## - x5    1 6.0540e+10 5.2757e+11 1761.2
##
## Step:  AIC=1753.47
## y ~ x2 + x4 + x5
##
##        Df  Sum of Sq        RSS    AIC
## - x2    1 2.3032e+10 4.9989e+11 1752.8
## <none>               4.7686e+11 1753.5
## - x4    1 2.8423e+10 5.0528e+11 1753.6
```

```
## - x5    1 5.3254e+10 5.3011e+11 1757.3
##
## Step:  AIC=1752.76
## y ~ x4 + x5
##
##        Df  Sum of Sq        RSS    AIC
## <none>               4.9989e+11 1752.8
## - x4    1 2.9733e+10 5.2963e+11 1752.9
## - x5    1 3.0760e+10 5.3065e+11 1753.0
```

```
BIC_loader4$coefficients
```

```
## (Intercept)          x4          x5
##  247217.155    3934.480   -3686.151
```

Question 3:

Estimate a 90% confidence interval for parameters Bj, j=1,. . . ,p

90% Confidence Intervals for each region listed below:

```
confint(AIC_loader1, level = 0.9)
```

```
##                        5 %         95 %
## (Intercept) -2.250082e+05 2.739312e+05
## x1          -2.425978e-01 7.494175e+00
## x3           8.445211e+03 1.715264e+04
## x4           3.059355e+02 6.432111e+03
## x5          -6.082087e+03 8.646288e+01
```

x1 -2.425978e-01 7.494175e+00 x3 8.445211e+03 1.715264e+04 x4 3.059355e+02 6.432111e+03 x5 -6.082087e+03 8.646288e+01

```
confint(AIC_loader2, level = 0.9)
```

```
##                        5 %          95 %
## (Intercept) -2.654457e+05 -139423.35914
## x1           6.132835e+00      11.73224
## x3           5.385754e+03    9984.27016
```

x1 6.132835e+00 11.73224 x3 5.385754e+03 9984.27016

```
confint(AIC_loader3, level = 0.9)
```

```
##                        5 %          95 %
## (Intercept) -1.008800e+05 -28694.890532
## x1           2.325627e+00      5.503114
## x3           1.125616e+03   3288.192401
```

x1 2.325627e+00 5.503114 x3 1.125616e+03 3288.192401

```
confint(AIC_loader4, level = 0.9)
```

```
##                       5 %       95 %
## (Intercept) 232751.4801 930483.708
## x2          -17151.0469  -1024.712
## x4             774.6599   6921.321
## x5          -11179.7778  -2940.668
```

x2 -17151.0469 -1024.712 x4 774.6599 6921.321 x5 -11179.7778 -2940.668

```
confint(BIC_loader1, level = 0.9)
```

```
##                     5 %       95 %
## (Intercept) -65355.845 361293.608
## x3            7131.412  15137.883
## x4            3379.490   7571.454
## x5           -6959.379  -1197.701
```

x3 7131.412 15137.883 x4 3379.490 7571.454 x5 -6959.379 -1197.701

```
confint(BIC_loader2, level = 0.9)
```

```
##                         5 %          95 %
## (Intercept) -2.654457e+05 -139423.35914
## x1           6.132835e+00      11.73224
## x3           5.385754e+03    9984.27016
```

x1 6.132835e+00 11.73224 x3 5.385754e+03 9984.27016

```
confint(BIC_loader3, level = 0.9)
```

```
##                         5 %          95 %
## (Intercept) -1.008800e+05 -28694.890532
## x1           2.325627e+00      5.503114
## x3           1.125616e+03   3288.192401
```

x1 2.325627e+00 5.503114 x3 1.125616e+03 3288.192401

```
confint(BIC_loader4, level = 0.9)
```

```
##                     5 %        95 %
## (Intercept) 60623.396 433810.9147
## x4            810.656   7058.3043
## x5          -6563.551   -808.7508
```

x4 810.656 7058.3043 x5 -6563.551 -808.7508

Using alpha = 0.01, compute the p-value for the two alterntives in the formula:

The P-Value computations and comparisons are listed below:

```
summary(AIC_loader1)
```

```
##
## Call:
## lm(formula = y ~ x1 + x3 + x4 + x5, data = loader)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -125137  -16763   -1610   13404  481036
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  24461.535 150233.059   0.163   0.8710
## x1               3.626      2.330   1.556   0.1228
## x3           12798.924   2621.848   4.882 4.09e-06 ***
## x4            3369.023   1844.621   1.826   0.0708 .
## x5           -2997.812   1857.380  -1.614   0.1097
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 56780 on 98 degrees of freedom
## Multiple R-squared:  0.332,  Adjusted R-squared:  0.3047
## F-statistic: 12.18 on 4 and 98 DF,  p-value: 4.481e-08
```

x1 P-Value: 0.1228 > 0.01 x3 P-Value: 4.09e-0 < 0.01 x4 P-Value: 0.0708 > 0.01 x5 P-Value: 0.1097 > 0.01

From this test, We can conclude that we can reject H0 for x1, x4, and x5, but fail to reject H0 for x3

```
summary(AIC_loader2)
```

```
##
## Call:
## lm(formula = y ~ x1 + x3, data = loader)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -53569 -14118  -3370   6912 359972
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.024e+05  3.797e+04  -5.331 5.62e-07 ***
## x1           8.933e+00  1.687e+00   5.295 6.59e-07 ***
## x3           7.685e+03  1.386e+03   5.547 2.19e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 40860 on 105 degrees of freedom
## Multiple R-squared:  0.2692, Adjusted R-squared:  0.2553
## F-statistic: 19.34 on 2 and 105 DF,  p-value: 7.076e-08
```

x1 P-Value: 6.59e-07 < 0.01 x3 P-Value: 2.19e-07 < 0.01

From this test we can conclude that we fail to reject H0 for both x1 and x3

```
summary(AIC_loader3)
```

```
##
## Call:
## lm(formula = y ~ x1 + x3, data = loader)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -61130 -15702  -7705   2106 214330
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.479e+04  2.181e+04  -2.971 0.003461 **
## x1           3.914e+00  9.599e-01   4.078 7.37e-05 ***
## x3           2.207e+03  6.533e+02   3.378 0.000931 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36530 on 149 degrees of freedom
## Multiple R-squared:  0.1081, Adjusted R-squared:  0.09612
## F-statistic: 9.029 on 2 and 149 DF,  p-value: 0.0001989
```

x1 P-Value: 7.37e-05 < 0.01 x3 P-Value: 0.000931 < 0.01

From this test, we conclude that we fail to reject H0 for either x1 or x3

```
summary(AIC_loader4)
```

```
##
## Call:
## lm(formula = y ~ x2 + x4 + x5, data = loader)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -92749 -32731 -10488   9860 588427
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   581618     209404   2.777  0.00696 **
## x2             -9088       4840  -1.878  0.06441 .
## x4              3848       1845   2.086  0.04048 *
## x5             -7060       2473  -2.855  0.00560 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 80820 on 73 degrees of freedom
## Multiple R-squared:  0.1088, Adjusted R-squared:  0.07217
## F-statistic:  2.97 on 3 and 73 DF,  p-value: 0.03733
```

x2 P-Value: 0.06441 > 0.01 x4 P-Value: 0.04048 > 0.01 x5 P-Value: 0.00560 < 0.01

From this test we can conclude that we reject H0 for x2 and x4, but fail to reject H0 for x5

```
summary(BIC_loader1)
```

```
##
## Call:
## lm(formula = y ~ x3 + x4 + x5, data = loader)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -114724  -17914   -1981   11517  484782
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   147969     128479   1.152   0.2522
## x3             11135       2411   4.618 1.17e-05 ***
## x4              5476       1262   4.338 3.47e-05 ***
## x5             -4078       1735  -2.351   0.0207 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 57190 on 99 degrees of freedom
## Multiple R-squared:  0.3155, Adjusted R-squared:  0.2947
## F-statistic: 15.21 on 3 and 99 DF,  p-value: 3.258e-08
```

x3 P-Value: 1.17e-05 < 0.01 x4 P-Value: 3.47e-05 < 0.01 x5 P-Value: 0.0207 > 0.01

From this test we conclude that we reject H0 for x3 and x4, but fail to reject H0 for x5

```
summary(BIC_loader2)
```

```
##
## Call:
## lm(formula = y ~ x1 + x3, data = loader)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -53569 -14118  -3370   6912 359972
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.024e+05  3.797e+04  -5.331 5.62e-07 ***
## x1           8.933e+00  1.687e+00   5.295 6.59e-07 ***
## x3           7.685e+03  1.386e+03   5.547 2.19e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 40860 on 105 degrees of freedom
## Multiple R-squared:  0.2692, Adjusted R-squared:  0.2553
## F-statistic: 19.34 on 2 and 105 DF,  p-value: 7.076e-08
```

x1 P-Value: 6.59e-07 $<$ 0.01 x3 P-Value: 2.19e-07 $<$ 0.01

From this test we conclude that we fail to reject H0 for either parameter

```
summary(BIC_loader3)
```

```
##
## Call:
## lm(formula = y ~ x1 + x3, data = loader)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -61130 -15702  -7705   2106 214330
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.479e+04  2.181e+04  -2.971 0.003461 **
## x1           3.914e+00  9.599e-01   4.078 7.37e-05 ***
## x3           2.207e+03  6.533e+02   3.378 0.000931 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36530 on 149 degrees of freedom
## Multiple R-squared:  0.1081, Adjusted R-squared:  0.09612
## F-statistic: 9.029 on 2 and 149 DF,  p-value: 0.0001989
```

x1 P-Value: 7.37e-05 $<$ 0.01 x3 P-Value: 0.000931 $<$ 0.01

From this test we conclude that we fail to reject H0 for either parameter

```
summary(BIC_loader4)
```

```
##
## Call:
## lm(formula = y ~ x4 + x5, data = loader)
```

```
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -81236 -28955 -14033   3792 612011
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   247217     112021   2.207   0.0304 *
## x4              3934       1875   2.098   0.0393 *
## x5             -3686       1727  -2.134   0.0362 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 82190 on 74 degrees of freedom
## Multiple R-squared:  0.06575,    Adjusted R-squared:  0.0405
## F-statistic: 2.604 on 2 and 74 DF,  p-value: 0.08076
```

x4 P-Value: 0.0393 > 0.01 x5 P-Value: 0.0362 > 0.01

From this test we conclude that we reject H0 for both x4 and x4

Test whether or not B1 = B2 = ... = Bp-1 = 0 with a = 0.05. State the decision rule and conclusion. Are these measures similar for the four regions?

H0: B1 = B2 = ... = Bp-1 = 0 Ha: B1 != B2 != ... = Bp-1 != 0

We fail to reject H0 with the majority of the tests, with the exceptions of x1, x4, x5 of Region 1, and x2 of Region 4. For these parameters, we fail to reject H0 and conclude that these parameters are not the same and do not equal 0.

```
summary(AIC_loader1)
```

```
##
## Call:
## lm(formula = y ~ x1 + x3 + x4 + x5, data = loader)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -125137  -16763   -1610   13404  481036
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  24461.535 150233.059   0.163   0.8710
## x1               3.626      2.330   1.556   0.1228
## x3           12798.924   2621.848   4.882 4.09e-06 ***
## x4            3369.023   1844.621   1.826   0.0708 .
## x5           -2997.812   1857.380  -1.614   0.1097
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 56780 on 98 degrees of freedom
## Multiple R-squared:  0.332,  Adjusted R-squared:  0.3047
## F-statistic: 12.18 on 4 and 98 DF,  p-value: 4.481e-08
```

```
summary(AIC_loader2)
```

```
##
```

```
## Call:
## lm(formula = y ~ x1 + x3, data = loader)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -53569 -14118  -3370   6912 359972
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.024e+05  3.797e+04  -5.331 5.62e-07 ***
## x1           8.933e+00  1.687e+00   5.295 6.59e-07 ***
## x3           7.685e+03  1.386e+03   5.547 2.19e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 40860 on 105 degrees of freedom
## Multiple R-squared:  0.2692, Adjusted R-squared:  0.2553
## F-statistic: 19.34 on 2 and 105 DF,  p-value: 7.076e-08
```

```
summary(AIC_loader3)
```

```
##
## Call:
## lm(formula = y ~ x1 + x3, data = loader)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -61130 -15702  -7705   2106 214330
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.479e+04  2.181e+04  -2.971 0.003461 **
## x1           3.914e+00  9.599e-01   4.078 7.37e-05 ***
## x3           2.207e+03  6.533e+02   3.378 0.000931 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36530 on 149 degrees of freedom
## Multiple R-squared:  0.1081, Adjusted R-squared:  0.09612
## F-statistic: 9.029 on 2 and 149 DF,  p-value: 0.0001989
```

```
summary(AIC_loader4)
```

```
##
## Call:
## lm(formula = y ~ x2 + x4 + x5, data = loader)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -92749 -32731 -10488   9860 588427
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   581618     209404   2.777  0.00696 **
## x2             -9088       4840  -1.878  0.06441 .
```

```
## x4              3848       1845   2.086  0.04048 *
## x5             -7060       2473  -2.855  0.00560 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 80820 on 73 degrees of freedom
## Multiple R-squared:  0.1088, Adjusted R-squared:  0.07217
## F-statistic:  2.97 on 3 and 73 DF,  p-value: 0.03733
```

```
summary(BIC_loader1)
```

```
##
## Call:
## lm(formula = y ~ x3 + x4 + x5, data = loader)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -114724  -17914   -1981   11517  484782
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   147969     128479   1.152   0.2522
## x3             11135       2411   4.618 1.17e-05 ***
## x4              5476       1262   4.338 3.47e-05 ***
## x5             -4078       1735  -2.351   0.0207 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 57190 on 99 degrees of freedom
## Multiple R-squared:  0.3155, Adjusted R-squared:  0.2947
## F-statistic: 15.21 on 3 and 99 DF,  p-value: 3.258e-08
```

```
summary(BIC_loader2)
```

```
##
## Call:
## lm(formula = y ~ x1 + x3, data = loader)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -53569 -14118  -3370   6912 359972
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.024e+05  3.797e+04  -5.331 5.62e-07 ***
## x1           8.933e+00  1.687e+00   5.295 6.59e-07 ***
## x3           7.685e+03  1.386e+03   5.547 2.19e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 40860 on 105 degrees of freedom
## Multiple R-squared:  0.2692, Adjusted R-squared:  0.2553
## F-statistic: 19.34 on 2 and 105 DF,  p-value: 7.076e-08
```

```
summary(BIC_loader3)
```

```
##
## Call:
## lm(formula = y ~ x1 + x3, data = loader)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -61130 -15702  -7705   2106 214330
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.479e+04  2.181e+04  -2.971 0.003461 **
## x1           3.914e+00  9.599e-01   4.078 7.37e-05 ***
## x3           2.207e+03  6.533e+02   3.378 0.000931 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36530 on 149 degrees of freedom
## Multiple R-squared:  0.1081, Adjusted R-squared:  0.09612
## F-statistic: 9.029 on 2 and 149 DF,  p-value: 0.0001989
```

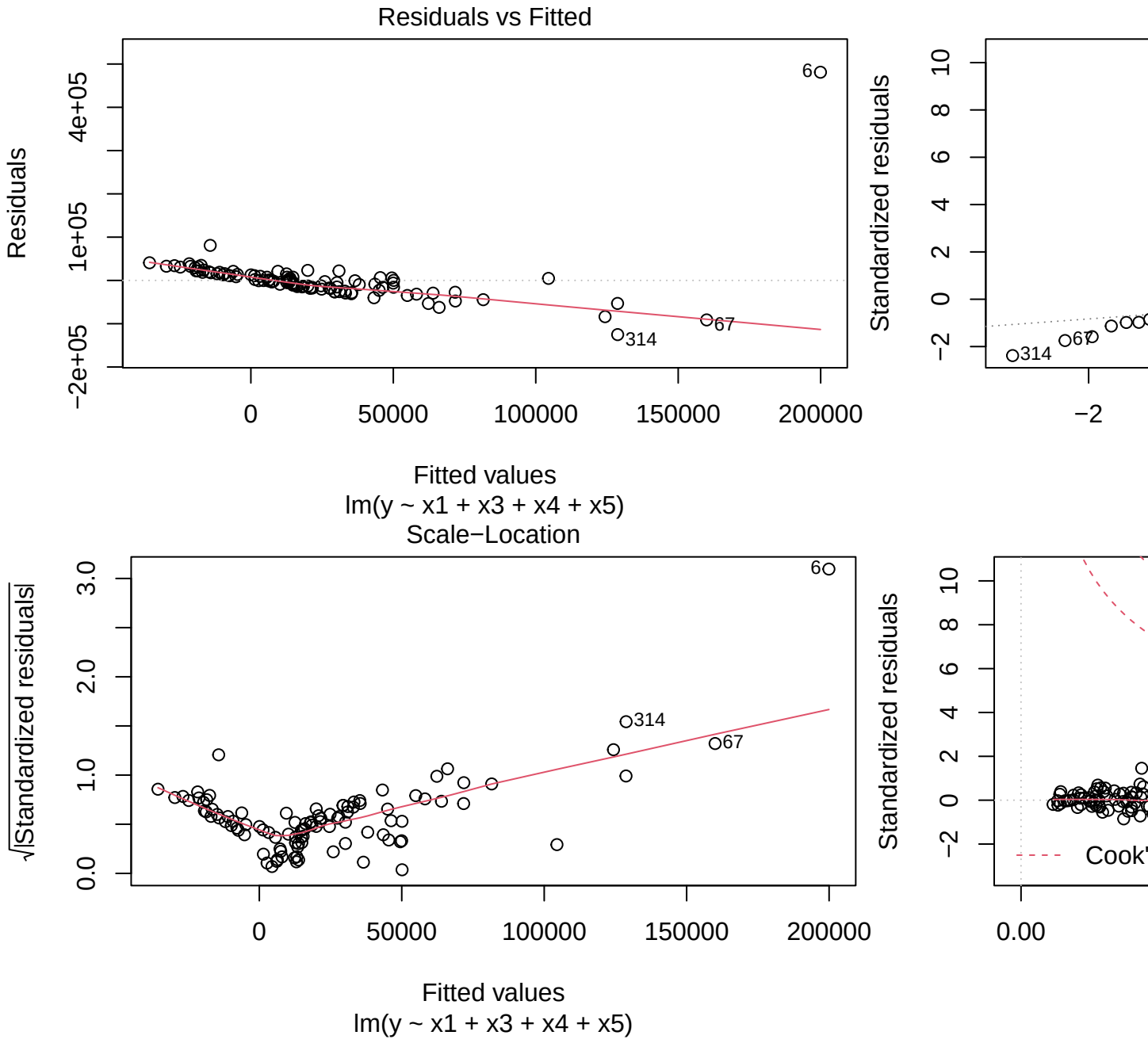
```
summary(BIC_loader4)
```

```
##
## Call:
## lm(formula = y ~ x4 + x5, data = loader)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -81236 -28955 -14033   3792 612011
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   247217     112021   2.207   0.0304 *
## x4              3934       1875   2.098   0.0393 *
## x5             -3686       1727  -2.134   0.0362 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 82190 on 74 degrees of freedom
## Multiple R-squared:  0.06575,    Adjusted R-squared:  0.0405
## F-statistic: 2.604 on 2 and 74 DF,  p-value: 0.08076
```

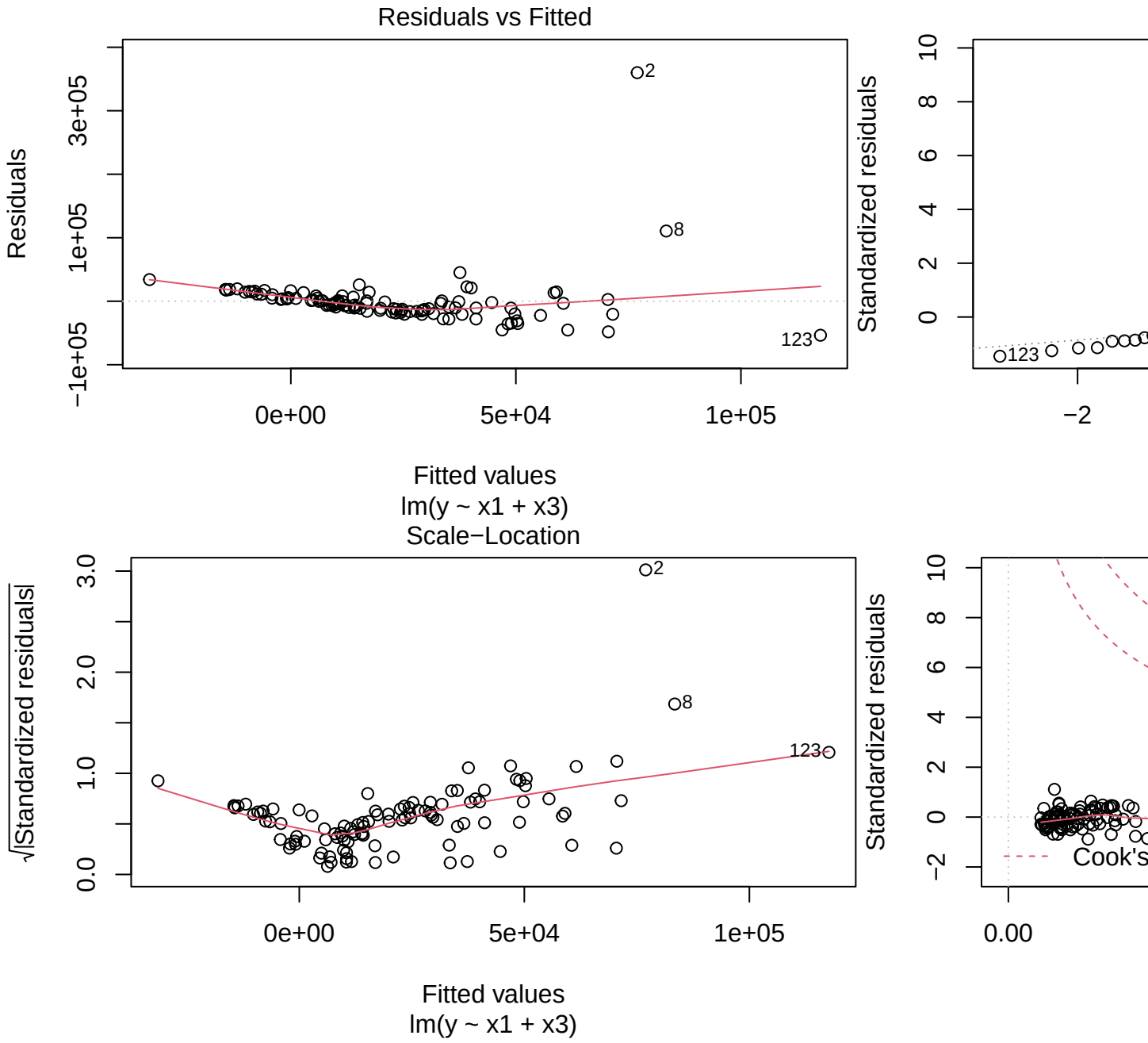
Obtain the residuals for each fitted model and prepare the diagnostic plots for each fitted model. State the conclusions.

Despite being categorized differently because of the AIC vs BIC method, the residual plots for the 4 geographic regions remain the same between AIC and BIC models. In terms of the plots compared between regions, Region 1 and Region 2 have very similar residual plots, with the only noticeable difference being that Region 2 has less extreme outliers in the Cook's Distance plot when compared to Region 1's plot. Region 4 also closely follows the trends of Region's 1 and 2. The most deviation comes from region 3, who's residuals deviate from the trends in a much more extreme fashion.

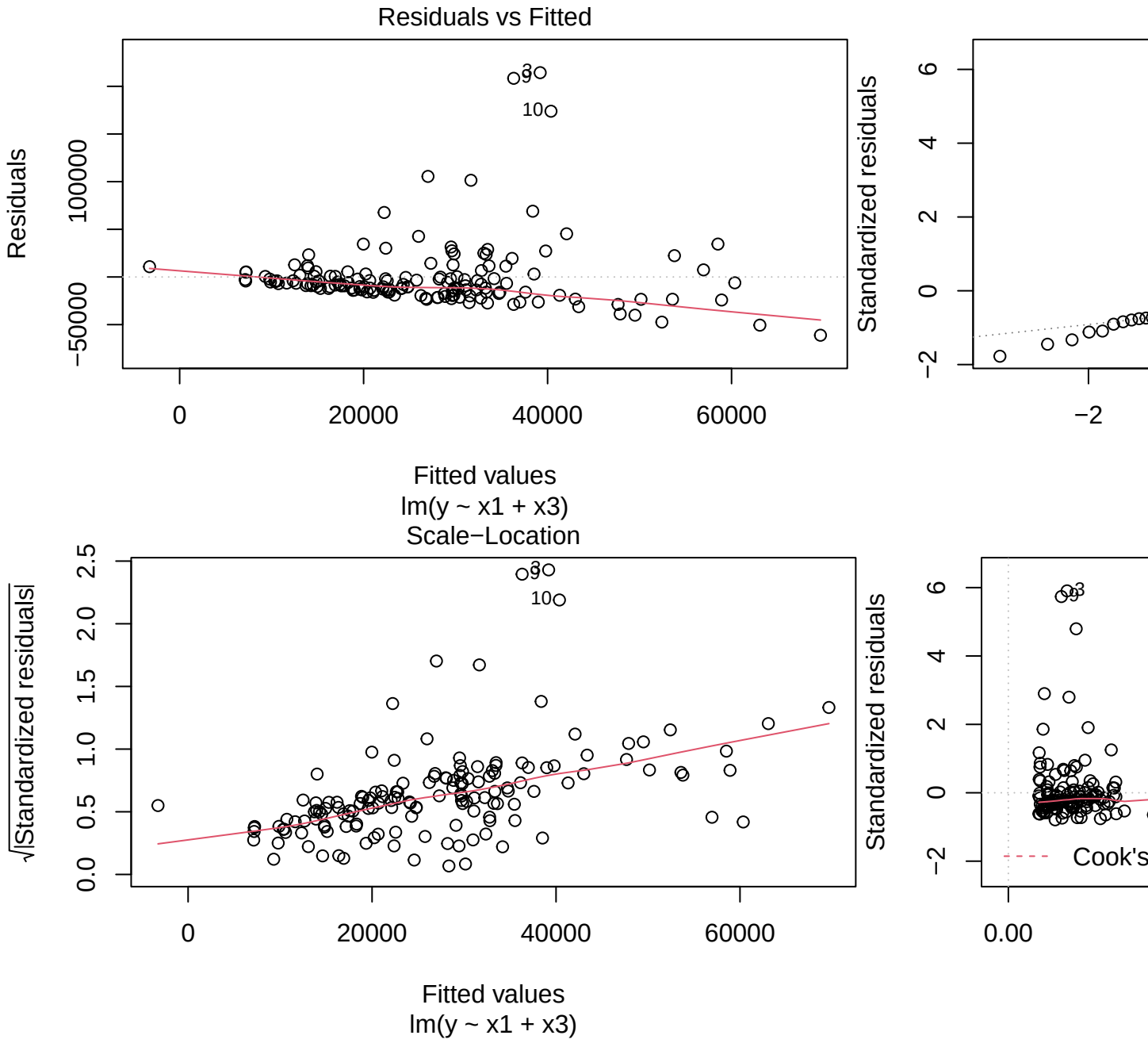
```
plot(AIC_loader1)
```



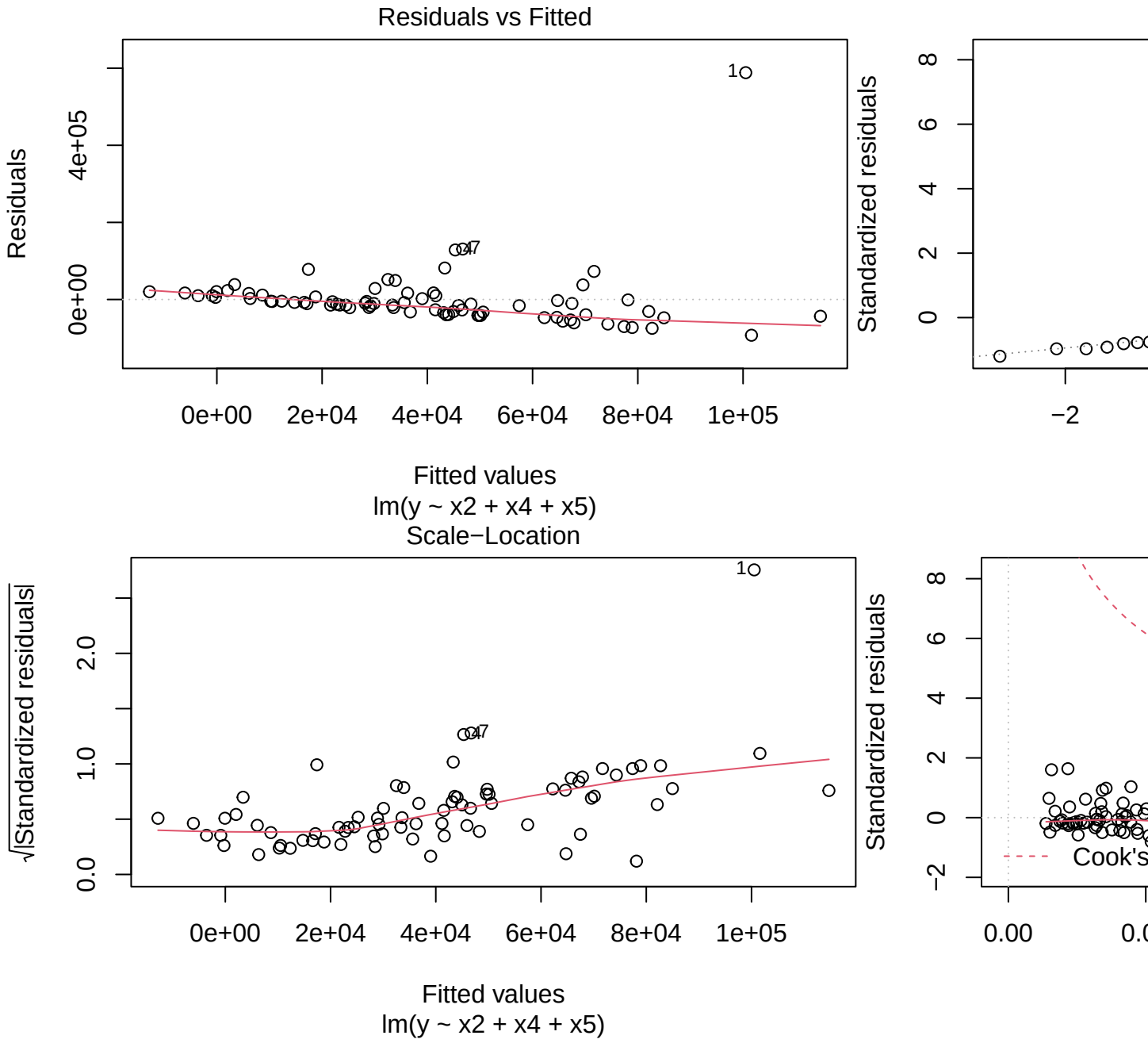

```
plot(AIC_loader2)
```



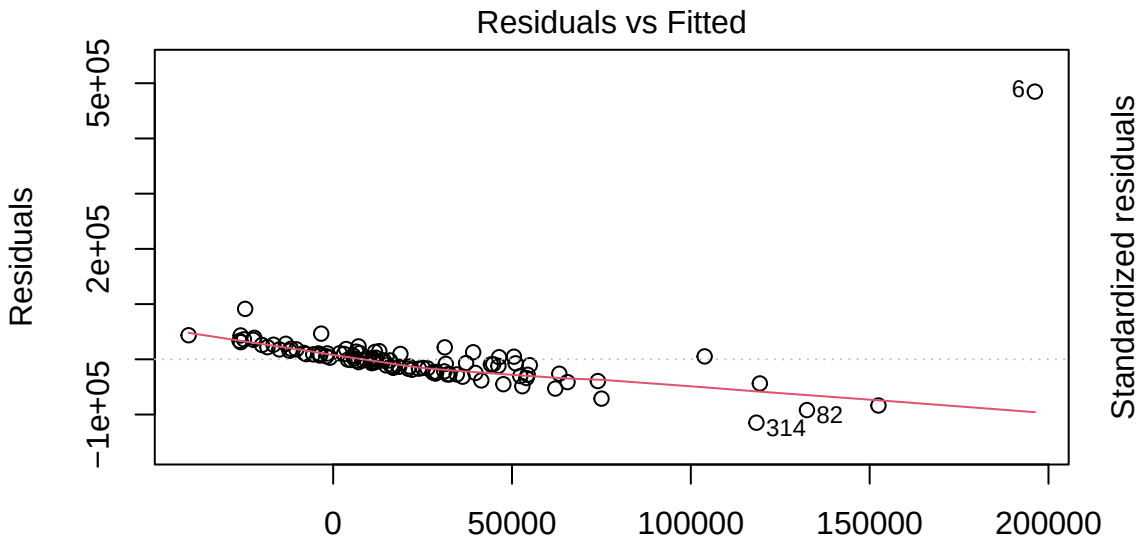

```
plot(AIC_loader3)
```

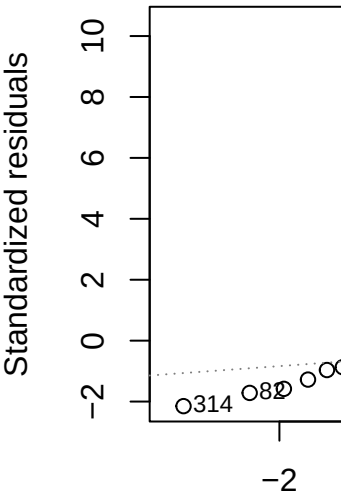


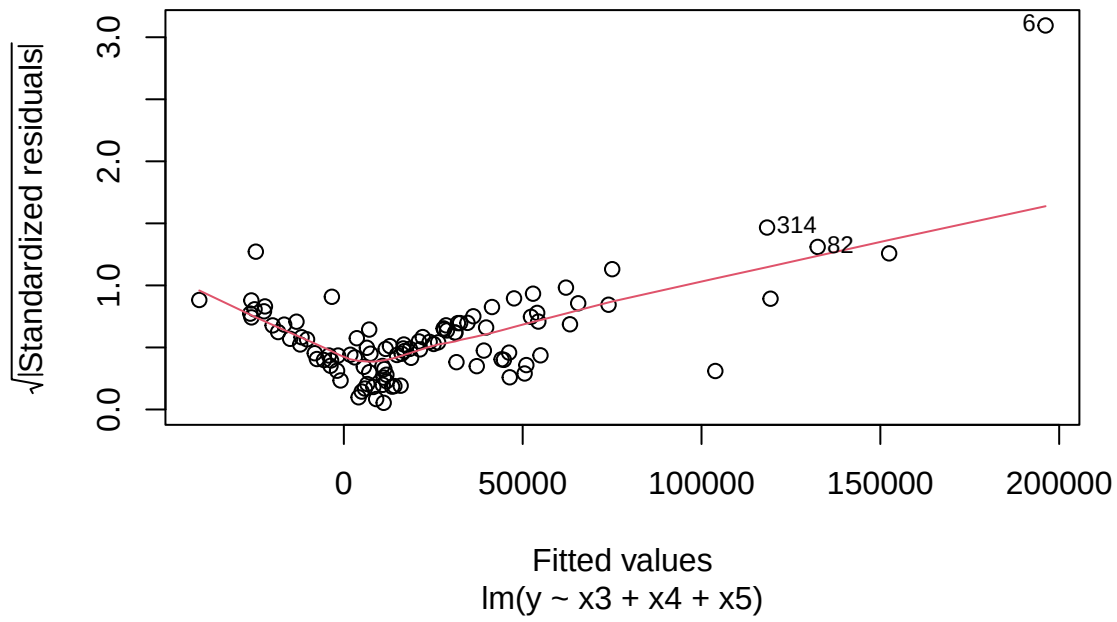

```
plot(AIC_loader4)
```

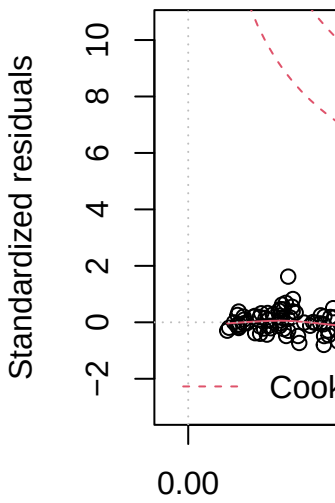



```
plot(BIC_loader1)
```

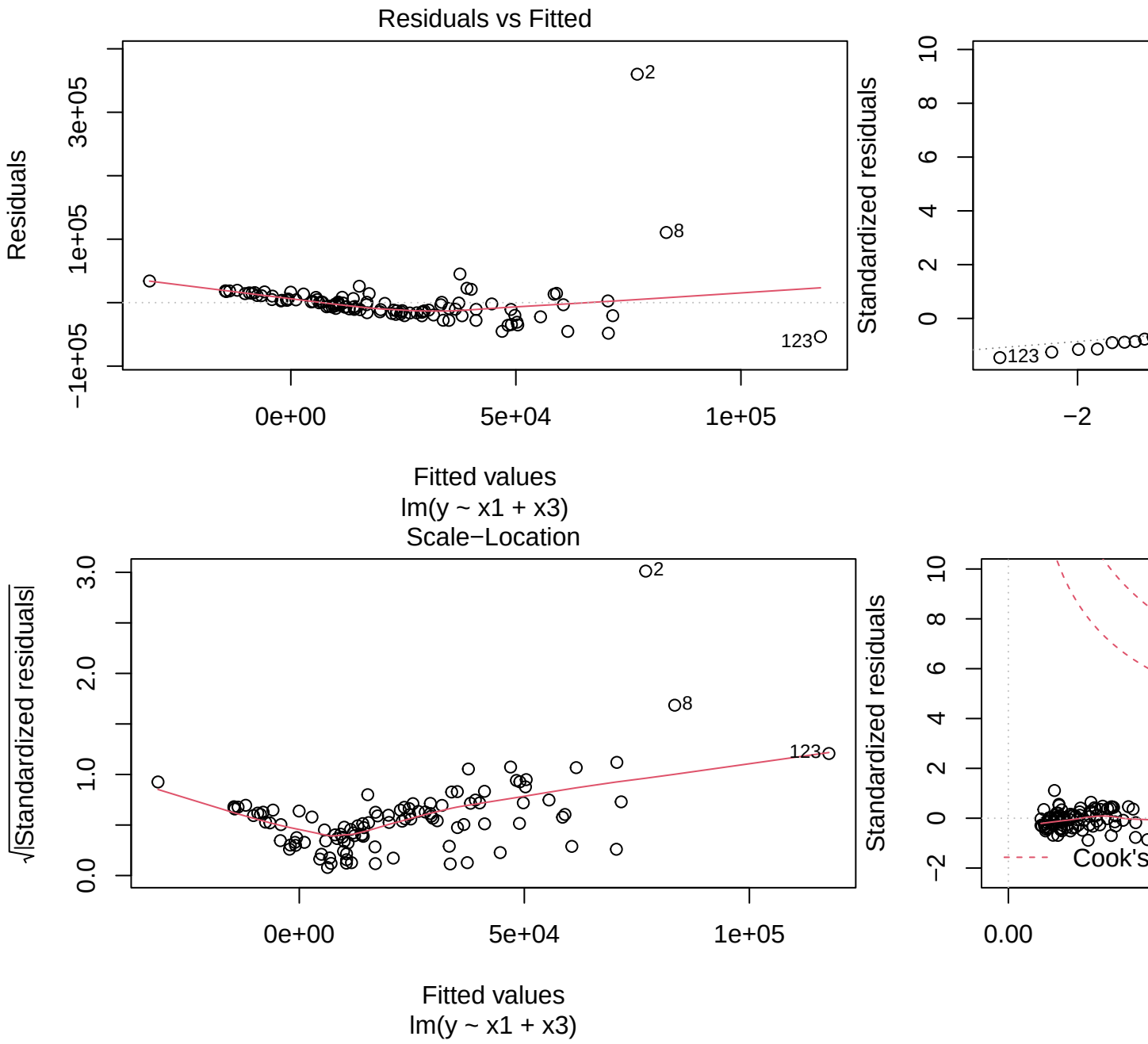




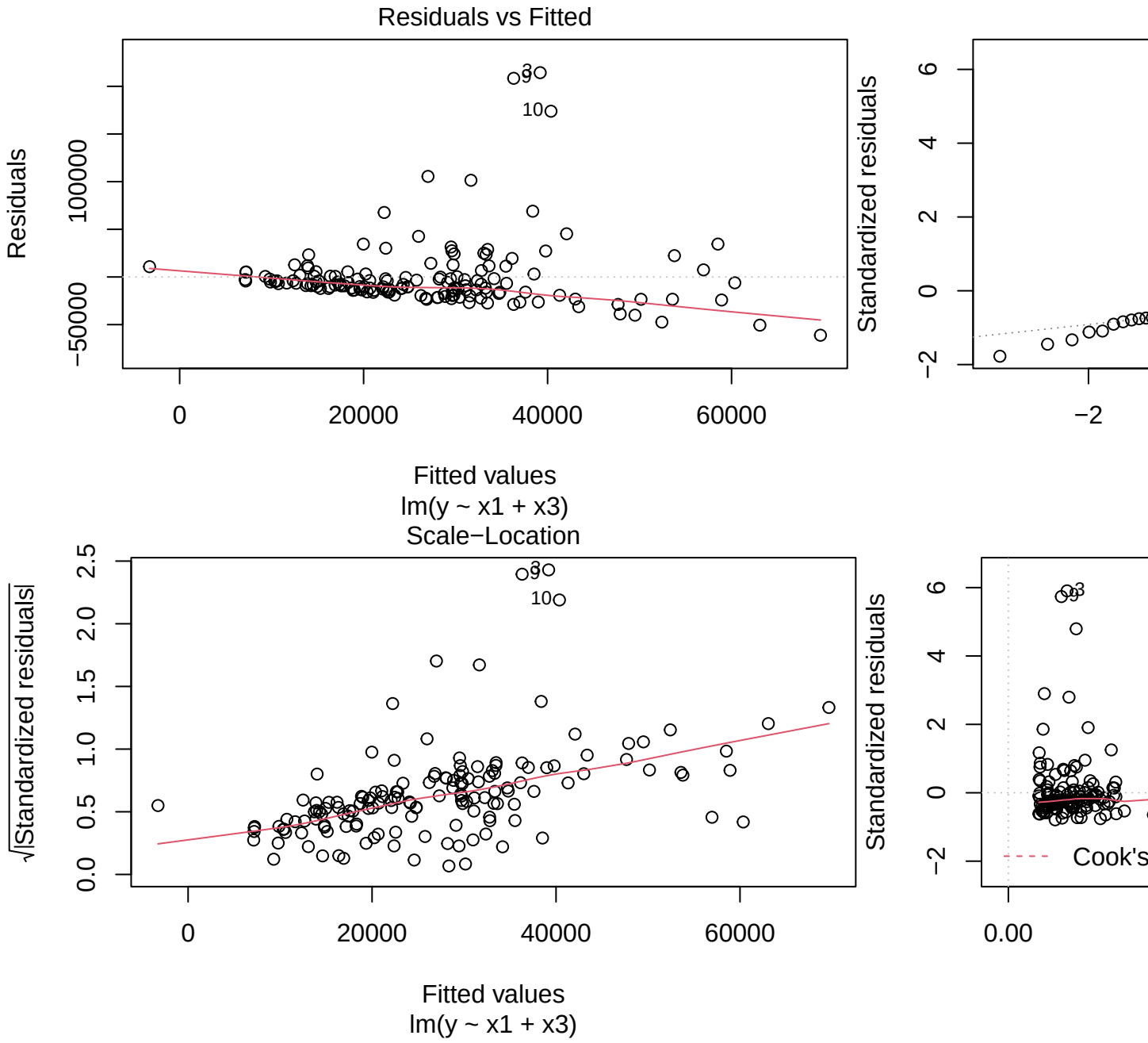




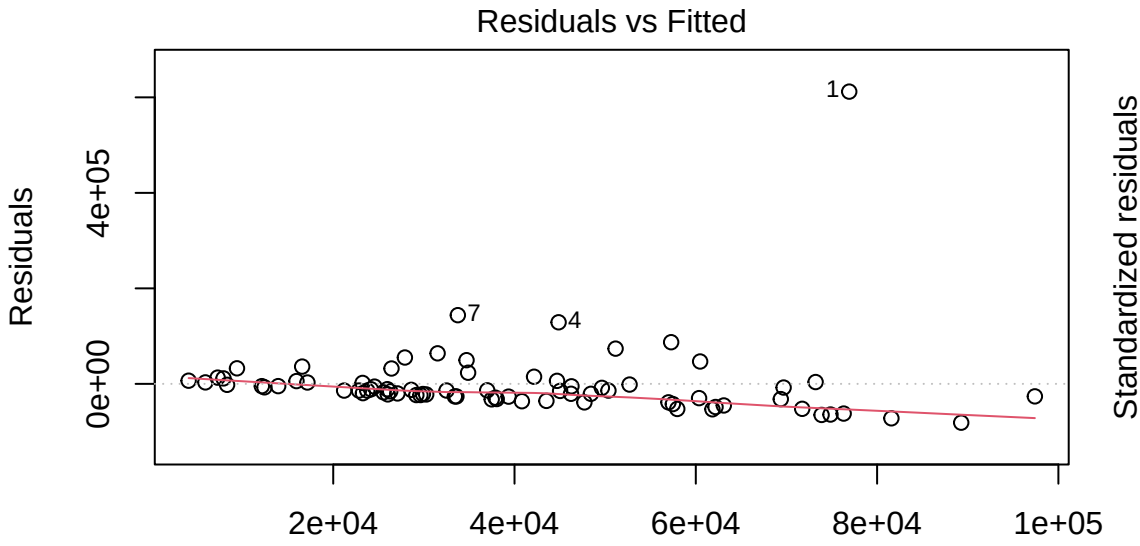

```
plot(BIC_loader2)
```

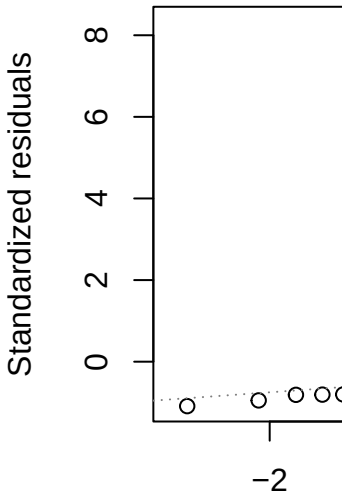


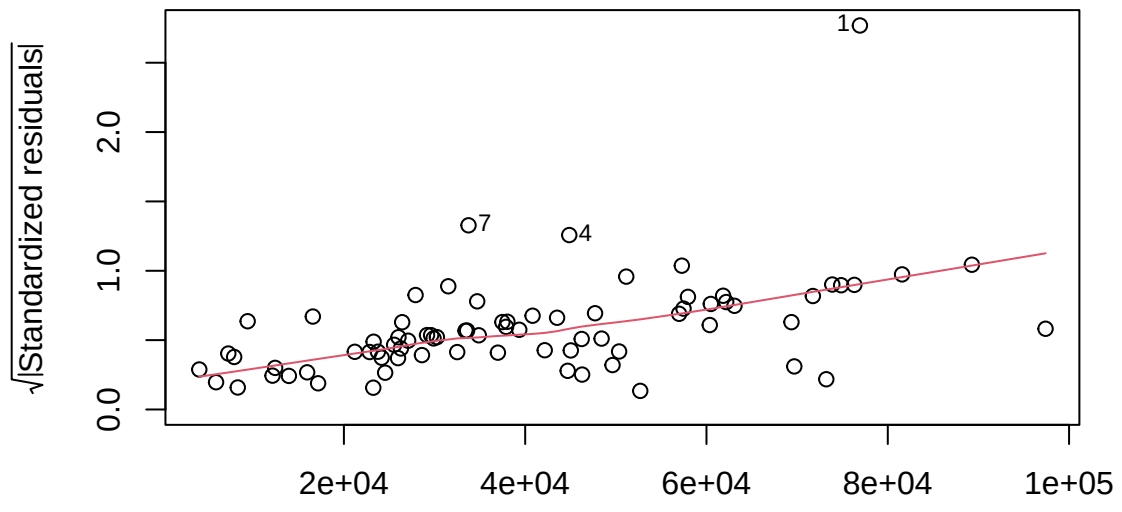

```
plot(BIC_loader3)
```

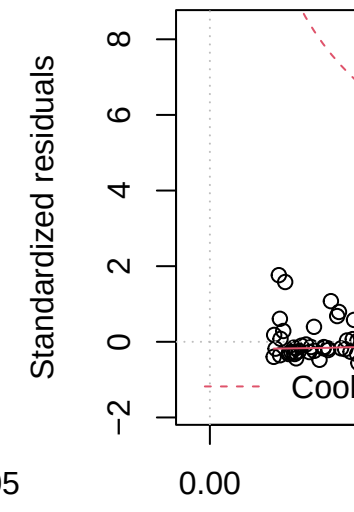



```
plot(BIC_loader4)
```


Residuals vs Fitted
Residuals
0e+00
4e+05
2e+04
4e+04
6e+04
8e+04
1e+05
1
7
4
Fitted values
lm(y ~ x4 + x5)


Standardized residuals
0
2
4
6
8
−2


Scale−Location
√|Standardized residuals|
0.0
1.0
2.0
2e+04
4e+04
6e+04
8e+04
1e+05
1
7
4
Fitted values
lm(y ~ x4 + x5)


Standardized residuals
−2
0
2
4
6
8
0.00
Cook