```
> #1(a)
> #Problem 1 Using ordinary least squares
> data(sat, package="faraway")
> lmod=lm(total~takers+ratio+salary+expend, sat)
> summary(lmod)

Call:
lm(formula = total ~ takers + ratio + salary + expend, data = sat)

Residuals:
    Min      1Q  Median      3Q     Max
-90.531 -20.855  -1.746  15.979  66.571

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1045.9715    52.8698  19.784  < 2e-16 ***
takers        -2.9045     0.2313 -12.559 2.61e-16 ***
ratio         -3.6242     3.2154  -1.127    0.266
salary         1.6379     2.3872   0.686    0.496
expend         4.4626    10.5465   0.423    0.674
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 32.7 on 45 degrees of freedom
Multiple R-squared:  0.8246,    Adjusted R-squared:  0.809
F-statistic: 52.88 on 4 and 45 DF,  p-value: < 2.2e-16
```

```
> #1(b)
> #Using robust regression
> rlmod=rlm(total~takers+ratio+salary+expend, sat)
> summary(rlmod)

Call: rlm(formula = total ~ takers + ratio + salary + expend, data = sat)
Residuals:
    Min      1Q  Median      3Q     Max
-92.510 -17.701  -1.002  15.015  77.058

Coefficients:
             Value     Std. Error t value
(Intercept) 1060.2074   49.8845    21.2533
takers        -2.9778    0.2182   -13.6470
ratio         -5.1254    3.0339    -1.6894
salary         2.0933    2.2525     0.9293
expend         3.9158    9.9510     0.3935

Residual standard error: 25.58 on 45 degrees of freedom
> pt(-13.6470, 45)
[1] 6.645373e-18
> pt(-1.6894, 45)
[1] 0.04903044
> pt(0.9293, 45)
[1] 0.821154
> pt(0.3935,45)
[1] 0.6520957
> #Takers and ratio are significant, salary and expend are insignificant.
```

```
> #2(a)
> data(happy, package="faraway")
> lmod=lm(happy~money+sex+work+love, happy)
> summary(lmod)

Call:
lm(formula = happy ~ money + sex + work + love, data = happy)

Residuals:
    Min      1Q  Median      3Q     Max
-2.7186 -0.5779 -0.1172  0.6340  2.0651

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.072081   0.852543  -0.085   0.9331
money        0.009578   0.005213   1.837   0.0749 .
sex         -0.149008   0.418525  -0.356   0.7240
work         0.476079   0.199389   2.388   0.0227 *
love         1.919279   0.295451   6.496 1.97e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.058 on 34 degrees of freedom
Multiple R-squared:  0.7102,    Adjusted R-squared:  0.6761
F-statistic: 20.83 on 4 and 34 DF,  p-value: 9.364e-09

> #Consider all the polynomial terms and interaction terms
> lmod5=lm(happy~I(money*work)+I(money*sex)+I(money*love)+I(money*work)+I(sex*love
)+I(sex*work)+I(love*work)+money+sex+work+love+I(money^2)+I(money^3)+I(sex^2)+I(se
x^3)+I(work^2)+I(work^3)+I(love^2)+I(love^3), happy)
> summary(lmod5)

Call:
lm(formula = happy ~ I(money * work) + I(money * sex) + I(money *
    love) + I(money * work) + I(sex * love) + I(sex * work) +
```

```
Call:
lm(formula = happy ~ I(money * work) + I(money * sex) + I(money *
    love) + I(money * work) + I(sex * love) + I(sex * work) +
    I(love * work) + money + sex + work + love + I(money^2) +
    I(money^3) + I(sex^2) + I(sex^3) + I(work^2) + I(work^3) +
    I(love^2) + I(love^3), data = happy)

Residuals:
    Min      1Q   Median      3Q     Max
-1.19961 -0.54026 -0.01203  0.37433  1.47274

Coefficients: (3 not defined because of singularities)
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -9.835e+00  5.523e+00  -1.781   0.0882 .
I(money * work)   8.168e-03  8.138e-03   1.004   0.3260
I(money * sex)    2.822e-02  2.638e-02   1.070   0.2959
I(money * love)  -4.591e-02  1.661e-02  -2.764   0.0111 *
I(sex * love)    -2.031e-01  9.652e-01  -0.210   0.8352
I(sex * work)    -4.165e-01  7.834e-01  -0.532   0.6000
I(love * work)   -3.864e-01  2.798e-01  -1.381   0.1806
money             1.508e-01  6.477e-02   2.329   0.0290 *
sex              -3.355e-02  3.163e+00  -0.011   0.9916
work             -3.183e+00  3.049e+00  -1.044   0.3074
love              8.541e+00  3.407e+00   2.507   0.0197 *
I(money^2)       -8.048e-04  8.243e-04  -0.976   0.3390
I(money^3)        2.669e-06  2.684e-06   0.994   0.3304
I(sex^2)                 NA         NA      NA       NA
I(sex^3)                 NA         NA      NA       NA
I(work^2)         1.720e+00  1.100e+00   1.564   0.1316
I(work^3)        -2.021e-01  1.284e-01  -1.573   0.1293
I(love^2)        -5.241e-01  5.276e-01  -0.993   0.3309
I(love^3)                NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Step:  AIC=-4.44
happy ~ I(money * love) + I(sex * work) + I(love * work) + money +
    love + I(work^2) + I(work^3)


                   Df Sum of Sq    RSS      AIC
<none>                          23.089  -4.4434
- I(sex * work)     1    1.8222 24.912  -3.4810
- I(love * work)    1    3.2087 26.298  -1.3686
- I(work^3)         1    6.4096 29.499   3.1109
- I(work^2)         1    9.9004 32.990   7.4727
- I(money * love)   1   12.6519 35.741  10.5970
- money             1   15.0833 38.173  13.1637
- love              1   22.8757 45.965  20.4085


Call:
lm(formula = happy ~ I(money * love) + I(sex * work) + I(love *
    work) + money + love + I(work^2) + I(work^3), data = happy)


Coefficients:
    (Intercept)  I(money * love)     I(sex * work)    I(love * work)
       -9.05691         -0.03646          -0.15679          -0.45701
          money             love         I(work^2)         I(work^3)
        0.11117          5.72296           0.74993          -0.09492

> #use log to transform the response happy
> lmod_final=lm(log(happy)~I(money*love)+I(love*work)+love+I(work^2)+I(work^3)+mon
ey, happy)
> summary(lmod_final)


Call:
lm(formula = log(happy) ~ I(money * love) + I(love * work) +
    love + I(work^2) + I(work^3) + money, data = happy)


Residuals:
```

```
> #use log to transform the response happy
> lmod_final=lm(log(happy)~I(money*love)+I(love*work)+love+I(work^2)+I(work^3)+mon
ey, happy)
> summary(lmod_final)

Call:
lm(formula = log(happy) ~ I(money * love) + I(love * work) +
    love + I(work^2) + I(work^3) + money, data = happy)

Residuals:
     Min      1Q   Median      3Q      Max
-0.35850 -0.09742 -0.03041  0.07352  0.40285

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -1.295698   0.420699  -3.080 0.004232 **
I(money * love)  -0.007607   0.001651  -4.606 6.21e-05 ***
I(love * work)   -0.117103   0.041221  -2.841 0.007762 **
love              1.174708   0.192147   6.114 7.82e-07 ***
I(work^2)         0.151376   0.035642   4.247 0.000174 ***
I(work^3)        -0.017396   0.005417  -3.211 0.003007 **
money             0.022697   0.004581   4.954 2.27e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1618 on 32 degrees of freedom
Multiple R-squared:  0.811,     Adjusted R-squared:  0.7756
F-statistic: 22.89 on 6 and 32 DF,  p-value: 2.732e-10
```

```
> #2(b)
> data(cornnit, package="faraway")
> #use polynomial transformation
> lmod_good=lm(yield~nitrogen+I(nitrogen^2)+I(nitrogen^3), cornnit)
> summary(lmod_good)

Call:
lm(formula = yield ~ nitrogen + I(nitrogen^2) + I(nitrogen^3),
    data = cornnit)

Residuals:
    Min      1Q  Median      3Q     Max
-42.885  -7.892  -0.111   9.709  26.573

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.989e+01  4.423e+00  20.322  < 2e-16 ***
nitrogen       9.497e-01  1.588e-01   5.982 5.01e-07 ***
I(nitrogen^2) -5.184e-03  1.424e-03  -3.640 0.000773 ***
I(nitrogen^3)  8.618e-06  3.261e-06   2.643 0.011673 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.93 on 40 degrees of freedom
Multiple R-squared:  0.6959,    Adjusted R-squared:  0.6731
F-statistic: 30.52 on 3 and 40 DF,  p-value: 1.976e-10
```

```
> #3(a)
> data(gala, package="faraway")
> lmod=lm(Species~Area, gala)
> summary(lmod)

Call:
lm(formula = Species ~ Area, data = gala)

Residuals:
    Min      1Q  Median      3Q     Max
-99.495 -53.431 -29.045   3.423 306.137

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 63.78286   17.52442   3.640 0.001094 **
Area         0.08196    0.01971   4.158 0.000275 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 91.73 on 28 degrees of freedom
Multiple R-squared:  0.3817,    Adjusted R-squared:  0.3596
F-statistic: 17.29 on 1 and 28 DF,  p-value: 0.0002748

> #Try transform the response, add log to it
> lmod_final=lm(log(Species)~Area+I(Area^2)+I(Area^3)+I(Area^4),gala)
> summary(lmod_final)

Call:
lm(formula = log(Species) ~ Area + I(Area^2) + I(Area^3) + I(Area^4),
    data = gala)

Residuals:
    Min      1Q  Median      3Q     Max
-1.9539 -0.4277  0.2544  0.5686  1.5287
```

```
Residual standard error: 91.73 on 28 degrees of freedom
Multiple R-squared:  0.3817,    Adjusted R-squared:  0.3596
F-statistic: 17.29 on 1 and 28 DF,  p-value: 0.0002748

> #Try transform the response, add log to it
> lmod_final=lm(log(Species)~Area+I(Area^2)+I(Area^3)+I(Area^4),gala)
> summary(lmod_final)

Call:
lm(formula = log(Species) ~ Area + I(Area^2) + I(Area^3) + I(Area^4),
    data = gala)

Residuals:
    Min      1Q  Median      3Q     Max
-1.9539 -0.4277  0.2544  0.5686  1.5287

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.642e+00  2.456e-01  10.758 7.2e-11 ***
Area         2.986e-02  8.135e-03   3.671 0.00115 **
I(Area^2)   -7.639e-05  2.579e-05  -2.961 0.00662 **
I(Area^3)    6.166e-08  2.311e-08   2.668 0.01320 *
I(Area^4)   -9.988e-12  3.862e-12  -2.586 0.01592 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.073 on 25 degrees of freedom
Multiple R-squared:  0.5943,    Adjusted R-squared:  0.5294
F-statistic: 9.155 on 4 and 25 DF,  p-value: 0.0001068

> #lmod_final is a good model
> par(mfrow=c(2,2))
> plot(lmod_final)
```
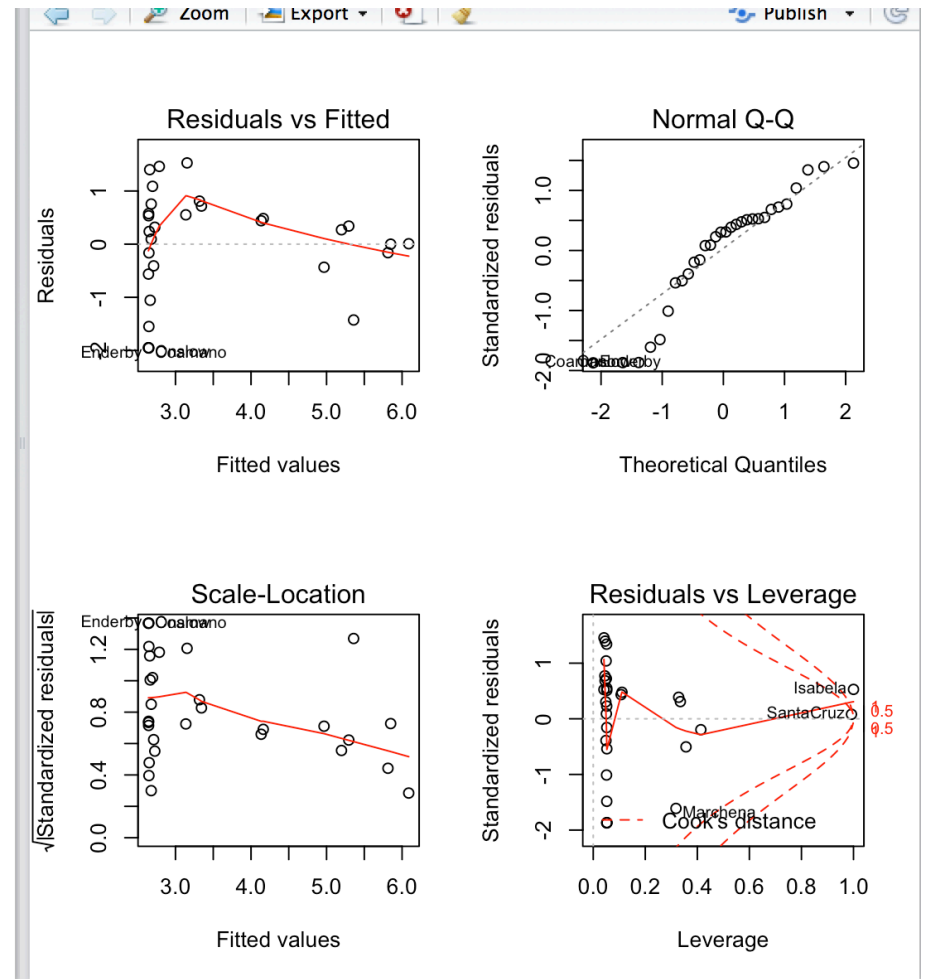
```
> #3(b)
> lmod2=lm(log(Species)~Area+I(Area^2)+I(Area^3),gala)
> summary(lmod2)

Call:
lm(formula = log(Species) ~ Area + I(Area^2) + I(Area^3), data = gala)

Residuals:
     Min      1Q    Median      3Q      Max
-2.16072 -0.54319  0.09771  0.95492  1.63210

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.852e+00  2.558e-01  11.148 2.11e-11 ***
Area         1.125e-02  4.185e-03   2.687   0.0124 *
I(Area^2)   -1.142e-05  6.457e-06  -1.769   0.0887 .
I(Area^3)    1.959e-09  1.198e-09   1.636   0.1139
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.184 on 26 degrees of freedom
Multiple R-squared:  0.4857,    Adjusted R-squared:  0.4264
F-statistic: 8.186 on 3 and 26 DF,  p-value: 0.0005319

> #It's not a better model then the first model because the first model has a bigger Adjust
ed R-squared
> plot(lmod2)
```

Residuals vs Fitted

Normal Q-Q

Scale-Location

Residuals vs Leverage