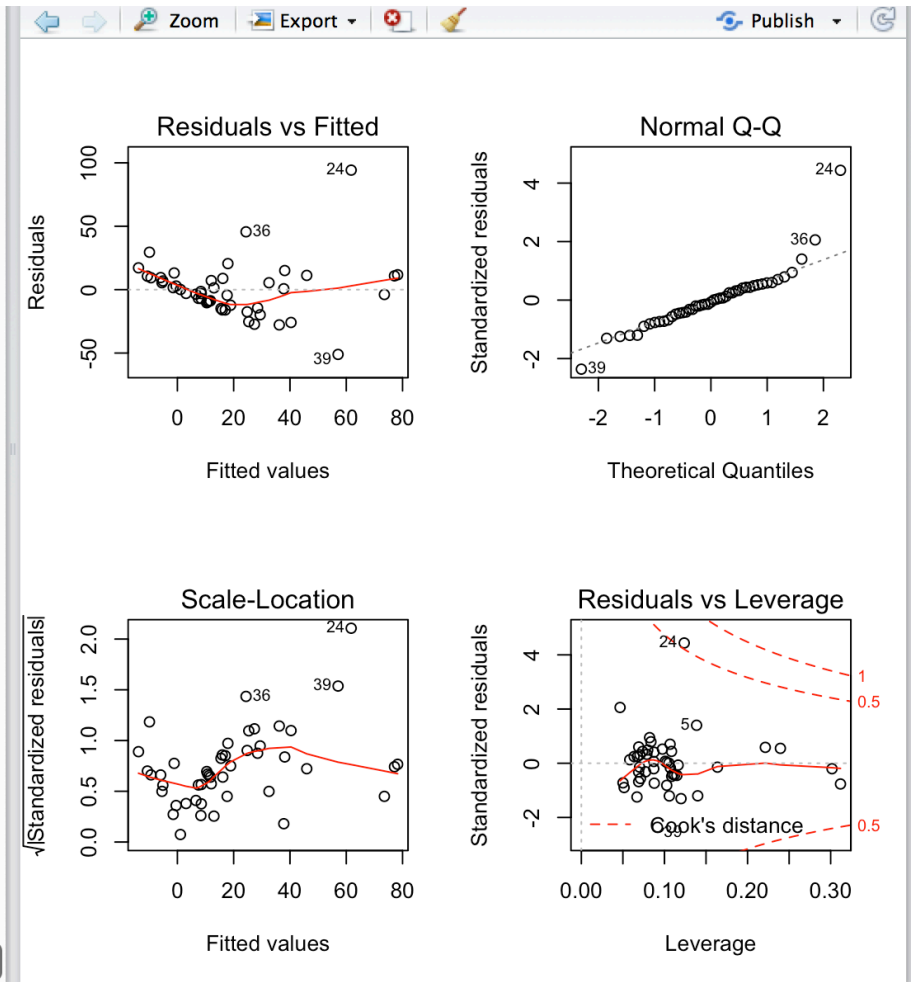


```

> #1(teengamb)
> data(teengamb, package="faraway")
> lmod=lm(gamble~sex+status+income+verbal,teengamb)
> #check the constant variance assumption for the errors
> par(mfrow=c(2,2))
> plot(lmod)
> #We can check the constant variance assumption for the errors from the Residuals
vs Fitted graph
> #We can check the normality assumption from the Normal Q-Q plot
> #We can check the large leverage points from the scale-location plot
> #From the residuals vs Fitted plot, and Normal Q-Q plot, we can see that they're
three outliers, 24, 36 and 39.
> #According the Residuals vs Fitted Plot, and the Scale-Location plot, we can see
there's a linear relationship between predictors and response.
> #1(teengamb)
> data(teengamb, package="faraway")
> lmod=lm(gamble~sex+status+income+verbal,teengamb)
> #check the constant variance assumption for the errors
> par(mfrow=c(2,2))
> plot(lmod)
> #We can check the constant variance assumption for the errors from the Residuals
vs Fitted graph
> #We can check the normality assumption from the Normal Q-Q plot
> #We can check the large leverage points from the scale-location plot
> #From the residuals vs Fitted plot, and Normal Q-Q plot, we can see that they're
three outliers, 24, 36 and 39.
> #We can check for influential points(cook's distance) on the residuals vs leverage
ge plot
> #According the Residuals vs Fitted Plot, and the Scale-Location plot, we can see
there's a linear relationship between predictors and response.
>
>
>
>
> |

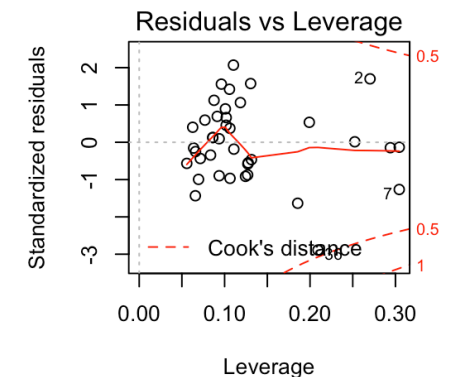
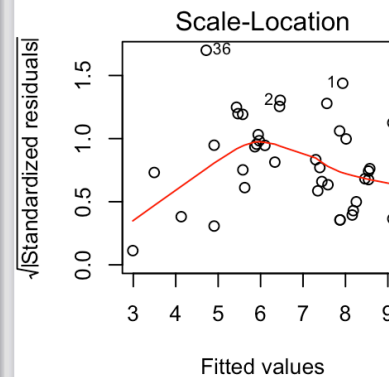
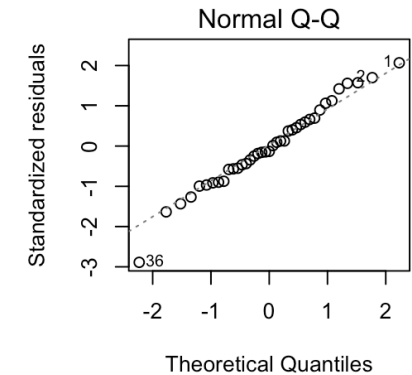
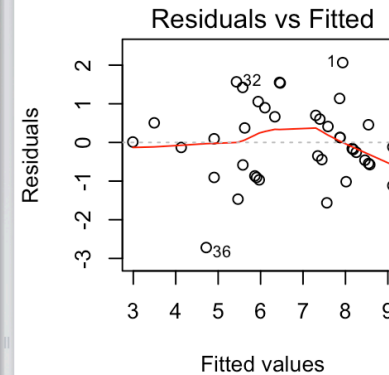
```



```

>
> #1(happy)
> data(happy, package="faraway")
> lmod=lm(happy~money+sex+love+work, happy)
> #check the constant variance assumption for the errors
> par(mfrow=c(2,2))
> plot(lmod)
> #We can check the constant variance assumption for the errors from the Residuals
vs Fitted graph
> #We can check the normality assumption from the Normal Q-Q plot
> #We can check the large leverage points from the scale-location plot
> #From the residuals vs Fitted plot, and Normal Q-Q plot, we can see that there's
one outlier, 36.
> #We can check for influential points(cook's distance) on the residuals vs leverage
ge plot
> #According the Residuals vs Fitted Plot, and the Scale-Location plot, we can see
there's a linear relationship between predictors and response.

```



```

-
> #2
> data(swiss, package="datasets")
> par(mfrow=c(1,1))
> lmod=lm(Fertility~Agriculture+Examination+Education+Catholic+Infant.Mortality, swiss)
> plot(residuals(lmod)~rstandard(lmod), xlab=expression(r[i]), ylab=expression(hat(epsilon[
i])))
> #The sums of residuals are not exact, but also depends on the leverage and the sigma hat
> internal_studentized=rstandard(lmod)
> external_studentized=rstudent(lmod)
> external_studentized
  Courtelary      Delemont Franches-Mnt      Moutier      Neuveville
0.84584030  0.08684212  1.01074888  1.31796269  1.81614211
  Porrentruy      Broye      Glane      Gruyere      Sarine
-2.36721827  0.61438409  1.66113713  0.15298493  0.50049516
  Veveyse      Aigle      Aubonne      Avenches      Cossonay
0.51646133  0.73437530  0.07248412  0.42104970 -0.54614038
  Echallens      Grandson      Lausanne      La Vallee      Lavaux
-0.74687721  0.01511285  0.02925763  0.61245225  0.24191247
  Morges      Moudon      Nyone      Orbe      Oron
0.47865752 -1.60418331 -0.71558271 -0.99565881 -0.15431971
  Payerne Paysd'enhaut      Rolle      Vevey      Yverdon
0.24875019  0.07942498 -0.21167094 -0.77601876 -1.02128559
  Conthey      Entremont      Herens      Martigwy      Monthey
-0.11516331 -1.12779737 -0.18491467 -0.85417952 -0.56297196
  St Maurice      Sierre      Sion      Boudry La Chauxdfnd
-1.20389559  2.44522679  1.22241299  0.69339739 -1.00648626
  Le Locle      Neuchatel      Val de Ruz ValdeTravers V. De Geneve
0.60322648  1.74541155  0.73018972 -0.83329565  0.03781782
  Rive Droite      Rive Gauche
-1.54311369 -2.39447054
> critical_value=qt(0.975, 40)
> which(abs(external_studentized)>critical_value)
  Porrentruy      Sierre Rive Gauche
           6           37           47
>
> #Neuveville, Porrentruy, Sierre, Neuchatel, Rive Gauche are the outliers
>

```

```
>
> #Use Bonferroni
> #Find Bonferroni critical value
> critical_value1=qt(0.05/(47*2), 40)
> critical_value1
[1] -3.529468
> which(abs(external_studentized)>-(critical_value1))
named integer(0)
> #According to Bonferroni, there are no outliers
>
```

```
> #4
> data(seatpos, package="faraway")
> lmod=lm(hipcenter~Leg+Thigh+Arm+Seated+Ht+HtShoes+Weight+Age, seatpos)
> summary(lmod)
```

Call:

```
lm(formula = hipcenter ~ Leg + Thigh + Arm + Seated + Ht + HtShoes +
    Weight + Age, data = seatpos)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-73.827	-22.833	-3.678	25.017	62.337

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	436.43213	166.57162	2.620	0.0138 *
Leg	-6.43905	4.71386	-1.366	0.1824
Thigh	-1.14312	2.66002	-0.430	0.6706
Arm	-1.32807	3.90020	-0.341	0.7359
Seated	0.53375	3.76189	0.142	0.8882
Ht	0.60134	10.12987	0.059	0.9531
HtShoes	-2.69241	9.75304	-0.276	0.7845
Weight	0.02631	0.33097	0.080	0.9372
Age	0.77572	0.57033	1.360	0.1843

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37.72 on 29 degrees of freedom

Multiple R-squared: 0.6866, Adjusted R-squared: 0.6001

F-statistic: 7.94 on 8 and 29 DF, p-value: 1.306e-05

```
> #No, they don't look significant
```

```
> #Find the F-score on the respective degrees of Freedom
```

```
> qf(0.95,8, 29)
```

```
[1] 2.278251
```

```
> #F-statistic greater than 2.278251, so at least one of the variables is significant
```

```
> vif(lmod)
      Leg      Thigh      Arm      Seated      Ht      HtShoes      Weight
6.694291  2.762886  4.496368  8.951054 333.137832 307.429378  3.647030
      Age
1.997931
> #threshold of 10, so we leave out Ht&HtShoes
> lmod1=lm(hipcenter~Leg+Thigh+Arm+Seated+Weight+Age, seatpos)
> summary(lmod1)
```

```
Call:
lm(formula = hipcenter ~ Leg + Thigh + Arm + Seated + Weight +
    Age, data = seatpos)
```

```
Residuals:
      Min       1Q   Median       3Q      Max
-68.296 -23.340  -5.672   24.183   74.065
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  409.00851  159.49517   2.564   0.0154 *
Leg          -8.40876   3.91939  -2.145   0.0399 *
Thigh        -1.91970   2.24858  -0.854   0.3998
Arm          -2.00541   3.69731  -0.542   0.5914
Seated       -1.73576   2.48225  -0.699   0.4896
Weight       -0.03251   0.31254  -0.104   0.9178
Age           0.83110   0.52771   1.575   0.1254
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 36.91 on 31 degrees of freedom
Multiple R-squared:  0.6791,    Adjusted R-squared:  0.617
F-statistic: 10.94 on 6 and 31 DF,  p-value: 1.571e-06
```

```
> qf(0.95,6,31)
[1] 2.409432
> #Leg seems significant, also for the overall F test it is significant
> vif(lmod1)
      Leg      Thigh      Arm      Seated      Weight      Age
4.832701 2.061632 4.219519 4.069626 3.396124 1.786192
> #Yes, they have changed. On a threshold of 10, none of them have problems
> |
```