

1)

高p值表示我们接受原假设。

数据表明，只有报纸接受原假设，即报纸对销售的影响不显著，其余因素都显著影响销售。

5)

$$\therefore \hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\therefore \hat{y}_i = \hat{\beta} x_i = \frac{x_i \sum x_j y_j}{\sum x_j^2}$$

$$\therefore a_{i'} = \frac{x_i x_{i'}}{\sum x_i^2}$$

8.a)

```
> Auto=read.csv("Auto.csv",header=T,na.strings="?")
> Auto=na.omit(Auto)
> attach(Auto)
> summary(Auto)
```

mpg	cylinders	displacement	horsepower	weight
acceleration	year	origin		name
Min. : 9.00	Min. :3.000	Min. : 68.0	Min. : 46.0	Min. :1613
Min. : 8.00	Min. :70.00	Min. :1.000	amc matador	: 5
1st Qu.:17.00	1st Qu.:4.000	1st Qu.:105.0	1st Qu.: 75.0	1st Qu.:2225
1st Qu.:13.78	1st Qu.:73.00	1st Qu.:1.000	ford pinto	: 5
Median :22.75	Median :4.000	Median :151.0	Median : 93.5	Median :2804
Median :15.50	Median :76.00	Median :1.000	toyota corolla	: 5
Mean :23.45	Mean :5.472	Mean :194.4	Mean :104.5	Mean :2978
Mean :15.54	Mean :75.98	Mean :1.577	amc gremlin	: 4
3rd Qu.:29.00	3rd Qu.:8.000	3rd Qu.:275.8	3rd Qu.:126.0	3rd Qu.:3615
3rd Qu.:17.02	3rd Qu.:79.00	3rd Qu.:2.000	amc hornet	: 4
Max. :46.60	Max. :8.000	Max. :455.0	Max. :230.0	Max. :5140
Max. :24.80	Max. :82.00	Max. :3.000	chevrolet chevette	: 4
			(Other)	:365

```
> lm.fit=lm(mpg~horsepower)
> summary(lm.fit)
```

Call:

```
lm(formula = mpg ~ horsepower)
```

Residuals:

Min	1Q	Median	3Q	Max
-13.5710	-3.2592	-0.3435	2.7630	16.9240

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	39.935861	0.717499	55.66	<2e-16 ***

```
horsepower -0.157845 0.006446 -24.49 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.906 on 390 degrees of freedom
Multiple R-squared: 0.6059, Adjusted R-squared: 0.6049
F-statistic: 599.7 on 1 and 390 DF, p-value: < 2.2e-16
```

零假设  $H_0: \beta = 0$

p值接近0，所以拒绝原假设，两者显著相关

$R^2 = 0.6059$ ，所以有60.59%的mpg能被horsepower解释

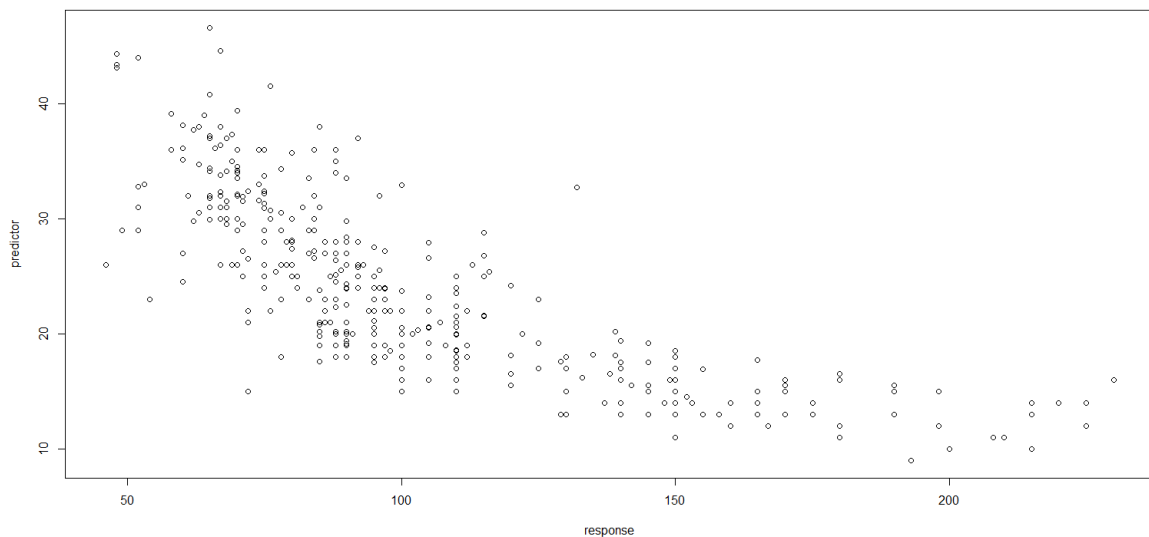
回归系数小于零，说明二者之间是消极的

对于98的预测结果和置信区间如下：

```
> predict(lm.fit2,data.frame(response=c(98)),interval="prediction",level=0.95)
      fit      lwr      upr
1 24.46708 14.8094 34.12476
```

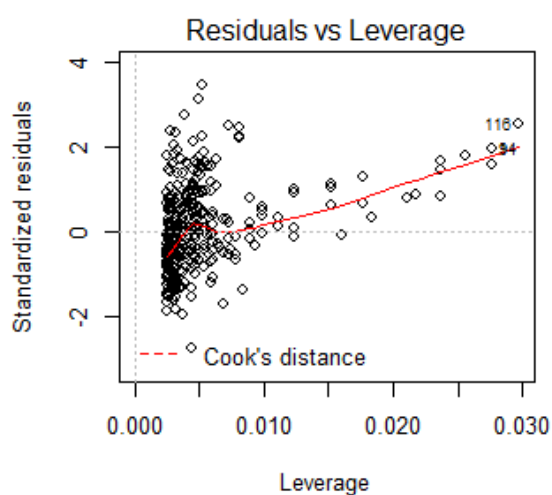
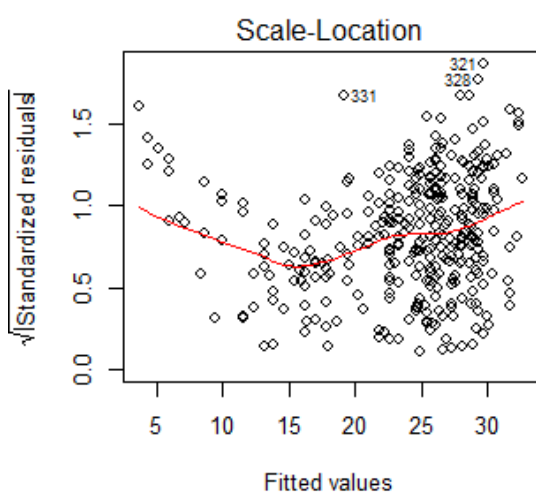
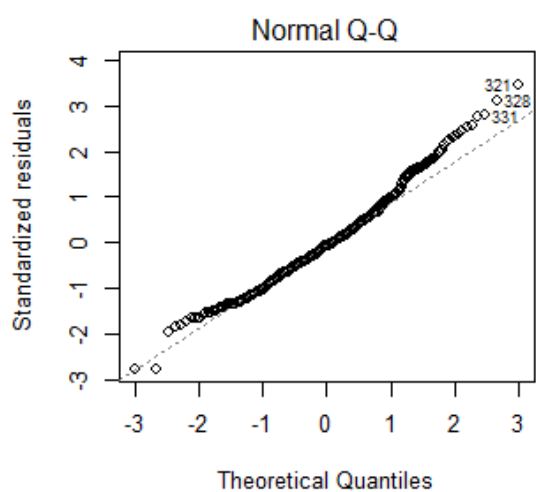
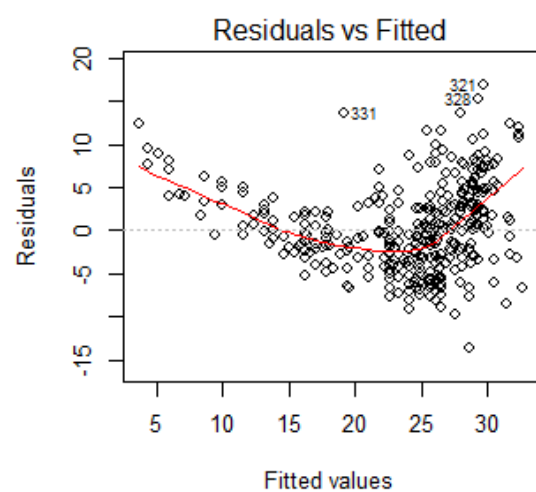
## 8.b)

```
> plot(response,predictor)
> abline(lm.fit2,lwd=3,col="red")
```



## 8.c)

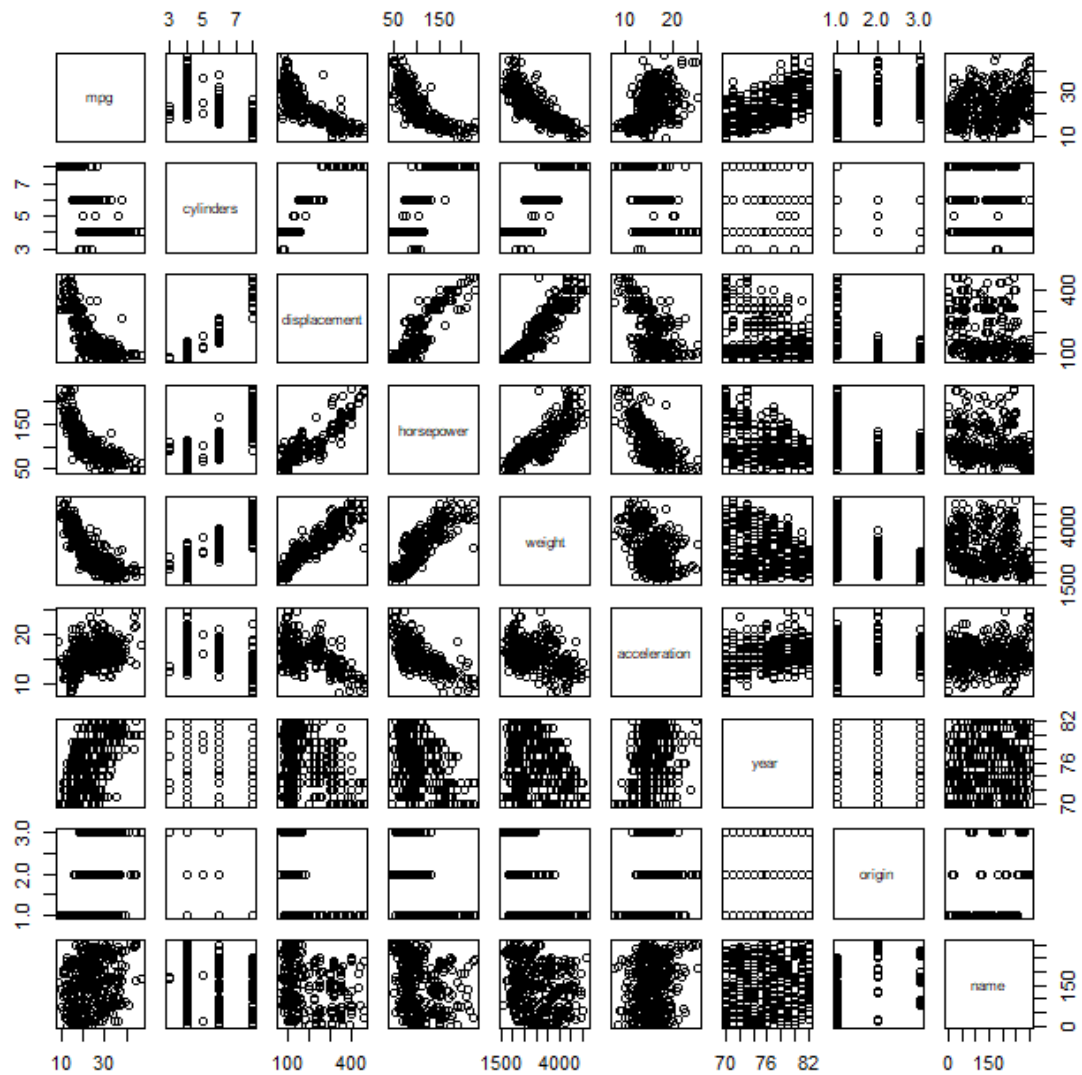
```
> par(mfrow=c(2,2))
> plot(lm.fit2)
```



可以看出两者大概率是非线性相关

## 9.a)

```
> pairs(Auto)
```



9.b)

```
> cor(subset(Auto,select=-name))
```

	mpg	cylinders	displacement	horsepower	weight
mpg	1.0000000	-0.7776175	-0.8051269	-0.7784268	-0.8322442
cylinders	-0.7776175	1.0000000	0.9508233	0.8429834	0.8975273
displacement	-0.8051269	0.9508233	1.0000000	0.8972570	0.9329944
horsepower	-0.7784268	0.8429834	0.8972570	1.0000000	0.8645377
weight	-0.8322442	0.8975273	0.9329944	0.8645377	1.0000000
acceleration	0.4233285	-0.5046834	-0.5438005	-0.6891955	-0.4168392
year	0.5805410	-0.3456474	-0.3698552	-0.4163615	-0.3091199
origin	0.5652088	-0.5689316	-0.6145351	-0.4551715	-0.5850054

	acceleration	year	origin
mpg	0.4233285	0.5805410	0.5652088
cylinders	-0.5046834	-0.3456474	-0.5689316
displacement	-0.5438005	-0.3698552	-0.6145351
horsepower	-0.6891955	-0.4163615	-0.4551715
weight	-0.4168392	-0.3091199	-0.5850054
acceleration	1.0000000	0.2903161	0.2127458
year	0.2903161	1.0000000	0.1815277
origin	0.2127458	0.1815277	1.0000000

## 9.c)

```
> lm.fit3=lm(mpg~.-name,data=Auto)
> summary(lm.fit3)

Call:
lm(formula = mpg ~ . - name, data = Auto)

Residuals:
    Min       1Q   Median       3Q      Max
-9.5903 -2.1565 -0.1169  1.8690 13.0604

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -17.218435   4.644294  -3.707  0.00024 ***
cylinders    -0.493376   0.323282  -1.526  0.12780
displacement  0.019896   0.007515   2.647  0.00844 **
horsepower   -0.016951   0.013787  -1.230  0.21963
weight       -0.006474   0.000652  -9.929 < 2e-16 ***
acceleration  0.080576   0.098845   0.815  0.41548
year          0.750773   0.050973  14.729 < 2e-16 ***
origin        1.426141   0.278136   5.127 4.67e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.328 on 384 degrees of freedom
Multiple R-squared:  0.8215,    Adjusted R-squared:  0.8182
F-statistic: 252.4 on 7 and 384 DF,  p-value: < 2.2e-16
```

零假设  $H_0: \beta_i = 0$

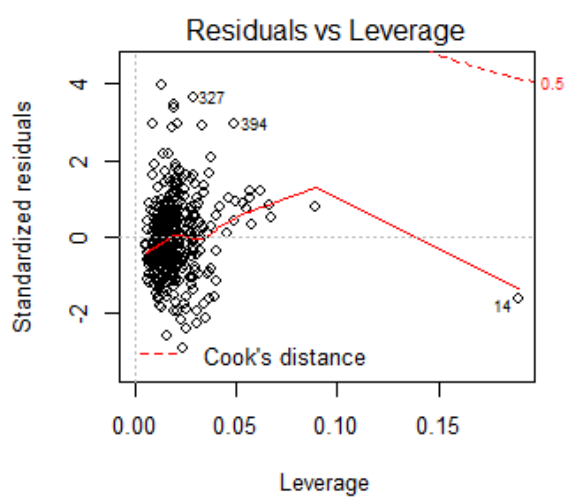
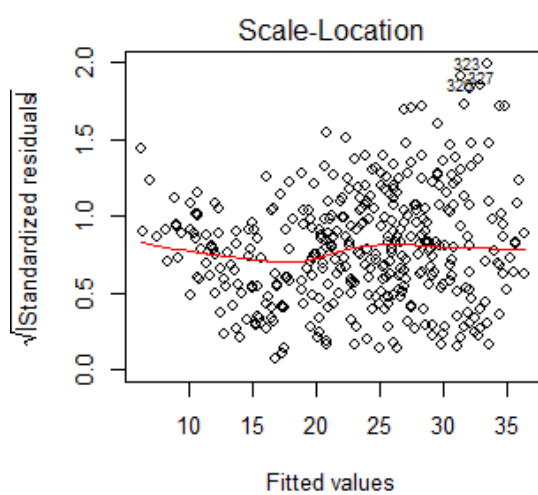
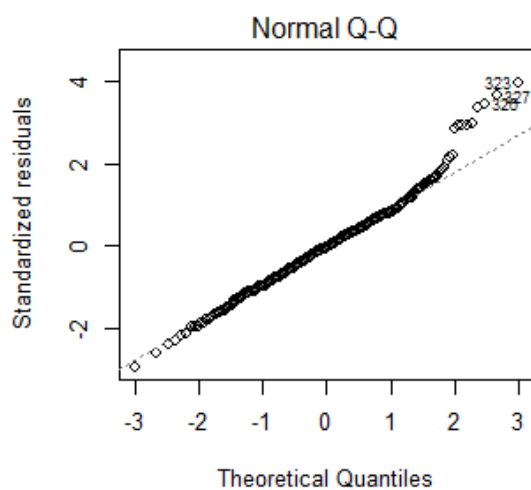
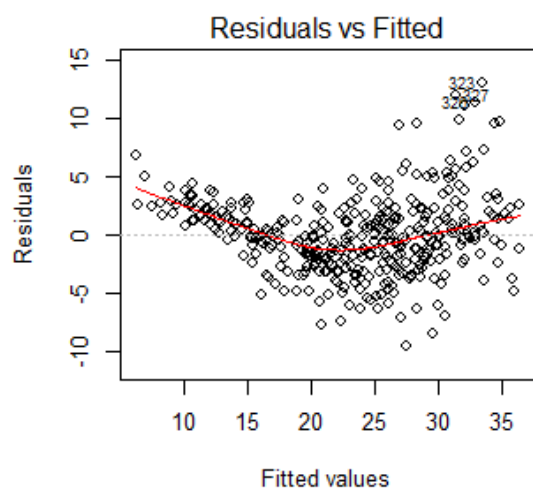
p值接近0，所以拒绝原假设，mpg和其他变量有显著关系

由单个变量的p值可知，displacement、weight、year、origin和mpg有显著关系

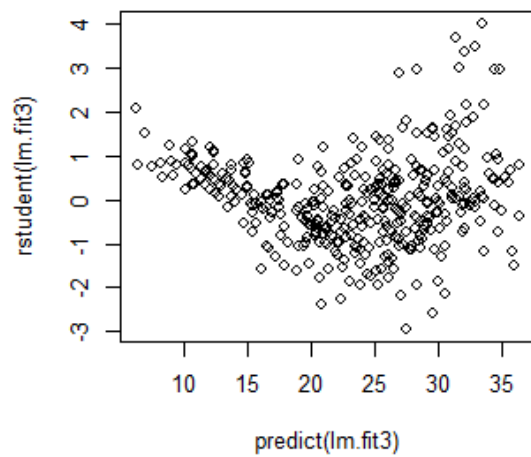
根据year得到结论，能源利用率逐年增长

## 9.d)

```
> par(mfrow=c(2,2))
> plot(lm.fit3)
```



```
> plot(predict(lm.fit3), rstudent(lm.fit3))
```



说明多元回归模型不正确，14号点没有较大的残差但有较大的权重

## 9.e)

```
> lm.fit4=lm(mpg~displacement*weight+year*origin)
> summary(lm.fit4)
```

Call:  
lm(formula = mpg ~ displacement \* weight + year \* origin)

Residuals:

Min	1Q	Median	3Q	Max
-9.5758	-1.6211	-0.0537	1.3264	13.3266

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.793e+01	8.044e+00	2.229	0.026394 *
displacement	-7.519e-02	9.091e-03	-8.271	2.19e-15 ***
weight	-1.035e-02	6.450e-04	-16.053	< 2e-16 ***
year	4.864e-01	1.017e-01	4.782	2.47e-06 ***
origin	-1.503e+01	4.232e+00	-3.551	0.000432 ***
displacement:weight	2.098e-05	2.179e-06	9.625	< 2e-16 ***

```

year:origin      1.980e-01  5.436e-02  3.642 0.000308 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.969 on 385 degrees of freedom
Multiple R-squared:  0.8575,    Adjusted R-squared:  0.8553
F-statistic: 386.2 on 6 and 385 DF,  p-value: < 2.2e-16

```

统计关系显著，且残差变小

## 9.f)

```

> lm.fit5 = lm(mpg~log(horsepower)+sqrt(horsepower)+horsepower+I(horsepower^2))
> summary(lm.fit5)

Call:
lm(formula = mpg ~ log(horsepower) + sqrt(horsepower) + horsepower +
    I(horsepower^2))

Residuals:
    Min       1Q   Median       3Q      Max
-15.3450  -2.4725  -0.1594   2.1068  16.2564

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -6.839e+02  2.439e+02  -2.804  0.00530 **
log(horsepower)  6.515e+02  2.111e+02   3.085  0.00218 **
sqrt(horsepower) -3.385e+02  1.092e+02  -3.101  0.00207 **
horsepower      1.165e+01  3.898e+00   2.988  0.00299 **
I(horsepower^2) -7.425e-03  2.796e-03  -2.655  0.00825 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.331 on 387 degrees of freedom
Multiple R-squared:  0.6952,    Adjusted R-squared:  0.692
F-statistic: 220.6 on 4 and 387 DF,  p-value: < 2.2e-16

> par(mfrow=c(2,2))
> plot(lm.fit5)

```



