**Chapter 3: Linear Regression**

**1**

表3-4中的P值对应的零假设，从上到下分别为：截距估计值为0；TV的系数估计值为0，即当在TV，radio和newspaper都投放广告时，TV对销售没有影响；radio的系数估计值为0，即当在TV，radio和newspaper都投放广告时，radio对销售没有影响；newspaper的系数估计值为0，即当在TV，radio和newspaper都投放广告时，newspaper对销售没有影响。

P值表明，TV和radio的系数估计值是显著不为0的，说明在TV，radio和newspaper都投放广告时，TV和radio对销售有正向影响，而newspaper对销售没有什么影响。

**2**

KNN classifier: Given a positive integer K and a test observation x0, the KNN classifier first identifies the K points in the training data that are closest to x0, represented by N0. It then estimates the conditional probability for class j as the fraction of points in N0 whose response values equal j:



Finally, KNN applies Bayes rule and classifies the test observation x0 to the class with the largest probability.

KNN regression: Given a value for K and a prediction point x0, KNN regression first identifies the K training observations that are closest to x0, represented by N0. It then estimates f(x0) using the average of all the training responses in N0. In other words,

.

In other word, the final result of KNN classifier is the classification output for Y (qualitative), but KNN regression predicts the quantitative value for f(X).

**3**

首先把模型写出来



X1=GPA, X2=IQ, X3=Gender, (1 for Female and 0 for Male), X4=Interaction between GPA and IQ, X5=Interaction between GPA and Gender.

**(a)**

**iii**是对的。

给定GPA和IQ的值，对于女性来说，模型可以写为



对于男性来说，模型可以写为



如果GPA较高，高于3.5，则平均来看，女性的收入低于男性，反之，如果GPA较低，则平均来看，女性的收入高于男性。

**(b)**

X1=4.0，X2=110时，X3=1时，代入模型得到f(X)=137.1。

**(c)**

False，应该先对变量进行标准化，然后进行假设检验，根据P值判断每个系数统计上是否显著不为0。

**4(a)**

含三次项的回归Training RSS更小，因为含三次项的模型相比线性模型更加flexible，更接近原始观测数据，对训练集拟合得更好。

**4(b)**

线性回归的Test RSS更小，因为X和Y真实的关系本来就是线性的，线性模型能给出一个较为精确的估计，三次模型会出现过拟合的问题。

**4(c)**

含三次项的回归Training RSS更小，回归模型越flexible，越接近训练集的数据，对训练集拟合得越好，Training RSS会更小。

**4(d)**

无法确定哪个模型的Test RSS更小，如果真实的数据更接近线性关系，那么，线性回归的Test RSS更小，反之，如果如果真实的数据更接近三次关系，那么含三次项的回归的Test RSS更小。

**5**





**6**

线性回归模型：

由公式(3.4)知 

即

则线性回归模型可化为

时，。即简单线性回归模型中，最小二乘线一定过点。

**7**

证明：





从而



又。

从而

**8(a)**

> attach(Auto)

> fit8<-lm(mpg~horsepower)

> summary(fit8)

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

**i.**

Yes, there is a relationship between horsepower and mpg as determined by testing the null hypothesis of all regression coefficients equal to zero. Since the F-statistic is far larger than 1 and the p-value of the F-statistic is close to zero we can reject the null hypothesis and state there is a statistically significant relationship between horsepower and mpg.

F-statistic远大于1，其P值非常小，可以拒绝原假设认为mpg和horsepowe之间是有相关性的。

**ii.**

mpg均值为23.44592，RSE的值为4.906，误差百分比约为20.92%。R2值为0.6059，表明mpg的60.59%的变异可以由horsepower解释。

**iii.**

负相关

**iv.**

> predict(fit8,data.frame(horsepower = c(98)),interval = "confidence")

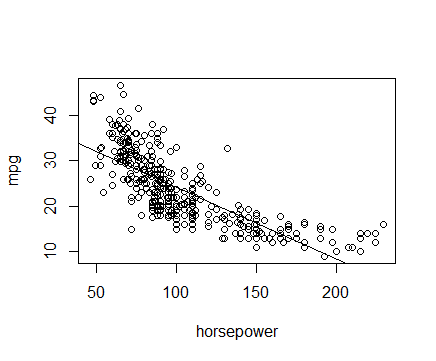
fit lwr upr

1 24.46708 23.97308 24.96108

**8(b).**

> plot(horsepower,mpg)

> abline(fit8)



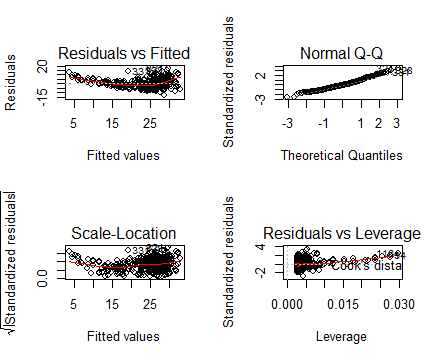
**8(c).**

> opar<-par(no.readonly = TRUE)

> par(mfrow=c(2,2))

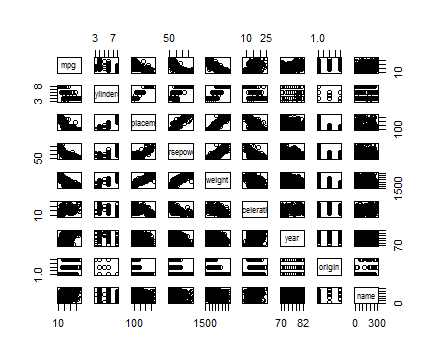
> plot(fit8)

> par(opar)



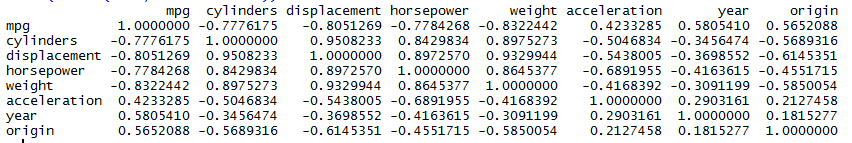
漏斗状残差显示存在异方差，且数据中存在非线性关系。

**9(a).**



**9(b).**

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



**9(c).**

> fit9<-lm(mpg~.-name,data = Auto)

> summary(fit9)

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

**i.**

F-statistic远大于1，其P值非常小，可以拒绝零假设，即至少有一个预测变量与相应变量是显著相关的。

**ii.**

displacement, weight, year, and origin

**iii.**

系数是0.750773，表示每过一年，mpg增长为0.750773。

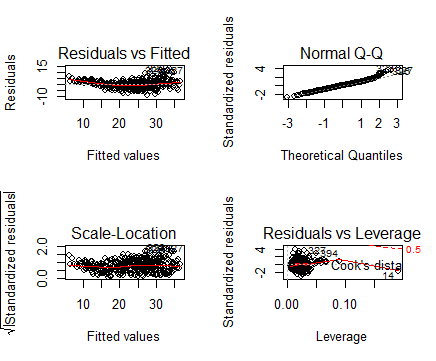
**9(d).**

> opar<-par(no.readonly = TRUE)

> par(mfrow=c(2,2))

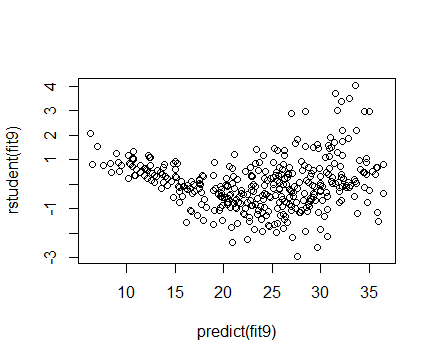
> plot(fit9)

> par(opar)



残差图呈U型，数据存在非线性关系。点14有高杠杆，但是残差值不高。

> plot(predict(fit9), rstudent(fit9))



学生化残差绝对值大于3的都可以认为是离群点。

**9(e).**

> fit\_9 = lm(mpg~cylinders\*displacement+displacement\*weight)

> summary(fit\_9)

Call:

lm(formula = mpg ~ cylinders \* displacement + displacement \*

weight)

Residuals:

Min 1Q Median 3Q Max

-13.2934 -2.5184 -0.3476 1.8399 17.7723

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 5.262e+01 2.237e+00 23.519 < 2e-16 \*\*\*

cylinders 7.606e-01 7.669e-01 0.992 0.322

displacement -7.351e-02 1.669e-02 -4.403 1.38e-05 \*\*\*

weight -9.888e-03 1.329e-03 -7.438 6.69e-13 \*\*\*

cylinders:displacement -2.986e-03 3.426e-03 -0.872 0.384

displacement:weight 2.128e-05 5.002e-06 4.254 2.64e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.103 on 386 degrees of freedom

Multiple R-squared: 0.7272, Adjusted R-squared: 0.7237

F-statistic: 205.8 on 5 and 386 DF, p-value: < 2.2e-16

找到相关系数最大的两对变量进行回归分析，displacement\*weight项的系数的P值是显著的，cylinders\*displacement项不是。即displacement和weight的interaction统计上是显著的。

**9(f)**

> fit9=lm(mpg~year+displacement+poly(horsepower,2)+origin+I(acceleration^2)+poly(weight,2))

> summary(fit9)

Call:

lm(formula = mpg ~ year + displacement + poly(horsepower, 2) +

origin + I(acceleration^2) + poly(weight, 2))

Residuals:

Min 1Q Median 3Q Max

-8.861 -1.671 -0.192 1.477 12.087

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -36.613268 3.954699 -9.258 < 2e-16 \*\*\*

year 0.776721 0.045347 17.129 < 2e-16 \*\*\*

displacement 0.003134 0.005109 0.613 0.539948

poly(horsepower, 2)1 -34.318036 10.022455 -3.424 0.000683 \*\*\*

poly(horsepower, 2)2 20.153941 4.542625 4.437 1.20e-05 \*\*\*

origin 0.753115 0.251999 2.989 0.002984 \*\*

I(acceleration^2) -0.003020 0.002965 -1.018 0.309089

poly(weight, 2)1 -78.646081 11.370216 -6.917 1.95e-11 \*\*\*

poly(weight, 2)2 20.314874 3.887637 5.226 2.86e-07 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.916 on 383 degrees of freedom

Multiple R-squared: 0.8633, Adjusted R-squared: 0.8605

F-statistic: 302.4 on 8 and 383 DF, p-value: < 2.2e-16

> fit9=lm(mpg~year+I(displacement^2)+poly(horsepower,2)+origin+poly(acceleration,2)+poly(weight,2))

> summary(fit9)

Call:

lm(formula = mpg ~ year + I(displacement^2) + poly(horsepower,

2) + origin + poly(acceleration, 2) + poly(weight, 2))

Residuals:

Min 1Q Median 3Q Max

-9.3986 -1.5413 0.0441 1.4558 12.4138

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -3.707e+01 3.467e+00 -10.692 < 2e-16 \*\*\*

year 7.840e-01 4.433e-02 17.685 < 2e-16 \*\*\*

I(displacement^2) -3.098e-06 1.056e-05 -0.293 0.769423

poly(horsepower, 2)1 -4.382e+01 1.015e+01 -4.316 2.03e-05 \*\*\*

poly(horsepower, 2)2 1.491e+01 4.765e+00 3.128 0.001892 \*\*

origin 6.957e-01 2.405e-01 2.893 0.004038 \*\*

poly(acceleration, 2)1 -1.372e+01 5.522e+00 -2.485 0.013390 \*

poly(acceleration, 2)2 1.400e+01 3.652e+00 3.832 0.000149 \*\*\*

poly(weight, 2)1 -6.984e+01 1.068e+01 -6.541 1.97e-10 \*\*\*

poly(weight, 2)2 2.323e+01 3.887e+00 5.977 5.23e-09 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.855 on 382 degrees of freedom

Multiple R-squared: 0.8693, Adjusted R-squared: 0.8662

F-statistic: 282.2 on 9 and 382 DF, p-value: < 2.2e-16

> fit9=lm(mpg~year+I(displacement^2)+poly(horsepower,2)+origin+log(acceleration)+poly(weight,2))

> summary(fit9)

Call:

lm(formula = mpg ~ year + I(displacement^2) + poly(horsepower,

2) + origin + log(acceleration) + poly(weight, 2))

Residuals:

Min 1Q Median 3Q Max

-9.0740 -1.5626 -0.1437 1.4685 12.1186

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -2.679e+01 5.968e+00 -4.489 9.48e-06 \*\*\*

year 7.729e-01 4.480e-02 17.251 < 2e-16 \*\*\*

I(displacement^2) 4.543e-06 1.050e-05 0.433 0.6654

poly(horsepower, 2)1 -4.570e+01 1.027e+01 -4.450 1.13e-05 \*\*\*

poly(horsepower, 2)2 2.113e+01 4.418e+00 4.783 2.47e-06 \*\*\*

origin 7.641e-01 2.430e-01 3.144 0.0018 \*\*

log(acceleration) -3.636e+00 1.582e+00 -2.298 0.0221 \*

poly(weight, 2)1 -7.023e+01 1.086e+01 -6.464 3.10e-10 \*\*\*

poly(weight, 2)2 1.933e+01 3.747e+00 5.158 4.02e-07 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.896 on 383 degrees of freedom

Multiple R-squared: 0.8652, Adjusted R-squared: 0.8623

F-statistic: 307.2 on 8 and 383 DF, p-value: < 2.2e-16

> fit9=lm(mpg~year+sqrt(displacement)+poly(horsepower,2)+origin+log(acceleration)+poly(weight,2))

> summary(fit9)

Call:

lm(formula = mpg ~ year + sqrt(displacement) + poly(horsepower,

2) + origin + log(acceleration) + poly(weight, 2))

Residuals:

Min 1Q Median 3Q Max

-9.147 -1.544 -0.160 1.467 12.155

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -25.02373 6.67832 -3.747 0.000207 \*\*\*

year 0.76986 0.04486 17.161 < 2e-16 \*\*\*

sqrt(displacement) -0.02546 0.15195 -0.168 0.867018

poly(horsepower, 2)1 -45.04029 10.19395 -4.418 1.30e-05 \*\*\*

poly(horsepower, 2)2 22.11314 4.06559 5.439 9.57e-08 \*\*\*

origin 0.72559 0.25505 2.845 0.004680 \*\*

log(acceleration) -3.96888 1.55107 -2.559 0.010887 \*

poly(weight, 2)1 -65.88933 11.83757 -5.566 4.91e-08 \*\*\*

poly(weight, 2)2 19.30865 3.86188 5.000 8.75e-07 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.896 on 383 degrees of freedom

Multiple R-squared: 0.8651, Adjusted R-squared: 0.8623

F-statistic: 307 on 8 and 383 DF, p-value: < 2.2e-16

对horsepower，acceleration以及weight做变换后，发现displacement的系数在统计上不显著不为0了，而horsepower，horsepower^2，weight，weight^2的系数在统计上都显著不为0了，log(acceleration)，sqrt(acceleration)的系数在统计上也是显著不为0的。不过有log(acceleration)的模型对数据拟合得更好。

**10(a)**

> attach(Carseats)

> fit10<-lm(Sales~Price+Urban+US)

> summary(fit10)

Call:

lm(formula = Sales ~ Price + Urban + US)

Residuals:

Min 1Q Median 3Q Max

-6.9206 -1.6220 -0.0564 1.5786 7.0581

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 13.043469 0.651012 20.036 < 2e-16 \*\*\*

Price -0.054459 0.005242 -10.389 < 2e-16 \*\*\*

UrbanYes -0.021916 0.271650 -0.081 0.936

USYes 1.200573 0.259042 4.635 4.86e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.472 on 396 degrees of freedom

Multiple R-squared: 0.2393, Adjusted R-squared: 0.2335

F-statistic: 41.52 on 3 and 396 DF, p-value: < 2.2e-16

**10(b)**

根据t统计量和P值，可以认为Price和Sales在统计上是显著相关的，且随着Price增大，Sales在减小。门店位置与销售量的关系在统计上并不显著。门店是否在美国与销售量是显著相关的，如果门店在美国，销售量会增加。

**10(c)**

Sales = 13.043469+ -0.054459\*Price + -0.021916\*UrbanYes + 1.200573\*USYes

**10(d)**

Price and USYes

**10(e)**

> fit\_10<-lm(Sales ~ Price + US)

> summary(fit\_10)

Call:

lm(formula = Sales ~ Price + US)

Residuals:

Min 1Q Median 3Q Max

-6.9269 -1.6286 -0.0574 1.5766 7.0515

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 13.03079 0.63098 20.652 < 2e-16 \*\*\*

Price -0.05448 0.00523 -10.416 < 2e-16 \*\*\*

USYes 1.19964 0.25846 4.641 4.71e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.469 on 397 degrees of freedom

Multiple R-squared: 0.2393, Adjusted R-squared: 0.2354

F-statistic: 62.43 on 2 and 397 DF, p-value: < 2.2e-16

**10(f)**

根据RSE和R2，两个模型拟合的效果差不多，(e)的模型略好。

**10(g)**

> confint(fit\_10)

2.5 % 97.5 %

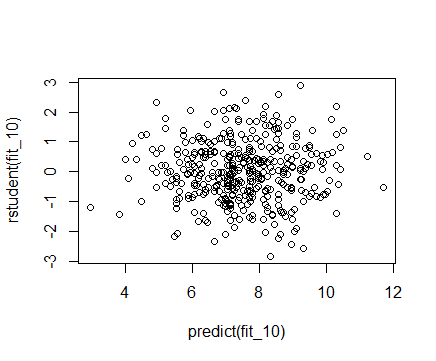
(Intercept) 11.79032020 14.27126531

Price -0.06475984 -0.04419543

USYes 0.69151957 1.70776632

**10(h)**

plot(predict(fit\_10), rstudent(fit\_10))



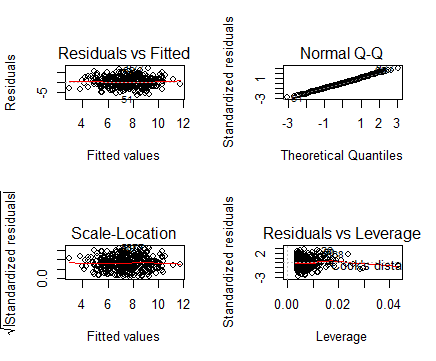
所有点的学生残差值的绝对值都小于3，因此可以认为没有离群点。

> opar<-par(no.readonly = TRUE)

> par(mfrow=c(2,2))

> plot(fit\_10)

> par(opar)



有一些点的杠杆值较大，大于了(P+1)/n。

**11(a)**

> set.seed(1)

> x<-rnorm(100)

> y<-2\*x+rnorm(100)

> fit11<-lm(y~x+0)

> summary(fit11)

Call:

lm(formula = y ~ x + 0)

Residuals:

Min 1Q Median 3Q Max

-1.9154 -0.6472 -0.1771 0.5056 2.3109

Coefficients:

Estimate Std. Error t value Pr(>|t|)

x 1.9939 0.1065 18.73 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9586 on 99 degrees of freedom

Multiple R-squared: 0.7798, Adjusted R-squared: 0.7776

F-statistic: 350.7 on 1 and 99 DF, p-value: < 2.2e-16

P值很小，拒绝零假设。

**11(b)**

> fit\_11<-lm(x~y+0)

> summary(fit\_11)

Call:

lm(formula = x ~ y + 0)

Residuals:

Min 1Q Median 3Q Max

-0.8699 -0.2368 0.1030 0.2858 0.8938

Coefficients:

Estimate Std. Error t value Pr(>|t|)

y 0.39111 0.02089 18.73 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4246 on 99 degrees of freedom

Multiple R-squared: 0.7798, Adjusted R-squared: 0.7776

F-statistic: 350.7 on 1 and 99 DF, p-value: < 2.2e-16

P值很小，拒绝0假设。

**11(c)**

两个模型反映了同一条直线。

**11(d)**



> t<-(sqrt(length(x)-1) \* sum(x\*y)) / (sqrt(sum(x\*x) \* sum(y\*y) - (sum(x\*y))^2))

> t

[1] 18.72593

结果是对的。

**11(e)**

把x和y对换一下就行了，显然是对的。

**11(f)**

> fit\_11\_<-lm(y~x)

> fit\_11\_\_<-lm(x~y)

> summary(fit\_11\_)

Call:

lm(formula = y ~ x)

Residuals:

Min 1Q Median 3Q Max

-1.8768 -0.6138 -0.1395 0.5394 2.3462

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.03769 0.09699 -0.389 0.698

x 1.99894 0.10773 18.556 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9628 on 98 degrees of freedom

Multiple R-squared: 0.7784, Adjusted R-squared: 0.7762

F-statistic: 344.3 on 1 and 98 DF, p-value: < 2.2e-16

> summary(fit\_11\_\_)

Call:

lm(formula = x ~ y)

Residuals:

Min 1Q Median 3Q Max

-0.90848 -0.28101 0.06274 0.24570 0.85736

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.03880 0.04266 0.91 0.365

y 0.38942 0.02099 18.56 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4249 on 98 degrees of freedom

Multiple R-squared: 0.7784, Adjusted R-squared: 0.7762

F-statistic: 344.3 on 1 and 98 DF, p-value: < 2.2e-16

t值是一样的。

**12(a)**

x的所有观测值的平方和与y的所有观测值的平方和相等时系数的估计值是相同的。

**12(b)**

> set.seed(1)

> x<-rnorm(100)

> y<-2\*x

> fit12<-lm(y~x+0)

> fit\_12<-lm(x~y+0)

> summary(fit12)

Call:

lm(formula = y ~ x + 0)

Residuals:

Min 1Q Median 3Q Max

-1.143e-14 -5.650e-17 2.000e-19 6.460e-17 3.469e-16

Coefficients:

Estimate Std. Error t value Pr(>|t|)

x 2.00e+00 1.21e-16 1.652e+16 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.154e-15 on 99 degrees of freedom

Multiple R-squared: 1, Adjusted R-squared: 1

F-statistic: 2.73e+32 on 1 and 99 DF, p-value: < 2.2e-16

> summary(fit\_12)

Call:

lm(formula = x ~ y + 0)

Residuals:

Min 1Q Median 3Q Max

-5.716e-15 -2.830e-17 1.000e-19 3.230e-17 1.734e-16

Coefficients:

Estimate Std. Error t value Pr(>|t|)

y 5.000e-01 3.026e-17 1.652e+16 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.772e-16 on 99 degrees of freedom

Multiple R-squared: 1, Adjusted R-squared: 1

F-statistic: 2.73e+32 on 1 and 99 DF, p-value: < 2.2e-16

系数估计值不同。

**12(c)**

> set.seed(1)

> x <- rnorm(100)

> y <- -sample(x, 100)

> sum(x^2)

[1] 81.05509

> sum(y^2)

[1] 81.05509

> fit\_12\_<-lm(y~x)

> fit\_12\_\_<-lm(x~y)

> summary(fit\_12\_)

Call:

lm(formula = y ~ x)

Residuals:

Min 1Q Median 3Q Max

-2.29058 -0.58434 -0.00216 0.60584 2.32827

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.108130 0.090942 -1.189 0.237

x -0.006955 0.101013 -0.069 0.945

Residual standard error: 0.9027 on 98 degrees of freedom

Multiple R-squared: 4.837e-05, Adjusted R-squared: -0.01016

F-statistic: 0.00474 on 1 and 98 DF, p-value: 0.9452

> summary(fit\_12\_\_)

Call:

lm(formula = x ~ y)

Residuals:

Min 1Q Median 3Q Max

-2.33102 -0.60922 0.00922 0.57929 2.29163

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.108130 0.090942 1.189 0.237

y -0.006955 0.101013 -0.069 0.945

Residual standard error: 0.9027 on 98 degrees of freedom

Multiple R-squared: 4.837e-05, Adjusted R-squared: -0.01016

F-statistic: 0.00474 on 1 and 98 DF, p-value: 0.9452

系数估计值相同。

**13(a)**

> set.seed(1)

> x<-rnorm(100)

**13(b)**

> eps<-rnorm(100,0,sqrt(0.25))

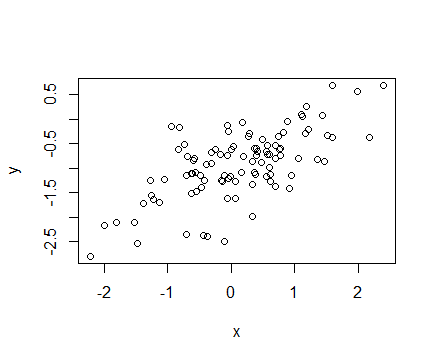
**13(c)**

> y<--1+0.5\*x+eps

The length of y is 100，is -1， is 0.5。

**13(d)**

> plot(x,y)



x和y之间存在线性关系。

**13(e)**

> fit13<-lm(y~x)

> summary(fit13)

Call:

lm(formula = y ~ x)

Residuals:

Min 1Q Median 3Q Max

-1.45706 -0.24115 -0.02266 0.32462 1.32079

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.98632 0.05235 -18.840 < 2e-16 \*\*\*

x 0.51058 0.05815 8.781 5.34e-14 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5197 on 98 degrees of freedom

Multiple R-squared: 0.4403, Adjusted R-squared: 0.4346

F-statistic: 77.1 on 1 and 98 DF, p-value: 5.336e-14

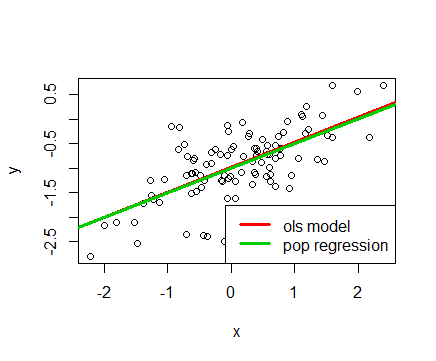
，，系数和斜率的估计值都非常接近和。F统计量很大，P值很小，拒绝零假设。

**13(f)**

> abline(fit13,lwd = 3,col = 2)

> abline(-1,0.5,lwd = 3,col = 3)

legend("bottomright", legend = c("ols model", "pop regression"), col=2:3, lwd=3)



**13(g)**

> fit\_13<-lm(y~x+I(x^2))

> summary(fit\_13)

Call:

lm(formula = y ~ x + I(x^2))

Residuals:

Min 1Q Median 3Q Max

-1.46371 -0.24761 -0.01792 0.32978 1.32271

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.979434 0.064138 -15.271 < 2e-16 \*\*\*

x 0.511912 0.058865 8.696 8.71e-14 \*\*\*

I(x^2) -0.008668 0.046208 -0.188 0.852

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5223 on 97 degrees of freedom

Multiple R-squared: 0.4405, Adjusted R-squared: 0.429

F-statistic: 38.19 on 2 and 97 DF, p-value: 5.856e-13

RSE减小，R2增大，与线性模型相比拟合得更好，不过P值显示y与x2的相关性在统计上并不显著。

**13(h)**

> set.seed(1)

> x1 = rnorm(100)

> eps1 = rnorm(100, 0, 0.01)

> y1 = -1 + 0.5\*x1 + eps1

> plot(x1, y1)

> fit13\_ = lm(y1~x1)

> summary(fit13\_)

Call:

lm(formula = y1 ~ x1)

Residuals:

Min 1Q Median 3Q Max

-0.018768 -0.006138 -0.001395 0.005394 0.023462

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.0003769 0.0009699 -1031.5 <2e-16 \*\*\*

x1 0.4999894 0.0010773 464.1 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.009628 on 98 degrees of freedom

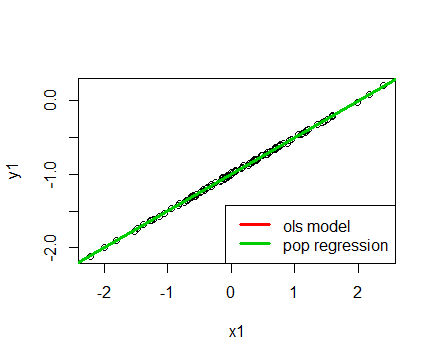
Multiple R-squared: 0.9995, Adjusted R-squared: 0.9995

F-statistic: 2.154e+05 on 1 and 98 DF, p-value: < 2.2e-16

> abline(fit13\_,lwd = 3,col = 2)

> abline(-1,0.5,lwd = 3,col = 3)

> legend("bottomright", legend = c("ols model", "pop regression"), col=2:3, lwd=3)



RSE减小，R2增大。

**13(i)**

> set.seed(1)

> x2 = rnorm(100)

> eps2 = rnorm(100, 0, 1)

> y2 = -1 + 0.5\*x2 + eps2

> plot(x2, y2)

> fit13\_\_ = lm(y2~x2)

> summary(fit13\_\_)

Call:

lm(formula = y2 ~ x2)

Residuals:

Min 1Q Median 3Q Max

-1.8768 -0.6138 -0.1395 0.5394 2.3462

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.03769 0.09699 -10.699 < 2e-16 \*\*\*

x2 0.49894 0.10773 4.632 1.12e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9628 on 98 degrees of freedom

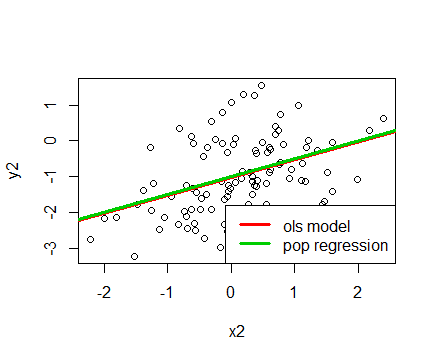
Multiple R-squared: 0.1796, Adjusted R-squared: 0.1712

F-statistic: 21.45 on 1 and 98 DF, p-value: 1.117e-05

> abline(fit13\_\_,lwd = 3,col = 2)

> abline(-1,0.5,lwd = 3,col = 3)

> legend("bottomright", legend = c("ols model", "pop regression"), col=2:3, lwd=3)



RSE减小，R2增大。

**13(j)**

> confint(fit13)

2.5 % 97.5 %

(Intercept) -1.0902064 -0.8824249

x 0.3951885 0.6259784

> confint(fit13\_)

2.5 % 97.5 %

(Intercept) -1.0023016 -0.9984522

x1 0.4978516 0.5021272

> confint(fit13\_\_)

2.5 % 97.5 %

(Intercept) -1.2301607 -0.8452245

x2 0.2851588 0.7127204

less noisy data set的置信区间最小，noisier data set的置信区间最大。区间中点都约为0.5。

**14(a)**

> set.seed(1)

> x1 = runif(100)

> x2 = 0.5 \* x1 + rnorm(100)/10

> y = 2 + 2\*x1 + 0.3\*x2 + rnorm(100)

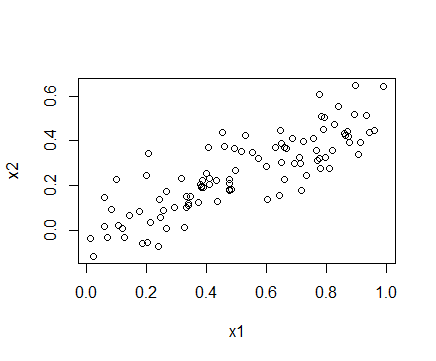
Y=2+2X1+0.3X2+，β0=2, β1=2, β2=0.3

**14(b)**

> cor(x1,x2)

[1] 0.8351212

> plot(x1,x2)



相关系数为0.8351212。

**14(c)**

> fit14<-lm(y~x1+x2)

> summary(fit14)

Call:

lm(formula = y ~ x1 + x2)

Residuals:

Min 1Q Median 3Q Max

-2.8311 -0.7273 -0.0537 0.6338 2.3359

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.1305 0.2319 9.188 7.61e-15 \*\*\*

x1 1.4396 0.7212 1.996 0.0487 \*

x2 1.0097 1.1337 0.891 0.3754

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.056 on 97 degrees of freedom

Multiple R-squared: 0.2088, Adjusted R-squared: 0.1925

F-statistic: 12.8 on 2 and 97 DF, p-value: 1.164e-05

，，。，，。系数的估计值与真实值很接近。根据P值，可以拒绝零假设，但不能拒绝零假设。

**14(d)**

> fit14\_<-lm(y~x1)

> summary(fit14\_)

Call:

lm(formula = y ~ x1)

Residuals:

Min 1Q Median 3Q Max

-2.89495 -0.66874 -0.07785 0.59221 2.45560

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.1124 0.2307 9.155 8.27e-15 \*\*\*

x1 1.9759 0.3963 4.986 2.66e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.055 on 98 degrees of freedom

Multiple R-squared: 0.2024, Adjusted R-squared: 0.1942

F-statistic: 24.86 on 1 and 98 DF, p-value: 2.661e-06

根据F统计量和P值，可以拒绝零假设。

**14(e)**

> fit14\_\_<-lm(y~x2)

> summary(fit14\_\_)

Call:

lm(formula = y ~ x2)

Residuals:

Min 1Q Median 3Q Max

-2.62687 -0.75156 -0.03598 0.72383 2.44890

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.3899 0.1949 12.26 < 2e-16 \*\*\*

x2 2.8996 0.6330 4.58 1.37e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.072 on 98 degrees of freedom

Multiple R-squared: 0.1763, Adjusted R-squared: 0.1679

F-statistic: 20.98 on 1 and 98 DF, p-value: 1.366e-05

根据F统计量和P值，可以拒绝零假设。

**14(f)**

不矛盾，X1与X2存在共线性。当X1和X2共同拟合最小二乘回归来预测Y时，很难分辨X1和X2的影响，当分别拟合最小二乘回归来预测Y时，Y与预测变量的线性关系可以更清楚地看出来。

**14(g)**

> x1 = c(x1, 0.1)

> x2 = c(x2, 0.8)

> y = c(y, 6)

> fit\_14<-lm(y~x1+x2)

> summary(fit\_14)

Call:

lm(formula = y ~ x1 + x2)

Residuals:

Min 1Q Median 3Q Max

-2.73348 -0.69318 -0.05263 0.66385 2.30619

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.2267 0.2314 9.624 7.91e-16 \*\*\*

x1 0.5394 0.5922 0.911 0.36458

x2 2.5146 0.8977 2.801 0.00614 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.075 on 98 degrees of freedom

Multiple R-squared: 0.2188, Adjusted R-squared: 0.2029

F-statistic: 13.72 on 2 and 98 DF, p-value: 5.564e-06

> fit\_14\_<-lm(y~x1)

> summary(fit\_14\_)

Call:

lm(formula = y ~ x1)

Residuals:

Min 1Q Median 3Q Max

-2.8897 -0.6556 -0.0909 0.5682 3.5665

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.2569 0.2390 9.445 1.78e-15 \*\*\*

x1 1.7657 0.4124 4.282 4.29e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.111 on 99 degrees of freedom

Multiple R-squared: 0.1562, Adjusted R-squared: 0.1477

F-statistic: 18.33 on 1 and 99 DF, p-value: 4.295e-05

> fit\_14\_\_<-lm(y~x2)

> summary(fit\_14\_\_)

Call:

lm(formula = y ~ x2)

Residuals:

Min 1Q Median 3Q Max

-2.64729 -0.71021 -0.06899 0.72699 2.38074

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.3451 0.1912 12.264 < 2e-16 \*\*\*

x2 3.1190 0.6040 5.164 1.25e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.074 on 99 degrees of freedom

Multiple R-squared: 0.2122, Adjusted R-squared: 0.2042

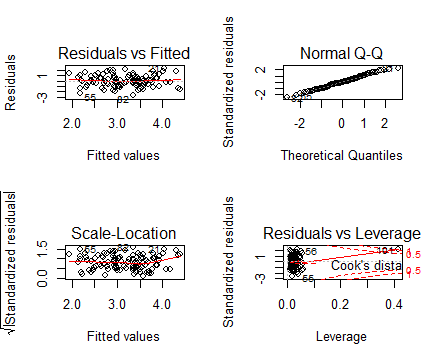
F-statistic: 26.66 on 1 and 99 DF, p-value: 1.253e-06

X1统计上不显著了，而X2统计上显著了。

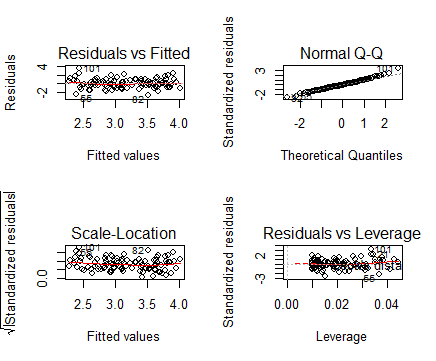
> opar<-par(no.readonly = TRUE)

> par(mfrow=c(2,2))

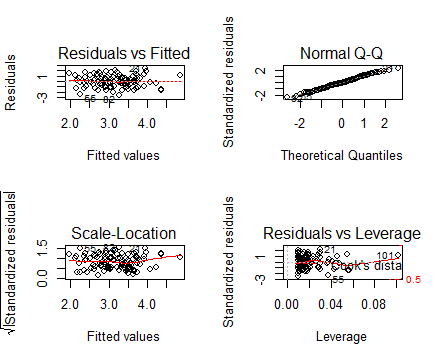
> plot(fit\_14)



> plot(fit\_14\_)

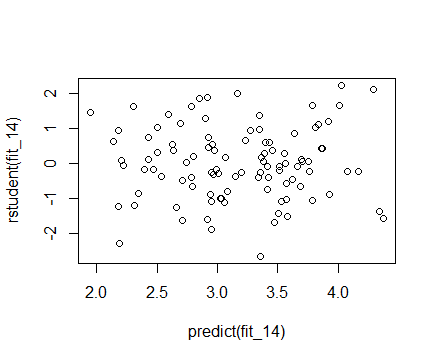


> plot(fit\_14\_\_)

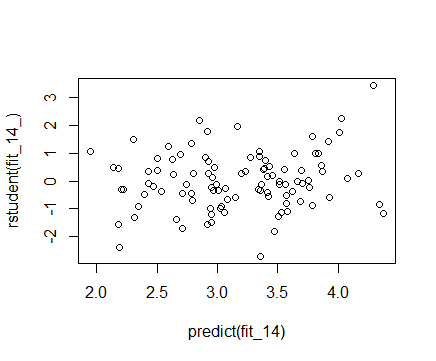


第一个和第三个模型存在高杠杆的值。

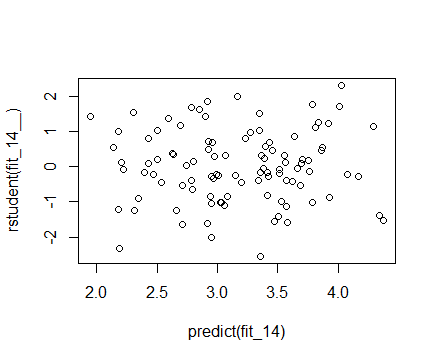
> plot(predict(fit\_14), rstudent(fit\_14))



> plot(predict(fit\_14), rstudent(fit\_14\_))



> plot(predict(fit\_14), rstudent(fit\_14\_\_))

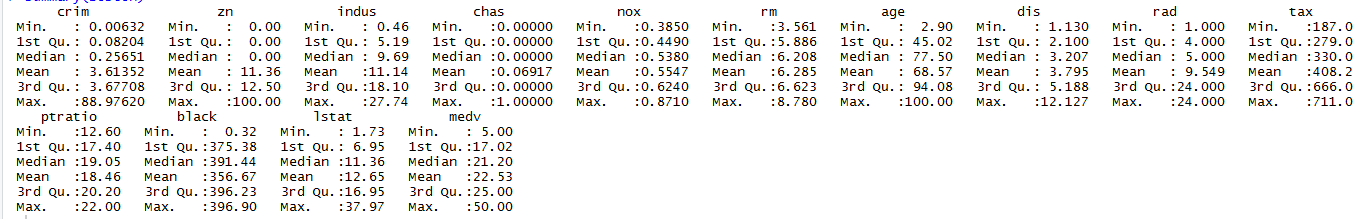


第一个和第三个模型可以认为没有离群点，第二个模型有个别值可能是离群点。

**15(a)**

> library(Boston)

> summary(Boston)



> attach(Boston)

> fit15\_1<-lm(crim~zn)

> summary(fit15\_1)

Call:

lm(formula = crim ~ zn)

Residuals:

Min 1Q Median 3Q Max

-4.429 -4.222 -2.620 1.250 84.523

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.45369 0.41722 10.675 < 2e-16 \*\*\*

zn -0.07393 0.01609 -4.594 5.51e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.435 on 504 degrees of freedom

Multiple R-squared: 0.04019, Adjusted R-squared: 0.03828

F-statistic: 21.1 on 1 and 504 DF, p-value: 5.506e-06

> fit15\_2<-lm(crim~indus)

> summary(fit15\_2)

Call:

lm(formula = crim ~ indus)

Residuals:

Min 1Q Median 3Q Max

-11.972 -2.698 -0.736 0.712 81.813

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -2.06374 0.66723 -3.093 0.00209 \*\*

indus 0.50978 0.05102 9.991 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.866 on 504 degrees of freedom

Multiple R-squared: 0.1653, Adjusted R-squared: 0.1637

F-statistic: 99.82 on 1 and 504 DF, p-value: < 2.2e-16

> fit15\_3<-lm(crim~chas)

> summary(fit15\_3)

Call:

lm(formula = crim ~ chas)

Residuals:

Min 1Q Median 3Q Max

-3.738 -3.661 -3.435 0.018 85.232

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.7444 0.3961 9.453 <2e-16 \*\*\*

chas -1.8928 1.5061 -1.257 0.209

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.597 on 504 degrees of freedom

Multiple R-squared: 0.003124, Adjusted R-squared: 0.001146

F-statistic: 1.579 on 1 and 504 DF, p-value: 0.2094

> fit15\_4<-lm(crim~nox)

> summary(fit15\_4)

Call:

lm(formula = crim ~ nox)

Residuals:

Min 1Q Median 3Q Max

-12.371 -2.738 -0.974 0.559 81.728

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -13.720 1.699 -8.073 5.08e-15 \*\*\*

nox 31.249 2.999 10.419 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.81 on 504 degrees of freedom

Multiple R-squared: 0.1772, Adjusted R-squared: 0.1756

F-statistic: 108.6 on 1 and 504 DF, p-value: < 2.2e-16

> fit15\_5<-lm(crim~rm)

> summary(fit15\_5)

Call:

lm(formula = crim ~ rm)

Residuals:

Min 1Q Median 3Q Max

-6.604 -3.952 -2.654 0.989 87.197

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 20.482 3.365 6.088 2.27e-09 \*\*\*

rm -2.684 0.532 -5.045 6.35e-07 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.401 on 504 degrees of freedom

Multiple R-squared: 0.04807, Adjusted R-squared: 0.04618

F-statistic: 25.45 on 1 and 504 DF, p-value: 6.347e-07

> fit15\_6<-lm(crim~age)

> summary(fit15\_6)

Call:

lm(formula = crim ~ age)

Residuals:

Min 1Q Median 3Q Max

-6.789 -4.257 -1.230 1.527 82.849

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -3.77791 0.94398 -4.002 7.22e-05 \*\*\*

age 0.10779 0.01274 8.463 2.85e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.057 on 504 degrees of freedom

Multiple R-squared: 0.1244, Adjusted R-squared: 0.1227

F-statistic: 71.62 on 1 and 504 DF, p-value: 2.855e-16

> fit15\_7<-lm(crim~dis)

> summary(fit15\_7)

Call:

lm(formula = crim ~ dis)

Residuals:

Min 1Q Median 3Q Max

-6.708 -4.134 -1.527 1.516 81.674

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 9.4993 0.7304 13.006 <2e-16 \*\*\*

dis -1.5509 0.1683 -9.213 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.965 on 504 degrees of freedom

Multiple R-squared: 0.1441, Adjusted R-squared: 0.1425

F-statistic: 84.89 on 1 and 504 DF, p-value: < 2.2e-16

> fit15\_8<-lm(crim~rad)

> summary(fit15\_8)

Call:

lm(formula = crim ~ rad)

Residuals:

Min 1Q Median 3Q Max

-10.164 -1.381 -0.141 0.660 76.433

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -2.28716 0.44348 -5.157 3.61e-07 \*\*\*

rad 0.61791 0.03433 17.998 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.718 on 504 degrees of freedom

Multiple R-squared: 0.3913, Adjusted R-squared: 0.39

F-statistic: 323.9 on 1 and 504 DF, p-value: < 2.2e-16

> fit15\_9<-lm(crim~tax)

> summary(fit15\_9)

Call:

lm(formula = crim ~ tax)

Residuals:

Min 1Q Median 3Q Max

-12.513 -2.738 -0.194 1.065 77.696

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -8.528369 0.815809 -10.45 <2e-16 \*\*\*

tax 0.029742 0.001847 16.10 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.997 on 504 degrees of freedom

Multiple R-squared: 0.3396, Adjusted R-squared: 0.3383

F-statistic: 259.2 on 1 and 504 DF, p-value: < 2.2e-16

> fit15\_10<-lm(crim~ptratio)

> summary(fit15\_10)

Call:

lm(formula = crim ~ ptratio)

Residuals:

Min 1Q Median 3Q Max

-7.654 -3.985 -1.912 1.825 83.353

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -17.6469 3.1473 -5.607 3.40e-08 \*\*\*

ptratio 1.1520 0.1694 6.801 2.94e-11 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.24 on 504 degrees of freedom

Multiple R-squared: 0.08407, Adjusted R-squared: 0.08225

F-statistic: 46.26 on 1 and 504 DF, p-value: 2.943e-11

> fit15\_11<-lm(crim~black)

> summary(fit15\_11)

Call:

lm(formula = crim ~ black)

Residuals:

Min 1Q Median 3Q Max

-13.756 -2.299 -2.095 -1.296 86.822

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 16.553529 1.425903 11.609 <2e-16 \*\*\*

black -0.036280 0.003873 -9.367 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.946 on 504 degrees of freedom

Multiple R-squared: 0.1483, Adjusted R-squared: 0.1466

F-statistic: 87.74 on 1 and 504 DF, p-value: < 2.2e-16

> fit15\_12<-lm(crim~lstat)

> summary(fit15\_12)

Call:

lm(formula = crim ~ lstat)

Residuals:

Min 1Q Median 3Q Max

-13.925 -2.822 -0.664 1.079 82.862

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -3.33054 0.69376 -4.801 2.09e-06 \*\*\*

lstat 0.54880 0.04776 11.491 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.664 on 504 degrees of freedom

Multiple R-squared: 0.2076, Adjusted R-squared: 0.206

F-statistic: 132 on 1 and 504 DF, p-value: < 2.2e-16

> fit15\_13<-lm(crim~medv)

> summary(fit15\_13)

Call:

lm(formula = crim ~ medv)

Residuals:

Min 1Q Median 3Q Max

-9.071 -4.022 -2.343 1.298 80.957

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 11.79654 0.93419 12.63 <2e-16 \*\*\*

medv -0.36316 0.03839 -9.46 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.934 on 504 degrees of freedom

Multiple R-squared: 0.1508, Adjusted R-squared: 0.1491

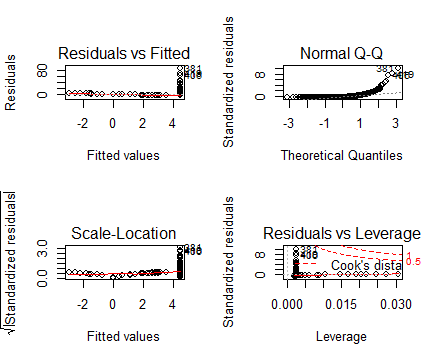
F-statistic: 89.49 on 1 and 504 DF, p-value: < 2.2e-16

除了chas统计上都是显著的。

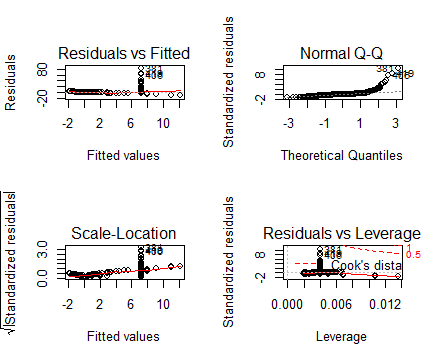
> opar<-par(no.readonly = TRUE)

> par(mfrow=c(2,2))

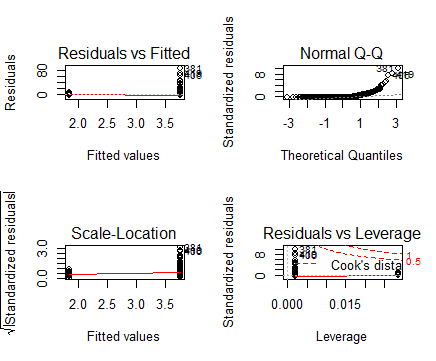
> plot(fit15\_1)



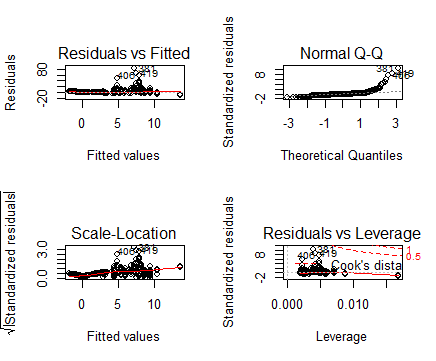
> plot(fit15\_2)



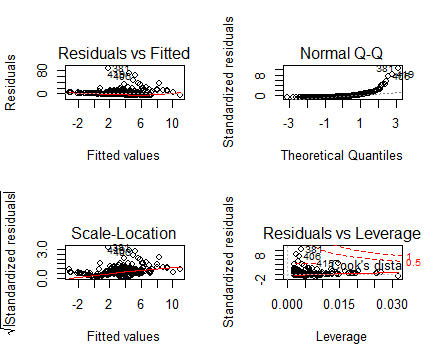
> plot(fit15\_3)



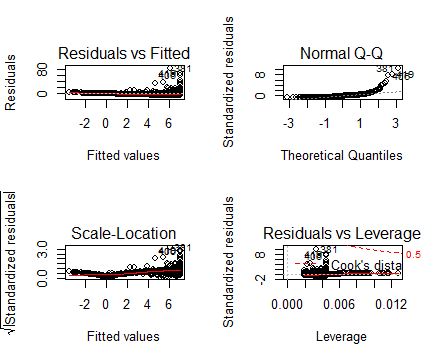
> plot(fit15\_4)



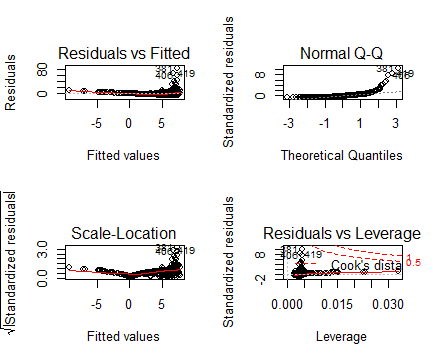
> plot(fit15\_5)



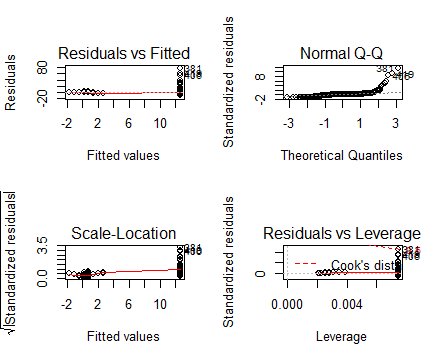
> plot(fit15\_6)



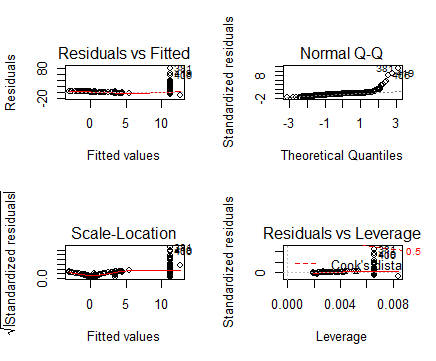
> plot(fit15\_7)



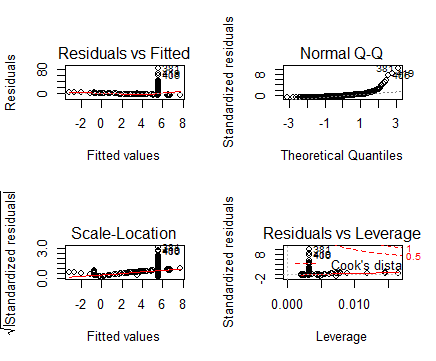
> plot(fit15\_8)



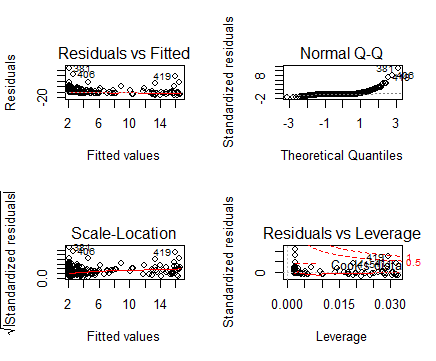
> plot(fit15\_9)



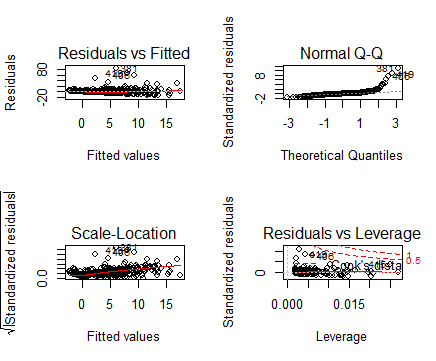
> plot(fit15\_10)



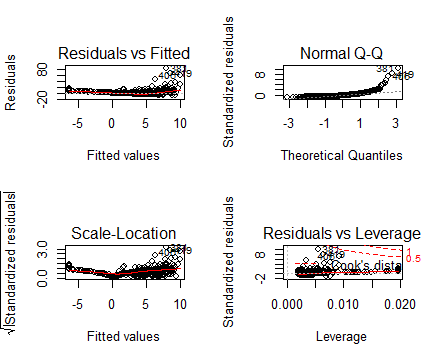
> plot(fit15\_11)



> plot(fit15\_12)



> plot(fit15\_13)



可以看到，除了chas的残差图拟合值全部集中在1.5和3.5左右，其余预测变量的残差图都没有明显规律。

**15(b)**

> fit15<-lm(crim~.,data = Boston)

> summary(fit15)

Call:

lm(formula = crim ~ ., data = Boston)

Residuals:

Min 1Q Median 3Q Max

-9.924 -2.120 -0.353 1.019 75.051

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 17.033228 7.234903 2.354 0.018949 \*

zn 0.044855 0.018734 2.394 0.017025 \*

indus -0.063855 0.083407 -0.766 0.444294

chas -0.749134 1.180147 -0.635 0.525867

nox -10.313535 5.275536 -1.955 0.051152 .

rm 0.430131 0.612830 0.702 0.483089

age 0.001452 0.017925 0.081 0.935488

dis -0.987176 0.281817 -3.503 0.000502 \*\*\*

rad 0.588209 0.088049 6.680 6.46e-11 \*\*\*

tax -0.003780 0.005156 -0.733 0.463793

ptratio -0.271081 0.186450 -1.454 0.146611

black -0.007538 0.003673 -2.052 0.040702 \*

lstat 0.126211 0.075725 1.667 0.096208 .

medv -0.198887 0.060516 -3.287 0.001087 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.439 on 492 degrees of freedom

Multiple R-squared: 0.454, Adjusted R-squared: 0.4396

F-statistic: 31.47 on 13 and 492 DF, p-value: < 2.2e-16

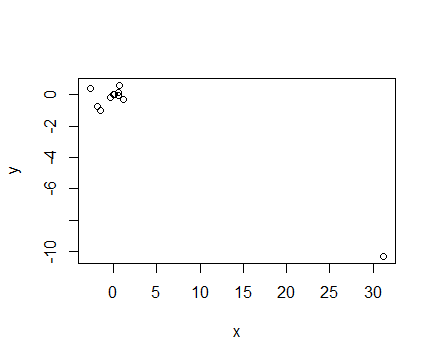
zn, dis, rad, black, medv，这几个预测变量可以拒绝零假设*H0*：。

**15(c)**

> x<-c(coefficients(fit15\_1)[2],coefficients(fit15\_2)[2],coefficients(fit15\_3)[2],coefficients(fit15\_4)[2],coefficients(fit15\_5)[2],coefficients(fit15\_6)[2],coefficients(fit15\_7)[2],coefficients(fit15\_8)[2],coefficients(fit15\_9)[2],coefficients(fit15\_10)[2],coefficients(fit15\_11)[2],coefficients(fit15\_12)[2],coefficients(fit15\_13)[2])

> y<-c(coefficients(fit15)[2:14])

> plot(x,y)



对于预测变量nox来说，单变量回归模型中，其参数估计值为31.249，在多元回归中其参数估计值为-10.313535。

**15(d)**

> fit15\_1\_<-lm(crim~poly(zn,3))

> summary(fit15\_1\_)

Call:

lm(formula = crim ~ poly(zn, 3))

Residuals:

Min 1Q Median 3Q Max

-4.821 -4.614 -1.294 0.473 84.130

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.6135 0.3722 9.709 < 2e-16 \*\*\*

poly(zn, 3)1 -38.7498 8.3722 -4.628 4.7e-06 \*\*\*

poly(zn, 3)2 23.9398 8.3722 2.859 0.00442 \*\*

poly(zn, 3)3 -10.0719 8.3722 -1.203 0.22954

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.372 on 502 degrees of freedom

Multiple R-squared: 0.05824, Adjusted R-squared: 0.05261

F-statistic: 10.35 on 3 and 502 DF, p-value: 1.281e-06

> fit15\_2\_<-lm(crim~poly(indus,3))

> summary(fit15\_2\_)

Call:

lm(formula = crim ~ poly(indus, 3))

Residuals:

Min 1Q Median 3Q Max

-8.278 -2.514 0.054 0.764 79.713

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.614 0.330 10.950 < 2e-16 \*\*\*

poly(indus, 3)1 78.591 7.423 10.587 < 2e-16 \*\*\*

poly(indus, 3)2 -24.395 7.423 -3.286 0.00109 \*\*

poly(indus, 3)3 -54.130 7.423 -7.292 1.2e-12 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.423 on 502 degrees of freedom

Multiple R-squared: 0.2597, Adjusted R-squared: 0.2552

F-statistic: 58.69 on 3 and 502 DF, p-value: < 2.2e-16

> fit15\_4\_<-lm(crim~poly(nox,3))

> summary(fit15\_4\_)

Call:

lm(formula = crim ~ poly(nox, 3))

Residuals:

Min 1Q Median 3Q Max

-9.110 -2.068 -0.255 0.739 78.302

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.6135 0.3216 11.237 < 2e-16 \*\*\*

poly(nox, 3)1 81.3720 7.2336 11.249 < 2e-16 \*\*\*

poly(nox, 3)2 -28.8286 7.2336 -3.985 7.74e-05 \*\*\*

poly(nox, 3)3 -60.3619 7.2336 -8.345 6.96e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.234 on 502 degrees of freedom

Multiple R-squared: 0.297, Adjusted R-squared: 0.2928

F-statistic: 70.69 on 3 and 502 DF, p-value: < 2.2e-16

> fit15\_5\_<-lm(crim~poly(rm,3))

> summary(fit15\_5\_)

Call:

lm(formula = crim ~ poly(rm, 3))

Residuals:

Min 1Q Median 3Q Max

-18.485 -3.468 -2.221 -0.015 87.219

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.6135 0.3703 9.758 < 2e-16 \*\*\*

poly(rm, 3)1 -42.3794 8.3297 -5.088 5.13e-07 \*\*\*

poly(rm, 3)2 26.5768 8.3297 3.191 0.00151 \*\*

poly(rm, 3)3 -5.5103 8.3297 -0.662 0.50858

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.33 on 502 degrees of freedom

Multiple R-squared: 0.06779, Adjusted R-squared: 0.06222

F-statistic: 12.17 on 3 and 502 DF, p-value: 1.067e-07

> fit15\_6\_<-lm(crim~poly(age,3))

> summary(fit15\_6\_)

Call:

lm(formula = crim ~ poly(age, 3))

Residuals:

Min 1Q Median 3Q Max

-9.762 -2.673 -0.516 0.019 82.842

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.6135 0.3485 10.368 < 2e-16 \*\*\*

poly(age, 3)1 68.1820 7.8397 8.697 < 2e-16 \*\*\*

poly(age, 3)2 37.4845 7.8397 4.781 2.29e-06 \*\*\*

poly(age, 3)3 21.3532 7.8397 2.724 0.00668 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.84 on 502 degrees of freedom

Multiple R-squared: 0.1742, Adjusted R-squared: 0.1693

F-statistic: 35.31 on 3 and 502 DF, p-value: < 2.2e-16

> fit15\_7\_<-lm(crim~poly(dis,3))

> summary(fit15\_7\_)

Call:

lm(formula = crim ~ poly(dis, 3))

Residuals:

Min 1Q Median 3Q Max

-10.757 -2.588 0.031 1.267 76.378

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.6135 0.3259 11.087 < 2e-16 \*\*\*

poly(dis, 3)1 -73.3886 7.3315 -10.010 < 2e-16 \*\*\*

poly(dis, 3)2 56.3730 7.3315 7.689 7.87e-14 \*\*\*

poly(dis, 3)3 -42.6219 7.3315 -5.814 1.09e-08 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.331 on 502 degrees of freedom

Multiple R-squared: 0.2778, Adjusted R-squared: 0.2735

F-statistic: 64.37 on 3 and 502 DF, p-value: < 2.2e-16

> fit15\_8\_<-lm(crim~poly(rad,3))

> summary(fit15\_8\_)

Call:

lm(formula = crim ~ poly(rad, 3))

Residuals:

Min 1Q Median 3Q Max

-10.381 -0.412 -0.269 0.179 76.217

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.6135 0.2971 12.164 < 2e-16 \*\*\*

poly(rad, 3)1 120.9074 6.6824 18.093 < 2e-16 \*\*\*

poly(rad, 3)2 17.4923 6.6824 2.618 0.00912 \*\*

poly(rad, 3)3 4.6985 6.6824 0.703 0.48231

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.682 on 502 degrees of freedom

Multiple R-squared: 0.4, Adjusted R-squared: 0.3965

F-statistic: 111.6 on 3 and 502 DF, p-value: < 2.2e-16

> fit15\_9\_<-lm(crim~poly(tax,3))

> summary(fit15\_9\_)

Call:

lm(formula = crim ~ poly(tax, 3))

Residuals:

Min 1Q Median 3Q Max

-13.273 -1.389 0.046 0.536 76.950

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.6135 0.3047 11.860 < 2e-16 \*\*\*

poly(tax, 3)1 112.6458 6.8537 16.436 < 2e-16 \*\*\*

poly(tax, 3)2 32.0873 6.8537 4.682 3.67e-06 \*\*\*

poly(tax, 3)3 -7.9968 6.8537 -1.167 0.244

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.854 on 502 degrees of freedom

Multiple R-squared: 0.3689, Adjusted R-squared: 0.3651

F-statistic: 97.8 on 3 and 502 DF, p-value: < 2.2e-16

> fit15\_10\_<-lm(crim~poly(ptratio,3))

> summary(fit15\_10\_)

Call:

lm(formula = crim ~ poly(ptratio, 3))

Residuals:

Min 1Q Median 3Q Max

-6.833 -4.146 -1.655 1.408 82.697

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.614 0.361 10.008 < 2e-16 \*\*\*

poly(ptratio, 3)1 56.045 8.122 6.901 1.57e-11 \*\*\*

poly(ptratio, 3)2 24.775 8.122 3.050 0.00241 \*\*

poly(ptratio, 3)3 -22.280 8.122 -2.743 0.00630 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.122 on 502 degrees of freedom

Multiple R-squared: 0.1138, Adjusted R-squared: 0.1085

F-statistic: 21.48 on 3 and 502 DF, p-value: 4.171e-13

> fit15\_11\_<-lm(crim~poly(black,3))

> summary(fit15\_11\_)

Call:

lm(formula = crim ~ poly(black, 3))

Residuals:

Min 1Q Median 3Q Max

-13.096 -2.343 -2.128 -1.439 86.790

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.6135 0.3536 10.218 <2e-16 \*\*\*

poly(black, 3)1 -74.4312 7.9546 -9.357 <2e-16 \*\*\*

poly(black, 3)2 5.9264 7.9546 0.745 0.457

poly(black, 3)3 -4.8346 7.9546 -0.608 0.544

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.955 on 502 degrees of freedom

Multiple R-squared: 0.1498, Adjusted R-squared: 0.1448

F-statistic: 29.49 on 3 and 502 DF, p-value: < 2.2e-16

> fit15\_12\_<-lm(crim~poly(lstat,3))

> summary(fit15\_12\_)

Call:

lm(formula = crim ~ poly(lstat, 3))

Residuals:

Min 1Q Median 3Q Max

-15.234 -2.151 -0.486 0.066 83.353

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.6135 0.3392 10.654 <2e-16 \*\*\*

poly(lstat, 3)1 88.0697 7.6294 11.543 <2e-16 \*\*\*

poly(lstat, 3)2 15.8882 7.6294 2.082 0.0378 \*

poly(lstat, 3)3 -11.5740 7.6294 -1.517 0.1299

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.629 on 502 degrees of freedom

Multiple R-squared: 0.2179, Adjusted R-squared: 0.2133

F-statistic: 46.63 on 3 and 502 DF, p-value: < 2.2e-16

> fit15\_13\_<-lm(crim~poly(medv,3))

> summary(fit15\_13\_)

Call:

lm(formula = crim ~ poly(medv, 3))

Residuals:

Min 1Q Median 3Q Max

-24.427 -1.976 -0.437 0.439 73.655

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.614 0.292 12.374 < 2e-16 \*\*\*

poly(medv, 3)1 -75.058 6.569 -11.426 < 2e-16 \*\*\*

poly(medv, 3)2 88.086 6.569 13.409 < 2e-16 \*\*\*

poly(medv, 3)3 -48.033 6.569 -7.312 1.05e-12 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.569 on 502 degrees of freedom

Multiple R-squared: 0.4202, Adjusted R-squared: 0.4167

F-statistic: 121.3 on 3 and 502 DF, p-value: < 2.2e-16

除了chas和black，其余预测变量二次或三次项系数统计上都是显著的，可以拒绝零假设。