1. **回归分析与相关分析**

9.1相关分析

男生身高与体重的相关性分析

> x<-c(185, 173, 175, 182, 173, 181, 184, 179, 181, 187, 169, 178,

+ 183, 168, 181, 175, 175, 186, 186, 182, 178, 177, 172, 168,

+ 173.5, 184, 183, 175, 168, 174, 181, 170, 166, 178, 177, 181,

+ 163, 172, 160, 173, 185, 172, 183, 180, 175, 178, 169, 175,

+ 165, 169, 170, 183, 184, 174, 170, 173, 170, 182, 178, 170,

+ 179)

> y<-c(65, 62, 80.3, 74.3, 55.7, 60, 59, 79, 62, 80, 56, 60.5,

+ 73, 46, 65, 91, 64, 88, 63, 64, 65, 75.6, 64, 65.5,

+ 58.8, 59, 71, 75, 60, 61, 75, 58, 56, 94.5, 87, 71,

+ 47.5, 59, 57, 65, 67, 60, 85, 65, 73, 70, 55, 75,

+ 55, 65, 65, 72, 99, 75, 53, 70, 58, 63, 92, 48,

+ 69)

> level <- data.frame(x,y)

> plot(level)

> attach(level)

The following objects are masked \_by\_ .GlobalEnv:

x, y

The following objects are masked from level (pos = 3):

x, y

> cor.test(x, y)

Pearson's product-moment correlation

data: x and y

t = 4.5635, df = 59, p-value = 2.601e-05

alternative hypothesis: true correlation is not equal to 0

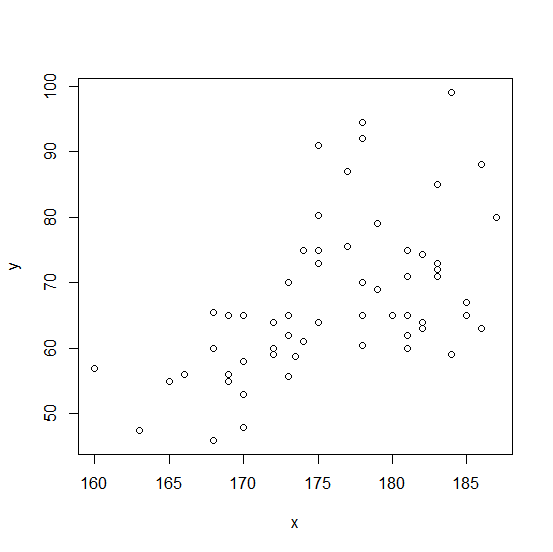
95 percent confidence interval:

0.2971703 0.6756816

sample estimates:

cor

0.5107684



结论: 因为p值＝2.601e-05≤0.05, 故拒绝原假设, 从而认为变量x与y相关.



在R中, 由函数lm( )可以非常方便地求出回归方程, 函数confint( )可求

出参数的置信区间.与回归分析有关的函数还有summary( )

9.2一元线性回归分析

相关分析只能得出两个变量之间是否相关, 但却不能回答在两个变量之间存在相关关系时, 它们之间是如何联系的, 即无法找出刻画它们之间因果关系的函数关系. 回归分析就可以解决这一问题。

9.2.1估计与检验

9.2.1.1β0，β1的估计（最小二乘法）

9.2.1.2回归方程的显著性检验（t、F、相关系数检验法）（对β1）

9.2.1.3β0，β1的区间估计

计算分析过程如下：

1) 建立数据集, 并画出散点图: 考查数据点的分布趋势, 看是否呈直线条状

分布. 程序如下

x<-c(185, 173, 175, 182, 173, 181, 184, 179, 181, 187, 169, 178,

183, 168, 181, 175, 175, 186, 186, 182, 178, 177, 172, 168,

173.5, 184, 183, 175, 168, 174, 181, 170, 166, 178, 177, 181,

163, 172, 160, 173, 185, 172, 183, 180, 175, 178, 169, 175,

165, 169, 170, 183, 184, 174, 170, 173, 170, 182, 178, 170,

179)

y<-c(65, 62, 80.3, 74.3, 55.7, 60, 59, 79, 62, 80, 56, 60.5,

73, 46, 65, 91, 64, 88, 63, 64, 65, 75.6, 64, 65.5,

58.8, 59, 71, 75, 60, 61, 75, 58, 56, 94.5, 87, 71,

47.5, 59, 57, 65, 67, 60, 85, 65, 73, 70, 55, 75,

55, 65, 65, 72, 99, 75, 53, 70, 58, 63, 92, 48,

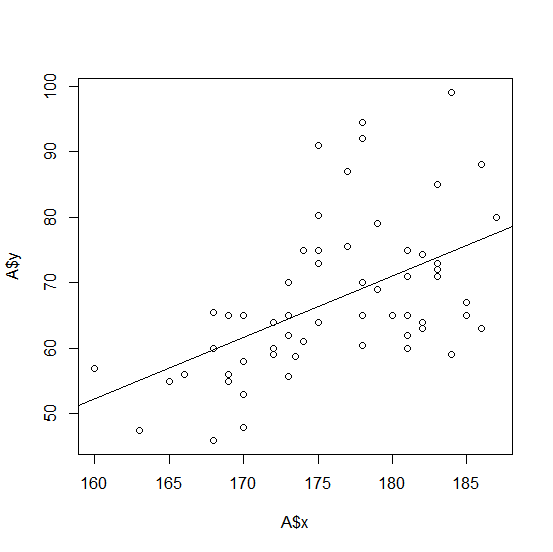
69)

A<-data.frame(x, y)

plot(A$x, A$y)

运行结果如图9.4所示（顺便画出了直线）, 可以看出, 这些点基本上(但不精确的)落在一条直

线上.



1. 进行回归分析, 并在散点图上显示回归直线. R程序为

lm.reg<-lm(y~1+x)

Call:

lm(formula = y ~ 1 + x)

Residuals:

Min 1Q Median 3Q Max

-15.710 -6.914 -1.935 5.542 25.382

Coefficients:

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

(Intercept) -96.7630 35.9796 -2.689 0.00929 \*\*

x 0.9319 0.2042 4.563 2.6e-05 \*\*\*

---

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

Residual standard error: 10.06 on 59 degrees of freedom

Multiple R-squared: 0.2609, Adjusted R-squared: 0.2484

F-statistic: 20.83 on 1 and 59 DF, p-value: 2.601e-05

confint(lm.reg, level=0.95)

2.5 % 97.5 %

(Intercept) -168.7579382 -24.767964

x 0.5232869 1.340544

程序中, 第三行函数lm( )表示使用线性回归模型, 第四行函

数summary( )为提取模型计算结果. 第五行为求95置信区间。

结论: 从上述输出结果p-值可以看出回归方程通过了回归参数的检验与回归方程

的检验, 由此得到回归方程Y=-96.7630+0.0.9319X

写出置信区间β0： -168.7579382 -24.767964。β1：0.5232869 1.340544

3）得到了回归方程, 还可以对误差项独立同正态分布的假设进行检验.

在R中只需再执行一个plot命令.

op<-par(mfrow=c(2, 2))

plot(lm.reg)

par(op)

运行结果见图9.3. 上面的命令plot(lm.reg)实际上使用了四次plot(x, y),

产生四个图形, 它们分别为：

1) Residual vs fitted为拟合值y帽子对残差的图形, 可以看出, 数据点都基本均匀

地分布在直线y=0的两侧, 无明显趋势;

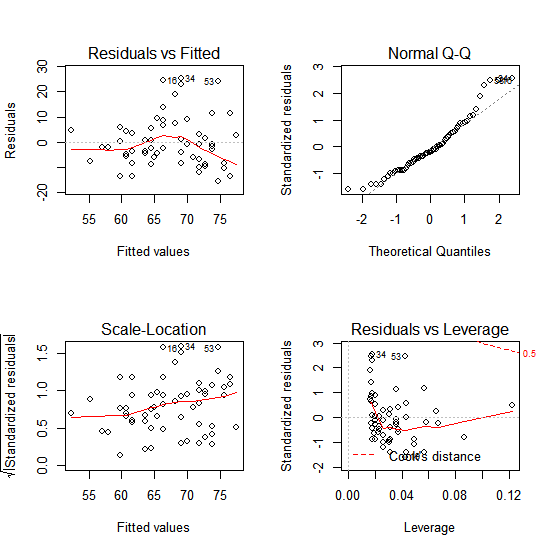
2) Normal QQ-plot图中数据点分布趋于一条直线, 说明残差是服从正态分

布的;

3) Scale－Location 图显示了标准化残差(standardized residuals)的平方根

的分布情况. 最高点为残差最大值点;

4) Cook距离(Cook's distance)图显示了对回归的影响点.



 回归系数的估计与检验: 回归系数的估计为β0帽子=-96.7630 ; β1帽子=0.9319; 相应的标准差为Sd（β0帽子）=35.9796; Sd（β1帽子）=0.2042. 它们

的p值均很小, 故是非常显著的

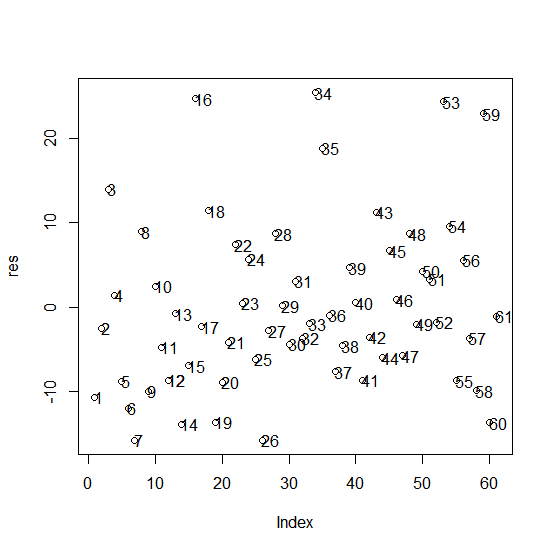
 相关分析: 相关系数的平方R2=0:9901, 表明数据中99%可由回归

方程来描述.

 方程的检验: F分布的p值为2:079×10^-14, 因此方程是非常显著的,

这与R^2的结果一致.

.



9.3多元线性回归分析

9.3.1估计与检验

9.3.1.1回归系数显著性检验

y<-c(185, 173, 175, 182, 173, 181, 184, 179, 181, 187, 169, 178,

183, 168, 181, 175, 175, 186, 186, 182, 178, 177, 172, 168,

173.5, 184, 183, 175, 168, 174, 181, 170, 166, 178, 177, 181,

163, 172, 160, 173, 185, 172, 183, 180, 175, 178, 169, 175,

165, 169, 170, 183, 184, 174, 170, 173, 170, 182, 178, 170,

179)

x1<-c(65, 62, 80.3, 74.3, 55.7, 60, 59, 79, 62, 80, 56, 60.5,

73, 46, 65, 91, 64, 88, 63, 64, 65, 75.6, 64, 65.5,

58.8, 59, 71, 75, 60, 61, 75, 58, 56, 94.5, 87, 71,

47.5, 59, 57, 65, 67, 60, 85, 65, 73, 70, 55, 75,

55, 65, 65, 72, 99, 75, 53, 70, 58, 63, 92, 48,

69)

x2<-c(188, 182, 183, 189, 185, 170, 179, 165, 171, 196, 168, 177,

179, 167, 175, 171, 170, 183, 180, 182, 180, 177, 176, 170,

167, 179, 186, 168, 160, 171.5, 181, 172, 146, 177, 168, 183,

162, 170.5, 166.7, 173, 176, 167, 188, 178, 178, 175, 169, 171,

164, 175, 161, 174, 188, 171, 166.6, 171, 169, 160, 175, 164,

169)

sales<-data.frame(y, x1, x2)

lm.reg<-lm(y~x1+x2, data=sales)

summary(lm.reg)

结果：

Call:

lm(formula = y ~ x1 + x2, data = sales)

Residuals:

Min 1Q Median 3Q Max

-11.696 -2.774 -0.249 2.855 12.162

Coefficients:

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

(Intercept) 95.79920 11.95024 8.017 5.75e-11 \*\*\*

x1 0.14382 0.05657 2.543 0.0137 \*

x2 0.40611 0.07563 5.369 1.45e-06 \*\*\*

---

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

Residual standard error: 4.547 on 58 degrees of freedom

Multiple R-squared: 0.5063, Adjusted R-squared: 0.4893

F-statistic: 29.74 on 2 and 58 DF, p-value: 1.29e-09

结论：由于用于回归方程检验的F统计量的p值与用于回归系数检验的t统计量

的p值均很小(<0.05), 因此回归方程与回归系数的检验都是显著的. 回归方程

为

9.3.3四个数据全做自变量（删除不显著的系数）

1) 建立数据集:

> male\_height<-c(185, 173, 175, 182, 173, 181, 184, 179, 181, 187, 169, 178,

183, 168, 181, 175, 175, 186, 186, 182, 178, 177, 172, 168,

173.5, 184, 183, 175, 168, 174, 181, 170, 166, 178, 177, 181,

163, 172, **160**, 173, 185, 172, 183, 180, 175, 178, 169, 175,

165, 169, 170, 183, 184, 174, 170, 173, 170, **182**, 178, 170,

179)

male\_weight<-c(65, 62, 80.3, 74.3, 55.7, 60, 59, 79, 62, 80, 56, 60.5,

73, 46, 65, 91, 64, 88, 63, 64, 65, 75.6, 64, 65.5,

58.8, 59, 71, 75, 60, 61, 75, 58, 56, 94.5, 87, 71,

47.5, 59, **57**, 65, 67, 60, 85, 65, 73, 70, 55, 75,

55, 65, 65, 72, 99, 75, 53, 70, 58, **63**, 92, 48,

69)

male\_armspan<-c(188, 182, 183, 189, 185, 170, 179, 165, 171, 196, 168, 177,

179, 167, 175, 171, 170, 183, 180, 182, 180, 177, 176, 170,

167, 179, 186, 168, 160, 171.5, 181, 172, 146, 177, 168, 183,

162, 170.5, **166.7**, 173, 176, 167, 188, 178, 178, 175, 169, 171,

164, 175, 161, 174, 188, 171, 166.6, 171, 169, **160**, 175, 164,

169)

male\_leglength<-c(102, 93, 107, 114, 107, 101, 98, 99, 98, 119, 97, 101,

102, 99, 103, 97, 82, 101, 95, 97, 91, 95.5, 95, 82,

83, 98, 98, 105, 93, 96.5, 104.5, 85, 77, 105, 94, 102,

77, 95.8, **83.5**, 89, 106, 98, 102, 93, 98, 90, 86, 98,

99, 99, 92, 100, 108, 102, 83, 101, 102, **80**, 85, 92,

98)

male\_footlength<-c(25.1, 23.5, 26.5, 26, 26, 25, 26.5, 26, 26, 29, 26, 26.5,

27, 24.5, 25.5, 22, 24, 28, 26.5, 26.5, 26.5, 27, 25, 25.5,

24, 26.5, 26, 27, 22, 23.4, 25.5, 25.4, 22, 26.5, 22, 26,

23.8, 23, **22.8**, 25.5, 25.8, 24.5, 26.5, 24, 26.5, 29, 23, 26,

22, 23, 22, 27.5, 26.5, 26, 20, 25.5, 24.5, **24**, 25.5, 23,

25)

2) 建立多元线性回归方程:

male\_weight<-data.frame(male\_height, male\_weight, male\_armspan,

male\_leglength, male\_footlength)

lm.reg<-lm(male\_weight~male\_height+male\_armspan

+male\_leglength+male\_footlength, data=blood)

summary(lm.reg)

Call:

lm(formula = male\_height ~ male\_weight + male\_armspan + male\_leglength +

male\_footlength, data = blood)

Residuals:

Min 1Q Median 3Q Max

-10.1002 -3.2960 -0.3567 2.7437 11.8329

Coefficients:

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

(Intercept) 93.75114 11.51422 8.142 4.51e-11 \*\*\*

male\_weight 0.11318 0.05566 2.034 0.0468 \*

male\_armspan 0.25330 0.09718 2.606 0.0117 \*

male\_leglength 0.04707 0.08875 0.530 0.5980

male\_footlength 1.04138 0.40275 2.586 0.0123 \*

---

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

Residual standard error: 4.345 on 56 degrees of freedom

Multiple R-squared: 0.5648, Adjusted R-squared: 0.5337

F-statistic: 18.17 on 4 and 56 DF, p-value: 1.288e-09

结论: 回归方程的系数的显著性不高, 有的甚至没有通过检验(X1与X2),

这说明如果选择全部变量构造方程, 效果并不好. 这就涉及到变量选择的

问题, 以建立“最优”的回归方程.

3) 变量选择与最优回归: R软件提供了获得“最优”回归方程的方法,“逐步

回归法”的计算函数step( ), 它是以Akaike信息统计量为准则(简称AIC准

则), 通过选择最小的AIC信息统计量, 来达到删除或增加变量的目的. 用函数step( )作逐步回归：

> lm.step<-step(lm.reg)

Start: AIC=184

male\_height ~ male\_weight + male\_armspan + male\_leglength + male\_footlength

Df Sum of Sq RSS AIC

- male\_leglength 1 5.309 1062.5 182.31

<none> 1057.2 184.00

- male\_weight 1 78.063 1135.2 186.35

- male\_footlength 1 126.215 1183.4 188.88

- male\_armspan 1 128.249 1185.4 188.99

Step: AIC=182.31

male\_height ~ male\_weight + male\_armspan + male\_footlength

Df Sum of Sq RSS AIC

<none> 1062.5 182.31

- male\_weight 1 87.291 1149.8 185.12

- male\_footlength 1 136.692 1199.2 187.69

- male\_armspan 1 191.091 1253.6 190.40

> summary(lm.step)

Call:

lm(formula = male\_height ~ male\_weight + male\_armspan + male\_footlength,

data = blood)

Residuals:

Min 1Q Median 3Q Max

-10.4162 -2.9971 -0.0533 2.7918 11.4406

Coefficients:

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

(Intercept) 93.16423 11.38842 8.181 3.45e-11 \*\*\*

male\_weight 0.11804 0.05455 2.164 0.03467 \*

male\_armspan 0.27639 0.08632 3.202 0.00223 \*\*

male\_footlength 1.07232 0.39598 2.708 0.00892 \*\*

---

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

Residual standard error: 4.317 on 57 degrees of freedom

Multiple R-squared: 0.5626, Adjusted R-squared: 0.5396

F-statistic: 24.44 on 3 and 57 DF, p-value: 2.704e-10

结论: 用全部变量作回归方程时, AIC统计量的值为42.16, 如果去掉

变量X1, AIC统计量的值为40.34; 如果去掉变量X2, AIC统计量的值

为43.568, 依次类推. 由于去掉X1使AIC统计量达到最小, 因此R软件

会自动去掉变量X1, 进入下一轮计算. 在下一轮中, 无论去掉哪一个变量,

AIC统计量的值均会升高, 因此R软件自动终止计算, 得到“最优”回归方

程.（数据没改）

回归系数的显著性水平有很大提高, 所有的检验均是显著的, 由此

得到“最优”的回归方程

Y \_x0010\_ 6:4996 􀀀 0:4023X2 0:2870X3 􀀀 0:6632X4:（数据没改）

9.4回归诊断

前面介绍了如何得到回归模型, 但没有对回归模型的一些特性作进一步的

研究, 并且没有研究对回归模型产生较大影响的异常值问题. 异常值的存在往

往会给回归模型带来不稳定, 为此, 人们提出了所谓回归诊断的问题, 其主要

内容有：残差分析、异常点识别、影响分析、共线性诊断等.

9.4.1残差分析

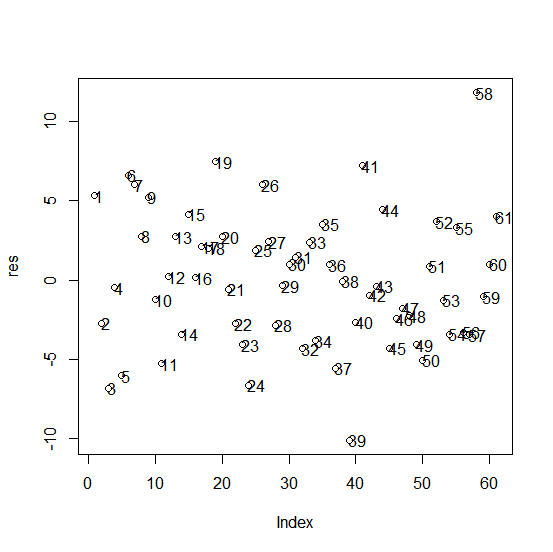
计算例9.3.1的残差和标准残差, 并画出相应的残差散点图

res<-residuals(lm.reg)

plot(res)

for(i in 1:61)

{text(i, res[i], labels=i, adj=(.05))}

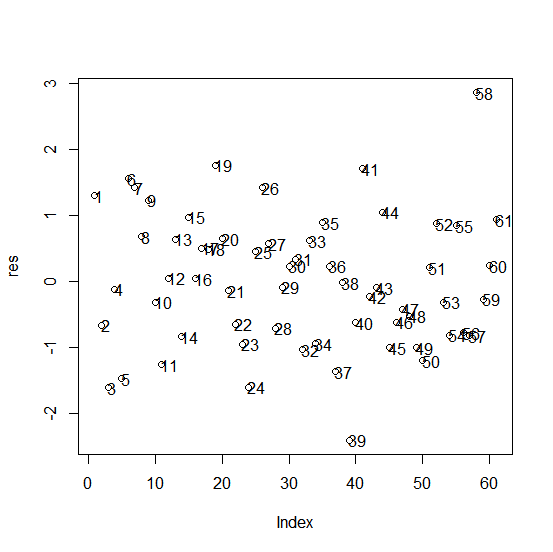


res<-rstandard(lm.reg)

plot(res)

for(i in 1:61)

{text(i, res[i], labels=i, adj=(.05))}



如果多元线性回归模型的假定成立, 从理论上可证明r1; r2;…… ; rn相互独立且近似

服从N(0,1), 故关于观测值等的残差图中散点应随机的分布在-2到+2的带子

里, 并称之为正常残差图, 否则称为异常残差图.这张图很正常。

9.4.2异常点的识别

如果拟合后的模型能够很好地描述这组数据, 那么残差对预测值的散点图

应该像一些随机散布的点. 可是, 若某个观测不能和其它数据一起用这个模型

表示, 那么那个观测的残差通常很大. 这里“很大”指的是残差的绝对值. 因为

一个“很大”的残差可能是正的也可能是负的. 如果只有占很小百分比的观测

出现大的残差, 那么这些观测可能是异常点(outliers) | 它们不能用来与其

余数据一起拟合模型. 因此对数据中有残差“很大”的观测点, 必须仔细地检

查.

一般把标准化残差的绝对值≥2的观测点认为是可疑点; 而标准化残差的

绝对值≥3的观测点认为是异常点。

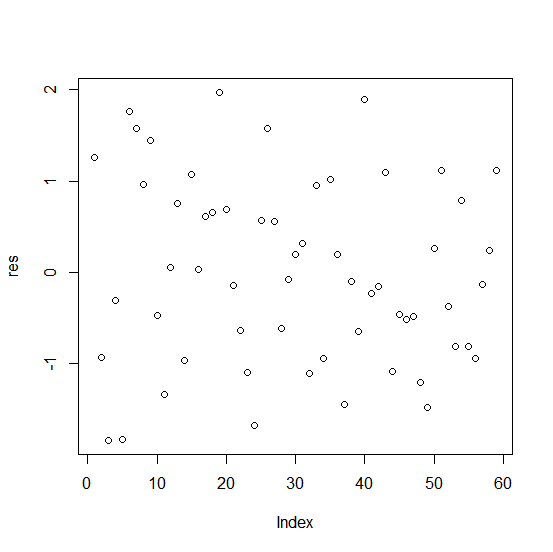
由例的计算结果并结合图形可以看出, 第58和39个点的残差比较大, 被认

定为异常点。

这里再做一个简单处理, 去掉第58和39观测样本点, 并重复上述回归分析及残差分

析的过程, 得到新的标准化残差图9.11. 与图9.10相比, 现在残差点的分布已有

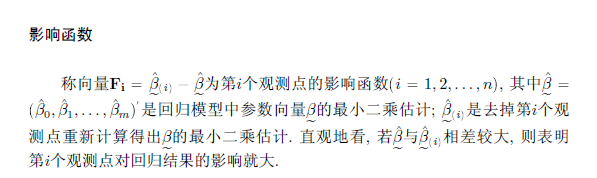
了很大的改进, 它们基本上落在[-2, 2]的带状区域内.



9.4.3 影响分析

从分析观测点对回归结果的影响入手, 找出对回归结果影响很大的观测点

的分析方法称为影响分析.



在R软件中, 函

数influence.measures( )可以做回归诊断中影响分析的概括

male\_height<-c(185, 173, 175, 182, 173, 181, 184, 179, 181, 187, 169, 178,

183, 168, 181, 175, 175, 186, 186, 182, 178, 177, 172, 168,

173.5, 184, 183, 175, 168, 174, 181, 170, 166, 178, 177, 181,

163, 172, 160, 173, 185, 172, 183, 180, 175, 178, 169, 175,

165, 169, 170, 183, 184, 174, 170, 173, 170, 182, 178, 170,

179)

male\_weight<-c(65, 62, 80.3, 74.3, 55.7, 60, 59, 79, 62, 80, 56, 60.5,

73, 46, 65, 91, 64, 88, 63, 64, 65, 75.6, 64, 65.5,

58.8, 59, 71, 75, 60, 61, 75, 58, 56, 94.5, 87, 71,

47.5, 59, 57, 65, 67, 60, 85, 65, 73, 70, 55, 75,

55, 65, 65, 72, 99, 75, 53, 70, 58, 63, 92, 48,

69)

male\_armspan<-c(188, 182, 183, 189, 185, 170, 179, 165, 171, 196, 168, 177,

179, 167, 175, 171, 170, 183, 180, 182, 180, 177, 176, 170,

167, 179, 186, 168, 160, 171.5, 181, 172, 146, 177, 168, 183,

162, 170.5, 166.7, 173, 176, 167, 188, 178, 178, 175, 169, 171,

164, 175, 161, 174, 188, 171, 166.6, 171, 169, 160, 175, 164,

169)

male\_leglength<-c(102, 93, 107, 114, 107, 101, 98, 99, 98, 119, 97, 101,

102, 99, 103, 97, 82, 101, 95, 97, 91, 95.5, 95, 82,

83, 98, 98, 105, 93, 96.5, 104.5, 85, 77, 105, 94, 102,

77, 95.8, 83.5, 89, 106, 98, 102, 93, 98, 90, 86, 98,

99, 99, 92, 100, 108, 102, 83, 101, 102, 80, 85, 92,

98)

male\_footlength<-c(25.1, 23.5, 26.5, 26, 26, 25, 26.5, 26, 26, 29, 26, 26.5,

27, 24.5, 25.5, 22, 24, 28, 26.5, 26.5, 26.5, 27, 25, 25.5,

24, 26.5, 26, 27, 22, 23.4, 25.5, 25.4, 22, 26.5, 22, 26,

23.8, 23, 22.8, 25.5, 25.8, 24.5, 26.5, 24, 26.5, 29, 23, 26,

22, 23, 22, 27.5, 26.5, 26, 20, 25.5, 24.5, 24, 25.5, 23,

25)

height<-data.frame(male\_height, male\_weight, male\_armspan,

male\_leglength, male\_footlength)

lm.reg<-lm(male\_height~male\_weight+male\_armspan

+male\_leglength+male\_footlength, data=blood)

lm.step<-step(lm.reg)

summary(lm.step)

influence.measures(lm.reg)

最后一句的结果如下：

Influence measures of

lm(formula = male\_height ~ male\_weight + male\_armspan + male\_leglength + male\_footlength, data = blood) :

dfb.1\_ dfb.ml\_w dfb.ml\_r dfb.ml\_l dfb.ml\_f dffit cov.r cook.d

1 -0.28362 -0.139347 3.66e-01 -0.021482 -0.20362 0.4327 1.057 3.70e-02

2 0.11743 0.066427 -2.85e-01 0.118043 0.20755 -0.3341 1.145 2.24e-02

3 0.20558 -0.147290 -4.98e-02 -0.164003 0.02763 -0.4360 0.837 3.63e-02

4 0.06737 0.013978 -4.38e-02 -0.064053 0.04300 -0.1131 1.233 2.60e-03

5 0.37499 0.479750 -2.90e-01 -0.249670 0.07470 -0.6695 0.898 8.57e-02

6 0.11077 -0.172490 -1.99e-01 0.269900 0.04887 0.4065 0.860 3.18e-02

7 -0.16136 -0.246137 1.03e-01 -0.032603 0.14620 0.3700 0.915 2.66e-02

8 0.17612 0.168512 -2.94e-01 0.100977 0.14567 0.3499 1.138 2.45e-02

9 0.05689 -0.110240 -1.55e-01 0.084180 0.17191 0.2981 0.940 1.74e-02

10 0.16678 0.027030 -5.79e-02 -0.093319 -0.02921 -0.2123 1.292 9.15e-03

11 -0.09989 0.190432 2.10e-01 -0.108040 -0.22194 -0.3725 0.998 2.73e-02

12 -0.00320 -0.007001 -6.56e-04 0.003086 0.00509 0.0112 1.150 2.55e-05

13 -0.05210 0.007088 -1.86e-02 0.020868 0.07620 0.1435 1.079 4.15e-03

14 -0.08084 0.255161 1.10e-01 -0.174761 -0.04085 -0.3383 1.128 2.29e-02

15 -0.01068 -0.068379 -4.90e-02 0.135452 0.01042 0.2039 1.021 8.29e-03

16 0.00426 0.011232 9.32e-05 0.000845 -0.00974 0.0147 1.396 4.41e-05

17 0.03880 0.023748 5.88e-02 -0.157960 -0.01082 0.1864 1.159 7.03e-03

18 -0.07240 0.120310 7.53e-04 -0.052674 0.08849 0.2053 1.162 8.52e-03

19 -0.22424 -0.191368 2.14e-01 -0.209856 0.16907 0.4638 0.796 4.07e-02

20 -0.09554 -0.066326 9.03e-02 -0.056601 0.03809 0.1581 1.106 5.05e-03

21 0.01547 0.007569 -1.94e-02 0.027368 -0.01299 -0.0384 1.179 3.00e-04

22 0.02338 -0.047197 1.12e-02 0.061629 -0.08626 -0.1426 1.112 4.11e-03

23 0.03404 0.049610 -7.93e-02 0.050016 0.02368 -0.1720 1.005 5.90e-03

24 -0.09378 -0.067342 -5.00e-02 0.487912 -0.21756 -0.5736 0.934 6.35e-02

25 0.05688 -0.012570 1.57e-02 -0.108935 0.01088 0.1512 1.142 4.63e-03

26 -0.16136 -0.246137 1.03e-01 -0.032603 0.14620 0.3700 0.915 2.66e-02

27 -0.10215 -0.009644 1.23e-01 -0.060711 -0.02706 0.1484 1.143 4.46e-03

28 -0.07392 -0.044129 2.00e-01 -0.125498 -0.12811 -0.2390 1.221 1.16e-02

29 -0.02018 -0.000978 1.21e-02 -0.009819 0.01130 -0.0259 1.207 1.36e-04

30 0.00938 -0.009923 5.52e-03 0.011230 -0.02470 0.0392 1.140 3.13e-04

31 -0.02609 0.012102 1.78e-02 0.024634 -0.02304 0.0639 1.132 8.29e-04

32 -0.00838 0.088035 -7.74e-02 0.241951 -0.11084 -0.3208 1.060 2.05e-02

33 0.46713 0.076246 -3.30e-01 -0.085470 0.06538 0.5016 1.286 5.04e-02

34 -0.01793 -0.290469 1.24e-01 -0.069965 -0.02578 -0.3543 1.154 2.52e-02

35 0.19116 0.364587 -3.30e-02 0.000948 -0.29334 0.4762 1.213 4.53e-02

36 -0.02479 -0.003745 2.16e-02 0.001305 -0.00579 0.0372 1.136 2.82e-04

37 -0.23303 0.195306 -7.32e-04 0.375407 -0.13705 -0.5834 1.041 6.66e-02

38 -0.00635 0.006666 -3.34e-03 -0.006348 0.01520 -0.0230 1.155 1.08e-04

39 -0.00635 0.003536 -2.26e-02 0.099351 -0.04673 -0.1366 1.104 3.77e-03

40 -0.04156 -0.109398 -1.30e-01 0.335766 0.03032 0.4370 0.818 3.63e-02

41 -0.02513 0.015047 2.98e-02 -0.027477 -0.00422 -0.0495 1.146 5.00e-04

42 0.02885 -0.021714 -2.78e-02 0.012535 0.00912 -0.0471 1.198 4.51e-04

43 -0.06015 -0.022175 1.94e-01 -0.106445 -0.14600 0.2611 1.039 1.36e-02

44 0.05715 -0.037151 -9.98e-03 0.044821 -0.08242 -0.1863 1.011 6.92e-03

45 0.01907 -0.001495 4.14e-02 0.103539 -0.18373 -0.2105 1.307 8.99e-03

46 -0.03604 0.040215 -5.00e-02 0.064098 0.05127 -0.1352 1.146 3.70e-03

47 -0.03284 -0.045425 6.51e-02 -0.014020 -0.04547 -0.1019 1.123 2.11e-03

48 -0.21378 0.115182 1.20e-01 -0.273961 0.23181 -0.4381 1.084 3.81e-02

49 -0.01066 0.025452 -1.43e-01 -0.112766 0.32750 -0.3950 0.956 3.05e-02

50 0.06216 0.020382 -3.27e-02 0.019786 -0.03988 0.0797 1.194 1.29e-03

51 0.00329 0.019009 -1.58e-01 0.032803 0.23398 0.2933 1.045 1.71e-02

52 0.05763 -0.118390 -3.56e-02 -0.007227 0.03996 -0.1587 1.282 5.12e-03

53 -0.05916 -0.066010 1.41e-01 -0.094379 -0.06784 -0.1970 1.092 7.81e-03

54 0.10763 -0.038090 1.69e-01 -0.102181 -0.29855 0.3819 1.279 2.94e-02

55 -0.05182 -0.020186 1.06e-01 -0.090208 -0.03803 -0.1613 1.073 5.24e-03

56 -0.07414 0.109356 1.12e-01 -0.182591 0.00952 -0.2482 1.082 1.23e-02

57 -0.00511 -0.051442 -8.18e-03 0.046188 -0.00239 -0.0669 1.378 9.13e-04

58 0.03350 -0.041119 -1.38e-02 0.016618 -0.01312 0.0676 1.182 9.32e-04

59 0.10581 0.041077 -1.41e-01 0.083860 0.03356 0.2077 1.012 8.59e-03

hat inf

1 0.1048

2 0.1151

3 0.0510

4 0.1175

5 0.1132

6 0.0489

7 0.0509

8 0.1159

9 0.0402

10 0.1677 \*

11 0.0708

12 0.0457

13 0.0350

14 0.1086

15 0.0346

16 0.2133 \*

17 0.0851

18 0.0914

19 0.0498

20 0.0502

21 0.0705

22 0.0490

23 0.0239

24 0.1010

25 0.0669

26 0.0509

27 0.0669

28 0.1320

29 0.0908

30 0.0399

31 0.0391

32 0.0768

33 0.2161 \*

34 0.1243

35 0.1786

36 0.0366

37 0.1363

38 0.0500

39 0.0431

40 0.0482

41 0.0461

42 0.0853

43 0.0541

44 0.0283

45 0.1764 \*

46 0.0651

47 0.0434

48 0.1159

49 0.0654

50 0.0865

51 0.0647

52 0.1548 \*

53 0.0552

54 0.1902 \*

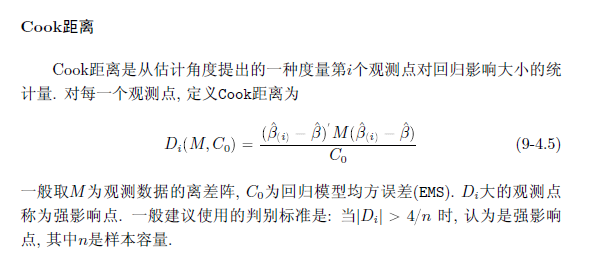
55 0.0380

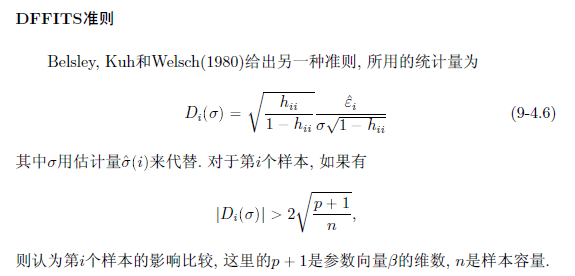
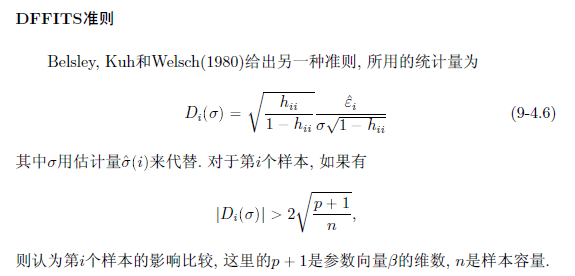
56 0.0655

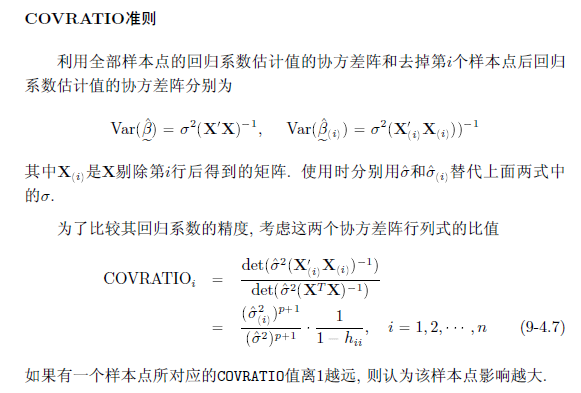
57 0.2048 \*

58 0.0759

59 0.0336







结论: 可以看出, 第10、16、33、45、52、54、57个观测点为强影响点, 结果中已用“\*”号标出. 其

中, 第一个样本点的cook.d值为2.2838比4/n=4/8=0.5 大得多; 第四个样本点

