

5.9 建立国家财政收入回归模型，由定性分析知，所选自变量都与因变量 y 具有较强的相关性，分别用后退法与逐步回归法换元。

1. 调用 step 函数

• 后退法

```
1 a<-read.table("D:/R/data/5.9.txt",sep="\t",header=T)
2 z<-lm(y~.,data=a)
3 library(leaps)
4 z.bwd<-step(z,direction="backward")
5 summary(z.bwd)
6
7 Call:
8 lm(formula = y ~ x1 + x2 + x5, data = a)
9
10 Residuals:
11      Min       1Q   Median       3Q      Max
12 -372.26 -102.79   -7.77  157.98  313.69
13
14 Coefficients:
15             Estimate Std. Error t value Pr(>|t|)
16 (Intercept)  874.60021   106.86563    8.184 2.67e-07 ***
17 x1           -0.61119    0.12382   -4.936 0.000125 ***
18 x2           -0.35305    0.08840   -3.994 0.000940 ***
19 x5            0.63671    0.08914    7.143 1.65e-06 ***
20 ---
21 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
22
23 Residual standard error: 183.1 on 17 degrees of freedom
24 Multiple R-squared:  0.9958, Adjusted R-squared:  0.9951
25 F-statistic: 1356 on 3 and 17 DF, p-value: < 2.2e-16
```

• 逐步回归法

```
1 z.both<-step(z,direction="both")
2 summary(z.both)
3 Call:
4 lm(formula = y ~ x1 + x2 + x5, data = a)
5
6 Residuals:
7      Min       1Q   Median       3Q      Max
8 -372.26 -102.79   -7.77  157.98  313.69
9
10 Coefficients:
11             Estimate Std. Error t value Pr(>|t|)
12 (Intercept)  874.60021   106.86563    8.184 2.67e-07 ***
13 x1           -0.61119    0.12382   -4.936 0.000125 ***
14 x2           -0.35305    0.08840   -3.994 0.000940 ***
15 x5            0.63671    0.08914    7.143 1.65e-06 ***
16 ---
17 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
18
19 Residual standard error: 183.1 on 17 degrees of freedom
```

```

20 Multiple R-squared:  0.9958,    Adjusted R-squared:  0.9951
21 F-statistic: 1356 on 3 and 17 DF,  p-value: < 2.2e-16

```

两方法得到最终结果均为:

$$\hat{y} = 874.6 - 0.611x_1 - 0.353x_2 + 0.637x_5.$$

2. 编程解决

• 后退法

```

1 x<-as.matrix(cbind(a[,1],a[,2],a[,3],a[,4],a[,5],a[,6]))
2 y<-as.martix(a[,7])
3 p<-ncol(x)
4 P<-vector(length=p)
5 P<-summary(z)$coefficients[,4]
6 max(P)
7 [1] 0.9319438
8 which.max(P)
9 x4
10 5

```

剔除最不显著的 x_4 再次回归。

```

1 z<-lm(y~x[,1]+x[,2]+x[,3]+x[,5]+x[,6])
2 P<-summary(z)$coefficients[,4]
3 max(P)
4 [1] 0.4282595
5 which.max(P)
6 x[, 3]
7 4

```

剔除 x_3 再次回归。

```

1 z<-lm(y~x[,1]+x[,2]+x[,5]+x[,6])
2 P<-summary(z)$coefficients[,4]
3 max(P)
4 [1] 0.3454117
5 which.max(P)
6 x[, 6]
7 5

```

剔除最不显著的 x_6 再次回归。

```

1 z<-lm(y~x[,1]+x[,2]+x[,5])
2 P<-summary(z)$coefficients[,4]
3 max(P)
4 [1] 0.0009398812

```

无变量可以继续剔除, 完成后退回过程。

• 逐步回归法

```

1 for(i in 1:p){
2 z<-lm(y~x[,i])
3 P[i]<-summary(z)$coefficients[2,4]
4 }
5 min(P)

```

```

6 [1] 5.896705e-20
7 which.min(P)
8 [1] 5

```

首先将 x_5 引入模型, 接下来进行双变量回归。

```

1 for(i in 1:p){
2 z<-lm(y~x[,i]+x[,5])
3 P[i]<-summary(z)$coefficients[2,4]
4 }
5 P[5]=1
6 min(P)
7 [1] 0.0148347
8 t<-which.min(P)
9 summary(lm(y~x[,t]+x[,5]))$coefficients[,4]
10 (Intercept)      x[, t]      x[, 5]
11 7.417462e-07 1.483470e-02 5.283411e-06
12 t
13 [1] 1

```

x_5 仍然更显著。将 x_5, x_1 加入模型。下面再加入一个变量进行回归。

```

1 for(i in 1:p){
2 z<-lm(y~x[,i]+x[,1]+x[,5])
3 P[i]<-summary(z)$coefficients[2,4]
4 }
5 P[5]<-1
6 P[1]<-1
7 min(P)
8 [1] 0.0009398812
9 t<-which.min(P)
10 summary(lm(y~x[,t]+x[,1]+x[,5]))$coefficients[,4]
11 (Intercept)      x[, t]      x[, 1]      x[, 5]
12 2.673898e-07 9.398812e-04 1.252522e-04 1.648223e-06
13 t
14 [1] 2

```

x_5, x_1 仍然更显著。将 x_5, x_1, x_2 加入模型。加入第四个变量进行回归。

```

1 for(i in 1:p){
2 z<-lm(y~x[,i]+x[,1]+x[,2]+x[,5])
3 P[i]<-summary(z)$coefficients[2,4]
4 }
5 P[5]<-1
6 P[1]<-1
7 P[2]<-1
8 min(P)
9 [1] 0.3454117

```

不再能加入显著的变量, 至此完成逐步回归过程。