实验5

谭铭瑞

2023-05-04

实验5-9

#实验5-9  
library(haven)  
xt5\_9 <- read\_sav("C:/Users/Administrator/Desktop/xt5.9.sav")  
View(xt5\_9)  
  
#后退法  
lm5.9 = lm(y~x1+x2+x3+x4+x5+x6,data = xt5\_9)  
summary(lm5.9)

##   
## Call:  
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6, data = xt5\_9)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -374.18 -82.44 -3.00 91.05 237.52   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.348e+03 2.211e+03 0.610 0.551859   
## x1 -6.410e-01 1.669e-01 -3.840 0.001804 \*\*   
## x2 -3.170e-01 2.044e-01 -1.551 0.143216   
## x3 -4.127e-01 5.485e-01 -0.752 0.464294   
## x4 -2.110e-03 2.428e-02 -0.087 0.931962   
## x5 6.711e-01 1.280e-01 5.241 0.000125 \*\*\*  
## x6 -7.541e-03 8.128e-03 -0.928 0.369220   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 191.8 on 14 degrees of freedom  
## Multiple R-squared: 0.9962, Adjusted R-squared: 0.9946   
## F-statistic: 618 on 6 and 14 DF, p-value: 3.81e-16

lm5.9.back = step(lm5.9,direction = 'backward')#使用step后退法

## Start: AIC=226.27  
## y ~ x1 + x2 + x3 + x4 + x5 + x6  
##   
## Df Sum of Sq RSS AIC  
## - x4 1 278 515554 224.28  
## - x3 1 20834 536110 225.10  
## - x6 1 31684 546960 225.52  
## <none> 515276 226.27  
## - x2 1 88536 603812 227.60  
## - x1 1 542591 1057867 239.37  
## - x5 1 1011046 1526322 247.07  
##   
## Step: AIC=224.28  
## y ~ x1 + x2 + x3 + x5 + x6  
##   
## Df Sum of Sq RSS AIC  
## - x3 1 22801 538355 223.19  
## - x6 1 39639 555193 223.83  
## <none> 515554 224.28  
## - x2 1 190111 705666 228.87  
## - x1 1 870111 1385665 243.04  
## - x5 1 1713154 2228708 253.02  
##   
## Step: AIC=223.19  
## y ~ x1 + x2 + x5 + x6  
##   
## Df Sum of Sq RSS AIC  
## - x6 1 31792 570147 222.39  
## <none> 538355 223.19  
## - x2 1 566230 1104585 236.28  
## - x1 1 847748 1386103 241.05  
## - x5 1 1704674 2243029 251.16  
##   
## Step: AIC=222.39  
## y ~ x1 + x2 + x5  
##   
## Df Sum of Sq RSS AIC  
## <none> 570147 222.39  
## - x2 1 534873 1105019 234.29  
## - x1 1 817120 1387267 239.06  
## - x5 1 1710882 2281029 249.51

summary(lm5.9.back)

所以此时用后退法自选出的变量为x1，x2，x5

##   
## Call:  
## lm(formula = y ~ x1 + x2 + x5, data = xt5\_9)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -372.27 -102.79 -7.78 157.94 313.69   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 874.58627 106.86620 8.184 2.67e-07 \*\*\*  
## x1 -0.61116 0.12382 -4.936 0.000125 \*\*\*  
## x2 -0.35304 0.08840 -3.994 0.000940 \*\*\*  
## x5 0.63669 0.08914 7.142 1.65e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 183.1 on 17 degrees of freedom  
## Multiple R-squared: 0.9958, Adjusted R-squared: 0.9951   
## F-statistic: 1356 on 3 and 17 DF, p-value: < 2.2e-16

X1，x2，x5显著

#逐步法  
  
lm5.9.both = step(lm5.9,direction = 'both')

## Start: AIC=226.27  
## y ~ x1 + x2 + x3 + x4 + x5 + x6  
##   
## Df Sum of Sq RSS AIC  
## - x4 1 278 515554 224.28  
## - x3 1 20834 536110 225.10  
## - x6 1 31684 546960 225.52  
## <none> 515276 226.27  
## - x2 1 88536 603812 227.60  
## - x1 1 542591 1057867 239.37  
## - x5 1 1011046 1526322 247.07  
##   
## Step: AIC=224.28  
## y ~ x1 + x2 + x3 + x5 + x6  
##   
## Df Sum of Sq RSS AIC  
## - x3 1 22801 538355 223.19  
## - x6 1 39639 555193 223.83  
## <none> 515554 224.28  
## + x4 1 278 515276 226.27  
## - x2 1 190111 705666 228.87  
## - x1 1 870111 1385665 243.04  
## - x5 1 1713154 2228708 253.02  
##   
## Step: AIC=223.19  
## y ~ x1 + x2 + x5 + x6  
##   
## Df Sum of Sq RSS AIC  
## - x6 1 31792 570147 222.39  
## <none> 538355 223.19  
## + x3 1 22801 515554 224.28  
## + x4 1 2245 536110 225.10  
## - x2 1 566230 1104585 236.28  
## - x1 1 847748 1386103 241.05  
## - x5 1 1704674 2243029 251.16  
##   
## Step: AIC=222.39  
## y ~ x1 + x2 + x5  
##   
## Df Sum of Sq RSS AIC  
## <none> 570147 222.39  
## + x6 1 31792 538355 223.19  
## + x3 1 14954 555193 223.83  
## + x4 1 11386 558761 223.97  
## - x2 1 534873 1105019 234.29  
## - x1 1 817120 1387267 239.06  
## - x5 1 1710882 2281029 249.51

summary(lm5.9.both)

所以此时用逐步回归法自选出的变量为x1，x2，x5

##   
## Call:  
## lm(formula = y ~ x1 + x2 + x5, data = xt5\_9)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -372.27 -102.79 -7.78 157.94 313.69   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 874.58627 106.86620 8.184 2.67e-07 \*\*\*  
## x1 -0.61116 0.12382 -4.936 0.000125 \*\*\*  
## x2 -0.35304 0.08840 -3.994 0.000940 \*\*\*  
## x5 0.63669 0.08914 7.142 1.65e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 183.1 on 17 degrees of freedom  
## Multiple R-squared: 0.9958, Adjusted R-squared: 0.9951   
## F-statistic: 1356 on 3 and 17 DF, p-value: < 2.2e-16

X1，x2，x5显著

实验5-10

library(haven)  
xt5\_10 <- read\_sav("C:/Users/Administrator/Desktop/xt5.10.sav")  
View(xt5\_10)  
  
#建立Y对x2~x6回归方程  
lm5.10 = lm(y~x2+x3+x4+x5+x6,data = xt5\_10)  
summary(lm5.10)

##   
## Call:  
## lm(formula = y ~ x2 + x3 + x4 + x5 + x6, data = xt5\_10)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -879.00 -272.40 -18.98 180.43 1153.61   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5922.8274 2504.3154 2.365 0.03961 \*   
## x2 4.8642 2.5074 1.940 0.08109 .   
## x3 2.3741 0.8424 2.818 0.01821 \*   
## x4 -817.9013 187.2787 -4.367 0.00141 \*\*  
## x5 14.5387 147.0778 0.099 0.92321   
## x6 -846.8669 291.6336 -2.904 0.01573 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 625.9 on 10 degrees of freedom  
## Multiple R-squared: 0.8237, Adjusted R-squared: 0.7356   
## F-statistic: 9.346 on 5 and 10 DF, p-value: 0.001574

#后退法  
  
lm5.10.back = step(lm5.10,direction = 'backward')#使用step后退法

## Start: AIC=210.53  
## y ~ x2 + x3 + x4 + x5 + x6  
##   
## Df Sum of Sq RSS AIC  
## - x5 1 3828 3921126 208.55  
## <none> 3917299 210.53  
## - x2 1 1474198 5391497 213.64  
## - x3 1 3111477 7028776 217.89  
## - x6 1 3303253 7220552 218.32  
## - x4 1 7471569 11388868 225.61  
##   
## Step: AIC=208.55  
## y ~ x2 + x3 + x4 + x6  
##   
## Df Sum of Sq RSS AIC  
## <none> 3921126 208.55  
## - x6 1 4908320 8829446 219.54  
## - x2 1 4950381 8871507 219.61  
## - x3 1 8042696 11963822 224.40  
## - x4 1 8603727 12524853 225.13

summary(lm5.10.back)

此时筛选出自变量x2，x3，x4，x6

##   
## Call:  
## lm(formula = y ~ x2 + x3 + x4 + x6, data = xt5\_10)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -877.60 -262.89 -19.53 172.95 1165.46   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6007.3203 2245.4805 2.675 0.021589 \*   
## x2 5.0681 1.3600 3.727 0.003343 \*\*   
## x3 2.3078 0.4858 4.750 0.000600 \*\*\*  
## x4 -824.2614 167.7764 -4.913 0.000462 \*\*\*  
## x6 -862.6990 232.4888 -3.711 0.003437 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 597 on 11 degrees of freedom  
## Multiple R-squared: 0.8235, Adjusted R-squared: 0.7594   
## F-statistic: 12.84 on 4 and 11 DF, p-value: 0.0003973

X2,x3,x4,x6显著

#逐步法  
  
lm5.10.both = step(lm5.10,direction = 'both')

## Start: AIC=210.53  
## y ~ x2 + x3 + x4 + x5 + x6  
##   
## Df Sum of Sq RSS AIC  
## - x5 1 3828 3921126 208.55  
## <none> 3917299 210.53  
## - x2 1 1474198 5391497 213.64  
## - x3 1 3111477 7028776 217.89  
## - x6 1 3303253 7220552 218.32  
## - x4 1 7471569 11388868 225.61  
##   
## Step: AIC=208.55  
## y ~ x2 + x3 + x4 + x6  
##   
## Df Sum of Sq RSS AIC  
## <none> 3921126 208.55  
## + x5 1 3828 3917299 210.53  
## - x6 1 4908320 8829446 219.54  
## - x2 1 4950381 8871507 219.61  
## - x3 1 8042696 11963822 224.40  
## - x4 1 8603727 12524853 225.13

summary(lm5.10.both)

此时筛选出自变量x2，x3，x4，x6

##   
## Call:  
## lm(formula = y ~ x2 + x3 + x4 + x6, data = xt5\_10)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -877.60 -262.89 -19.53 172.95 1165.46   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6007.3203 2245.4805 2.675 0.021589 \*   
## x2 5.0681 1.3600 3.727 0.003343 \*\*   
## x3 2.3078 0.4858 4.750 0.000600 \*\*\*  
## x4 -824.2614 167.7764 -4.913 0.000462 \*\*\*  
## x6 -862.6990 232.4888 -3.711 0.003437 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 597 on 11 degrees of freedom  
## Multiple R-squared: 0.8235, Adjusted R-squared: 0.7594   
## F-statistic: 12.84 on 4 and 11 DF, p-value: 0.0003973

X2,x3,x4,x6显著