lineargression

why

*线性回归模型*

##   
## Call:  
## lm(formula = formula, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.6960 -0.6525 -0.1114 0.8768 2.2384   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 35.3507 5.2684 6.710 8.75e-05 \*\*\*  
## x1 164.3917 26.3805 6.232 0.000153 \*\*\*  
## x2 -0.2365 0.1740 -1.359 0.207150   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.267 on 9 degrees of freedom  
## Multiple R-squared: 0.9569, Adjusted R-squared: 0.9473   
## F-statistic: 99.94 on 2 and 9 DF, p-value: 7.155e-07

*回归方程的显著性F检验*

## Anova Table (Type III tests)  
##   
## Response: y  
## Sum Sq Df F value Pr(>F)   
## (Intercept) 72.258 1 45.0228 8.754e-05 \*\*\*  
## x1 62.323 1 38.8322 0.000153 \*\*\*  
## x2 2.965 1 1.8475 0.207150   
## Residuals 14.444 9   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

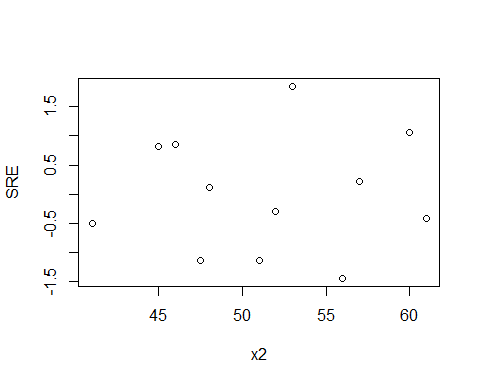
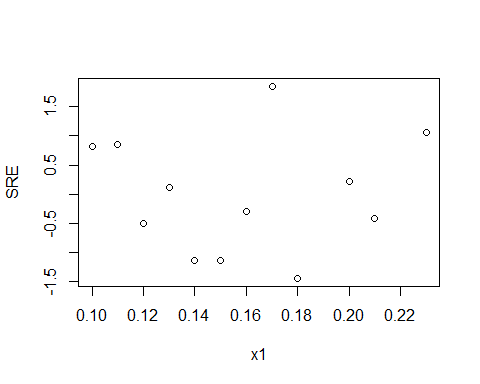
*相关系数显著性检验F检验*

##   
## Pearson's product-moment correlation  
##   
## data: x and data$y  
## t = 13.511, df = 10, p-value = 9.505e-08  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.9061336 0.9928074  
## sample estimates:  
## cor   
## 0.9736872   
##   
##   
## Pearson's product-moment correlation  
##   
## data: x and data$y  
## t = 5.8024, df = 10, p-value = 0.0001724  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.6131406 0.9654532  
## sample estimates:  
## cor   
## 0.8780662

## [,1] [,2] [,3]  
## [1,] 1.0000000 0.9010229 -0.4126973  
## [2,] 0.9010229 1.0000000 0.7416070  
## [3,] -0.4126973 0.7416070 1.0000000

*回归系数置信区间估计*

## 2.5 % 97.5 %  
## (Intercept) 23.4326582 47.2687200  
## x1 104.7147360 224.0686329  
## x2 -0.6301606 0.1571161

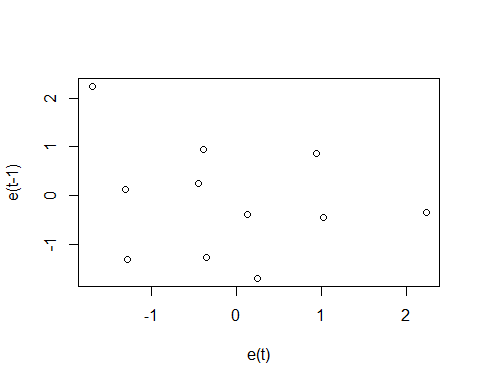
*残差分析*  *异方差等级相关系数检验 x与SRE*

##   
## Spearman's rank correlation rho  
##   
## data: x and abse  
## S = 246, p-value = 0.6672  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 0.1398601   
##   
##   
## Spearman's rank correlation rho  
##   
## data: x and abse  
## S = 256, p-value = 0.7495  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 0.1048951

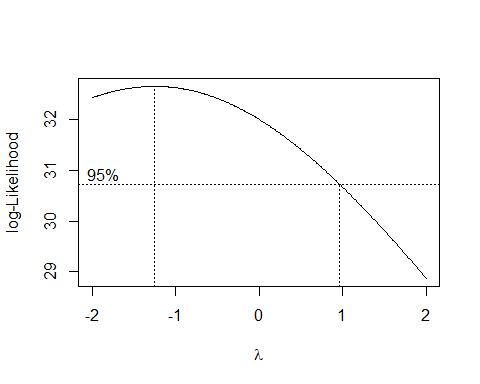
*异方差消除 一元加权最小二乘法*

## [1] -2

## [[1]]  
##   
## Call:  
## lm(formula = formula, data = data, weights = c(w))  
##   
## Weighted Residuals:  
## Min 1Q Median 3Q Max   
## -0.27961 -0.09583 -0.01391 0.12714 0.38031   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 38.0462 5.8165 6.541 0.000106 \*\*\*  
## x1 187.2907 28.5125 6.569 0.000103 \*\*\*  
## x2 -0.3608 0.1909 -1.890 0.091388 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2022 on 9 degrees of freedom  
## Multiple R-squared: 0.9601, Adjusted R-squared: 0.9512   
## F-statistic: 108.2 on 2 and 9 DF, p-value: 5.079e-07

*自相关判断 残差散点图 DW检验* 

##   
## Durbin-Watson test  
##   
## data: LinearRegression  
## DW = 2.326, p-value = 0.7575  
## alternative hypothesis: true autocorrelation is not 0

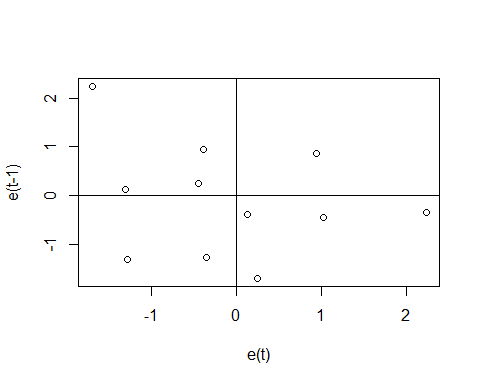
*box-cox变换* 

## [1] -1.25

##   
## Call:  
## lm(formula = formula, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.6960 -0.6525 -0.1114 0.8768 2.2384   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 35.3507 5.2684 6.710 8.75e-05 \*\*\*  
## x1 164.3917 26.3805 6.232 0.000153 \*\*\*  
## x2 -0.2365 0.1740 -1.359 0.207150   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.267 on 9 degrees of freedom  
## Multiple R-squared: 0.9569, Adjusted R-squared: 0.9473   
## F-statistic: 99.94 on 2 and 9 DF, p-value: 7.155e-07

*bc异方差等级相关系数检验*

##   
## Spearman's rank correlation rho  
##   
## data: as.vector(as.matrix(data[i])) and abse\_bc  
## S = 246, p-value = 0.6672  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 0.1398601   
##   
##   
## Spearman's rank correlation rho  
##   
## data: as.vector(as.matrix(data[i])) and abse\_bc  
## S = 256, p-value = 0.7495  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 0.1048951

*bc自相关判断* 

##   
## Durbin-Watson test  
##   
## data: lm\_bc  
## DW = 2.326, p-value = 0.7575  
## alternative hypothesis: true autocorrelation is not 0

*自变量选择 全子集回归*

## x1 x2 调整R平方  
## 1 ( 1 ) \* 0.9428734  
## 2 ( 1 ) \* \* 0.9473368

*前进法 后退法*

## Start: AIC=41.96  
## y ~ 1  
##   
## Df Sum of Sq RSS AIC  
## + x1 1 317.82 17.41 8.465  
## + x2 1 258.46 76.77 26.270  
## <none> 335.23 41.959  
##   
## Step: AIC=8.47  
## y ~ x1  
##   
## Df Sum of Sq RSS AIC  
## + x2 1 2.9652 14.444 8.2248  
## <none> 17.410 8.4654  
##   
## Step: AIC=8.22  
## y ~ x1 + x2

##   
## Call:  
## lm(formula = y ~ x1 + x2, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.6960 -0.6525 -0.1114 0.8768 2.2384   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 35.3507 5.2684 6.710 8.75e-05 \*\*\*  
## x1 164.3917 26.3805 6.232 0.000153 \*\*\*  
## x2 -0.2365 0.1740 -1.359 0.207150   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.267 on 9 degrees of freedom  
## Multiple R-squared: 0.9569, Adjusted R-squared: 0.9473   
## F-statistic: 99.94 on 2 and 9 DF, p-value: 7.155e-07

*逐步回归*

## Start: AIC=8.47  
## y ~ x1  
##   
## Df Sum of Sq RSS AIC  
## + x2 1 2.97 14.44 8.225  
## <none> 17.41 8.465  
## - x1 1 317.82 335.23 41.959  
##   
## Step: AIC=8.22  
## y ~ x1 + x2  
##   
## Df Sum of Sq RSS AIC  
## <none> 14.444 8.2248  
## - x2 1 2.965 17.410 8.4654  
## - x1 1 62.323 76.767 26.2705

##   
## Call:  
## lm(formula = y ~ x1 + x2, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.6960 -0.6525 -0.1114 0.8768 2.2384   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 35.3507 5.2684 6.710 8.75e-05 \*\*\*  
## x1 164.3917 26.3805 6.232 0.000153 \*\*\*  
## x2 -0.2365 0.1740 -1.359 0.207150   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.267 on 9 degrees of freedom  
## Multiple R-squared: 0.9569, Adjusted R-squared: 0.9473   
## F-statistic: 99.94 on 2 and 9 DF, p-value: 7.155e-07

*多重共线性判断 特征根*

## x1 x2  
## x1 1.0000000 0.9358368  
## x2 0.9358368 1.0000000

## [1] 30.17051