OLS

用Stata系统自带的auto.dta导出一份csv，读到R里面来展示OLS

autos <- tbl\_df(read.csv("d:/auto.csv",header = TRUE)) %>%  
 print

## Source: local data frame [74 x 12]  
##   
## make price mpg rep78 headroom trunk weight length turn  
## 1 AMC Concord 4099 22 3 2.5 11 2930 186 40  
## 2 AMC Pacer 4749 17 3 3.0 11 3350 173 40  
## 3 AMC Spirit 3799 22 NA 3.0 12 2640 168 35  
## 4 Buick Century 4816 20 3 4.5 16 3250 196 40  
## 5 Buick Electra 7827 15 4 4.0 20 4080 222 43  
## 6 Buick LeSabre 5788 18 3 4.0 21 3670 218 43  
## 7 Buick Opel 4453 26 NA 3.0 10 2230 170 34  
## 8 Buick Regal 5189 20 3 2.0 16 3280 200 42  
## 9 Buick Riviera 10372 16 3 3.5 17 3880 207 43  
## 10 Buick Skylark 4082 19 3 3.5 13 3400 200 42  
## .. ... ... ... ... ... ... ... ... ...  
## Variables not shown: displacement (int), gear\_ratio (dbl), foreign (fctr)

运行OLS回归，将结果存入到ols1变量中,I表示隔离,:表示交互项的相乘关系，而\*表示完全交互，隐含了主效用，事实上ols1和ols2的运行结果是相同的，里面的公式含义一致。

ols1<-lm(price~headroom+weight+I(headroom^2)+length+foreign:weight+foreign,data=autos)  
ols2<-lm(price~headroom+I(headroom^2)+length+foreign\*weight,data=autos)  
  
#展示回归结果  
summary(ols1)

##   
## Call:  
## lm(formula = price ~ headroom + weight + I(headroom^2) + length +   
## foreign:weight + foreign, data = autos)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3673 -1029 -317 999 5927   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 3426.451 4933.163 0.69 0.490  
## headroom 1214.780 1942.218 0.63 0.534  
## weight 5.342 0.966 5.53 5.7e-07  
## I(headroom^2) -262.396 317.003 -0.83 0.411  
## length -81.987 33.451 -2.45 0.017  
## foreignForeign -2537.421 2730.801 -0.93 0.356  
## weight:foreignForeign 2.426 1.078 2.25 0.028  
##   
## Residual standard error: 1960 on 67 degrees of freedom  
## Multiple R-squared: 0.597, Adjusted R-squared: 0.56   
## F-statistic: 16.5 on 6 and 67 DF, p-value: 1.41e-11

summary(ols2)

##   
## Call:  
## lm(formula = price ~ headroom + I(headroom^2) + length + foreign \*   
## weight, data = autos)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3673 -1029 -317 999 5927   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 3426.451 4933.163 0.69 0.490  
## headroom 1214.780 1942.218 0.63 0.534  
## I(headroom^2) -262.396 317.003 -0.83 0.411  
## length -81.987 33.451 -2.45 0.017  
## foreignForeign -2537.421 2730.801 -0.93 0.356  
## weight 5.342 0.966 5.53 5.7e-07  
## foreignForeign:weight 2.426 1.078 2.25 0.028  
##   
## Residual standard error: 1960 on 67 degrees of freedom  
## Multiple R-squared: 0.597, Adjusted R-squared: 0.56   
## F-statistic: 16.5 on 6 and 67 DF, p-value: 1.41e-11

#展示factors类别变量的  
#beta  
coef(ols1)

## (Intercept) headroom weight   
## 3426.451 1214.780 5.342   
## I(headroom^2) length foreignForeign   
## -262.396 -81.987 -2537.421   
## weight:foreignForeign   
## 2.426

#残差  
resid(ols1)

## 1 2 3 4 5 6 7 8   
## -1127.79 -3673.19 -1240.09 -56.47 144.28 -32.32 1768.23 -742.78   
## 9 10 11 12 13 14 15 16   
## 2151.34 -2148.26 2284.72 5926.54 5003.64 567.20 -714.09 -1124.86   
## 17 18 19 20 21 22 23 24   
## -507.24 -1155.09 -2681.49 1215.77 -2420.20 -544.20 819.48 2166.72   
## 25 26 27 28 29 30 31 32   
## -100.85 279.39 2411.91 5020.54 -807.56 -2655.45 -790.95 -1963.36   
## 33 34 35 36 37 38 39 40   
## -2678.76 -304.08 1074.15 -1084.02 -242.62 -1037.17 -1004.60 -438.25   
## 41 42 43 44 45 46 47 48   
## 1267.01 -3637.54 2857.39 2169.55 3286.80 -190.93 -510.54 -2028.59   
## 49 50 51 52 53 54 55 56   
## -253.83 -520.79 -1667.14 -329.55 1029.20 2194.17 1374.47 -365.09   
## 57 58 59 60 61 62 63 64   
## 155.71 -981.13 -434.46 -1336.76 327.77 756.70 -686.82 237.21   
## 65 66 67 68 69 70 71 72   
## -850.80 -967.33 -843.08 -1986.40 -2943.87 2176.05 85.67 240.19   
## 73 74   
## 1911.88 906.70

#对数似然值  
logLik(ols1)

## 'log Lik.' -662.1 (df=8)

#AIC  
AIC(ols1)

## [1] 1340

#拟合值或者预测时间序列的新值  
predict(ols1)

## 1 2 3 4 5 6 7 8 9 10 11 12   
## 5227 8422 5039 4872 7683 5820 2685 5932 8221 6230 9100 8573   
## 13 14 15 16 17 18 19 20 21 22 23 24   
## 10902 2732 6419 5629 5611 4822 6636 2768 6430 6430 5523 2222   
## 25 26 27 28 29 30 31 32 33 34 35 36   
## 4288 11218 11182 8445 4637 8034 6956 6479 8982 3595 7740 6256   
## 37 38 39 40 41 42 43 44 45 46 47 48   
## 4976 5927 5186 4633 9104 8285 1568 2312 3199 4251 6309 6963   
## 49 50 51 52 53 54 55 56 57 58 59 60   
## 5476 5244 6091 4502 8661 4101 8361 6594 4433 6060 8563 5633   
## 61 62 63 64 65 66 67 68 69 70 71 72   
## 5471 3742 4682 12753 4746 4765 6742 5734 8663 4964 5311 4457   
## 73 74   
## 4938 11088

#wald-test来模型对比,去掉length变量，两个模型对比  
ols3<-update(ols1, .~.-length)  
summary(ols3)

##   
## Call:  
## lm(formula = price ~ headroom + weight + I(headroom^2) + foreign +   
## weight:foreign, data = autos)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3176 -1248 -286 975 6463   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -6575.452 2872.067 -2.29 0.025  
## headroom 2102.662 1977.140 1.06 0.291  
## weight 3.233 0.454 7.11 8.9e-10  
## I(headroom^2) -437.032 320.062 -1.37 0.177  
## foreignForeign -2031.640 2821.466 -0.72 0.474  
## weight:foreignForeign 2.199 1.113 1.98 0.052  
##   
## Residual standard error: 2030 on 68 degrees of freedom  
## Multiple R-squared: 0.56, Adjusted R-squared: 0.528   
## F-statistic: 17.3 on 5 and 68 DF, p-value: 5e-11

#只显示系数检验  
coeftest(ols3)

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -6575.452 2872.067 -2.29 0.025  
## headroom 2102.662 1977.140 1.06 0.291  
## weight 3.233 0.454 7.11 8.9e-10  
## I(headroom^2) -437.032 320.062 -1.37 0.177  
## foreignForeign -2031.640 2821.466 -0.72 0.474  
## weight:foreignForeign 2.199 1.113 1.98 0.052

#结果显著，表明length这个变量效果是显著的  
waldtest(ols1,ols3)

## Wald test  
##   
## Model 1: price ~ headroom + weight + I(headroom^2) + length + foreign:weight +   
## foreign  
## Model 2: price ~ headroom + weight + I(headroom^2) + foreign + weight:foreign  
## Res.Df Df F Pr(>F)  
## 1 67   
## 2 68 -1 6.01 0.017