

Multiple Regression Analysis: Further Issues

实验1：数据测度单位改变对OLS统计量的影响

统计量类型	因变量单位变换影响	自变量单位变换影响	对数变量单位变换单位
系数估计值	等比例缩放	反比例缩放	斜率系数不变
标准误	等比例缩放	反比例缩放	斜率系数的SE不变
置信区间宽度	等比例缩放	反比例缩放	斜率系数的CI不变
t、F	不变	不变	斜率系数的t不变
R^2	不变	不变	不变
SSR、SER	改变	不变	不变

实验2：标准化系数

*标准化变量

```
egen z_x = std( x )
```

*比较解释变量的重要程度

```
reg z_y z_x1 z_x2 z_x3, noc
```

*等价于：

```
reg y x1 x2 x3, beta
```

如何比较解释变量的系数相对大小？

实验3：含对数的模型

精确百分比的计算与使用条件

```
dis 100*( exp(_b[x]) - 1 )
```

实验4：含二次项的模型

$$y = ax^2 + bx + c$$

$$dy/dx = 2ax + b$$

*生成交互项

*foreach (已讲)

```
reg y c.x##c.x //生成x的一次项与x的平方项
```

```
reg y c.x#(c.p2-p9) //只生成x与p2-p9的系列交互项
```

```
reg y c.x##(c.p2-p9) //生成x与p2-p9的系列交互项，和x与p2-p9本身
```

*二次拟合

*qfit (已讲)

*中心化处理

```
sum x
```

```
gen x_centered = x - r(mean)
```

```
sysuse auto, clear
reg price mpg
est store m1
reg price c.mpg##c.wei //基本回归+交乘项
est store m2
esttab m1 m2

center mpg wei, prefix(C_)

reg price mpg wei c.C_mpg#c.C_wei //只对交乘项去心
est store m3
reg price c.C_mpg##c.C_wei //对主变量、调节变量和交乘项都做去心
est store m4
esttab m1 m2 m3 m4, mtitle(ols nocenter center_inter center_all)
```

*-Notes:

- *-中心化仅是方便一次项系数的解释，不能克服共线性，也不能解决内生性；
- *-只关注交乘项的系数，中心化与否均可；
- *-虚拟变量无需中心化

实验5：含交叉项的模型

$$y = a + b_1X + e$$

$$dy/dX = b_1$$

$$y = a + b_1X + b_2Z + b_3(XZ) + e$$

$$dy/dX = b_1 + b_3Z$$

X 的边际效果依赖于 Z:

- 若 b_1 和 b_3 符号相同, 则表明随着 Z 的增加, X 对 y 的边际影响得以"加强";
- 若 b_1 和 b_3 符号不同, 则表明随着 Z 的增加, X 对 y 的边际影响会"减弱".

主效应项系数的方向和显著性重要么？

不重要，b1的含义：当Z=0时, X的变动平均会带来y的变动

主效应项要不要？

交互项？分组回归？

```
help margin
```

*再参数化

```
sum x1
```

```
scalar x1_mean = r(mean)
```

```
gen x1cx2 = ( x1 - x1_mean ) * x2
```


实验6：拟合优度和变量选择

$$R^2 = [\text{corr}(y, \hat{y})]^2 \quad (\text{证明})$$

R2 越高越好吗？

R2分解：相对重要性分析 (Dominance Analysis) `domin`

实验7：预测和残差分析

```
reg y $x, r  
predict yhat //加不加xb?  
gen resid = y - yhat  
  
predict uhat, residual
```

```
bcuse gpa2, clear
regress colgpa sat hsperc hsize hsize_sq
```

*预测: sat = 1200、hsperc = 30、hsize = 5时的预测值:

*数据集中没有符合条件的观测值, 需添加一条新观测:

```
set obs `=_N+1'
replace sat = 1200 in `=_N'      // `=_N' 表示最后一行
replace hsperc = 30 in `=_N'
replace hsize = 5 in `=_N'
replace hsize_sq = 25 in `=_N'   // hsize=5时平方为25
```

*生成预测值

```
predict yhat if _n == _N, xb
list yhat if _n == _N
```

*-Note: 利用回归模型预测时, 解释变量的值最好不要离开样本范围太远

*6-4d Predicting y When the Dependent Variable Is $\log(y)$:

```
reg lny x1 x2 x3
predict lnyh
predict uh, res
gen eu = exp(uh)
egen sumeu = sum(eu)
dis sumeu/_N //smearing estimator偏误修正估计值
gen m=exp(lnyh)
dis a0h*m //y的预测值
```

*过原点的回归

```
reg y m, noc //斜率即为 $\alpha_0\_check$ 的估计值
```

```
corr y yf
```

*-Note: 因变量为 $\ln y$ 时转换为 y 后的可决系数= y 的预测值与 y 观测值相关系数的平方

```

*Baum_4.6.1 Computing interval predictions (self-reading)
use http://www.stata-press.com/data/imeus/hprice2a, clear
quietly regress lprice lnox if _n<=100
predict double xb if e(sample)
predict double stdpred if e(sample), stdp
scalar tval = invttail(e(df_r), 0.025)
generate double uplim = xb + tval * stdpred
generate double lowlim = xb - tval * stdpred
summarize lnox if e(sample), meanonly
local lnoxbar = r(mean)
label var xb "Pred"
label var uplim "95% prediction interval"
label var lowlim "95% prediction interval"
twoway (scatter lprice lnox if e(sample), ///
sort ms(0h) xline(`lnoxbar')) ///
(connection xb lnox if e(sample), sort msize(small)) ///
(rline uplim lowlim lnox if e(sample), sort), ///
yttitle(Actual and predicted log price) legend(cols(3))

```

Bootstrap

Stata: Bootstrap-自抽样-自举法

Stata: 手动实现置换检验(permutation)和自抽样(bootstrap)

本章主要参考资料:

[课件/open5_regress.zip · lianxh/Stata公开课-连享会 - Gitee.com](#)

[线性回归中相关系数与决定系数相等的证明 - 知乎](#)

[相关系数和R方的关系是什么? - 知乎](#)

[Stata数据处理：各种求和方式一览](#)