

Multiple Regression Analysis with Qualitative Information

Summary

Binary Dummy:

1. 虚拟变量陷阱：有 m 个虚拟变量类别
 - 模型带常数项，至多能设 $m-1$ 个虚拟变量；
 - 模型不带常数项，可以设 m 个虚拟变量。
2. 虚拟变量的作用：
 - 测量截距变动
 - 测量斜率变动
 - 测量截距和斜率变动
3. 截距和虚拟变量系数的含义、显著性检验

Multiple Categories Dummies:

1. 回顾 `c.` : 连续型变量 `c.x#c.x` `c.x#c.y`
`i.` : 虚拟变量 `i.Dummy1#i.Dummy2`
2. 如何设基础组
3. 包含序数信息的虚拟变量
4. 虚拟变量只能代表质的因素, 不代表数量因素!
5. 生成虚拟变量: `gen dummy = (判断条件)`

Interactions with Dummies:

*Example 7.10: Log Hourly Wage Equation

```
reg lwage i.female##i.educ c.exper##c.exper c.tenure##c.tenure
```

```
testparm 1.female 1.female#c.educ //检验男女的平均工资是否相同
```

```
testparm 1.female#c.educ //检验男女的教育回报率是否相同
```

Chow statistic:

$$F = \frac{[SSR_P - (SSR_1 + SSR_2)]}{SSR_1 + SSR_2} \cdot \frac{[n - 2(k + 1)]}{k + 1}$$

Attention: 含 `i.x` 回归方程的检验不能用 `test` 而是用 `testparm` 。

Chow test:

```
chowtest y x1 x2 x3, group(组别)
```

LPM:

1. 系数解释
2. y 的预测值是 $y=1$ 的预测概率
3. 正确预测百分比
4. LPM 的优缺点与解决
 - robust
 - logit , probit

```
reg y $x, r
predict yhat
replace yhat = max(0,min(1,yhat))
gen pred_class = (yhat >= 0.5)
```