知乎

R语言笔记.formula

瑶瑶

7 人赞同了该文章

使用lm、glm和plm等函数进行回归估计，或是使用cast等函数进行数据重塑的时候，都要用到一类特殊的对象：formula。回归模型的表达式就是一个formula对象。

```
f <- formula(y~x+z)
f <- as.formula(y~x+z)
f <- y~x+z
#以上三行代码完全等价,创建的都是一个formula对象
```

y~x+z是一个简单的formula。~和+是formula中的运算符，但它们与人们通常理解中的数学运算符相去甚远。

以下是formula中运算符的含义：

- ~： ~连接公式两侧，~的左侧是因变量，右侧是自变量。
- +： 模型中不同的项用+分隔。注意R语言中默认表达式带常数项，因此估计 $y=\beta_0+\beta_1 x$ 只需要写y~x。
- -： -表示从模型中移除某一项，y~x-1表示从模型中移除常数项，估计的是一个不带截距项的过原点的回归方程。此外，y~x+0或y~0+x也可以表示不带截距项的回归方程。
- ：： 冒号在formula中表示交互项
- *： *不表示乘法， $a*b$ 与 $a+b+a:b$ 是等价的， $(a+b+c)*(a+b+c)$ 与 $a+b+c+a:b+b:c+c:a$ 等价
- ^： $(a+b)$ ^2与 $(a+b)*(a+b)$ 等价，所以 a ^2在formula中并不是 a 的平方的意思

如果想要在表达式中加入数学运算符，应该怎么办呢？对某一变量取对数，可以直接写log(y)~log(x)，这一表达式的含义就是估计 $log(y)=\beta_1 log(x)+\beta_0$ ；自然指数同样也可以直接表示为exp()；但如果想要表示加减乘除和平方之类，需要用到I()这个运算符。(←是大写的i不是小写的L)

y~x+I(z^2）的含义： $y=\beta_0+\beta_1 x+\beta_2 z^2$
y~x+z^2的含义： $y=\beta_0+\beta_1 x+\beta_2 z$ （因为z没法和自己交互）

那么，y~x+w+z和y~x+I(w+z)有什么区别呢？

y~x+w+z的含义： $y=\beta_0+\beta_1 x+\beta_2 w+\beta_3 z$
y~x+I(w+z)的含义： $y=\beta_0+\beta_1 x+\beta_2(w+z)$

可以看到，第二个式子将w+z作为一个整体估计这一变量的参数。

如果要估计动态面板模型，在plm包中，滞后变量(lagged variable)用运算符lag()表示，如lag(x,1)表示x滞后一期的滞后变量，lag(log(z),2)表示log(z)滞后两期的滞后变量；差分项则使用运算符diff()表示。正如我们使用formula和as.formula函数创建formula对象一样，dynformula函数用于创建动态面板模型的表达式，在这里不详细展开。

发布于 2018-07-24