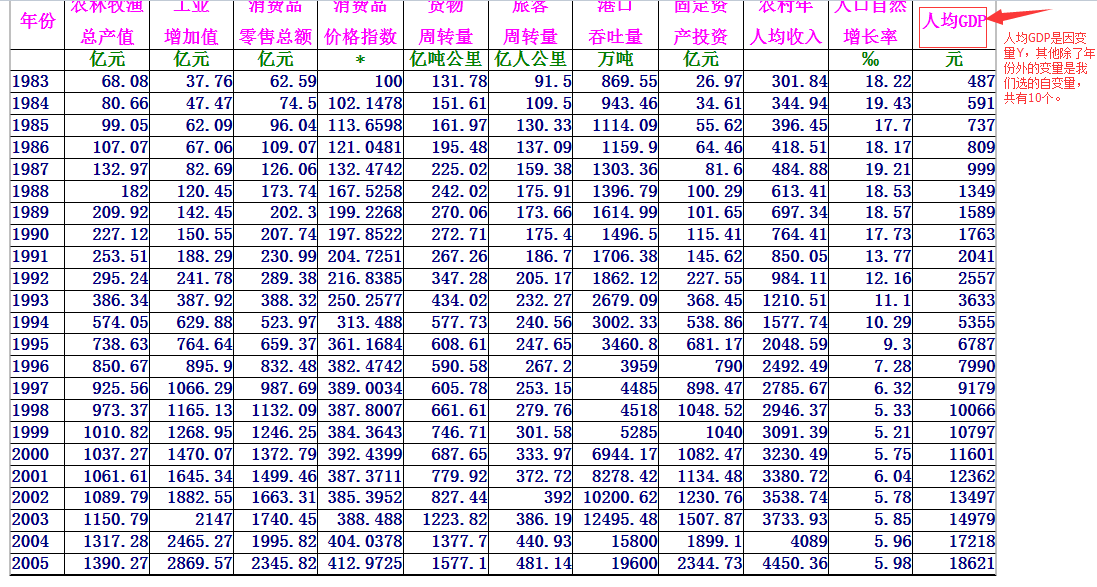
多因素滑移回归预测

1、前移回归（把新的年份的因变量值作为旧的年份的预测值）

（注：我直接用例子进行说明）



【注】这里的年份是从1983到2005，说明是升序的。考虑到我们遇到的数据可能年份是从2005-1983，即降序的。所以，在前端，会让用户选择他的表格的时间是升序还是降序的。

这里我们假设是按照这个表格，即升序来进行前移回归分析。

【要做的】

如果是升序，我们就是把因变量整列的数据**向上**移动一位，即2005年的18621移动到了2004年，2004年的17218移动到了2003……，1984的591移动到1983,1983的因变量向上就没了（不考虑）。这时候2005年对应的行的因变量就空缺了。自变量的数据位置不要动。

向上移动一位后，得到的**完整数据**就从1983-2004了，用这些数据代入你的逐步回归方程中，得到逐步回归方程式子，用2005年的自变量数据，就能得到2006年的预测值了。

如果是降序的（2005到1983），我们还是一样需要把2005的值给2004,2004的值给2003……,1984的值给1983，只是这时候我们数据要**向下**移动一位才能实现这种效果，得到2004-1983的完整数据，代入逐步回归的方法中，得到逐步回归的方程，用2005的自变量值代入，就能得到2006年的预测值了。（一样）

【结果显示】

可以直接显示最后得到的预测回归的方程里的系数表、R方情况，anova。然后显示得到的回归方程，再给出得到的预测值是什么。

1. 后移回归（把旧的年份的因变量值作为新的年份的因变量值）

同样利用的是上面的数据，前移回归是预测下一年即2006年的值，后移回归就是为了得到早些年的值，如这里我们想要知道1982年的人均GDP是多少（因变量），就需要用到后移动回归。

后移回归也是要考虑原始数据的表格的年份是如何排列的，如果是升序（1983-2005），我们想要用1983的值作为1984的因变量，1984作为1985的应变量……，2004作为2005的因变量，自变量的数据不要变动，1983的因变量是空缺的，需要我们**向下**移动一位，这时候能够得到完整信息的是1984-2005的数据，把1984-2005的数据代入逐步回归中，得到逐步回归的方程式，用1983的自变量的值代入公式中去，就能得到1982的因变量值。

如果时间序列是降序的，即（2005-1983），我们还是需要用2004的值作为2005的因变量的值……1983因变量的值作为1984的值，就需要对因变量的这一列进行**向上**移动一位。同样利用完整的2005-1984的数据，代入逐步回归方程，得出结果，用1983的自变量数据代入，就能得出1982年的因变量的值。

【总结】

1. 前移回归法

①时间升序（因变量向上移动一位）

②时间降序（因变量向下移动一位）

1. 后移回归法

①时间升序（因变量向下移动一位）

②时间降序（因变量向上移动一位）

【注】

在写代码之前要先判断用户选择的数据是否能进行处理，例如用户如果只选择一行，就显示数据选择太少，无法进行滑移回归分析。因为如果只有1行数据，移动之后就没了。这里至少2行数据。

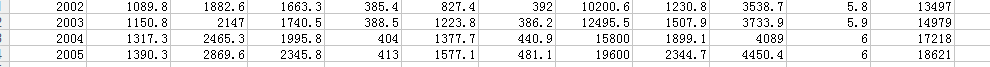
【测试数据：预测性回归分析.xls】

【注】

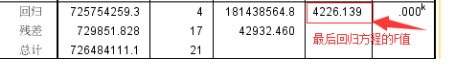
**这边我年份的值是默认全选的**，有时候，用户可能不按照顺序进行选择，例如用户只选择了1983、1987、2004的这样的数据，我们还是把它当做行的数据取得进行分析。

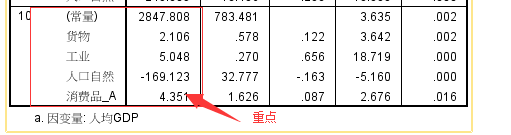
利用前移回归法预测2006年的人均GDP（时间升序）：

（人均GDP为因变量y ,除了年份外都是自变量。F引入=4.747 F剔除=4.667）



结果是“逐步回归”文档中的例子“逐步回归3.xls”的结果。





最后的回归方程是：

人均GDP= 2847.808+ 5.048\*工业 + 4.351\*消费品\_A + 2.106\*货物 -169.123\*人口自然

F值为4226.139 （我们把系数从大到小，方便用户知道哪个变量的权重占比较大。）

**预测第2006年的人均GDP为：**

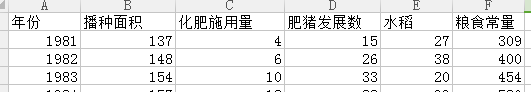
2847.808+5.049\*2869.6 + 4.351\*413 + 2.106\*1577.1 - 169.123\*6 = **21440.016**

所以我们也要返回给用户最后预测的2006年的人均GDP为21440.016

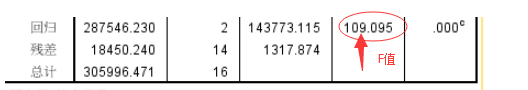
【测试数据：逐步回归2.xls】

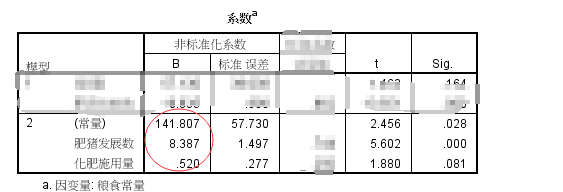
利用后移回归法推算出1980年的粮食常量（时间是升序）：

（粮食常量为因变量y，除了年份外的是自变量。F引入为2.5，F剔除为2.49）



结果：





后移回归法之后，得到的最终回归方程为：

粮食产量=141.807+8.387\*肥猪发展数+0.520\*化肥施用量

则**1980年的粮食产量的推测值**为：141.807+8.387\*15+0.520\*4 = **269.692**

所以我们也要返回给用户最后推测的1980年的粮食产量为269.692

多因素趋势回归

1. 说明及与多因素滑移回归的异同

所谓的多因素趋势回归，和多因素滑移回归是一样的，区别在于多因素滑移回归一次只移动1位，而多因素趋势回归使用移动1位、移动2位、移动3位后得到的预测值，如Z1，Z2，Z3按照一定的比例（默认5:3:2）得出最后的趋势回归结果：Y=Z1/10\*5+Z2/10\*3+Z3/10\*2。

也要考虑时间是升序还是降序，已经判断数据是否大于3个，否则无法进行多因素回归分析。

多因素趋势回归，也分前移（升序、降序）和后移（升序、降序），这4种~

多因素趋势回归，对于前移，还是预测下一年的值，如果数据是从1983-2008，利用前移，我们就是要预测出2009的值，移动一位则是用2008的值来预测2009年；移动两位就是用2007的值来预测2009的值；移动三位就是用2006年的值来预测2009的值。

如果是后移，数据从1983-2008，我们要推算1982的值，移动一位，用1983年推测1982，移动两位，用1984年来推测1982，移动三位，用1985年来推测1982年。

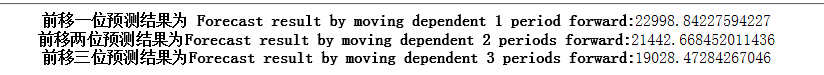
再根据权重（5:3:2）综合得出2009年的预测值，或1982年的推测值。

1. 结果显示

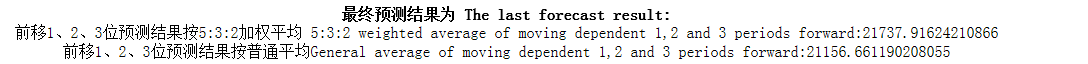
①显示移动1位的系数结果(有t值的那个）、R方结果、anova（有F值的那个）的值，以及回归方程，预测的值。

②显示移动2位的系数结果(有t值的那个）、R方结果、anova（有F值的那个）的值，以及回归方程，预测的值。

③显示移动3位的系数结果(有t值的那个）、R方结果、anova（有F值的那个）的值，以及回归方程，预测的值。



④然后显示按比例5:3:2得出的预测结果和按照普通比例的预测结果（除以3）



1. 测试数据

以“预测型回归分析.xls”（这里时间是升序的，人均GDP为因变量y ,除了年份外都是自变量。F引入=4.747 F剔除=4.667，预测2006年的人均GDP）为数据进行分析：

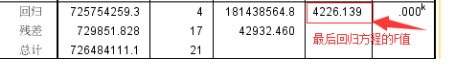


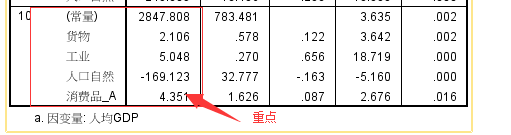
……



①前移动一位：

与“前移回归”得到的结果相同：





**前移一位回归方程是**：

人均GDP= 2847.808+ 5.048\*工业 + 4.351\*消费品\_A + 2.106\*货物 -169.123\*人口自然

F值为4226.139 （我们把系数从大到小，方便用户知道哪个变量的权重占比较大。）

**前移一位预测结果为：**

（用2005年的值代入上述方程）

2847.808+5.049\*2869.6 + 4.351\*413 + 2.106\*1577.1 - 169.123\*6 = **21440.016**

②前移动两位：





**前移两位回归方程是：**

人均GDP=-2590.570+26.174\*旅客+7.214\*固定资

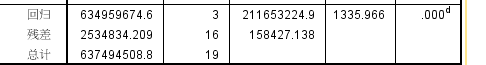
**前移两位预测结果为：**

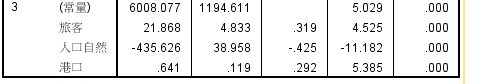
（用2004年的值代入上述方程）



-2590.570+26.174\*440.9 +7.214\*1899.1= **22649.654**

③前移动三位：





**前移三位回归方程是：**

人均GDP=6008.077 +21.868\*旅客+0.641\*港口-435.626\*人口自然

**前移三位预测结果为：**

（用2003年的值代入上述方程）



6008.077 +21.868\*386.2+0.641\*12495.5-435.626\*5.9= **19892.9207**

④按照（5:3:2）的权重得到最后的2006年的预测结果：

**21440.016/10\*5 + 22649.654/10\*3 + 19892.9207/10\*2 = 21493.4883**

⑤按照普通（1:1:1）的权重得到最后的2006年的预测结果：

**（21440.016+22649.654+19892.9207）/3 = 21327.5302**

后移（略）