



多元非线性回归

《美国数学建模竞赛》

完整课程请长按下方二维码





目录页

目录 CONTENTS

- 1 多元非线性回归
- 2 逐步回归
- 3 标准化回归





多元非线性回归





1.4 多元非线性回归

建立多元非线性回归方程在科学研究中应用广泛，其重要方法是将非线性回归方程转化为线性回归方程。转化时应首先选择适合的非线性回归形式，并将其线性化。再确定线性化回归方程的系数，最后确定非线性回归方程中未知的系数或参数。



1.4 多元非线性回归

实例：湖北省油菜投入与产出的统计分析

1.投入指标

(1) 土地 (S)。土地用播种面积来表示。农作物播种面积是指当年从事农业

(2) 劳动 (L)。劳动用劳动用工数 (成年劳动力一人劳动一天为一个工) 来表示。劳动用工中包含着直接和间接生产用工。

(3) 资本 (K)。资本用物质费用来表示。物质费用包含直接费用和间接费用。主要有种子秧苗费、农家肥费、化肥费、农药费、畜力、固定资产折旧费和管理及其他费用等。

2.产出指标

产出指标用湖北省历年油菜生产的总产量(Y)来表示。



1.4 多元非线性回归

年份	产量 (万吨) Y	物质费用 (万元) K	播种面积 (万亩) S	劳动用工 (万个) L	年份序号 t
1990	70.8972	40076.5884	825.1305	15347.4273	1
1991	83.7506	48008.7690	915.1500	15832.0950	2

$$Y = A_0 e^{\lambda t} K^{\alpha} L^{\beta} S^{\gamma}$$

$$\ln Y = \ln A_0 + \lambda t + \alpha \ln K + \beta \ln L + \gamma \ln S + \mu$$



1.4 多元非线性回归

```
data ex;input y k s l t @@;
```

```
x1=log(k);x2=log(s);x3=log(l);y1=(y);
```

```
cards;
```

70.8972	40076.5884	825.1305	15347.4273	1
83.7506	48008.7690	915.1500	15832.0950	2
70.8627	44593.8425	801.6150	13306.8090	3
78.3451	43460.3229	783.2100	13314.5700	4
98.0749	72657.2633	923.8050	14596.1190	5
134.8767	146108.3421	1282.8900	20911.1070	7
147.5315	162433.3500	1244.7000	18670.5000	8
154.7607	166979.6325	1330.5150	18627.2100	9
159.9743	190395.5262	1505.4600	20775.3480	10
198.4942	205914.6645	1738.4100	22599.3300	11
194.7943	189762.7335	1677.0900	20963.6250	12
187.1013	193461.5610	1761.9450	21936.2153	14
235.1184	183768.4035	1779.1500	19606.2330	15

```
;
```

```
proc reg;model y1=x1 x2 x3 t ; /*selection=stepwise*/
```

```
run;
```



1.4 多元非线性回归

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2.15231	0.53808	148.95	<.0001
Error	8	0.02890	0.00361		
Corrected Total	12	2.18121			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	0.93016	1.91920	0.48	0.6409
x1	1	0.24781	0.09610	2.58	0.0327
x2	1	1.28223	0.57122	2.24	0.0550
x3	1	-0.82102	0.55591	-1.48	0.1780
t	1	-0.00168	0.02437	-0.07	0.9466

变量t 的显著性概率为0.9466，远大于0.05，因此将
model y1=x1 x2 x3 t;去掉他，即改为**model y1=x1 x2 x3**
;



1.4 多元非线性回归

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2.15229	0.71743	223.29	<.0001
Error	9	0.02892	0.00321		
Corrected Total	12	2.18121			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	0.87950	1.67253	0.53	0.6117
x1	1	0.24554	0.08518	2.88	0.0181
x2	1	1.24568	0.20239	6.15	0.0002
x3	1	-0.78798	0.26689	-2.95	0.0162

截距项Intercept 的显著性概率为0.6117，大于0.05，因此将model $y_1 = x_1 x_2 x_3$; 改为**model $y_1 = x_1 x_2 x_3 / \text{noint}$** ;



1.4 多元非线性回归

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	309.07415	103.02472	34565.8	<.0001
Error	10	0.02981	0.00298		
Uncorrected Total	13	309.10395			

Root MSE	0.05459	R-Square	0.9999
Dependent Mean	4.85895	Adj R-Sq	0.9999
Coeff Var	1.12358		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
x1	1	0.22851	0.07588	3.01	0.0131
x2	1	1.21016	0.18376	6.59	<.0001
x3	1	-0.65225	0.06539	-9.98	<.0001



1.4 多元非线性回归

$$\ln \hat{Y} = 0.244189 \ln K + 1.172185 \ln S - 0.643284 \ln L$$

$$F=34565.8 \quad R^2=0.9999$$

K, S, L的t值分别为 (3.01) (6.59) (-9.98)

$$\hat{Y} = K^{0.22851} S^{1.21016} L^{-0.65225}$$

要善于解释经济含义、本模型虽然满足数学规则，但不能通过经济检验。助于如何继续修正模型，需要学习数学与经济的交叉学科《计量经济学》。



逐步回归





1.5 逐步回归

逐步回归的**基本思想**是，从当前在圈外的全部变量中，挑选其偏回归平方和贡献最大的变量，用方差比进行显著性检验的办法，判别是否选入；而当前在圈内的全部变量中，寻找偏回归平方和贡献最小的变量，用方差比进行显著性检验的办法，判别是否从回归方程中剔除。选入和剔除循环反复进行，直至圈外无符合条件的选入项，圈内无符合条件的剔除项为止。

逐步回归选择变量快捷，但对于存在多重共线的自变量选择，有时并不准确，使用时注意分辨。

还是用上面的例子，将`model y1=x1 x2 x3 t`；改为**`model y1=x1 x2 x3 t /selection=stepwise;`**



1.5 逐步回归

注意，为了筛选变量宽容，程序中默认显著度为0.15，而不是0.05，以避免条件过于严格只用筛选无法进行。

All variables left in the model are significant at the 0.1500 level.

No other variable met the 0.1500 significance level for entry into the model.

Summary of Stepwise Selection

Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x2		1	0.9668	0.9668	11.0349	320.51	<.0001
2	t		2	0.0086	0.9754	7.8417	3.50	0.0909
3	x1		3	0.0077	0.9831	5.1812	4.12	0.0730

从程序结果中不难看出，x2、x1、t进入模型。
因此model y1=x1 x2 x3 t /selection=stepwise;
改为model y1=x1 x2 t /noint;再运行一遍即可。



1.5 逐步回归

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	309.06293	103.02098	25111.1	<.0001
Error	10	0.04103	0.00410		
Uncorrected Total	13	309.10395			
Root MSE					
		0.06405	R-Square	0.9999	
Dependent Mean		4.85895	Adj R-Sq	0.9998	
Coeff Var		1.31822			
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
x1	1	0.21081	0.09041	2.33	0.0419
x2	1	0.29433	0.14500	2.03	0.0698
t	1	0.04175	0.00501	8.34	<.0001

思考：为什么这个结果与前面计算的结果不一样？



标准化回归





1.6 标准化回归

由于单位量纲不一样，偏回归系数的大小不能完全反映自变量对因变量影响的大小。要想真实反映自变量的贡献，标准化回归是个好的选择。

标准化回归系数（Beta值）在多元回归中被用来比较变量间的重要性。但注意这种重要性是相对的，请同学们百度一下，看看相关资料。



1.6 标准化回归

还是用上面的例子，将model y1=x1 x2 t /noint;改为model y1=x1 x2 t /noint stb;运行即可。

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
x1	1	0.21081	0.09041	2.33	0.0419	0.50200
x2	1	0.29433	0.14500	2.03	0.0698	0.42920
t	1	0.04175	0.00501	8.34	<.0001	0.07640

从最后一列（标准化回归系数）可看出，x1重要性超过x2和t。与原先的参数大小比有变化。