·科研方法专题·

加性与广义加性模型回归分析

胡良平[1,2*]

(1. 军事科学院研究生院,北京 100850;
2. 世界中医药学会联合会临床科研统计学专业委员会,北京 100029
* 通信作者:胡良平,E-mail:lphu812@sina.com)

【摘要】 本文目的是介绍加性与广义加性模型回归分析的概念、作用以及用软件实现计算的方法。先介绍有关的基本概念,再介绍基本原理,最后通过一个实例并基于 SAS 软件演示如何实施加性模型回归分析。结果表明:相对于引入派生变量进行常规多重线性回归分析而言,加性与广义加性模型回归分析能够更好地提升模型对资料的拟合效果。

【关键词】 加性模型;广义加性模型;样条函数;光滑分量

中图分类号:R195.1 文献标识码:A doi:10.11886/j.issn.1007-3256.2018.04.001

Regression analysis based on the additive model and generalized additive model

Hu Liangping[1,2*]

(1. *Graduate School, Academy of Military Sciences PLA China, Beijing* 100850, *China*;
2. *Specialty Committee of Clinical Scientific Research Statistics of World Federation of Chinese Medicine Societies, Beijing* 100029, *China*
* *Corresponding author: Hu Liangping, E-mail: lphu812@sina.com*)

【Abstract】 The purpose of this paper was to introduce the concepts, functions and the calculation methods by using the statistical software of the regression analysis based on the additive model and generalized additive model. Firstly, the basic concepts of the regression analysis were introduced. Secondly, the basic principles of the regression analysis were given. Finally, the regression analysis based on the additive model was demonstrated through one example by using the SAS software. The results showed that the additive model regression analysis could improve the fitted effect greatly than the result acquired by using derived variables in conventional multiple linear regression analysis.

【Keywords】 Additive model; Generalized additive model; Spline function; Smoothing component

1 概 述[1-2]

1.1 加性模型

将多重线性回归模型进行推广,使其表达式成为下面的式(1)形式:

$$Y = s_0 + s_1(X_1) + s_2(X_2) + \cdots + s_p(X_p) + \varepsilon \quad (1)$$

在式(1)中,$s_j(X_j)$,$j=1,2,\cdots,p$,是 P 个"光滑函数";误差"ε"满足如下条件:它的期望为 0 [$E(\varepsilon)=0$]、方差为 σ^2 [$\mathrm{Var}(\varepsilon)=\sigma^2$]。为了使式(1)成为可以估计的,要求光滑函数 $s_i(X_i)$ 必须满足如下的标准化条件:即期望为 0,$E[s_j(X_j)]=0$。式(1)中的 P 个光滑函数不以参数形式呈现,而以非参数形式呈现。

1.2 广义加性模型

在加性模型式(1)中,假定因变量 y 服从正态分布。然而,在很多场合下,因变量不服从正态分布,而可能服从其他某种分布。现假定式(1)中的因变量 y 具有下面的指数族分布密度,见式(2):

$$f_Y(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\} \quad (2)$$

在式(2)中,θ 被称为"自然参数",ϕ 被称为"尺度参数";以因变量 y 的平均值 μ 为自变量构造出的函数 g(μ) 被称为"连接函数",它与协变量"X_1、X_2、$\cdots$、X_p"之间建立了关系。下面的数量定义了"加性分量(或成分)",见式(3):

$$\eta = s_0 + \sum_{j=1}^{p} s_j(X_j) \quad (3)$$

式(3)中,$S_1()$、$S_2()$、$\cdots$、$S_p()$ 都是"光滑函数"。μ 与 η 之间的关系由下式来定义:

$$g(\mu) = \eta \quad (4)$$

最常用的连接函数为"典型连接函数",即 $\eta=\theta$。

1.3 加性与广义加性模型回归分析应用的场合

广义线性模型强调对模型中参数的估计和推断,而广义加性模型则聚焦于如何用非参数法探测

项目基金:国家高技术研究发展计划课题资助(2015AA020102)

数据。换言之，广义加性模型更适合于探查数据并可视化因变量与自变量之间的关系。

1.4 加性与广义加性模型回归分析的计算原理

1.4.1 加性模型回归分析的计算原理

基于加性模型式(1)，可以构造如下的残差，见式(5)：

$$R_k = Y - s_0 - \sum_{j \neq k} s_j(X_j) \tag{5}$$

式(5)被称为"第 k 项光滑参数与因变量 y"之间的"残差"，即 $R_k \approx s_k(X_k)$。严格地说，应该有下式成立，见式(6)：

$$E(R_k | X_k) = s_k(X_k) \tag{6}$$

由式(5)可知：对于所有其他的项"j"($j \neq k$)，在给定 $\{\hat{s}_j(),j \neq k\}$"的估计值时，其观测值提供了一种用于估计每个光滑函数"$s_k()$"的方法。依此做法求得结果的迭代过程被称为"后退拟合算法"，此法最早由 Friedman 和 Stuetzle 提出。

1.4.2 后退拟合算法

1.4.2.1 未加权的后退拟合算法

未加权的后退拟合算法步骤如下：

第 1 步：初始化。

$$s_0 = E(Y), s_1^{(1)} = s_2^{(1)} = \cdots = s_p^1 = 0, m = 0$$

第 2 步：迭代。令 $m = m + 1$；让 j 从 1 到 p 循环，循环体内的计算公式为：

$$R_j = Y - s_0 - \sum_{k=1}^{j-1} s_k^{(m)}(X_k) - \sum_{k=j+1}^{p} s_k^{(m-1)}(X_k); s_j^{(m)} = E(R_j | X_j);$$

第 3 步：终止。

$$RSS = \frac{1}{n} \| Y - s_0 - \sum_{j=1}^{p} s_j^{(m)}(X_j) \|^2$$

直到上式的计算结果不再下降，或满足收敛的临界值，就停止迭代。

在前面的表达式中，$s_j^{(m)}()$ 代表在第 m 次迭代时 $s_j()$ 的估计值。可以看到：对于许多"光滑器"(包括线性回归、单变量或双变量样条以及它们的组合)而言，在任何一步中，RSS 从来都不会增加。这意味着上述算法总是会收敛的。

值得注意的是：对于除正态分布之外的其他分布而言，具有权重的数值不稳定可能会引起收敛问题。即使当算法收敛时，各个个体函数并不需要彼此完全不同，即便对于同一个拟合曲面来说，由于协变量之间的依赖性会导致多于一个表达式出现。

1.4.2.2 加权的后退拟合算法

除了要对光滑器进行加权之外，加权的后退拟合算法具有与未加权的后退拟合算法相同的形式。在 SAS 的 GAM 过程步中，具体是在采用"局部计分过程"中且资料为非正态分布时，使用了"权重"。

GAM 过程使用下面的"条件"作为后退拟合算法的"收敛临界值"：

$$\frac{\sum_{i=1}^{n} \sum_{j=1}^{p} [s_j^{(m-1)}(X_{ij}) - s_j^{(m)}(X_{ij})]^2}{1 + \sum_{i=1}^{n} \sum_{j=1}^{p} [s_j^{(m-1)}(X_{ij})]^2} \leqslant \varepsilon$$

此处，$\varepsilon = 10^{-8}$ 是缺省的界值。用户可以通过修改模型语句中的选项"EPSILON ="来改变此界值。

1.4.3 广义加性模型回归分析的计算原理

广义加性模型回归分析的计算原理比前面介绍的加性模型回归分析的计算原理稍复杂，其中，最关键的内容为"局部计分算法"。该算法的重要内容取决于与每个特定分布对应的"连接函数"。它们之间的关系见表 1。

表 1 局部计分算法涉及到的重要内容

Distribution	Link	Adjusted Dependent (z)	Weights (w)
Normal	μ	y	1
Binomial	$\log\left(\frac{\mu}{n-\mu}\right)$	$\eta + (y-\mu)/n\mu(1-\mu)$	$n\mu(1-\mu)$
Gamma	$-1/\mu$	$\eta + (y-\mu)/\mu^2$	μ^2
Poisson	$\log(\mu)$	$\eta + (y-u)/\mu$	μ
Inverse Gaussian	$1/\mu^2$	$\mu - 2(y-\mu)/\mu^3$	$\mu^3/4$

注：表 1 摘录自 SAS 软件 GAM 过程帮助信息

由表 1 可知：一旦分布被指定，相应的"那些量"也就被定义了。于是，可按下面的步骤实施"局部计分算法"。

1.4.4 一般的局部计分算法

第 1 步:初始化。

$s_i = g[E(y)], s_1^0 = s_2^0 = \cdots = s_p^0 = 0, m = 0$

第 2 步:迭代。令 $m = m+1$;从前一次迭代中获得各变量的相应数值,这些变量分别是:预测量 η、均值 μ、权重 w 和校正后的因变量 z:

$$\eta_i^{(m-1)} = s_0 + \sum_{j=1}^{p} s_j^{(m-1)}(x_{ij})$$

$$\mu_i^{(m-1)} = g^{-1}(\eta_i^{(m-1)})$$

$$w_i = [V_i^{(m-1)}]^{-1} \cdot \left[\left(\frac{\partial\mu}{\partial\eta}\right)_i^{(m-1)}\right]^2$$

$$z_i = \eta_i^{(m-1)} + [y_i - \mu_i^{(m-1)}] \cdot \left(\frac{\partial\mu}{\partial\eta}\right)_i^{(m-1)}$$

此处,$V_i^{(m-1)}$ 是在 $\mu_i^{(m-1)}$ 处 Y 的方差。通过使用带有权重 w 的后退拟合算法,对 z 拟合一个加性模型,从而获得被估计的函数 $s_j^{(m)}(\ ), j=1,\cdots,p$。

第 3 步:终止。直到达到收敛临界值或离差不再减少时,停止迭代。这里所说的"离差",实际上是广义线性模型中"*RSS*"的一个扩展或推广。

GAM 过程使用下面的"条件"作为局部计分算法的"收敛临界值":

$$\frac{\sum_{i=1}^{n} w_i \sum_{j=1}^{p} [s_j^{(m-1)}(X_{ij}) - s_j^{(m)}(X_{ij})]^2}{\sum_{i=1}^{n} w_i \{1 + \sum_{j=1}^{p} [s_j^{(m-1)}(X_{ij})]^2\}} \leq \varepsilon^3$$

此处,$\varepsilon = 10^{-8}$ 是缺省的界值。用户可以通过修改模型语句中的选项"EPSILON ="来改变此界值。

算法小结:广义加性模型的估计过程由两个循环构成。在局部计分算法(外循环)每一步内部,使用加权后退拟合算法(内循环),直到收敛或 *RSS* 不再减少。然后,基于来自这个加权后退拟合算法得到的估计量,计算出一组新的权重,开始计分算法的下轮迭代。当达到收敛临界值或估计量的离差停止下降时,计分算法也就停止了。

2 基于加性模型回归分析解决实际问题[2]

2.1 问题与数据结构

【例 1】下面是一个假设的来自化学试验的例子:每次试验,研究者将某种催化剂加入到某种化学溶液中,从而合成一种新化合物。其数据是测量溶液的温度(temperature)、加入的催化剂量(catalyst)和化学反应量(yield)的结果。试验数据的结构很简单,两个计量原因变量和一个计量结果变量及其取值,前 6 次试验数据见表 2。

表 2 不同"溶液温度"和"催化剂量"条件下化学反应量的测定结果

试验号	溶液的温度	催化剂量	化学反应量
1	80	0.0015	6.039
2	80	0.0100	4.719
3	80	0.0150	6.301
4	80	0.0200	4.558
5	80	0.0250	5.917
6	80	0.0300	4.365

注:表中仅列出了少量数据,详细试验数据见下面的 SAS 程序,此处从略

【对数据结构的分析】对于每次试验而言,可以观测到 3 个计量的数据,即溶液的温度(temperature)、加入的催化剂量(catalyst)和化学反应的量(yield)。

【统计分析方法的选择】若希望考察化学反应的量(yield)是如何随溶液的温度(temperature)和催化剂量(catalyst)变化而变化的依赖关系,可选择多重线性回归分析。因本例中的因变量为计量变量,故可以考虑选用"加性模型回归分析"。若因变量是定性变量或计数变量,就可能需要选用"广义加性模型回归分析"。因篇幅所限,本文只介绍如何用 SAS 实现"加性模型回归分析"。

2.2 基于常规方法构建多重线性回归模型[3-4]

2.2.1 创建 SAS 数据集

创建一个名为"ExperimentA"的临时 SAS 数据集所需的数据步程序:

```
data ExperimentA;
    format Temperature f4.0 Catalyst f6.3 Yield f8.3;
    input Temperature Catalyst Yield @@;
    x1 = temperature; x2 = Catalyst; y = Yield;
datalines;
80 0.005 6.039 80 0.010 4.719 80 0.015 6.301
80 0.020 4.558 80 0.025 5.917 80 0.030 4.365
80 0.035 6.540 80 0.040 5.063 80 0.045 4.668
80 0.050 7.641 80 0.055 6.736 80 0.060 7.255
80 0.065 5.515 80 0.070 5.260 80 0.075 4.813
80 0.080 4.465 90 0.005 4.540 90 0.010 3.553
90 0.015 5.611 90 0.020 4.586 90 0.025 6.503
90 0.030 4.671 90 0.035 4.919 90 0.040 6.536
```

```
90 0.045 4.799 90 0.050 6.002 90 0.055 6.988
90 0.060 6.206 90 0.065 5.193 90 0.070 5.783
90 0.075 6.482 90 0.080 5.222 100 0.005 5.042
100 0.010 5.551 100 0.015 4.804 100 0.020 5.313
100 0.025 4.957 100 0.030 6.177 100 0.035 5.433
100 0.040 6.139 100 0.045 6.217 100 0.050 6.498
100 0.055 7.037 100 0.060 5.589 100 0.065 5.593
100 0.070 7.438 100 0.075 4.794 100 0.080 3.692
110 0.005 6.005 110 0.010 5.493 110 0.015 5.107
110 0.020 5.511 110 0.025 5.692 110 0.030 5.969
110 0.035 6.244 110 0.040 7.364 110 0.045 6.412
110 0.050 6.928 110 0.055 6.814 110 0.060 8.071
110 0.065 6.038 110 0.070 6.295 110 0.075 4.308
110 0.080 7.020 120 0.005 5.409 120 0.010 7.009
120 0.015 6.160 120 0.020 7.408 120 0.025 7.123
120 0.030 7.009 120 0.035 7.708 120 0.040 5.278
120 0.045 8.111 120 0.050 8.547 120 0.055 8.279
120 0.060 8.736 120 0.065 6.988 120 0.070 6.283
120 0.075 7.367 120 0.080 6.579 130 0.005 7.629
130 0.010 7.171 130 0.015 5.997 130 0.020 6.587
130 0.025 7.335 130 0.030 7.209 130 0.035 8.259
130 0.040 6.530 130 0.045 8.400 130 0.050 7.218
130 0.055 9.167 130 0.060 9.082 130 0.065 7.680
130 0.070 7.139 130 0.075 7.275 130 0.080 7.544
140 0.005 4.860 140 0.010 5.932 140 0.015 3.685
140 0.020 5.581 140 0.025 4.935 140 0.030 5.197
140 0.035 5.559 140 0.040 4.836 140 0.045 5.795
140 0.050 5.524 140 0.055 7.736 140 0.060 5.628
140 0.065 6.644 140 0.070 3.785 140 0.075 4.853
140 0.080 6.006
;
run;
```

【SAS 程序说明】数据中每行上有 3 次试验的结果，每次试验结果都有 3 个数据，即温度数值(temperature)、催化剂量(catalyst)与产量(yield)。

创建一个名为“ExperimentB”的临时 SAS 数据集的 SAS 数据步程序：

```
data ExperimentB;
    set ExperimentA;
x3 = x1* x1; x4 = x2* x2; x5 = x1* x2; x6 = x3* x1;
x7 = x4* x2; x8 = x3* x2; x9 = x4* x1;
run;
```

【SAS 程序说明】以上 SAS 程序产生 7 个“派生变量”，它们分别为 x1 与 x2 两个原始自变量的平方项、立方项、交叉乘积项，具体地说，$x_3 = x_1^2$、$x_4 = x_2^2$、$x_5 = x_1 \times x_2$、$x_6 = x_1^3$、$x_7 = x_2^3$、$x_8 = x_1^2 \times x^2$、$x_9 = x_2^2 \times x_1$。**其中，由前面的 SAS 程序可知：** x_1 = temperature、x_2 = Catalyst、y = Yield。

2.2.2 基于常规方法构建多重线性回归模型

利用下面的两个 SAS 过程步程序可以创建两个二重线性回归模型：

```
proc reg data = ExperimentA;
    model y = x1 x2/r;
run;
```

记以上 SAS 程序创建的二重线性回归模型为模型(1)。

```
proc reg data = ExperimentA;
    model y = x1 x2/noint r;
run;
```

记以上 SAS 程序创建的二重线性回归模型为模型(2)。

经比较，模型(1)优于模型(2)。具体方法详见下文，此处从略。

引入自变量的“二次项”，利用下面的两个 SAS 过程步程序可以创建两个多重线性回归模型：

```
proc reg data = ExperimentB;
    model y = x1 - x5/selection = backward sle =
0.05 r;
run;
```

记以上 SAS 程序创建的多重线性回归模型为模型(3)。

```
proc reg data = ExperimentB;
    model y = x1 - x5/noint selection = backward sle
 =0.05 r;
run;
```

记以上 SAS 程序创建的多重线性回归模型为模型(4)。

经比较，模型(4)优于模型(3)。具体方法详见下文，此处从略。

引入自变量的“三次项”，利用下面的两个 SAS 过程步程序可以创建两个多重线性回归模型：

```
proc reg data = ExperimentB;
    model y = x1 - x9/selection = backward sle
 =0.05 r;
run;
```

记以上 SAS 程序创建的多重线性回归模型为模型(5)。

```
proc reg data = ExperimentB;
    model y = x1 - x9/noint selection = backward sle
```

=0.05 r;
run;

记以上SAS程序创建的多重线性回归模型为模型(6)。

经比较,模型(5)优于模型(6)。具体方法详见下文,此处从略。

将模型(4)与模型(1)比较,得出模型(4)优于模型(1)。最后,需要将模型(5)与模型(4)作比较,具体方法如下:

模型(4)的有关信息为: $SS_\varepsilon = 128.48055$(模型误差的离均差平方和)、$df_\varepsilon = 108$(误差的自由度);

模型(5)的有关信息为: $SS_\varepsilon = 78.07028$(模型误差的离均差平方和)、$df_\varepsilon = 106$(误差的自由度)。

利用下面的 F 检验对上述回归模型(5)与模型(4)进行拟合优度比较:

$$F = \frac{(128.48055 - 78.07028)/(108 - 106)}{78.07028/106} = 34.222$$

对应的 F 临界值 $F_{((2,106)(0.01))} < 4.82$,因 $F = 34.222 > 4.82$,说明 $P < 0.01$,故需要选择参数多的回归模型(5)。模型(5)的输出结果如下:

方差分析

源	自由度	平方和	均方	F值	Pr > F
模型	5	88.66611	17.73322	24.08	<.0001
误差	106	78.07028	0.73651		
校正合计	111	166.73639			

均方根误差	0.85820	R方	0.5318
因变量均值	6.14077	调整R方	0.5097
变异系数	13.97550		

参数估计值

变量	自由度	参数估计值	标准误差	t值	Pr > |t|
Intercept	1	122.10320	18.51622	6.59	<.0001
x1	1	-3.45961	0.52226	-6.62	<.0001
x3	1	0.03322	0.00482	6.89	<.0001
x4	1	1437.83096	246.93235	5.82	<.0001
x6	1	-0.00010355	0.00001460	-7.09	<.0001
x7	1	-17556.00000	3130.36119	-5.61	<.0001

根据最后的"参数估计值",请读者写出相应的"五重线性回归模型"的表达式,此处从略。

2.3 基于加性模型构建多重回归模型[2]

利用下面的SAS程序可基于加性模型构建多重回归模型:

```
proc gam data=ExperimentA;
model y=spline(x1) spline(x2);
output out=a3 residual;
run;
```

【SAS程序说明】以上SAS程序调用GAM过程拟合加性模型。模型语句等号右边的两项分别用"三次样条函数"拟合自变量x1与x2。

【SAS输出结果及其解释】

因变量: y

平滑模型成分: spline(x1) spline(x2)

输入数据集的汇总

观测数	112
缺失观测数	0
分布	Gaussian
关联函数	Identity

以上是关于数据集一般情况的描述,并告知:假定因变量y服从正态分布(或高斯分布)、采用恒等的关联函数,实际上,就是没有对y作任何变量变换。

迭代汇总和拟合统计量

最终 Backfitting 迭代次数	2
最终 Backfitting 准则	2.673726E-14
最终估计的偏差	68.464845603
Backfitting 算法收敛	

以上是关于"迭代汇总和拟合统计量"的信息，关键是倒数第二行：最终估计的偏差为 68.464846，此值相当于通常回归分析给出的"模型误差的离差平方法和"。

回归模型分析

参数估计值

参数	参数估计值	标准误差	t 值	Pr > |t|
Intercept	3.77618	0.45348	8.33	<.0001
Linear(x1)	0.01765	0.00385	4.58	<.0001
Linear(x2)	9.95091	3.34239	2.98	0.0036

以上给出的是"加性模型"中"参数分析部分"的结果，即

$\hat{y}_1 = 3.77618 + 0.01765x_1 + 9.95091x_2$ 模型(7)

平滑模型分析

平滑成分的拟合汇总

成分	平滑参数	自由度	GCV	唯一观测数
Spline(x1)	0.161745	3.000000	0.304733	7
Spline(x2)	0.884237	3.000000	0.153507	16

以上给出的是"加性模型"中"非参数分析部分"的结果，即

$\hat{y}_2 = 0.161745\text{Spline}(x_1) + 0.884237\text{Spline}(x_2)$ 模型(8)

将模型(7)与模型(8)合并成一个模型，见模型(9)：

$\hat{y} = \hat{y}_1 + \hat{y}_2 = 3.77618 + 0.01765x_1 + 9.95091x_2 + 0.161745\text{Spline}(x_1) + 0.884237\text{Spline}(x_2)$ 模型(9)

平滑模型分析

偏差分析

源	自由度	平方和	卡方	Pr > 卡方
Spline(x1)	3.00000	53.010023	79.7494	<.0001
Spline(x2)	3.00000	25.411103	38.2290	<.0001

以上是关于加性模型中两个非参数项(即样条函数)的假设检验结果，两项各占用了 3 个自由度，经卡方检验，说明两个非参数项都具有统计学意义。

图 1 左边的曲线描述的是模型(8)中的第 1 项；图 1 右边的曲线描述的是模型(8)中的第 2 项。其中，第 1 项比第 2 项更复杂。

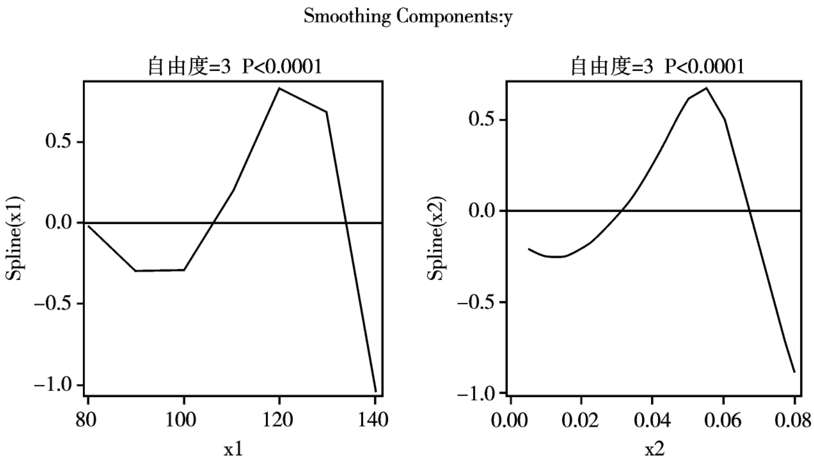

图 1 两个非参数项分别与 x1、x2 之间的函数曲线

2.4 两类回归模型拟合效果比较

常规多重线性回归模型与加性多重回归模型对同一个资料究竟谁的拟合效果更好？这个问题尚无公认的评判方法，但可以近似地采用下面的方法进行比较：

将常规多重线性回归模型中拟合效果最好的模型(5)与加性模型(9)进行比较，用类似于模型(5)与模型(4)比较的 F 检验：

已知：模型(9)的 $SS_\varepsilon = 68.464856$、$df_\varepsilon = 103$；模型(5)的 $SS_\varepsilon = 78.07028$、$df_\varepsilon = 106$。利用下面的 SAS 程序可以求出检验统计量 F 的数值以及对应的 F 临界值：

```
data abc;
```

```
v1 = (78.07028 - 68.464856) /(106 - 103);
v2 = 68.464856/103;
F = v1/v2;
F3_103 = FINV(0.95,3,103);
proc print data = abc;
    var F F3_103;
run;
```

【SAS 输出结果】

Obs	F	F3_103
1	4.81687	2.69284

因 $F = 4.817 > F_{(3,103)(0.95)} = 2.693$，所以，$P < 0.05$，说明不能用含参数个数少的模型(5)取代含参数个数多的模型(9)。

【结论】本例以加性模型的回归分析结果为优。

参考文献

[1] Armitage P, Colton T. Encyclopedia of biostatistics [M]. 2nd. John Wiley & Sons, 2005: 73, 2253 - 2257.

[2] SAS Institute Inc. STAT SAS 9.3 User´s Guide [M]. Cary, NC: SAS Institute Inc, 2011: 2549 - 2604.

[3] 胡良平. 多重线性回归分析的核心内容与关键技术[J]. 四川精神卫生, 2018, 31(1): 1 - 6.

[4] 谷恒明, 胡良平. 基于经典统计思想实现多重线性回归分析[J]. 四川精神卫生, 2018, 31(1): 7 - 11.

(收稿日期: 2018 - 08 - 10)
(本文编辑: 陈　霞)

科研方法专题策划人——胡良平教授简介

胡良平，男，1955 年 8 月出生，教授，博士生导师，曾任军事医学科学院研究生部医学统计学教研室主任和生物医学统计咨询中心主任、国际一般系统论研究会中国分会概率统计系统专业理事会常务理事和北京大学口腔医学院客座教授；现任世界中医药学会联合会临床科研统计学专业委员会会长、中国生物医学统计学会副会长，《中华医学杂志》等 10 余种杂志编委和国家食品药品监督管理局评审专家。主编统计学专著 48 部，参编统计学专著 10 部；发表第一作者学术论文260余篇，发表合作论文 130 余篇，获军队科技成果和省部级科技成果多项；参加并完成三项国家标准的撰写工作；参加三项国家科技重大专项课题研究工作。在从事统计学工作的 30 年中，为几千名研究生、医学科研人员、临床医生和杂志编辑讲授生物医学统计学，在全国各地作统计学学术报告 100 余场，举办数十期全国统计学培训班，培养多名统计学专业硕士和博士研究生。近几年来，参加国家级新药和医疗器械项目评审数十项、参加 100 多项全军重大重点课题的统计学检查工作。归纳并提炼出有利于透过现象看本质的“八性”和“八思维”的统计学思想，独创了逆向统计学教学法和三型理论。擅长于科研课题的研究设计、复杂科研资料的统计分析与 SAS 实现、各种层次的统计学教学培训和咨询工作。