第4卷第1期 上海水产大学学报 Vol. 4, No. 1
1995年3月 JOURNAL OF SHANGHAI FISHERIES UNIVERSITY March, 1995

研究简报

数据回归分析通用程序的研制

DEVELOPMENT ON GENERAL PROGRAM OF DATA REGRESSION ANALYSIS

楼文高

（上海水产大学工程技术学院，200090）

Lou Wen-gao

(*College of Engineering Technology, SFU*, 200090)

关键词 回归分析，通用程序，最佳回归方程

KEYWORDS regression analysis, general program, optimum regression equation

在科学实(试)验和研究中，为找出变量之间的变化规律以及化简繁复的理论公式，大量用到数据回归分析处理技术[刘国安等，1988；徐庆登等，1992；楼文高，1992；臧维玲等，1993；戴祥庆等，1988]。此时，一般可以根据理论、专业知识或实验数据的"散点图"确定待回归问题的数学模型。但是，在实际应用中，特别是对未知现象变化规律的探索研究时，回归现象的数学模型是未知的，必须通过比较各种模型的回归结果即相关系数、F 检验值和剩余标准离差等，确定经检验是显著的最佳回归方程。为此，作者研制了常用函数类型回归分析的通用程序，以满足科学研究的需要。

1 材料与方法

1.1 常用的函数类型

用于回归分析的常用函数类型有：

(1) 直线：$y = b_0 + b_1x_1 + \cdots\cdots + b_px_p$

(2) 双曲线：$1/y = b_0 + b_1/x$

(3) 抛物线：$y = b_0 + b_1\sqrt{x}$

(4) 幂函数：$y = b_0x_1^{b_1}x_2^{b_2}\cdots\cdots x_p^{b_p}$

(5) 指数函数 I：$y = e^{b_0+b_1x_1+\cdots\cdots+b_px_p}$

(6) 指数函数 II：$y = e^{b_0+b_1/x}$

(7) 对数函数：$y = b_0 + b_1^{\ln x_1} + \cdots\cdots + b_p^{\ln x_p}$

1994-11-20收到。

(8)S 型曲线：$y = 1/(b_0 + b_1e^{-x})$

(9) 多项式：$y = b_0 + b_1x + b_2x^2 + \cdots\cdots + b_px^p$　　$p(>1)$

对于函数(1)(4)(5)(7)P 是自变量个数，函数(9)中 P 为多项式最高次数。

1.2 回归分析理论

1.2.1 函数的线性化处理

上述函数类型经表1的线性化变换均可化为线性函数：

$$Y=B_0+B_1X_1+B_2X_2+\cdots\cdots+B_pX_p \qquad (1)$$

其中 P 为自变量个数或多项式最高次数。

1.2.2 线性回归分析

设有 P+1个相关的变量 $Y,X_1,X_2,\cdots\cdots,X_P$，则线性回归方程可表示为(1)式，其中 $B_0,B_1,B_2,\cdots\cdots,B_P$ 为待定的回归常数。

表1 常用函数的线性变换

Tab. 1 Linearization transformation of functions

函数类型	线性化变换关系式									
序　号	X_1	X_2	……	X_P	Y	B_0	B_1	B_2	……	B_P
1	x_1	x_2	……	x_p	y	b_0	b_1	b_2	……	b_p
2	$1/x$	—	……	—	$1/y$	b_0	b_1	—	……	—
3	$\sqrt{x}$	—	……	—	y	b_0	b_1	—	……	—
4	$\ln x_1$	$\ln x_2$	……	$\ln x_p$	$\ln y$	$\ln b_0$	b_1	b_2	……	b_p
5	x_1	x_2	……	x_p	$\ln y$	b_0	b_1	b_2	……	b_p
6	$1/x$	—	……	—	$\ln y$	b_0	b_1	—	……	—
7	$\ln x_1$	$\ln x_2$	……	$\ln x_p$	y	b_0	b_1	b_2	……	b_p
8	x	—	……	—	$1/y$	b_0	b_1	—	……	—
9	x	x^2	……	x^p	y	b_0	b_1	b_2	……	b_p

由样本数据(X_{1i}, X_{2i},……, X_{Pi}, Y_i)，可求得上述回归常数 $B_0,B_1,B_2,\cdots\cdots,B_P$，则对于 n 组样本数据，其估计值 $\hat{Y}_i$ 为：

$$\hat{Y}_i = B_0 + B_1X_{1i} + B_2X_{2i} + \cdots\cdots + B_PX_{Pi} \qquad (i = 1,2,\cdots\cdots,n) \qquad (2)$$

则根据最小二乘法原理即误差平方和最小得函数：

$$F = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n}[Y_i - (B_0 + B_1X_{1i} + B_2X_{2i} + \cdots\cdots + B_PX_{Pi})]^2 \to \min$$

由函数极值定理知，F 有极小值的条件为：

$$\partial F/\partial B_i = 0 \qquad (i = 0,1,2,\cdots\cdots,P) \qquad (3)$$

由(3)式得正规方程组：

$$AB=D \qquad (4)$$

其中 $A = X^TX$, $B = [B_0, B_1, B_2,\cdots\cdots, B_P]^T$, $D = X^TY$, $Y = [Y_1, Y_2\cdots\cdots Y_n]^T$,

$$X=\begin{bmatrix}1 & 1 & & 1\\ X_{11} & X_{12} & \cdots\cdots & X_{1n}\\ X_{21} & X_{22} & \cdots\cdots & X_{2n}\\ & \cdots\cdots\cdots & & \\ X_{P1} & X_{P2} & \cdots\cdots & X_{Pn}\end{bmatrix}^{T}$$

解方程组(4)得 $B_i(i=0,1,2,\cdots\cdots p)$,再利用线性化变换的逆变换可求得 b_i,即得到样本数据的回归预测和控制方程。

1.2.3 回归方程的相关系数 R、F 检验值和剩余标准离差 S

相关系数 R、F 检验值和剩余标准离差 S 为:

$$R=\sqrt{1-Q/L_{yy}} \tag{5}$$

$$F=U(n-p-1)/(L_{yy}-U)/P \tag{6}$$

$$S=\sqrt{Q/(n-p-1)} \tag{7}$$

其中 $L_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2$, $U=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$, $Q=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$, $\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$, 公式(5)(6)(7)中的 y_i, $\bar{y}$, $\hat{y}_i$ 不能用线性化后的 Y_i, $\bar{Y}$和 $\hat{Y}_i$ 代替,对于非线性函数类型,$L_{yy}=U+Q$ 不一定成立,有时 $U>L_{yy}$,因此不宜用式 $R=\sqrt{U/L_{yy}}$计算相关系数(另文讨论)。

1.2.4 回归方程的显著性检验

对于给定的置信度 α,可查 F 分布表[白新桂,1986],得临界值 F_α,从而由下式求得相关系数 R 的临界值 R_α。

$$R_\alpha=pF_\alpha/[(n-p-1)+pF_\alpha] \tag{8}$$

若 $R>R_\alpha$, $F>F_\alpha$,则上述回归方程对于置信度 α 时是显著的,即求得的回归方程能较好地描述样本数据变量间的关系,回归方程有实用价值,可用于预测或控制。

又由回归理论知,R_α、F 值越大和 S 值越小,回归方程的适用性越好。即在未知样本数据函数类型时,可以选用几种函数类型分别进行回归分析计算,比较 R、F 和 S 值,即 R、F 大的和 S 小的为最佳,其回归方程为最佳回归方程。

2 计算实例分析

文献(1)中鳙体重与体长的科研实测数据,采用1、4、5三种函数类型,其回归分析的计算结果如表2所示。

表2 不同函数类型回归计算结果比较(α=0.01)

Tab. 2 Comparison of regression results of several functions

函数类型序号	回归计算结果						
	b_0	b_1	R	F	S	F_α	R_α
1	−9136.43	253.39	0.9779	525.3	1169	5.66	0.44
4	0.04344	2.8248	0.9814	705.8	1073	5.66	0.44
5	5.4907	0.04615	0.8858	175.6	2596	5.66	0.44

由表2知,鳙体重与体长关系确以指数函数表示为最佳,即要求满足 von Bertalaffy 生长方程的基本假设 $W=aL^b$。

3 源程序清单

```
2 REM ********************
4 REM 数据回归分析通用程序
6 REM ********************
8 INPUT "输入数据组数 N=";N
10 DIM X(N,10),Y(N),A(11,12),Y1(N),
   XX(N,12),XXT(12,N),X1(N,10),BB(11),
   YY(N)
12 PRINT "------总菜单------"
14 PRINT "1--一元函数回归"
16 PRINT "2--多元函数回归"
18 INPUT "请输入编号";CH0%:CLS
20 IF CH0%=1 THEN P=1:GOTO 24
22 INPUT "输入自变量维数(P<10)",P
24 PRINT "-----数据调用方式-----"
26 PRINT "1---input 方式"
28 PRINT "2---顺序文件方式"
30 INPUT "请输入编号";CH1%
32 ON CH1% GOTO 34,50
34 FOR I=1 TO N
36 FOR J=1 TO P
38 PRINT "x(";I;",";J;")";
40 INPUT X1(I,J):X(I,J)=X1(I,J)
42 NEXT J
44 PRINT "y(";I;")=";
46 INPUT Y1(I):Y(I)=Y1(I)
48 NEXT I:CLS:GOTO 70
50 INPUT "输入文件名 F$ =";F$
52 F$ =F$ +".dta"
54 OPEN F$ FOR INPUT AS # 1
56 FOR I=1 TO N
58 FOR J=1 TO P
60 INPUT #1,X1(I,J):X(I,J)=X1(I,J)
62 NEXT J
64 INPUT #1 ,Y1(I):Y(I)=Y1(I)
66 IF EOF(1) THEN CLOSE #1:
   N=I:GOTO 70
68 NEXT I:CLOSE #1:CLS
70 PRINT "一元函数类型"
72 PRINT "1--y=b0+b1*x1+...+bp*xp"
74 PRINT "2--1/y=b0+b1/x"
76 PRINT "3--y=b0+b1*x^(1/2)"
78 PRINT "4--y=b0*x1^b1*...*xp^bp"
80 PRINT "5--y=exp(b0+b1*x1+...+bp*
   xp)"
82 PRINT "6--y=exp(b0+b1/x)"
84 PRINT "7--y=b0+b1*ln(x1)+...+
   bp*ln(xp)"
86 PRINT "8--y=1/(b0+b1*exp(-x)"
88 PRINT "9--y=b0+b1*x+...+bp*x^p"
90 INPUT "请输入编号";CH%
92 ON CH% GOSUB 102,116,130,144,166,184,
   196,214,100
94 GOSUB 306:GOSUB 340
96 END
98 REM---数据线性化处理---
100 INPUT "输入多项式次数(P<10)";P
102 GOSUB 230
104 LPRINT "y=";BB(0);
106 FOR J=1 TO P
108 IF BB(J)>0 THEN LPRINT "+";
110 LPRINT BB(J);
112 IF CH%=9 THEN LPRINT "*x";J
   ELSE LPRINT "*x^";J
114 NEXT J:LPRINT:RETURN
116 FOR I=1 TO N
118 X(I,1)=1/X(I,1):Y(I)=1/Y(I)
120 NEXT I
122 GOSUB 230
124 LPRINT "1/y=";BB(0);
126 IF BB(1)>0 THEN LPRINT "+";
128 LPRINT BB(1);"/x":RETURN
130 FOR I=1 TO N
132 X(I,1)=SQR(X(I,1)):Y(I)=Y(I)
134 NEXT I:GOSUB 230
136 LPRINT "y=";BB(0);
138 IF BB(1)>0 THEN LPRINT "+";
140 LPRINT BB(1);"*x^(1/2)"
142 RETURN
144 FOR I=1 TO N
146 PRINT Y(I);
```

```
148 FOR J=1 TO P
150 PRINT X(I,J)
152 X(I,J)=LOG(X(I,J))
154 NEXT J:Y(I)=LOG(Y(I))
156 NEXT I:GOSUB 230
158 LPRINT "y=";EXP(BB(0));
160 FOR I=1 TO P
162 LPRINT "*x";I;"^";BB(I);
164 NEXT I:LPRINT:RETURN
166 FOR I=1 TO N
168 Y(I)=LOG(Y(I))
170 NEXT I:GOSUB 230
172 LPRINT "y=";"exp(";BB(0));
174 FOR I=1 TO P
176 IF BB(I)>0 THEN LPRINT "+";
178 LPRINT BB(I);"*x";I;
180 NEXT I
182 LPRINT ")":LPRINT:RETURN
184 FOR I=1 TO N
186 X(I,1)=1/X(I,1):Y(I)=LOG(Y(I))
188 NEXT I:GOSUB 230
190 LPRINT "y=exp(";BB(0));
192 IF BB(1)>0 THEN LPRINT "+";
194 LPRINT BB(1);"/x)":RETURN
196 FOR I=1 TO N
198 FOR J=1 TO P
200 X(I,J)=LOG(X(I,J))
202 NEXT J,I:GOSUB 230
204 LPRINT "y=";BB(0);
206 FOR I=1 TO P
208 IF BB(I)>0 THEN LPRINT "+";
210 LPRINT BB(I);"ln(x";I;")";
212 NEXT I:LPRINT:RETURN
214 FOR I=1 TO N
216 X(I,1)=EXP(-X(I,1))
218 Y(I)=1/Y(I):NEXT I:RETURN
220 LPRINT "y=1/(";BB(0);
222 IF BB(1)>0 THEN LPRINT "+";
224 LPRINT BB(1);"*exp(-x))"
226 RETURN
228 REM 求系数矩阵
230 FOR I=1 TO N
232 XX(I,1)=1:XXT(1,I)=XX(I,1)
234 FOR J=2 TO P+2
236 IF CH%<>9 THEN XX(I,J)=X(I,J-1)
    ELSE XX(I,J)=XX(I,J-1)*X(I,1)
238 IF J=P+2 THEN XX(I,J)=Y(I)
240 XXT(J,I)=XX(I,J)
242 NEXT J,I
244 FOR I=1 TO P+1
246 FOR J=1 TO P+2
248 A(I,J)=0
250 FOR K=1 TO N
252 A(I,J)=XXT(I,K)*XX(K,J)+A(I,J)
254 NEXT K,J,I
256 REM --- 解方程 ---
258 FOR I=1 TO P+1
260 FOR J=I TO P+1
262 IF A(J,I)<>0 THEN 270
264 NEXT J
266 PRINT "矩阵某列元数全为0,方程无解!":
    GOTO 96
268 REM 求方程的系数 BB(I)
270 FOR K=1 TO P+2
272 SWAP A(I,K),A(J,K):NEXT K
274 T=1/A(I,I)
276 FOR L=1 TO P+2
278 A(I,L)=A(I,L)*T
280 NEXT L
282 FOR L=1 TO P+1
284 IF L=I THEN 294
286 T=-A(L,I)
288 FOR K=1 TO P+2
290 A(L,K)=A(L,K)+A(I,K)*T
292 NEXT K
294 NEXT L,I
296 FOR I=1 TO P+1
298 BB(I-1)=A(I,P+2)
300 LPRINT "b(";I-1;")=";BB(I-1),
302 NEXT I:LPRINT
304 RETURN
306 REM 求Y(I)的近似值 YY(I)
308 FOR I=1 TO N
310 YY(I)=BB(0)
312 FOR J=1 TO P
314 IF CH%=9 THEN Y=BB(J)*X(I,1)^J
```

```
    ELSE Y=BB(J)*X(I,J)
316 YY(I)=YY(I)+Y
318 NEXT J,I
320 ON CH% GOTO 328,322,328,322,322,322,
    328,322,328
322 FOR I=1 TO N
324 IF CH%=8 OR CH%=2 THEN YY(I)=
    1/YY(I) ELSE YY(I)=EXP(YY(I))
326 NEXT I
328 IF CH%=9 THEN P=1
330 FOR I=1 TO N
332 FOR J=1 TO P
334 LPRINT X1(I,J);:NEXT J
336 LPRINT Y1(I),YY(I)
338 NEXT I:RETURN
340 REM 回归方程显著性和相关性检验
342 Y=0:M=N-P-1
344 FOR I=1 TO N:Y=Y+Y1(I):NEXT I
346 Y=Y/N:U=0:LYY=0:Q=0
348 FOR I=1 TO N
350 U=U+(YY(I)-Y)^2
352 LYY=LYY+(Y1(I)-Y)^2
354 Q=Q+(Y1(I)-YY(I))^2
356 NEXT I
358 R=SQR(U/LYY):R1=SQR(1-Q/LYY)
360 F=U*M/(P*Q):S=SQR(Q/M)
362 LPRINT "R,R1,F,S=";R;R1;F;S
364 INPUT "输入某一置信度的F值";FO
366 RO=SQR(P*FO/(M+P*FO))
368 LPRINT "FO=";FO,"RO=";RO
370 IF F<FO OR R<RO THEN 374
372 LPRINT "回归方程显著":GOTO 376
374 LPRINT "回归方程不显著!重新回归"
376 RETURN
```

参 考 文 献

[1] 白新桂，1986.数据分析与试验优化设计，126－127.清华大学出版社(京).

[2] 刘国安等，1988.乌龟性腺发育的研究.水产学报，12(1)：13－20.

[3] 徐庆登等，1992.高邮杂交鲫杂种优势利用及其遗传性状.上海水产大学学报，1(1－2)：10－19.

[4] 楼文高，1992.分压式拦鱼电栅实用设计方程的初步研究.渔业机械仪器，19(5)：29－31.

[5] 臧维玲等，1993.河口区中国对虾幼虾中间培育池水化学状况.上海水产大学学报，2(2)：101－112.

[6] 戴祥庆等，1988.青鱼饲料最适能量蛋白比的研究.水产学报，12(1)：35－41.