郑宗成　王振堂　编

实用预测方法BASIC程序库

实用预测方法BASIC程序库

郑宗成　王振堂　编

中 山 大 学 出 版 社

实用预测方法BASIC程序库

郑宗成　王振堂　编

●

中山大学出版社出版
广东省新华书店发行
中山大学印刷厂印刷

787×1092　32开本　9.75印张　205,000字
1985年9月第1版　1985年9月第1次印刷
印数：1—11,000册

书号：13339·7　　定价：1.95元

内 容 簡 介

本书包括回归分析、时间序列分析、投入产出分析等常用预测方法的BASIC语言程序，共22个。每个程序均有 方法说明、程序说明、程序清单、上机操作说明以及实际应用例子。

本书可供经济计量研究部门、国家经济计划部门、商业、厂矿企业、财贸系统的经济管理人员、技术人员使用，亦可作为大专院校经济系、企业管理系师生的工具书。

前　言

近年来我国许多企业都在开展市场调查和市场预测工作，各行各业特别是经济计划部门都在为实现2000年的宏伟奋斗目标而制定长远规划和近期计划，以便作出最优决策。无论是规划、计划、决策以及市场预测，都要以科学的预测方法作为依据。目前我国已经翻译及编写了一些预测方法的著作，为了便于使用这些预测方法，我们将常用的预测方法编写成苹果Ⅱ型微型计算机上的BASIC语言程序。

本书的各章内容安排如下：第一章是回归分析程序。回归分析模型是利用一个或多个自变量和被预测变量之间的依赖关系来进行预测的。按照自变量的个数可分为一元回归分析和多元回归分析两种，根据依赖关系又可分为线性和非线性回归分析两种。本章中除给出各种回归分析的预测程序外，还给出了拟合幂函数，Logistic曲线和Gompertz曲线的程序。第二章至第五章是时间序列分析程序。时间序列分析模型是利用被预测变量的过去资料来对其未来进行预测的。第二章给出了过去资料数据在某一水平线上、下随机波动的水平时间序列模型的预测程序。如果过去资料数据有明显的直线上升或下降趋势，或者在某条二次曲线上、下随机波动，这就是趋势时间序列模型，其预测程序在第三章中给出。第四章是过去资料数据呈现季节性或周期性的季节性时间序列模型的预测程序。第五章给出了预测精度较高的Box-Jenkin时间序列模型及门限自回归模型的预测程序。第六章是投入产出分析计算程序，它是一种常用的综合平衡

预测方法。

本书编写时从实用角度出发，对每一个程序除概要地叙述该程序所使用的预测方法外，还详细地说明了使用该程序的上机操作步骤，以便从未接触过微型计算机的读者只须稍加学习苹果Ⅱ型微型计算机的上机操作方法（可参见本书附录）就可以顺利地使用这些程序。使用程序时并不要求读者懂得 BASIC 语言。本书每一个程序后面均附有通程序时所用的数据以及计算机运行结果，读者将程序打入计算机后，可用它们来检查所打入的程序是否正确，只有当计算结果和原来的计算结果完全一致时，才能使用所打入的程序。为了方便读者使用，我们已将本书中所有程序存贮在一个软磁盘内，有了这个软磁盘，读者使用时只需将盘中的程序调入内存即可进行计算。读者如需要软磁盘可向中山大学出版社联系。

为了便于读者使用本书所介绍的程序，书中对每一程序都给出若干利用该程序作预测的实际例子。因而本书既是一本使用预测方法的工具书，也是一本学习预测方法的有益参考资料。

参加本书编写的还有中山大学数学系八〇届毕业班的邵全喜、芦茂群、黄志明同学。本书编写过程中得到中山大学出版社、计算机科学系微型机实验室同志的帮助和支持，特此表示感谢。

由于我们水平有限，缺点错误定难避免，敬请读者批评指正。

编　　者

1984年9月

目　录

第一章　回归分析模型 …………………………………（1）

§1　一元线性回归 ………………………………………（1）

§2　一元非线性回归 ……………………………………（14）

§3　用非线性最小二乘法拟合幂函数 …………………（34）

§4　用非线性最小二乘法拟合 Logistic 曲线…（44）

§5　用非线性最小二乘法拟合 Gompertz 曲线（56）

§6　多元线性回归 ………………………………………（67）

第二章　水平时间序列模型 ………………………………（88）

§1　一次滑动平均模型 …………………………………（89）

§2　一次指数平滑模型 …………………………………（94）

第三章　趋势时间序列模型 ………………………………（100）

§1　线性回归模型 ………………………………………（102）

§2　二次滑动平均模型 …………………………………（108）

§3　二次指数平滑模型 …………………………………（116）

§4　一次平滑模型 ………………………………………（123）

§5　二次回归模型 ………………………………………（129）

§6　三次指数平滑模型 …………………………………（136）

第四章　季节时间序列模型 ………………………………（144）

§1　季节性水平模型 ……………………………………（148）

§2　季节性交乘趋势模型 ………………………………（165）

§3　季节性迭加趋势模型 ………………………… (181)
§4　三角函数模型 ……………………………… (198)

第五章　Box-Jenkins 模型 ……………………… (209)

§1　自回归模型 ………………………………… (212)
§2　自回归滑动平均模型 ………………………… (229)
§3　门限自回归模型 …………………………… (246)

第六章　投入产出分析 ……………………………… (281)

附录　APPLE Ⅱ 上机操作方法 ……………………… (301)

第一章　回归分析模型

事物的变化和发展常受许多因素的影响，例如收录机的销售量与其质量、性能、售价、社会拥有量、居民的收入水平，其它竞争品的性能和价格，磁带的价格等许多因素有关。又如家具和地毯的销售量与上年新建住宅落成面积有关。婴儿服装的需求量与婴儿出生人数有关等等。我们把影响因素用变量表示，并称它们为自变量或解释变量。自变量往往是非随机变量。随自变量而变的变量称为因变量或应变量。因变量常是随机变量。例如两年中婴儿出生人数相同时，但每一年婴儿服装的需求量可能不同，因而它们不能用普通的函数关系来表达。我们把给定了自变量，在不同的现实下因变量可以有不同的值与之对应的非确定性关系称为回归关系，有关回归关系的理论和方法称为回归分析。回归分析广泛用于科学实验和某个因变量的预测。本章给出了一元线性回归、一元非线性回归和多元线性回归的计算程序。作为非线性回归的特例，本章对在科学实验和经济预测中经常使用的幂函数、Logistic 曲线和Gompertz 曲线给出了用非线性最小二乘法确定参数的计算程序，利用这种方法可以大大减少曲线和实际观测值的拟合误差，因而提高了预测精度。

§1　一元线性回归

一、方法概要

设影响因变量 y 的自变量为 x，共作 n 次试验，其数据

为 (x_k, y_k)，$k=1, 2, \cdots, n$。根据这些数据要求我们在 y 与 x 之间配一线性回归方程

$$\hat{y} = a + bx$$

其中 a 是常数项，b 为 y 对 x 的回归系数，要求确定 a 和 b 使得总误差

$$Q = \sum_{k=1}^{n} (y_k - \hat{y}_k)^2 = \sum_{k=1}^{n} [y_k - (a + bx_k)]^2$$

达到极小。使上述 Q 达到极小的 a 和 b 分别记为 $\hat{a}$ 和 $\hat{b}$。由数学分析中求极值的方法推知

$$\hat{b} = l_{xy}/l_{xx}$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

其中

$$\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$$

$$\bar{y} = \frac{1}{n}\sum_{k=1}^{n} y_k$$

$$l_{xx} = \sum_{k=1}^{n} (x_k - \bar{x})^2 = \sum_{k=1}^{n} x_k^2 - \frac{1}{n}\left(\sum_{k=1}^{n} x_k\right)^2$$

$$l_{xy} = \sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y})$$

$$= \sum_{k=1}^{n} x_k y_k - \frac{1}{n}\left(\sum_{k=1}^{n} x_k\right)\left(\sum_{k=1}^{n} y_k\right)$$

为了下面的需要，还要计算

$$l_{yy} = \sum_{k=1}^{n} (y_k - \bar{y})^2 = \sum_{k=1}^{n} y_k^2 - \frac{1}{n}\left(\sum_{k=1}^{n} y_k\right)^2$$

从上面的讨论中，我们可以看出，对任何两个变量 x 和 y 的一组实验数据(x_k, y_k) $(k=1, 2, \cdots, n)$，都可以按上述计算步骤配出一条直线。但是 x 和 y 之间究竟有没有线性相关关系呢？回归直线有没有实际应用的价值呢？这可用相关系数来确定。相关系数定义为

$$r=\frac{l_{xy}}{\sqrt{l_{xx}\, l_{yy}}} \tag{1-1-1}$$

可以证明，相关系数 r 的取值范围为

$$-1\leqslant r\leqslant 1$$

相关系数是衡量两个变量线性相关关系是否密切的一个指标。其绝对值越接近于1，x 和 y 的线性关系越好，如果它接近于0，就可以认为 x 和 y 之间没有线性关系，这时有两种情况：或者两者没有关系，或者两者有非线性关系。表1-1-1是相关系数检验表，表中的数叫做相关系数的起码值。求出的相关系数要大于表上的数，才能考虑用直线来描述 x 和 y 之间的关系。

检验回归直线有没有实际应用的价值的另一方法是对回归进行方差分析。可以证明总离差 l_{yy} 可以分解为

$$l_{yy}=Q+U \tag{1-1-2}$$

其中

$$Q=\sum_{k=1}^{n}[y_k-(\hat{a}+\hat{b}x_k)]^2=\sum_{k=1}^{n}(y_k-\hat{y}_k)^2$$

称剩余平方和，而

$$U=\sum_{k=1}^{n}(\hat{y}_k-\bar{y})^2$$

称回归平方和。

表1-1-1　　相关系数检验表

n-2	5%	1%	n-2	5%	1%	n-2	5%	1%
1	0.997	1.000	16	0.468	0.590	35	0.325	0.418
2	0.950	0.990	17	0.456	0.575	40	0.304	0.393
3	0.878	0.959	18	0.444	0.561	45	0.288	0.372
4	0.811	0.917	19	0.433	0.549	50	0.273	0.354
5	0.754	0.874	20	0.423	0.537	60	0.250	0.325
6	0.707	0.834	21	0.413	0.526	70	0.232	0.302
7	0.666	0.798	22	0.404	0.515	80	0.217	0.283
8	0.632	0.765	23	0.396	0.505	90	0.205	0.267
9	0.602	0.735	24	0.388	0.496	100	0.195	0.254
10	0.576	0.708	25	0.381	0.487	125	0.174	0.228
11	0.553	0.684	26	0.374	0.478	150	0.159	0.208
12	0.532	0.661	27	0.367	0.470	200	0.138	0.181
13	0.514	0.641	28	0.361	0.463	300	0.113	0.148
14	0.497	0.623	29	0.355	0.456	400	0.098	0.128
15	0.482	0.606	30	0.349	0.449	1000	0.062	0.081

Q 表示因变量 y 的实际值与根据回归方程的计算值之间的误差平方和，它是自变量 x 对因变量 y 的线性影响以外的一切因素对这个因变量的作用造成的。U 是据根回归方程计算出的因变量的值与因变量的平均值之差的平方和。我们假设变量y与x之间有一定的线性关系，y是随 x 的取值不同而变化的，当x的值不等于它的平均值$\bar{x}$时，自然相应的y值也应与$\bar{y}$的值有所不同。式(1-1-2)说明因变量 y 取不同值 $y_1, y_2, \cdots, y_n$是由两类原因造成的，其一是由于自变量 x 取值不同(由U反映)；其二是由于其它原因，如误差(由

Q反映）。回归分析必须分清何者是主要原因。如果前者是主要的，即在总变差中回归平方和U所占的比重很大，则线性回归的效果就好；如果剩余平方和Q所占的比重很大，则线性回归的效果就不好。令

$$F=\frac{U}{Q/n-2}$$

计算出F的值后，利用F表可以进行检验。给定显著水平α，在F分布表中查出第一个及第二个自由度分别为1及$(n-2)$的值F_α，如果从某一个实际问题计算的F值，大于或等于查表得出的F_α值，则说明y与x的线性相关关系密切；如果小于查表得出的F_α值，则说明线性相关关系不密切。

最后，我们来讨论利用回归方程进行预报的精度问题。由于x和y之间是回归关系，知道了x的值，并不能精确地知道y的值，但由回归线可以知道y的平均值是$\hat{y}$，那么实际的值离开$\hat{y}$可能有多远呢？利用剩余标准离差可以回答这个问题，它定义为

$$s=\sqrt{\frac{1}{n-2}Q}=\sqrt{\frac{1}{n-2}\sum_{k=1}^{n}(y_k-\hat{y}_k)^2}$$

当x和y是线性关系时，s亦可用

$$s=\sqrt{\frac{l_{yy}-l_{xy}b}{n-2}}$$

计算。可以预料，有95.4%的y值落在两条平行于回归线的直线

$$y'=a-2s+bx$$
$$y''=a+2s+bx$$

之间。

二、程序中主要变量的意义

N　资料数据个数

X　贮存$\sum_{k=1}^{N}x_k$

Y　贮存$\sum_{k=1}^{N}y_k$

XX　贮存$\sum_{k=1}^{N}x_k^2$

XY　贮存$\sum_{k=1}^{N}x_k\cdot y_k$

YY　贮存$\sum_{k=1}^{N}y_k^2$

A　贮存线性回归方程的常数a

B　贮存回归系数b

R　贮存相关系数r

U　贮存回归平方和U

Q　贮存剩余平方和Q

F　贮存F比

S　贮存剩余标准离差s

X(N)　贮存自变量x的资料数据

Y(N)　贮存因变量y的资料数据

YK(N)　贮存拟合值$\hat{y}$

A$　用来回答是否要求预测。YES要求进行预测，NO不要求进行预测。（其中$表示“字符串型”的符号，下同。）

K, Y1, Q1, YU 工作单元。

三、程序

文件名 LR

```
10 INPUT "N=? "; N
20 DIM X(N), Y(N), YK(N)
30 FOR K=1 TO N
40 READ X(K), Y(K)
50 X=X+X(K)
60 Y=Y+Y(K)
70 XX=XX+X(K)*X(K)
80 XY=XY+X(K)*Y(K)
90 YY=YY+Y(K)*Y(K)
100 NEXT K
110 B=N*XX-X*X
120 IF B=0 THEN PRINT "NO UNIQUE
    SOLUTION" : STOP
130 B=(N*XY-X*Y)/B
140 A=Y/N-B*X/N
150 A=INT (A*10000+0.5)/10000
160 B=INT (B*10000+0.5)/10000
170 PRINT"Y="; A; "+"; B; "*X"
180 R=(N*XY-X*Y)/SQR ((N*XX-X*X)
    *(N*YY-Y*Y))
190 PRINT "R="; R
200 Y1=Y/N
210 FOR K=1 TO N
220 YK(K)=A+B*X(K)
230 Q1=Y(K)-YK(K)
```

```
240 YU=YK(K)-Y1
250 U=U+YU*YU
260 Q=Q+Q1*Q1
270 NEXT K
280 F=(N-2)*U/Q
290 S=SQR (Q/(N-2))
300 PRINT "U="; U
310 PRINT "Q="; Q
320 PRINT "F="; F
330 PRINT "S="; S
340 PRINT
350 INPUT "PREDICTION? <YES OR NO>"; A$
360 IF A$="NO" GOTO 420
370 INPUT "X=?"; X
380 Y=A+B*X
390 PRINT
400 PRINT "PREDICTION OF Y="; INT
    (Y * 100+0.5)/100
410 GOTO 340
420 PRINT
430 PRINT "FITTED VALUES OF Y(K)="
440 FOR K=1 TO N
450 PRINT INT (YK (K) * 100+0.5)/100; "     ";
460 IF INT (K/5)=K/5 THEN PRINT
470 NEXT K
480 END
1000 DATA 121, 360, 118, 260, 271, 440, 190, 400,
     75, 360, 263, 500, 334, 580, 368, 560, 305, 505,
     210, 480
1010 DATA 387, 602, 270, 540, 218, 414, 342, 590,
```

```
    173, 492, 370, 660, 170, 360, 205, 410, 339, 680,
    283, 594
■RUN
N=?20
Y=218.4147+1.0811*X
R=.884074985
U=188078.229
Q=52562.1419
F=64.4077276
S=54.0381254
PREDICTION? <YES OR NO> YES
X=?400
PREDICTION OF Y=650.85
PREDICTION ? <YES OR NO> NO
FITTED VALUES OF Y(K)=
349.23   345.98   511.39   423.82   299.5
502.74   579.5    616.26   548.15   445.45
636.8    510.31   454.09   588.15   405.45
618.42   402.2    440.04   584.91   524.37
```

四、上机操作说明

1．将本程序调入机内：从键盘打入LOAD LR，按**RETURN**键。*

2．输入资料数据：从键盘打入1000DATA，再打入资料数据，**每一程序行可打入20个数据，各个数据之间用逗号“，”隔开，但最后一个数据后面不能加逗号。若数据多于20个，可以另起一个程序行，但每行开头都必须有**DATA，

---

*若程序直接从键盘打入机内，此步骤可省略，请参见附录。