

《衡阳农业科技通讯》

· 增 刊 ·

微型电子计算机 在农业回归分析中的应用讲座

湖南省衡阳市农业科学研究所

周庭波 编著

《衡阳农业科技通讯》编辑室

1984.9

衡阳农业科技通讯

(增刊)



1984年9月出刊

微型电子计算机 在农业回归分析中的应用讲座

目

录

- 第一讲 绪 论..... (1)
- 第二讲 用 F X-3600 P袖珍计算器作
直线回归分析..... (3)
- 第三讲 用 P C-1500袖珍程序计算机
作多元线性回归分析..... (7)
- 第四讲 用 F X-702 P袖珍程序计算器
作曲线回归分析..... (19)
- 第五讲 用 A P P L E II 微型计算机
作正交多项式回归分析..... (26)
- 第六讲 用 A P P L E II 微型计算机
作逐步回归分析..... (39)
- 第七讲 用 A P P L E II 微型计算机
作多元非线性回归分析..... (50)
- 第八讲 用 A P P L E II 微型计算机
作回归设计试验分析..... (51)
- 第九讲 用 A P P L E II 微型计算机
作积分回归分析..... (63)
- 第十讲 农业综合回归分析..... (70)

第一讲 绪 论

§ 1.1 回归分析是研究农业问题的有效方法 农业科研和生产中的课题是十分复杂的，往往一个课题中包含许多的因子。这些因子，有的是可控变量，即非随机变量，如施肥量、插植密度、插秧叶龄等；有的是不可控变量，即随机变量，如气温、土壤速效养分、植株高度、作物产量等。这些可控或不可控变量之间的复杂的数量关系，不能不是农业科技工作者关注的重心。因为只有把握这种数量关系，才能真正谈得上定量研究，才能比较准确地控制农业生产过程中的某些条件或预测其结果。而要把握这种数量关系，对农业生产过程进行控制和预测，简单的组群比较 t 测验或方差分析是无济于事的，只有应用回归分析方法才能办到。

现代回归分析方法，是数理统计的一个重要分支，不仅能研究二个变量间的简单的线性关系，而且能研究一个变量与一组变量间的线性或非线性关系，还能研究二组变量间的复杂的关系（因篇幅所限，二组变量间的回归在本书中没有讨论）。当我们对某一农业课题进行回归分析，获得回归方程或称数学模型后，就不难对其过程进行控制和预测。所以，我们说回归分析是研究农业问题的有效方法。我们农业科技工作者不能老停留在 t 测验和方差分析阶段，必须再前进一步，将回归分析普遍应用于农业科研和生产。

§ 1.2 微型计算机开拓了农业回归分析的广阔前景 统计学很早就产生和发展了回归分析方法，然而随着变量个数的增加和分析方法的不断复杂化，回归分析的计算量使人望而却步，从而阻碍了回归分析方法的应用。例如，要解一个10元线性回归问题，要做几十万次的运算，是人力难以胜任的。有了电子计算机，这个问题便迎刃而解了。特别是近年飞跃发展的微型电子计算机，因为其价格的急速降低和易于掌握，更促进了农业回归分析的广泛应用。

国外如美国针对不同作物不同土壤所建议的施肥公式，都是通过回归试验设计和分析，借助电子计算机而得到的。国内北京师范大学刘来福等早几年就应用微型计算机计算遗传距离和选择指数，后者实际上也是回归问题。我省娄底地区农科所庄郁华等首先将回归试验设计应用于水稻规范化栽培，通过电子计算机寻找优化方案，取得了可喜的成果。我所在育种、肥料试验、栽培试验诸方面也都应用了回归分析方法和微型计算机，收到了一定的成效。今年，由省农业厅组织的，有我所参加的“微型计算机在苗情预报中的应用研究”课题协作组，已经在苗情预报的关键技术上取得了重大进展。此外，安徽、上海、四川、广东等省市也正在农业中应用微型计算机方面大踏步前进。

现在，微型计算机正在向更小型化，但计算能力更强的方向发展，与之相应的软件研究也在发展。可以相信，成套的适于微机的农业回归分析软件将会应运而生。这更开拓了农业回归分析的广泛前景。

§1.3 本书的宗旨和主要内容 本书旨在以简短的篇幅，将微型计算机的使用和农业回归分析紧密结合，有必要的公式推导，有用相应的微机的BASIC语言编写的计算机程序，有实际例子的数据资料，也有对电算结果的简要分析，尽量使之适合如同我初用回归分析和微机时的同行们参考。

本书第二讲介绍如何用FX-3600 P袖珍计算器作直线回归分析。通过这一讲引入回归分析的一般概念，介绍最小二乘法原理和回归方程的显著性测验。第三讲介绍如何用PC-1500袖珍程序计算机作多元线性回归分析。这一讲是第二讲的直接推广，也是下面各讲的基础。这一讲着重讨论了用矩阵形式来解决多元线性回归的计算问题。第四讲介绍如何用FX-702 P袖珍程序计算器作曲线回归分析，但只举了生长曲线回归和二次曲线回归二个例子。第五讲至第九讲分别介绍应用APPLE II微型计算机作正交多项式回归分析、逐步回归分析、多元非线性回归分析、回归设计试验分析及积分回归分析。这几讲的次序都是按它们之间的有机联系安排的。其中积分回归分析在一般的统计学书籍中鲜有介绍。第十讲简述了农业综合回归分析，其中也讲到了回归分析的困难。

本书列举了一种计算器和三种微型计算机的使用方法。因为现在许多农业科技工作者手中握有如同FX-3600 P一类的计算器，所以顺便加以介绍。其他三种微型计算机也是现阶段一般农业科技工作者所能遇到的。尽管如此，当中用基本或扩展BASIC语言编写的程序，并不失其通用性。

最后，趁讲座刊行之机，向我的大学老师北京农业大学吴仲贤教授和研究生指导老师湖南农学院裴新澍教授致以衷心的感谢，因为是他们鼓励和指导我在农业科研和生产中运用数学。另衡阳市蔬菜研究所旷碧锋同志协助机上操作，我所王开元同志为本书作图，在此一并致谢。

虽然主观上尽力避免出现错误，但由于本人的数学水平有限以及回归分析和微机应用的不断发展，错误和不当之处在所难免，欢迎同行及专家们批评指教。

5/26

第二讲 用FX-3600P袖珍计算器

作直线回归分析

§ 2.1 引言 描写非随机变量之间的数量关系，我们使用函数式。例如不计空气阻力的自由落体公式 $h = \frac{1}{2}gt^2$ ，式中时间 t 为自变量，落下高度 h 为因变量，均为非随机变量。 g 为常数。函数式所表示的数量关系是严格的。描写随机变量之间或随机变量与非随机变量之间的数量关系，我们使用回归模式（注意“使用”二字，是指实际问题如此，而不是数学定义）。例如儿子和父亲的身高的关系可表示为： $y = \beta_0 + \beta x + \varepsilon$ 。式中 y 表儿子身高，为因变量， x 表父亲身高，为自变量，均为随机变量。 β_0 、 β 为常数， ε 是一个随机误差。回归模式所表示的数量关系是不严格的，原因是随机因变量除了同自变量有关之外，还有一个随机误差 ε 的干扰。这里顺便声明一句，为简便计，以后我们把随机变量也称为变量。

直线回归即一元一次回归，它是为研究两个变量之间的近似的线性关系而设的。所谓线性关系，是指一个变量随着另一个变量的变化而成比例地变化。

§ 2.2 一个直线回归的例子及其数学模型
对应杂交早稻“威优98”的5种插植密度，有5个穗平总粒数。以 i 代表试验号或样品号，以 x 代表插植密度（自变量），以 y 代表穗平总粒数（因变量），列如表2.1。

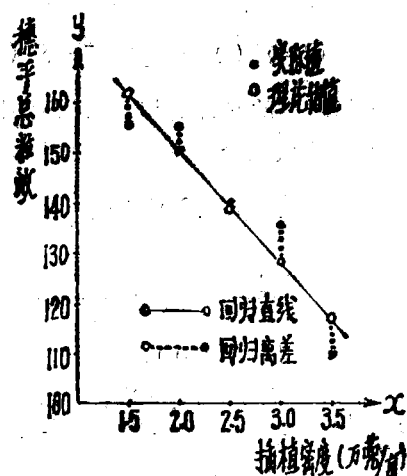


图2.1 穗平总粒数与插植密度的关系

表2.1

不同插植密度的穗平总粒数

试验号 i	1	2	3	4	5
插植密度 x (万蔸/亩)	1.5	2.0	2.5	3.0	3.5
穗平总粒数 y	155.3	155.1	139.7	135.2	109.8

先将 x 和 y 的5个数对描在图2.1的坐标平面上，可以看出这些点大致在一条直线上。穗平总粒数 y 随着插植密度 x 的增加相应地有所降低。看来我们可以用直线回归来描述它们之间的关系，其数学模型为：

$$y_i = \beta_0 + \beta x_i + \varepsilon_i \quad (i = 1, 2, \dots, N) \quad (2.1)$$

(2.1) 式中 β_0 是回归直线在 y 轴上的截距，称回归截距，而 β 是回归直线的斜率，称回归

系数, ε 是随机误差, X 、 y 、 ε 的下标 i 是指第 i 个试验, 或第 i 对数据。这里, $N=5$, 即资料包含 5 对数据。假如 $i=3$, 则从表 2.1 可知, X_3 为 2.5, 而 y_3 为 139.7。

式 (2.1) 即直线回归模型。插平总粒数 y 随插植密度 X 的变化而线性地变化, 这就是直线回归效应; 但总粒数并不完全取决于插植密度, 它还受到其他我们没有研究或不可能研究的因素的影响, 也即机误影响, 这就是随机误差 ε 。

§ 2.3 直线回归公式的推导 (2.1) 式是我们假设的数学模型。那么, 这样的假设符不符合实际情况, 又怎样求得 β_0 和 β 的合理的估值 b_0 和 b 呢? 我们先来解决后一个问题。现设想已从样本的 N 对数据中求出了 b_0 和 b 的值, 其估计的回归方程便为:

$$\hat{y}_i = b_0 + b X_i \quad (i = 1, 2, \dots, N) \quad (2.2)$$

比较 (2.2) 式和 (2.1) 式, (2.2) 式少了机误一项, 因为 $\hat{y}_i$ 是以 (2.2) 式算得的实际值 y_i 的理论估计值, y_i 与 $\hat{y}_i$ 之差才是机误。这个差又称回归离差。令

$$Q = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - b_0 - b X_i)^2 \quad (2.3)$$

Q 即回归离差的平方的和, 叫做离回归平方和。所谓合理的估值 b_0 和 b , 便应使得这个 Q 值为最少。从微积分可知, b_0 和 b 便是方程组

$$\begin{cases} \frac{\partial Q}{\partial b_0} = -2 \sum (y_i - b_0 - b X_i) = 0 \\ \frac{\partial Q}{\partial b} = -2 \sum (y_i - b_0 - b X_i) X_i = 0 \end{cases} \quad (2.4)$$

的解。方程组中 $\partial Q / \partial b_0$ 是 Q 对 b_0 的偏导数, 而 $\partial Q / \partial b$ 是 Q 对 b 的偏导数。解此方程得:

$$b = \frac{\sum X_i y_i - \sum Y_i \sum X_i / N}{\sum X_i^2 - (\sum X_i)^2 / N} \quad (2.5)$$

$$b_0 = \sum y_i / N - b \sum X_i / N \quad (2.6)$$

这样就可以求得回归方程 (2.2)。上面推导计算 b 、 b_0 的公式的方法叫最小二乘法。从图形上直观地看, 这条直线回归方程使得各点实际值到它的距离的和为最小。

从 (2.2) 式也可看出, 由于少了机误一项, 其形式与同名函数完全一样。事实上, 有什么样的函数, 就可以建立什么样的回归。不过, 究竟采用何种回归模式, 应以实际问题中变量之间的具体的数量关系而定, 这就需要专业知识。所以农业回归分析是农业专业知识和数理统计的综合运用。

§ 2.4 用 FX-3600 P 袖珍计算器进行直线回归的计算 从式 (2.5) 和 (2.6) 可知, 要计算出回归式中的 b 和 b_0 以及其后的回归显著性测验, 必须首先计算出 $\sum X_i$ (X 的总和)、 $\sum y_i$ (y 的总和)、 $\sum X_i^2$ (X 的平方总和)、 $\sum y_i^2$ (y 的平方总和) 和 $\sum X_i y_i$ (X 和 y 的乘积和), 然后按式计算即可。如果用手算, 这也是够麻烦的。但直线回归是回归分析中一种最基本最简单的回归。杀鸡焉用牛刀, 这种回归不必用较高级的微型计算机运算。有些电子计算器 (如 CASIO FX-3600 P、SHARP EL-5100、SHARP

EL-5103 S等)设有现成的回归和相关运算键,可以用来解决直线回归运算的各种问题。现以CASIO FX-3600 P计算器为例,进行表2.1的直线回归实例运算来说明它的操作过程(见表2.2)。这种操作虽然不使用程序语言,但从广义上说也是一种程序,是使用计算器执行某种运算的操作程序。并述附表中,SA...

从表2.2中可以查得b=-22.18(第9步)。所以,其回归方程为:

$$\hat{y}_i = 194.47 - 22.18x_i \quad (2.7)$$

表2.2 CASIO FX-3600 P直线回归运算操作步骤

步骤	操 作	显示器内容
1	ON [MODE] [2]	LR
2	[INV] [AC] 1.5 [X ₀ , y ₀]	1.5
3	KAC 155.3 [DATA]	155.3
4	2.0 [X ₀ , y ₀] 155.1 [DATA]	155.1
5	2.5 [X ₀ , y ₀] 139.7 [DATA]	139.7
6	3.0 [X ₀ , y ₀] 135.2 [DATA]	135.2
7	3.5 [X ₀ , y ₀] 109.8 [DATA]	109.8
8	[INV] [A]	194.47
9	[INV] [B]	-22.18
10	[INV] [r]	-0.9398
11	1.5 [Y]	161.20
12	2.0 [Y]	150.11
13	2.5 [Y]	139.02
14	3.0 [Y]	127.93
15	3.5 [Y]	116.84
16	[Kout] [ΣY]	695.1
17	[×] [×] [=]	483164.01
18	[+] [Kout] [n] [=]	96632.802
19	[=] [Kout] [ΣY ²] [=]	1392.468
20	[Kout] [ΣX] [=]	12.5
21	[×] [Kout] [ΣY] [=]	8688.75
22	[+] [Kout] [n] [=]	1737.75
23	[=] [Kout] [ΣXY] [=]	-55.45
24	[×] [INV] [B] [=]	1229.881

另外也可以从表2.2中查得根据(2.7)式算得的各种插植密度的穗平总粒数的理论估值。当插植密度分别为1.5、2.0、2.5、3.0、3.5(万蔸)时,穗平总粒数的回归理论估值分别为

161.20、150.11、139.02、127.93、116.84(第11步到15步)。同时, 相关系数 $r = -0.9306$ 也计算出来了(第10步)。

§2.5 直线回归的显著性测定 不论表2.1中的数据怎样, 我们都可以计算出 b_0 和 b_1 , 获得一条回归直线。但正如2.3节开头所提出的问题, 直线回归数学模型的假设是否符合表1.1的实际资料, 也即所配直线回归方程是否有意义, 必须得到回答。回归方程的显著性检验就是回答这个问题的。表2.2第16步之后的操作也就是为了达到这个目的的。

这里介绍用方差分析F测验来进行直线回归方程的显著性检验。先将因变量 Y 的离均差平方和分解为回归平方和和离回归平方和, 它们分别归诸于回归效应和机误影响, 其自由度也作相应的分解。计算公式列如表2.3。

表2.3 直线回归平方和及自由度的分解

变 异 来 源	平 方 和	自 由 度
总 的	$\sum Y_i^2 - (\sum Y_i)^2 / N$	$N - 1$
回 归	$b (\sum X_i Y_i - \sum X_i \sum Y_i / N)$	1
误 差	总的平方和 - 回归平方和	$N - 2$

对照表2.2, 显然第19步为 Y 的离均差平方和, 即总的平方和; 最后一步为回归平方和。由此可得方差分析表2.4。

表2.4 实例的直线回归方程显著性检验方差分析

变 异 来 源	平 方 和	自 由 度	均 方	F
回 归	1229.881	1	1229.881	22.69
误 差	162.587	3	54.196	
总 的	1392.468	4		

表2.4里, 各项的均方为各项的平方和与相应的自由度之商, 而 F 为回归均方与误差均方之商。查 F 表可知, 当大、小均方自由度分别为 1、3 时, 0.05 水准的 F 值为 8.65, 0.01 水准的 F 值为 34.1。所以该回归方程超过了 0.05 显著水准, 而没有达到 0.01 极显著水准。结论是, 穗平总粒数明显地因插植密度的加大而减少。我们可以用 (2.7) 的近似关系式来预测早稻“威优98”在某种密度下的穗平总粒数。不过, 直线回归在已试验的范围内作预测比较可靠, 超出试验范围太远的预测是会要犯错误的。

另外, 在此补充一点。在直线回归情形, 回归方程的显著性等价于相关系数的显著性。所以也可查相关系数的显著性测定表来作相关系数的显著性检验, 以代替直线回归方程的方差分析显著性检验。读者不妨自己用此例验证这一点。

第三讲 用PG-1500袖珍程序计算机

作多元线性回归分析

§3.1 引言 在农业科研和生产中，一个因变量只受一个自变量的影响的情况是很少的，往往是一个因变量受到许多个自变量的影响。例如，水稻亩产量不仅受插植密度的影响，而且受播种量、插秧叶龄、施肥量以及气温、雨量等的影响。这就要用多元回归分析来进行研究。而多元线性回归分析是多元回归分析中最基本的一种回归分析。所谓多元线性回归，即是多元一次回归。回归模式中所有自变量都是一次的，其意义是因变量与每一个自变量的数量关系都是线性的。多元线性回归既是多元非线性回归的基础，又是直线回归的直接扩展。

矩阵是多维的算术。多元回归是多维空间中的数学，所以，多元回归只有用矩阵来研究才会显得简便。不熟悉矩阵的同志可任找一本高等代数进行学习。

对于兴趣只在实用的同志，即使数学推导没有弄通，但只要明了多元线性回归的意义，明了数据结构及计算机上资料输入的方法，是不难使用本讲的现有程序去计算多元线性回归的。

§3.2 多元线性回归模型和一个实际例子 多元线性回归数学模型为：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_P X_{iP} + \varepsilon_i \quad (i=1, 2, \dots, N) \quad (3.1)$$

(3.1)式中 β_0 为回归常数， $\beta_1, \beta_2, \dots, \beta_P$ 为相应的自变量的偏回归系数， ε 是随机误差。 y, ε 的下标及诸 x 的第一个下标是指第 i 个试验，或第 i 组数据，诸 x 的第二个下标是自变量 x 的序号。这个模式表明，因变量 y 受到 P 个自变量和机误的影响，共做了 N 个试验，或观测了 N 组数据。

(3.1)式还可进一步简写为：

$$Y_i = \beta_0 + \sum \beta_j X_{ij} + \varepsilon_i \quad (i=1, 2, \dots, N; j=1, 2, \dots, P) \quad (3.2)$$

这里有一个杂交水稻“威优98”实际亩产量与有效穗、穗平实粒数和千粒重的资料（见表3.1）。虽然三者的乘积便是理论产量，但往往理论产量与实际产量相隔很远。我们想用三元线性回归来研究实际亩产与这三个产量构成因子之间的关系，以提高估产的精度。显然针对这个具体问题，其回归模式为：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad (i=1, 2, \dots, 16) \quad (3.3)$$

表3.1 威优98实际亩产与有效穗、穗平实粒数和千粒重

样本号	有效穗 x_1 (万/亩)	穗平实粒数 x_2	千粒重 x_3 (克)	实际亩产 y (斤/亩)
1	20.25	106.4	28.4	1103.3
2	19.95	93.5	28.5	1023.3
3	16.75	129.4	29.4	1083.3
4	17.55	127.0	29.9	1130.0
5	18.75	98.5	28.5	1123.3
6	19.73	116.6	29.4	1126.7
7	16.25	132.1	30.1	1080.0
8	15.70	119.1	29.1	946.7
9	21.45	104.4	29.3	1120.0
10	20.48	107.6	28.8	1090.0
11	17.40	125.7	29.4	1083.3
12	17.00	132.2	28.5	1050.0
13	18.53	114.8	29.5	1130.0
14	18.98	98.6	29.0	1093.3
15	16.75	114.8	29.6	1106.7
16	15.95	126.5	28.9	1046.7

3.3 多元线性回归系数的最小二乘估计及其矩阵化 我们可以仿照直线回归用最小二乘法来寻求多元线性回归模型中诸回归系数 β_j ($j=0,1,2,\dots,P$) 的合理估值 b_j ($j=0,1,2,\dots,P$)。亦假设我们已求得诸 b_j , 那么

$$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_p x_{ip} \quad (i=1,2,\dots,N) \quad (3.4)$$

设

$$\begin{aligned} Q &= \sum (y_i - \hat{y}_i)^2 \\ &= \sum (y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \dots - b_p x_{ip})^2 \end{aligned} \quad (3.5)$$

分别求 Q 对 $b_0, b_1, b_2, \dots, b_p$ 的偏导数并令其为 0, 整理得方程组:

方阵为增广方阵。(3.7)的增广方阵为:

$$\begin{pmatrix} N & \sum x_{i1} & \sum x_{i2} & \cdots \sum x_{ip} & \sum y_i \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1} x_{i2} & \cdots \sum x_{i1} x_{ip} & \sum x_{i1} y_i \\ \sum x_{i2} & \sum x_{i2} x_{i1} & \sum x_{i2}^2 & \cdots \sum x_{i2} x_{ip} & \sum x_{i2} y_i \\ \vdots & \vdots & \vdots & \cdots \vdots & \vdots \\ \sum x_{ip} & \sum x_{ip} x_{i1} & \sum x_{ip} x_{i2} & \cdots \sum x_{ip}^2 & \sum x_{ip} y_i \\ \sum y_i & \sum y_i x_{i1} & \sum y_i x_{i2} & \cdots \sum y_i x_{ip} & \sum y_i^2 \end{pmatrix} \quad (3.10)$$

考察这个增广方阵就可知是一个对称方阵。对角线元素除第一个为N外,其余都是相应变量的平方和。对角线两边的元素,除第一行、第一列为相应变量总和外,其余都是两两间的乘积和。这个增广方阵又可分解为下面两个矩阵的乘积:

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & x_{31} & \cdots & x_{N1} \\ x_{12} & x_{22} & x_{32} & \cdots & x_{N2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{1p} & x_{2p} & x_{3p} & \cdots & x_{Np} \\ y_1 & y_2 & y_3 & \cdots & y_N \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} & y_1 \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} & y_2 \\ 1 & x_{31} & x_{32} & \cdots & x_{3p} & y_3 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 1 & x_{N1} & x_{N2} & \cdots & x_{Np} & y_N \end{pmatrix} \quad (3.11)$$

显然,(3.11)中左边的矩阵是右边的矩阵的转置。右边这个矩阵我们称之为增广结构矩阵。把增广结构矩阵与表3.1的资料比较,仅仅是比资料表的左边多一列1。

如果我们用X表增广结构矩阵,用 X^T 表X的转置矩阵,用D表增广方阵,显然:

$$X^T X = D \quad (3.12)$$

§3.5 求解求逆紧凑变换 为了叙述方便,我们将增广结构矩阵最后一列因变量Y也写成X,并把它当作第P个变量,于是 y_i 便改为 x_{ip} 。这样前P-1个变量才是自变量。由这样的结构矩阵所求得的增广方阵便是:

$$\begin{pmatrix} N & \sum x_{i1} & \sum x_{i2} & \cdots & \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1} x_{i2} & \cdots & \sum x_{i1} x_{ip} \\ \sum x_{i2} & \sum x_{i2} x_{i1} & \sum x_{i2}^2 & \cdots & \sum x_{i2} x_{ip} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \sum x_{ip} & \sum x_{ip} x_{i1} & \sum x_{ip} x_{i2} & \cdots & \sum x_{ip}^2 \end{pmatrix} \quad (3.13)$$

这样一改,很清楚,(3.13)中的最后一列(或行),除去 $\sum x_{ip}^2$ 外,正是方程(3.6)相应的常数项。

进一步将(3.13), 各元素用 a_{ij} 表示为:

$$\begin{pmatrix} a_{00} & a_{01} & a_{02} & \dots & a_{0p} \\ a_{10} & a_{11} & a_{12} & \dots & a_{1p} \\ a_{20} & a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ a_{p0} & a_{p1} & a_{p2} & \dots & a_{pp} \end{pmatrix} \quad (3.14)$$

求解求逆紧凑变换可按任意次序对(3.14)的前 P 个对角元素进行(注意, (3.14)乃 $(P+1) \times (P+1)$ 方阵). 用 L_{kl} 表示第 l 次对第 $(l-1)$ 次变换后的对角元素 $a_{kk}^{(l-1)}$ 进行变换, 则 L_{kl} 变换为:

$$L_{kl}: \begin{cases} a_{ij}^{(l)} = a_{ij}^{(l-1)} - a_{ik}^{(l-1)} a_{kj}^{(l-1)} / a_{kk}^{(l-1)}, & i, j \neq k \\ a_{kj}^{(l)} = a_{kj}^{(l-1)} / a_{kk}^{(l-1)}, & j \neq k \\ a_{ik}^{(l)} = -a_{ik}^{(l-1)} / a_{kk}^{(l-1)}, & i \neq k \\ a_{kk}^{(l)} = 1 / a_{kk}^{(l-1)} \end{cases} \quad (3.15)$$

将变换(3.15)翻译为文字就是:

(1) 与 $a_{kk}^{(l-1)}$ 不同行又不同列的元素等于变换前该元素与一个分数之差, 这个分数的分子为变换前与该元素同行又与 $a_{kk}^{(l-1)}$ 同列的元素乘上与该元素同列又与 $a_{kk}^{(l-1)}$ 同行的元素的积, 而分母就是变换前的 $a_{kk}^{(l-1)}$.

(2) 与 $a_{kk}^{(l-1)}$ 同行但不同列的元素等于变换前该元素与变换前的 $a_{kk}^{(l-1)}$ 的商.

(3) 与 $a_{kk}^{(l-1)}$ 同列但不同行的元素等于变换前该元素与变换前 $a_{kk}^{(l-1)}$ 的商的反数.

(4) $a_{kk}^{(l-1)}$ 就等于变换前的 $a_{kk}^{(l-1)}$ 的倒数.

设增广方阵(3.14)通过对前 P 个对角元素的变换成为下面的方阵:

$$\begin{pmatrix} a_{00}^{(p)} & a_{01}^{(p)} & a_{02}^{(p)} & \dots & a_{0p}^{(p)} \\ a_{10}^{(p)} & a_{11}^{(p)} & a_{12}^{(p)} & \dots & a_{1p}^{(p)} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ a_{p0}^{(p)} & a_{p1}^{(p)} & a_{p2}^{(p)} & \dots & a_{pp}^{(p)} \end{pmatrix} \quad (3.16)$$

则(3.16)中前 P 阶方阵就是方程组(3.7)的系数矩阵的逆矩阵, 也即(3.9)式中的 A^{-1} ; 而最后一列(或行)前 P 个元素就是方程组(3.7)的解, 也即(3.9)式中的 B . B 中包含了诸回归系

数 b_j ($j=0, 1, 2, \dots, P-1$), 这样便可求得回归方程:

$$\hat{y}_i = b_0 + \sum b_j x_{ij} \quad (i=1, 2, \dots, N; j=1, 2, \dots, P-1) \quad (3.17)$$

(3.17) 中 $b_j = a_{jp}^{(p)}$ ($j=1, 2, \dots, P-1$), 而 b_0 等于 $a_{0p}^{(p)}$.

§ 3.6 全回归方程及诸回归系数的显著性测验 同直线回归一样, 也存在多元线性回归模型是不是符合实际, 整个回归方程是否有意义的问题。全回归方程的显著性检验便是针对此问题而设的。下面来叙述这个检验。

如果第一步对增广方阵 (3.14) 的第一个对角元素 a_{00} 进行变换 (3.15), 那么所得方阵的最后一个对角元素为:

$$a_{pp}^{(1)} = a_{pp} - a_{p0} a_{0p} / a_{00}$$

如用 (3.13) 中的相应元素表示, 则为:

$$a_{pp}^{(1)} = \sum x_{ip}^2 - \sum x_{ip} \sum x_{ip} / N$$

进一步, 如用 (3.10) 中相应的元素表示, 则为:

$$a_{pp}^{(1)} = \sum y_i^2 - \sum y_i \sum y_i / N \quad (3.18)$$

显然 $a_{pp}^{(1)}$ 正是因变量的离均差平方和, 也即总平方和 SS_y 。而通过对前 P 个对角元素变换之后的方阵 (3.16) 的最后一个对角元素 $a_{pp}^{(p)}$ 恰恰是误差平方和 SS_e (证明从略)。

所以回归平方和 SS_R 为:

$$SS_R = a_{pp}^{(1)} - a_{pp}^{(p)} \quad (3.19)$$

总平方和、回归平方和、误差平方和的自由度分别为 $(N-1)$, $(P-1)$ 和 $(N-P)$ 。于是我们可以完成多元线性全回归方程的方差分析和显著性检验, 列如表 3.2。由表 3.2 中的 F 值来判定全回归方程的显著性。

表 3.2 多元线性回归全回归显著性测验

变异来源	平方和	自由度	均方	F
回 归	$SS_R = a_{pp}^{(1)} - a_{pp}^{(p)}$	$P-1$	$S_R = SS_R / (P-1)$	S_R / S_e
误 差	$SS_e = a_{pp}^{(p)}$	$N-P$	$S_e = SS_e / (N-P)$	
总 的	$SS_y = a_{pp}^{(1)}$	$N-1$		

在作全回归显著性测验时, 可顺带计算出误差方差根 E 和复相关系数 R :

$$E = \sqrt{a_{pp}^{(p)} / (N-P)} \quad (3.20)$$

$$R = \sqrt{(a_{pp}^{(1)} - a_{pp}^{(p)}) / a_{pp}^{(1)}} \quad (3.21)$$

复相关系数 R 和误差方差根 E 也是全回归方程精度的度量。

与直线回归不同的是,多元线性回归除了对全回归方程进行显著性测验以外,还必须对每一个自变量的偏回归系数的显著性进行检验,以判明每一个自变量与因变量之间是不是存在显著的线性关系。下面又来叙述这个检验。

总的回归平方和 SS_R 可以再分为一个个自变量所引起的平方和,称之为偏回归平方和。可以证明,归诸于某一个自变量 x_j 的偏回归平方和为:

$$SS_j = b_j^2 / a_{jj}^{(p)} = (a_{jp}^{(p)})^2 / a_{jj}^{(p)} \quad (j=1, 2, \dots, P-1) \quad (3.22)$$

(3.22) 式中 b_j 为 x_j 的偏回归系数,它等于方阵(3.16)最后一列(或行)相应的元素 $a_{jp}^{(p)}$ (除去 $a_{op}^{(p)}$ 和 $a_{pp}^{(p)}$), $a_{jj}^{(p)}$ 为方阵(3.16)主对角线上的元素(除去 $a_{oo}^{(p)}$ 和 $a_{pp}^{(p)}$,也即为系数矩阵的逆矩阵 A^{-1} 主对角线上的元素)。偏回归平方和描述了与之相应的自变量对因变量 Y 的变异的贡献大小。对应于偏回归平方和的自由度为 1, 所以,

$$F_j = SS_j / S_e = (b_j^2 / a_{jj}^{(p)}) / (SS_e / (N-P)) \\ = (a_{jp}^{(p)})^2 (N-P) / a_{jj}^{(p)} a_{pp}^{(p)} \quad (j=1, 2, \dots, P-1) \quad (3.23)$$

可以用来作诸回归系数 b_j 的显著性测验。计算 F_j 的数据均可在方阵(3.16)中找到。 F_j 的大、小均方自由度分别为 1 和 $N-P$ 。

上述的所有计算完成后,不妨将各试验点的自变量值代入回归方程(3.17),求得 Y_i 的理论估值 $\hat{Y}_i$ 。 $\hat{Y}_i$ 又称回归拟合值。实际值与拟合值之差 $Y_i - \hat{Y}_i$ 便是拟合差,乃回归误差效应所造成。这个拟合差也最好一并算出。

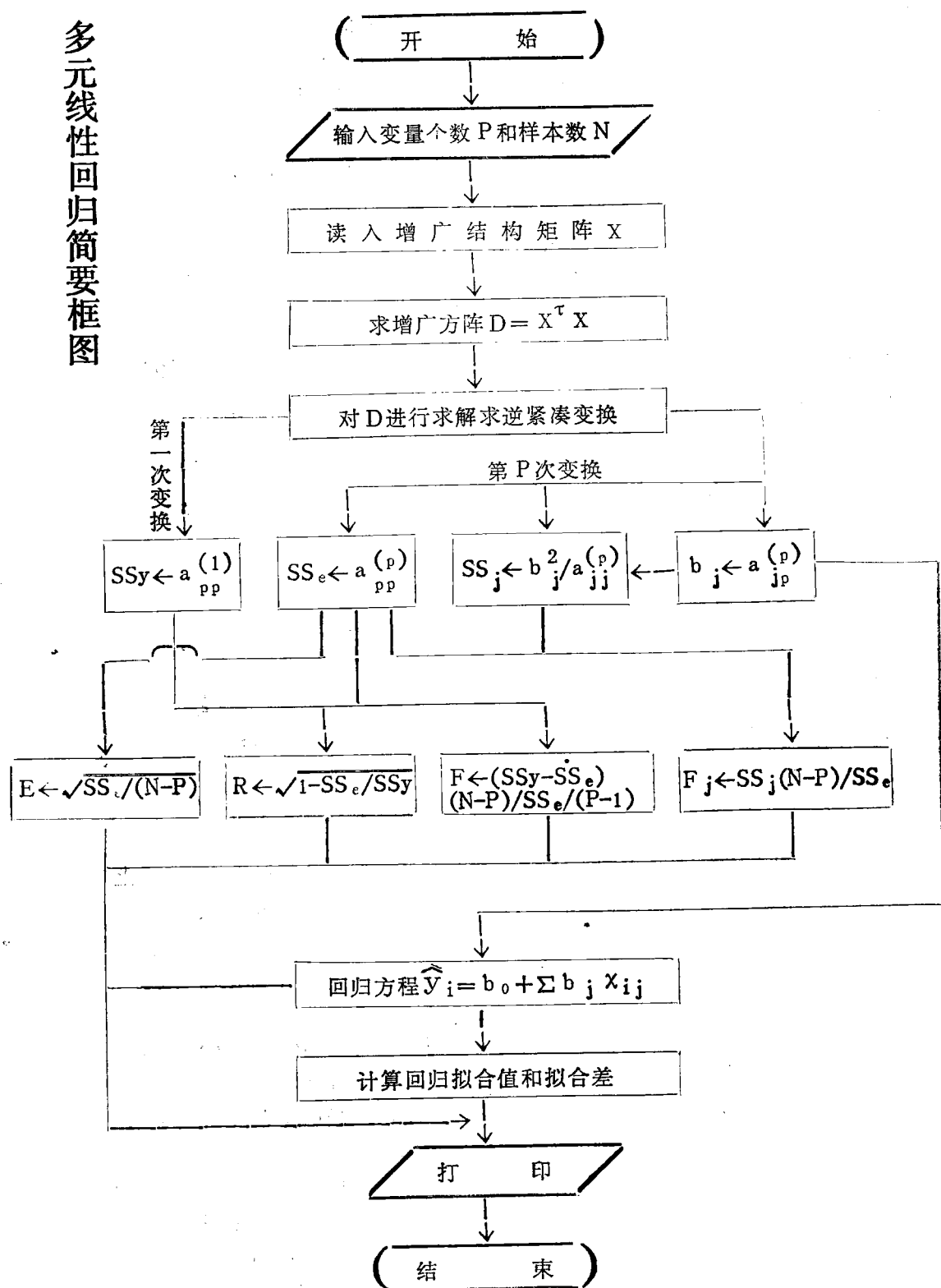
§3.7 用 PC-1500 袖珍程序计算机作多元线性回归分析的程序 从以上的讲述可以总结出多元线性回归分析计算的简要框图,如下页。根据这个框图和 SHARP PC-1500 袖珍程序计算机指令系统,将多元线性回归分析的 BASIC 程序编辑如下。该程序已在 PC-1500 微机上通过,亦可用于 Apple II 及其他微型计算机。

多元线性回归 BASIC 程序(在 SHARP PC-1500 机上通过)

```

5: "A": CLEAR          40: INPUT N: LPRINT N
10: LPRINT "THE NUMBER  50: DIM X(N-1,P), A(P,P)
    OF VARIABLE P=" ;   60: FOR J=0 TO P
20: INPUT P: LPRINT P   70: FOR I=0 TO N-1
30: LPRINT "THE NUMBER  80: READ X(I,J)
    OF SAMPLE N=" ;     90: NEXT I
                        100: NEXT J
    
```

多元线性回归简要框图



```

110: FOR J=0 TO P
120: FOR I=0 TO P
140: FOR K=0 TO N-1
150: A(I, J)=A(I, J)+
      X(K, J)*X(K, I)
160: NEXT K
170: NEXT I
180: NEXT J
190: FOR K=0 TO P-1
200: FOR J=0 TO P
210: IF J=K THEN 260
220: FOR I=0 TO P
230: IF I=K THEN 250
240: LET A(I, J)=A(I, J)-
      A(I, K)*A(K, J)/A(K,
      K)
250: NEXT I
260: NEXT J
270: FOR I=0 TO P
280: IF I=K THEN 300
290: LET A(I, K)=-A(I, K)
      /A(K, K)
300: NEXT I
310: FOR J=0 TO P
320: IF J=K THEN 340
330: LET A(K, J)=A(K, J)
      /A(K, K)
340: NEXT J
350: A(K, K)=1/A(K, K)
360: IF K<>0 THEN 380
370: LET Y2=A(P, P)
380: NEXT K
390: E2=A(P, P)
400: E=SQR(E2/(N-P))
410: R=SQR((Y2-E2)/Y2)

420: F=(Y2-E2)*(N-P)
      /(E2*(P-1))
430: FOR K=1 TO P-1
440: F1=A(K, P)^2/A(K, K)
      /(E*E)
460: LPRINT "F(" ; K ; "
      " ; F1
470: NEXT K
480: LPRINT
490: LPRINT "F=" ; F
500: LPRINT "R=" ; R
510: LPRINT "E=" ; E
520: LPRINT "Y=" ; A(0
      , P) ;
530: FOR K=1 TO P-1
540: LPRINT "+(" ; A(K, P
      ) ; " ) X(" ; K ; " )" ;
550: NEXT K
560: FOR I=0 TO N-1 :
      Y=0
580: FOR J=0 TO P-1
590: Y=Y+X(I, J)*A(J, P)
600: NEXT J
610: Y1=X(I, P)-Y
613: LPRINT
615: LPRINT "(" ; I+1 ;
      ")"
620: LPRINT "Y=" ;
      X(I, P)
630: LPRINT "EY=" ; Y
640: LPRINT "DY=" ; Y1
650: NEXT I
660: END
1000: DATA...
      :

```

该程序在行号1000后安排资料输入。首先输入N个1 再输入第一个变量的N个值，第二个变量的N个值等等，直至第P个变量（即因变量）的N个值。不过，要注意的是，每个