电子计算机在畜牧业上的应用

DIANZIJISUANJI ZAI XUMUYESHANGDE YINGYONG

DIANZIJISUANJI ZAI XUMUYESHANGDE YINGYONG

DIANZIJISUANJI ZAI XUMUYESHANGDE YINGYONG

陈幼春 张子仪 编著

农业出版社

电子计算机在畜牧业上的应用

陈幼春　张子仪　编著

电子计算机在畜牧业上的应用
陈幼春　张子仪　编著
*　*　*
责任编辑　刘振生

农业出版社出版（北京朝阳区枣营路）
新华书店北京发行所发行　　农业出版社印刷厂印刷

787×1092mm16 开本　　18 印张　　326 千字
1990 年12月第 1 版　　1990 年12月北京第 1 次印刷
印数　1—2,490 册　　定价 9.30 元
ISBN 7-109-01081-3/S·786

序　言

半个世纪以来，随着各项科学技术的发展，特别是各专门学科的相互渗透，在许多学科的边缘地带涌现出大量的突破性的成果，最令人瞩目的是计算机技术在许多方面所起着的重大作用。在畜牧业方面也和其他行业一样，基于客观上的信息量的增加，对信息处理技术的不断改革，一些先驱者们在计算机技术应用于畜牧方面积累了不少经验。特别是在遗传、育种、饲料、营养等方面起到了积极作用。

近年来，随着频繁的国际学术交流，以及国内一些学者的跨行业协作，在估测种畜个体育种值、分析计算各种经济性状的遗传参数、优选以最低成本为指标的饲料配方程序等方面，都进行过有益的尝试，并在生产上试用，获得明显的经济或社会效益。

当前计算机已从简单的数据处理发展到信息处理阶段。有人预测90年代计算机不仅是用于处理信息，而且还将主要用于更高层次的更系统的信息，即处理知识的时代。也可以说，计算机将发展成为知识处理机或知识信息处理系统。无疑到那时又将有一次重大变革，进入到智能革命时代——21世纪。

据了解，目前我国现有大中型计算机约5000台，微型机约13万台，袖珍计算机则更多。如何发挥这批硬件的作用，无疑的，启蒙式的读物将发挥应用的作用。作者们都是从事畜牧科研或教学的工作者，在计算机专业这一领域应该承认是小学生。由于在实践中尝到了甜头，愿把一部分心得体会整理出来，与广大读者共同学习，不断切磋提高。

本稿承蒙畜牧战线上应用电子计算机的先驱者们的支持和允诺，采纳他们辛勤劳动的成果，得以为广大读者提供系统的软件程序。在此一并表示感谢，并望提出宝贵意见。

鉴于我们在电脑应用上纯粹是小学生，编写不免会有许多差错，欢迎读者、老师和同行们批评指正。

编　者

1988年元月

承蒙各位先生热忱提供稿源，特此向：

于汉周	王桂英	甘　杰	刘松元	刘忠君	刘信福
卢书勤	庞之洪	许尚忠	毕中立	安树森	冯海清
李冀钊	李润藩	汪　强	吴常信	陈引杰	陈道炎
陈绪生	张　斌	张　瑜	张成武	张忠盛	苏文练
周必文	林诚玉	杨德新	杨静华	胡迪先	俞宗源
柳　楠	苗泽荣	封朝璧	武　彬	徐慧如	夏增权
高文仲	常宝山	商建国	彭中镇	喻传洲	焦　骅

等致谢。

编　者

目　　录

第一章　FORTRAN 语言的基本程序

FORTRAN 一字是由 FOR mula TRAN slator 两字演变而来，是公式译码的简写，广泛用于科学研究，是研究的计算机语言。尤其是使用最新的 IBM 机常与 FORTRAN 编译程序相联系，随着 IBM 机的日益普及和科研在生产中的不断深入和发挥越来越大的作用，FORTRAN 语言将成为最常见的计算语言。由于世界上 IBM 机占有领先地位，FORTRAN 的介绍十分广泛，为便于吸收世界科学发展的先进技术，本章主要通过畜牧业上常用的统计方法，介绍基本的几个典型的 FORTRAN 程序，为改进和自编新的程序作参考。有关的程序主要是 FORTRAN Ⅳ。这是以美国国家标准研究所（ANSI）和计算机协会（ACM）协调制定的标准为根据的，便于与新的发展作对比。下面将介绍基本数理统计方法的程序和矩阵代数程序，以及关于畜牧业管理和繁育方面应用的程序。

一、基本数理统计方法程序

畜牧生产和科研经常需要科学地处理一些数据，涉及一定数量的个体，在一定的范围内这些个体是个群体。该群体的某性状有着数量分布的范围，这种分布可能是正态的，也可能是偏态的。这种分布有许多特征值。当我们说明的是群体时，这些特征值叫做参数；当我们指的是样本时，这些特征值叫统计值。抽样的统计值可用来估计原有群体的参数，并用于鉴验所涉群体的假说。

对群体或样本的最重要的度量首先是平均数。还有一些度量值在实践上用得不多，如众数，是指出现在频率最高的地方的那个值。如图 1.1.1 所示，众数就是频率曲线上最高点的值。中数是频率分布中间的值。在图 1.1.1 上，分布曲线下有一半面积是在中数的右边，一半是在左边。均数是算术平均数的另一个用词，并定义为所有观察值之和被观察数所除的商。在非对称频率曲线中，中数位于均数和众数之间；在对称曲线中，诸如正态曲线，这三个度量合一。

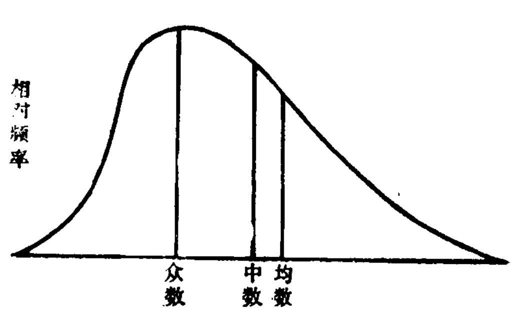

图 1.1.1　众数、中数和均数

传统上度量分布曲线有几个常用的符号。群体分布的符号用希腊字母，而抽样分布的符号用罗马字母。如样品均数用 $\bar{x}$，而群体均数用 μ(mu)。样品均数有两个很理想的特

性，它比中数和众数能更好地估测平均数。首先，样品均数是群体均数的一个无偏估计，如果统计量是得自很大的样品系列，那末统计值相当于相应参数的非偏估计值。第二，当对称分布为正态分布时，样品均数极接近于群体均数。

下面是用FORTRAN语言编写的计算平均数、标准差、方差、协方差、相关系数以及方差分析的程序。

1. 平均数

计算平均数的程序如下：

```
C      PROGRAM1.1.1 XMEAN
C
C      PROGRAM TO COMPUTE THE MEA OF 'N' SAMPLES
C
C      SET SUM TO ZERO
       SUMX=0.0
C      READ NUMBER OF SAMPLES TO BE USED
       READ (5,1000) NS
       DO 100 I=1,NS
C      READ A SAMPLE AND ADD TO SUM
       READ (5,1001) X
       SUMX=SUMX+X
  100  CONTINUE
C      COMPUTE THE MEAN
       AMEAN=SUMX/FLOAT (NS)
C      PRINT RESULTS
       WRITE (6,2000) NS, SUMX, AMEAN
       CALL EXIT
 1000  FORMAT (14)
 1001  FORMAT (F10.1)
 2000  FORMAT (21H 1NUMBER OF SAMPLES= , I10, //,
       121H  SUM OF SAMPLES  = , F10.1, //,
       221H  MEAN OF SAMPLES  = ,F10.3)
       END
```

例：若抽样测定某地所产箭筈豌豆的干物质中可消化粗蛋白的含量，经五次抽样测定，分别获得每公斤干物质中的克数为205、255、195、220、235。

通过上述程序的运算可得出两个结果：1，总数值1110；2，平均数222。

2. 标准差和方差

描述分布曲线特征的另一个方面是变数在均数两侧的分布情况，有两种度量较普遍地

被用于生产和科研。一个是方差，另一个是方差的平方根，称做标准差。方差也可认作是所有可能的观察值偏离群体均值的平均的离差平方，可定义为公式

$$\sigma^2 = \frac{\sum_{i=1}(X_i - \mu)^2}{n} \tag{1.1.1}$$

σ^2 为群体的方差。而样品的方差用 S^2 表达。如果 $X_L \cdots\cdots X_n$ 各观察值是一个常态分布中的随机样品，S^2 就是对 σ^2 的有效估计。

在常规情况下，平方偏差的均值在使用上并不方便，而标准差却比较方便。在实际应用上以计算式代替（1.1.1）的定义式

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1} \tag{1.1.2}$$

或

$$S^2 = \frac{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)} \tag{1.1.3}$$

这里 ΣX_i^2 是非校正的平方和，而校正的平方和（SS）可定义为

$$SS = \sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{1.1.4}$$

其计算式是

$$SS = \sum_{i=1}^{n} X^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n} \tag{1.1.5}$$

在取得这个数值后，方差值可用（n－1）除 SS 获得。

$$S^2 = \frac{SS}{n-1} = \frac{\sum_{i=1}^{n} X_i^2 - \left[\left(\sum_{i=1}^{n} X_i\right)^2/n\right]}{n-1} = \frac{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)} \tag{1.1.6}$$

用（n－1）作为除数的原因是使 S^2 的估测为一个非偏差值。

例：其计算沿用上例的数据（表 1.1.1）。

表 1.1.1　平方和与方差计算过程

X	X^2
205	42025
255	65025
195	38025
220	48400
235	55225
$\Sigma X_i = 1110$	$\Sigma X_i^2 = 248700$

$(\Sigma X_i)^2 = 1232100$

$SS = 248700 - \frac{1232100}{5} = 2280$

$S^2 = \frac{2280}{4} = 570$

$S = \sqrt{570} = 23.88$

假定这组分析的值是近乎正态分布，我们可以预期其中有2/3是处于198至246之间，而实际的计算是60%的值落在这个范围之内。

其计算可用程序1.1.2进行。此程序可计算方差和标准差。

```
C       PROGRAM 1.1.2 VAR
C
C       PROGRAM TO COMPUTE THE VARIANCE OF 'N' SAMPLES
C
C       SET SUMS TO ZERO
        SUMX=O.O
        SUMX2= .O
C       READ NUMBER OF SAMPLES TO BE USED
        READ (5,1000) NS
        DO 100 I=1,NS
C       READ A SAMPLE AND ADD TO SUM
        READ (5,1001) X
        SUMX=SUMX+X
        SUMX2=SUMX2+X*X
   100  CONTINUE
C       COMPUTE THE VARIANCE
        VAR=(FLOAT(NS)*SUMX2-SUMX*SUMX)/FLOAT (NS* (NS-
          1))
C       PRINT RESULTS
        WRITE (6,2000) NS, SUMX, SUMX2, VAR
C       COMPUTE THE STANDARD DEVIATION AND MEAN
        STDEV=SQRT(VAR)
        AMEAN=SUMX/ELOAT(NS)
C       PRINT STANDARD DEVIATION AND MEAN
        WRITE (6,2001) STDEV, AMEAN
        STOP
  1000  FORMAT (I4)
  1001  FORMAT (F1C.1)
  2000  FORMAT (1H1, 8X, 20H NUMBER OF SAMPLES = , I10, //,
       112 X,17H SUM OF SAMPLES = , F10.3,//,
       229H SUM OF SQUARES OF SAMPLES = , F10.3//,
       37X,22H VARIA NCE OF SAMPLES = , F10.3)
  2001  FORMAT (/,8X, 21HSTANDARD DEVIATION = , F10.3, //,
```

```
111X, 18HMEAN OF SAMPLES = , F10.3)
END
```

3.协方差

例：如果上例箭筈豌豆样品的各自同株干秧上叶子干物质的蛋白质含量如表1.1.2，现要问豆粒和叶内蛋白质的含量间是否有相应关系。

表 1.1.2 蛋白质的量

豆粒内		叶内
	205	130
	255	165
	195	100
	220	135
	235	145
总计	1110	675
平均	222	135

这两组变数发生在不同的部位，如果相互之间没有关系，这两组数就相互独立。如果相互间有关系，这两组共变量就有协方差的关系，即产生对比的变化关系。这可用图1.1.2表示变数 X_1 与 X_2 的相互依变关系。

在计算协方差时，首先要象计算平方和那样求出一个数量，即校正的乘积和（SP），其定义式如：

$$SP_{jk}=\sum_{i=1}^{n}(X_{ij}-\overline{X}_j)(X_{ik}-\overline{X}_k) \qquad (1.1.7)$$

这里X_{ij}指的是 j 变数的第i个度量，X_{ik} 指的是 k 变数的第 i 个度量。SP_{jk} 是指变数 j 和变数 k 之间的乘积和。其计算式为

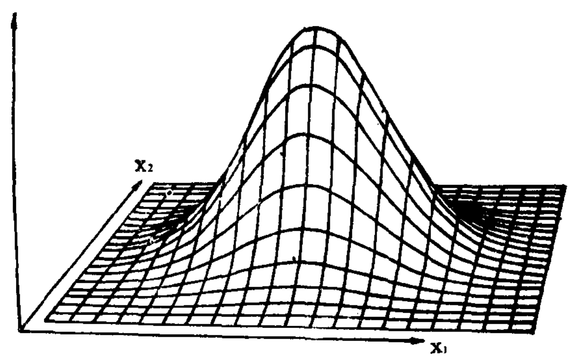

图 1.1.2 两个独立的常态分布曲线（X_1 和 X_2 变数的）联合分布图

$$SP_{jk}=\sum_{i=1}^{n}(X_{ij}X_{ik})-\frac{\sum_{i=1}^{n}X_{ij}\sum_{i=1}^{n}X_{ik}}{n} \qquad (1.1.8)$$

式内 $\Sigma(X_{ij}X_{ik})$ 称做非校正的乘积和。

如果方差可用 SS/(n－1) 来表示，那末协方差就可用 SP/(n－1) 来表示。

$$COV_{jk}=\frac{SP_{jk}}{n-1}=\frac{\sum_{i=1}^{n}X_{ij}X_{ik}-\left(\sum_{i=1}^{n}X_{ij}\sum_{i=1}^{n}X_{ik}/n\right)}{n-1}$$

$$= \frac{n\sum_{i=1}^{n} X_{ij}X_{ik} - \sum_{i=1}^{n} X_{ij}\sum_{i=1}^{n} X_{ik}}{n(n-1)} \qquad (1.1.9)$$

豆粒内和叶子内干物质中蛋白质含量的协方差的计算如下表1.1.3

表1.1.3　豆粒（X_1）和叶（X_2）内蛋白质的协方差

X_1^2	X_1	X_1X_2	X_2	X_2^2
42025	205	26650	130	16900
65025	255	42075	165	27225
38025	195	19500	100	10000
48400	220	29700	135	18225
55225	235	34075	145	21025
$\Sigma X_1^2=$ 248700	$\Sigma_1=$ 1110	$\Sigma X_1X_2=$ 152000	$\Sigma X_2=$ 675	$\Sigma X_2^2=$ 93375

$SP_{12} = 15200 - \frac{(1110)(675)}{5} = 2150$

$COV_{12} = \frac{2150}{4} = 537.5$

此计算值可用程序1.1.3来运算。

```
C       PROGRAM 1.1.3 BICOV
C
C       PROGRAM TO COMPUTE--
C         A. VARIANCE OF EACH VARIABLE
C         B. MEAN OF EACH VARIABLE
C         C. COVARIANCE BETWEEN VARIABLES
C
C       SET SUMS TO ZERO
        SUMX1=0.0
        SUMX2=0.0
        SX1SQ=0.0
        SX2SQ=0.0
        SX1X2=0.0
C       READ NUMBER OF SAMPLES TO BE USED
        READ (5,1000) NS
        DO 100 I=1, NS
C       READ A SAMPLE AND ADD TO SUM
        READ (5,1001) X1, X2
        SUMX1=SUMX1+X1
        SUMX2=SUMX2+X2
```

```
      SX1SQ = SX1SQ + X1*X1
      SX2SQ = SX2SQ + X2*X2
      SX1X2 = SX1X2 + X1*X2
  100 CONTINUE
C     PRINT SUMS
      WRITE (6,2000) NS, SUMX1, SXISQ, SUMX2, SX2SQ, SX1X2
C     COMPUTE AND PRINT MEAN, VARIANCE AND STANDARD
        DEVIATION
C     OF VARIABLE X1
      AMEAN = SUMX1/FLOAT (NS)
      VAR = (FLOAT(NS)*SX1SQ-SUMX1*SUMX1)/FLOAT(NS* (NS-
        1))
      STDEV1 = SQRT (VAR)
      WRITE (6,2001) AMEAN, VAR, STDEV1
C     COMPUTE AND PRINT MEAN, VARIANCE AND STANDARD
        DEVIATION
C     OF VARIABLE X2
      AMEAN = SUMX2/FLOAT(NS)
      VAR = (FLOAT(NS)*SX2SQ-SUMX2*SUMX2)/FLOAT(NS* (NS-
        1))
      WRITE (6,2002) AMEAN, VA R, STDEV2
C     COMPUTE AND PRINT COVARIANCE BETWEEN X1 AND X2
      COV = (FLOAT(NS)*SX1X2-SUMX1* SUMX2)/FLOAT (NS*(NS-
        1))
      WRITE (6,2003) COV
      CALL EXIT
 1000 FORMAT (I4)
 1001 FORMAT (2F10.0)
 2000 FORMAT (1H1, 11X, 21H NUMBER OF SAMPLES = .I10,//,
     121X, 12HSUM OF X1 = , F10.2, //,
     210X, 23HSUM OF SQUARES OF X1 = , F10.2,//,
     321X, 12HSUM OF X2 = , F10.2, //,
     410X, 23HSUM OF SQUARES OF X2 = , F10.2, //,
     59X, 24HSUM OF CROSS PRODUCTS = , F10.2 )
 2001 FORMAT (/, 20X, 13HMEAN OF X1 = , F10.2, //,
     116X, 17HVARIANCE OF X1 = , F10.2)
```

```
      26X, 27HSTANDARD DEVIATION OF X1 = , F10.2)
 2002 FORMAT(/, 20X, 13HMEAN OF X2= , F10.2, //,
     116X, 17HVARIANCE OF X2= , F10.2, //,
      26X,27HSTANDARD DEVIATION OF X2=,F10.2)
 2003 FORMAT (/, 2X, 31HCOVARIANCE BETWEEN X1 AND X2
      = , F10.2)
       END
```

程序 1.1.3 与程序 1.1.2 相比是一种发展。程序 1.1.3 可用三组数据之间，即三对数据间协方差的运算，如果表 1.1.2 上的数据有第三种蛋白质含量要与豆粒内和叶内蛋白质含量间进行协方差计算的话，即可以在计算豆粒内 X_1 与叶内 X_2 间有关的协方差 COV_{12} 之后，又计算 COV_{13} 和 COV_{23}。

协方差值高和低可用图示得出。如图 1.1.3 说明两组变数之间存在着高协方差关系。而图 1.1.4 的两组变数间为低协方差关系。尽管两组变数的各自方差与图 1.1.3 上的一样。

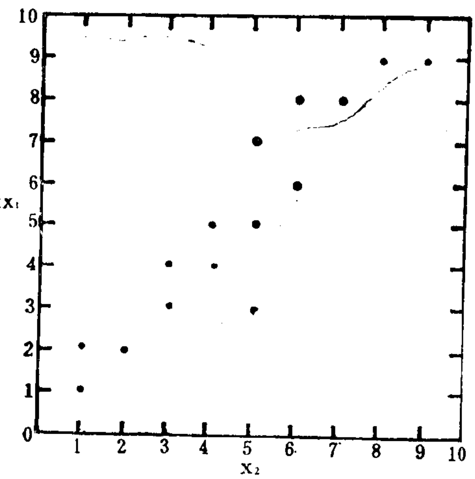

图 1.1.3 具高协方差关系的两组变数的分布情况，X_1 的方差 = 5.7，X_2 的方差 = 7.1，协方差 = 5.9

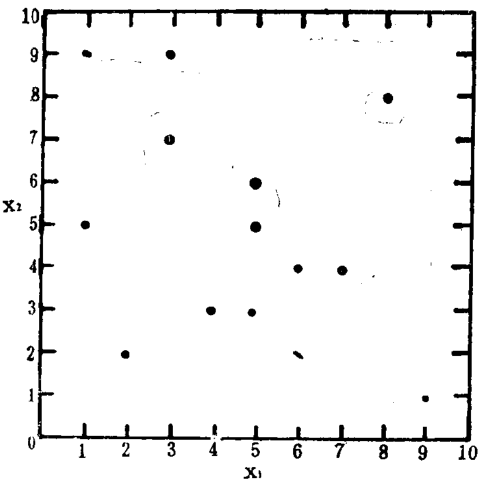

图 1.1.4 具低协方差关系的两组变数的分布情况，X_1 的方差 = 5.7，X_2 的方差 = 7.1，协方差 = -2.3

4. 相关系数

在估测不同度量单位的两组变数间相互关系时，为了不受度量单位的影响，使用相关系数这个统计量。相关系数是两组变数的协方差与这两组变数的标准差的乘积的比，公式为

$$r_{jk} = \frac{COV_{jk}}{S_j S_k} \tag{1.1.10}$$

相关系数是个无单位的值，协方差值不会超过两组变数的标准差的乘积，所以相关的范围处于 +1 到 -1 之间。+1 这个相关系数表示两组变数是完全直接相关，-1 这个相关系数表示两组变数是相互有逆变的关系。在两个值之间是一个不完全相关的景象，其中零相关表示不存在任何线性相关。

在实践上，相关系数的计算用下式

$$r_{jk}=\frac{SP_{jk}}{\sqrt{SS_j\cdot SS_k}}$$

$$=\frac{\sum_{i=1}^{n}X_{ij}X_{ik}-\left(\sum_{i=1}^{n}X_{ij}\sum_{i=1}^{n}X_{ik}\right)/n}{\sqrt{\left\{\sum_{i=1}^{n}X_{ij}^2-\left[\left(\sum_{i=1}^{n}X_{ij}\right)^2/n\right]\right\}\left\{\sum_{i=1}^{n}X_{ik}^2-\left[\left(\sum_{i=1}^{n}X_{ik}\right)^2/n\right]\right\}}} \quad (1.1.11)$$

在一群有机体群落内许多生物学特征都是具有紧密的联系。譬如向日葵叶子的长和宽之间存在着强相关，如对6对叶片的测定（表1.1.4）。

表 1.1.4　向日葵叶片的长和宽

长　　度（cm）	宽　　度（cm）
18.4	15.4
16.9	15.1
13.6	10.9
11.4	9.7
7.8	7.4
6.3	5.3

由此数据我们可以建立下列计算表：

表 1.1.5　向日葵叶片长度和宽度相关计算

	X_1^2	X_1	X_1X_2	X_2	X_2^2
	338.56	18.4	283.36	15.4	237.16
	285.61	16.9	255.19	15.1	228.01
	184.96	13.6	148.24	10.9	118.81
	129.96	11.4	110.58	9.7	94.09
	60.84	7.8	57.72	7.4	54.76
	39.69	6.3	33.39	5.3	28.09
总和	1039.62	74.4	888.48	63.8	760.92

$$SP_{12}=888.48-\frac{(74.4)(63.8)}{6}=97.37$$

$$COV_{12}=\frac{97.37}{5}=19.47$$

$$SS_1=1039.62-\frac{74.4^2}{6}=117.06$$

$$SS_2=760.92-\frac{63.8^2}{6}=82.51$$

$$S_1^2=\frac{117.06}{5}=23.41 \qquad S_1=\sqrt{23.41}=4.84$$

$$S_2^2=\frac{82.51}{5}=16.50 \qquad S_2=\sqrt{16.50}=4.06$$

$$r_{12}=\frac{19.47}{(4.84)(4.06)}=0.99$$