统计知识（二）

张俊杰 主编

需要全本请访问 www.ertongbook.com

目　录

统计小常识（续） 1
统计名言 1
如何阅读《统计公报》 5
什么是季节变动及调整？ 5
什么是“贡献率” 6
CPI 是怎么得来的？ 8
统计方法 9
非线性回归分析技术要点 9
非线性回归分析实例研究 12
模型参数灵敏度分析 21
周期组合模型 24
模糊数学方法 32
灰色系统方法 58
BP 神经网络模型 79
其它统计方法 87
模糊聚类分析 88
模糊模式识别 93
模糊相似优先比方法 98
模糊综合评判 103
模糊关系方程求解 107
综合评判的逆问题 110
中华人民共和国统计法实施细则 114
总则 114
统计调查计划和统计制度 116
统计资料的管理和公布 120
统计机构和统计人员 121
奖励和惩罚 127

附则 129
统计报告案例 129
2003 年安西县社会统计分析报告 129
《绍兴晚报》读者调查问卷的统计分析报告 135
全省中小学校图书馆（室）建设综合统计分析报告（2001 年度） 139

统计小常识（续）

统计名言

至于统计，如果没有坚强的组织，尤其是——这一点在最初的章程中曾专门指出——如果没有总的领导，这项工作是无法完成的。马克思：《所谓国际内部的分裂》《马克思恩格斯全集》第 18 卷第 44 页。

一般的社会统计，特别是经济统计，最近二三十年来作出了巨大的成绩。有许多问题，而且是涉及到现代国家的经济制度和这种制度的发展的最根本的问题，过去是根据一般的估计和大致的材料加以解决的，现在如果不根据某一个一定的纲要收集并经统计专家综合的关于某一国家全国情况的浩繁材料，就无法加以比较认真地研究。尤其是争论最多的农业经济问题，更加要求根据精确的和大量的材料作出回答，况且在欧美各国所有农户进行定期调查，已经愈来愈成为一种惯例。列宁：《现代农业的资本主义制度》(1910 年)《列宁全集》第一版，第 16 卷，第 420 页。

由于分类的方法不同，同一个材料竟得出完全相反的结论。

如果根据土地面积来判断农户规模，结论是随着农户规模的扩大，农业的集约程度在下降；如果根据农户产值来判断农户规模，结论是随着农户规模的扩大，农业的集约程度在提高……。对于这个特别重要的方面，

现时的经济学家和统计学家注意得最不够。列宁：《关于农业中资本主义发展规律的新材料》(1915 年)，《列宁全集》第 1 版，第 22 卷，第 57—60 页。

任何建设工作，任何国家工作，任何计划工作，没有正确的计算是不可想象的。而没有统计，计算是不可想象的。斯大林：《俄共(布)第十三次代表大会》(1924 年 5 月 23—31 日)《斯大林全集》第 6 卷，第 189 页。

计算和监督，——这就是把共产主义社会第一阶段“调整好”，使它能正常地运转所必需的主要条件。列宁：《国家与革命》《列宁选集》第 3 卷，第 258 页。

当然，在形式上，叙述方法必须与研究方法不同。研究必须充分地占有材料，分析它的各种发展形式，探寻这些形式的内在联系。只有这项工作完成以后，现实的运动才能适当地叙述出来。这点一旦做到，材料的生命一旦观念地反映出来，呈现在我们面前的就好象是一个先验的结构了。马克思：(《资本论》第一卷第二版跋)(1873 年 1 月 24 日)《马克思恩格斯选集》第三卷第 217 页。

统计工作不是把数字随便填到几个格格里去，而应当是用数字来说明所研究的现象在实际生活中已经充分呈现出来或正在呈现出来的各种社会类型。列宁：《莫斯科省的工作日和工作年》(1912 年 8 月 12 日)《列宁全集》第一版，第 18 卷第 254 页。

在我生病期间(现在但愿病很快会痊愈)，我是无法写作的，但是，我吞下了大批统计学方面和其他方面的“材料”，对于那些肠胃不习惯于这类食物并且不能把它们迅速消化的人来说，这些材料本身就足以致病。马克思：《致路德维希·库格曼》(1868 年 3 月 6 日)《马

需要全本请访问 www.ertongbook.com

克思思格斯全集》第 32 卷第 526 页。

现在我们看到了一个可叹的事实：地方自治局的统计汇编化费了无穷无尽的劳动，收集了极丰富的、有价值的、新的(实行 11 月 9 日法会的结果！)然而没有加以结算、总结分组和复合的资料，这个汇编只有几乎等于零的一点微不足道的科学价值。列宁：《谈谈关于地方自治局统计任务的问题》(1914 年 1 月《列宁全集》第一版第 20 卷第 74 页。

为了从事实际工作，我们必须掌据数字，而中央统计局应该比谁都更早地掌握这些字。列宁：《给中央统计局的信》》《列宁全集》第 33 卷第 12 页。

我很清楚知道，干巴巴的统计数字是非常不适合作口头报告的，是会把听众吓跑的。但是为了使你们有可能来估价整个运动的真正客观基础，我还是不能不引证一些化成整数的数字。列宁：《关于 1905 年革命的报告》(1917 年 1 月 9 日)《列宁全集》第 1 版，第 23 卷，第 245 页。

我们说，现在我们的任务已经不是剥夺者，而是统计、监督、提高劳动生产率和加强纪律。列宁：《莫斯科省第七次党代表会议》(1921 年 10 月 29 日一 31 日)《论新经济政策》。《列宁全集》第一版，第 33 卷，第 65 一 66 页。

量和质。数字是我们所知道的最纯粹的量的规定。但是它充满了质的差异。恩格斯：《自然辩证法》《马克思恩格斯选集》第 3 卷，第 69 页。

精明能干的经济学家不会去编制毫无意义的提纲，而会去细心研究事实、数字和材料，分析我们自己的实际经验，然后指出：我们在某某地方犯了错误，要如此

这般加以改正。精明能干的行政管理人员一定会根据这种研究，提出建议或自行采取措施，来调换工作人员改变月报制度，改组机构等等。在我们这里还没有看到过有人有这两种切实的态度来对待统一的经济计划。列宁：《论统一的经济计划》(1921 年 2 月 21 日)《列宁全集》第二版，第 40 卷，第 351—352 页或《列宁全集》第 4 卷，第 474 页。

研究政治经济学不能随随便便，不能没有任何基础知识，不能不了解很多极重要的历史问题、统计学问题及其他问题。列宁：《书评》(亚，波格丹诺夫)《政治经济学简明教程》《列宁全集》第二版，第 4 卷，第 5 页。

我们相信中央统计局是科学的保垒。我们认为如果没有中央统计局的数字，任何一个管理机关都不能进行计算和计划工作。我们认为中央统计局应该提供不受任何偏见影响的客观材料，因为使数字适合于某种偏见的企图是一种带有刑事性质的犯罪行为。既然如此，如果中央统计局自己都不相信自己的数字，那么人家又怎能相信它的数字呢?斯大林《联共(布)第十四次代表大会》(1925 年 12 月 18—31 日)《斯大林全集》第 7 卷，第 271—272 页。

统计是社会主义建设的一项重要基础工作。我国要实现工业，农业、科学技术和国防现代化，必须实现统计工作的现代化。(《国务院关于加强统计工作的决定》1984 年 1 月 6 日)。

胸中有“数”。就是说，对情况和问题一定要注意到它们的数量方面，要有基本的数量分析。任何质量都表现为一定的数量，没有数量也就没有质量。我们有许

多同志至今不懂得注意事物的数量方面，不懂得注意基本的统计、主要的百分比，不懂得注意成决定事物质量的数量界限，一切都是胸中无“数”，结果就不能不犯错误。毛泽东：《党委会的工作方法》《毛泽东选集》第4卷，第1380—1381页。

如何阅读《统计公报》

阅读《统计公报》，首先应将几个部分联系起来，不能割裂地单看某一个部分，否则就不能从总体上把握国民经济和社会发展的面貌。其次，要弄清公报中一些概念和术语含义。如什么是第一产业？什么是第二产业？国内生产总值（GDP）到底代表什么含义？等等。第三，最好是将每一年的公报连续对照起来看，从历年数量变化的轨迹中找到规律性的东西，将定量分析与定性分析结合起来，从而在更高层上观察、分析、把握社会经济形势。第四，可以对公报中有关数据进行必要的加工处理，从而了解社会经济生活的种种特色。如可从中计算出全国每天生产多少煤、电，每天完成多少基本建设项目、新建多少公路，每个人每天创造多少收入，每天每时出生多少人口等等，这不仅使得枯燥的数字可以生动化、生活化，也可增添您对生活的量化概念。

什么是季节变动及调整？

在一年之内，由于季节的变动，会使某些社会经济现象（一定的时间序列）产生规律性的变化，这种规律性变化通常称之为季节变动。时间序列的变化会受到社

会经济系统中各种因素的影响。从总体上看，影响时间序列变化的因素一般可以分为三类：第一类为纯粹的时间因素，如气候、日历天数和节假日等有关的季节性因素（用 S 来表示）;第二类是与不同的发展阶段有关的趋势性因素（用 T 来表示）;第三类是经济变量有规律的相互影响，这种相互影响随时间的延续而有规律地发展，表现在时间序列上就是有一定规律的周期波动（用 C 来表示）。另外，还有一些偶然性因素，如社会心理、政治因素等（用 I 来表示）。因此时间序列一般可以表示为：Y=T+C+I+S。为更准确地反映客观经济现象的本质，必须事先对季节变动因素作一定消除和调整。

季节调整就是将一个时间序列分解成以上各部分。目前世界上主要有以下两种季节调整方法；X-11 方法、贝叶斯方法。我国一般采用美国商业部普查局开发的为官方使用并在世界上有着广泛影响的 X-11 季节调整程序。

什么是“贡献率”

在统计分析中经常使用“贡献率”，那么“贡献率”是什么含义？它是怎样计算的？

贡献率是分析经济效益的一个指标。它是指有效或有用成果数量与资源消耗及占用量之比，即产出量与投入量之比，或所得量与所费量之比。计算公式：

贡献率（%）=贡献量（产出量，所得量）/投入量（消耗量，占用量）×100%

贡献率也用于分析经济增长中各因素作用大小的程度。

计算方法是：

贡献率（%）=某因素贡献量（增量或增长程度）/总贡献量（总增量或增长程度）×100%

上式实际上是指某因素的增长量（程度）占总增长量（程度）的比重。

举例说明如下：

总资产贡献率（%）=（利润总额+税金总额+利息支出）/平均资产总额×100%

（1)总资产贡献率 :反映企业资金占用的经济效益，说明企业运用全部资产的收益能力。

（2）社会贡献率：是衡量企业运用全部资产为社会创造或支付价值的能力。

社会贡献率（%）=社会贡献总额/平均资产总额×100%

社会贡献总额包括工资、劳保退休统筹及其他社会福利支出、利息支出净额、应交增值税、产品销售税金及附加、应交所得税及其他税、净利润等。为了反映企业对国家所作贡献的程度，可按上述原则计算贡献率。

企业对国家的贡献率（%）=(税金总额+上缴利润)/社会贡献总额×100%

技术进步对产出增长速度的贡献率

这个指标是指在产出增长速度中，技术进步因素所占的比重,综合反映了技术进步对经济增长作用的大小。

技术进步对产出增长速度的贡献率（%）=技术进步速度/产出增长速度×100%

上式贡献率越大则表明技术进步对经济增长的贡献和作用就越大，反之则小。

（4）各产业贡献率：

第一、二、三产业增量与国内生产总值增量之比，即为各产业的贡献率。

第三产业贡献率=第三产业当年增量/国内生产总值当年增量×100%

应该注意的是，贡献率指标比较抽象，在使用时，应说明具体含义，但也不能任意使用，要符合常规，做到标准化、规范化、通俗化。如资本收益率、资金利税率以及某些对增量因素分析的指标，已有专用名称，就没有必要改称为贡献率。另外，在计算各产业贡献率时应剔除价格变动因素，分子、分母均用可比价格的增量计算。

CPI是怎么得来的？

大多数国家都编制居民消费价格指数（ＣＰＩ），反映城乡居民购买并用于消费的消费品及服务价格水平的变动情况，并用它来反映通货膨胀程度。

从2001年起，我国采用国际通用做法，逐月编制并公布以2000年价格水平为基期的居民消费价格定基指数，作为反映我国通货膨胀（或紧缩）程度的主要指标。经国务院批准，国家统计局城调总队负责全国居民消费价格指数的编制及相关工作，并组织、指导和管理各省区市的消费价格调查统计工作。

我国编制价格指数的商品和服务项目，根据全国城乡近11万户居民家庭消费支出构成资料和有关规定确定，目前共包括食品、烟酒及用品、衣着、家庭设备用品及服务、医疗保健及个人用品、交通和通讯、娱乐教育文化用品及服务、居住八大类，251个基本分类，约

需要全本请访问 www.ertongbook.com

700 个代表品种。居民消费价格指数就是在对全国 550 个样本市县近 3 万个采价点进行价格调查的基础上，根据国际规范的流程和公式算出来的。

统计方法

非线性回归分析技术要点

1.求解参数的操作过程

在进行建模分析之前，用户必须从实际问题出发，确定一个合理的模型表达式。这一点对于从事相关学科的专业人员来讲，根据实验数据、并定义数学模型是不难的。要注意的是，在 DPS 数据处理平台上正确输入相应的公式和参数，操作步骤如下：

首先，在电子表格中编辑数据和定义数据块，一行为一个样本，一列为一个变量，并将待分析的数据定义成数据块。在数据块中，第 1 列为 X_1，第 2 列为 X_2，余此类推，最后一列为 X_p。

定义数据块后,再在屏幕下部文本编辑器窗口中写入数学模型表达式，并用鼠标定义产成公式块。对需要进行模拟和参数求解的数学模型,必须将其定义成系统能够识别处理的形式，第 1 行输入待拟合的数学方程表达式,第 2 行输入待估参数的初始值和加权的变量(或表达式)。

数学方程表达式分两个部分 :等号左边为因变量(或是因变量的计算表达式),等号右边是由自变量和待估参

数组合起来的表达式。公式中待求参数用 $c_1,c_2,\cdots,c_m$ 表示，它们必须从 1 开始按顺序定义。用 $x_1,x_2,\cdots,x_p$ 代表数据矩阵中的各列数据(c、x 大小写字母输入均可)。至如此，一个基本数学模型在本系统中定义完毕。

第 2 行中，先输入各待求参数的初始值，各初始值之间用空格隔开。然后，根据用户需要与否，用“ //= ”引导，输入在进行非线性最小二乘分析时要加权的变量或表达式。当然，第 2 行也可只放入各待求参数的初始值，或只放入由“ //= ”引导的加权变量或表达式，或者什么也不输入(公式块只定义第 1 行)。

公式(包括待求参数的初值或加权变量)按上述格式编辑好之后，拖动鼠标——把公式及其参数初值这两行拖黑。

为充分利用系统资源，在定义公式时需要注意以下事项：

(1)公式块中可直接引用本系统提供的全部标准函数。

(2)表达式右边可允许使用差分变量；

(3)等号左端只能输入表达式,等号右端则可允许输入函数或表达式。

本系统可以分析处理 3000 个数据，512 个样本，求解 20 个未知参数，一般可满足常规研究需要。

2.关于参数初值的定义

如果用户定义的数学模型过于复杂，系统在自动给出参数初值的基础上仍不能完成参数的拟合求解，这时用户需要根据自己掌握的专业知识和研究经验及试验数据的特征，对待拟合的参数给出较为合理的初值，以使得收敛尽快发生，使模型拟合能够顺利进行。对初始值

的确定，有以下几个简单而有用的原则：

(1)分析模型的性质。非线性回归模型的参数对科学家或研究人员来说通常是有意义的。这种意义可能表现在图象、生物学、物理学、化学或其它合适的形式上，因而对确定初始值可能很有用。某些参数的初始值可能从相关试验中获得，也可以根据函数图象，了解参数怎样影响函数的表现。根据模型的基本特征如拐点、极值等特征确定待模拟参数的初始值。生物学中常用的许多曲线方程，如逻辑斯蒂方程，方程本身含有参数的极限值，其大体范围是不难确定的。若是最大值，欲设定的初值可比观察值中的最大值略大一些;如是极小值,待设定初值可比观察值中的最小值略小一些。

例如在Michaelis-Menten酶反应模型 $f=\theta_1 x/(\theta_2+x)$ 中，参数 θ_1 是酶反应的渐近速度，因而能够由试验数值的最大观察值估计；类似地，θ_2 表示“半浓度”，即这样一个 x 值，当浓度达到该值时，速度是最大值的一半。

(2)分析一下，是否可对模型作适当转换，以获得更简单的关系，或使之线性化。将模型线性化的一般方法是取对数，如方程中有指数项，可对方程两边同时取对数，使方程线性化后，再对该线性方程进行计算以获得初始值。如前述柯布—道格拉斯方程两边取对数后可转化为线性方程。

例如，Bard(1974)在研究一个化学动力学的实例时应用了模型 $f=\exp(-\theta_1 x_1\exp(-\theta_2/x_2))$。对此模型，取 2 次对数后可得到 $\ln(\ln(f))=\ln x_1+\ln(-\theta_1)-\theta_2/x_2$。该模型是一个线性模型。这时可应用线性最小二乘法来获得初始值。

(3)简化原方程。这意味着将方程从简到繁逐步求

解。有些在生物学研究中应用的模型是由一些较简单(基本)的方程演化而来的。例如从逻辑斯蒂方程衍生出来的模型有：

$$Y=\frac{K}{1+e^{a+b_1x+b_2x^2}}$$

$$Y=\frac{K}{1+\exp(-r(T-T_0))}(1-\exp(-\frac{T-T_1}{\delta_1}))(1-\exp(-\frac{T_M-T}{\delta_M}))$$

对此类模型，可先求前面的基本部分，然后再逐步增加，最后求出完整的参数。

(4)在采用以上几个方法都不能得到参数的精估计值时，可采用网格法寻求参数的初值，即对每个待估计的参数给出一个区间范围和搜索步长，进行计算，该法计算量较大，但在运算速度很快的计算机上是不难完成的。

非线性回归分析实例研究

1．普通非线性模型

这类模型一般不含指数或复杂的数学函数，而仅仅是一些加减乘除的组合。这类模型建模时，只需要将试验数据编辑定义之后，将希望建立的数学模型(公式)进行编辑定义，一般无需给出各个参数的初值(因系统默认各个参数的初值为 0.01)就可以使用麦夸特法快速地求出模型中的各个参数。

例 1 研究“岱字棉”自播种至齐苗(以 80%出苗为准)期的天数(*Y*)和日平均土温(*X*,℃)的关系，经试验得到数据后欲建非线性经验模型(莫惠栋 1984)。根据有效积温模型，描述自播种至齐苗期天数和日平均土温相互关系

最直观的回归方程的数学表达形式为

$Y=a/(X-b)$(21.16)

数据和公式块的定义见图 21–1 所示。图中上部为数据区，下部为公式区，数据块以灰底蓝字显示，公式块以黑底白字（反相）显示。

由于因变量 Y 位于数据块的第 2 列，自变量 X 位于数据块的第 1 列，因此令 $x_2=Y$，$x_1=X$。因为式(21.16)中有两个待求参数 a 和 b,定义模型(方程)时则分别用 $c1$ 和 $c2$ 表示，最后定义公式块,见图 21–1 下部（黑底白字）反相显示部分。



图 21－1 播种至齐苗天数和日均土温相互关系模型分析

编辑定义数据和公式块后，进入菜单，选择“数学模型→单因变量参数估计→麦夸特方法”后，按回车键即刻进入分析提示界面，在屏幕的右偏下部分显示当前各参数估计、拟合误差的动态信息，并有文字提示。如果迭代结果已达到要求,或用户需要提前终止迭代计算，

可点击第一排快捷按钮最右边的那个按钮（该按钮只有在进行计算时才显示出来）。在本例中，用麦夸特迭代模拟方法成功地给出拟合结果（图 21－1）。

如图 21－1 所示，返回的计算结果包括方差分析表、系数 c_i 的协方差阵（略）、系数 c_i 的相关阵（略）、参数拟合值及其标准误差、t 测验值和相应的显著水平，最后给出各样本因变量的观察值、拟合值、拟合误差、标准残差、cook 距离和杠杆率 H，供模型诊断使用。

从以上拟合结果看，F 检验值为 1108.34，达极显著水平；相关系数为 0.9983，拟合度 99.64%，均优于经线性化转换后的拟合效果。将拟合结果代回原方程，有 $Y=\frac{50.7517}{X-14.6138}$。如果进行区间估计，岱字棉从播种到齐苗的发育起点温度为(14.61±0.19)℃，有效积温为(50.75±2.36)℃。

如果求置信概率为 95%的起点温度，需先求出自由度为 4 时的 $t_{0.05}$ 临界值，可在电子表格中键入“＝ttest(4,0.05)”，而后立即显示结果是 2.7764，由此求得置信概率 95%的发育起点温度为(14.61±2.7764×0.19)℃，有效积温为(50.75±2.7764×2.36)℃。

2 含有指数或某些函数的非线性模型

此类非线性模型因含指数或其它数学函数，往往不宜或无法转换为线性模型求解。因此对这类模型的参数估计，必须借助于非线性的迭代方法(如 Newton-Raphson 方法)求解。这对不十分熟悉或精通数理知识的生物学工作者来说是很比较困难的。但是，在我们提供的 DPS 处理平台上，用户只需将试验数据和希望建立的数学模型(公式)编辑定义成数据和公式块，便

需要全本请访问 www.ertongbook.com