

统计论文选编

1 9 8 0

沈阳市统计学会

PDG

前 言

为推进我市统计学术研究，开展学术交流，逐步提高统计科学理论水平和实际工作的水平，更好地为“四化”建设服务，我们选编了统计学术论文。共二十四篇。

这些论文，是沈阳地区有关高等院校、中专理论教育工作者和从事统计工作多年具有实践经验的统计学会会员撰写的。内容有综合、农业、工业、财贸、劳动工资、物资、基建、交通运输、卫生等各专业统计方面的文章。

由于时间所限，广大学会会员已撰写的学术论文，尚未全部编入，今后拟将不定期地继续选编这样的材料，以供学会会员，统计理论和实际工作者研究参考。

沈阳市统计学会

一九八〇年九月

目 录

统计方法论的探讨·····	沈阳农学院	谭启栋 (1)
马克思主义再生产理论与最终产品·····	辽宁大学 张幼华	廖群笙 (15)
加强工业企业统计分析的科学性·····	沈阳气体压缩机厂	邓子善 (29)
地区国民经济综合平衡统计浅见·····	沈阳市统计局	宋克辉 (37)
国民生产总值计算方法初步探讨·····	沈阳市统计局	刘年金 (47)
职工生活费指数编制方法浅见·····	沈阳市统计局	曹伟刚 (52)
对统计监督的粗浅认识·····	沈阳市财经学校	李文选 (63)
关于农村社队考核指标的商榷·····	沈阳农学院	姚普祥 (68)
对反映农村致富统计指标的探讨·····	沈阳市统计局 顾树德	杨洪祺 (73)
工人健康状况的评价方法·····	中国医科大学 丁道芳	刘延令 (77)
评价工业企业经济效果的主要指标·····	沈阳风动工具厂	曹锦逊 (86)
机械制造企业统计综合分析方法·····	沈阳重型机器厂	顾振光 (90)
有色金属冶炼金属回收率计算方法探讨·····	沈阳冶炼厂	李鸿勳 (98)
关于工业净产值核算及时性的探讨·····	沈阳线材厂	佟一勳 (106)
工业净产值的核算根据和方法问题浅见·····	沈阳市统计局	苏 萱 (115)
关于社会商品零售额包括范围问题的探讨·····	沈阳市统计局	周玉生 (127)

试谈基本建设投资效果分析·····	沈阳市统计局	姚昭文(130)
关于建筑工程量计算方法的探讨·····	沈阳市建工局	吴风鸣(133)
对以净产值计算劳动生产率的探讨·····	沈阳市建工局	徐俊德(143)
对现行劳动工资统计人员分类的改革意见·····	沈阳市建二公司	范传新(151)
试论生产定额统计·····	沈阳市财经学校	许世海(163)
物资统计资料搜集、积累、管理与利用·····	沈阳变压器厂	刘政兴(171)
浅谈运输统计预测·····	沈阳市统计局	赵国魁(180)
论抽样调查在城乡人民生活调查中的运用·····	沈阳市统计局	吴秉林(192)

统计方法论的探讨

沈阳农学院 谭启栋

现在的统计工作包括搜集资料，整理归纳与统计分析决断，是基于概率论与数理统计理论基础上的。概率论与数理统计实际是来源于统计实践。在我国人口，土地，农情，军事等统计有二千多年历史，但概率论只有二百多年。数理统计又是在概率论基础上发展起来的，它成为一个有体系的学科还不到一百年。从概率论而言，由于抽象而简化，古典概型与扔硬币及掷骰联系起来。由于当时欧洲社会风气而参进了赌博中的某些条件与要求。这些东西不仅保留在概率论中，它也插入到数理统计及统计工作中。当然这些理论促使近百年来的统计理论与统计工作有了极大发展。从我国二千多年统计实践经验与目前我国群众要求来说，统计方法有进一步探索的必要。我们认为在一定统计总体与工作要求前提下，建立一种中国的，群众的，科学的统计方法，不仅是必要的，也是可能的。

拿典型调查来说，就是我国运用最久而行之有效的一种统计方法。我国的农情估算，产品检验等，群众都是采用这一方法。不仅已用了二千多年，而且至今仍在采用。主要特点是由有经验的工作人员选择具有代表性的少数典型单位。进行实测后估算总体。误差一般在5%左右而不超过10%。可是若按数理统计的抽样理论说，那是不符合不参加人的意志而随机抽取调查单位原则的。可千百年的实践已证明了中国典型调查是既经济而有效的统计方法，我们在千百年实践的经验基础上建立起科学的统计理论是完全必要的，也是我国统计工作者义不容辞的使命。我们在1979年《沈阳农学院学报》上发表的“典型中心抽样的理论与分析”就是在这一思想基础上研讨与实践写

出来的。

从我国目前群众要求说，尤其是广大农村，不论是生产上，科学实验上都需要有科学的统计判断，可是若按统计工作说，要经过调查（实验）、整理资料、统计分析三个步骤，不仅有繁琐的设计，计算分析，还必备一些必要统计用表。这不是一般社队科研人员能短时运用裕如，而不仅知其然还知其所以然的。群众要求我们的统计理论工作者可否设计出一套取样极少，取到结果后用不着任何分析计算就可作出统计决断呢？这实际是要求我们设计出寓分析于实验设计的最少实验单位的方案。我们认为这一群众要求是合理的，也是我国目前四个现代化群众科研长征途中的必需！这篇东西在这两个方面介绍两个方法，这仍只是开始。

无分析计算的 8—1 试验设计方案

在任何科学试验取得数值结果后，必须经过整理归纳，再采用适当的统计分析方法进行计算作出决策判断。有些试验在试验前还需按要求误差算出取样的单位数。这些统计方法的选择与计算，对社队常用的简单对比科学实验说，不仅是繁杂而且是不易为社队科研及生产人员所完全理解。有些很有前途的科学实验结果，由于没作出统计分析判断而失去了科学价值。我们认为群众需要一种没有计算寓分析于实验的试验方案，只要得出试验结果后即可作出科学的统计判断。我们认为不同条件有不同的无分析试验设计。现提出简单对比试验的 8—1 无分析试验方案，我们认为它不仅试验个体少，而且决断方法也极简。具体作法及论证分别写在下面：

具体实施步骤这种试验方案是实用于一种新措施或技术与常用较优措施，进行对比试验。

实验安排步骤，设一新育或引进品种，要判断它是否优于原常用品种，试验地要如下安排。根据被实验作物的特点可选择下列两种实验地的一种进行实验。

图1：条形方案

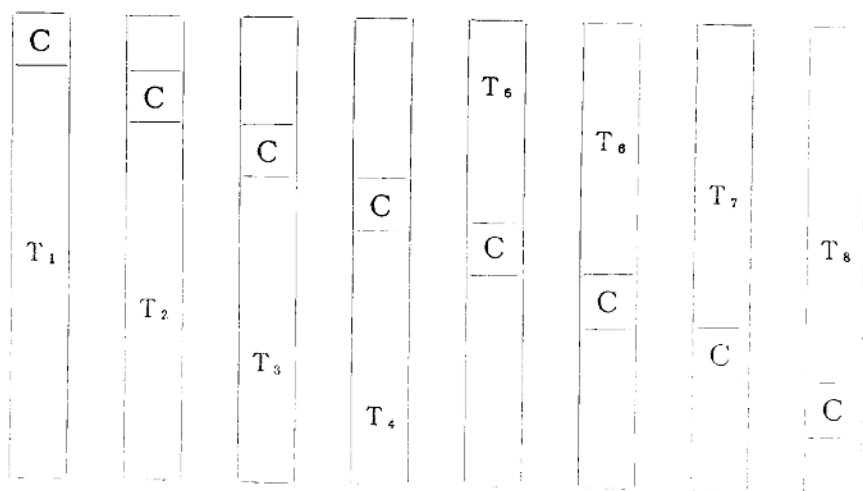


图2：方形方案

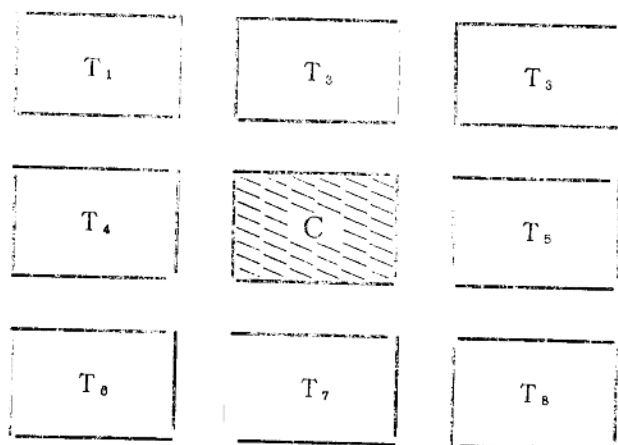


图1及图2中的C是原措施的实验地作为一个计算单位, $T_1, T_2, \dots, T_8$ 为试验措施的实验地, 单打单收分别为8个计算单位, 由此可得1个C的数值8个不同的 T_i 值。

统计判断在上试验地的一个C值及8个 T_i 值中, 若8个 T_i 值完全高于(或优于)C值, 则即可作出我们试验的措施是优于原措施的判断。它的显著

水准是 5%，也就是说新措施有 95% 的把握优于原措施。若 8 个 T_i 有 1 个低于 C 值，按 5% 水准就不能说新措施优于原措施。

这种试验方案不仅实用于可计量的实验，对于质的比较也可以根据质的差别如色彩、程度等，进行比较后也可作出决断，只要优劣是可分的都可用这一方案进行实验。

理论依据 8—1 的比例关系是怎样来的？为什么只看 8 个处理结果比对照为优，不经分析即可作出判断呢？下面提出它的论证。

我们用农业科研上常用的 5% 显著水准及无效假设方法来说明。

假设新措施与原措施效果是相等（来自同一总体）的，则新措施处理的 8 个试验与对照结果的优劣从理论上说应是相等。新措施值大于对照的概率应是 $1/2$ ，也就是说 8 个处理中有一半大于对照，一半小于对照。如果证明了新措施与对照相等的可能小于 5%，则我们就可判断新措施效果不是与对照相同，也就是说新措施有 95% 把握优于原措施。

请看下面：

设原措施与对照效果相等，则

$$T_i < C \text{ 之概率 } P, \quad P = \frac{1}{2}$$

$$\because \text{二项分布是 } \sum_{r=0}^n \binom{n}{r} p^r (1-p)^{n-r}$$

$\therefore$ 在 n 个事件中出现 r 个优的概率是 P_r ,

$$P_r = \binom{n}{r} p^r (1-p)^{n-r}, \quad r = 0, 1, 2, \dots, n,$$

$$\text{在 } P = \frac{1}{2} \text{ 时, } P_r = \binom{n}{r} \left(\frac{1}{2}\right)^r \left(1 - \frac{1}{2}\right)^{n-r} = \binom{n}{r} \left(\frac{1}{2}\right)^n$$

在我们试验设计中, $n = 8$ 、 $r = 8$

$$\therefore 8 \text{ 个 } T \text{ 皆优于 } C \text{ 之概率是 } P_8 = \binom{8}{8} \left(\frac{1}{2}\right)^8 = \frac{1}{256} = 0.00396$$

这是微小样本它的误差是比较大的，其抽样误差 E 是：

$$\therefore E = \sqrt{\frac{p(1-p)}{n}}$$

$$\therefore = \sqrt{\frac{0.00396(1-0.00396)}{8}} = 0.0222$$

因此出现 8 个 Ti 优于 Ci 概率是

$$P = p^8 + 2E = 0.00396 + 2 \times 0.0222 = 0.04837 < 0.05$$

根据显著水准，原假设新措施与对照效果相等是无效的，因此判断新措施有 95% 概率优于对照，也就是说新措施优于原措施。通过这一证明也说明了 8—1 方案是可信的。

再谈典型抽样

典型抽样我们在 1979 年沈阳农学院学报中已提出了这样几点认识：

1、典型调查是取样小、经费省、人工少、花时短，精度好的一种调查方法。

2、典型调查可用于“知其概”的社会、经济、自然事物的调查总体。对于目前卫星遥感技术而言，对于一般静态总体知其概是不难的。

3、典型调查是我国千百年用之有效的一种调查统计方法，由它的生命力就说明了它的可靠性。

4、典型调查是在人选择后的一种抽样，并不是“主观调查方法”。

根据上述几点认识，我们提出了典型中心抽样方法。

我们所提的典型中心抽样是依据各地历史数据或某种估测结果作为“概知”。因此这类概知是各地区的历史或估测平均值，由此可得出总体平均值，或各地区的数量大小排序。由于这一概知条件我们提出了以概知的总体均值为中心或以中位值单元为中心的典型抽样方法，因此也可称为均值与中位值典型抽样。

实际情况是大量总体的大小顺序与均值是不知其概的，因此前述的方法就用不上了，在不少情况下，由调查总体中找出常见的单元及稀少与小的单元是完全可能的。根据这种概知条件我们提出它的典型调查方法与论证。我们可称这一调查方法是典型众数抽样法。我们在下面来探讨这一问题。

典型众数抽样步骤及其计算，凡是能辨别统计总体的单元，那些是常出现的或“中不溜”的单元，并可找出最大值（或较优）与最小值（或较差）的单元的社会经济或自然现象的统计总体，都可用这一典型抽样方法进行统计估计与分析，它的具体方法与步骤分述如下：

这种典型抽样是在调查总体中，只抽取6个调查单元，必要时只抽取3个单元即可作出总体的统计估计，在一些调查单元中还有必要进一步典型抽取6个或3个实测单位进行实测。因此我们也可称这一典型抽样为“六六典型抽样法”。现将其抽选调查单元的具体步骤说明如下：

1、在调查总体中抽取典型常见的或“中不溜”的单元4个，分别记作 b_1 、 b_2 、 b_3 、及 b_4 ，在某种条件限制下，也可只抽选一个常见（或中不溜）单元记作 b 。

2、在调查总体中尽可能的找出一个最差的单元记作 a ，再找出一个最好的单元记作 c 。

3、在必要时，可在 a 、 b 、 c 、中再找出二层的实测的 a 、 b 、 c 。

在 a 、 b 、 c 值实测得到后，就完成了调查工作，下一步乃是内业的计算与统计分析了。

4、下面介绍几个基本统计指标的计算

设： A = 典型众数抽样平均值。

b_1 、 b_2 、 b_3 、 b_4 = 调查总体中抽选4个常见单位的值。

b_s = 在某种条件限制，只抽取到一个常见单元的值。

a = 调查总体中可找到最差的单元值。

c = 调查总体中可找到最好的单元值。

V = 方差。

B = 标准差。

则可列计算式：

$$A = \frac{a + b_1 + b_2 + b_3 + b_4 + c}{6} \dots\dots (1-1)$$

$$A = \frac{a + 4b + c}{6} \quad (1-2)$$

方差V与标准差B为：

$$V = \frac{c - a}{6} \quad (1-3)$$

$$B = \frac{c - a}{6} \quad (1-4)$$

现用一实例说明它的调查与计算

例1、现要测某林场某种树木的平均胸高直径，经林场组织一实测小组，调查6株典型立木径实测是：

找到最小胸径立木的胸高直径值 $a = 15$ 公分，

四株“中不溜”立木胸径分别是

$$b_1 = 26 \text{公分} \quad b_2 = 28 \text{公分}$$

$$b_3 = 29 \text{公分} \quad b_4 = 30 \text{公分}$$

找到最大胸高直径立木测得 $C = 44$ 公分

试求这林场中某种立木的平均胸高直径及标准差。

由实测结果依上计算公式可得：

$$A = \frac{15 + 26 + 28 + 29 + 30 + 44}{6} = \frac{172}{6} = 28.67 \text{公分}$$

$$B = \frac{44 - 15}{6} = \frac{29}{6} = 4.83 \text{公分}$$

由典型抽样过程可看出它与随机抽样的根本差别是随机取样是由随机样本总体去估计统计总体，而典型抽样是由统计总体中抽选单元，而直接估算总体，换句话说随机取样是由总体中随机抽取部分形成一个样本，由样本再去估计总体，而典型抽样是由统计总体，通过人的认识，由总体中选抽能代表各部分的单元而直接计算统计总体。因此它的误差小于同大小的随机抽样。

根据上述理由及典型抽样实践经验我们可得到它们分析指标的计算式。

总体平均值 μ 有95%概率在:

$$\mu = A \pm A/10\text{之间。} \dots\dots\dots (2-1)$$

据此上例中林场所调查总体之平均胸高直径为95%可能在:

$$28.67 \pm 28.67/10\text{厘米之间,}$$

$$\text{即} 25.803 \text{至} 31.537 \text{厘米之间。}$$

在两种处理的实验地, 或两个不同品种、地区某类事物进行差异分析时, 也可用这一调查结果进行计算后判断。

设两种处理之典型抽样单元值, 平均值及标准差分别为:

第一种处理为: a_1, b_1, c_1, A_1, B_1

第二种处理为: a_2, b_2, c_2, A_2, B_2

计算:

$$d = |A_1 - A_2| / \frac{A_1 + A_2}{20} \dots\dots\dots (2-2)$$

在 $d < 1$ 时, A_1 与 A_2 的差异是显著的 (水准为95%)

在 $b < 1$ 时, A_1 与 A_2 的差异不显著。

下面用一个例说明这一分析步骤

例2、某生产队进行高粱育种对比试验, 在这两品种地块中进行了典型抽样, 每一典型抽样单元是长3米的一条垅, 垅宽都是0.6米, 经实测典型取样结果如下表, 试判断两品种产量的差异是否显著?

表1 两品种对比试验典型抽样实测值

单位: 斤 垅宽0.6米、长3米

典型抽样	甲品种	乙品种
a	3.7	2.9
b	3.1	2.2
c	2.1	1.8

由上实测取样可得：

$$\text{由 (1-2) 式: } A_1 = \frac{3.7 + 4 \times 3.1 + 2.1}{6}$$

$$= \frac{18.2}{6} = 3.03 \text{ 斤,}$$

$$A_2 = \frac{2.9 + 4 \times 2.2 + 1.8}{6}$$

$$= \frac{13.5}{6} = 2.25 \text{ 斤.}$$

由 (2-2) 式计算 d 值：

$$d = \frac{\frac{3.03 - 2.25}{\frac{3.03 + 2.25}{20}}}{\frac{0.78}{0.264}} = 2.95 > 1,$$

$\because d > 1$, $\therefore$ 甲品种产量高于乙品种，它们之间的差异是显著的。

根据上述设想，其它分析指标，有些也可建立起来，这留待以后去探讨。

上面的计算方法是简单而行的，可是它有何依据呢？这是我们关心的，也是下面要谈的。

理论依据，典型众数抽样的基本问题是依照：差：中（常见）：优 = 1：4：1 的比例进行选取典型单元，这一比例关系是美国“化非肯定型为肯定型”方法中提出的。华罗庚先生在《统筹方法平话》中介绍过。根据华先生介绍在美国用 β 分布作过证明，但并未说清这一比率关系。华先生用加权平均数后求平均的方法，说明了平均数及方差的公式，现引述如下：

设：a = 保守值，C = 最乐观值，b = 最可能值，

$$\text{则：平均值 } A = \frac{a + 4b + c}{6}$$

$$\text{方差 } V = \left(\frac{c - a}{6} \right)^2$$

华证：因 a 是保守值估计，b 是最可能估计，

用加权平均求其平均值为： $\frac{a + 2b}{3}$

同理得最乐观值与最可能值平均为： $\frac{2b + c}{3}$

∴上两平均值之平均为调查总体平均。

$$\text{故： } A = \frac{\frac{a + 2b}{3} + \frac{2b + c}{3}}{2} = \frac{a + 4b + c}{6}。$$

同理可得方差V为：

$$\begin{aligned} V &= \frac{1}{2} \left[\left(\frac{a + 4b + c}{6} - \frac{a + 2b}{3} \right)^2 \right. \\ &\quad \left. + \left(\frac{a + 4b + c}{6} - \frac{2b + c}{3} \right)^2 \right] \\ &= \left(\frac{c - a}{6} \right)^2。 \end{aligned}$$

$$\text{显然可得： } B = \frac{c - a}{6}。$$

这一证明说明了乐观值、保守值与最可能值间的1 : 2权数关系，由此说明了1 : 4 : 1的比例。

由上面的介绍可以看出我们的典型众数抽样计算就是循此而来的，我们认为化非肯定型为肯定型的实质就是我们所定义的典型抽样，因为最乐观值及保守值就是我们称为概知的典型最优最差单元值，所谓最可能值就是典型的常见单元值，我们认为从正态分布也可说明：差 : 中 : 优 = 1 : 4 : 1的比例关系。

典型抽样的调查总体要求是近似正态分布的，正态分布式为：

$$Y = \frac{1}{B\sqrt{2\pi}} e^{-\frac{(x - \bar{x})^2}{2B^2}}$$

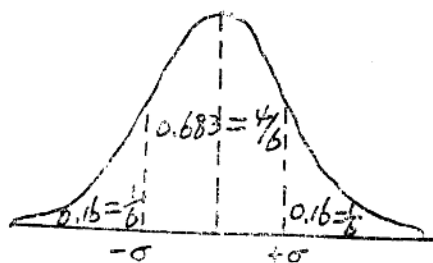
式中B为标准差， $\bar{x}$ 为x的平均值。

$$\text{以 } \bar{x} \text{ 为点原，以 } B = 1, \quad \text{使 } y = Y / \frac{1}{B\sqrt{2\pi}}$$

则上式简化为： $y = e^{-\frac{x^2}{2}}$

正态曲线图形如下：

图 3 正 态 分 布



由正态分布曲线可求得在 $X = \pm B$ 时为拐点，求上式各阶导数：

$$Y' = -X e^{-\frac{x^2}{2}}$$

$$Y'' = -e^{-\frac{x^2}{2}} + x^2 e^{-\frac{x^2}{2}}$$

$$\text{令 } Y'' = 0 \quad \text{即 } -e^{-\frac{x^2}{2}} + x^2 e^{-\frac{x^2}{2}} = 0$$

$$\text{则 } x^2 = 1$$

$$\therefore x = -1, \quad x = 1,$$

$$\text{亦 } x = -B, \quad x = B,$$

由其三阶导数可验证 $X = -B$ 及 $X = B$ 均为拐点，说明在 $X \pm B$ 之间 Y 的变化率是较小的，因此在选取后抽样，不论何种中心典型抽样，它都是最大可能落在这一区间，据此众数取样在6个中取4个常见单元或取一个单元给以4的权数就是这个道理。

$$\text{因为 } \frac{4}{6} = 0.667 \approx 0.683$$

在 ± 6 以外各取一个单元作为较优及较差的典型，它们各占

$$\frac{1}{6} = 0.1667 \approx 0.16$$

这是我们从分布上认识 差：中：优 = 1：4：1 的关系。

为什么典型抽样平均数加减 $A/10$ 就可估计总体平均数呢？

在我国群众的估量一般是较准确的，如农产量估产老农看过庄稼以后一般说法是产量可达七、八百斤，这就是说不得少过700斤，也不得多800斤，这与我们抽样调查后的估计区间是一致的，因此根据历年经验，典型抽样的误差不会超过10%，因为是估计我们只可说有95%的可能，由此得估计总体均值经验公式为有95%可能在 $A \pm A/10$ 之间，同理典型抽样精度是90%。

典型抽样应用示例农村人民公社进行某种作物的品种产量试验，在收割前要进行预估单位面积产量，问用典型抽样如何进行估产？

这种调查要求只是单位面积产量，不求这一试验地的总产，因此可不管试验地大小。

在试验地中找出最好的地段，多数情况的代表地段，及最差的地段。

设试验地为垅作，经测垅宽为0.6米，

在选好的三个地段中选取最可代表的一条垅实测3米长的各一个典型抽样，并进行实测得其结果如下表：

表2 典型抽样实测结果

取样 $3 \times 0.6 \text{米}^2$ 垅

样 别	取 样 产 量 (斤)
a (最好的)	3.7
b (一般的)	3.1
c (最差的)	2.1

依 (1-2) 式算出每垅样平均产量为

$$A = \frac{3.7 + 4 \times 3.1 + 2.1}{6} = 3.03 \text{斤}$$

$$\therefore \text{每亩} = \frac{10000}{15} = \frac{2000}{3} \text{米}^2$$

由上计算结果可得平均每米²产量为

$$\text{每平方米产量} = \frac{3.03}{3 \times 0.6} = 1.68 \text{斤}$$

由此得每亩产量为

$$\text{亩产} = \frac{2000}{3} \times 1.68 = 1,120 \text{斤}$$

由此得垅样单位面积产量为

$$\text{亩产} = \frac{2000}{3} \times \frac{\text{平均垅样产量 (斤)}}{\text{垅长} \times \text{垅宽 (米)}}$$

上例为了计算方便，只是在一块试验地中估计产量，若是一个公社的估产则可进行两级典型抽样，那就是说在公社各大队中先选出优、中（多数）、劣的三个大队，在中等大队中抽选出常见的（中不溜）4个地块，在最优及最劣的大队中各再选出最优最劣的地块各一，然后在抽样地块中实测后，分级计算最后得出全公社的单位面积产量。当然全县也可仿此进行估计，其它社会、经济、自然事物，只要是符合这种概知条件的，都可估计总体。

小 结 与 说 明

根据我国的历代实践，中国的典型调查是一种调查单位很少，精确度能达到要求的调查方法。也可以说是一种既经济、也易为群众接受的调查统计方法。我们认为根据千百年调查经验来总结出一个调查分析体系，不仅是必要的，也是可能的，我们认为统计抽样中排除人的认识，在抽取调查单位时的作用是相对的，是有其局限性的，把它绝对化了会增加过多的调查单位，而造成不必要的浪费及复杂化，我们认为合理的运用人的认识，在统计调查工作中，是减少调查单位及具有经济效益的最佳措施。调查方法由随机抽样到分层系统抽样，就是很好例证。我们也认为，把人对客观的认识参加到调