

农产量多目标抽样方案的探讨

张尧庭 张璋 梅江

自1984年以来,对称等距抽样法以其独特的优点和性能受到广大使用单位的青睐;但是任何一种抽样方法都不是十全十美的。对称等距抽样由于只考虑了一个指标,因而产生了粮食作物分季分品种的产量难以统计准确的问题;另外由于每层中只抽了一个样品,因而使得样本的代表性相对较差。为了弥补这些不足,我们以聚类分析和回归分析为工具,首先对粮食作物各项指标进行了反复筛选。在此基础上又对湖北省各县市的粮食结构做了进一步分析,从而产生出较好的抽样方案,经分析比较,其优点是十分明显的。

抽样调查的目的很少是只了解一个指标的情况,一般都是要了解调查对象的很多方面、很多指标。以农产量的抽样调查为例,不仅要调查全年粮食的总产量,还需了解小麦、早稻、中稻、晚稻、大豆等各项作物的产量情况,也需了解夏粮、秋粮的情况。因此在选定的一套抽样点上能否同时满足上述各项指标调查的要求,这就是多目标抽样调查方案研究的问题。

在论述、介绍具体的方法之前,我们先对一些名词术语作一些规定和解释,然后对抽样方案好坏的比较标准作一些建议和说明。

指标(标志值):抽样调查中被调查的量,如总产量、平均亩产量……等,指标也称为变量。

样本点:被抽中作为样本的单位。以农产量调查为例,把县作为单位,则被抽中的县就是样本点。

一套样本:按一定的规则所得的全部被调查的对象称为一套样本,一套样本中的单位就是样本点,一种规则可以获得许多套样本。如对称等距抽样法就可以获得若干套样本。

其他的一些术语和平常使用的意义相同,不另作说明,如总体就是被研究的全部对象,总体中的每个个体称为单位。

一、引言

怎样来获得较好的改进方案呢?必须针对原有方案的问题指出改进的方法,才有可能达到目的。

从对称等距抽样方案的具体方法来看,按全年粮食亩产量排队,按播种面积累计,实际上相当于考虑全年粮食产量,因此全年的代表性很好(每年相对误差均不超过2%),然而

对其他的指标就有问题：夏粮偏低（-7%以上），小麦偏低（-14%以上），……等等。这原因是分组中的不合理造成的，对称等距抽样是每一组中只取一个样，因此这个样的各项指标必须和本组其他单位的各项指标很相近才能有代表性，如果只有一部分指标是相近的，那就无法反映其他指标的值。因此必须在分组上先想办法改进，这是第一点。另一方面，按亩产量的大小排队时，如从大到小排列，则分组后第一组内亩产量都很大，而最后的几组亩产量都很低。容易看出：低产的那些组，组内各单位的值相差不可能大，也即方差较小，而高产的组，组内的差别大，也即方差大。然而每组只取一个样，对高产的组代表性就差，对低产的组又嫌样本点多了（相对而言），不符合抽样数应与方差成比例的抽样原则。这样一分析，我们就从两方面着手改进，先聚类，分选出特殊的单位，然后再决定样本点。但是，在做聚类分析时，聚类中第一项内容就是按哪些指标进行聚类。指标越多，如仍要求每一类之内相对差别很小，就必须分成很多类。每类之中只有少数几个单位，这样就失去了聚类的作用；如果指标太少，虽然聚类效果好了，但是没有被考虑的指标在类内是否会差别太大，这样也不能达到多目标的目的。因此要将被调查的这么多指标进行适当的筛选，然后得出比较合适的一组聚类指标。我们根据这一特点，首先进行了指标筛工作。

二、指 标 筛 选

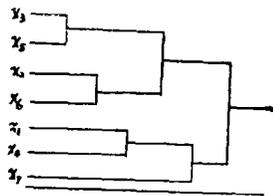
要能够真正筛选出主要指标，就必须对指标的结构有一个清楚的了解，切实弄清各指标间的相互关系。为此，我们的筛选工作，分两个步骤进行。

2.1、指标结构分析

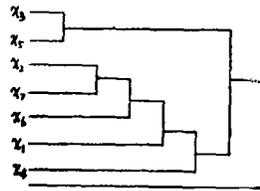
湖北省粮食作物的指标共有十项，其中三项为综合指标：全年粮食总产量、夏季粮食总产量、秋季粮食总产量（以下简称全年粮、夏粮、秋粮），这是统计部门非常关心的；另外七项为实割实测的单项指标：小麦、洋芋、早稻、中稻、晚稻、玉米、大豆。现要从这七个指标中筛选出部分主要指标，并能较全面地反映三个综合指标的情况。为此，我们首先对单项指标做R型聚类分析，即进行指标聚类，以弄清各指标间的相互联系。

首先，我们以相关系数为距离，采用最长距离法进行聚类。令 X_1 为小麦， X_2 为洋芋， X_3 为早稻， X_4 为中稻， X_5 为晚稻， X_6 为玉米， X_7 为大豆。聚类结果如图一所示。这个结果是符合实际种植情况的，因为早稻和晚稻的相关性极强，其相关系数高达0.9908，自然最先聚成一类；而玉米和洋芋，小麦和中稻之间，由于地块种植的衔接关系，也具有一定的相关性（相关系数分别为0.7612和0.6033）。但是如果根据这个结果来挑选指标则会产生一定的误差；因为如果根据图一的结果将指标分为四类，则一类为早、晚稻，一类为玉米、洋芋，一类为小麦、中稻，一类为大豆。如果在每类中各选出一个指标，则必须在小麦和中稻中挑选一个，但湖北省各县（市）的小麦和中稻产量都很高，对全年粮都有着举足轻重的影响，从中只选一个指标则不能反映出综合指标的概貌；另外，大豆虽然在指标性质上有一定的特殊性，但湖北省的大豆产量普遍较低，因此不能做为主要指标考虑。

根据以上的分析，我们又选用马氏距离为度量指标，采用最短距离法进行聚类，其结果如图二所示。按照这个结果将指标分为四类，则一类为早、晚稻，一类为洋芋、大豆、玉米，小麦和中稻各自成一类。这时再来挑选指标，便可满足实际需要，因为早、晚稻的相关性极强，当然可以只选一个；小麦和中稻都是自然入选的；而洋芋、玉米、大豆均属杂粮，而且产量都很低，在这一类中只选一个指标，虽然有一定的误差，但不会对全年粮产生很大的影



图一



图二

响。因此，我们初步决定根据这个聚类结果来挑选指标。

2.2、指标筛选

通过聚类分析，我们对指标的结构有了一定的了解。另外，我们又运用回归分析对指标做了进一步的筛选。

1. 以三个综合指标为因变量，以七个单项指标为自变量做多因变量双重逐步回归。从回归的结果中，我们了解到如下的情况。

(1) 三个综合指标中，秋粮与夏粮归为一组，全年粮归为另一组。这与实际情况是相吻合的。因为夏、秋两季的农作物产量已构成全年粮食产量，自然就与其分开考虑了。

(2) 在全年粮与单项指标的筛选过程中，由于这七个指标都是粮食产量中的重要统计指标，所以全都入选了，回归结果如下：

$$y = 1.021925x_1 + 1.07885x_2 + 1.108108x_3 + 1.082031x_4 + 0.975779x_5 + 0.937807x_6 + 2.195807x_7 + 8082.875$$

其中， y 为全年粮， x_1 为玉米， x_2 为小麦， x_3 为洋芋， x_4 为早稻， x_5 为中稻， x_6 为晚稻， x_7 为大豆。方程的相关系数为0.9956。各自变量与 y 的 F 检验值见表一。从表中可以知道，对全年粮影响较大的指标为：中稻、小麦、玉米、晚稻。而后三个指标的作用相对较小。

各自变量检验值

表一

x_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7
F_i	107.2229	2098.66	64.1783	80.8118	7431.966	97.3286	44.9318

因为夏粮的情况通过小麦可基本上反映出来，但秋粮的结构则比较复杂，因此，我们又对秋粮做了进一步分析。

2. 以秋粮为因变量 y ，所有七个单项指标为自变量，进行逐步回归。结果如下：

$$y = 1.456632x_1 + 1.085893x_2 + 0.991499x_3 + 0.919849x_4 + 1.138564x_5 + 8426.406$$

其中， y 为秋粮， x_1 为大豆， x_2 为早稻， x_3 为中稻， x_4 为晚稻， x_5 为玉米。方程的相关系数为0.9913。各自变量关于 y 的 F 检验值见表二。

各自变量检验值

表二

x_i	x_1	x_2	x_3	x_4	x_5
F_i	18.4416	72.5343	12211.26	91.176	251.146

从这个结果可以看出，筛选的情况与统计指标的性质是非常一致的，因为没有入选的两个指标，小麦和洋芋恰好是夏粮的统计指标。这说明筛选的结果是好的。另外，从表二知道，对秋粮影响较大的指标为：中稻、玉米、晚稻，这与全年粮算出的结果是一致的。

综上所述，并结合聚类图二的结果，我们初步选定四个基本指标：中稻、小麦、玉米、晚稻。从回归分析来看，它们对三个综合指标的影响都很大；从聚类分析来看，它们又恰好

来自不同的类。但为了进一步验证这个结果，我们又做了如下的考虑：

1. 为了考查四个基本指标对其它指标的代表性，我们将全年粮，秋粮、夏粮、早稻，大豆、洋芋作为因变量，以小麦、中稻、晚稻、玉米为自变量，逐一做多元线性回归，其结果如下：

$$\begin{aligned}
 y_1 &= 0.08579x_1 + 0.001329x_2 + 0.010175x_3 + 0.028705x_4 - 273.5205 \\
 y_2 &= 0.177597x_1 + 0.025392x_2 - 0.022539x_3 + 0.926567x_4 + 8183.532 \\
 y_3 &= 1.534512x_1 + 1.048523x_2 + 0.981642x_3 + 2.000038x_4 + 17927.51 \\
 y_4 &= 1.067423x_1 + 0.08516x_2 + 0.963979x_3 + 1.9732x_4 + 15729.76 \\
 y_5 &= 0.466018x_1 - 0.054798x_2 - 0.007142x_3 - 0.003039x_4 + 1434.106 \\
 y_6 &= 0.467089x_1 + 0.963362x_2 + 0.017662x_3 + 0.026838x_4 + 2197.814
 \end{aligned}$$

其中： y_1 为大豆， y_2 为早稻， y_3 为全年粮， y_4 为秋粮， y_5 为洋芋， y_6 为夏粮； x_1 为玉米， x_2 为小麦， x_3 为中稻， x_4 为晚稻。

各因变量 y_i 的复相关系数及 F 统计量如表三所示。从这个结果可以看到，所选出的四个指标对全年粮，秋粮、夏粮、早稻的代表性都很好，复相关系数均在 0.98 以上；洋芋的复相关系数为 0.63404，也可令人满意；但大豆的复相关系数仅为 0.32152，说明这四个指标还不能反映大豆的情况，这与图一的聚类结果是一致的，也进一步说明了大豆指标的特殊性。

回归方程统计检验表 表三

因变量	复相关系数	F 统计量
y_1	0.32152	8.0558
y_2	0.98384	1034.433
y_3	0.99427	2948.26
y_4	0.99496	3353.202
y_5	0.63404	29.1532
y_6	0.98238	947.5016

2. 我们将大豆也划归为自变量，再次进行回归分析，计算结果如下：

$$\begin{aligned}
 y_1 &= 0.502547x_1 - 0.054232x_2 \\
 &\quad - 0.425794x_3 + 0.011474x_4 \\
 &\quad + 0.009148x_5 + 1317.64 \\
 y_2 &= -0.144121x_1 + 0.02591x_2 \\
 &\quad - 0.390209x_3 - 0.018568x_4 \\
 &\quad + 0.937768x_5 + 8076.812 \\
 y_3 &= 1.422822x_1 + 1.046793x_2
 \end{aligned}$$

$$+ 1.301898x_3 + 0.968395x_4 + 1.962667x_5 + 18283.61$$

$$y_4 = 0.636057x_1 - 0.527704x_2 + 1.470205x_3 + 1.050805x_4 + 1.876689x_5 + 38237.85$$

$$y_5 = 0.50371x_1 + 0.986634x_2 + 0.00681x_3 + 0.01804x_4 + 0.02179x_5 + 2226.311$$

其中： y_1 为洋芋， y_2 为早稻， y_3 为全年粮， y_4 为秋粮， y_5 为夏粮； x_1 为玉米， x_2 为小麦， x_3 为大豆， x_4 为中稻， x_5 为晚稻。

各因变量 y_i 的复相关系数及 F 统计量如表四所示。从这个结果看出，各因变量的回归

回归方程统计检验表 表四

因变量	复相关系数	F 统计量
y_1	0.6459	21.4422
y_2	0.98412	830.0272
y_3	0.99463	2481.83
y_4	0.95989	320.6662
y_5	0.98441	832.133

效果均可令人满意。因此，从定量的角度应该筛选出五个主要指标：小麦、中稻、晚稻、玉米和大豆。

最后，我们又结合定性分析，了解各指标的实际作用。根据湖北省自然条件知道，筛选出的五个基本指标中，小麦、中稻、晚稻的产量都很高，在全年粮中的作用都比较大；而玉米和大豆的产量均偏低，但玉米与

洋芋有一定的相关性，其作用相对较大，而大豆虽然在指标性质上有一定的特殊性，但对全年粮没有太大的影响，且因为播种面积较小，在统计数据时也有一定的困难。因此，综合定性和定量的各种分析结果，最后确定四个主要指标：小麦、中稻、晚稻、玉米。由此便可了解各种粮食作物的概貌。便于更好地做出聚类分析。

三、改进方案的研究

我们把湖北省73个县(市)作为单位，依这四个单项指标作系统聚类，聚类的距离是马氏距离，采用最短距离法进行归并，有关的理论和公式可以参看〔1〕。

我们做了下面两种聚类。

(i) 以中稻、晚稻、玉米、小麦、大豆为聚类指标；

(ii) 用全部十项指标作为聚类指标，其结果也大体上是一致的。在大的类上没有差别，在分细时归类有一些差别：如黄冈和大冶在原聚类中较早就并为一类，而现在合并较迟，因为两县在大豆和洋芋的产量上差别较大，大冶县洋芋产量是71吨，而黄冈为0，大冶县大豆产量为3618吨，而黄冈县是1875吨，其他情况也和这类似。

根据上面的各种聚类分析的结果，可以认为以四个单项指标聚类的结论是可信的。我们就以这一结果作为抽样基础。

3.1 确定改进的抽样方案

从聚类分析图上可以明显看到，就类之间的距离来看，73个县分成33类(因为湖北省现有的国家调查县是33个，为便于比较，在改进的抽样方案中，仍将抽样数目定为33个)，每一个类的方差，除了只有一个单位的类以外，几乎都相同。于是可以考虑以下的改进方法：

(i) 在每一类中都抽一个单位，一共33个类，共得33个样本点，组成一套样本。因为每一类中各项指标都很相近，选取一个样本点就能代表这一类，这个样本点可以在这一个类中随机地选取。这是一种抽样的方案。

(ii) 在聚类分析的基础上，区别两种不同的类型，然后对一部分再用对称等距抽样的方法。

实际上，这33类可以分为两部分。一部分是与其他的县(市)差异很大的，而且实际上是粮食产量的“大户”或结构独特的单位，这些县(市)的产量能否估准对全省的产量有重大的影响，因此，这些“特殊的”单位必须作为样本点。这一部分共有十六个县(市)，它们是：

监利县、仙桃市、洪湖市、安陆市、南漳县、当阳市、广水市、宜城县、江陵县、京山县、潜江市、荆门市、枣阳市、襄阳县、钟祥县、随州市

另一部分是其余的57个县(市)。总的说来差异小的57个县(市)分别是公安县、孝感市、浠水县、黄陂县、老河口市、新洲县、天门市、大悟县、谷城县、蕲春县、应城市、黄梅县、阳新县、松滋县、武昌县、红安县、鄂州市辖区、石首市、黄冈县、武穴市、大冶县、云梦县、汉阳县、通城县、蒲圻市、嘉鱼县、崇阳县、咸宁市、郧县、丹江口市、恩施市、利川市、宜昌县、枝江县、枝城市、罗田县、郧西县、英山县、建始县、巴东县、长阳县、房县、竹山县、竹溪县、秭归县、保康县、通山县、神农架林区、五峰县、鹤峰县、兴山县、咸丰县、宜恩县、来凤县、远安县、汉川县、麻城市，它们之间的差异比较小，我们对这57个县(市)仍然采用原来的以全年粮食亩产排队、播种面积累计的对称等距抽样方法，从中抽出17个样本点，这样一共有 $17 + 16 = 33$ 个样本点，仍然保持样本点的数目是33。这是另一种抽样方案，

我们对上述两种改进的方案的各种可能的每一套样本，均用1989年的数据进行了实际的计算。计算结果表明后一种方案比前一种好。后一种方案一共有33套样本，这33套样本的估计值与真值的相对误差都比较小。因此我们建议可以采用后一种方案。

3.2 改进方案的比较分析

从以上的研究可以看出，无论是现行的对称等距抽样方案，还是这里新介绍的抽样方案，各种方案给出的样本套数是不多的，我们很容易将全部不同的各套样本均列举出来，研究各套样本给出的估计值和真值的差别就能衡量这一套样本的好坏。我们引入一些符号来进一步说明这一点，设 X_1, \dots, X_p 分别表示被研究的指标，如果一共有 k 套样本，第 i 套样本给出的对指标 x_α 的总体真值 μ_α 的估计值为 $\hat{\mu}_\alpha(i)$ ，则 $\frac{1}{\mu_\alpha} |\hat{\mu}_\alpha(i) - \mu_\alpha|$ 就反映了第 i 套样本对指标 X_α 而言代表性的好坏，这个值越小就越好。第 i 套样本对 X_1, \dots, X_p 而言，代表性最差的一定是 $\frac{1}{\hat{\mu}_\alpha} |\hat{\mu}_\alpha(i) - \mu_\alpha|$ ， $\alpha = 1, \dots, p$ 中的最大值，记

$$(1) \quad \Delta_i = \max_{1 \leq \alpha \leq p} |\hat{\mu}_\alpha(i) - \mu_\alpha| / \mu_\alpha \quad i = 1, 2, \dots, k$$

则 Δ_i 就反映了第 i 套样本的总的情况，我们用 $\max_{1 \leq i \leq k} \Delta_i$ (记 Δ)这个数反映这一种抽样方法的好坏，如果 Δ 值小，这一种抽样方案的精度就高， Δ 值大，这一种抽样方案的精度就差。由于农产量的变化一般说来不是大起大落的，除非有严重的灾情，这时无无论那一种方法都需另行调整，因此我们对历史资料进行上述精度的分析是有现实意义的。而且从下面的介绍可以看出，单纯用随机抽样的理论去研究精度是行不通的。因此我们把(1)式中的 Δ_i 称为第 i 套样本对 X_1, \dots, X_p 这 P 个指标的历史精度或历年精度，相应的 Δ 就称这种抽样方案的历史精度或历年精度。事实上 Δ 反映的是这种抽样方案中最差的一套样本的情况，我们也可引入

$$(2) \quad \Delta_* = \min_{1 \leq i \leq k} \Delta_i$$

Δ_* 反映了这种抽样方案中最好的一套样本的情况，从 Δ 和 Δ_* 就能了解这种抽样方案的概貌。

下面我们列举对称等距抽样方案和改进方案的历史精度数值，以便比较。根据湖北省1989年的资料，计算的结果是这样的：

最差的一套样本的 Δ_i (单位：%)

	全年粮食	夏粮	小麦	秋粮	早稻	中稻	晚稻	Δ
对称等距抽样	-0.10	-14.51	-20.62	0.45	2.07	-8.08	-1.21	20.62
改进的抽样	0.08	-2.57	6.57	-0.51	-0.03	1.11	-1.00	-6.57

最好的一套样本的 Δ_i (单位：%)

	全年粮食	夏粮	小麦	秋粮	早稻	中稻	晚稻	Δ_*
对称等距抽样	0.12	-7.81	-13.43	2.34	2.26	-6.23	-1.91	13.43
改进的抽样	0.07	-0.21	-1.73	-1.00	1.59	0.59	-0.75	-1.73

从这里可以看出：对称等距抽样方案的最大相对误差在13.43%—20.62%之间；而改进方案的最大相对误差在1.73%—6.57%之间。由此可见改进方案的最差的一套样本也比原来

对称等距抽样中最好的一套样本要好。值得注意的是，改进方案和对称等距抽样的抽样点的数目是一样的。我们列举上述数字，说明多目标的抽样方案是可以找到的，原用的对称等距抽样方案是应予以改进的。

四、小 结

从前面的介绍和讨论可以看出，多目标的抽样方案是有办法制定的。抽样调查的方法在我国真正的施行还是在改革开放以后近十年的事，特别是随着改革开放的不断深入，抽样调查的方法正发挥着越来越大的作用，应用的范围也在迅速扩大，因此认真研究多目标抽样调查方案的理论和方法，对改进我国的统计工作，加速社会科学的深入研究，是有明显的作用并能迅速获得效益，我们的工作是在这一方面能起推动和促进的作用。

我国的统计调查中，长期以来，积累了不少全面普查和典型调查的经验，而西方国家的抽样技术和理论是以随机抽样为基础的。要能将典型调查与随机抽样相结合，走出一条新的路子，建立适合我国国情的一套抽样调查的理论和方法，是我们统计工作者面临的共同任务。本项目也是想在这一方面探索出一些经验和方法。随着电子计算机的广泛作用，一些复杂的多元统计分析方法的应用也日趋成熟，在分析研究统计资料上是可以发挥它的功能的，在这一方面本项目也作了一些探索。

从本项目的研究中我们感到有下列几点体会：

1、聚类分析是整理资料，发现典型的一个重要的手段。所谓典型是与分类密切相关的，如果类还没有形成时，典型就没有意义。而不同的调查目标，它所考察的对象分类的标准自然不同，因此按调查项目进行聚类，找出合理的分类，再从各类中去选择典型。

以湖北省农产量调查的问题为例，聚类分析的结果区分出特殊和一般。特殊的一些单位必需调查，它们无法由别的单位来“代表”，而且是重要的单位。一般的单位，可以用对称等距抽样的方法去再分类和选取典型——也就是样本点。这样就能把“特殊”与“一般”给出定量的明显的区分。这类情况在人口、工业和其他行业的抽样调查中也是存在的。例如人口抽样调查，如将300万以上的特大城市作为一类，则这一类中的差别就很大；如将5万人口以下的小城市归为一类，虽然有一定数目的城市，它们之间的差别就小。因此适合不同目标的抽样调查，在有可能时，先用历史资料作一些聚类分析是很有意义的。

2、抽样方案应采用典型与随机相结合的方法。抽样方案的好坏取决于所得的各套可能的样本是否具有“代表”性。对于一些特殊的单位，用随机方法去抽取样本，是不合适的。对于一般的单位是完全可以的。完全符合“不同性质的矛盾应用不同的方法来处理”这一原则。改进方案所以有明显的成效，也正是贯彻了这一原则的结果。

然而，随着这种抽样方案的产生，就伴随一个评价方案的标准如何确定的问题。因为这不是纯随机的抽样，用随机抽样的方法原理来评价会有困难，而且也不易比较。事实上，对称等距抽样也不是真的随机选一个 r 的值而确定的，往往先确定 r 若干个值，得几套样本，比较不同套的样本的分布那一套更合理些，交通困难少些，费用小一些……等各种其他因素，从而才确定了某一套样本，因此也不是随机的。正因如此，我们充分利用了对称等距抽样的特点，把各种可能出现的每一套样本都作了分析比较，充分利用了历史资料，给出了比较可信的评判标准。

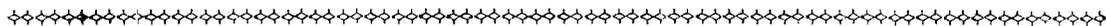
(下转29页)

承包的内容,将承包者的利益与企业经济效益紧密联系起来;二是对因技术进步而提高经济效益所带来的级差收入,应按一定的比例分配给承包者,并合理确定基数,避免出现“鞭打快牛”,激发承包者提高经济效益的积极性。(5)全面推行工资总额与经济效益挂钩的工资改革,建立反映劳动质和量的工资制度。通过将职工工资总额与企业经济效益挂钩,同时,实行反映劳动的质和量的个人消费品的按劳分配,使劳动者从个人利益的角度去关心企业经济效益的提高。(6)把经济效益指标纳入财政包干体制。通过把地区经济效益指标纳入财政包干体制,把经济效益的提高作为划分中央和地方利税递增分割比例的一个权数,将地方财政收入与本地区经济效益联系起来,使地方政府从本地区的利益的角度去关心经济效益的提高。(7)加强国有资产的管理。开展清产核资,解决国有资产状况不清,管理混乱和被侵占流失的状况,确保国有资产的完整和增殖。同时,实行“税后还贷”,硬化企业预算约束,改变企业速度扩张冲动制约条件松弛的状况。

如果建立起了地方政府、企业、企业职工和承包者的经济利益与经济效益的紧密联系,建立起了有效的效益利益机制,经济效益长期低下的顽症就能得到根治。

注释:

①详见刘光杰:《社会主义政治经济学基本理论问题研究》第68页。武汉大学出版社1988年版。



(上接36页)

当然,如果没有历史资料可用,以上这两点就很难做到。然而中国有几十年比较丰富、全面的调查资料,这正是我们可以充分利用的信息,所以我们提出的改进方法,对于相当一部分的抽样调查问题是有现实意义的。

从我们的研究和探索中,虽然积累了一点经验,提出了看来是可行的改进方案,但也感到还有一些问题没有很好解决。在本项目的研究中,我们较多地应用了数理统计中的多元分析方法的某几种处理资料的技术,如聚类分析,多对多、多对一的回归分析,取得了一些明显的成效,然而还有大量的方法尚未引入多目标抽样调查的领域,还需作进一步的探索 and 开发。从实际工作的需要来看,就农产量调查本身,还有以下的几个问题有待进一步的研究:

(1)遇到有重灾或重大政策变化的年份,应怎样补充一些样本点,或怎样进行折算,……等给出一些补救的方法来保证估计的精度。

(2)怎样在一套样本点使用若干年的基础上,及时发现应予以调整才能保证精度?如何调整使人力、物力等花费较小?

(3)对于典型调查与随机抽样调查相结合的方案,应如何评价它的好坏?怎样从理论上给出一般性的原则来指导这一类调查方案?

(4)如何将经济作物的调查,农家计的调查……等等,许多不同类型的多目标调查形成的一个特大的多目标调查选择抽样方案?这一问题的重要性是十分明显的。

从我们的工作来看,深深感到这一方面还有许多问题可以研究,还有很多潜力可发掘,希望在这一方面能有更多更好的成果不断涌现出来。