

文章编号:0253-9993(2007)10-1117-04

# 全国乡镇煤矿原煤产量抽样调查与预测模型

胡予红<sup>1</sup>, 刘文革<sup>2</sup>, 赵一归<sup>2</sup>, 黄盛初<sup>2</sup>

(1. 中国矿业大学(北京)化学与环境工程学院, 北京 100083; 2. 国家安全生产监督总局信息研究院能源所, 北京 100029)

**摘要:** 建立乡镇煤矿二阶段不等概分层随机抽样方法, 抽取分布于山西、内蒙古、黑龙江等13省的236个样本煤矿进行实地调查, 采取系数修正法和直接推算法测算2006年1—6月乡镇煤矿原煤产量, 并建立回归模型, 预测2006年全年乡镇煤矿原煤产量和全国煤炭产量, 为国家制定煤矿整顿关闭战略和煤炭行业发展战略提供依据。

**关键词:** 乡镇煤矿; 原煤产量; 抽样调查; 预测模型

**中图分类号:** F407.21      **文献标识码:** A

## Survey sampling and forecasting model on coal output of town and village coal mines in China

HU Yu-hong<sup>1</sup>, LIU Wen-ge<sup>2</sup>, ZHAO Yi-gui<sup>2</sup>, HUANG Sheng-chu<sup>2</sup>

(1. School of Chemical and Environmental Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China; 2. Energy and Safety Division Information Institute, State Administration of Work Safety China, Beijing 100029, China)

**Abstract:** Established the method of two-stage unequal probability stratified sampling on town and village coal mines (TVCM), selected and surveyed 236 coal mine samples from 13 Provinces such as Shanxi, Inner Mongolia and Heilongjiang, and calculated the town and village coal mines (TVCM) coal output from January to June 2006 by using the methods of coefficient modification and direct calculation. Moreover, forecasted the national and TVCM coal outputs of the whole year by setting up the regression model, which provided the basis for the government to make out the national strategies on coal mine rectification and closure and the development of coal industry.

**Key words:** TVCM; coal output; survey sampling; forecasting model

近年来由于煤炭行业管理弱化, 统计渠道不畅, 乡镇煤矿产量统计误差较大. 从近10 a的统计数据看, 乡镇煤矿原煤产量占全国原煤总产量的比重一直在40%左右. 乡镇煤矿原煤统计产量与实际产量的差异是造成全国煤炭总产量存在较大误差的主要原因. 为解决这个问题, 笔者于2006年9—12月组织开展了全国乡镇煤矿原煤产量抽样调查. 采用抽样调查这一科学方法, 通过抽样设计, 抽取具有代表性的样本, 将抽样误差控制在允许的范围之内, 从而测算出相对准确的乡镇煤矿原煤产量, 为制定煤矿整顿关闭战略和煤炭行业发展战略提供可靠依据.

## 1 抽样方案设计

### 1.1 调查范围与样本规模

测算全国乡镇煤矿原煤产量, 须以全国24个产煤省的所有乡镇煤矿作为调查范围(总体), 以样本

煤矿为调查对象,调查样本煤矿2006年1—6月的统计产量、实际产量及基本信息。

经验和公式检验表明,当误差和置信区间一定时,不同的样本量计算公式计算出来的样本量是十分相近的。按在90% ( $t=1.64$ )的概率保证度下,最大相对误差不超过10% ( $P$ ),根据经验,取变异程度  $V=0.86$ ,用简单随机抽样方法样本量计算公式:  $n = t^2 V^2 / P^2 \leq t^2 / P^2$  计算出所需的样本量约为197个。考虑数据回收率不足的可能性,放大样本量5%,初步确定总样本量为210个<sup>[1]</sup>。

## 1.2 样本抽取方法与分配

以全国乡镇煤矿为总体,采用二阶段不等概分层随机抽样方法选取样本。

(1) 第1阶段不等概率抽样——抽取调查省份 按与2006年上半年各省乡镇煤矿原煤统计产量成比例,采用PPS抽样方法,抽取山西、内蒙古、贵州、陕西、四川、云南、河南、湖南、黑龙江、重庆、辽宁、吉林、甘肃13个产煤省为一级抽样单元。抽样显示,2005年,13个省乡镇煤矿原煤统计报表产量为7.5亿t,占全国乡镇煤矿产量的近90%,样本单元产量占总体足够大的比重,大小样本单元比较齐全。结合实际情况,将样本总量增大到236个,按2005年各省乡镇煤矿数量比例分配到13个省。

(2) 第2阶段分层随机抽样——抽取样本煤矿 以抽中的产煤省为单位编制抽样框。抽样框包括全省乡镇煤矿的名称、地址、2005年核定产能信息,乡镇煤矿按2005年核定产能( $X$ )分为3层 ( $X \leq 3$ 万t,  $3 < X \leq 6$ 万t,  $X > 6$ 万t)。乡镇煤矿产能普遍较高的省份,分层准则可以调整。然后按各层煤矿数量比例分配各层中的样本量;采用简单随机抽样方法从各层中抽取样本。

## 1.3 总量推算方法<sup>[2]</sup>

### (1) 系数修正法

系数修正法是根据抽样调查数据与统计数据相对比,求出差错率作为修正系数,来修正总体指标的一种方法。

$$\text{差错率} = (\text{抽样调查数} - \text{全面统计数}) / \text{全面统计数} \times 100\%$$

$$\text{总体修正后总数} = \text{全面统计数} + \text{总体差错数} \quad (\text{总体差错数} = \text{全面统计数} \times \text{差错率})$$

### (2) 直接推算法

① 采用分层随机抽样估计公式推算各省总量和误差。省总量估计、估计量方差、估计量标准差、估计量置信区间分别为

$$\hat{Y} = \sum_{h=1}^L N_h \bar{y}_h,$$

$$V(\hat{Y}) = N^2 \sum_{h=1}^L W_h^2 V(\bar{y}_h) = \sum_{h=1}^L N_h^2 \frac{1-f_h}{n_h} s_h^2,$$

$$s(\hat{Y}) = \sqrt{V(\hat{Y})},$$

$$\hat{Y} \pm ts(\hat{Y}),$$

式中,  $h$  为层,  $h=1, 2, \dots, L$ ;  $L$  为分层总数;  $N$  为省乡镇矿点总数;  $n$  为样本乡镇矿点总数;  $N_h$  为第  $h$  层乡镇矿点总数;  $n_h$  为第  $h$  层样本乡镇矿点总数。

② 2006年1—6月全国乡镇煤矿原煤产量推算。采用汉森-赫维茨(Hansen-Hurwitz)估计公式推算全国总量和误差。全国总量估计量、估计量方差、标准差、最大相对误差分别为

$$\hat{Y}_{\text{HH}} = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{z_i},$$

$$V(\hat{Y}_{\text{HH}}) = \frac{1}{n(n-1)} \sum_{i=1}^n \left( \frac{y_i}{z_i} - \hat{Y}_{\text{HH}} \right)^2,$$

$$s(\hat{Y}_{\text{HH}}) = \sqrt{V(\hat{Y}_{\text{HH}})},$$

$$\hat{r}_{HH} = \frac{t \sqrt{V(\hat{Y}_{HH})}}{\hat{Y}_{HH}}$$

式中,  $n$  为抽取省(一级抽样单元)个数;  $y_i$  为第  $i$  个省(一级抽样单元)产量无偏估计;  $z_i$  为第  $i$  个省(一级抽样单元)入样概率;  $m_i$  为第  $i$  个省(一级抽样单元)2006年1—6月乡镇煤矿原煤统计产量;  $m_0$  为2006年1—6月全国乡镇煤矿统计产量。

## 2 总量推算

通过对236个样本煤矿的实地调查,对收回的调查数据仔细审核录入,采用Excel软件和统计软件SPSS进行数据处理测算<sup>[3]</sup>。

### 2.1 用系数修正法推算总量

2006年1—6月236个样本矿点原煤统计产量671.1万t,通过对所有样本实地调查核实,实际产量为743.7万t,采用系数修正法,计算2006年1—6月全国乡镇煤矿修正产量为 $31\,060.67 \times (1 + 10.8\%) = 34\,422.64$ 万t。

### 2.2 用直接推算法推算总量

采用直接推算法,推算各省乡镇煤矿2006年1—6月原煤产量,推算全国乡镇煤矿2006年1—6月原煤产量,结果见表1。

表1 全国乡镇煤矿2006年1—6月原煤产量推算

Table 1 The calculated coal output of TVCM from 2006.1 to 2006.6

总量估计值 $Y_{HH}/Mt$	$Y_{HH}$ 方差 $V(Y_{HH})$	$Y_{HH}$ 标准差 $s(Y_{HH})/Mt$	90% 概率保证度置 信区间 ( $t=1.64$ )	最大相对 误差 $r$
34 620.89	176 072.89	419.61	33 932.73      35 309.05	0.02

通过系数修正法和直接推算法推算2006年上半年全国乡镇煤矿原煤产量分别为34 422.64万t和34 620.89万t(概率保证度90%,最大相对误差为2%),两种计算方法得到的总量非常接近,仅相差198.25万t。取两种方法推算结果的均值,2006年1—6月乡镇煤矿原煤产量应为34 521.77万t。

## 3 全国煤炭产量预测模型

### 3.1 2006年全年乡镇煤矿原煤产量预测

根据1996—2005年全国乡镇煤矿原煤产量统计数据(表2),以乡镇煤矿1—6月累计产量为自变量  $x$ , 乡镇煤矿全年产量为因变量  $y$ , 建立回归预测模型,得回归方程  $y = 2.5444x - 1\,600.9$ , 决定系数  $R^2 = 0.9342$ , 说明方程具有很好的统计意义,如图1所示。

表2 1996—2005年乡镇煤矿1—6月产量与全年产量<sup>[4]</sup>

Table 2 The coal output before July and the whole-year coal output of TVCM from 1996 to 2005

年 份	乡镇煤矿1—6月累计产量/Mt	乡镇煤矿全年产量/Mt	年 份	乡镇煤矿1—6月累计产量/Mt	乡镇煤矿全年产量/Mt
1996	26 913.29	61 477.04	2001	12 024.30	26 385.47
1997	25 556.66	57 042.14	2002	14 476.19	43 350.60
1998	21 118.43	51 617.06	2003	22 816.30	63 389.61
1999	14 778.28	31 700.57	2004	30 817.82	76 175.03
2000	11 356.02	26 917.28	2005	31 437.00	83 551.47

将2006年1—6月的全国乡镇煤矿原煤统计产量  $x = 31\,060.67$ 万t,代入回归方程计算得  $y =$

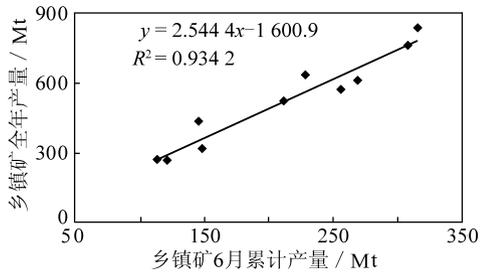


图1 1996—2005年乡镇煤矿上半年产量与全年产量关系

Fig. 1 The relation between the coal output before July and the whole-year coal output of TVCM from 1996 to 2005

以及资源整合、关闭整顿政策的实施, 还有因事故造成的停产整顿的因素, 绝大部分煤矿生产很不正常, 笔者认为这个比例比较符合实际情况. 按 36.94% 计算, 则 2006 年全年乡镇煤矿原煤预测产量  $y_x = 34\ 521.77 / 36.94\% = 93\ 453.62$  万 t.

### 3.2 2006 年全国煤炭产量预测

根据 1996—2005 年全国国有煤矿原煤产量统计数据, 以国有煤矿 1—10 月原煤产量为自变量  $x$ , 以国有煤矿全年产量为因变量  $y$ , 建立回归预测模型, 得回归方程  $y = 1.1733x + 4\ 514.8$ , 决定系数  $R^2 = 0.9977$ , 如图 2 所示.

预测 2006 年全国国有煤矿原煤产量为  $y_g = 1.1733 \times 116\ 648.20 + 4\ 514.8 = 141\ 378.13$  万 t. 2006 年全国煤炭预测产量为  $y_q = y_x + y_g = 93\ 453.62 + 141\ 378.13 \approx 23.5$  亿 t. 国家统计局公布的数据为 23.8 亿 t.

## 4 结 语

通过抽样调查方法测算 2006 年 1—6 月全国乡镇煤矿原煤产量为 34 521.77 万 t. 根据 1996—2005 年煤炭产量统计数据, 建立回归预测模型, 预测 2006 年乡镇煤矿原煤总产量为 93 453.62 万 t, 国有煤矿原煤总产量为 141 378.13 万 t, 全国煤炭产量为 23.5 亿 t.

这次全国乡镇煤矿抽样调查成果及煤炭产量预测值为国家安全生产监督总局调度统计司确定 2006 年全国煤炭产量提供了重要参考. 2007 年 1 月调度统计司公布 2006 年全国煤炭产量为 23.25 亿 t.

采用抽样调查方法测算乡镇煤矿原煤产量, 不仅节省大量的人力、物力和财力, 而且能够在短时间内获取相对准确的产量数据. 建议煤炭行业统计部门建立抽样调查长效机制, 将抽样调查与全面统计结合起来, 修正统计报表数据误差, 从而测算相对准确的煤炭实际产量.

### 参考文献:

- [1] 孙山泽. 抽样调查 [M]. 北京: 北京大学出版社, 2004.
- [2] 金勇进. 抽样技术 [M]. 北京: 中国人民大学出版社, 2002.
- [3] 张建同. 以 Excel 和 SPSS 为工具的管理统计 [M]. 北京: 清华大学出版社, 2005.
- [4] 黄盛初. 2005 中国煤炭发展报告 [M]. 北京: 煤炭工业出版社, 2006.

77 429.88 万 t, 则  $x/y \approx 40\%$ , 初步预测乡镇煤矿 2006 年 1—6 月原煤产量占全年产量的 40% 左右.

根据乡镇煤矿 2006 年 1—10 月原煤统计产量, 加上乡镇煤矿 2005 年 11—12 月原煤统计产量及翘尾量推算 2006 年全年统计产量, 预测 2006 年 1—6 月统计产量占全年统计产量比例为:  $31\ 060 / (61\ 634 + 7\ 184 + 10\ 958 + 4\ 313) = 36.94\%$ .

根据调查, 今年上半年由于节假日休息,

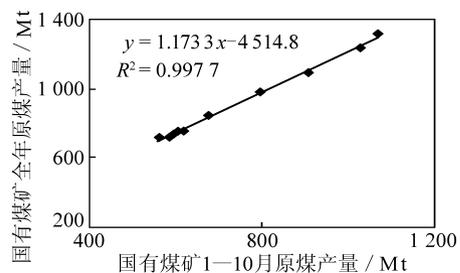


图2 1996—2005年国有煤矿1—10月产量与全年产量的关系

Fig. 2 The relation between the coal output before November and the whole-year coal output of NCM from 1996 to 2005