

## 影响山西省农机化水平的多因素逐步回归分析

陈宝峰<sup>1,2</sup> 白人朴<sup>2</sup> 刘广利<sup>3</sup>

(1. 中国农业大学 经济管理学院, 北京 100083;

2. 中国农业大学 工学院, 北京 100083;

3. 中国农业大学 信息与电气工程学院, 北京 100083)

**摘要** 针对定量分析农机化发展影响因素困难的问题,依据2002年山西省115个县(市)影响农机化水平的10个指标数据,采用逐步回归分析法,建立了山西省农业机械化发展水平影响因素逐步回归模型。模型分析结果表明,山西省农机化发展水平的主要影响因素按程度高低依次为:每 $\text{hm}^2$ 农业机械总动力、人均国内生产总值、丘陵山地比例、农民人均纯收入、每 $\text{hm}^2$ 农业机械原值、农业劳动力占总劳动力比重和无霜期7个因素。提升综合农机化作业水平的关键在于在加强农机动力配备的同时还有赖于国民经济的发展、农业劳动力的合理转移,以及农民收入的提高和适合丘陵山地作业的小型农业机械的研制和推广。

**关键词** 农机化水平;影响因素;逐步回归分析

中图分类号 S 23

文章编号 1007-4333(2005)04-0115-04

文献标识码 A

## Stepwise regression analysis on the influencing factors of Shanxi agriculture mechanization level

Chen Baofeng<sup>1,2</sup>, Bai Renpu<sup>2</sup>, Liu Guangli<sup>3</sup>

(1. Economic & Management College, China Agricultural University, Beijing 100083, China;

2. College of Engineering, China Agricultural University, Beijing 100083, China;

3. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China)

**Abstract** This paper introduces the stepwise regression analysis in view of the difficulties to quantitatively analyze the influencing factors of the development of agricultural mechanization. Referring to the data in 2002 of 115 counties (cities) in Shanxi province, this paper set up the stepwise regression model of influencing factors of development level of agriculture mechanization of Shanxi province. As the analyzing result of the model shows, there are 7 main factors to influence the development level of agriculture mechanization of Shanxi province. In accordance with their influence, they are arranged successively as total motive force of agricultural machinery of each hectare, per capita gross domestic product, proportion of mountain and hill, per capita net income of farmers, agricultural machinery initial value of each hectare, proportion of agricultural workforce and frost-free period. Thus the key to improve the agriculture mechanization level is to strengthen the motive force of agricultural machinery. Meanwhile, the national economy and so on should be attached importance to.

**Key words** agriculture mechanization level; influencing factor; stepwise regression analysis

农业机械化的发展受社会、经济、自然环境等多种因素的影响<sup>[1]</sup>,在研究农业机械化发展问题时,需要了解影响其发展的主要因素,以便采取相应措施,有效地促进农机化的发展。如何在众多的影响

因素中找出主要影响因素是一个非常复杂的问题,人们对这个问题的研究多采用定性的比较分析,也有用综合评价方法。这些方法能大致确定各种因素的重要程度,但难以量化分析各影响因素对农机化

收稿日期:2005-02-25

基金项目:山西省软科学研究项目(051011-3)

作者简介:陈宝峰,教授,博士生导师,主要从事项目管理、农机化与农村发展阶段的研究,E-mail:cbfisu@163.com。

水平的影响程度<sup>[2-3]</sup>。本研究旨在采用逐步回归数学模型方法量化分析各种影响因素对山西省农机化发展的影响程度。

### 1 逐步回归数学模型

在多种影响因素中寻找主要因素,建立最优回归方程的最佳方法之一是逐步回归。其基本原理是:对所有的自变量,依其对因变量作用的大小逐个引进回归方程,通过反复筛选、替代,仅留下对因变量贡献较大的自变量,最终得到最优回归方程。以此对经济变量之间的关系进行解释说明,逐步回归的实质是利用数理统计的原理寻找一种简捷有效的科学方法揭示复杂经济现象的内在规律性<sup>[4-6]</sup>。

#### 1.1 相关系数矩阵的生成

设经济变量  $y$  与  $n$  个自变量  $x_1, x_2, \dots, x_n$  呈线性关系,取  $n$  个  $m$  组自变量  $x_{ki}$ , 则  $m$  组经济变量的观察值为  $y_k (k = 1, 2, \dots, m; i = 1, 2, \dots, n)$ 。

根据数理统计的原理,为了求得多元线性回归方程建立方程组

$$\left. \begin{aligned} B_{11} C_1 + B_{12} C_2 + \dots + B_{1n} C_n &= B_{1y} \\ B_{21} C_1 + B_{22} C_2 + \dots + B_{2n} C_n &= B_{2y} \\ \dots & \dots \dots \dots \\ B_{m1} C_1 + B_{m2} C_2 + \dots + B_{mn} C_n &= B_{my} \end{aligned} \right\} \quad (1)$$

其中:  $B_{ij} (i, j = 1, 2, \dots, n, n + 1)$  为方程组的系数;  $C_j (j = 1, \dots, n)$  为回归系数。

如果定义,  $\bar{x}_i = \frac{1}{m} \sum_{k=1}^m x_{ki}, \bar{y}_i = \frac{1}{m} \sum_{k=1}^m y_{ki} (i = 1, 2, \dots, n)$ ; 并令  $\bar{y} = \bar{x}_{n+1}, y_k = x_{k, n-1}$ , 则  $B_{ij} = (x_{ki}x_{kj}) - m(\bar{x}_i\bar{x}_j)$ 。方程组(1)的系数矩阵便成为  $(n + 1) \times (n + 1)$  的以对角线对称的方阵。

#### 1.2 相关系数方程组的建立

由于  $B_{ij}$  的量纲不同,而且值的差异也很大,为便于计算,可将式(1)改写为相关系数方程组。因为

$$R_{ij} = \frac{B_{ij}}{\sqrt{B_{ii}} \sqrt{B_{jj}}}$$

则有

$$\left. \begin{aligned} R_{11} C_1 + R_{12} C_2 + \dots + R_{1n} C_n &= R_{1y} \\ R_{21} C_1 + R_{22} C_2 + \dots + R_{2n} C_n &= R_{2y} \\ \dots & \dots \dots \dots \\ R_{m1} C_1 + R_{m2} C_2 + \dots + R_{mn} C_n &= R_{my} \end{aligned} \right\} \quad (2)$$

其中:  $R_{iy} = R_{i, n+1} (i = 1, 2, \dots, m)$ ;  $C_i$  为新的回归系数;相关系数矩阵  $R$  为  $(n + 1) \times (n + 1)$  阶的以对角线对称的方阵。

#### 1.3 逐步回归过程的实现

根据生成的相关系数矩阵  $R$  对自变量逐个进行筛选、替代。步骤如下:

1) 计算全部自变量的贡献,并找出相应的极值。

$$U_i = \frac{R_{n+1,i} R_{i,n+1}}{R_{ii}} \quad i = 1, 2, \dots, n$$

当  $U_i < 0$  时,找出  $|U_i|$  中的最小值  $U_2$  及对应的  $x_i$  的序号:  $i_2 = \min\{|U_i|\}, k_2 = i$ ; 当  $U_i \geq 0$  时,找出其中最大值  $U_1$  及对应的  $x_i$  的序号:  $i_1 = \max\{U_i\}, k_1 = i$ 。

2) 对剔除变量进行  $F$  检验。先计算  $F = U_2^2 / (m - 1) / R_{n+1, n+1}$ , 其中  $U_2$  为统计引入的自变量的个数,其起始值为 0。根据用户给定的临界检验值  $F_a$ , 若  $F > F_a$ , 则剔除变量  $x_i (i = k_2)$ , 并使  $i = i_1, k = k_2, k = 0$  (剔除标记), 之后转 4); 若  $F < F_a$ , 执行 3)。

3) 对引入的变量做  $F$  检验, 计算  $F = U_1^2 / (m - 1) / R_{n+1, n+1}$ 。若  $F > F_a$ , 则引入该自变量  $x_i (i = k_1)$ , 且使  $i = i_1, k = k_1, k = 1$  (引入标记), 之后执行 4); 若  $F < F_a$ , 说明回归方程已为最佳, 即可进行最终结果处理。此时  $k$  为前次对  $k$  个自变量所做的剔除(或引入)标记。

4) 重新计算矩阵  $R$ 。

通过下面的消去变换即可实现自变量的引入或剔除。设第  $q$  步的矩阵为  $R_{ij}^q$ , 第  $q + 1$  步时:

$$R_{ij}^{q+1} = \begin{cases} R_{ij}^q - (R_{ik}^q R_{kj}^q) / R_{kk}^q & i, j \neq k \\ - R_{ik}^q / R_{kk}^q & i \neq k, j = k \\ R_{kj}^q / R_{kk}^q & i = k, j \neq k \\ 1 / R_{kk}^q & i = j = k \end{cases}$$

重新生成矩阵  $R$  后返回 1)。

#### 1.4 逐步回归结果的计算

经 1.3 中步骤 1) ~ 4) 的反复筛选、替代, 如果可判断已经获得最优回归方程, 则可进行逐步回归结果的计算。

1) 建立最优回归方程。关键是确定引入自变量的系数

$$b_i = \frac{R_{i, n+1} \sqrt{B_{n+1, n+1}}}{\sqrt{B_{ii}}} \quad i = 1, 2, \dots, n$$

引入标记  $i = 1, b_0 = \bar{y} - (b_i \bar{x}_i)$ 。最优回归方程为

$$y = b_0 + (b_i x_i) \quad i = 1$$

2) 计算复相关系数  $R_0$  和剩余标准差  $S$  :

$$R_0 = \sqrt{1 - R_{n+1, n+1}}$$

$$S = \sqrt{B_{n+1, n+1} / (m - 1)}$$

3) 计算第  $i$  个自变量的标准差 :

$$S_i = \sqrt{\frac{1}{m} \sum_{k=1}^m (x_{ki}^2 - \bar{x}_i^2)}$$

## 2 山西省农机化水平的多因素逐步回归分析

### 2.1 变量选择和回归模型的建立

农机化作业水平是反映整体农机化水平的核心指标。经过筛选,确定农业劳动力占总劳动力比重等反映国民经济、农业、农民收入、自然环境及农机装备水平情况的 10 个指标为影响农机化水平的自变量<sup>[7-9]</sup>。模型中各因素及其关系如下:

因变量:  $Y$ , 农机作业水平, %。

自变量:  $X_1$ , 农业劳动力占总劳动力比重, %;  $X_2$ , 人均耕地面积,  $\text{hm}^2/\text{人}$ ;  $X_3$ , 人均国内生产总值, 元/人;  $X_4$ , 人均地方财政预算内收入, 元/人;  $X_5$ , 人均农林牧渔业总产值, 元/人;  $X_6$ , 农村居民人均纯收入, 元/人;  $X_7$ , 每  $\text{hm}^2$  农业机械总动力,  $\text{W}/\text{hm}^2$ ;  $X_8$ , 每  $\text{hm}^2$  农业机械原值, 元/ $\text{hm}^2$ ;  $X_9$ , 丘陵山地比例, %;  $X_{10}$ , 无霜期, d。

各因素之间的关系可表示为

$$Y_i = b_0 + b_1 X_{1i} + \dots + b_{10i} X_{10i} + U_i$$

其中:  $i$  为样本容量,  $i = 1, 2, \dots, 115$ ;  $b_j$  为方程变量系数,  $j = 1, 2, \dots, 10$ ;  $U_i$  为残余项。

### 2.2 结果分析

依据 2002 年山西省 115 个县(市)的指标数据采用 SPSS 逐步回归程序进行计算,结果表明,山西省农机化作业水平与所选择的 7 个因素明显相关(表 1)。

在所选定的影响山西省农机化水平的 10 个因素中,农村居民人均纯收入、无霜期、丘陵山地比例、人均国内生产总值、农业劳动力占总劳动力比重、每  $\text{hm}^2$  农业机械总动力、每  $\text{hm}^2$  农业机械原值对农机

作业水平有显著的影响,而劳均农林牧渔业产值、人均耕地面积及人均地方财政预算内收入因影响不显著而被排除在回归模型之外。

表 1 山西省农机化水平逐步回归系数

Table 1 Agriculture mechanization level Coefficients of Shanxi Province

模型参数	回归系数 $b$	标准方差	标准回归系数 $b^*$
常数 $U_i$	- 12.516	12.742	
农村居民人均纯收入 $X_6$	$6.55 \times 10^{-3}$	0.002	0.295
无霜期 $X_{10}$	0.123	0.046	0.174
丘陵山地比例 $X_9$	- 0.242	0.058	- 0.307
人均国内生产总值 $X_3$	$2.82 \times 10^{-3}$	0.001	0.331
农业劳动力占总劳动力比重 $X_1$	- 0.403	0.12	- 0.27
每 $\text{hm}^2$ 农业机械总动力 $X_7$	$1.37 \times 10^{-3}$	0	0.419
每 $\text{hm}^2$ 农业机械原值 $X_8$	$1.64 \times 10^{-3}$	0.001	0.291

注:农机作业水平  $Y$  为独立变量。

由于模型中的变量具有不同的计量单位,因此回归系数的大小并不说明对因变量的影响程度。为了能定量分析变量之间的关系,将回归系数  $b$  变为标准回归系数  $b^*$  进行比较分析(表 1)。

标准回归系数  $b^*$  绝对值的大小代表着不同影响因素对农机化作业水平影响的大小,因此,按对农机化作业水平影响程度从小到大的各影响因素的排列顺序分别为:每  $\text{hm}^2$  农业机械总动力、人均国内生产总值、丘陵山地比例、农村居民人均纯收入、每  $\text{hm}^2$  农业机械原值、农业劳动力占总劳动力比重、无霜期。

正相关关系的因素有每  $\text{hm}^2$  农业机械总动力、人均国内生产总值、农村居民人均纯收入、每  $\text{hm}^2$  农业机械原值和无霜期,即这些数值越大,农机化作业水平也就越高。

呈负相关关系的因素有丘陵山地比例和农业劳动力占总劳动力比重,丘陵山地积所占比例和农业劳动力占总劳动力比重,这两者越大农机化水平越低。

## 3 结束语

应用逐步回归分析方法对影响农业机械化水平的因素进行量化与分析是可行的。

按影响程度高低顺序,山西省农机化水平的影响因素依次为:每  $\text{hm}^2$  农业机械总动力、人均国内生产总值、丘陵山地比例、农村居民人均纯收入、每  $\text{hm}^2$  农业机械原值、农业劳动力占总劳动力比重及无霜期。对山西省农机化作业水平影响最大的因素是每  $\text{hm}^2$  农业机械总动力。

笔者建议:在推进山西省农机化发展的过程中,要提高综合农机化作业水平,首要的是加大农业机械的投入,大幅度提高农机动力配备水平;考虑到国民经济发展水平、农业劳动力比重、农民收入、地理条件等因素对农机化作业水平也有较大影响的现实,山西省农机化作业水平真正大幅度的提高还有赖于国民经济的发展、农业劳动力的大量转移、农民收入的大幅度提高和适合丘陵山地作业的小型农业机械的研制和推广。

逐步回归只是因子分析的方法之一,进一步可以结合其他方法进行研究。

## 参 考 文 献

- [1] 傅宾忠. 农机化投入影响因素的研究[J]. 中国农机化, 2004, 11(5): 11-14
- [2] 金秀满, 吴华, 朱晓华. 影响农业机械化发展因素的灰色关联度分析法[J]. 农机化研究, 1994, 8(3): 10-12
- [3] 余锦华, 王振堂. 经济预测方法程序和实例[M]. 广州: 中山大学出版社, 1996. 206-208
- [4] 卢纹岱. SPSS for Windows 统计分析[M]. 北京: 电子工业出版社, 2002. 21-32
- [5] 彭志捌, 蒋丽娟, 张凤. 财政收入的逐步回归分析[J]. 乌鲁木齐齐职业大学学报, 2004, 13(2): 39-41
- [6] 方开泰. 实用回归分析[M]. 北京: 科学出版社, 2002. 15-18
- [7] 俞大刚. 线性回归模型分析[M]. 北京: 中国统计出版社, 1996. 58-67
- [8] 山西地图集编委会. 山西省自然地图集[M]. 上海: 上海中华印刷厂, 1984. 21-23
- [9] 山西省统计局. 山西省农村经济统计年鉴(2002年)[M]. 北京: 中国统计出版社, 2003. 213-217