Advances in Social Sciences 社会科学前沿, 2016, 5(4), 579-585
Published Online September 2016 in Hans. http://www.hanspub.org/journal/ass
http://dx.doi.org/10.12677/ass.2016.54081

Hans汉斯

The Prediction of Grain Production in Shanxi Province

Mengdi Zhai

Yunnan University of Finance and Economics, Kunming Yunnan
Email: 540532394@qq.com

Received: Aug. 10th, 2016; accepted: Aug. 24th, 2016; published: Aug. 31st, 2016

Copyright © 2016 by author and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

In this paper, the cities in Shanxi Province are divided into southern part, middle part and northern part, and the grain planting area is used as the auxiliary variable. Then, the total grain yield is estimated by five methods, and the most suitable method is chosen for estimating the total grain output by comparing the accuracy.

Keywords

GDP, Regression Estimate, Ratio Estimator

山西省粮食产量预测

翟梦迪

云南财经大学，云南 昆明
Email: 540532394@qq.com

收稿日期：2016年8月10日；录用日期：2016年8月24日；发布日期：2016年8月31日

摘 要

本文将山西省的城市按南部、中部和北部分为三层，选取粮食种植面积为辅助变量，进而用五种方法对粮食总产量进行估计，通过比较精度选择出估计粮食总产量最适合的方法。

文章引用: 翟梦迪. 山西省粮食产量预测[J]. 社会科学前沿, 2016, 5(4): 579-585.
http://dx.doi.org/10.12677/ass.2016.54081

关键词

GDP，回归估计，比估计

1. 引言

保证国家粮食安全是一项长期的、须臾不可放松的历史重任，粮食关系着国计民生，是一个具有时间和空间永恒性的问题。合理利用与保护耕地资源，增加粮食生产能力，是实现社会经济可持续发展的重要课题。在以往的研究中可知，相同情况下，分层抽样的误差要小于其他概率抽样(简单随机抽样、整群抽样、系统抽样等)。因此本文选取分层抽样的方法，将山西省城市按照南部、中部和北部分为三层。分别利用比估计和回归估计的方法来计算粮食总产量。已知粮食种植面积与粮食总产量有显著的相关关系，所以本文选择粮食种植面积作为辅助变量，进而利用简单估计、比估计和回归估计来预测粮食总产量。

对粮食产量的预测方面，国内学者主要采用了多元线性模型、灰色预测等时间序列方法，如周永生等[1]对影响粮食产量的各种因素进行分析，应用多元线性回归分析法建立了广西粮食产量的预测模型。孙东升等[2]利用 HP 滤波法将粮食产量分离为时间趋势序列和波动序列，对趋势序列建立了关于时间 t 的趋势模型，以及王步祥[3]系统评价了国内外有关粮食产量及灰色系统理论研究的状况，回顾了我国粮食生产的历史阶段，深入系统剖析了我国粮食发展的现状，从现有数据及资料出发，分析了我国各大区域及粮食主产区的生产情况。但是利用比估计和回归估计方法研究粮食产量的成果较少，但这种方法存在其合理性和实用性，因此本文从此角度出发，从山西省统计信息网中，得到 2013 年粮食产量与种植面积数据，利用简单估计、比估计和回归估计的方法来预测粮食产量，并进行精度比较。

2. 理论基础

2.1. 分层抽样

分层抽样又称为类型抽样或分类抽样，即在每一层中独立进行抽样，最后将各层样本组成总的样本，由于总体参数未知，所以利用各层抽样得到的样本对参数进行估计，这种抽样就称为分层抽样。

2.2. 比率估计

实际中我们真正关心的变量 Y 通常不易获得衡量数据，那么，遇到这种问题通常退而寻找一个与 Y 有关的变量 X，称 X 为辅助变量，并且 X 的总体总值需为已知的。在实际抽样调查中，选取辅助变量 X 通常出于以下几个原因：

1) 同一个变量的前一期调查结果存在着当期与前一期相比变化不会太大的假设，即不会因为该量造成很大的估计误差；

2) 与主要变量之间整体上存在某种比值关系，即隐含着两者比值关系的变化不会太大的假设，即不会因为利用二者的比值而造成很大的估计误差。

比率估计通常分为分别比估计和联合比估计，简而言之，分别比估计就是先“比”后“加权”，而联合比估计就是先“加权”后“比”，具体计算过程如下：

分别比估计，总体均值和总体总量的分别比估计为：

$$\bar{y}_{RS}=\sum_{h=1}^{L}W_h\bar{y}_{Rh}=\sum_{h=1}^{L}W_h\frac{\bar{y}_h}{\bar{x}_h}\bar{X}_h$$

$$\hat{Y}_{RS}=N\overline{y}_{RS}=\sum_{h=1}^{L}N_h\frac{\overline{y}_h}{\overline{x}_h}\overline{X}_h=\sum_{h=1}^{L}\frac{\overline{y}_h}{\overline{x}_h}X_h=\sum_{h=1}^{L}\hat{Y}_{Rh}$$

统计量的方差为：

$$v\left(\overline{y}_{RS}\right)\approx\sum_{h=1}^{L}\frac{W_h^2\left(1-f_h\right)}{n_h}\left(s_{yh}^2+\hat{R}_h^2s_{xh}^2-2\hat{R}_hs_{xyh}\right)$$

式中，W_h 为层权；f_h 为第 h 层的抽样比；$S_{xh}^2,S_{yh}^2,S_{xyh}^2$ 分别为第 h 层指标 X,Y 的方差以及它们的协方差，$\hat{R}_h=\frac{\overline{y}_h}{\overline{x}_h}$。

联合比估计，总体均值和总体总量的联合比估计为：

$$\overline{y}_{RC}=\frac{\overline{y}_{st}}{\overline{x}_{st}}\overline{X}\overset{def}{=}\hat{R}_c\overline{X}$$

$$\hat{Y}_{RC}=N\overline{y}_{RC}=N\frac{\overline{y}_{st}}{\overline{x}_{st}}\overline{X}=\frac{\overline{y}_{st}}{\overline{x}_{st}}X\overset{def}{=}\hat{R}_cX$$

式中，$\overline{y}_{st}=\sum_{h=1}^{L}W_h\overline{y}_h,\overline{x}_{st}=\sum_{h=1}^{L}W_h\overline{x}_h$ 分别为 $\overline{Y}$ 和 $\overline{X}$ 的分层简单估计量；而 $\hat{R}_c=\frac{\overline{y}_{st}}{\overline{x}_{st}}$。

统计量的方差为：

$$v\left(\overline{y}_{RS}\right)\approx\sum_{h=1}^{L}\frac{W_h^2\left(1-f_h\right)}{n_h}\left(s_{yh}^2+\hat{R}_c^2s_{xh}^2-2\hat{R}_cs_{xyh}\right)$$

2.3. 回归估计

当通过分析发现和之间存在近似的线性关系，但不通过和构成的坐标系的原点，也就是所谓截距为非 0 数，那么这时比率估计就不再适用，违背了其最初的假设，但是两者的线性关系仍为最好的解决问题的入手点，所以利用对的线性回归关系进行估计。

将线性回归估计的思想与分层随机样本的实际情况相结合，类似于比率估计，同样有两个类型，一种是分别回归比估计，主要思想为对每层样本先求回归估计量，然后对各层的回归估计量进行加权平均；另一种是联合回归比估计，方法是对两个变量先分别计算出总体均值或总体总量的分层简单估计量，然后再对它们的分层简单估计量来构造回归估计，具体计算过程如下。

分别回归估计，对于的分别回归估计为：

$$\hat{Y}_{lrs}=\sum_{h=1}^{L}N_h\overline{y}_{lrh}=\sum_{h=1}^{L}N_h\left[\overline{y}_h+b_h\left(\overline{X}_h-\overline{x}_h\right)\right]$$

式中，b_h 为样本回归系数；

对于统计量 y_{lrs} 的方差为：

$$v\left(y_{lrs}\right)=N^2\sum_{h=1}^{L}\frac{W_h^2\left(1-f_h\right)}{n_h\left(n_h-2\right)}\left(n_h-1\right)s_{yh}^2\left(1-r_h^2\right)$$

式中，r_h^2 是第 h 层样本相关系数的平方。

联合回归估计，对于的分别回归估计为：

$$\hat{Y}_{lrc}=N\overline{y}_{lrc}=N\left[\overline{y}_{st}+b_c\left(\overline{X}-\overline{x}_{st}\right)\right]$$

对于统计量 y_{lrc} 的方差为：

$$v\left(y_{lrc}\right)=N^2\sum_{h=1}^{L}\frac{W_h^2\left(1-f_h\right)}{n_h}\left(s_{yh}^2+b_c^2s_{xh}^2-2b_cs_{xyh}\right)。[4]$$

3. 不同的抽样方式用于粮食产量分析的实证

3.1. 总样本量

总体包括山西省的 119 个县(县级市，地级市市区)，拟利用其中的 11 个县(县级市，地级市市区)调查粮食种植面积和粮食产量，因此样本量 $n=11$。

3.2. 层的划分

按山西省不同地理位置，将总体划分为 3 个层，分别对应山西省北部地区、山西省中部地区和山西省南部地区。其中北部地区包括大同、朔州、忻州；中部地区包括太原、阳泉、晋中、吕梁、长治；南部地区包括临汾、晋城、运城。

3.3. 各层样本量

采用比例分配原则确定各层样本量，根据层的大小 N_1， N_2 和 N_3，在总体样本量的基础上确定各层样本量 $n_1=3, n_2=5, n_3=3$。

3.4. 样本抽取

按照随机抽样的准则，利用 SPSS 软件在各层内随机地抽取县(县级市、地级市市区)进行粮食种植面积和粮食产量的统计，最终入选的 11 个样本点分别对应为大同市灵丘县、朔州市平鲁县、忻州市忻府区、太原市晋源区、晋中市左权县、吕梁市交城县、长治市长治县、长治市壶关县、临汾市乡宁县、晋城市泽州县以及运城市芮城县。对上述 11 个样本点进行数据搜集，得表 1。其中 X_{hi} 代表第 h 层的第 i 个样本县(县级市、地级市市区) 的 2013 年的粮食种植面积，Y_{hi} 代表该县(县级市、地级市市区) 2013 年粮食产量。

3.5. 数据整理

根据表 1 中的调查数据，计算得出表 2 中的相关统计量的值。

3.6. 总体总值估计

基于上述数据整理的结果，采用分层随机抽样的分别比估计、联合比估计、分别回归估计和联合回归估计对总体总值做出估计。

1) 简单估计

$$\hat{Y}_{st}=N\sum_{h=1}^{L}W_h\bar{y}_h=16299883.5\approx16299884$$

$$s_h^2=\frac{\sum_{i=1}^{n_h}\left(y_{hi}-\bar{y}_h\right)^2}{n_h-1}$$

$s_1^2=1.964\times10^{10}$， $s_2^2=2350326754$， $s_3^2=1.472\times10^{10}$

$$\begin{aligned}v\left(\bar{y}_{st}\right)&=\sum_{h=1}^{L}\frac{W_h^2s_h^2}{n_h}-\sum_{h=1}^{L}\frac{W_hs_h^2}{N}\\&=1005502958-89044846.1\\&=916458111.506107\end{aligned}$$

Table 1. Data table
表 1. 数据表

h_1			h_2			h_3		
i	x_{hi}	y_{hi}	i	x_{hi}	y_{hi}	i	x_{hi}	y_{hi}
1	31,348	79,918	1	3390	22,973	1	27,473	81,339
2	44,975	66,130	2	11,592	56,252	2	66,668	227,984
3	50,094	315,455	3	9522	46,384	3	66,176	322,099
			4	19,411	132,760			
			5	16,174	121,833			

Table 2. Calculation result
表 2. 计算结果

	h_1	h_2	h_3	总计
N_h	31	52	36	N = 119
n_h	3	5	3	n = 11
$W_h = N_h/N$	0.2605	0.43697	0.30252	1
f_h	0.09677	0.09615	0.08333	-
$(1-f_h)/n_h$	0.301076667	0.18077	0.305556667	-
X_h	980,170	1,013,457	1,396,507	3,390,134
$\overline{X}_h = X_h/N_h$	31,618.387	19,489.557	38,791.861	89,899.81
$\overline{y}_h = \sum_{i=1}^{n_h} y_{hi}/n_h$	153,834.33	76,040.4	210,474	-
$\overline{x}_h = \sum_{i=1}^{n_h} x_{hi}/n_h$	42,139	12,017.8	53,439	-
s_{yh}^2	19,638,457,156	2,350,326,754	14,721,294,475	-
s_{xh}^2	93,885,301	38,195,665	505,735,383	-
$s_{xyh} = r_h s_{yh} s_{xh}$	644,755.2664	466,424.969	3092515.181	-
$\hat{R}_h = \overline{y}_h/\overline{x}_h$	3.6506403	6.3273145	3.9385842	-
b_h	9.77	7.527	4.95	-
r_h^2	0.456976	0.9216	0.840889	-

$$\sqrt{v\left(\hat{Y}_{RS}\right)} = N\sqrt{v\left(\overline{y}_{RS}\right)} = 119 \times 30273.05917 = 3602494.041$$

2) 分别比估计

$$\hat{Y}_{RS} = \sum_{h=1}^{3} \frac{\overline{y}_h}{\overline{x}_h} X_h = \sum_{h=1}^{3} \hat{R}_h X_h$$
$$= 980170 \times 3.65064034 + 1013457 \times 6.327314484 + 1396507 \times 3.93858418$$
$$= 15490969.67 \approx 15490970$$

$$v\left(\overline{y}_{RS}\right) \approx \sum_{h=1}^{3} \frac{W_h^2\left(1-f_h\right)}{n_h}\left(s_{yh}^2 + \hat{R}_h^2 s_{xh}^2 - 2\hat{R}_h s_{xyh}\right) = 1190776243$$

所以

$$\sqrt{v\left(\hat{Y}_{RS}\right)}=N\sqrt{v\left(\bar{y}_{RS}\right)}=4106407.478$$

3) 联合比估计

$$\begin{aligned}\bar{y}_{st}&=\sum_{h=1}^{3}W_h\bar{y}_h\\&=0.26050\times153834.3333+0.43697\times76040.4+0.30252\times210474\\&=136973.8119\approx136973.81\end{aligned}$$

$$\begin{aligned}\bar{x}_{st}&=\sum_{h=1}^{3}W_h\bar{x}_h\\&=0.26050\times42139+0.43697\times12017.8+0.30252\times53439\\&=32394.99385\approx32394.99\end{aligned}$$

所以

$$\hat{R}_c=\frac{\bar{y}_{st}}{\bar{x}_{st}}=\frac{136973.81}{32394.99}=4.228240096\approx4.22824$$

$$\hat{Y}_{RC}=\hat{R}_cX=4.22824\times3390134=14334300.18\approx14334300$$

$$v\left(\bar{y}_{RC}\right)\approx\sum_{h=1}^{3}\frac{W_h^2\left(1-f_h\right)}{n_h}\left(s_{yh}^2+\hat{R}_c^2s_{xh}^2-2\hat{R}_cs_{xyh}\right)=1203751269$$

所以

$$\sqrt{v\left(\hat{Y}_{RC}\right)}=N\sqrt{v\left(\bar{y}_{RC}\right)}=119\times\sqrt{1203751269}\approx4128719.14$$

4) 分别回归估计

$$\hat{Y}_{lrs}=\sum_{h=1}^{3}N_h\bar{y}_{lrh}=\sum_{h=1}^{3}N_h\left[\bar{y}_h+b_h\left(\bar{X}_h-\bar{x}_h\right)\right]=13428006.8\approx13428007$$

$$v\left(\bar{y}_{lrs}\right)=\sum_{h=1}^{3}\frac{W_h^2\left(1-f_h\right)}{n_h\left(n_h-1\right)}\left(n_h-1\right)s_{yh}^2\left(1-r_h^2\right)=296102015.6$$

所以

$$\sqrt{v\left(\hat{Y}_{lrs}\right)}=N\sqrt{v\left(\bar{y}_{lrs}\right)}=119\times\sqrt{296102015.6}=2047706.19$$

5) 联合回归估计

$$b_c=\frac{\sum_{h=1}^{3}\left(W_h^2\left(1-f_h\right)s_{xyh}/n_h\right)}{\sum_{h=1}^{3}\left(W_h^2\left(1-f_h\right)s_{xh}^2/n_h\right)}=0.006660447$$

$$\hat{Y}_{lrc}=N\bar{y}_{lrc}=N\left[\bar{y}_{st}+b_c\left(\bar{X}-\bar{x}_{st}\right)\right]=16345461.54\approx16345462$$

$$v\left(\bar{y}_{lrc}\right)=\sum_{h=1}^{3}\frac{W_h^2\left(1-f_h\right)}{n_h}\left(s_{yh}^2+b_c^2s_{xh}^2-2b_cs_{xyh}\right)=894145033.4$$

所以

Table 3. Result comparison
表 3. 结果对比

估计方法	$\hat{Y}$	$\sqrt{v(\hat{Y})}$
简单估计	16,299,884	3602,494.04
分别比估计	15,490,970	4,106,407.48
联合比估计	14,334,300	4,128,719.14
分别回归估计	13,428,007	2,047,706.19
联合回归估计	16,345,462	3,558,368.70

$$\sqrt{v\left(\hat{Y}_{lrc}\right)} = N\sqrt{v\left(\bar{y}_{lrc}\right)} = 119\times\sqrt{894145033.4} = 3558368.702$$

因此，运用五种方法对总体总值进行估计，得到的估计量分别为：简单估计为 16,299,884，分别比估计为 15,490,970，联合比估计为 14,334,300，分别回归估计为 13,428,007，联合回归估计为 16,345,462。

3.7. 精度比较

对以上五种方法所得的结果总结于表 3 进行比较。

从表 3 可以看出，针对本问题来说，有：1) 回归估计的效果均好于比估计和简单估计；2) 对于粮食产量的预测，分别回归估计的误差最小，效果较优。

4. 结论

本文采用分层随机抽样方法抽取了山西省 11 个样本县(县级市、地级市市区)，然后收集样本区的粮食种植面积，运用统计方法中的简单估计、比估计法和回归估计法对下一年的粮食总产量进行了预测，对总体做出了有效估计。

在本项调查研究中，相较于简单估计和比估计而言，回归估计法的误差更小，估计的精度更高，具有更高的可信度。这为今后基于粮食播种面积调查的样本数据进行总体估计提供了一条新的优化技术路线，即充分利用可以得到的辅助信息，如粮食种植面积，巧妙借助回归估计法，尤其是利用分别回归估计的方法，提高总量估计的精确性和可靠性，如估计粮食总产量。

参考文献 (References)

[1] 周永生, 肖玉欢, 黄润生. 基于多元线性回归的广西粮食产量预测[J]. 南方农业学报, 2011, 42(9): 1165-1167.

[2] 孙东升, 梁仕莹. 我国粮食产量预测的时间序列模型与应用研究[J]. 农业技术经济, 2010(3): 97-106.

[3] 王步祥. 基于灰色系统理论的我国粮食产量预测研究[D]: [硕士学位论文]. 镇江: 江苏大学, 2009.

[4] 金勇进. 抽样技术[M]. 北京: 中国人民大学出版社, 2012.

期刊投稿者将享受如下服务：

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击：http://www.hanspub.org/Submission.aspx