DOCUMENT RESUME

ED 091 098 RC 007 854

AUTHOR Storms, Judith
TITLE Population Estimates for Rural Ccmmunities.
INSTITUTION Arizona State Dept. of Economic Planning and Development, Phoenix.
SPONS AGENCY Four Corners Regional Commission, Santa Fe, N. Mex.
PUB DATE May 73
NOTE 13p.

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS Area Studies; Community Size; *Community Surveys; Demography; Electrical Systems; Elementary Schools; Geographic Regions; *Multiple Regression Analysis; *Population Trends; Residence Requirements; *Rural Population; Statistical Data; Student Enrollment; Telephone Communication Systems; Urban Areas; *Utilities
IDENTIFIERS *Arizona

ABSTRACT

Initial investigations were made into the feasibility of estimating populations through a multiple regression technique during summer 1972. There was a relationship between population and the 2 independent variables tested--school enrollment and residential electrical hook-ups--when allowance was made for size through the use of dummy variables. The sample was limited to 14 communities. Research continued into fall 1972 with the sample size increased to 35 communities. The expanded sample was tested to see if a relationship still existed between the population and electrical hook-ups and school enrollment. Since satisfactory results could not be obtained using just the variables, dummy variables were added as independents. The problem was then approached from a regional perspective, using 5 basic regions which were defined on the basis of geographical proximity. An inherent weakness was that 4 regions included 6 or less communities. Thus, the simple approach to calculating people per residential electrical hook-up was used for the following dates: January 1970; January 1972; and January 1973. Thirty-six communities were used in January 1973. This report presents these telephone, electrical, and school data on appendix form. (NQ)

OFFICE OF THE GOVERNOR
By and Through The Planning Division
Office of Economic Planning and Development

POPULATION ESTIMATES FOR RURAL COMMUNITIES

State of Arizona
Office of the Governor

by the

Planning Division
Department of Economic Planning and Development

Principal Researcher
Judith Storms

May, 1973

The preparation of this report has been financed in part through coordinated Technical Assistance Grants from the Four Corners Regional Commission

TABLE OF CONTENTS

INTRODUCTION

Last summer, some initial investigations were made into the feasibility of estimating populations through a multiple regression technique. The results, reported in SUMMARY REPORT ON SMALL COMMUNITY ANALYSIS, were significant enough to warrant further study of the method.

Research was continued into fall, 1972. Sample size was increased and alternative regression relationships were tested. The results have not been encouraging. While significant relationships have been found, none of the multiple regression equations yielded results with the precision necessary for confidently applying them in the numerous small, nonmetropolitan communities in the state.

The results of the various regression approaches tried are presented in this paper. In addition, the results of the more simplistic approach of people per utility hook-up are presented at the community level. It appears as though this type of approach, while the simplest, is also the recommended.

A by-product of this study has been the generation of significant amounts of consistent data on the communities used. In the hopes that these data will be of use to other researchers in their work, it is presented in summary form in an appendix.

TESTING FOR A SIZE RELATIONSHIP

The work done in the summer of 1972 concluded that there was a relationship between population and the two independent variables tested–school enrollment and residential electrical hook-ups–when allowance was made for size through the use of dummy variables. The sample upon which these conclusions were based was limited, however, to 14 communities.

When the study was continued into the fall, one of the first steps was to increase the sample size to 35. Data on utility connections and elementary school enrollment were collected for each community. The expanded sample was tested to see if a relationship still existed between the population figure and electrical hook-ups and school enrollment.

Significant relationships were found--the best fit being expressed in the equation:

$$\underset{}{Pop_{70}} = \underset{(.587)}{1.96\ Sch_{70}} + \underset{(.351)}{2.14\ Util_{70}}$$

$$R^2 = .988$$

Standard Error of the Estimate = 1317.2

$$\overline{Pop} = 5778.9$$

Where Pop_{70} = the estimated 1970 population
$Util_{70}$ = residential electrical hook-ups in 1970
Sch_{70} = elementary school enrollment in 1970

While this regression yielded significant t statistics and a good coefficient of correlation, the standard error of the estimate was much too large for the precision needed in estimating populations for communities as small as 773.

Several alternative sample sizes, which eliminated communities above or below certain levels, were tested to see if this would increase the precision of the results. Unfortunately, this did not occur. In no case were the percentage deviations of the predicted from the observed all within what was judged to be a reasonable range.

For the purposes of these population estimates, 10 percent was decided upon as the maximum percentage deviation acceptable.

Since satisfactory results could not be obtained using just the two independent variables, school enrollment and utility connections, a selection of dummy variables were added as independent variables. As this had produced satisfactory results during the summer, it was hoped it would do so again.

In each attempt to add dummy variables, t statistics were obtained where they were not significant at the 90 or 80 percent levels of confidence. This says that segmenting communities within a sample according to size, did not better define a relationship. While these results were disappointing, the concept of a multiple regression approach was not yet abandoned. Instead, it was decided to approach the problem from a different perspective–from a regional outlook.

THE REGIONAL APPROACH

Since satisfactory results could not be obtained through the previous methods, the regional approach was attempted. This approach was intuitively appealing. If the state was divided into regions, it seemed plausible that communities in the same general area might exhibit similar functional relationships: that is, one might find the same relationship between population, utilities, and enrollment within a set of communities.

All of the communities for which data were available were plotted on a map. Regions were then defined on the basis of geographical proximity. Several communities were initially placed in two regions until proper placement could be determined. Five basic regions were sketched out with a sixth potential region allowed for. As it turned out, the sixth region was not needed.

Separate regressions were run for each region. The results were satisfactory. In four of the five regions, t statistics were significant at least the 90 percent level; which in the other region, the t statistic was significant at the 80 percent level. In all regions, the standard errors of the estimate were much improved over previous methods. With the exception of three communities, percentage deviation of the predicted from the observed was under 10 percent.

The regions as defined and used were as follows:

Region I:

$$Pop_{70} = \underset{}{-578} + \underset{(.333)}{1.24\ Sch_{70}} + \underset{(.244)}{2.68\ Util_{70}}$$

$$R^2 = .999$$

Standard Error of the estimate = 223.9

$$\overline{Pop} = 5487.2$$

COMMUNITIES: Tombstone, Bisbee, Douglas, Benson, Willcox

Region II:

$$Pop_{70} = 323 + 2.75\ Sch_{70} + 1.24\ Util_{70}$$
$$(.668) \qquad (.314)$$

$$R^2 = .9995$$

Standard Error of the Estimate = 168.5

$$\overline{Pop} = 4141$$

COMMUNITIES: Clifton-Morenci, Duncan, Safford, Thatcher

Region III:

$$Pop_{70} = -261 + 2.11\ Sch_{70} + 2.09\ Util_{70}$$
$$(.393) \qquad (.228)$$

$$R^2 = .9993$$

Standard Error of the Estimate = 325.45

$$\overline{Pop} = 7259$$

COMMUNITIES: Florence, Sommerton, Hayden-Winkelman, Miami, Superior, Coolidge, Eloy, Globe, Casa Grande, Yuma

Region IV:

$$Pop_{70} = 235.59 + 1.27\ Sch_{70} + 2.06\ Util_{70}$$
$$(.382) \qquad (.226)$$

$$R^2 = .9988$$

Standard Error of the Estimate = 179.57

$$\overline{Pop} = 3871$$

COMMUNITIES: Springerville-Eagar, Showlow, Snowflake, Holbrook, Winslow

Region V:

$$Pop_{70} = -1264 + 14.60\ Sch_{70} - 3.18\ Util_{70}$$

$$R^2 = .9975$$

Standard Error of the Estimate = 874.52

$$\overline{Pop} = 7188$$

COMMUNITIES: Williams, Wickenburg, Parker, Cottonwood, Prescott, Flagstaff

From a statistical point of view, these regression equations met the criterion necessary for acceptability. Despite the low degrees of freedom, the t statistics were sufficiently high to still be significant at 80 or 90 percent confidence levels. It would have been simple at this point to generate 1972 population estimates based on these equations and call the work done. However, after the misleading results of the summer's work, a closer analytical look at the results seemed appropriate.

The equations do seem to confirm a set of regional relationships. However, even given a regional relationship, the question can be raised as to whether or not these equations should be used to freely estimate population. It is an inherent weakness that four regions have six or less communities in them. This results in only 1, 2, or 3 degrees of freedom—a definite constraint.

Two solutions to this problem seem apparent. First, increase the number of communities within the region by gathering more data. This is unworkable for the data simply are not available. Second, combine regions so as to increase degrees of freedom. This was tried and the result was a breakdown in the relationship between population, school enrollment, and utility connections.

Since the means of improving the degrees of freedom problem were stymied, one is faced with a set of regression equations, statistically significant, that could be inaccurate due to too few observations.

As a result, it was decided to use the findings selectively. The findings were interpreted to identify general regional similarities in the relationship between dependent and independent variables. However, specific population estimates were not generated from the equations due to the smallness of sample size. Instead, the general knowledge of regions was combined with a more simplistic approach toward generating specific population figures.

SIMPLE POPULATION ESTIMATES

Rather than produce no population estimates, the simple approach of calculating people per residential electrical hook-up in 1970 was used. The people per hook-up figure was applied to number of hook-ups on January 1, 1972, and 1973 to produce population estimates for three dates.

These results are presented in Table 1 for 36 communities in the state for which residential utility hook-up figures were available. In addition, average figures were calculated for each region as defined in the regression equations. These are shown in Table II. While these averages were not significantly different when tested, it was felt they might be of use in estimating population for communities with no 1970 population figures.

DISCUSSION

It is disappointing to find that an approach which appeared workable in preliminary testing is inappropriate for general application. The people per hook-up approach is not as satisfying as a regression approach would have been for it does not recognize any functional relationships across communities in the state. Each community is treated on an individual basis; whereas, if the regression equation had proved applicable, functional relationships would have been established that could have been applied confidently on a broader basis than simply the sample communities.

As it stands, the people per hook-up approach appears to be the best alternative available. Its weaknesses are recongized, particularly the continuing tendency of apartment complexes to have one utility hook-up for all apartments. Other alternatives simply are not workable.

The best alternative would be population estimates based on actual birth, death, and migration statistics. Migration estimation proves a problem at any level of analysis in Arizona; however, at the community level, even birth and death figures are problematic. This is due to the fact that included in births and deaths for a community are those figures on births and deaths occurring outside the city limit that are reported to the community.

One side product of this continued study, has been the large amount of statistical data which has been collected. Residential electrical hook-ups, elementary school enrollment figures within city limits, and residential telephone hook-ups have been gathered for approximately 36 communities. The telephone figures were gathered for use in a second phase once the regression equations had been established. They also were tried as independent variables in the regressions, but this was not successful.

These telephone, electrical and school data are presented in this report in appendix form. One of the foremost problems in analyzing any small community is the difficulty in obtaining data, and it is hoped that having these figures in one central volume will be a step towards alleviating the problem.

TABLE I

	People Per Utility Connection	Population April 70	Population Est. Jan 72	Population Est. Jan 73
Benson	2.85	2,839	3,311	2,867
Bisbee	3.08	8,328	8,676	8,500
Casa Grande	3.94	10,536	11,221	12,619
Clarkdale-Jerome	2.18	1,182	1,308	1,168
Clifton-Morenci	2.67	8,140	6,912	6,955
Coolidge	3.62	5,314	5,484	5,756
Cottonwood	2.49	2,815	3,244	3,329
Douglas	3.47	12,462	12,703	12,637
Duncan	3.37	773	839	775
Eloy	4.67	5,381	5,893	6,337
Flagstaff	3.51	22,612*	24,373*	24,892*
Florence	3.53	2,173	2,562	2,619
Globe	3.55	7,333	7,490	7,920
Hayden-Winkelman	3.21	2,257	2,275	2,314
Holbrook	3.74	4,759	4,884	4,944
Kearny	4.37	2,829	3,056	3,068
Mammoth	4.07	1,953	2,079	2,356
Miami	3.54	3,394	3,416	3,465
Page	2.19	1,439	3,111	3,177
Parker	2.53	1,948	2,074	1,725
Patagonia	2.54	630	744	662
Pima	3.14	1,184	1,208	1,275
Prescott	2.53	12,700*	13,153*	13,813*
Safford	3.25	5,333	4,989	5,429
San Manuel	3.81	4,332	5,097	5,246
Show Low	3.78	2,129	2,509	2,790
Snowflake	4.37	1,977	2,189	2,189
Sommerton	4.54	2,225	2,655	2,869
Springerville-Eagar	3.34	2,430	2,631	2,785
Superior	3.45	4,975	5,164	5,275
Thatcher	4.29	2,082*	2,243*	2,372*
Tombstone	2.94	1,241	1,390	1,387
Willcox	2.54	2,568	2,646	2,085
Williams	3.37	2,386	2,470	2,544
Winslow	2.95	8,066	8,009	7,478
Yuma	3.32	29,007	30,341	31,274

*Population adjusted for institutional inmates and those in group quarters

TABLE II

Regional Averages - People Per Utility Hookup

Region	Average
I	2.97
II	3.40
III	3.70
IV	3.63
V*	2.94

*Wickenburg Omitted

APPENDIX A

Residential Electrical Hookups

	January 1970	January 1972	January 1973
Benson	995	1162	1006
Bisbee	2702	2817	2760
Casa Grande	2669	2848	3203
Clarkdale - Jerome	540	600	536
Clifton-Morenci	3039	2589	2605
Coolidge	1466	1515	1590
Cottonwood	1130	1303	1337
Douglas	3589	3661	3642
Duncan	229	249	230
Eloy	1151	1262	1357
Flagstaff	6431	6944	7092
Florence	615	726	742
Gila Bend	490	514	517
Globe	2060	2110	2231
Hayden-Winkelman	703	709	721
Holbrook	1272	1306	1322
Kearny	647	699	702
Mammoth	479	511	579
Miami	958	965	979
Page	656	1421	1451
Parker	769	820	682
Patagonia	248	293	261
Pima	376	385	406
Prescott	5010	5189	5460
Safford	1746	1636	1780
San Manuel	1135	1338	1377
Show Low	563	664	738
Sierra Vista	1883	2391	3552
Snowflake	452	501	501
Sommerton	490	585	632
Springerville - Eagar	726	788	883
Superior	1440	1497	1529
Thatcher	485	523	553
Tombstone	421	473	472
Willcox	1009	1042	820
Williams	707	733	758
Winslow	2732	2715	2535
Yuma	8735	9139	9420

APPENDIX B

Elementary School Enrollment Within City Limits

	January 1970	January 1972
Benson	515	569
Bisbee	1306	1088
Casa Grande	2298	1905
Clarkdale - Jerome	216	225
Clifton - Morenci	1441	1446
Coolidge	1121	1067
Cottonwood	469	755
Douglas	2778	2805
Eloy	1536	1550
Flagstaff	3260	3934
Florence	585	595
Gila Bend	363	475
Globe	1806	1755
Hayden - Winkelman	531	491
Holbrook	1378	1307
Kearny	583	393
Kingman	1639	1651
Mammoth	435	484
Miami	671	679
Nogales	2004	2078
Page	331	644
Parker	369	373
Pima	286	279
Prescott	2266	2266
Safford	1067	992
San Manuel	495	721
Show Low	527	599
Snowflake	782	893
Sommerton	557	651
Springerville - Eagar	513	547
Superior	1118	1090
Thatcher	466	460
Tombstone	471	590
Wickenburg	259	282
Willcox	576	562
Williams	483	464
Winslow	1767	1647
Yuma	5214	5177

APPENDIX C

Residential Urban Telephone Hookups [1]

	January 1970	January 1972
Ajo	1433	1558
Benson	830	980
Bisbee	2689	2911
Camp Verde	572	802
Casa Grande	2449	2710
Coolidge	1524	1750
Cottonwood	1206	1568
Douglas	3188	3550
Duncan	221	307
Eloy	937	1089
Flagstaff	5435	6479
Florence	538	599
Gila Bend	320	363
Globe	3226	3570
Holbrook	846 [2]	930 [2]
Joseph City	101	120
Mammoth	87	99
Miami	1434	1564
Nogales	2311	2818
Page	335	904
Parker	413 [2]	448 [2]
Patagonia	129	176
Pima	254	292
Prescott	4957	5828
Safford	1948	2085
San Manuel	832 [2]	978 [2]
St. Johns	250 [2]	275 [2]
Show Low	407	509
Snowflake	502 [2]	640 [2]
Sommerton	351	405
Springerville - Eagar	449	532
Superior	949	1041
Tombstone	318	399
Welton	198	178
Wickenburg	998	1203
Willcox	780	845
Williams	526	572
Winslow	1969	2044
Yuma	9600	11441

1 - Source: Mountain Bell unless otherwise specified
2 - Supplied by Western States Telephone