A MODEL FOR PREDICTING AIR TRAVEL DEMAND IN SMALL COMMUNITIES

by

Rick Donnelly

B.S., Kansas State University. 1983

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF REGIONAL AND COMMUNITY PLANNING

Department of Regional and Community Planning

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1985

Approved by:

Claude A. Keithly

Major Professor

"Statistics is the art of beating

data until it confesses."

R. Wasserstein, 1985

I would like to acknowledge the assistance of Mr. Ron Wasserstein, good friend and eminent statistician, without whose help this thesis would have been literally impossible.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Planners have used secondary data sources, such as census data. to predict urban travel demand for several years. In the early 1960s, extensive use of origin-destination surveys were made to gather information about the travelling public. In the 1970s, planners set about to discover cost saving and more efficient ways to collect data. Several researchers found that secondary data sources, such as traffic volume counts and censal data. estimated travel demand fairly well (Weiner. 1983). The ability to use such short-cut methods of data collection have enabled planners to employ sketch planning[1] on a wide number of alternatives to policy issues. Without the use of such secondary data sources, the cost of collecting the data to examine such a large number of alternatives would be prohibitive.

While urban transportation planners have used these secondary sources for quite some time. there does not appear to be such a trend in air travel demand forecasting. The development of large airports seems to occur without the guidance of any one governing body. and is frequently tied mainly to observed passenger demands (Meyer & Oster. 1984). Smaller communities generally do not enjoy the luxury of specialized airport planning. Most of the financing for their airport improvements comes from local funding and bond issues[2]. For such communities, the undertaking of airport construction or renovation is a very major burden. A sketch planning tool that planners in these communities could utilize might prove most helpful. as most do not have the resources or personnel to conduct market analyses and surveys. Such a sketch tool would enable planners to determine the feasibility of engaging in airport planning for their community.

A second prospective user for this model might be commuter airline[3] operators. Most commuter services do not have a planner or market analyst in their employ. but could conduct simple studies if the methodology were available. A model such as this would enable them to single out promising prospective markets to which they can devote greater attention.

Existing models of air travel demand for small communities that are reported in the literature are all based in whole or in part on survey data. As previously noted, these studies were expensive and time-consuming to complete. The model developed by Chan (1976) is probably the best model described, and would appear to be useful as a sketch planning tool. However. two factors distract from its utility. First. the data requirements call for survey data. Secondly. the model was constructed during the period of time that carriers were controlled by the Civil Aeronautics Board (CAB). Once a carrier established service to a community. they was frequently required to continue such service although they proved to be unprofitable routes.

When the CAB was disbanded and carriers permitted to operate under economic rationale, many of these routes were immediately abandoned. It would seem reasonable that a model constructed under such conditions may prove a poor estimater in todays environment (Meyer & Oster, 1984).

Purpose Of This Study

This study will attempt to show the utility of using census data to construct models of demand for air travel in smaller communities. Census data is published and widely distributed by the Bureau of the Census, and planners at every level have ready access to such. Using the number of enplaned passengers[4] and aircraft departures for commuter airline operations in Kansas in 1983, a model will be constructed to describe demand. The structure of the model will be compared to the traditional models of demand, and irregularities examined. From this, conclusions will be reached concerning the usefulness of secondary data sources in model construction.

Traditional Determinants of Air Travel Demand

The demand for intercity travel is influenced by the levels of economic and social activities that occur there, as well as other attractions. Such other attractions may include civic, governmental, cultural, and recreational activities. Experience has shown that a communities aviation activity is primarily affected by three factors. The first is the size of the population and its structure. The economic character of the community is the other factor (Kanafani, 1984). Another factor having an influence on travel is the cost, both in terms of money and time (Paquette & Ashford, 1982). Proximity to large metropolitan areas also plays an important role in demand. In order to describe those factors above, a large number of candidate socioeconomic variables could be studied. The three that are cited the most in the literature are population, income, and employment counts.

Trip purpose is perhaps the most influential behavioral component of travel demand (Kanafani. 1983). However. trip purpose can only be obtained from surveys, which are costly and time consuming. Another important aspect of demand is price. which can be obtained through industry sources. For short haul markets, demand is sensitive to price, especially in the presence of competing modes (Kanafani. 1983; Meyer & Oster. 1984).

The demand for air travel is generally measured in one of three manners: by passenger enplanements, by aircraft departures, or in passenger-miles travelled to the nearest hub airport. Variables such as those described above are almost always included in models reported in the literature. although the effect of population is questioned by at least one researcher. Meyer and Oster (1984) found in a survey of 135 New England towns that total population was not a good index of the scale of travel generation taking place. In addition, some also include level of education, fares, and travel time. Other possible factors might include the level of service and frequency of flights.

Hypothesis and Expected Findings

It was anticipated that distance from major airports and population would have the greatest effect on air travel demand. As the distance between nodes increases, it becomes less convenient to commute by automobile, thus steering travelers towards air transport. It was also anticipated that the number of persons in professional and technical occupations would influence demand, as business travelers have been widely reported as comprising the bulk of air travelers on trunk carriers. The existance of certain groups, most notably college students and military personnel. was also seen as influencing factors.

The following chapters describe the process used in this study to arrive at such a model. The appendices contain the output from the computer analysis of the data.

[1] Sketch planning techniques are designed to "analyze a large set of alternatives in a quick and broad based manner- to examine the full set of alternatives under consideration with the objective of identifying a small set of the most promising alternatives, which can then be analyzed in greater detail" (Meyer & Miller. 1984).

[2] This is not entirely true, as some communities have acquired airports as the result of the closing of nearby Air Force bases or fields. In many instances the capacity of these airports far exceeds the potential demand that these communities will generate.

[3] A commuter airline is one which serves a localized geographical area. providing service between major hub airports and small outlying communities. These communities are too small to attract service by certificated carriers. Commuter airlines may have either scheduled or "on demand" operations.

[4] The number of enplaned passengers is defined as the total number of passengers departing from the study area in question. Incoming passengers are not included in this total. as it would have the effect of double counting the number of passengers. It is assumed that most, if not all. incoming passengers will depart using the same mode at some point in the future (Meyer & Oster. 1984).

STUDY METHODOLOGY

The goal of this study was to arrive at a satisfactory mathematical model to describe air travel demand. The most appropriate technique to isolate such, given the character of the data, is regression analysis. Regression analysis evaluates the relationship of one or more independent variables to a dependent variable. By its very definition, this dependent variable is uncontrollable, but predicted by the values of the independent variables. It is worth noting that a causal relationship is not shown by regression alone, but that additional testing and experimentation are necessary to establish such (Kleinbaum and Kupper, 1978; Willemain, 1980). The purpose of this study was not to establish causal factors, but rather to develop a simple descriptive model.

The first step in any study is to carefully evaluate the subject(s) or area under study. In this study, all communities in Kansas served exclusively by commuter air service were considered. Table 1 lists the cities included in the study. Figure 1 shows the cities and counties included in the study. Wichita was excluded from the study, as it is a major hub, with service by various trunk carriers. It was expected that Topeka, the state capital, would need to be excluded from the survey. However, since it filled the requirement above, it was included at the onset.

Figure 1: Cities and Counties Included in the Study

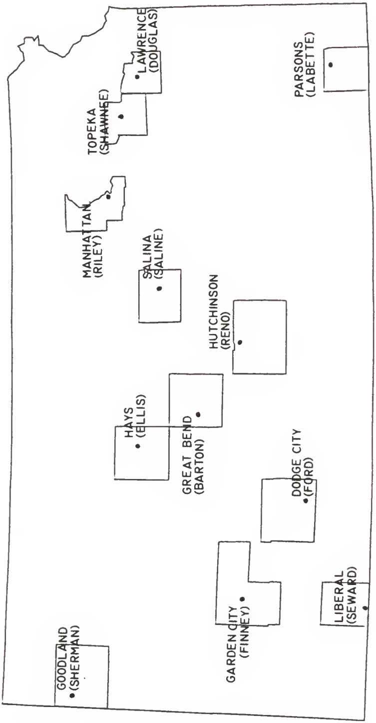

TABLE 1

Cities Included in the Study

Dodge City	Lawrence
Garden City	Liberal
Goodland	Manhattan
Great Bend	Parsons
Hays	Salina
Hutchinson	Topeka

Next, a list of variables that might have an effect on air travel demand was compiled. Table 2 lists the variables selected. Some of the variables were chosen because they had been shown in other studies to have an influence on demand, while others were selected in order to determine the significance of their inclusion in the model, if any.

The delineation of the study area was considered at great length. It was decided that the county in which the airport was located would be the study area. Various other schemes, such as including all cities or persons in a certain radius, appeared to be difficult to implement. The census data is broken down into county units and for places of over 2500 in population. The relative ease of using the county data was felt to be a significant factor in constructing a simple model.

The measures of these variables for each county in the study were obtained from censal publications in the Kansas State University Library. Once the observations for each independent variable were recorded, the data for the dependent variables EP and AD were obtained. Attempts to solicit this information from

TABLE 2

Variables Included in the Study

Variable Name	Description
Dependent Variables	
AD	number of aircraft departures
EP	number of enplaned passengers
Independent Variables	
AGE	median age
AGRI	number of persons employed in agriculture
COLLEGE	number of college graduates
DISTANCE	road distance from the airport to the nearest hub
HOUSES	number of households
HSGRAD	number of high school graduates
HUNITS	number of housing units
HVALUE	median value of houses, in 1979 dollars
INCOME	median income of households
LABOR	number of persons employed in trade and labor jobs
OVER15	number of households earning >= $15.000 per year
PHOUSES	number of persons per household
POP	county population
PROF	number of persons employed in professional jobs
SERVICE	number of persons employed in service jobs
SPGEN	number of special generators (college + military)
TECH	number of persons employed in technical jobs

the commuter airlines themselves proved unsuccessful. The FAA provided the information requested for the cities from calender year 1983. These observations, combined with those of the independent variables, were placed together in a computer dataset for subsequent computer analysis. The Statistical Analysis System (SAS) computer package was used to analyze the data for this study.

It was realized from the start of this study that the small number of observations in the dataset would make drawing conclusions from the results obtained difficult. Unfortunately. the data for AD and EP were hard to collect[1], which dictated doing the study with scant data. The original plan was to obtain data for all cities served by commuter air service in Kansas, Nebraska. Oklahoma. and Missouri A study area of that size would have provided the necessary number of observations to satisfy the minimum sample size requirements. A detailed discussion of the problems of a small sample size is presented in the next section.

The strategy for deriving the desired model consisted of the following steps, recommended by Kleinbaum and Kupper (1978):

1. Begin by assuming that a straight line is the appropriate model.
2. Find as the best fitting straight line that line. among all possible straight lines, which best agrees with the data.
3. Determine whether or not the straight line found in step 2 significantly helps to describe the dependent variables.
4. Examine whether or not the assumption of a straight line model is correct.
5. If step 4 finds the assumption of a straight line model to be invalid, fit a new model (e.g., a parabola) to the data. determine how well it describes the dependent variables, and then decide whether or not the new model is appropriate.

6. Continue to try new models until an appropriate one is found.

This process is described in detail below.

Dealing With a Small Sample Size

Thirty observations is a widely proclaimed minimum number of observations required upon which base statistical inferences. The results obtained from analyses using fewer observations are frequently suspect. There are no hard and fast rules that can be applied indiscriminately to all situations concerning sample size. When performing exploratory data analysis and original research. it is not uncommon to encounter small sample sizes which may be statistically relevant (Boyer. 1985; Milliken, 1985; Wasserstein, 1985; Willemain, 1980).

The most important change in methodology encountered when using small (<30) samples is that the Students t-distribution is used in place of the traditional Gaussian function to describe probability. When using a small number of observations in regression analysis, the model constructed is generally not considered accurate (Boyer. 1985). The accuracy of the model decreases as the number of independent variables increases. However. when the correlation coefficient r approaches the values of -1. 0. or 1 (within 0.1 or so), the validity of the model obtained is enhanced such that it may be considered reliable (Boyer. 1985; Milliken, 1985; Wasserstein, 1985).

The most apparent impact the small number of observations had on this study was in reducing the scope of the study. Instead of producing a precise mathematical model of demand, a more general descriptive model had to be resorted to.

Selecting a Subset of Variables

It was necessary to determine which of the variables listed in Table 2 actually would be of benefit in predicting the values of AD and EP. The SAS procedure RSQUARE evaluated each combination of the dependent variables with the independent variables. RSQUARE produces the r^2 and Mallows Cp statistic for each possible model. Due to the immense amount of computing time that would have been required to perform this procedure for all of the independent variables, the search was limited to models of up to 4 independent variables. Appendix A contains the listing produced for this study.

SAS uses matrix methods during the computation of regression cases. When the procedure RSQUARE is used, it identifies a condition known as singularity if it is encountered it in the data. Singularity[2] is present when the values of one independent variable can be derived from a simple mathematical operation[3] on another independent variable. In other words, the values of one independent variable can be directly computed knowing the values of another. Such is the ultimate form of collinearity. in that there is a perfect correlation between the two variables.

Evaluation of Influence

A test was performed to measure the influence of each observation on the model. The SAS procedure REG provided diagnostics to evaluate such influence. As a result of this test. Topeka was found to have a large influence on the model derived. Because of its influence, Topeka had to be removed from the dataset and the entire analysis to this point repeated. A discussion of this finding is presented in the next chapter.

Finding the Best Fitting Line

Several statistical tests to isolate the best fitting line are available. The most commonly used methods are forward selection, backward elimination, stepwise regressions, and the maximum r^2 improvement method (Draper and Smith. 1980). In addition, a procedure known as the minimum r^2 improvement method (SAS Institute. 1982) was employed. These methods are adequately described in the literature. and are not elaborated upon here. Green (1978, p.78) presents an excellent description of these techniques. Using the SAS computer package, the data was fitted using all of the above procedures.

Checking for Collinearity

As previously noted, a check for collinearity is required to prove the true independence of the independent variables. There were two methods available to check for collinearity using SAS. The approach used in the study consisted of examining the collinearity diagnostics produced by the SAS procedure REG, which is based on work performed by Belsley ėt al (1980). A second method consists of having SAS print a correlation matrix. This method was not conducted due to the large number of variables present in the dataset. Examination by this method has little advantage over examination of the condition index (Wasserstein, 1985).

Evaluating the Model

The objective of the study was to obtain an acceptable model for description of demand. At this point in the study. it was apparent that an acceptable model had been produced. This is not to say it was the best model. but rather that it had filled the criteria we had defined earlier in the study.

If a suitable straight-line model had not been produced, it would have been necessary to evaluate other models. Having no empirical basis on which to form such a model, proceeding from this point would have required estimating endless varieties of nonlinear models. SAS does not have the capability to fit nonlinear models. Rather. the user must estimate an equation, declare parameter names, guess starting values for them. and specify the derivatives of the model. It would have been both very expensive and time consuming to adopt this approach.

[1] Hard to collect is an understatement. The carriers themselves appeared quite reluctant to supply such information. When the FAA was contacted to supply the information, it advised the author that the search for the requested data would be very costly. as the necessary reports were stored at the CAB warehouse in Washington, D.C. The data was made available for Kansas only because a private party had already requested the data, and it was readily available.

[2] Letting k denote the number of independent variables in the study. we construct a k x k matrix to determine the correlation coefficient for each combination of variables. A square matrix is said to be singular when the determinant is zero. Nonsingularity is required for the existance of a matrix inverse, which is required to perform regression analysis (Searle, 1982).

[3] Such as addition, subtraction, multiplication, taking the square root, and the like.

SUMMARY OF FINDINGS

A straight line model with three variables was selected as the model to describe air travel demand based on the available data. The derived model fits the data better than anticipated, and appears adequate. considering the size of the sample.

Creating a Subset of Variables

The creation of a subset of variables to be considered by the model proved relatively simple. Although 17 variables were selected for consideration in the study. all but 7 were eliminated due to singularity. As previously noted, these variables were expected to be predictors of demand.

It was unfortunate that both income and the presence of special generators had to be eliminated due to singularity. Most texts on demand state that income is an important component of demand. It is not known why such is not the case in this study. Special generators were included in the study to help account for the large number of travelers in Lawrence, Manhattan, and Topeka. It was felt that the presence of colleges and/or military installations would have an effect on demand, as both of these groups are more likely to demand air transportation (Chan, 1982). Table 3 lists the variables that were used in the regression analysis.

TABLE 3

Classification of Study Variables

Used in Regression	Rejected due to Singularity
COLLEGE	AGE
DISTANCE	AGRI
HOUSES	HSGRAD
POP	HUNITS
PROF	HVALUE
SERVICE	INCOME
TECH	LABOR
	OVER15
	PHOUSES
	SPGEN

They may not have been significant in this study for two reasons. First, the number of special generators may have been underestimated. Perhaps including state and federal employees in the count would have helped to arrive at a more accurate count. Secondly, it may not have been weighted properly in the straight line model. It is possible that special generators have more of an influence than other factors, and that a nonlinear model would have demonstrated this fact.

Removal of Topeka from the Study

The check for influence found that Topeka had a profound effect on the data. This is evidenced by examining the DFFITS statistic generated by the SAS REG procedure. DFFITS measures the change in the predicted value for the ith observation created by deleting that observation (SAS Institute, 1982). A large value indicates that the observation is very influential and should be removed from the dataset (Wasserstein, 1985). As seen in Table 4, Topeka (observation 12) generated a DFFITS score of 23.6285, which dwarfs the values for the other cities in the study.

TABLE 4

DFFITS Statistics With and Without Topeka

OBS	DFFITS	
	With Topeka	Topeka Removed
1	0.1186	0.1279
2	0.1770	0.4633
3	0.0861	-2.1983
4	0.0654	0.1474
5	0.0409	0.0242
6	-2.9398	-3.5603
7	-7.7364	-4.6013
8	0.2392	0.2195
9	11.2408	6.0469
10	0.0866	-0.1040
11	-0.3191	0.7139
12	23.6285	

Aside from the statistical justification for its removal, Topeka might also be deleted from consideration on the basis of its size. Topeka is over twice the size of the next largest observation in the dataset, as seen in Appendix A. By removing it from the study, we arrive at communities closer together in size.

After Topeka was removed from the dataset, another measure of influence was performed. As seen in Table 4, the values of DFFITS with Topeka removed are not nearly as extreme, and denote that there are no observations present that unduly influence the model (Boyer. 1985; Wasserstein, 1985). From this point forward, the analysis proceeded without Topeka included.

Exploring Collinearity

A check for collinearity. discussed in the previous chapter. revealed no evidence of such in the subset (Boyer. 1985; Wasserstein, 1985).

Derivation of the Model

The regression fitting techniques discussed in the last chapter produced several models for comparison. In all cases, AD was found to be better predicted by the data than EP. The first procedure used to analyze the data was RSQUARE. which found better fits for a three variable model using AD. Appendix A contains the computer listing of the procedures results. A good candidate model appears to be one which (1) has a high value for r^2. and (2) has a value of Cp that approaches the number of independent variables in the model (Hocking, 1976; Kleinbaum & Kupper. 1978). The best models suggested by procedure RSQUARE used the dependent variable AD and several combinations of 3 independent variables. Table 5 lists the best 3 models suggested by RSQUARE for both AD and EP. On the advice of several statisticians, three variables were selected as the maximum number for this study due to the number of observations (Boyer. 1985; Milliken, 1985; Wasserstein, 1985).

TABLE 5

Three Best Models Proposed By RSQUARE

<u>Dependent Variable AD</u>

r^2	Cp	Variables Included
0.88	2.39	POP PROF HOUSES
0.88	2.37	COLLEGE PROF HOUSES
0.90	1.48	POP COLLEGE PROF

<u>Dependent Variable EP</u>

r^2	Cp	Variables Included
0.55	41.39	POP PROF DISTANCE
0.55	41.34	DISTANCE POP HOUSES
0.57	39.37	DISTANCE POP SERV

All five of the regression fitting methods described in chapter 2 were utilized to arrive at the best model. As EP proved to be poorly predicted by the available data. it was excluded from study from this point forward. The stepwise, maxr. and minr methods arrived at the following three variable model:

$$AD = 1925.726 + 0.101*POP + 0.496*COLLEGE - 1.918*PROF$$

This model had the highest value of r^2 (0.90) and best value for Cp (1.48). The forward and backward regressions each produced a different four variable equation, which were considered unacceptable because of the previously set limit of 3 variables. It is interesting to note that the forward regression included DISTANCE to the three variables in the accepted model. Table 6 summarizes the results of the regression analysis. The computer listing of these procedures is found in Appendix C.

TABLE 6

Model Selections By Procedure REG

Method	Model	r^2	Cp
Forward	COLLEGE PROF POP DISTANCE	0.93	2.35
Backward	DISTANCE COLLEGE PROF HOUSES	0.93	2.18
Stepwise.			
MaxR.	COLLEGE PROF POP	0.90	1.48
MinR			

Utility of the Model

While the model does appear to be statistically pleasing, there were several problems with accepting it as a valid model for planning use. Perhaps the largest problem was that of constructing a model with such a small set of observations. Although an acceptable model was produced, it is heavily biased in respect to place and time. Its ability to produce accurate results within the study area is questionable, and would be even more so if applied elsewhere.

The inaccuracy of the model is seen in Table 7, which shows the actual and predicted values of AD using the derived equation. The large error noted for Goodland may be due to its small size relative to the other communities in the survey. It is also the furthest from a major airport, and distance may be more of an important factor there. The predicted values for Salina and Hutchinson deviated from the actual values as a result of the inadequacy of the model (Boyer. 1985; Wasserstein, 1985). This is in turn the result of using a model constructed from so few observations. This lends justification for reducing the scope of the study to describing the structure of demand only.

TABLE 7

Comparison of Actual and Predicted Values of AD

City	AD Actual	AD Predicted	Error
Dodge City	2170	1998.6	-171.4
Garden City	2300	1955.4	-344.6
Goodland	1000	1897.5	897.5
Great Bend	2250	2060.0	-190.0
Hays	2000	1969.3	- 30.7
Hutchinson	1400	1954.9	554.9
Lawrence	2570	2768.9	198.9
Liberal	2040	1810.6	-229.4
Manhattan	6180	6048.0	-132.0
Parsons	2000	2123.1	123.1
Salina	1700	1241.3	-458.7

Note: A positive error denotes overestimation by the model; while a negative value represents underestimation.

CONCLUSIONS

The model is able to demonstrate the structure of demand, rather than the values of the components. While a model that could be used to accurately describe demand was desired, the limitations of the database precluded arriving at such. The variables population, number of college graduates, and the number of persons employed in professional occupations were shown to influence demand, as predicted in the hypothesis. Other likely measures, such as distance and income. were not found to be significant in this analysis, but it is felt that further studies may find them to be significant. It is worth noting that these variables have been found to describe demand in other studies (Kanafani. 1983; Liew & Liew. 1981; Meyer & Clinton, 1984; Thorsen & Brewer. 1978).

The model derived cannot be used to describe the demand accurately. as seen by the comparisons in Chapter 3. This is due to several factors. Probably most significant among them is the small sample size. which has been discussed. If more data observations had been available and a larger area encompassed in the study. there is every reason to believe that a more accurate model would have been produced. As it is constructed herein, it is limited in both time and space, and is not likely to remain conditioned for very long. A larger database would hopefully overcome this limitation.

It would be important to calibrate the data in actual applications, although that was not accomplished in the study. Since 1980 censal data was being used with 1983 passenger movements, a method to convert from the current year to the base year should have been used. Such a transformation is commonly used in economic analysis to convert current prices to baseyear prices for the purpose of studying price increases and purchasing power. Such techniques are widely published, and are not within the scope of this report.

Although an accurate model could not be obtained, it shows nonetheless that such is probably obtainable. The ability to fit such scant data to a model suggests that indirect data sources might be as capable of describing demand as survey methods. None of the models noted above perform much better than the one arrived at in this study, and cost much more to complete. The advantages of using secondary sources such as the census data lies in the ease of data acquisition and the simplicity of the study scheme. It is conceivable that a planner could perform the entire study described within a short period of time. The survey methods, on the other hand, require a long period just to collect needed data. It might also be argued that the quality of data present in the Census database is superior to that which would be obtained in a survey.

The inconsistencies noted in the model may relate not just to the scant data used to arrive at its derivation, but also from external factors. The changing character of the market under deregulation is blamed by some writers for the difficulty encountered in describing travel behavior (Liew & Liew. 1980; Meyer & Oster. 1984). Because the use of secondary data sources does not take into account pricing and the market sensitivity to cost, some irregularities may result no matter how well the model is constructed. Several writers have reported that the demand for commuter air travel is very sensitive to cost, both economic and temporal (Kanafani. 1983; Meyer & Oster. 1984; Thorsen & Brewer. 1980). Such data is available for those communities already served by commuter air service. However. it would be difficult to contrive such information for communities that are contemplating establishing service.

The difficulty in modelling intercity demand in general may also contribute to the inconsistency of the model (Kanafani. 1983; Chan, 1982). There has been a considerable amount of research performed on intracity transportation models, while intercity movements have not received such attention (Weiner. 1983). The result is that intercity demand forecasting has not progressed at the same rate as intracity methods. A considerable amount of study still needs to be conducted in this area. Perhaps from it we will obtain a better understanding of the factors influencing intercity travel demand.

This study has shown the feasibility of using indirect data sources to describe air travel demand in small communities. Although the data and results obtained are localized in time and place, it nevertheless represents the potential for constructing such models. The traditional models have all used survey data either exclusively or in part, have been costly to construct, and have varying levels of accuracy. By contrast, the model described herein is easy to apply, requiring readily available data and simple statistical techniques.

BIBLIOGRAPHY

Belsley. D., Kuh. E., and R. Welsch. Regression Diagnostics. New York: John Wiley & Sons, 1980.

Boyer. J., PhD. Associate Professor of Statistics, Kansas State University. Personal Communication, 1985.

Chan, Y. "Airline Deregulation and Service to Small Communities." In Transportation Research Record 851, 1982.

Draper. E., and R. Smith. Applied Regression Analysis, Second Edition. New York: McGraw-Hill. 1980.

Green, P. Analyzing Multivariate Data. Hinsdale. IL: Dryden Press, 1978.

Hocking, R. "The Analysis and Selection of Variables in a Linear Regression." Biometrics 32:1-49. March 76.

Kanafani. A. Transportation Demand Analysis. New York: McGraw-Hill 1983.

Kleinbaum. D., and L. Kupper. Applied Regression Analysis and Other Multivariable Methods. Belmont, CA: Duxbury Press, 1978.

Liew. C., and C. Liew. "Airline Deregulation and Its Impact on Intercity Travel Demand." In Transportation Research Record 855. 1978.

Meyer. J., and C. Oster. Jr. Deregulation and the New Airline Entrepreneurs. Cambridge: MIT Press, 1984.

Meyer. M., and E. Miller. Urban Transportation Planning: A Decision-Oriented Approach. New York: McGraw-Hill. 1984.

Milliken, G, PhD. Professor of Statistics, Kansas State University. Personal Communication, 1985.

Paquette. R., and N. Ashford. Transportation Engineering, Second Edition. New York: John Wiley & Sons, 1982.

SAS Institute. SAS Users Guide: Statistics, 1982 Edition. Cary. NC: SAS Institute. 1982.

Searle. S. Matrix Algebra Useful for Statistics. New York: John Wiley & Sons, 1982.

Thorsen, B., and K. Brewer. "Model to Estimate Commuter Airline Demand in Small Cities." In Transportation Research Record 791, 1978.

U.S. Department of Commerce. Census of Population: Characteristics of the Population. Vol. 1. Ch. B, Part 18, Kansas. Washington, DC: GPO, 1982.

Wasserstein, R. PhD Candidate in Statistics, Kansas State University. and Assistant Professor of Statistics, Washburn University. Personal Communication, 1985.

Weiner. E. Urban Transportation Planning in the United States: An Historical Overview. US DOT Report DOT-I-83-43. August, 1983.

Willemain, T. Statistical Methods for Planners. Cambridge: MIT Press, 1980.

Appendix A

LISTING OF DATA OBSERVATIONS

```
1       S A S   L O G   CMS SAS 82.3   VM/CMS  CMS USER VMMQE

NOTE: CMSSAS RELEASE 82.3 AT KANSAS STATE UNIVERSITY

NOTE: NO OPTIONS SPECIFIED.

   1 cms FILEDEF TDATA DISK THESIS DATA C1 ;
   1
   2 options linesize=64 pagesize=58 nodate nonumber;
   3 data one;
   4   input city $ county $ ep ad distance pop college under18
   5     over65 #2 prof tech serv agri labor #3 houses phouses hu
nits hvalue
   6     hincome fincome age f15 h15 spgen;
   7   infile tdata;


NOTE: INFILE TDATA IS FILE THESIS DATA C1
NOTE: 36 LINES WERE READ FROM INFILE TDATA.
NOTE: DATA SET WORK.ONE HAS 12 OBSERVATIONS AND 24 VARIABLES.
NOTE: THE DATA STATEMENT USED 0.20 SECONDS AND 536K.

   8 proc print;

NOTE: THE PROCEDURE PRINT USED 0.22 SECONDS AND 600K
      AND PRINTED PAGE 1.

NOTE: SAS INSTITUTE INC.
      SAS CIRCLE
      PO BOX 8000
      CARY, N.C.  27511-8000
```

SAS

OBS	CITY	COUNTY	EP	AD	DISTANCE	POP	COLLEGE
1	DODGE	FORD	5224	2170	155	24315	4109
2	GARDEN	FINNEY	9058	2300	208	23825	3788
3	GOODLAND	SHERMAN	2147	1000	191	7759	954
4	GREATB	BARTON	4404	2250	115	31343	3886
5	HAYS	ELLIS	7367	2000	175	26098	5454
6	HUTCH	RENO	1574	1400	52	64983	8578
7	LAWRENCE	DOUGLAS	4222	2570	37	67640	23742
8	LIBERAL	SEWARD	30464	2040	214	17071	2270
9	MANHATTA	RILEY	45038	6180	114	63505	20068
10	PARSONS	LABETTE	4000	2000	144	25682	2620
11	SALINA	SALINE	22300	1700	90	48905	7629
12	TOPEKA	SHAWNEE	44465	9900	61	154916	30828

OBS	UNDER18	OVER65	PROF	TECH	SERV	AGRI	LABOR	HOUSES	PHOUSES
1	7056	2992	2305	3521	1586	913	3333	8776	2.68
2	7980	1951	2323	2980	1519	926	3730	8104	2.89
3	2253	1030	670	971	588	547	725	2861	2.65
4	8678	4282	2585	4052	2029	851	5485	11797	2.61
5	6888	2647	2762	4033	2031	867	3617	9200	2.63
6	17619	8922	5625	8249	3904	1525	10995	24448	2.58
7	14058	5056	9262	9854	4863	844	7641	23817	2.46
8	5345	1475	1546	2228	870	386	3185	6125	2.75
9	13744	3487	6384	7498	3781	1009	3803	19269	2.59
10	7342	4361	1927	2514	1529	671	3982	9702	2.57
11	13422	5949	4903	7175	3362	668	7892	19613	2.57
12	42451	18529	18557	27234	10212	1029	18939	58832	2.55

OBS	HUNITS	HVALUE	HINCOME	FINCOME	AGE	F15	H15	SPGEN
1	9802	38400	16505	19858	28.7	4458	4903	1224
2	8938	40600	17511	19715	26.2	4168	4729	890
3	3240	33400	14221	16908	30.7	1250	1358	214
4	12804	37000	16292	19333	30.8	5740	6363	1008
5	10231	42000	15623	19527	25.6	4375	4842	3629
6	26519	35300	15949	19091	30.6	11802	13110	2761
7	25479	46900	14156	19672	24.3	9516	11322	20518
8	6690	36700	17630	20768	28.7	3267	3673	528
9	20765	46500	12359	15645	22.6	6557	7879	28165
10	10595	24700	13330	16925	32.3	3968	4354	1018
11	20302	35900	16201	19473	29.6	9016	10082	2577
12	64393	43700	17713	21371	30.3	29650	34459	8454

Appendix B

LISTING FROM SAS PROCEDURE RSQUARE

```
NOTE: CMSSAS RELEASE 82.3 AT KANSAS STATE UNIVERSITY

NOTE: NO OPTIONS SPECIFIED.

  1 options linesize=64 pagesize=58 nodate nonumber;
  2 cms FILEDEF TDATA DISK THESIS DATA C1 ;
  2
  3 title1 Data Region = County;
  4 title2 All Cities included in analysis;
  5 data one;
  6   input city $ county $ ep ad distance pop college under18
  7     over65 #2 prof tech serv agri labor #3 houses phouses hu
nits hvalue
  8     hincome fincome age f15 h15 spgen;
  9   infile tdata;


NOTE: INFILE TDATA IS FILE THESIS DATA C1
NOTE: 36 LINES WERE READ FROM INFILE TDATA.
NOTE: DATA SET WORK.ONE HAS 12 OBSERVATIONS AND 24 VARIABLES.
NOTE: THE DATA STATEMENT USED 0.20 SECONDS AND 536K.

 10 proc rsquare cp;
 11   model ad ep = distance pop college prof tech serv  houses
/stop=3;

NOTE: THE PROCEDURE RSQUARE USED 0.47 SECONDS AND 536K
      AND PRINTED PAGES 1 TO 4.

 12 data two;
 13   set one;
 14   if city='TOPEKA' then delete;
 15 title2 Data with Topeka removed;

NOTE: DATA SET WORK.TWO HAS 11 OBSERVATIONS AND 24 VARIABLES.
NOTE: THE DATA STATEMENT USED 0.14 SECONDS AND 536K.

 16 proc rsquare cp;
 17   model ad ep = distance pop college prof tech serv  houses
/stop=3;

NOTE: THE PROCEDURE RSQUARE USED 0.49 SECONDS AND 536K
      AND PRINTED PAGES 5 TO 8.
NOTE: SAS USED 536K MEMORY.

NOTE: SAS INSTITUTE INC.
      SAS CIRCLE
      PO BOX 8000
      CARY, N.C.  27511-8000
```

Data Region = County
All Cities included in analysis

N= 12 REGRESSION MODELS FOR DEPENDENT VARIABLE AD

NUMBER IN MODEL	R-SQUARE	C(P)	VARIABLES IN MODEL
1	0.13843947	82.11597700	DISTANCE
1	0.68433451	25.01741774	COLLEGE
1	0.69385861	24.02123324	HOUSES
1	0.71296840	22.02242162	SERV
1	0.73454684	19.76539792	PROF
1	0.74174602	19.01239079	POP
1	0.74782405	18.37665143	TECH
2	0.71543352	23.76457859	SERV HOUSES
2	0.73117728	22.11784165	COLLEGE SERV
2	0.73635974	21.57577526	PROF HOUSES
2	0.73661159	21.54943255	COLLEGE HOUSES
2	0.73948306	21.24908827	COLLEGE PROF
2	0.73981659	21.21420225	PROF SERV
2	0.74365870	20.81233210	POP PROF
2	0.74823279	20.33389969	PROF TECH
2	0.74997806	20.15135068	POP TECH
2	0.75373664	19.75821765	POP COLLEGE
2	0.75619525	19.50105590	TECH SERV
2	0.76063285	19.03690021	POP SERV
2	0.76889068	18.17316235	COLLEGE TECH
2	0.78296803	16.70072390	DISTANCE COLLEGE
2	0.78654648	16.32643174	TECH HOUSES
2	0.80317696	14.58694627	POP HOUSES
2	0.80406696	14.49385571	DISTANCE HOUSES
2	0.81313437	13.54543903	DISTANCE TECH
2	0.83701849	11.04725009	DISTANCE PROF
2	0.84442740	10.27230653	DISTANCE SERV
2	0.87741494	6.82193450	DISTANCE POP
3	0.73959143	23.23775330	COLLEGE PROF HOUSES
3	0.74118717	23.07084501	PROF SERV HOUSES
3	0.74174273	23.01273474	COLLEGE PROF SERV
3	0.74208892	22.97652453	COLLEGE SERV HOUSES
3	0.74998324	22.15080901	POP PROF TECH
3	0.75457858	21.67015363	POP COLLEGE PROF
3	0.77059688	19.99470025	POP COLLEGE TECH
3	0.77376354	19.66347918	PROF TECH SERV
3	0.78641815	18.33985517	POP COLLEGE SERV
3	0.78787523	18.18744971	TECH SERV HOUSES
3	0.78812683	18.16113377	PROF TECH HOUSES
3	0.79647773	17.28766063	POP TECH SERV
3	0.80412333	16.48796020	POP PROF HOUSES
3	0.80838628	16.04207110	POP SERV HOUSES
3	0.81036387	15.83522291	COLLEGE TECH HOUSES
3	0.81313449	15.54542608	DISTANCE TECH HOUSES
3	0.81559762	15.28779226	POP PROF SERV
3	0.81843507	14.99100633	POP COLLEGE HOUSES

Data Region = County
All Cities included in analysis

N= 12 REGRESSION MODELS FOR DEPENDENT VARIABLE AD

NUMBER IN MODEL	R-SQUARE	C(P)	VARIABLES IN MODEL
3	0.83712073	13.03655611	DISTANCE PROF TECH
3	0.83860516	12.88129042	DISTANCE PROF HOUSES
3	0.84512853	12.19897118	DISTANCE PROF SERV
3	0.85016254	11.67243270	DISTANCE SERV HOUSES
3	0.85306149	11.36921286	COLLEGE TECH SERV
3	0.85924791	10.72213749	DISTANCE TECH SERV
3	0.85951321	10.69438736	DISTANCE COLLEGE PROF
3	0.87750673	8.81233376	DISTANCE POP PROF
3	0.88292422	8.24568418	DISTANCE COLLEGE SERV
3	0.88387549	8.14618492	DISTANCE COLLEGE TECH
3	0.88411214	8.12143285	DISTANCE POP SERV
3	0.88694945	7.82466080	COLLEGE PROF TECH
3	0.89281897	7.21073055	POP TECH HOUSES
3	0.89434105	7.05152725	DISTANCE COLLEGE HOUSES
3	0.90040919	6.41682239	DISTANCE POP TECH
3	0.91166069	5.23995785	DISTANCE POP COLLEGE
3	0.93974733	2.30220168	DISTANCE POP HOUSES

Data Region = County
All Cities included in analysis

N= 12 REGRESSION MODELS FOR DEPENDENT VARIABLE EP

NUMBER IN MODEL	R-SQUARE	C(P)	VARIABLES IN MODEL
1	0.01362635	134.17159804	DISTANCE
1	0.30426819	92.27975080	HOUSES
1	0.30549577	92.10281260	SERV
1	0.32357912	89.49635643	PROF
1	0.33268574	88.18376840	TECH
1	0.34401508	86.55080746	POP
1	0.34879218	85.86225717	COLLEGE
2	0.30599842	94.03036254	SERV HOUSES
2	0.32479075	91.32171775	PROF HOUSES
2	0.33270366	90.18118613	PROF TECH
2	0.33803200	89.41318165	PROF SERV
2	0.34424213	88.51808046	POP TECH
2	0.34790674	87.98987992	POP PROF
2	0.34969698	87.73184323	COLLEGE SERV
2	0.35081574	87.57058997	COLLEGE PROF
2	0.35179595	87.42930652	TECH SERV
2	0.35448594	87.04158314	COLLEGE HOUSES
2	0.35918259	86.36462747	TECH HOUSES
2	0.36338729	85.75858152	POP COLLEGE
2	0.36407031	85.66013417	COLLEGE TECH
2	0.45113928	73.11039126	DISTANCE TECH
2	0.45254593	72.90764371	POP SERV
2	0.45263125	72.89534601	POP HOUSES
2	0.47017557	70.36658445	DISTANCE HOUSES
2	0.47957559	69.01170546	DISTANCE PROF
2	0.48503869	68.22427912	DISTANCE SERV
2	0.53996807	60.30699652	DISTANCE COLLEGE
2	0.54614609	59.41652462	DISTANCE POP
3	0.34826509	89.93822943	POP PROF TECH
3	0.35398190	89.11423224	COLLEGE PROF SERV
3	0.35485060	88.98902224	PROF SERV HOUSES
3	0.35932245	88.34446927	PROF TECH HOUSES
3	0.36157450	88.01986879	TECH SERV HOUSES
3	0.36418645	87.64339354	POP COLLEGE TECH
3	0.36452626	87.59441505	COLLEGE PROF HOUSES
3	0.37373369	86.26729605	PROF TECH SERV
3	0.39333927	83.44143335	COLLEGE TECH HOUSES
3	0.39786624	82.78893581	POP COLLEGE PROF
3	0.45899887	73.97754420	POP PROF HOUSES
3	0.46208295	73.53301826	COLLEGE SERV HOUSES
3	0.47705015	71.37571164	DISTANCE TECH HOUSES
3	0.48112413	70.78850540	DISTANCE PROF TECH
3	0.48297435	70.52182363	POP COLLEGE HOUSES
3	0.48412793	70.35555144	DISTANCE PROF HOUSES
3	0.48510037	70.21538867	DISTANCE SERV HOUSES
3	0.48539617	70.17275293	DISTANCE PROF SERV

Data Region = County
All Cities included in analysis

N= 12 REGRESSION MODELS FOR DEPENDENT VARIABLE EP

NUMBER IN MODEL	R-SQUARE	C(P)	VARIABLES IN MODEL
3	0.48564919	70.13628328	POP TECH SERV
3	0.50481965	67.37313713	POP TECH HOUSES
3	0.51439879	65.99244199	POP SERV HOUSES
3	0.51762214	65.52784158	POP COLLEGE SERV
3	0.51859290	65.38792119	COLLEGE TECH SERV
3	0.51997931	65.18809041	DISTANCE TECH SERV
3	0.54247544	61.94559631	POP PROF SERV
3	0.55015903	60.83811716	DISTANCE COLLEGE PROF
3	0.56270645	59.02958565	DISTANCE POP PROF
3	0.56486853	58.71795304	DISTANCE COLLEGE SERV
3	0.56558889	58.61412418	DISTANCE COLLEGE TECH
3	0.58445345	55.89506854	DISTANCE COLLEGE HOUSES
3	0.59974532	53.69096517	DISTANCE POP COLLEGE
3	0.61559098	51.40703999	DISTANCE POP SERV
3	0.61884576	50.93791009	COLLEGE PROF TECH
3	0.65622513	45.55021121	DISTANCE POP HOUSES
3	0.66155279	44.78230498	DISTANCE POP TECH

Data Region = County
Data with Topeka removed

N= 11 REGRESSION MODELS FOR DEPENDENT VARIABLE AD

NUMBER IN MODEL	R-SQUARE	C(P)	VARIABLES IN MODEL
1	0.02046532	37.26930726	DISTANCE
1	0.09372829	33.95824429	HOUSES
1	0.13065303	32.28945957	TECH
1	0.15391707	31.23805912	SERV
1	0.18681557	29.75123699	PROF
1	0.21154894	28.63343185	POP
1	0.40951605	19.68646247	COLLEGE
2	0.18352891	31.89977522	TECH HOUSES
2	0.20692632	30.84234751	PROF SERV
2	0.21170386	30.62643039	POP PROF
2	0.22457263	30.04483625	TECH SERV
2	0.24383326	29.17436708	PROF TECH
2	0.24394929	29.16912296	PROF HOUSES
2	0.24571866	29.08915769	DISTANCE HOUSES
2	0.27087235	27.95235657	SERV HOUSES
2	0.30477789	26.42002207	DISTANCE TECH
2	0.31330006	26.03486904	POP SERV
2	0.36777478	23.57292647	DISTANCE PROF
2	0.39824267	22.19595420	POP TECH
2	0.40877792	21.71982141	DISTANCE SERV
2	0.43875854	20.36487061	POP COLLEGE
2	0.50128377	17.53909151	COLLEGE HOUSES
2	0.53983981	15.79658173	COLLEGE SERV
2	0.54467546	15.57803814	COLLEGE TECH
2	0.55441984	15.13764821	DISTANCE POP
2	0.58519244	13.74690464	DISTANCE COLLEGE
2	0.68755957	9.12050214	COLLEGE PROF
2	0.85588383	1.51321811	POP HOUSES
3	0.24710930	31.02630904	PROF TECH HOUSES
3	0.27046560	29.97073924	PROF TECH SERV
3	0.27104697	29.94446472	PROF SERV HOUSES
3	0.29185898	29.00388222	TECH SERV HOUSES
3	0.30482887	28.41771777	DISTANCE TECH HOUSES
3	0.36798719	25.56332674	DISTANCE PROF TECH
3	0.37776924	25.12123471	DISTANCE PROF HOUSES
3	0.40733447	23.78505706	POP TECH SERV
3	0.40947837	23.68816545	DISTANCE PROF SERV
3	0.41921556	23.24810078	DISTANCE SERV HOUSES
3	0.44691752	21.99613209	POP PROF SERV
3	0.49141407	19.98514526	DISTANCE TECH SERV
3	0.52012845	18.68742096	POP PROF TECH
3	0.54468812	17.57746573	COLLEGE TECH SERV
3	0.55450725	17.13369779	DISTANCE POP PROF
3	0.56556566	16.63392157	DISTANCE POP SERV
3	0.58406822	15.79771282	COLLEGE TECH HOUSES
3	0.58520631	15.74627804	DISTANCE COLLEGE SERV

Data Region = County
Data with Topeka removed

N= 11 REGRESSION MODELS FOR DEPENDENT VARIABLE AD

NUMBER IN MODEL	R-SQUARE	C(P)	VARIABLES IN MODEL
3	0.58680532	15.67401199	DISTANCE COLLEGE TECH
3	0.58698539	15.66587377	COLLEGE SERV HOUSES
3	0.59917028	15.11518691	DISTANCE COLLEGE HOUSES
3	0.62606451	13.89972348	DISTANCE POP TECH
3	0.65725744	12.48998336	DISTANCE POP COLLEGE
3	0.68816841	11.09298579	DISTANCE COLLEGE PROF
3	0.72506321	9.42555428	POP COLLEGE TECH
3	0.75175693	8.21915218	POP COLLEGE SERV
3	0.83055267	4.65804010	COLLEGE PROF SERV
3	0.85164833	3.70463829	COLLEGE PROF TECH
3	0.85971675	3.33999213	DISTANCE POP HOUSES
3	0.86514955	3.09446119	POP TECH HOUSES
3	0.86916943	2.91278574	POP SERV HOUSES
3	0.86992614	2.87858685	POP COLLEGE HOUSES
3	0.88079413	2.38741667	POP PROF HOUSES
3	0.88125083	2.36677613	COLLEGE PROF HOUSES
3	0.90093530	1.47715190	POP COLLEGE PROF

Data Region = County
Data with Topeka removed

N= 11 REGRESSION MODELS FOR DEPENDENT VARIABLE EP

NUMBER IN MODEL	R-SQUARE	C(P)	VARIABLES IN MODEL
1	0.01165439	89.93422357	DISTANCE
1	0.01797151	89.31465817	HOUSES
1	0.02339807	88.78243618	SERV
1	0.02953209	88.18082793	TECH
1	0.03537394	87.60787586	PROF
1	0.05515158	85.66813880	POP
1	0.11356578	79.93903425	COLLEGE
2	0.02795132	90.33586587	SERV HOUSES
2	0.03729481	89.41948229	PROF TECH
2	0.04574652	88.59056139	PROF HOUSES
2	0.05652514	87.53342390	PROF SERV
2	0.05674500	87.51186019	TECH HOUSES
2	0.05719620	87.46760776	TECH SERV
2	0.06705747	86.50044186	POP PROF
2	0.12445075	80.87146691	POP COLLEGE
2	0.13497618	79.83916094	POP TECH
2	0.15463130	77.91144147	COLLEGE HOUSES
2	0.16634302	76.76278780	COLLEGE TECH
2	0.20552717	72.91971366	COLLEGE SERV
2	0.22867791	70.64915311	POP SERV
2	0.27238533	66.36244915	COLLEGE PROF
2	0.29343763	64.29769767	DISTANCE PROF
2	0.33833263	59.89451919	POP HOUSES
2	0.33969781	59.76062612	DISTANCE COLLEGE
2	0.36607743	57.17338528	DISTANCE SERV
2	0.39135476	54.69425436	DISTANCE TECH
2	0.42619806	51.27691866	DISTANCE HOUSES
2	0.54051270	40.06525293	DISTANCE POP
3	0.05674651	89.51171270	PROF TECH HOUSES
3	0.06073809	89.12022893	PROF SERV HOUSES
3	0.07976454	87.25416776	TECH SERV HOUSES
3	0.09472946	85.78644880	PROF TECH SERV
3	0.14420010	80.93450466	POP PROF TECH
3	0.17167022	78.24031046	COLLEGE TECH HOUSES
3	0.24023345	71.51581701	POP COLLEGE TECH
3	0.24824156	70.73040340	POP TECH SERV
3	0.31204922	64.47232340	COLLEGE PROF SERV
3	0.32058054	63.63559531	COLLEGE TECH SERV
3	0.33876870	61.85175036	POP PROF SERV
3	0.33988710	61.74206063	DISTANCE COLLEGE PROF
3	0.34187055	61.54752990	POP TECH HOUSES
3	0.36154731	59.61768650	POP COLLEGE HOUSES
3	0.36996918	58.79169309	DISTANCE PROF SERV
3	0.37597499	58.20266017	COLLEGE SERV HOUSES
3	0.38123605	57.68666931	POP PROF HOUSES
3	0.38664475	57.15619897	DISTANCE COLLEGE SERV

Data Region = County
Data with Topeka removed

N= 11 REGRESSION MODELS FOR DEPENDENT VARIABLE EP

NUMBER IN MODEL	R-SQUARE	C(P)	VARIABLES IN MODEL
3	0.39141451	56.68839393	DISTANCE TECH SERV
3	0.39331160	56.50233296	DISTANCE PROF TECH
3	0.42035925	53.84957335	DISTANCE COLLEGE TECH
3	0.42170119	53.71795947	POP SERV HOUSES
3	0.42676774	53.22104632	POP COLLEGE PROF
3	0.43026867	52.87768430	COLLEGE PROF HOUSES
3	0.43479581	52.43367502	DISTANCE SERV HOUSES
3	0.43635327	52.28092387	DISTANCE TECH HOUSES
3	0.44208488	51.71878355	DISTANCE PROF HOUSES
3	0.45142068	50.80315385	POP COLLEGE SERV
3	0.47913834	48.08468145	COLLEGE PROF TECH
3	0.51173098	44.88808508	DISTANCE COLLEGE HOUSES
3	0.54461220	41.66318507	DISTANCE POP COLLEGE
3	0.54622210	41.50529054	DISTANCE POP TECH
3	0.54744276	41.38557113	DISTANCE POP PROF
3	0.54792147	41.33862096	DISTANCE POP HOUSES
3	0.56794213	39.37504969	DISTANCE POP SERV

Appendix C

LISTING FROM SAS PROCEDURE REG

```
1        S A S   L O G   CMS SAS 82.3   VM/CMS  CMS USER VMMQE

NOTE: CMSSAS RELEASE 82.3 AT KANSAS STATE UNIVERSITY

NOTE: NO OPTIONS SPECIFIED.

   1 cms FILEDEF TDATA DISK THESIS DATA C1 ;
   1
   2 title1 Data Region = County;
   3 options linesize=64 pagesize=58 nodate nonumber;
   4 data one;
   5   input city $ county $ ep ad distance pop college under18
   6     over65 #2 prof tech serv agri labor #3 houses phouses hu
nits hvalue
   7     hincome fincome age f15 h15 spgen;
   8   infile tdata;


NOTE: INFILE TDATA IS FILE THESIS DATA C1
NOTE: 36 LINES WERE READ FROM INFILE TDATA.
NOTE: DATA SET WORK.ONE HAS 12 OBSERVATIONS AND 24 VARIABLES.
NOTE: THE DATA STATEMENT USED 0.20 SECONDS AND 536K.

   9 proc reg ;
  10   model ad = college prof pop /influence collin xpx i;

NOTE: THE PROCEDURE REG USED 0.28 SECONDS AND 536K
      AND PRINTED PAGES 1 TO 3.

  11 data two;
  12   set one;
  13   if city='TOPEKA' then delete;
  14 title2 Topeka Removed from dataset;

NOTE: DATA SET WORK.TWO HAS 11 OBSERVATIONS AND 24 VARIABLES.
NOTE: THE DATA STATEMENT USED 0.14 SECONDS AND 536K.

  15 proc reg;
  16   model ad = college prof pop /influence collin xpx i;

NOTE: THE PROCEDURE REG USED 0.28 SECONDS AND 536K
      AND PRINTED PAGES 4 TO 6.
NOTE: SAS USED 536K MEMORY.

NOTE: SAS INSTITUTE INC.
      SAS CIRCLE
      PO BOX 8000
      CARY, N.C.  27511-8000
```

MODEL CROSSPRODUCTS X'X X'Y Y'Y

X'X	INTERCEP	COLLEGE	PROF
INTERCEP	12	113926	58849
COLLEGE	113926	2137563850	1058321316
PROF	58849	1058321316	558154951
POP	556042	9154335008	4857559793
AD	35510	563318110	292543700

X'X	POP	AD
INTERCEP	556042	35510
COLLEGE	9154335008	563318110
PROF	4857559793	292543700
POP	43055036344	2598305440
AD	2598305440	175880300

X'X INVERSE, B, SSE

INVERSE	INTERCEP	COLLEGE	PROF
INTERCEP	0.3237831	-.0000140359	0.0001597551
COLLEGE	-.0000140359	8.49254E-09	-2.68093E-08
PROF	0.0001597551	-2.68093E-08	2.27836E-07
POP	-.0000192212	1.40027E-09	-2.20680E-08
AD	383.7719	0.0810299	-0.116539

INVERSE	POP	AD
INTERCEP	-.0000192212	383.7719
COLLEGE	1.40027E-09	0.0810299
PROF	-2.20680E-08	-0.116539
POP	2.46349E-09	0.05131179
AD	0.05131179	17375908

Data Region = County

DEP VARIABLE: AD

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	3	53424384	17808128	8.199	0.0080
ERROR	8	17375908	2171989		
C TOTAL	11	70800292			

ROOT MSE	1473.767	R-SQUARE	0.7546
DEP MEAN	2959.167	ADJ R-SQ	0.6625
C.V.	49.80344		

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > ^T^
INTERCEP	1	383.772	838.602	0.458	0.6594
COLLEGE	1	0.081030	0.135815	0.597	0.5673
PROF	1	-0.116539	0.703461	-0.166	0.8725
POP	1	0.051312	0.073148	0.701	0.5029

COLLINEARITY DIAGNOSTICS VARIANCE PROPORTIONS

NUMBER	EIGENVALUE	CONDITION INDEX	PORTION INTERCEP	PORTION COLLEGE	PORTION PROF
1	3.572	1.000	0.0142	0.0040	0.0006
2	0.375544	3.084	0.5423	0.0136	0.0019
3	0.047731	8.651	0.0274	0.7210	0.0167
4	0.004322	28.750	0.4161	0.2613	0.9808

NUMBER	PORTION POP
1	0.0007
2	0.0006
3	0.0531
4	0.9456

OBS	RESIDUAL	RSTUDENT	HAT DIAG H	COV RATIO	DFFITS	DFBETAS INTERCEP
1	474.252	0.3231	0.1189	1.8248	0.1186	0.1010
2	657.503	0.4544	0.1317	1.7490	0.1770	0.1575
3	218.878	0.1584	0.2281	2.1787	0.0861	0.0850
4	244.333	0.1669	0.1330	1.9367	0.0654	0.0251
5	157.036	0.1068	0.1280	1.9436	0.0409	0.0376
6	-2357.71	-3.1874	0.4597	0.0874	-2.9398	1.1205
7	-2128.93	-5.3606	0.6756	0.0077	-7.7364	-2.8132
8	776.515	0.5491	0.1595	1.7147	0.2392	0.2218
9	1655.55	5.5154	0.8060	0.0108	11.2408	-3.4911
10	310.71	0.2135	0.1413	1.9357	0.0866	0.0410

Data Region = County

OBS	RESIDUAL	RSTUDENT	HAT DIAG H	COV RATIO	DFFITS	DFBETAS INTERCEP
11	-1239.96	-0.8822	0.1157	1.2654	-0.3191	-0.0989
12	1231.83	7.7649	0.9025	0.0020	23.6285	-5.3328

OBS	DFBETAS COLLEGE	DFBETAS PROF	DFBETAS POP
1	-0.0257	0.0275	-0.0307
2	-0.0571	0.0630	-0.0637
3	-0.0255	0.0437	-0.0492
4	-0.0125	-0.0177	0.0231
5	-0.0065	0.0160	-0.0190
6	0.0468	2.0229	-2.5111
7	-3.4726	-2.7762	4.5016
8	-0.0799	0.0934	-0.0978
9	10.0769	-8.5299	6.1658
10	-0.0187	-0.0184	0.0241
11	0.1243	0.0136	-0.0707
12	-8.8533	7.4280	-1.9717

Data Region = County
Topeka Removed from dataset

MODEL CROSSPRODUCTS X'X X'Y Y'Y

X'X	INTERCEP	COLLEGE	PROF
INTERCEP	11	83098	40292
COLLEGE	83098	1187198266	486246120
PROF	40292	486246120	213792702
POP	401126	4378584560	1982783581
AD	25610	258120910	108829400

X'X	POP	AD
INTERCEP	401126	25610
COLLEGE	4378584560	258120910
PROF	1982783581	108829400
POP	19056069288	1064637040
AD	1064637040	77870300

X'X INVERSE, B, SSE

INVERSE	INTERCEP	COLLEGE	PROF
INTERCEP	0.4765038	.00002702615	-.0000186866
COLLEGE	.00002702615	1.95329E-08	-7.47869E-08
PROF	-.0000186866	-7.47869E-08	4.36331E-07
POP	-.0000142959	2.72454E-09	-2.78228E-08
AD	1925.725	0.495615	-1.91819

INVERSE	POP	AD
INTERCEP	-.0000142959	1925.725
COLLEGE	2.72454E-09	0.495615
PROF	-2.78228E-08	-1.91819
POP	2.62234E-09	0.1010404
AD	0.1010404	1807491

Data Region = County
Topeka Removed from dataset

DEP VARIABLE: AD

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	3	16438072	5479357	21.220	0.0007
ERROR	7	1807491	258213		
C TOTAL	10	18245564			

ROOT MSE	508.147	R-SQUARE	0.9009
DEP MEAN	2328.182	ADJ R-SQ	0.8585
C.V.	21.8259		

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > ^T^
INTERCEP	1	1925.725	350.770	5.490	0.0009
COLLEGE	1	0.495615	0.071019	6.979	0.0002
PROF	1	-1.918186	0.335658	-5.715	0.0007
POP	1	0.101040	0.026022	3.883	0.0060

COLLINEARITY DIAGNOSTICS VARIANCE PROPORTIONS

NUMBER	EIGENVALUE	CONDITION INDEX	PORTION INTERCEP	PORTION COLLEGE	PORTION PROF
1	3.657	1.000	0.0114	0.0029	0.0008
2	0.294315	3.525	0.4269	0.0431	0.0017
3	0.041850	9.348	0.5574	0.3791	0.0120
4	0.006939	22.957	0.0043	0.5749	0.9855

NUMBER	PORTION POP
1	0.0015
2	0.0000
3	0.2213
4	0.7772

OBS	RESIDUAL	RSTUDENT	HAT DIAG H	COV RATIO	DFFITS	DFBETAS INTERCEP
1	172.413	0.3390	0.1247	1.9619	0.1279	0.1033
2	545.542	1.1856	0.1325	0.9202	0.4633	0.3593
3	-897.33	-3.2938	0.3082	0.0431	-2.1983	-2.1733
4	189.914	0.3760	0.1332	1.9473	0.1474	0.0498
5	32.2662	0.0630	0.1290	2.1213	0.0242	0.0195
6	-553.227	-2.5054	0.6688	0.3191	-3.5603	2.0545
7	-190.755	-1.3851	0.9169	7.3495	-4.6013	0.1524
8	229.882	0.4706	0.1787	1.9513	0.2195	0.1993
9	137.398	1.3277	0.9540	14.3754	6.0469	-0.0743

Data Region = County
Topeka Removed from dataset

OBS	RESIDUAL	RSTUDENT	HAT DIAG H	COV RATIO	DFFITS	DFBETAS INTERCEP
10	-122.813	-0.2444	0.1534	2.1032	-0.1040	-0.0555
11	456.71	1.0889	0.3006	1.2877	0.7139	-0.2038

OBS	DFBETAS COLLEGE	DFBETAS PROF	DFBETAS POP
1	0.0030	0.0018	-0.0246
2	-0.0711	0.0939	-0.1524
3	-0.4730	0.0813	0.7726
4	-0.0144	-0.0327	0.0518
5	-0.0009	0.0054	-0.0103
6	1.5278	0.0913	-1.9536
7	0.6056	-2.6558	2.8085
8	0.0084	0.0088	-0.0645
9	5.0762	-4.6942	3.5411
10	-0.0077	0.0355	-0.0341
11	-0.5347	0.3734	-0.0427

Appendix D

LISTING FROM SAS PROCEDURE STEPWISE

```
1        S A S   L O G   CMS SAS 82.3   VM/CMS   CMS USER VMMQE

NOTE: CMSSAS RELEASE 82.3 AT KANSAS STATE UNIVERSITY

NOTE: NO OPTIONS SPECIFIED.

   1 options linesize=64 pagesize=58 nodate nonumber;
   2 cms FILEDEF TDATA DISK THESIS DATA C1 ;
   2
   3 data one;
   4   input city $ county $ ep ad distance pop college under18
   5     over65 #2 prof tech serv agri labor #3 houses phouses hu
nits hvalue
   6     hincome fincome age f15 h15 spgen;
   7   if city='TOPEKA' then delete;
   8   infile tdata;


NOTE: INFILE TDATA IS FILE THESIS DATA C1
NOTE: 36 LINES WERE READ FROM INFILE TDATA.
NOTE: DATA SET WORK.ONE HAS 11 OBSERVATIONS AND 24 VARIABLES.
NOTE: THE DATA STATEMENT USED 0.20 SECONDS AND 536K.

   9 proc stepwise;
  10   model ad ep = college prof pop distance tech serv houses/
  11   f b stepwise maxr minr stop=3;

NOTE: SLENTRY AND SLSTAY HAVE BEEN SET TO .15 FOR THE STEPWISE T
ECHNIQUE.
NOTE: SLENTRY AND SLSTAY HAVE BEEN SET TO .15 FOR THE STEPWISE T
ECHNIQUE.
NOTE: THE PROCEDURE STEPWISE USED 1.80 SECONDS AND 536K
      AND PRINTED PAGES 1 TO 27.
NOTE: SAS USED 536K MEMORY.

NOTE: SAS INSTITUTE INC.
      SAS CIRCLE
      PO BOX 8000
      CARY, N.C.  27511-8000
```

FORWARD SELECTION PROCEDURE FOR DEPENDENT VARIABLE AD

STEP 1 VARIABLE COLLEGE ENTERED R SQUARE = 0.40951605
C(P) = 19.68646247

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	7471851.11306	7471851.113	6.24	0.0340
ERROR	9	10773712.52331	1197079.169		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	1455.14412				
COLLEGE	0.11557	0.04625753	7471851.113	6.24	0.0340

STEP 2 VARIABLE PROF ENTERED R SQUARE = 0.68755957
C(P) = 9.12050214

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	12544911.81822	6272455.909	8.80	0.0095
ERROR	8	5700651.81814	712581.477		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	2476.55484				
COLLEGE	0.39064	0.10909470	9136356.373	12.82	0.0072
PROF	-0.84615	0.31712572	5073060.705	7.12	0.0284

STEP 3 VARIABLE POP ENTERED R SQUARE = 0.90093530
C(P) = 1.47715190

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	16438072.38230	5479357.461	21.22	0.0007
ERROR	7	1807491.25406	258213.036		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	1925.72537				
COLLEGE	0.49562	0.07101864	12575416.17	48.70	0.0002
PROF	-1.91819	0.33565807	8432675.44	32.66	0.0007
POP	0.10104	0.02602155	3893160.56	15.08	0.0060

SAS

FORWARD SELECTION PROCEDURE FOR DEPENDENT VARIABLE AD

STEP 4 VARIABLE DISTANCE ENTERED R SQUARE = 0.92576738
C(P) = 2.35488536

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	4	16891147.68522	4222786.921	18.71	0.0015
ERROR	6	1354415.95114	225735.992		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-278.002579				
COLLEGE	0.434265	0.07927516	6773850.294	30.01	0.0015
PROF	-1.671700	0.35883932	4899115.325	21.70	0.0035
POP	0.114781	0.02619214	4335127.117	19.20	0.0047
DISTANCE	9.294881	6.56083413	453075.303	2.01	0.2063

NO OTHER VARIABLES MET THE 0.5000 SIGNIFICANCE LEVEL FOR ENTRY

BACKWARD ELIMINATION PROCEDURE FOR DEPENDENT VARIABLE AD

STEP 0 ALL VARIABLES ENTERED R SQUARE = 0.93361983
C(P) = 8.00000000

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	7	17034420.06080	2433488.580	6.03	0.0840
ERROR	3	1211143.57556	403714.525		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-1042.02875				
COLLEGE	0.60869	0.4083438	897034.034	2.22	0.2328
PROF	-2.41383	1.4376472	1138111.702	2.82	0.1917
POP	0.01478	0.2353880	1591.737	0.00	0.9539
DISTANCE	13.00937	13.8682463	355257.776	0.88	0.4174
TECH	0.17432	0.8417296	17315.052	0.04	0.8492
SERV	0.32931	2.0024166	10918.646	0.03	0.8798
HOUSES	0.28208	0.8262022	47058.186	0.12	0.7553

STEP 1 VARIABLE POP REMOVED R SQUARE = 0.93353259
C(P) = 6.00394273

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	6	17032828.32383	2838804.721	9.36	0.0242
ERROR	4	1212735.31253	303183.828		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-1156.03713				
COLLEGE	0.63273	0.12298820	8024409.796	26.47	0.0068
PROF	-2.49022	0.66397444	4264597.956	14.07	0.0199
DISTANCE	13.59862	8.84874607	716031.377	2.36	0.1992
TECH	0.16728	0.72294332	16233.042	0.05	0.8284
SERV	0.34982	1.71203437	12658.161	0.04	0.8481
HOUSES	0.33232	0.17830877	1053109.997	3.47	0.1358

BACKWARD ELIMINATION PROCEDURE FOR DEPENDENT VARIABLE AD

STEP 2 VARIABLE SERV REMOVED R SQUARE = 0.93283883
C(P) = 4.03529697

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	5	17020170.16332	3404034.033	13.89	0.0059
ERROR	5	1225393.47304	245078.695		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-974.243942				
COLLEGE	0.642722	0.10145586	9835528.137	40.13	0.0014
PROF	-2.479532	0.59511337	4254470.687	17.36	0.0088
DISTANCE	12.993423	7.49685287	736198.229	3.00	0.1436
TECH	0.254818	0.52357258	58051.175	0.24	0.6471
HOUSES	0.347037	0.14665270	1372391.630	5.60	0.0642

STEP 3 VARIABLE TECH REMOVED R SQUARE = 0.92965717
C(P) = 2.17908960

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	4	16962118.98793	4240529.747	19.82	0.0013
ERROR	6	1283444.64844	213907.441		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-1147.08699				
COLLEGE	0.63068	0.09192128	10069506.27	47.07	0.0005
PROF	-2.29770	0.43276332	6029919.43	28.19	0.0018
DISTANCE	13.84170	6.81196994	883200.83	4.13	0.0884
HOUSES	0.40154	0.08847286	4406098.42	20.60	0.0039

ALL VARIABLES IN THE MODEL ARE SIGNIFICANT AT THE 0.1000 LEVEL.

STEPWISE REGRESSION PROCEDURE FOR DEPENDENT VARIABLE AD

STEP 1 VARIABLE COLLEGE ENTERED R SQUARE = 0.40951605
C(P) = 19.68646247

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	7471851.11306	7471851.113	6.24	0.0340
ERROR	9	10773712.52331	1197079.169		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	1455.14412				
COLLEGE	0.11557	0.04625753	7471851.113	6.24	0.0340

STEP 2 VARIABLE PROF ENTERED R SQUARE = 0.68755957
C(P) = 9.12050214

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	12544911.81822	6272455.909	8.80	0.0095
ERROR	8	5700651.81814	712581.477		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	2476.55484				
COLLEGE	0.39064	0.10909470	9136356.373	12.82	0.0072
PROF	-0.84615	0.31712572	5073060.705	7.12	0.0284

STEP 3 VARIABLE POP ENTERED R SQUARE = 0.90093530
C(P) = 1.47715190

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	16438072.38230	5479357.461	21.22	0.0007
ERROR	7	1807491.25406	258213.036		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	1925.72537				
COLLEGE	0.49562	0.07101864	12575416.17	48.70	0.0002
PROF	-1.91819	0.33565807	8432675.44	32.66	0.0007
POP	0.10104	0.02602155	3893160.56	15.08	0.0060

NO OTHER VARIABLES MET THE 0.1500 SIGNIFICANCE LEVEL FOR ENTRY

MAXIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE AD

STEP 1 VARIABLE COLLEGE ENTERED R SQUARE = 0.40951605
C(P) = 19.68646247

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	7471851.11306	7471851.113	6.24	0.0340
ERROR	9	10773712.52331	1197079.169		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	1455.14412				
COLLEGE	0.11557	0.04625753	7471851.113	6.24	0.0340

THE ABOVE MODEL IS THE BEST 1 VARIABLE MODEL FOUND.

STEP 2 VARIABLE PROF ENTERED R SQUARE = 0.68755957
C(P) = 9.12050214

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	12544911.81822	6272455.909	8.80	0.0095
ERROR	8	5700651.81814	712581.477		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	2476.55484				
COLLEGE	0.39064	0.10909470	9136356.373	12.82	0.0072
PROF	-0.84615	0.31712572	5073060.705	7.12	0.0284

THE ABOVE MODEL IS THE BEST 2 VARIABLE MODEL FOUND.

STEP 3 VARIABLE POP ENTERED R SQUARE = 0.90093530
C(P) = 1.47715190

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	16438072.38230	5479357.461	21.22	0.0007
ERROR	7	1807491.25406	258213.036		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	1925.72537				
COLLEGE	0.49562	0.07101864	12575416.17	48.70	0.0002
PROF	-1.91819	0.33565807	8432675.44	32.66	0.0007
POP	0.10104	0.02602155	3893160.56	15.08	0.0060

THE ABOVE MODEL IS THE BEST 3 VARIABLE MODEL FOUND.

MINIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE AD

STEP 1 VARIABLE DISTANCE ENTERED R SQUARE = 0.02046532
C(P) = 37.26930726

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	373401.27705	373401.277	0.19	0.6748
ERROR	9	17872162.35931	1985795.818		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	2765.06055				
DISTANCE	-3.21449	7.41296235	373401.2771	0.19	0.6748

STEP 1 DISTANCE REPLACED BY HOUSES R SQUARE = 0.09372829
C(P) = 33.95824429

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	1710125.49232	1710125.492	0.93	0.3599
ERROR	9	16535438.14404	1837270.905		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	1599.47211				
HOUSES	0.05578	0.05781317	1710125.492	0.93	0.3599

STEP 1 HOUSES REPLACED BY TECH R SQUARE = 0.13065303
C(P) = 32.28945957

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	2383838.12192	2383838.122	1.35	0.2747
ERROR	9	15861725.51444	1762413.946		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	1510.22906				
TECH	0.16952	0.14576273	2383838.122	1.35	0.2747

MINIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE AD

STEP 1 TECH REPLACED BY SERV — R SQUARE = 0.15391707, C(P) = 31.23805912

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	2808303.75457	2808303.755	1.64	0.2327
ERROR	9	15437259.88179	1715251.098		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	1423.14487				
SERV	0.38199	0.29853350	2808303.755	1.64	0.2327

STEP 1 SERV REPLACED BY PROF — R SQUARE = 0.18681557, C(P) = 29.75123699

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	3408555.44504	3408555.445	2.07	0.1843
ERROR	9	14837008.19133	1648556.466		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	1497.06872				
PROF	0.22690	0.15779772	3408555.445	2.07	0.1843

STEP 1 PROF REPLACED BY POP — R SQUARE = 0.21154894, C(P) = 28.63343185

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	3859829.61561	3859829.616	2.41	0.1546
ERROR	9	14385734.02076	1598414.891		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	1251.62106				
POP	0.02952	0.01899814	3859829.616	2.41	0.1546

MINIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE AD

STEP 1 POP REPLACED BY COLLEGE R SQUARE = 0.40951605
C(P) = 19.68646247

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	7471851.11306	7471851.113	6.24	0.0340
ERROR	9	10773712.52331	1197079.169		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	1455.14412				
COLLEGE	0.11557	0.04625753	7471851.113	6.24	0.0340

THE ABOVE MODEL IS THE BEST 1 VARIABLE MODEL FOUND.

STEP 2 VARIABLE POP ENTERED R SQUARE = 0.43875854
C(P) = 20.36487061

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	8005396.93962	4002698.470	3.13	0.0992
ERROR	8	10240166.69674	1280020.837		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	1843.57567				
COLLEGE	0.16684	0.09270720	4145567.324	3.24	0.1096
POP	-0.02127	0.03295026	533545.827	0.42	0.5366

STEP 2 POP REPLACED BY HOUSES R SQUARE = 0.50128377
C(P) = 17.53909151

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	9146205.00735	4573102.504	4.02	0.0619
ERROR	8	9099358.62901	1137419.829		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	2090.23674				
COLLEGE	0.18301	0.07157548	7436079.515	6.54	0.0338
HOUSES	-0.08761	0.07220769	1674353.894	1.47	0.2596

MINIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE AD

STEP 2 HOUSES REPLACED BY SERV R SQUARE = 0.53983981
C(P) = 15.79658173

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	9849681.51646	4924840.758	4.69	0.0448
ERROR	8	8395882.11991	1049485.265		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	2297.11557				
COLLEGE	0.23666	0.09136489	7041377.762	6.71	0.0321
SERV	-0.74146	0.49259220	2377830.403	2.27	0.1707

STEP 2 SERV REPLACED BY TECH R SQUARE = 0.54467546
C(P) = 15.57803814

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	9937910.73975	4968955.370	4.78	0.0430
ERROR	8	8307652.89662	1038456.612		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	2238.55591				
COLLEGE	0.22695	0.08414764	7554072.618	7.27	0.0272
TECH	-0.33676	0.21853119	2466059.627	2.37	0.1619

STEP 2 TECH REPLACED BY DISTANCE R SQUARE = 0.58519244
C(P) = 13.74690464

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	10677165.93287	5338582.966	5.64	0.0296
ERROR	8	7568397.70349	946049.713		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-876.541245				
COLLEGE	0.188668	0.05716854	10303764.66	10.89	0.0109
DISTANCE	13.092994	7.11312755	3205314.82	3.39	0.1029

MINIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE AD

STEP 2 DISTANCE REPLACED BY PROF R SQUARE = 0.68755957
C(P) = 9.12050214

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	12544911.81822	6272455.909	8.80	0.0095
ERROR	8	5700651.81814	712581.477		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	2476.55484				
COLLEGE	0.39064	0.10909470	9136356.373	12.82	0.0072
PROF	-0.84615	0.31712572	5073060.705	7.12	0.0284

THE ABOVE MODEL IS THE BEST 2 VARIABLE MODEL FOUND.

STEP 3 VARIABLE DISTANCE ENTERED R SQUARE = 0.68816841
C(P) = 11.09298579

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	12556020.56789	4185340.189	5.15	0.0343
ERROR	7	5689543.06847	812791.867		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	2786.19652				
COLLEGE	0.40164	0.1497623	5845762.375	7.19	0.0315
PROF	-0.90321	0.5940628	1878854.635	2.31	0.1722
DISTANCE	-1.35196	11.5643651	11108.750	0.01	0.9102

STEP 3 DISTANCE REPLACED BY SERV R SQUARE = 0.83055267
C(P) = 4.65804010

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	15153901.61298	5051300.538	11.44	0.0043
ERROR	7	3091662.02339	441666.003		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	1917.99484				
COLLEGE	0.55549	0.10944234	11378414.28	25.76	0.0014
PROF	-2.46480	0.71124315	5304220.10	12.01	0.0105
SERV	2.21255	0.91033922	2608989.79	5.91	0.0454

MINIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE AD

STEP 3 SERV REPLACED BY TECH R SQUARE = 0.85164833
C(P) = 3.70463829

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	15538803.77199	5179601.257	13.40	0.0028
ERROR	7	2706759.86437	386679.981		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	2103.53286				
COLLEGE	0.62891	0.11743557	11089928.52	28.68	0.0011
PROF	-2.74454	0.72113601	5600893.03	14.48	0.0067
TECH	1.14542	0.41164443	2993891.95	7.74	0.0272

STEP 3 TECH REPLACED BY HOUSES R SQUARE = 0.88125083
C(P) = 2.36677613

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	16078918.15753	5359639.386	17.32	0.0013
ERROR	7	2166645.47884	309520.783		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	2122.86904				
COLLEGE	0.66570	0.10861116	11627925.77	37.57	0.0005
PROF	-2.43395	0.51428693	6932713.15	22.40	0.0021
HOUSES	0.31319	0.09268605	3534006.34	11.42	0.0118

STEP 3 HOUSES REPLACED BY POP R SQUARE = 0.90093530
C(P) = 1.47715190

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	16438072.38230	5479357.461	21.22	0.0007
ERROR	7	1807491.25406	258213.036		
TOTAL	10	18245563.63636			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	1925.72537				
COLLEGE	0.49562	0.07101864	12575416.17	48.70	0.0002
PROF	-1.91819	0.33565807	8432675.44	32.66	0.0007
POP	0.10104	0.02602155	3893160.56	15.08	0.0060

THE ABOVE MODEL IS THE BEST 3 VARIABLE MODEL FOUND.

FORWARD SELECTION PROCEDURE FOR DEPENDENT VARIABLE EP

STEP 1 VARIABLE COLLEGE ENTERED R SQUARE = 0.11356578
C(P) = 79.93903425

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	227270253.621	227270253.6	1.15	0.3109
ERROR	9	1773950954.560	197105661.6		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	7530.34035				
COLLEGE	0.63737	0.59356811	227270253.6	1.15	0.3109

STEP 2 VARIABLE DISTANCE ENTERED R SQUARE = 0.33969781
C(P) = 59.76062612

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	679810457.241	339905228.6	2.06	0.1901
ERROR	8	1321410750.941	165176343.9		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-20174.9805				
COLLEGE	1.5060	0.7553949	656487438.4	3.97	0.0813
DISTANCE	155.5723	93.9891110	452540203.6	2.74	0.1365

STEP 3 VARIABLE POP ENTERED R SQUARE = 0.54461220
C(P) = 41.66318507

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	1089889486.227	363296495.4	2.79	0.1190
ERROR	7	911331721.955	130190246.0		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-72620.3434				
COLLEGE	0.2452	0.976915	8204009.6	0.06	0.8090
POP	0.9740	0.548786	410079029.0	3.15	0.1192
DISTANCE	350.2053	137.802237	840836012.6	6.46	0.0386

FORWARD SELECTION PROCEDURE FOR DEPENDENT VARIABLE EP

STEP 4 VARIABLE PROF ENTERED R SQUARE = 0.61715064
C(P) = 36.54881454

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	4	1235054940.376	308763735.1	2.42	0.1600
ERROR	6	766166267.806	127694378.0		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-53196.8208				
COLLEGE	1.9707	1.885483	139500869.6	1.09	0.3362
PROF	-9.0998	8.534646	145165454.1	1.14	0.3273
POP	1.2986	0.622955	554865662.8	4.35	0.0822
DISTANCE	269.5381	156.043090	380998290.1	2.98	0.1349

STEP 5 VARIABLE HOUSES ENTERED R SQUARE = 0.91451297
C(P) = 9.38433356

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	5	1830142743.279	366028548.7	10.70	0.0106
ERROR	5	171078464.902	34215693.0		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-120897.807				
COLLEGE	17.098	3.756357	708913168.6	20.72	0.0061
PROF	-57.488	12.415317	733600216.3	21.44	0.0057
POP	-7.507	2.135856	422649352.6	12.35	0.0170
DISTANCE	622.321	116.962624	968635176.6	28.31	0.0031
HOUSES	30.908	7.411368	595087802.9	17.39	0.0087

STEP 6 VARIABLE SERV ENTERED R SQUARE = 0.92745403
C(P) = 10.11510915

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	6	1856040682.883	309340113.8	8.52	0.0286
ERROR	4	145180525.299	36295131.3		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-115059.034				
COLLEGE	16.985	3.871143	698709925.0	19.25	0.0118
PROF	-53.785	13.517533	574609686.7	15.83	0.0164
POP	-7.311	2.212011	396448506.2	10.92	0.0298
DISTANCE	603.563	122.493962	881180596.4	24.28	0.0079
SERV	-12.816	15.172610	25897939.6	0.71	0.4458

FORWARD SELECTION PROCEDURE FOR DEPENDENT VARIABLE EP

HOUSES	31.461	7.661240	612061629.6	16.86	0.0148

STEP 7 VARIABLE TECH ENTERED R SQUARE = 0.96941187
C(P) = 8.00000000

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	7	1940007589.815	277143941.4	13.58	0.0278
ERROR	3	61213618.367	20404539.5		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-97628.8419				
COLLEGE	17.0935	2.9030344	707427796.3	34.67	0.0098
PROF	-56.4587	10.2206502	622631268.9	30.51	0.0117
POP	-6.8587	1.6734417	342755296.5	16.80	0.0263
DISTANCE	530.8394	98.5933806	591503814.2	28.99	0.0126
TECH	12.1392	5.9840998	83966906.9	4.12	0.1355
SERV	-30.1768	14.2357598	91687835.9	4.49	0.1242
HOUSES	28.9738	5.8737107	496491326.6	24.33	0.0160

NO OTHER VARIABLES MET THE 0.5000 SIGNIFICANCE LEVEL FOR ENTRY

BACKWARD ELIMINATION PROCEDURE FOR DEPENDENT VARIABLE EP

STEP 0   ALL VARIABLES ENTERED   R SQUARE = 0.96941187
C(P) = 8.00000000

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	7	1940007589.815	277143941.4	13.58	0.0278
ERROR	3	61213618.367	20404539.5		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-97628.8419				
COLLEGE	17.0935	2.9030344	707427796.3	34.67	0.0098
PROF	-56.4587	10.2206502	622631268.9	30.51	0.0117
POP	-6.8587	1.6734417	342755296.5	16.80	0.0263
DISTANCE	530.8394	98.5933806	591503814.2	28.99	0.0126
TECH	12.1392	5.9840998	83966906.9	4.12	0.1355
SERV	-30.1768	14.2357598	91687835.9	4.49	0.1242
HOUSES	28.9738	5.8737107	496491326.6	24.33	0.0160

STEP 1   VARIABLE TECH REMOVED   R SQUARE = 0.92745403
C(P) = 10.11510915

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	6	1856040682.883	309340113.8	8.52	0.0286
ERROR	4	145180525.299	36295131.3		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-115059.034				
COLLEGE	16.985	3.871143	698709925.0	19.25	0.0118
PROF	-53.785	13.517533	574609686.7	15.83	0.0164
POP	-7.311	2.212011	396448506.2	10.92	0.0298
DISTANCE	603.563	122.493962	881180596.4	24.28	0.0079
SERV	-12.816	15.172610	25897939.6	0.71	0.4458
HOUSES	31.461	7.661240	612061629.6	16.86	0.0148

BACKWARD ELIMINATION PROCEDURE FOR DEPENDENT VARIABLE EP

STEP 2 VARIABLE SERV REMOVED R SQUARE = 0.91451297
C(P) = 9.38433356

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	5	1830142743.279	366028548.7	10.70	0.0106
ERROR	5	171078464.902	34215693.0		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-120897.807				
COLLEGE	17.098	3.756357	708913168.6	20.72	0.0061
PROF	-57.488	12.415317	733600216.3	21.44	0.0057
POP	-7.507	2.135856	422649352.6	12.35	0.0170
DISTANCE	622.321	116.962624	968635176.6	28.31	0.0031
HOUSES	30.908	7.411368	595087802.9	17.39	0.0087

ALL VARIABLES IN THE MODEL ARE SIGNIFICANT AT THE 0.1000 LEVEL.

SAS

STEPWISE REGRESSION PROCEDURE FOR DEPENDENT VARIABLE EP

NO VARIABLES MET THE 0.1500 SIGNIFICANCE LEVEL FOR ENTRY

MAXIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE EP

STEP 1 VARIABLE COLLEGE ENTERED R SQUARE = 0.11356578
C(P) = 79.93903425

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	227270253.621	227270253.6	1.15	0.3109
ERROR	9	1773950954.560	197105661.6		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	7530.34035				
COLLEGE	0.63737	0.59356811	227270253.6	1.15	0.3109

THE ABOVE MODEL IS THE BEST 1 VARIABLE MODEL FOUND.

STEP 2 VARIABLE DISTANCE ENTERED R SQUARE = 0.33969781
C(P) = 59.76062612

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	679810457.241	339905228.6	2.06	0.1901
ERROR	8	1321410750.941	165176343.9		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-20174.9805				
COLLEGE	1.5060	0.7553949	656487438.4	3.97	0.0813
DISTANCE	155.5723	93.9891110	452540203.6	2.74	0.1365

STEP 2 COLLEGE REPLACED BY POP R SQUARE = 0.54051270
C(P) = 40.06525293

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	1081685476.669	540842738.3	4.71	0.0446
ERROR	8	919535731.513	114941966.4		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-75783.5821				
POP	1.0741	0.353986	1058362458	9.21	0.0162
DISTANCE	360.2337	123.920640	971314963	8.45	0.0197

THE ABOVE MODEL IS THE BEST 2 VARIABLE MODEL FOUND.

MAXIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE EP

STEP 3 VARIABLE SERV ENTERED R SQUARE = 0.56794213
C(P) = 39.37504969

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	1136577834.824	378859278.3	3.07	0.1004
ERROR	7	864643373.358	123520481.9		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-67718.7461				
POP	1.6182	0.894775	403975919.7	3.27	0.1135
DISTANCE	325.2877	138.746025	678942756.1	5.50	0.0515
SERV	-9.7724	14.659341	54892358.2	0.44	0.5264

THE ABOVE MODEL IS THE BEST 3 VARIABLE MODEL FOUND.

MINIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE EP

STEP 1 VARIABLE DISTANCE ENTERED R SQUARE = 0.01165439
C(P) = 89.93422357

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	23323018.841	23323018.8	0.11	0.7520
ERROR	9	1977898189.341	219766465.5		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	8892.52071				
DISTANCE	25.40486	77.9839466	23323018.84	0.11	0.7520

STEP 1 DISTANCE REPLACED BY HOUSES R SQUARE = 0.01797151
C(P) = 89.31465817

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	35964965.385	35964965.4	0.16	0.6943
ERROR	9	1965256242.797	218361804.8		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	9003.47323				
HOUSES	0.25579	0.63027279	35964965.38	0.16	0.6943

STEP 1 HOUSES REPLACED BY SERV R SQUARE = 0.02339807
C(P) = 88.78243618

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	46824709.983	46824710.0	0.22	0.6534
ERROR	9	1954396498.199	217155166.5		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	8649.69944				
SERV	1.55979	3.35903322	46824709.98	0.22	0.6534

MINIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE EP

STEP 1   SERV REPLACED BY TECH   R SQUARE = 0.02953209
C(P) = 88.18082793

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	59100249.247	59100249.2	0.27	0.6134
ERROR	9	1942120958.935	215791217.7		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	8272.55326				
TECH	0.84409	1.61290729	59100249.25	0.27	0.6134

STEP 1   TECH REPLACED BY PROF   R SQUARE = 0.03537394
C(P) = 87.60787586

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	70791072.471	70791072.5	0.33	0.5797
ERROR	9	1930430135.711	214492237.3		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	8557.67202				
PROF	1.03404	1.79992576	70791072.47	0.33	0.5797

STEP 1   PROF REPLACED BY POP   R SQUARE = 0.05515158
C(P) = 85.66813880

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	110370513.737	110370513.7	0.53	0.4870
ERROR	9	1890850694.444	210094521.6		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	6588.47445				
POP	0.15787	0.21780794	110370513.7	0.53	0.4870

MINIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE EP

STEP 1 VARIABLE DISTANCE ENTERED R SQUARE = 0.01165439
C(P) = 89.93422357

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	23323018.841	23323018.8	0.11	0.7520
ERROR	9	1977898189.341	219766465.5		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	8892.52071				
DISTANCE	25.40486	77.9839466	23323018.84	0.11	0.7520

STEP 1 DISTANCE REPLACED BY HOUSES R SQUARE = 0.01797151
C(P) = 89.31465817

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	35964965.385	35964965.4	0.16	0.6943
ERROR	9	1965256242.797	218361804.8		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	9003.47323				
HOUSES	0.25579	0.63027279	35964965.38	0.16	0.6943

STEP 1 HOUSES REPLACED BY SERV R SQUARE = 0.02339807
C(P) = 88.78243618

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	46824709.983	46824710.0	0.22	0.6534
ERROR	9	1954396498.199	217155166.5		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	8649.69944				
SERV	1.55979	3.35903322	46824709.98	0.22	0.6534

MINIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE EP

STEP 1 SERV REPLACED BY TECH — R SQUARE = 0.02953209, C(P) = 88.18082793

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	59100249.247	59100249.2	0.27	0.6134
ERROR	9	1942120958.935	215791217.7		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	8272.55326				
TECH	0.84409	1.61290729	59100249.25	0.27	0.6134

STEP 1 TECH REPLACED BY PROF — R SQUARE = 0.03537394, C(P) = 87.60787586

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	70791072.471	70791072.5	0.33	0.5797
ERROR	9	1930430135.711	214492237.3		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	8557.67202				
PROF	1.03404	1.79992576	70791072.47	0.33	0.5797

STEP 1 PROF REPLACED BY POP — R SQUARE = 0.05515158, C(P) = 85.66813880

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	110370513.737	110370513.7	0.53	0.4870
ERROR	9	1890850694.444	210094521.6		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	6588.47445				
POP	0.15787	0.21780794	110370513.7	0.53	0.4870

MINIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE EP

STEP 1 POP REPLACED BY COLLEGE R SQUARE = 0.11356578
C(P) = 79.93903425

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	1	227270253.621	227270253.6	1.15	0.3109
ERROR	9	1773950954.560	197105661.6		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	7530.34035				
COLLEGE	0.63737	0.59356811	227270253.6	1.15	0.3109

THE ABOVE MODEL IS THE BEST 1 VARIABLE MODEL FOUND.

STEP 2 VARIABLE POP ENTERED R SQUARE = 0.12445075
C(P) = 80.87146691

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	249053473.579	124526736.8	0.57	0.5877
ERROR	8	1752167734.602	219020966.8		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	10012.2712				
COLLEGE	0.9650	1.21268442	138682959.8	0.63	0.4492
POP	-0.1359	0.43101575	21783220.0	0.10	0.7606

STEP 2 COLLEGE REPLACED BY TECH R SQUARE = 0.13497618
C(P) = 79.83916094

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	270117201.443	135058600.7	0.62	0.5599
ERROR	8	1731104006.739	216388000.8		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	6186.64576				
POP	1.06343	1.07688167	211016952.2	0.98	0.3523
TECH	-6.76073	7.86853998	159746687.7	0.74	0.4152

MINIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE EP

STEP 2 POP REPLACED BY COLLEGE — R SQUARE = 0.16634302, C(P) = 76.76278780

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	332889178.577	166444589.3	0.80	0.4830
ERROR	8	1668332029.605	208541503.7		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	12657.2916				
COLLEGE	1.3663	1.19246047	273788929.3	1.31	0.2850
TECH	-2.2039	3.09681675	105618925.0	0.51	0.4969

STEP 2 TECH REPLACED BY SERV — R SQUARE = 0.20552717, C(P) = 72.91971366

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	411305336.383	205652668.2	1.03	0.3984
ERROR	8	1589915871.798	198739484.0		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	14937.5944				
COLLEGE	1.7027	1.25728354	364480626.4	1.83	0.2127
SERV	-6.5230	6.77862231	184035082.8	0.93	0.3641

STEP 2 COLLEGE REPLACED BY POP — R SQUARE = 0.22867791, C(P) = 70.64915311

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	457635078.718	228817539.4	1.19	0.3540
ERROR	8	1543586129.464	192948266.2		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	6760.75252				
POP	1.63176	1.1182935	410810368.7	2.13	0.1826
SERV	-22.75770	16.9636251	347264565.0	1.80	0.2166

SAS

MINIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE EP

STEP 2 SERV REPLACED BY HOUSES — R SQUARE = 0.33833263, C(P) = 59.89451919

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	677078430.854	338539215.4	2.05	0.1917
ERROR	8	1324142777.328	165517847.2		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	8998.14839				
POP	2.09030	1.06209749	641113465.5	3.87	0.0846
HOUSES	-5.57822	3.01465937	566707917.1	3.42	0.1014

STEP 2 POP REPLACED BY DISTANCE — R SQUARE = 0.42619806, C(P) = 51.27691866

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	852916600.881	426458300.4	2.97	0.1084
ERROR	8	1148304607.301	143538075.9		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-87789.0274				
DISTANCE	412.6615	172.973390	816951635.5	5.69	0.0442
HOUSES	3.3717	1.402475	829593582.0	5.78	0.0429

STEP 2 HOUSES REPLACED BY POP — R SQUARE = 0.54051270, C(P) = 40.06525293

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	2	1081685476.669	540842738.3	4.71	0.0446
ERROR	8	919535731.513	114941966.4		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-75783.5821				
POP	1.0741	0.353986	1058362458	9.21	0.0162
DISTANCE	360.2337	123.920640	971314963	8.45	0.0197

THE ABOVE MODEL IS THE BEST 2 VARIABLE MODEL FOUND.

MINIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE EP

STEP 3 VARIABLE COLLEGE ENTERED R SQUARE = 0.54461220
C(P) = 41.66318507

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	1089889486.227	363296495.4	2.79	0.1190
ERROR	7	911331721.955	130190246.0		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-72620.3434				
COLLEGE	0.2452	0.976915	8204009.6	0.06	0.8090
POP	0.9740	0.548786	410079029.0	3.15	0.1192
DISTANCE	350.2053	137.802237	840836012.6	6.46	0.0386

STEP 3 COLLEGE REPLACED BY TECH R SQUARE = 0.54622210
C(P) = 41.50529054

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	1093111251.337	364370417.1	2.81	0.1177
ERROR	7	908109956.845	129729993.8		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-73064.5416				
POP	1.2967	0.838947	309923811.5	2.39	0.1661
DISTANCE	347.8497	138.106285	822994049.9	6.34	0.0399
TECH	-1.8967	6.391263	11425774.7	0.09	0.7753

STEP 3 TECH REPLACED BY PROF R SQUARE = 0.54744276
C(P) = 41.38557113

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	1095554070.765	365184690.3	2.82	0.1166
ERROR	7	905667137.417	129381019.6		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-76232.9543				
PROF	-1.4433	4.408246	13868594.1	0.11	0.7529
POP	1.2374	0.624284	508320468.9	3.93	0.0879
DISTANCE	358.6308	131.565045	961357238.7	7.43	0.0295

MINIMUM R-SQUARE IMPROVEMENT FOR DEPENDENT VARIABLE EP

STEP 3 PROF REPLACED BY HOUSES

R SQUARE = 0.54792147
C(P) = 41.33862096

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	1096512067.304	365504022.4	2.83	0.1163
ERROR	7	904709140.878	129244163.0		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-66012.4520				
POP	1.3936	1.015091	243595466.4	1.88	0.2122
DISTANCE	319.8049	177.524770	419433636.4	3.25	0.1146
HOUSES	-1.2189	3.598906	14826590.6	0.11	0.7448

STEP 3 HOUSES REPLACED BY SERV

R SQUARE = 0.56794213
C(P) = 39.37504969

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	3	1136577834.824	378859278.3	3.07	0.1004
ERROR	7	864643373.358	123520481.9		
TOTAL	10	2001221208.182			

	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-67718.7461				
POP	1.6182	0.894775	403975919.7	3.27	0.1135
DISTANCE	325.2877	138.746025	678942756.1	5.50	0.0515
SERV	-9.7724	14.659341	54892358.2	0.44	0.5264

THE ABOVE MODEL IS THE BEST 3 VARIABLE MODEL FOUND.

A MODEL FOR PREDICTING AIR TRAVEL DEMAND IN SMALL COMMUNITIES

by

Rick Donnelly

B.S., Kansas State University. 1983

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF REGIONAL AND COMMUNITY PLANNING

Department of Regional and Community Planning

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1985

Planners have used secondary data sources, such as census data, to predict urban travel demand for several years. Several researchers found that secondary data sources, such as traffic volume counts and censal data, estimated travel demand fairly well. The ability to use such short-cut methods of data collection have enabled planners to employ sketch planning techniques to study a wide number of policy alternatives. While urban transportation planners have used these secondary sources for quite some time, there does not appear to be such a trend in air travel demand forecasting.

This study attempted to show the utility of using census data to construct models of demand for air travel in smaller communities.

A model of demand was constructed for the State of Kansas using available census data and statistics provided by the Federal Aviation Administration on passenger movements. Regression analysis was used to describe a relationship which suggests that the population size, number of college graduates, and the number of persons employed in professional categories all influence demand. The model obtained was able to demonstrate the structure of demand, rather than the values of the components. While a model that could be used to accurately describe demand was desired, the limitations of the database precluded arriving at such. Although an accurate model could not be obtained, it showed nonetheless that such is probably obtainable. The ability to fit such scant data to a model suggests that indirect data sources might be as capable of describing demand as survey methods. The advantages of using secondary sources such as the census data lies in the ease of data acquisition and the simplicity of the study scheme.

This study has shown the feasibility of using indirect data sources to describe air travel demand in small communities. Although the data and results obtained are localized in time and place, it nevertheless represents the potential for constructing such models. The traditional models have all used survey data either exclusively or in part, have been costly to construct, and have varying levels of accuracy. By contrast, the model described is easy to apply, requiring readily available data and simple statistical techniques.