RETAIL SITE SELECTION USING MULTIPLE REGRESSION ANALYSIS

DISSERTATION

Presented to the Graduate Council of the North Texas State University in Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Ronald D. Taylor, M.B.A.
Denton, Texas
December, 1978

Samples of stores were drawn from two chains, Pizza Hut and Zale Corporation. Two different samples were taken from Pizza Hut. Site specific material and sales data were furnished by the companies and demographic material relative to each site was gathered. Analysis of variance tests for linearity were run on the three regression equations developed from the data and each of the three regressions equations were found to have a statistically significant linear relationship. Statistically significant differences were found among similar variables used in the prediction of sales by using Fisher's Z' Transformations on the correlation coefficients. Eight of the eighteen variables used in the Pizza Hut study were found to be statistically different between the two regions used in the study. Additionally, analysis of variance tests were used to show that traffic pattern variables were not better predictors than demographic variables.
# TABLE OF CONTENTS

**LIST OF TABLES** ............................... v

**Chapter**

I. RETAIL LOCATION: AN INTRODUCTION .... 1
   The Importance of Retail Location
   Expenditures for Store Location
   Research
   Chain Research Responsibilities and
   Factors Studied
   Regression Analysis for Site Selection
   Proposed Study
   Methodology
   Research Hypotheses of the Study
   Outline of Dissertation

II. BASIC IDEAS IN SITE SELECTION ....... 10
   Theories of City Growth
   Selecting a City
   Store Trade Areas
   Generative and Suscipient Locations
   Nelson's Principles of Store
   Location
   The Rule of Retail Compatibility
   Major Considerations in Site
   Evaluation
   The Ratio of Advertising to Rent
   for Various Retail Types
   Strategy in Store Location

III. CURRENT APPROACHES TO SITE SELECTION .. 42
   Gravitational Models
   Analog Method
   Map Transformation Method
   Checklist Method
   Behavioral Models
   Miscellaneous Methods
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>Illustration of Applebaum's Site Evaluation Method</td>
<td>53</td>
</tr>
<tr>
<td>II.</td>
<td>Pizza Hut Independent Variables Used in the Study</td>
<td>66</td>
</tr>
<tr>
<td>III.</td>
<td>List of the Independent Variables Considered for Zale Corporation</td>
<td>67</td>
</tr>
<tr>
<td>IV.</td>
<td>Regression Equation for Pizza Hut Stores in Texas to Predict Sales of Stores</td>
<td>70</td>
</tr>
<tr>
<td>V.</td>
<td>Regression Equation Summary Table for Pizza Huts in Texas</td>
<td>71</td>
</tr>
<tr>
<td>VI.</td>
<td>The Residual as a Percent of the Actual Sales</td>
<td>72</td>
</tr>
<tr>
<td>VII.</td>
<td>Summary Table of Standardized Residuals for Pizza Huts in Texas</td>
<td>72</td>
</tr>
<tr>
<td>VIII.</td>
<td>Regression Equation for Pizza Hut Stores in Louisiana</td>
<td>73</td>
</tr>
<tr>
<td>IX.</td>
<td>Regression Equation Summary Table for Pizza Hut Stores in Louisiana</td>
<td>74</td>
</tr>
<tr>
<td>X.</td>
<td>The Residual as a Percent of the Actual Sales</td>
<td>75</td>
</tr>
<tr>
<td>XI.</td>
<td>Regression Equation Summary for Standardized Residuals for Pizza Hut Stores in Louisiana</td>
<td>75</td>
</tr>
<tr>
<td>XII.</td>
<td>Regression Equation for Zale Corporation to Predict Sales of Stores</td>
<td>76</td>
</tr>
<tr>
<td>XIII.</td>
<td>Summary Table for Zale Corporation Regression Equation</td>
<td>77</td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>XIV.</td>
<td>The Residual as a Percent of Actual Sales</td>
<td>78</td>
</tr>
<tr>
<td>XV.</td>
<td>Summary Table for Residuals for Zale Corporation</td>
<td>79</td>
</tr>
<tr>
<td>XVI.</td>
<td>Simple Correlations Between Original Independent Variables and Sales for Zale's Stores</td>
<td>80</td>
</tr>
<tr>
<td>XVII.</td>
<td>Simple Correlations Between Original Independent Variables and Sales for Pizza Hut Stores in Texas and Louisiana</td>
<td>81</td>
</tr>
<tr>
<td>XVIII.</td>
<td>Fisher's Z' Transformations as a Comparison Between Correlations of Similar Variables for Pizza Hut and Zales</td>
<td>82</td>
</tr>
<tr>
<td>IXX.</td>
<td>Fisher's Z' Transformations Between Correlations of Similar Variables for Pizza Hut Stores in Louisiana and Texas</td>
<td>83</td>
</tr>
<tr>
<td>XX.</td>
<td>Regression Equation for Pizza Hut Using All the Original Variables</td>
<td>84</td>
</tr>
<tr>
<td>XXI.</td>
<td>Regression Equation for Original Demographic Variables for Pizza Hut Stores</td>
<td>85</td>
</tr>
<tr>
<td>XXII.</td>
<td>Test for Linearity of Demographic Regression Equation</td>
<td>85</td>
</tr>
<tr>
<td>XXIII.</td>
<td>Regression Equation for Original Traffic Pattern Variables for Pizza Hut Stores</td>
<td>85</td>
</tr>
<tr>
<td>XXIV.</td>
<td>Test for Linearity of Traffic Pattern Regression Equation</td>
<td>86</td>
</tr>
<tr>
<td>XXV.</td>
<td>Pizza Hut Regression Equation for the Original Variables of Pizza Hut Stores Excluding Demographic Variables</td>
<td>86</td>
</tr>
<tr>
<td>XXVI.</td>
<td>Regression Equation for Original Variables of Pizza Hut Stores with Traffic Pattern Variables Excluded</td>
<td>87</td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>XXVII. Significance Test for Semi-Partial Correlation Between the Full Model Regression Equation and the Equation that Excludes the Traffic Patterns</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>XXVIII. Significance Test for Semi-Partial Correlation Between the Full Model Regression Equation and the Equation that Excludes the Demographic Variables</td>
<td>88</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER I

RETAIL LOCATION: AN INTRODUCTION

Because of the continually shifting population, spiraling costs of land and construction, the increasing number of planned shopping centers, and changing urban conditions, much emphasis is being placed on the selection of sites for retail stores. Too often new businesses will settle into a building vacated due to the failure of the previous tenant. Obviously, the new occupants have probably done little more research than finding an empty building. Another common error is the consideration of rental costs as the only criterion. Normally low rental costs are associated with poor sites. Promotional costs are considered to be polarized with rental costs. The worse the site, the more promotional expenses that will be needed for the store to be successful. Thus, the old saying that "if you build a better mousetrap, the world will beat a path to your door," is not quite correct. If the public can't find your door, then you have to promote your product in order for people to discover where to buy it.

The Importance of Retail Location

Some researchers consider the location of the store as the most important variable in the formation of the consumer's
image of the store.¹ Other studies have concluded that location is a negative factor only.² Thus a poor location could have some negative effects on the image of the store, while a good location might have little or no effect. For some companies the location of the store is their sole competitive advantage. A good example of this point is the 7-11 stores. This chain locates close to housing developments and apartments in hopes of getting numerous small transactions from the people who live close.

For small businesses, the failure rate due to poor location is between two to five percent. Certainly, poor location is a contributing factor to the failure of many other small businesses. For chain stores, the failure rate of their stores due to poor location is less than one percent, on the average. Yet, a failure for a chain store represents a loss of several thousand dollars.

Expenditures for Store Location Research

A 1964 survey of the store location procedures of retail chains indicated that the average expenditure for locational


research per store was $4,075.00, which was slightly over one percent of the average investment per location taken.\(^3\) However, none of the respondents to the survey indicated specifically what was included in the expense. Thus, it could be possible that part of this expenditure was for real estate acquisition, which would additionally reduce the amount of research done. Chains with annual sales over $500 million accounted for nearly two-thirds of the total amount spent for research expenditures. Chains with sales below $100 million accounted for less than nine percent of the total.\(^4\)

Chain Research Responsibilities and Factors Studied

Smaller companies tend to rely more on the decision of one person, whereas in the largest chains the decision is normally relegated to a group.\(^5\) One study indicated that the president had the final decision toward site selection in 21 percent of the companies polled, while a vice president made the decision in 26 percent of the cases, an executive committee in 35 percent, and a research department in 18 percent.\(^6\) In the same study, it was found that all but 10 percent of the respondents did some type of store location research.\(^7\) However, little information was given as to the amount of input


\(^4\)Ibid., p. 54.

\(^5\)Ibid., p. 55.

\(^6\)Ibid., p. 55.

\(^7\)Ibid., p. 55.
that various functional areas of the company had into the decisions. It is entirely possible that the only input considered came from the real estate department. The most commonly mentioned location factors that were considered by the respondents in the survey listed in descending order are the following: population (characteristics and income); level and quality of competition; automobile traffic patterns; economic situation of the area; and the occurrence of shopping centers.  

Regression Analysis for Site Selection

Chapter III presents a detailed discussion of the common methods of quantitatively determining the site selection decision. All of the methods have limitations. Huff's equation has received much publicity and study due to its chief advantage of simplicity of calculation. Unfortunately, computations are eased due to its dependence upon only two variables both of which are questionable at best--the driving time needed by customers to reach the site and the size of the store. It is certainly intuitively appealing to include more than two variables and in fact researchers have theorized and tested other variables successfully (see the section on Huff's formula). Another problem with Huff's equation is that computations requiring comparisons of several sites are quite cumbersome.

The use of the Central Place Theory and Reilly's law of retail gravitation were designed to determine the optimum

---

8Ibid., p. 55.
regions in which a new store should be placed. Consequently, the use of these models for the selection of a specific site should be limited to rural locations at best.

The survey method is interesting due to its reliance upon the consumer, rather than strictly attending to the physical characteristics associated with the site. Consumer preferences are hopefully determined from a sample garnered by interviewing people in the prospective area. The accuracy of the data is only as good as the construction of the questions asked, the sample selected, and the method of analysis. Too often subjectivity enters into the data compilation. Another problem associated with this method is that there is still not a concrete method of selecting the best site.

The analog method depends heavily upon the historical performance of the firm under a variety of market conditions. Unfortunately, this entire approach is not wrapped around an objective attempt at selection of the best site, but rather subjectivity upon the researcher’s part is the source of the decision.

The simulation model suggested by MacKay has been successfully employed, but only on grocery stores, so far. A problem with this model is that it relies on the quantification of human behavior, something that has not been successful in the past.

Each of the models discussed above is limited. All could gain by adapting the best facets of the others. Huff’s model
has the advantage of ease in calculation of the "best" site. The analog method suggests the incorporation of historical data relating to the operation of stores under a variety of conditions. The behavioral model's strong point is the dependence upon the characteristics of the consumer. This emphasis on the consumer is inline with the prevailing philosophies associated with the marketing concept. The proposed model is an attempt to incorporate the above features along with the traffic patterns of each of the sites in question. The approach used to tie these ideas together will be regression analysis.

The use of regression analysis in this area is not new. However, it has not been heavily researched. In fact, the format suggested has not been researched at all. One study using regression analysis attempted to forecast the sales associated with sites based on the following variables: gross selling area; rent expenses; distance to the nearest parking area; the accessibility of the store; and the urban growth rate.\(^9\) Another study coupled regression with the Automatic Interaction Detector (A.I.D.), a multivariate technique. The object of this approach was to establish a retrospective way for evaluating the economic health of existing stores. The significant variables found in this study were the quantity of parking spaces, the selling area in square feet, the

---

population of the catchment area, and the number of competitors in the region. Other factors considered were social class of the catchment area; age characteristics; car ownership; population of the catchment area; type of households; size of households; availability of public transportation; exterior appearance; frontage of store; displayed stock; appearance of interior; staff salaries; and number of checkouts.\textsuperscript{10} While this is a valid incorporation of regression techniques, the incorporation of the A.I.D. program in this manner is rhetorical at best.

Proposed Study

Regression techniques will be used to predict the sales associated with sites based on the independent variables of the surrounding demographic characteristics and traffic patterns. Data will be drawn from chain stores in order to establish an analog of data that are characteristic of the chain and its performance under a variety of conditions. Whenever possible the data used will reflect the demographic and traffic patterns around the site that were prevalent at the time of the store's opening. Also, the sales to be used in the formulation of the model will be the first years of operation of the store, if available. The above decisions are under

\begin{footnotesize}
\begin{enumerate}
\end{enumerate}
\end{footnotesize}
the assumption that most failures occur during the first years of operation. Possible confounding variables are the following: the size of the store in square feet; the amount of promotional dollars allotted for each store; and the number of parking spaces. The inclusion of these variables into the initial pool of items to be considered will hopefully assure that any importance associated with them will be included if the variable is found to make a significant contribution to the predictive power of the model.

Methodology

Sales data will be gathered from chain stores willing to cooperate in the study. Also, information relative to the traffic patterns, amount and quality of competition and accessibility of the site will be furnished by the stores, whenever possible. Demographic data for each store will either be furnished by the stores or found in the U. S. Census of Population. The sales data to be used will be the amount of total sales at each site for the initial years of operation.

Both factor analysis and stepwise regression technique will be used to establish a regression equation for the chain. The standard error of the regression equation will be calculated and the effectiveness of the equation will be measured by the percentage distribution of the amount each predicted sales figure missed the the actual sales figure.
Research Hypotheses of the Study

The hypotheses to be studied are the following.

1. Regression analysis will be found to be a viable tool for selecting retail sites.

2. There will be a difference among the variables that are important for different chains.

3. It will be possible to discover regional differences in the performance characteristics of a chain through the use of regression analysis.

4. Demographic variables will be better predictors of the sales associated with a site than traffic patterns.

Outline of Dissertation

Chapter II describes some of the theories of city growth, gives input toward the selection of cities, explains some principles of site selection, and discusses some of the major considerations of site selection. Chapter III discusses the historical background and computational philosophies of the most commonly used retail site selection models. Chapter IV contains the results of the study, a discussion of the computations, the equations found for each chain, and an interpretation of the importance of each variable used in the equations. Chapter V is composed of a discussion of the findings of the prior chapter, conclusions based on this material, and suggestions for further study.
CHAPTER II

BASIC IDEAS IN SITE SELECTION

Three decisions have to be made relative to the choice of a location. The three are the selection of a city, the choice of an area within the city, and finally the decision on a specific site.\(^1\) This is the logical sequence of decisions, but all decisions are not made in this order. In fact, with the rapid popularity of regional malls, the order is sometimes switched for chains that sell more specialized items. However, the following pages follow the traditional sequence of events by first discussing factors dealing with cities and their choice, followed by a discussion of the choice of an area within the city, and then specifics dealing with the individual sites.

Theories of City Growth

Large urban areas have been studied from the macro view and various theories have evolved that attempt to explain the structure of cities. The three theories that relate the most to cities in North America are the concentric circle theory, the sector theory and the multiple nuclei theory.\(^2\)


Concentric Circle Theory. Ernest Burgess, a sociologist, feels that under many conditions cities have taken the form of five concentric zones. The inner most circle is the central business district which tends to be crowded in the daytime but relatively empty at night. The second zone is a core of transition or deterioration, where the lowest socioeconomic classes live. The third circle is inhabited by the working people of nearby shops and offices. The next zone is composed of middle class workers, professional people, clerks, etc. who often live in apartments or closely built houses. Beyond this is the final circle where the commuters live. These people can afford the longer trip to their work. Of course, there are exceptions to all theories. For example, Chicago is bounded on the east by Lake Michigan which affects the application of this theory.

Sector Theory. This theory suggests that a trend will exist within each city for the high rent areas to move outward toward the outskirts in a path indicated by one or more sectors. Should a low-rent area develop in one sector, the remainder of that sector is likely to be occupied over a period of time by low-rent alternatives. Additionally, the following propositions form a basis for the theory.

---


4Ibid., pp. 75-76.

5Dickinson, p. 135.
A grouping occurs in the social order, primarily as a result of income and social position;

Low-priced housing is located near the business and industrial center, from which it fans out into the sector in question;

The main growth of cities in this country takes place by new buildings at the periphery rather than by rebuilding older areas;

High-priced residential areas will tend to develop along the faster transportation lines;

Office buildings, banks and stores will generally pull the high-priced residential neighborhoods in the same geographic direction.6

The sector theory places emphasis on transportation, income, and social class in the development of residential areas. There are many examples of U.S. cities that developed according to the theory. Among the best examples are San Francisco, Seattle, Boston and Minneapolis.7

Multiple Nuclei Theory. This approach proposes that in cities there are several nuclei around which the usage patterns of land have evolved. Four factors combine to create separate nuclei.

(1) Certain activities have specialized requirements. For example, the retail district should be most easily accessible to transportation, or a manufacturing area requires access to rail or truck transportation, as well as a large tract of land at a lower price than the core area would demand.

(2) Certain similar activities tend to locate close together because of the advantages

6Ibid., p. 135. 7Ibid., p. 136.
to them of cohesion. The clustering of financial establishments in the Wall Street area; the development of such facilities as the Merchandise Mart or the Furniture Mart in Chicago; and the greater profit shown to result from clustering of retail establishments are examples of this factor.

(3) Certain unlike activities are detrimental to each other. The presence of large air-polluting factories would detract from the success of a shopping center.

(4) Certain activities are unable to afford the high rents of the most desirable sites. A large wholesale establishment that requires storage space can hardly afford to locate in the high priced residential area along a lake front, nor can low-class housing afford to be out in the woods and hilly terrain of the expensive suburbs.8

The multiple nuclei theory seems to be relevant in many cities. The dominance of the central business district is disappearing, and areas are developing many strong centers.9

Selecting A City

Some communities are considered to be much more desirable than others for locating new stores. In fact, many large chains of stores have a "check list" or some other structured evaluational technique for comparing cities.10 Among the variables that are generally felt as basic in choosing cities are the following:

1. The size of the trading area from which the stores draw business;

---

8Mulirhill, pp. 77-78. 9Dickinson, p. 136.

10Davidson, Doody and Sweeney, p. 483.
2. The population of the trading area and all demographic trends that could effect the purchasing patterns of the city;

3. The total purchasing power of the area with consideration to the level of disposable income, the distribution of the income among the consumers, and the sources of income;

4. The total retail trade potential for the store;

5. The competitive conditions in respect to the number of competitors, their size and the level of business they have;

6. The level of progressiveness of the community which includes new construction, the level of the school system, the police and fire protection, the quality of living, etc.\(^1\)

Other factors that should also be considered are the following:

1. The quality of the available advertising media which should include their coverage of the market and their costs of use;

2. The availability of quality labor at reasonable costs;

3. The proximity of the city to adequate wholesale facilities;

4. Legal restrictions such as minimum wage laws, regulations of store hours, local taxes, and licensing requirements.\(^2\)

Types of Locations Within A City

Small town locational analysis is relatively simple due to the limited number of possibilities. Unfortunately, this cannot be said for larger cities and towns. Below is a

\(^{11}\text{Ibid., p. 484.}\) \(^{12}\text{Ibid., p. 484.}\)
classification of urban-retail locations, which has been found to be highly useful for study and analysis.\textsuperscript{13}

\textbf{Unplanned Business Districts.} This category consists of those retail locations which have developed from uncoordinated site decisions by independent retailers. These include the following types:

1) The central business district which is the intensively used center of the transportation facilities and contains a conglomerate of stores;

2) The frame of the central business district which contains retail stores and other types of businesses that are not capable of supporting the higher rentals associated with other areas of the city;

3) The secondary business districts, which serve major portions of the central city and some of the suburbs;

4) Strings of businesses which are normally along the major street and transportation lines;

5) Clusters of stores in neighborhoods aimed at offering convenience aspects to the consumer;

6) Isolated locations which are normally at the fringe of the city or suburbs.\textsuperscript{14}

In recent years there has been a strong trend toward the decline in importance of the central business district. Shifts


\textsuperscript{14}Davidson, Doody, and Sweeney, p. 484.
in population from urban areas to the suburbs is one reason for this decline. The suburban population growth rate has rapidly outpaced the metropolitan rate since 1945.\textsuperscript{15} Another reason for the central business district decline is the increasing reliance on automobiles as the major transportation mode for people. Downtown areas harbor traffic congestion and parking problems, while shopping in the other areas of the city aren't plagued with this problem as much. A final factor that affects the desirability of downtown shopping is the older, less attractive facilities which compare unfavorably with the more modern facilities of the suburban shopping centers.\textsuperscript{16}

\textbf{Planned Shopping Centers.} Practically all of the planned shopping centers have been built since World War II. Centers such as these differ from others in that they are planned, controlled, and operated by one organization. It is not uncommon that the developer of the shopping center is one of the major tenants.\textsuperscript{17}

Normally, these centers are classified on the basis of their size and the area of the market that is served. The first type is the neighborhood center, usually consisting of

\textsuperscript{15}Ibid., p. 485.
\textsuperscript{16}Ibid., p. 486.
a small strip of stores predominately selling convenience goods. "Neighborhood centers typically have between 30,000 and 75,000 square feet of gross floor area, require at least 1000 families in the trading area, and have a supermarket or drugstore as the leading tenant."\(^1\) The second type is the community shopping center. "Total floor space in the community center ranges between 100,000 and 300,000 square feet, and the leading tenant is typically a variety or junior department store."\(^2\) The final and largest is the regional shopping center. Normally, the regional center will serve several hundred thousand people and have over 500,000 square feet. Frequently, there will be several large department stores as major tenants.\(^3\)

The layouts of shopping centers have taken many forms over the years. However, it is still possible to group all of the centers into three categories: strip centers, malls, and cluster centers.\(^4\) The strip center is the most common and derives its name from the alignment of its stores. A grouping of a large number of stores in a strip center is usually not convenient for the consumers due to the long distance that they would be forced to walk. Large groups of stores such as in regional centers are normally housed in a mall. All or most of the stores face the mall with parking

\(^{1}\)Davidson, Doody, and Sweeney, p. 487.

\(^{2}\)Ibid., p. 487.

\(^{3}\)Ibid., p. 487.

\(^{4}\)Ibid., p. 487.
surrounding the exterior. This arrangement permits a large concentration of store fronts in limited areas. The cluster centers are the rarest of the three layouts. The arrangement for this type is such that all of the smaller stores are grouped around the single, large tenant.  

One of the greatest advantages of the planned shopping centers is that the number of direct competitors for each store is limited. Other advantages include: more than adequate parking; modern and attractive atmospherics; the combined attraction power of several stores; and the strong feelings of community interest. Occupancy of a planned shopping center is not without disadvantages and the following are the most common: not every planned shopping center is well conceived, some are, in fact, very poorly planned; it is often hard for a retailer to achieve an individualized image; many major shopping centers in the same metropolitan area have essentially the same tenant composition; and the rental costs are extremely high per square foot.  

Store Trade Areas

"A store trade area is that area from which a store gets its business--where the customers come from." The needs and wants of consumers generally move around in a relatively

---

22Ibid., pp. 487-488.  
23Ibid., pp. 488-489.  
24Ibid., pp. 489-490.  
confined area, they are more or less bound to the stores that are along their regular lines of travel. The movement of consumers may originate at home, work, or recreation, but most shopping trips start from the home. The movement patterns of consumers are fairly predictable, just like their movement patterns within a store are predictable. Thus, stores should not expect to lure consumers out of their way just to trade at the store unless they have a considerable uniqueness within their marketing mix.\textsuperscript{26}

"A store's trade area is determined by the convenience of the location and by what the store sells, the manner in which it sells, the shopping habits of its consumer, existing competition, and the type of location."\textsuperscript{27} Assuming that all stores offer similar goods and services, then the location of the stores will become the major factor in determining the trade area of each store.\textsuperscript{28}

The trade area of a store is affected by the trade area of neighboring stores. "In theory the extent of the most powerful store's trade area serves as the boundary of all other stores."\textsuperscript{29} However, in practice the volume of sales garnered by a store from the more remote parts of its trade area are only slightly more than if the store were free-standing.\textsuperscript{30} Also, the size and shape of the store trade areas are

\textsuperscript{26}Ibid., p. 33. \textsuperscript{27}Ibid., p. 33. \textsuperscript{28}Ibid., p. 33. \textsuperscript{29}Ibid., p. 33. \textsuperscript{30}Ibid., p. 33.
affected by population density, accessibility, geographical features, and store image.\textsuperscript{31} For example, the more dense the population, the larger the size of the trade area. Also, density in population means that a greater portion of sales will come from the close proximity of the store.\textsuperscript{32} Accessibility is important because trade areas are directly influenced by street patterns. "In general store trade area shapes are stretched out longest along main highways; depending on where the population is concentrated, access to the highway, and the location of competition."\textsuperscript{33} Geographic features may present barriers to the consumer that greatly effect the trade area. For example, a large river bordering the location of a store would limit the access of the public to the store. Store image definitely affects the trade area of the store, but is hard to define and measure. Competition, of course, also affects the trade areas of stores. However, a large portion of the impact of competition is decided by timing; whether the competition opens before or after the store in question. Once trade area boundaries have been established, the addition of nearby competing stores has little affect. Consumers form shopping habits and generally remain faithful.\textsuperscript{34}

Trade areas change, just as population shifts and consumer buying habits change. However, trade areas are very

\textsuperscript{31}Ibid., p. 33  \hspace{1cm}  \textsuperscript{32}Ibid., p. 33.
\textsuperscript{33}Ibid., p. 33.  \hspace{1cm}  \textsuperscript{34}Ibid., p. 35.
slow in their changes. One trend that has had a slow affect on trade areas is that of scrambled merchandising. Stores have adopted the policy of broadening their product lines to the point that it sometimes is difficult to tell a drug store from a food supermarket. Thus, trade areas are changing due to the addition of many more competitors. Currently, the trade areas of service stations are being altered by the tremendous gasoline sales of the convenience stores. Despite scrambled merchandising, population shifts, and changes in consumer buying habits the trade areas of most stores are only nominally different than they were in the past.

Generate and Suscipient Locations

According to Nelson there are two types of retail locations, suscipient and generative.35 A generative location is one that attracts consumers from their residence. The consumers' principle purpose for leaving their home is to shop at this particular location. A generative location is quite accessible and for this reason is chosen as the shopping place of many consumers.36 "A suscipient location is one to which the consumer is impulsively or coincidentally attracted while away from his place of residence for any primary purpose other than shopping."37 Thus, a suscipient location does not

36 Ibid., p. 45.
37 Ibid., p. 45.
generate business; it just takes or receives it. The key reason for stores existing in these locations is service. They service people away from home whose primary purpose is other than shopping. An example of a suscipient location could be a news-stand in an airport or a coffee shop in a large building. No location will be a generative or a suscipient site for all people. Thus, we will find that nearly all locations will have some generative trade and some suscipient trade.

Nelson's Principles of Store Location

Whatever the motive for choosing a particular location, Nelson contends that there are eight principles which must be considered. The first is the adequacy of the trading area potential. This can be accomplished by finding out the number of people in the trading area and how much they have to spend. After determining the total sales volume present, the firm next has to figure out what share of the market their store will capture.

Second, the firm needs to measure the accessibility of the site to the trading area. The potential business from generative, shared and suscipient trade must be estimated. Generative business which can be garnered through heavy promotional expenditures, wide product offerings, lower prices,

---

38Ibid., p. 45. 
39Ibid., p. 52. 
40Ibid., p. 52.
convenient locations or any combination of the firm's marketing mix must be figured for each segment of the trading area. Consumer knowledge of unique characteristics of the firm's marketing mix is often of a generative nature. Stores that depend heavily upon generation of their business need to be in the most accessible site within reasonable cost considerations. Shared business is gained by a store as the result of the generative strength of nearby stores. Many of the small stores in large shopping malls depend heavily upon shared business, as they do not individually have enough strength (or uniqueness in their marketing mix) to generate a sufficient volume of business. Suscipient business is that not generated by the store itself or its neighbors. Stores depending on suscipient trade often do little or nothing to generate business. Thus, their existence depends heavily upon the correct location.\[^{41}\]

The third principle to be observed is that the area's growth potential should be promising. There has been a recent trend toward failure of stores caused by shifts in population that have changed once prosperous trade areas into lums.\[^{42}\]

The idea of interception of business is the fourth principle. Consumers tend to habitually go to the same source of goods. In following the principle of interruption, the firm would position their store between the mass of consumers and

\[^{41}\text{Ibid.}, \ p. \ 53.\]  
\[^{42}\text{Ibid.}, \ p. \ 45.\]
and their typical shopping facilities. This principle assumes that it is much easier to stop the consumer on their route to their normal stores than to pull them away from that route to a new location. 43

The fifth principle is that of cumulative attraction, which loosely means that two or more stores have more drawing power than one store. One type of cumulative attraction is with similar units and the other type is with two complimentary units. Stores that are complimentary carry lines of products that depend upon each other for completeness. An example would be a clothing store and a shoe store; they carry different products, but both lines are needed to fully accessorize the consumer. 44

Compatibility among types of stores is the next principle to be considered. Consumers do not expect to see discount stores next to the nicer department stores; the images of the two strictly are not compatible. 45

The seventh principle is minimization of competitive hazards. "The principle of minimizing competitive hazard should lead the prospective retailer, other things equal, to: (1) select a location near which there are as few such competitive sites as possible; (2) consider the feasibility of controlling or earmarking the use of such sites for non-competitive purposes; and (3) select a location in which the

43 Ibid., p. 54. 44 Ibid., p. 54. 45 Ibid., p. 54.
competitive sites are in nonintercepting positions."\textsuperscript{46}

The final principle is the consideration of the cost of the site relative to its expected productivity. The typography of the site, its load-bearing qualities, visibility, and other off-site factors need to be viewed in depth. This also includes the building, its appearance, and the efficiency of use of the building. The lack of labor should also be pointed out here as should the advertising media and its cost-productivity ratio.\textsuperscript{47}

The Rule of Retail Compatibility

Some businesses operate well together while others do not. "The measure of their compatibility lies in the answer to the questions: 'Does it harm Business B, or does it apparently have no effect on Business B?'"\textsuperscript{48} It is entirely possible that the answers to the above may be the most important locational factor that the store will consider. The developers of shopping centers place emphasis on compatibility when considering placement of the various stores. The retailer should give it at least as much thought as the developers.\textsuperscript{49}

The degree of compatibility between two stores is very high, when their sales are greater with adjacent locations than with separated sites. High compatibility may be created by stores selling complementary or competing items: in either

\textsuperscript{46}Ibid., p. 46.  \textsuperscript{47}Ibid., p. 55.  \\
\textsuperscript{48}Ibid., p. 66.  \textsuperscript{49}Ibid., p. 66.
case the cumulative attraction of the two stores is enhanced by their close proximity. Thus, the measure of compatibility is relative to the amount of interchange of customers among the business.\(^50\)

Therefore, the rule of compatibility among retail stores may be stated as follows:

Two compatible businesses located in close proximity will show an increase in business volume directly proportionate to the incidence of total customer interchange between them, inversely proportionate to the ratio of the business volume of the larger store to that of the smaller store, and directly proportionate to the sum of the ratios of purposeful purchasing to total purchasing in each of the two stores.\(^51\)

In general two stores are considered to be highly compatible if from ten to twenty percent of the customers visit both stores (interchange) while a five to ten percent interchange between two stores means that the two are moderately compatible. While two stores that share one to five percent are considered to be slightly compatible. Two stores are incompatible if there is negligible customer and interchange and may be deleterious if one harms the other.\(^52\)

**Major Considerations in Site Evaluation**

Most evaluation of store sites begins with an estimate of the sales that will be generated by each location under consideration. The evaluation should consider the following:

\(^{50}\text{Ibid.}, \ p. \ 66.\)
\(^{51}\text{Ibid.}, \ p. \ 66.\)
\(^{52}\text{Applebaum, et al.}, \ p. \ 25.\)
demographic factors, the competitive situation, the economic stability of the area, and the accessibility of the site.  

Accessibility of Site. For many types of businesses, such as clothing stores, the ease with which consumers can reach a site and the traffic patterns around the site are the most important factors in selecting a location. The fact that a site is easy to reach implies short distances and ease of driving.

The road surface conditions of the roads that feed the site is an important variable. Feeder roads that are unpaved, have pot-holes, or inclined steeply so as to create winter driving hazards give poor accessibility to the site and adversely affects the sales of the stores.

The accessibility of the site may also be affected by the speed of the passing traffic. Speed is a double culprit, as it hinders the visibility of the site and makes it difficult for customers to exit from the road. Visibility is reduced because the consumers are physically in front of the site for a shorter period of time. Consequently, their chance to see the store is reduced.


55Cohen and Applebaum, p. 1.

56Applebaum, p. 83.
Another element that bothers accessibility is the distance from the consumers.\textsuperscript{57} Distance can not be separated from the time it takes the consumer to reach the site. It is possible that a site could be located next to one of the more heavily used corners in a town and still be poorly located because the site would take a number of minutes to reach. Distance and time are both functions of the size of the trade area and the population.

The amount of traffic flow into a site is another factor that can affect accessibility. Too little traffic flow will make it more difficult for consumers to cross major lines in traffic. Too much traffic can create lines which will discourage consumers from shopping. However, larger volumes of traffic is the least of the two evils, as a continuous line of traffic making turns makes it easier for the drivers to turn.\textsuperscript{58}

The peaks and valleys of the traffic flow need also to be considered. Normally, the "rush hour" traffic does not coincide with the prime shopping hours. Most shopping is done during the late mornings, mid-afternoons, and early evenings. Sometimes the release of factory shifts and schools can adversely affect shopping due to the increased congestion around the site. Shopping may also be reduced due to traffic congestion caused by special events such as ball games, large meetings and other special events. Some stores are highly

\textsuperscript{57}\textit{Ibid.}, p. 83. \hspace{1cm} \textsuperscript{58}\textit{Ibid.}, p. 85.
dependent upon seasonal business which must be taken into account during traffic flow studies.\textsuperscript{59}

As mentioned earlier, visibility affects the accessibility of a site. Retail sites that are located around a curve in the road or in the middle of numerous established stores are hard to see until the traffic is directly in front of the site. Low lying sites subject to smoke or fog may be hard to see.\textsuperscript{60} Visibility may also be decreased by regulations against signs or the inability to find good locations for the signs.

Accessibility must be considered in much the same manner for pedestrians. It is necessary to identify potential hazards for pedestrians. Positive factors toward the accessibility of the site are sidewalks and ease of entry. Sites that require climbing steps or other forms of extra physical exertion on the part of the consumer are not accessible and pedestrian traffic will be smaller.\textsuperscript{61} Pedestrian traffic will be negatively affected by the following six factors:

1) Dead spots which cause shoppers to lose interest in going farther. An example of a dead spot could be an empty building.

2) Driveways, curbs, and other physical breaks in the sidewalks.

3) Either vehicular or pedestrian cross traffic.

\textsuperscript{59}Ibid., p. 85.
\textsuperscript{60}Ibid., p. 85.
\textsuperscript{61}Pickle and Abrahamson, p. 143.
4) Areas that are hazardous, noisy, smelly, unsightly or other qualities that pedestrians might find to be unappealing.

5) Businesses that create traffic such as trucks, public vehicles, private automobiles or nonshopping pedestrians, all which tend to generate congestion.

6) Lengthly average parking time of customers of the business.

Parking is a final variable that affects accessibility. Consumers demand that adequate parking be close to the store. There are no uniform methods for determining the number of parking spaces that will be adequate for a site. One of the determining characteristics is the type of store. One study found that a supermarket needs 12 parking spaces for every 1000 square feet, while small retailers need no more than 5 spaces per 1000 square feet. Basic elements in the determination of parking adequacy include the following:

1) The type of neighborhood in which the store is situated.
2) The frequency and length of stay of shoppers.
3) Seasonal, weekly, and daily fluctuations of business.

Demographic Factors. These factors are extremely important in evaluating sites. Demographic variables form one

---

63 Pickle and Abrahamson, p. 145.
of the three major categories of factors on which all markets are segmented. Site selection is but one of the elements that fit into the marketing mix used by retail stores to attract the market segments that they have chosen. Among the demographic variables of importance are: the total population in the area; the average ages of the head of the household and children; the average income level of a family; the average educational level of the head of the household; occupations; the proportion of races within the area; the marital status of people within the area; the region's cost and type of housing; the predominant religions in the area; the social class of the area; and family size.

"People make markets." Shortly after World War II, the population growth was the most important factor for increasing demand in such areas as food, clothing, housing, and education. Even today, many firms are looking toward the growth of population as a stimulus for future marketing progress. Relative to a site the absolute number of people is a strong indicator of the ability of the area to support the store in question.

---


68 Ibid., p. 103.
Of course, firms don't want to overlook the growth potential of regions as an extra determinant in the situation of a new store.

The size of the family might indicate the responsiveness of the market to stores with juvenile goods or other items coveted by youth. Knowledge about the age of the market and the number of citizens in each age category can give the store additional information about the number of people who would be potential customers based on the characteristics of the market that the store is trying to attract. For example, stores selling conservative mens clothing will want to locate in areas with significant numbers of men over 30 years of age.

The income levels of the region are an indicator of how much money the people have to spend and what items they can afford to buy. Educational levels, occupation, and type of housing are highly correlated with income levels. As an example, most communities that have a higher than average income distribution will also have higher than average levels of education, more expensive homes, and a larger number of professional and semiprofessional people. Communities such as these are likely to have a greater demand for health facilities, stock brokers, high-fashion clothing stores and other similar facilities.

69Ibid., p. 109.
In short, the demographics of the region are descriptive of the degree of fit that the region has with the type of consumers the store is trying to attract.

**Competitive Situation.** "Competition is the aggregate of all retail facilities which together share the total market potential." 70 Thus the competition of a retailer is all stores selling similar types of goods. 71 Sometimes a firm will want to analyze other firms in direct competition and those that are indirect competitors. For example, a chain of pizza parlors may be interested in knowing the number of other pizza parlors in the area (direct competition) and the number of other restaurants (indirect competition).

Regarding competition, the firm will need to note the number of competitors in the area with respect to the total market potential. For example, a firm considering a site that has ten competitors each with $300,000 annual sales in an area with an estimated total sales potential of $3 million, might decide not to open a store in the area due to the apparent saturated condition of the area. Of course, decisions such as this are made in light of the firm's objectives and the possible ramifications of competitive retaliation.

---

70 Applebaum, et al., p. 88

71 Ibid., p. 88.
Economic Stability. It is possible to evaluate economic stability, but not predict it. By classifying the area by functional type such as residential, manufacturing, farming, recreational, educational, defense or diversified, it is possible to figure the value added by each of these categories obviously indicates that the region relies upon these categories. Should this category stop being productive for some reason, the entire economic situation of the area would have negative ramifications on the sales of stores. For example, stores in some communities in West Virginia and Kentucky have suffered through the recent coal strike due to the immense reduction of the income of many of their customers.

"Even completely accurate assessments of the economic prospects of an area cannot always indicate the desirability of such areas for store sites." However, poor economic prospects should not be the only determining factor. The decision to locate a store needs to be weighed in light of consumer buying habits and the number of competitors. Economic depression has far greater affect on the sales of material goods such as sold by department stores, car agencies, or real estate firms than on food stores or other outlets selling goods that are rapidly consumed.

\(^{72}\text{Ibid.}, p. 90.\)
\(^{73}\text{Ibid.}, p. 91.\)
\(^{74}\text{Ibid.}, p. 91.\)
The Ratio of Advertising to Rent
for Various Retail Types

"Rent and advertising are reciprocals." Thus, the
amount of rent that a retailer should consider paying depends
on how much money he can spend to attract customers and the
amount of advertising he does. Also, the more money allotted
for rent, the less money that is left for advertising. Of
course, the poorer the location the more advertising that is
needed to attract customers.

Discount Houses. Discount houses operate on a very low
markup and depend upon high turnover that is generated by high
promotional expenditures to remain in business. Better than
90 percent of their business is generated by their promotions,
which means that consumers are making special trips to reach
the store. Consequently, the site of the store needs only to
be reasonably accessible and the rent need not be too high.

Department Stores. Most department stores are heavy
advertisers. However, a smaller portion of their advertising
is price oriented than the discounters. Consequently, the
advertising of department stores tends to generate less bus-
iness than discounter advertising. Thus, department stores
need better sites than do discounters.

76 Ibid., p. 46.
77 Ibid., p. 46.
Apparel Stores. "Apparel stores typically have a lower proportion of self-generated business than department stores and, therefore, have greater need for a prime location (particularly with respect to other generators)."\textsuperscript{78} Some of the very high fashion stores are exceptions to this rule because their product mix is generative due to its uniqueness.\textsuperscript{79}

Variety Stores. Variety stores have one of the lowest proportions of self-generated business of all major types of stores. Thus, the specific location is very important to the variety store in order to help generate more business.\textsuperscript{80}

Convenience Stores. Convenience stores such as the 7-11 chain depend almost entirely on excellent sites. Consequently, these stores do little, if any advertising.

Supermarkets. The food market business rivals the discount stores in the amount of self-generated business. Supermarkets are heavy advertisers and draw consumers to purchase the promoted items. Thus, supermarkets need a site that is reasonably accessible due to their strong pulling power and market penetration within a very small radius.\textsuperscript{81}

\textsuperscript{78}Ibid., p. 48.  
\textsuperscript{79}Ibid., p. 48.  
\textsuperscript{80}Ibid., p. 48.  
\textsuperscript{81}Ibid., p. 48-49.
Drugstores, Bakeries, and Liquor Stores. This unlikely grouping of stores all have a greater reliance upon neighborhood-generated and suscipient business than other types of stores. Therefore, they need an identifiable location that is near generative stores. 82

Service Stores. Service such as barber shops, repair shops, and beauty shops normally have a large portion of self-generated business. Consequently, location is not too important for service stores. Typical locations are in lower rent areas, even though these stores are not heavy advertisers. 83

Restaurants. Restaurants tend to be quite variable relative to locational needs. Some advertise heavily and treat locational decisions much as do the discount stores. While other restaurants depend upon business by locating close to generative types of businesses. The latter could be typified by hot dog stands in ball parks. 84

As the above discussion indicates, locational requirements differ according to the type of store. Due to varied marketing mixes adopted by competitors, stores of the same type often vary as to locational requirements. Therefore,

82 Ibid., p. 49.
83 Ibid., p. 49.
84 Ibid., p. 49.
the above guidelines for rent to advertising ratios are simply
thoughts to be considered during site analysis.

Strategy in Store Location

Chains competing in a mature industry can't approach
store location in a haphazard manner. Some strategic thought
needs to be applied to the selection of new sites as well as
toward evaluation of old sites.

Setting Marketing Goals. The return on investment of
chains is influenced by numerous variables, some are con-
trollable while others are uncontrollable. The goal of ob-
taining larger market shares will be rewarded in the long
run by diminishing returns on investment. Thus, a chain may
not want to expand too much due to this threat. Yet, there
are several reasons why corporations want to add store facili-
ties despite the probable reduction in average sales per
square foot (which is often the measure of performance most
often used to evaluate stores). Additional stores will in-
crease the sales volume of a chain and enable them to

(1) Purchase items on more favorable terms
due to the larger volume.

(2) More adequately use warehouses.

(3) Provide more outlets for private label
merchandise.

(4) Spread the risks of merchandising.

(5) Reduce the average cost of advertising
and management.  

85Applebaum, et al., p. 30.
Objectives that are aimed at particular competitors will produce strategies aimed at locating stores close to the competitor's stores or trying to keep the competition out of lucrative areas. Objectives that include attraction of new market segments will produce new stores accessible to these segments. In short, the determination of the number of new stores to be added and their locations will be greatly influenced by the corporate objectives.

Disturbing Forces. There are several uncontrollable factors that disturb the balance between the available store facilities and their efficient use. The factors are

(1) Population changes.
(2) Purchasing power changes.
(3) Changes in what customers want or expect in services.
(4) Merchandise assortment changes.
(5) Shopping and buying habit changes.
(6) Changes in the facilities, decor, and atmosphere of retail stores.
(7) Changes in opportunities for investment.86

It is possible that the disturbance may be either short-run or long-run. Understoring, in the short-run may be corrected by simply building more. While short-run overstoring could be handled by the elimination of unprofitable stores or waiting for the population to increase. Imbalances due to

86Ibid., p. 31.
overstoring in the long-run can be corrected only through changes and innovations in store facilities, services, and merchandising. 87

"The great depression of the thirties precipitated, because of a reduced demand, a sudden overstoring and a long-run imbalance between the availability of and need for the then-existing store facilities. Out of this depression emerged scrambled merchandising, the supermarket, and the discount house." 88 Of course, economic depressions are not the only causes of long-run imbalance. Sometimes prosperity can be the culprit, too. Overly zealous entrepreneurs can expand too quickly during times of prosperity. Periods of imbalance are long, while the point of perfect balance is extremely short. 89

A single large chain may create an industry imbalance by overbuilding and ruin others. Weak chains may ruin themselves by miscalculating. Dispersion in stores can be an advantage for large chains and can afford to take additional risks. Large chains can also acquire smaller local chains, but not at the risk of weakening the performance and personality of the absorbed stores. Mergers are considered to be an alternative that minimizes the risk of overstoring due to

87 Ibid., p. 31.
88 Ibid., p. 31.
89 Ibid., p. 31.
the purchase of existing facilities. However, the effectiveness of the merger could be negative if the various outlets can't be efficiently used.  

Guideposts. Sound growth needs a solid store location strategy consisting of the following:

(1) A determined decision as to what kind of retailer a company wants to be, and the image or "personality" it seeks to represent.

(2) A thought-out plan of how the company wants to grow.

(3) A sound knowledge of the market areas where the company wants to do business.

(4) A careful, detailed study of alternative store location opportunities, and the risks as well as rewards which each presents.

(5) Good timing and flexibility to adjust to changing conditions.  

---

90 Ibid., p. 31

91 Ibid., p. 32.
A noted retail site researcher, David MacKay, feels that the three most commonly used methods of site determination fall under three categories: gravity models, analog methods, and map transformation methods. Philip Kotler adds a fourth to this list, the checklist method. For the purpose of this paper two other categories will also be discussed, behavioral models and miscellaneous methods.

Gravitational Models

Reilly's Law of Retail Gravitation. The basis of this model and all gravitational models stem from W. J. Reilly's law of retail gravitation which states that in the proximity of the breakeven point between two centers, the amount of trade to each center will be proportional to the population of the centre and inversely proportional to the square of the distance from that centre. Hence, the breakeven point will occur where

---


\[ \frac{W_A}{D^2A} = \frac{W_B}{D^2B} \]

where \( W \) is the population centre, and \( D \) the distance from the centre, the two centres being \( A \) and \( B \).

In 1949, P. D. Converse developed from Reilly's work a formula to determine where the breakeven points should be on a straight line between two centers. When dealing with behavior of people, Converse's formula may be written as

\[ I_{ij} = K W_j F(D_{ij}) \]

where \( D \) is the distance between the centers and \( I_{ij} \) is the attraction that center \( j \) has on individual \( i \). The assignment of the variable \( W \) has taken many forms, varying between the following: population, purchasing power, retail turnover, number of shops, number of telephones, or other measures of the size of the center. Sometimes, the cost of the trip is used in place of distance.

Converse's formula will give an approximate measure of how trade divides between two towns without making a survey. Of course, by using a survey, the town's merchants may discover "how well they are doing" in competition with other towns. This model is normally used for study of larger

---


4. Ibid., p. 37.

regional areas, but it may be applied to individual stores. However, the latter approach may be useless when there are many stores grouped within a close proximity. Consequently, when used in choosing the site for a store, Converse's formula is most appropriately in rural areas where population and store density tends to be lower.\(^6\)

**Central Place Theory.** Reilly's work was followed by the Central Place Theory which was developed by geographer Walter Christaller and economist August Losch. The theory accounts for the regularity of the functions of marketing as performed by the store within the areas served. It can determine "the spatial arrangement of stores required for optimal distribution of a single good to a dispersed population."\(^7\)

Christaller explained that since the cost of travel varies with the distance, it would be possible to assume that a "demand cone" exists for the store. As the distance from the store increased, the quantity demanded by customers would decrease due to rising costs of transportation and time. He further hypothesized that it would be possible to calculate the demand for a store within its "demand cone" and developed the following equation:

\[
D_i = S \int_0^\infty \int_0^\infty \left[ f(p_i + mt) \right] d\theta
\]

\(^6\)Dixon, p. 33.

where \( p + m t \) is a function of price and transportation costs and \( S \) is population density. It is possible to test a variety of prices with this formula and obtain correspondingly different "demand cones," as one would expect.\(^8\)

In more simple terms, the Central Place Theory examines two elements: first, the smallest area that will support a particular store and second, the maximum distance consumers are willing to travel. Like Reilly's Law, the Central Place Theory is more relevant for rural areas.

**Maximum Likelihood Model.** This approach is associated with the name of David Huff, who followed the retail gravitational ideas of Reilly and Converse by developing a series of mathematical models designed to evaluate a proposed site and to select an optimum site from a number of alternatives. He incorporated the basic gravitational notion that the attraction exerted on a customer in area \( i \) by a retail store at location \( j \) is proportional to the size of the retail store and inversely proportional to the distance that the consumer lives from the retail store.\(^9\) Huff's reasoning is as follows. Suppose two or more stores (or shopping centers) are pretty much alike except for the distance from the consumer and the size. A larger store means larger product assortment, more

\(^8\)Ibid., p. 61.

variety and hence a greater utility for the consumer. Distance represents a cost in both traveling time and expense (disutility) to the consumer. Assuming that the consumer desires to be an efficient shopper, he will be attracted to a particular store in proportion of the utility to disutility.\(^\text{10}\) Huff expressed the probability of consumer attraction by the following:

\[
P_{ij} = \frac{S_j / D_{ij}^\lambda}{\sum_{j=1}^{\infty} S_j D_{ij}^\lambda}
\]

where \(P_{ij}\) = the probability of a consumer in area \(i\) shopping at location \(j\)

\(S_j\) = the size of the store at location \(j\)

\(D_{ij}\) = the distance between area \(i\) and location \(j\)

\(\lambda\) = an estimated sensitivity parameter relating kinds of shopping trips and distances

\(n\) = the number of retail locations\(^\text{11}\)

Thus, the formula states that the probability of a customer at point \(i\) going to store \(j\) is equal to the utility derived from store \(j\) divided by the cost of traveling to store \(j\). This ratio is further divided by a similar ratio of utility and traveling costs of stores number one through \(N\).

As an illustration, suppose \(n\) is three and the consumer is able to shop at two retail stores, the first consisting of

\(^{10}\)Kotler, p. 316. \(^{11}\)Ibid., p. 317.
10,000 square feet at a distance of five miles and the second consisting of 1,000 square feet at a distance of one mile. According to Huff's formula, the consumer's shopping at each store are .29 and .71, respectively, as shown by the following calculations:\textsuperscript{12}

\[ p_{i1} = \frac{10000/5^2}{10000/5^2 + 1000/1^2} = .29 \]

\[ p_{i2} = \frac{1000/1^2}{10000/5^2 + 1000/1^2} = .71 \]

These probabilities are subject to the following interpretations which in practice lead to essentially the same result:

1. Each customer in location \( i \) will shop 29 percent of the time at store two.

2. Twenty-nine percent of the customers in area \( i \) will shop at store one while 71 percent will shop at store two.\textsuperscript{13}

A key to the model has to be the value assigned to \( \lambda \). Originally researchers used the squared distance as the negative factor (\( \lambda = 2 \)) because this was supported by the early work of Reilly and Converse. Later work proved that varied among stores and should be independently fitted.

\textsuperscript{12} Ibid., p. 317.

\textsuperscript{13} Ibid., p. 317.
Huff used an iterative method of estimating $\lambda$. Initially, $\lambda$ is assigned an arbitrary number greater than one. Then Huff's equation is used to figure the pij's for each consumer area. Then the actual number of consumers in each area is found from census data and used to estimate the expected number of consumers from each area who will shop at store $j$ using the following formula:

$$E_{ij} = pij \cdot Ci$$

where $E_{ij} =$ the expected number of consumers in area $i$ that shop at store $j$

$Ci =$ the total number of customers in area $i$\textsuperscript{14}

These expected values are compared with the actual relative frequencies obtained from survey data and a coefficient of correlation is obtained. The object is to obtain the highest correlation coefficient possible by successively adjusting the value of $\lambda$.\textsuperscript{15}

Assuming that the geographical area has been reduced to a number of specified sites, Huff's model would go through the following steps in choosing the optimum site:

1. Estimation of sales for specified scales of operation at an initial location.

2. Estimation of the costs associated with each scale of operation stated above at the location.

\textsuperscript{14} Ibid., p. 317.

\textsuperscript{15} Ibid., p. 317.
3. Calculation of net profit at each scale of operation (step 1 minus step 2 for corresponding figures).

4. Pick out the best scale of operation for that particular site.

5. Repeat the first four steps for each of the potential locations.

6. Select the optimum location by picking the site with the highest profit.¹⁶

Huff's model is best used to predict drawing power of a shopping mall or similar arrangement of stores and is unreliable in the prediction of the sales for a single store. Another weakness is the inability of the model to account for natural barriers and distortions of the terrain. Also, the different images, accessibilities and prices of stores are viewed differently by customers.

It has been suggested that images and prices associated with different stores could be included in the formula.¹⁷ In fact, one study tested the viability of adding image inputs to the model.¹⁸ This study concluded that driving time was more important as a predicting variable than the size of the store or the associated image. While image was more important


¹⁷Kotler, p. 319.

than size. Another weakness of the model is that if many stores or sites are being considered then the calculations can become quite complex.

Louis P. Bucklin, with the maximum likelihood model in mind, used discriminant analysis on survey data to determine and evaluate hypotheses about the role of mass in shopping (mass being defined as the component which defines the assortment of goods and services within a store). Bucklin suggests that the model could be improved by considering the motives of the shopper. The consumer may be undertaking the shopping venture with three possible perspectives about the shopping to be done. The consumer may be in "full search" where he is perfectly willing to make comparisons between stores on products, as the value derived from their searches is more than the perceived cost of the searches. The second alternative is "directed search." Here, the consumer is more likely to patronize the store where he perceives he has the best chance to obtain what he desires. Comparisons between stores is perceived to have low returns. Finally, the consumer may undertake "casual search" where product selection and comparisons between stores are not important. Bucklin believes that the measurement for these variables will greatly improve the predictive capacity of the model, but he offers no suggestions on how to incorporate these measures.  

The Klavac and Little Gravitational Model. This model attempts a similar format as Huff's model except with the use of additional factors in an effort to predict sales for potential sites for automobile dealerships. In this model, the attraction of an auto dealer is a function of the customer's make preference, the dealer's local image which is independent of the make, and the distance that the customer is from the dealer. Specifically, the probability that dealer \( j \) will sell to consumer \( i \) is given by

\[
P_{ij} = \frac{\lim_{m}^i \cdot L_{m}^j \cdot e^{-bjD_{ij}}}{\sum_{k=1}^{\lim_{m}^i} \cdot e^{-bD_{ik}}}\]

where
- \( P_{ij} \) = the probability of dealer \( j \) selling to consumer \( i \)
- \( L_{m}^i \) = the make preference of consumer \( i \) for auto maker \( m \)
- \( L_{j}^i \) = dealer \( j \)'s local image independent of make
- \( K \) = a subscript for the dealer
- \( e^{-bjD_{ij}} \) = the effect of distance on probability, where \( D_{ij} \) is the distance between consumer and dealer, \( b \) is the amount sales decrease with expanding distances, and \( e \) is the constant 2.71.

The key difference between Huff's formula and the Klavac and Little model is that the latter expresses the effect of distance purchase probability through an exponential decay factor.\(^{20}\)

\(^{20}\)Kotler, p. 320.
Analog Method

This method depends upon a combination of subjective evaluation and quantified experience. Obviously, the more and better the quantified data, the less dependence on subjectivity and the greater the chance the model will be improved. Quantified experience includes data obtained from store sales in comparison to consumer behavior, store characteristics, market characteristics, and other known, pertinent factors.\(^{21}\) The quantification of this data becomes an analog, hence the name of this technique. Potential store sites are then contrasted to similar segments of the company analog, then sales potential is projected based on the analogs as a measurement standard.\(^{22}\)

Two sites are never alike nor are the market factors affecting the site alike; thus, subjective judgment must be applied in spots. Due to changing market factors, consumer behavior and store characteristics, the analogs quickly become outdated, which is another source for some subjectivity in the decision on the applicability of the analog.\(^{23}\)

Analogs are derived from market analysis, penetration studies, and consumer spotting techniques. Since a firm's image generally varies over different parts of the country, a highly decentralized firm with many branches is in a better position to obtain more accurate analogs. The analog of one firm

\(^{21}\)Applebaum, p. 232.  \(^{22}\)Ibid., p. 232.  
\(^{23}\)Ibid., p. 233.
will probably not fit another firm within the same industry.

The computational procedure for this method calls for developing zones around the proposed site and estimating the sales that the proposed store will draw from each zone. Estimates of these sales are based on the company's analog which is nothing more than the customer attraction rate of similar stores in the company's chain.

Table I presents one of William Applebaum's illustrations for a real but disguised location (Applebaum's is a leading proponent of the Analog Method).

<table>
<thead>
<tr>
<th>Zone (miles)</th>
<th>2 Population in zone</th>
<th>3 Estimated per capita sales</th>
<th>4 Estimated weekly sales</th>
<th>5 Computed drawing power, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>.00 - .25</td>
<td>4,700</td>
<td>$2.00</td>
<td>$9,400</td>
<td>28</td>
</tr>
<tr>
<td>.25 - .50</td>
<td>12,900</td>
<td>.76</td>
<td>9,804</td>
<td>29</td>
</tr>
<tr>
<td>.50 - .75</td>
<td>23,000</td>
<td>.22</td>
<td>5,060</td>
<td>15</td>
</tr>
<tr>
<td>.75 - 1.00</td>
<td>36,300</td>
<td>.12</td>
<td>4,356</td>
<td>13</td>
</tr>
<tr>
<td>beyond</td>
<td>-</td>
<td></td>
<td>5,051</td>
<td>15</td>
</tr>
</tbody>
</table>

In this example the area surrounding the proposed site was divided into quartermile zones (column 1). An estimate of the population in each zone is derived by using census tract data (column 2). An analog of the drawing power of similar stores is used to estimate the expected weekly sales for every person in each zone (column 3). The estimated weekly sales from each zone, column 4, is found by multiplying zone population by the per capita expenditures. The total of column 4 represents an estimate of the store's weekly sales volume. Column 5, the computed drawing power, represents a check figure that can be compared against company analogs. The greater the degree of similarity between column 5 and the analogs, the greater the confidence in the sales estimate.  

The Analog Method offers the researcher more opportunity to express his own feelings without being tied to the rigidity maintained by the previous formulas. This same flexibility, while a strength in one aspect, is also a weakness, as the method is only as strong as the management or researcher's opinions. Additionally, the method will increase in subjectivity with more locations being considered. The analog method depends heavily upon the historical aspect of the firm and extrapolation of past trends, which are often not accurate due to changing market trends that affect the effectiveness of the company's marketing mix.

25Kotler, p. 315.
Map Transformation Method

Most often the effect of population densities has been of primary importance in the selection of store sites. This method takes into account the distribution of money that is allocated for spending on retail goods. It assumes the location of retail stores should be based upon the distribution of income. Expenditures on consumption are spread unevenly over space. Following the assumption that there are limits on the distance that consumers will travel for goods, it is possible to locate areas within which it would be feasible to open a retail store. This method is based on the following premises:

1. Market areas exist for retail stores.
2. There is a minimum overlap or no overlap at all of market areas.
3. Consumption expenditure location has a direct bearing on the location of retail stores.
4. Travel time or cost is the same for any unit distance from place to place.
5. Rent or any other economic factor except consumption expenditures for goods supplied in retail store has no bearing on the general location of the firm.

The first step in this method is to determine the boundaries for the study. The city limits of a smaller town may

---


27 Ibid., p. 16.
be natural boundaries, while for other areas actual distance in miles from a predetermined center might establish the boundaries. Next, the researcher through use of census tracts and other available data, should determine the percentage of income available for expenditure on goods in question. The third step is the distortion of the map to reflect the estimated consumption expenditure. In this stage, each cell is made rectangular whenever possible and the size of the cell should reflect the amount of money available. The rectangles are then segmented into hexagons and the center of each hexagon is transferred to the original map. The number of hexagons chosen normally approximates the number of retail stores already in the area. The center of each hexagon now represents the theoretical location of a store. The theoretical distribution is then compared to the real distribution of the stores, if an area of the map tested is shown to be under-populated according to the theoretical distribution, then a new store might be in order for that area.  

This method is not without limitations. First, it does not consider the cost of rent. Secondly, the method appears to be most valid for grocery stores and some of the assumptions would not work for other types of stores. Finally, it doesn't determine the number of stores needed, only the areas that do not have many stores in relation to the income of the regions.

---

\(^{28}\)Ibid., p. 16-19.  
\(^{29}\)Ibid., p. 21.
Checklist Method

This technique is one of the more simplistic evaluation methods for sites. Yet, it is probably the most frequently used technique due to the ease of evaluation. Because of the rapid growth of claims the site selection decision had to become more formalized than driving a car around in prospective site areas. Consequently, real estate advisors were hired to conduct systematic surveys of population, income, expenditure patterns, local zoning, highway plans, and so on. The real estate people would observe traffic patterns, quiz shoppers on their place of residence, study consumer travel patterns, and look at other pertinent variables. Due to the large number of important factors, checklists were developed that covered all of the variables that tended to affect the potential success of the site. It could be possible to replace the adjective rating system (shown in Appendix 1) with numeric values. Then individual weights could be assigned to each factor that would indicate the importance of each. The following would then represent a formula used to evaluate the value of each site.

\[ V_i = \sum_{j=1}^{n} W_j F_{ij} \]

where \( V_i \) = the value of site \( i \)
\( W_j \) = the weight of factor \( j \)
\( F_{ij} \) = the rating of site \( i \) on factor \( j \)
\( n \) = the total number of factors to be considered

Obviously, the weights associated with each factor tend to be subjective. Thus, the value of the site often has a subjective flavor. Yet, this quantification effort is better than the automobile tour that was used previously.\(^{31}\) Limitations of this method include the following: the subjectivity associated with the importance of each factor considered; the lack of reliance upon what surrounding elements have previously proved successful for the firm (analogs); and data collection techniques that could be less than objective.

Behavioral Models

MacKay's Microanalytic Approach. This model accounts for multistop shopping trips and is based on a framework that portrays consumers as going through a sequence of three steps during the process of store selection. The first step is the decision to either go or not go shopping. Stage two is a decision on how many stops to make, and the last step is the decision on which stores to visit on each stop.\(^{32}\)

\(^{31}\)Kotler, p. 314. \(^{32}\)MacKay, p. 134.
The final step is divided into two phases: trip generation and trip structure. Trip generation is based on the variety of stores to be visited during the trip. This phase is affected by demographic characteristics of the shopper, the consumer's image of the store, and variables peculiar to the shopping trip (such as mode of transportation, time of day, distance from home, establishments visited previously, etc.). The calculation is made by figuring "the probability of choosing each stop are computed as functions of the discriminant scores, and then Monte Carlo sampling is used to select the event chosen."^33

Trip structure requires that the types of establishments be replaced by stores considered on the trip. The use of "movement heuristics" aids in accomplishing the above, as the heuristics provide "rules" for the customers to follow in selecting stores.\(^{34} \) The heuristics considered for the model are the following:

"Single-stop distance minimization": the shops at which to stop are chosen on the basis of close proximity to home.

"Modified sequential distance minimization": small trips are minimized on distance, but on larger trips a series of distance minimization techniques are adopted as distance increases.

"Discriminant heuristics": the same as the previous heuristic except when a stop is about to be made. At the time of the stop,
discrimination input are used to pick the store. Discrimination inputs include demographic characteristics of the shopper, store image, and variables peculiar to that shopping trip.\textsuperscript{35}

The simulation can be used to determine the projected sales of a new location by inserting the spatial coordinate of the projected site. Efforts must be made on trying to estimate the prospective image of the new store.\textsuperscript{36}

The model was validated by MacKay through the use of "travel diaries" kept by some 200 consumers in the Chicago area. It was found that the simulation produced travel patterns for consumers that were not significantly different from those of the sampled population. The projected store sales were also not statistically significant from the average sales of each outlet of the respective chains.\textsuperscript{37}

The Questionnaire Approach. Through the use of a questionnaire the researcher estimates the number of customers the firm would have for its product and their preferences on location. The questionnaire tries to identify whether the consumer has made a purchase at one of the firm's stores within the last twelve months. This separation of the respondents into customer and non-customer segments allows statistical comparisons of the characteristics of the two groups.\textsuperscript{38}

\textsuperscript{35}Ibid., p. 135. \textsuperscript{36}Ibid., p. 137.
\textsuperscript{37}Ibid., p. 137.
Characteristics that are statistically significant become part of the "customer profile."

Should research be needed for the location of a store for a new firm, it would be necessary to test alternative attitudes toward product lines, advertising, and physical image of the store. Generally, this type of assessment is less valid than studies dealing with existing firms.\textsuperscript{39}

After the relevant segment of the population is identified, concentration must be placed on the shopping behavior and profile of these people. The following questions must be answered:

1. What is the nature and strength of each location's attractions?
2. What are the consumers' perceptions of the firm's product and how intense is their desire for it?
3. How would the firm's association with each potential new agglomeration (of stores) alter the consumer's perception of the firm's product?\textsuperscript{40}

The next major step is the forecasting of the number of customers at each alternative location. In order to accomplish this objective it will be necessary to determine the geographic region where the people who will shop at a particular region will live. Some data should be gathered on employers and their location of consumers. Consideration should be made for correlation of consumer preferences and merchandise offered by the firm.\textsuperscript{41}

\textsuperscript{39}Ibid., p. 322.  \textsuperscript{40}Ibid., p. 322.  \textsuperscript{41}Ibid., p. 322.
The next stage amounts to a demographic analysis of the region in question. Particular emphasis should be placed on determining the number within the region that resemble the consumer profile previously derived. This number represents an estimate of potential customers. An estimation of expected sales can now be generated by accounting for an allotted quantity of sales for each of the perspective customers. Estimated profits can be determined by subtracting the cost from the expected sales. Each prospective location can then be ranked by the estimated profits.\(^{42}\)

This technique obviously depends on survey material. The measurements taken may not be so precise or easily obtained as driving time or store size, but they do seem to be somewhat more pertinent.

Miscellaneous Methods

**Retail Saturation Index.** This is not really a method for determining a site, but a simplistic way of narrowing the number of sites being considered. It is estimated that about one half of all sites are rejected on the surface due to the demands imposed by the management of the chain store.\(^ {43}\)

The crux of this method is the division of population within the region by the square footage of sales space allotted

\(^{42}\)Ibid., p. 322.

to a particular line of goods. This figure is then compared to data on the national ratio of floor space to population for this good. If the computed ratio exceeds the national ratio then a site in this area is rejected.

**Systems Construct.** This model has three components: origins, which are areas of money generation; links, which are roads connecting origins and destinations; and the destinations. Each of the parts is defined by an equation, which is then calculated quickly by a computer.

The following equation is used for determining the amount of money flowing from each origin for grocery items:

\[ Y_i = 52(C_i)(P_i) \]

where \( Y_i \) = the yearly flow of money from origin area \( i \)
\( C_i \) = the per capita weekly food costs in area \( i \)
\( P_i \) = the population of area \( i \)
\( C_i \) is defined as:

\[ C_i = 2.2F_i - 1.31 \]

where \( F_i \) is the average family size.

The links are measured in terms of resistance to flow by the following formula:

\[ X_{ij} = R_{ij} Y_{ij} \]

---

where \( X_{ij} \) = the pressure required to cross the link.
\( R_{ij} \) = the link's resistance to flow.
\( Y_{ij} \) = the yearly net flow of money through the link.\(^{46}\)

The destinations are measured by the following formula:

\[
Y_i = A_j X_j
\]

where \( Y_i \) is the yearly sales supermarket \( j \).

\( A_j \) is the attraction of supermarket \( j \).

\( X_j \) is the propensity to shop at supermarket \( j \).\(^{47}\)

The system method has proven in a Canadian study to be a fairly accurate predictor of where money will be spent.\(^{48}\)

However, it appears to be seriously weakened due to the limited number of variables that are considered during the calculation. Also, the method is most valid for grocery stores where purchases are made on a regular basis and in reasonably large amounts.

\(^{46}\)Ibid., p. 45.
\(^{47}\)Ibid., p. 45.
\(^{48}\)Ibid., p. 45.
CHAPTER IV

ANALYSIS OF DATA

Two chains of stores agreed to participate in the study. The first chain was Pizza Hut, Incorporated, operators of fast food restaurants. The second was Zale Corporation, whose largest chain of stores deals in jewelry under the corporate name. The sample of stores from Pizza Hut included 42 stores from Texas and Louisiana (predominately Texas) with from one to four years of sales data for each store giving a total of 108 data points and 17 stores in Louisiana with multiple years of sales data for most stores giving a total of 45 data points. Data relative to the store sites were contributed by the corporate headquarters in Wichita, Kansas. Sales data for each store were contributed by the regional offices in Euless, Texas. The sales data were adjusted to account for inflation with 1977 serving as the base year.

The sample of stores for Zale included 92 stores from Texas. Data pertaining to the shopping center in which the Zale's store is located and the characteristics of this center were donated by the corporate headquarters. Additionally, sales data were furnished by the company. However, all sales data were coded to maintain company secrecy. Demographic data around each site were gathered from
1970 U. S. Census Data.

The decision to use stores in the Texas area from both chains was based on the belief that differences in regional images by consumers for each chain would effect the viability of the use of regression analysis.

Table II represents the variables considered in the study of the Pizza Hut data. Table II lists the variables used in the study for Zale Corporation.

TABLE II

Pizza Hut Independent Variables Used in the Study

1. The total amount of square feet in the lot of the store.

2. The number of parking spaces around the store.

3. Whether or not the store is connected to other buildings (dummy coded with a one indicating a free-standing store and a zero for a store that is connected to other buildings.)

4. Whether or not the lot is on a corner where two streets intersect (a one indicates a corner lot while a zero is for a lot that is not on a corner.)

5. The number of people who work within a mile and a half of the store.

6. A dummy coded variable with a one indicating that there is residential backup around the store and a zero meaning that few homes are nearby.

7. A dummy coded variable dealing with the traffic past the store. A zero indicates that the traffic tends to be predominately going away from the homes of the drivers, while a one means that the traffic is home-ward bound.

8. The average age of the people around the store.
9. The median income of the nearby residents around the store.

10. The total number of people within a mile and a half of the store.

11. The total number of people within three miles of the store.

12. The traffic count in front of the store.

13. The speed limit on the street in front of the store.

14. The percent of commuter traffic traveling by the store.

15. The number of other stores that sell pizza within a mile and a half.

16. The number of other restaurants within a mile and a half.

17. The value of the land on which the store is built.

18. A dummy coded variable with a one meaning that the store sells beer and a zero indicating no beer being sold.

**TABLE III**

List of the Independent Variables Considered for Zale Corporation

1. The amount of advertising money spent for the store.

2. The executive rating of the manager in charge of the store (a one was used to indicate the manager was poor, a nine would indicate that the manager was exceptionally good, while a five would be average, etc.).

3. A dummy variable to indicate whether the store was in a mall or not (a one indicated a mall while a zero indicated that another type of shopping facility was present).

4. The total square feet of shopping area in the shopping center.
5. The number of parking spaces around the shopping center.

6. The total number of stores in the shopping center.

7. The total number of anchor stores (major tenants) in the center.

8. The number of grocery stores in the center.

9. The number of national chain anchors in the center.

10. The number of jewelry competitors in the center.

11. The age of the shopping center.

12. The classification according to size and purpose of the center (dummy coded with a zero representing a neighborhood center, a one representing a community shopping facility, and a two representing a center with regional importance).

13. The nonworker to worker ratio with the city.

14. The percent of females over 16 years old that are in the labor force.

15. The percentage of the labor force that is employed by manufacturing companies.

16. The percentage of the labor force that is involved in "white collar" work.

17. The total number of people within the town.

18. The median income of the people within the town.

19. The average number of school years completed by the people within the town.

20. The percentage of the families in the city with an annual income exceeding $15,000.

21. The percent of males between the ages of 18 and 24 that are in the labor force.
Regression Analysis as a Viable Site Selection Analysis of Pizza Hut Data

Table IV shows a formula developed for the prediction of sales of Pizza Hut stores in Texas. Note that each of the variables in the equation are making a statistically significant contribution toward the amount of variance explained by the equation. Table V contains a summary of an analysis of variance test for linearity of this regression equation. It should be noted that the multiple R is .81384 and the multiple R$^2$ is .66234.

A summary of the residuals (the difference between the actual sales and the predicted sales) as a percentage of the actual sales is shown in Table VI. Nearly half of the predictions were within ten percent of the actual sales figures for the stores. The average percentage of residual to actual sales was 14.3 percent. Only once did the equation predict a sales figure that varied by over fifty percent from the actual sales. On this case the estimate was less than half of the actual sales. Underestimation occurred a total of 55 times while overestimation occurred 53 times. Underestimation might cause a company depending upon this approach for site selection to miss some potentially lucrative sites. While overestimation might cause a company to open some stores in sites that are not adequate enough to support the business.

Table VII shows a summary of the standardized residuals for the regression equation developed for Pizza Hut stores.
in Texas. (The average sales for a Pizza Hut store in this area was $186,231.) Well over half (88 out of 108) of the stores had standardized residual values less than one. Thus, most of the actual sales figures for the stores fell within one standard error (the standard error of the estimate was $33,618.77) of the predicted scores.

**TABLE IV**

Regression Equation for Pizza Hut Stores in Texas to Predict Sales of Stores

(Note: The predicted value needs to be multiplied by 1000)

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>Std. Error of B</th>
<th>F-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>3.877409</td>
<td>1.32569</td>
<td>8.555***</td>
</tr>
<tr>
<td>V2</td>
<td>-.003157</td>
<td>.00083</td>
<td>14.332***</td>
</tr>
<tr>
<td>V3</td>
<td>103.1981</td>
<td>19.56192</td>
<td>27.830***</td>
</tr>
<tr>
<td>V4</td>
<td>.069143</td>
<td>.00777</td>
<td>79.180***</td>
</tr>
<tr>
<td>V5</td>
<td>-.486243</td>
<td>.22181</td>
<td>4.806***</td>
</tr>
<tr>
<td>V6</td>
<td>-.106 x 10^-12</td>
<td>.1 x 10^-13</td>
<td>4.972***</td>
</tr>
<tr>
<td>V7</td>
<td>-24.35437</td>
<td>5.39869</td>
<td>20.551***</td>
</tr>
<tr>
<td>V8</td>
<td>-.6796 x 10^-9</td>
<td>.20 x 10^-10</td>
<td>70.126***</td>
</tr>
<tr>
<td>V9</td>
<td>.758076</td>
<td>.00045</td>
<td>2.788*</td>
</tr>
<tr>
<td>V10</td>
<td>-.046101</td>
<td>.00754</td>
<td>37.427***</td>
</tr>
<tr>
<td>V11</td>
<td>.3408 x 10^-7</td>
<td>.101 x 10^-7</td>
<td>20.508***</td>
</tr>
<tr>
<td>V12</td>
<td>-.1053 x 10^-10</td>
<td>.215 x 10^-8</td>
<td>12.914***</td>
</tr>
<tr>
<td>Constant</td>
<td>-5.536911</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant: $\alpha = .10$

**Significant: $\alpha = .05$

***Significant: $\alpha = .01$

Where V1 is the number of parking spaces on the site
V2 is the square feet taken by the lot
V3 is whether or not the store is free-standing
V4 is the number of people working within a mile and a half
V5 is the number of other stores selling pizza in the area
V6 is the cubed value of the traffic count past the store
V7 is whether or not the traffic past the store is homeward bound multiplied by the number of other restaurants in the area
V8 is the cubed value of the number of people working in the area
V9 is the square feet taken by the lot multiplied by whether the store is on a corner lot
V10 is whether or not the store is on a corner multiplied by the number of people working in the area
V11 is the number of people living within three miles multiplied by the traffic count in front of the store
V12 is the number of people living within three miles multiplied by the value of the site

TABLE V

(Regression Equation) Summary Table for Pizza Huts in Texas

Multiple $R^2 = \ 0.81384$
Multiple $R = \ 0.66234$
Standard Error $= 33618.77$

<table>
<thead>
<tr>
<th>Analysis of Variance</th>
<th>Df</th>
<th>Sum of Squares</th>
<th>Mean Squares</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>12</td>
<td>210617.69559</td>
<td>17551.47463</td>
<td>15.52923*</td>
</tr>
<tr>
<td>Residual</td>
<td>95</td>
<td>107371.05999</td>
<td>1150.22168</td>
<td></td>
</tr>
</tbody>
</table>

*Significant: $\alpha = .01$
TABLE VI

<table>
<thead>
<tr>
<th>Limits within which the percentage falls</th>
<th>Number over-estimated</th>
<th>Number underestimated</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-5</td>
<td>11</td>
<td>14</td>
<td>25</td>
</tr>
<tr>
<td>5-10</td>
<td>14</td>
<td>13</td>
<td>27</td>
</tr>
<tr>
<td>10-15</td>
<td>11</td>
<td>10</td>
<td>21</td>
</tr>
<tr>
<td>15-20</td>
<td>9</td>
<td>8</td>
<td>17</td>
</tr>
<tr>
<td>20-30</td>
<td>5</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>30-50</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>over 50</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Totals 53 55 108

Average Percentage = 14.3

TABLE VII

Summary Table of Standardized Residuals for Pizza Huts in Texas

<table>
<thead>
<tr>
<th>Standardized Value</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over -2</td>
<td>2</td>
</tr>
<tr>
<td>Between -2 and -1</td>
<td>9</td>
</tr>
<tr>
<td>Between -1 and -.5</td>
<td>11</td>
</tr>
<tr>
<td>Between -.5 and -.25</td>
<td>12</td>
</tr>
<tr>
<td>Between -.25 and 0</td>
<td>19</td>
</tr>
<tr>
<td>Between 0 and +.25</td>
<td>17</td>
</tr>
<tr>
<td>Between +.25 and +.5</td>
<td>14</td>
</tr>
<tr>
<td>Between +.5 and +1</td>
<td>11</td>
</tr>
<tr>
<td>Between +1 and +2</td>
<td>10</td>
</tr>
<tr>
<td>Over +2</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>108</td>
</tr>
</tbody>
</table>

Table VIII shows a formula developed for the prediction of sales of Pizza Hut stores in Louisiana. Again, all of the variables in the equation are significant contributors to the amount of variance explained. Table IX contains a
summary of an analysis of variance test for linearity of this regression equation. The test for linearity shows that it is a highly significant linear relationship. The multiple R for this equation is .86215.

Table X shows a summary of the residuals as a percent of the actual sales. Nearly half of the stores (21 out of 45) had predictions that missed the actual sales figure by less than ten percent. The average percent that the predicted sales figure missed the actual sales figure was 13.6 percent. It should be noted that the sales of 24 or the 45 stores were underestimated.

TABLE VIII

Regression Equation for Pizza Hut Stores in Louisiana

(Note: the predicted value needs to be multiplied by 1000)

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>Std. Error of B</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>.3798 x 10^{-7}</td>
<td>.306 x 10^{-8}</td>
<td>3.141*</td>
</tr>
<tr>
<td>V2</td>
<td>-.7256 x 10^{-10}</td>
<td>.827 x 10^{-11}</td>
<td>19.048***</td>
</tr>
<tr>
<td>V3</td>
<td>-.1036 x 10</td>
<td>.00003</td>
<td>10.324***</td>
</tr>
<tr>
<td>V4</td>
<td>.1270 x 10^{-6}</td>
<td>.153 x 10^{-7}</td>
<td>14.224***</td>
</tr>
<tr>
<td>V5</td>
<td>.1695 x 10</td>
<td>.189 x 10</td>
<td>5.203**</td>
</tr>
<tr>
<td>V6</td>
<td>25.75410</td>
<td>13.63722</td>
<td>3.566*</td>
</tr>
<tr>
<td>Constant</td>
<td>109.7062</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant: \( \alpha = .10 \)
**Significant: \( \alpha = .05 \)
***Significant: \( \alpha = .01 \)
Where $V_1$ is the median income of the people working in the area multiplied by the value of the land

$V_2$ is the cubed value of the number of people working in the area

$V_3$ is the cubed value of the number of stores selling pizza in the area

$V_4$ is the number of square feet the lot consumes multiplied by the number of people living within a mile and a half

$V_5$ is the number of parking spaces around the store multiplied by the value of the land

$V_6$ whether or not the traffic by the store is homeward bound

TABLE IX

Regression Equation Summary Table for Pizza Hut Stores in Louisiana

Multiple $R^2 = .86215$
Multiple $R = .74330$
Standard Error = 31980.42

Analysis of Variance

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum of Squares</th>
<th>Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>6</td>
<td>112464.53777</td>
<td>18744.08963</td>
</tr>
<tr>
<td>Residual</td>
<td>38</td>
<td>38840.09196</td>
<td>1022.10768</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>18.33866*</td>
</tr>
</tbody>
</table>

*Significant: $\alpha = .01$
Table X

The Residual as a Percent of the Actual Sales

<table>
<thead>
<tr>
<th>Limits within which the percentage falls</th>
<th>Number over-estimated</th>
<th>Number underestimated</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-5</td>
<td>4</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>5-10</td>
<td>7</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>10-15</td>
<td>3</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>15-20</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>20-30</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>30-50</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>over 50</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Totals 21 24 45

Average percentage = 13.6

Table XI

Regression Equation Summary for Standardized Residuals for Pizza Hut Stores in Louisiana

<table>
<thead>
<tr>
<th>Standard Value</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over -2</td>
<td>1</td>
</tr>
<tr>
<td>Between -2 and -1</td>
<td>2</td>
</tr>
<tr>
<td>Between -1 and -.5</td>
<td>5</td>
</tr>
<tr>
<td>Between -.5 and -.25</td>
<td>4</td>
</tr>
<tr>
<td>Between -.25 and 0</td>
<td>9</td>
</tr>
<tr>
<td>Between 0 and +.25</td>
<td>10</td>
</tr>
<tr>
<td>Between +.25 and +.5</td>
<td>8</td>
</tr>
<tr>
<td>Between +.5 and +1.0</td>
<td>4</td>
</tr>
<tr>
<td>Between +1 and +2</td>
<td>2</td>
</tr>
<tr>
<td>Over +2</td>
<td>0</td>
</tr>
</tbody>
</table>

Total 45
Analysis of Zale Corporation Data

A regression equation developed for the prediction of sales for stores belonging to Zale Corporation is in Table XII. All of the variables used in the equation make statistically significant contributions to the amount of variance explained by the equation. Table IXX contains a list of the original variables used for Zales and their simple correlations with sales for the stores. The variable that correlated most highly with the sales of the store was advertising expenditures ($r = .936$). It was assumed and later verified that Zales established their advertising budget for each store based on their expected sales for the store. Consequently, this variable was deemed meaningless as far as prediction of sales and was dropped from consideration.

Table XII

Regression Equation for Zale Corporation to Predict Sales of Stores

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>Std. Error of B</th>
<th>F-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>2356.452</td>
<td>352.11279</td>
<td>44.7872*</td>
</tr>
<tr>
<td>V2</td>
<td>1.188</td>
<td>.23430</td>
<td>25.7021*</td>
</tr>
<tr>
<td>V3</td>
<td>-30.025</td>
<td>12.69734</td>
<td>5.5916*</td>
</tr>
<tr>
<td>V4</td>
<td>-7.664</td>
<td>2.69734</td>
<td>8.0729*</td>
</tr>
<tr>
<td>V5</td>
<td>.002</td>
<td>.00073</td>
<td>5.8277*</td>
</tr>
<tr>
<td>Constant</td>
<td>-8009.797</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant: $\alpha = .01$
Where \( V_1 \) is the rating of the store manager
\( V_2 \) is the number of parking spaces around the center
\( V_3 \) is the number of stores in the shopping center
\( V_4 \) is the proportion of the females over 16 in the community's labor force
\( V_5 \) is the population of the community

Table XIII contains the results of an analysis of variance test for linearity on the regression equation. A statistically significant linear relationship was found to exist in the equation.

Table XIV shows a summary of the residuals as a percent of the actual sales. About one-fourth of the stores (26 out of 92) had predicted sales values that were within ten percent of the actual sales. The average percent that the predicted sales value missed the actual sales was 23.7 percent. Table XV contains a summary of the standardized residuals for the equation developed for Zale Corporation. Most of the stores (76 out of 92) had actual sales that fell within one standard error (the standard error of the estimate was 2394.3) of the predicted sales (the average sales for Zale stores was $6974.00).

**Table XIII**

Summary Table for Zale Corporation Regression Equation

- \( R^2 \) = 0.7799
- \( R^2 \) = 0.6082
- Standard Error = 2394.3013
### TABLE XIII--Continued

<table>
<thead>
<tr>
<th>Analysis of Variance</th>
<th>Df</th>
<th>Sum of Squares</th>
<th>Mean Squares</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>5</td>
<td>765300992.00</td>
<td>153060192.00</td>
<td>26.730*</td>
</tr>
<tr>
<td>Residual</td>
<td>86</td>
<td>493010432.00</td>
<td>5732679.00</td>
<td></td>
</tr>
</tbody>
</table>

*Significant: $\alpha = .01$

### TABLE XIV

The Residual as a Percent of Actual Sales

<table>
<thead>
<tr>
<th>Limits within which the percentage falls</th>
<th>Number over-estimated</th>
<th>Number under-estimated</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-5</td>
<td>7</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>5-10</td>
<td>5</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>10-15</td>
<td>4</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>15-20</td>
<td>6</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>20-30</td>
<td>8</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>30-50</td>
<td>17</td>
<td>12</td>
<td>19</td>
</tr>
<tr>
<td>over 50</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

**Totals**

|               | 50 | 42 | 92 |

Average Percentage = 23.7
### TABLE XV

Summary Table for Residuals for Zale Corporation

<table>
<thead>
<tr>
<th>Standardized Value</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over -2</td>
<td>1</td>
</tr>
<tr>
<td>Between -2 and -1</td>
<td>3</td>
</tr>
<tr>
<td>Between -1 and -.5</td>
<td>14</td>
</tr>
<tr>
<td>Between -.5 and -.25</td>
<td>14</td>
</tr>
<tr>
<td>Between -.25 and 0</td>
<td>18</td>
</tr>
<tr>
<td>Between 0 and +.25</td>
<td>12</td>
</tr>
<tr>
<td>Between +.25 and +.5</td>
<td>9</td>
</tr>
<tr>
<td>Between +.5 and +1.0</td>
<td>9</td>
</tr>
<tr>
<td>Between +1.0 and +2.0</td>
<td>8</td>
</tr>
<tr>
<td>Greater than +2.0</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>92</strong></td>
</tr>
</tbody>
</table>

Thus, each of the three regression equations developed have a statistically significant linear relationship. While, all overaged predictions within 25 percent of the actual sales figure.

**Determination of Differences Among Variables of Importance for Different Chains of Stores**

Table XVII contains the correlation coefficients for the relationship between sales and each of the variables in the Pizza Hut analysis (for both Louisiana and Texas). Table XVI has similar information for the Zales study. Table XVIII contains the results of a statistical analysis using Fisher's Z' Transformation to test the difference in Correlation between similar variables that existed in both the Pizza Hut study and the Zale study. Three of the four
variables were found to be statistically significant. Of the four, only the correlations between population and sales for the two chains was found to be statistically insignificant. To add further support to this point, it is interesting to note that none of the variables used in the Pizza Hut regression equations are present in the Zale regression equation. Of course, little emphasis can be placed on this due to the nature of regression analysis and the techniques used to derive the equations used in this paper.

TABLE XVI

Simple Correlations Between Original Independent Variables and Sales for Zales' Stores

<table>
<thead>
<tr>
<th>Variable</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Whether or not the store is in a mall</td>
<td>.495</td>
</tr>
<tr>
<td>2. The number of square feet in the shopping center</td>
<td>.575</td>
</tr>
<tr>
<td>3. The number of parking spaces around the center</td>
<td>.581</td>
</tr>
<tr>
<td>4. The number of stores in the center</td>
<td>.441</td>
</tr>
<tr>
<td>5. The number of anchor stores in the center</td>
<td>.533</td>
</tr>
<tr>
<td>6. The number of grocery stores in the center</td>
<td>-.322</td>
</tr>
<tr>
<td>7. The number of national chain store anchors in the center</td>
<td>.409</td>
</tr>
<tr>
<td>8. The number of other jewelry stores in the center</td>
<td>.439</td>
</tr>
<tr>
<td>9. The age of the center</td>
<td>-.050</td>
</tr>
<tr>
<td>10. The type of the center</td>
<td>.521</td>
</tr>
<tr>
<td>11. The nonworker to worker ratio in the area</td>
<td>.013</td>
</tr>
<tr>
<td>12. The percentage of females over 16 in the labor force</td>
<td>.036</td>
</tr>
<tr>
<td>13. The percentage of the labor force employed in manufacturing</td>
<td>-.172</td>
</tr>
<tr>
<td>14. The percentage of the labor force made up by white collar workers</td>
<td>.210</td>
</tr>
</tbody>
</table>
### TABLE XVI--Continued

| 15. The population in the area | .224 |
| 16. The median income in the area | .049 |
| 17. The average number of years of school completed | .069 |
| 18. The percentage of families with annual incomes over $15,000 | .032 |
| 19. The percentage of males 18 to 24 in the labor force | -.006 |
| 20. The advertising budget of the store | .936 |
| 21. The executive rating of the management of the store | .625 |

### TABLE XVII

Simple Correlations Between Original Independent Variables and Sales for Pizza Hut Stores in Texas and Louisiana

<table>
<thead>
<tr>
<th>Variable</th>
<th>Texas Correlation</th>
<th>Louisiana Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Number of square feet at the site</td>
<td>-.180</td>
<td>-.053</td>
</tr>
<tr>
<td>2. Number of parking spaces</td>
<td>-.097</td>
<td>-.030</td>
</tr>
<tr>
<td>3. Whether or not the store is free-standing</td>
<td>.229</td>
<td>.000</td>
</tr>
<tr>
<td>4. Whether or not the store is on a corner</td>
<td>-.273</td>
<td>-.508</td>
</tr>
<tr>
<td>5. The number of people working in the area</td>
<td>.204</td>
<td>-.197</td>
</tr>
<tr>
<td>6. Whether or not there is a residential backup around the store</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>7. Whether or not the traffic is homeward bound</td>
<td>-.161</td>
<td>.514</td>
</tr>
<tr>
<td>8. The average age of the people in the area</td>
<td>-.111</td>
<td>-.142</td>
</tr>
<tr>
<td>9. The median income of the people in the area</td>
<td>-.199</td>
<td>.598</td>
</tr>
<tr>
<td>10. The number of people living within a mile and a half</td>
<td>.161</td>
<td>.148</td>
</tr>
<tr>
<td>11. The number of people living within three miles</td>
<td>.031</td>
<td>.101</td>
</tr>
<tr>
<td>12. The traffic count in front of the store</td>
<td>.115</td>
<td>.027</td>
</tr>
</tbody>
</table>
TABLE XVII--Continued

<table>
<thead>
<tr>
<th>Variable</th>
<th>Texas Correlation</th>
<th>Louisiana Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>13. The speed limit in front of the store</td>
<td>.177</td>
<td>.326</td>
</tr>
<tr>
<td>14. The percentage of the traffic that is for commuter purposes</td>
<td>-.044</td>
<td>.118</td>
</tr>
<tr>
<td>15. The number of pizza stores in the area</td>
<td>.177</td>
<td>.326</td>
</tr>
<tr>
<td>16. The number of restaurants in the area</td>
<td>.210</td>
<td>-.267</td>
</tr>
<tr>
<td>17. The value of the land</td>
<td>-.108</td>
<td>.580</td>
</tr>
<tr>
<td>18. Whether or not the store sells beer</td>
<td>.270</td>
<td>0.000</td>
</tr>
</tbody>
</table>

TABLE XVIII

Fisher's Z' Transformations As A Comparison Between Correlations of Similar Variables for Pizza Hut and Zales

<table>
<thead>
<tr>
<th>Variable</th>
<th>Pizza Hut Correlation n = 108</th>
<th>Zale's Correlation n = 92</th>
<th>Z Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>The number of parking spaces</td>
<td>-.097</td>
<td>.581</td>
<td>-4.776***</td>
</tr>
<tr>
<td>The median income of residents</td>
<td>-.199</td>
<td>.049</td>
<td>-1.721*</td>
</tr>
<tr>
<td>The population in the area</td>
<td>.161</td>
<td>.224</td>
<td>-.437</td>
</tr>
<tr>
<td>The number of competitors</td>
<td>.117</td>
<td>.439</td>
<td>-2.235**</td>
</tr>
</tbody>
</table>

*Significant: $\alpha = .10$
**Significant: $\alpha = .05$
***Significant: $\alpha = .01$
Determination of Differences Among Variables for Different Regions Within the Same Chain

Table IXX contains the results of a statistical analysis using Fisher's Z' Transformations to test the difference in correlation between the same variables that were present in both the Texas and Louisiana studies for Pizza Hut. Eight of the eighteen variables on the correlation coefficients were statistically different. While only one variable (the cubed value of the number of people working in the area) was present in both of the regression equations developed for each region.

**TABLE IXX**

*Fisher's Z' Transformations As A Comparison Between Correlations of Similar Variables for Pizza Hut Stores in Louisiana and Texas*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Texas</th>
<th>Louisiana</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-.180</td>
<td>-.053</td>
<td>-.713</td>
</tr>
<tr>
<td>2</td>
<td>-.097</td>
<td>-.030</td>
<td>-.376</td>
</tr>
<tr>
<td>3</td>
<td>.229</td>
<td>.000</td>
<td>1.285*</td>
</tr>
<tr>
<td>4</td>
<td>-.273</td>
<td>-.508</td>
<td>1.319*</td>
</tr>
<tr>
<td>5</td>
<td>.204</td>
<td>-.197</td>
<td>2.251**</td>
</tr>
<tr>
<td>6</td>
<td>.000</td>
<td>.000</td>
<td>.00</td>
</tr>
<tr>
<td>7</td>
<td>-.161</td>
<td>.514</td>
<td>3.788***</td>
</tr>
<tr>
<td>8</td>
<td>-.111</td>
<td>-.142</td>
<td>.174</td>
</tr>
<tr>
<td>9</td>
<td>-.199</td>
<td>.598</td>
<td>4.473***</td>
</tr>
<tr>
<td>10</td>
<td>.161</td>
<td>.148</td>
<td>.073</td>
</tr>
<tr>
<td>11</td>
<td>.031</td>
<td>.101</td>
<td>-.393</td>
</tr>
<tr>
<td>12</td>
<td>.115</td>
<td>.027</td>
<td>.494</td>
</tr>
<tr>
<td>13</td>
<td>-.044</td>
<td>.118</td>
<td>-.909</td>
</tr>
<tr>
<td>14</td>
<td>.177</td>
<td>.020</td>
<td>.881</td>
</tr>
<tr>
<td>15</td>
<td>.117</td>
<td>.326</td>
<td>-1.173</td>
</tr>
<tr>
<td>16</td>
<td>.210</td>
<td>-.267</td>
<td>2.677***</td>
</tr>
<tr>
<td>17</td>
<td>-.108</td>
<td>.580</td>
<td>3.861**</td>
</tr>
<tr>
<td>18</td>
<td>.270</td>
<td>.000</td>
<td>1.515*</td>
</tr>
</tbody>
</table>

*Significant: α= .1  **Significant: α= .05  ***Significant: α= .01
Demographics Versus Traffic Patterns
as a Predictor Variable

Table XX contains a regression equation using all of the original variables for Pizza Hut. Tables XXI and XXIII contain regression equations for demographic variables and traffic pattern variables respectively. Table XXII gives an analysis of variance test for linearity of the regression equation for demographic variables. Table XXIV gives an analysis of variance test for linearity of the regression equation for traffic pattern variables.

**TABLE XX**

Regression Equation for Pizza Hut Using All the Original Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.0023</td>
</tr>
<tr>
<td>2</td>
<td>0.5576</td>
</tr>
<tr>
<td>3</td>
<td>95.1453</td>
</tr>
<tr>
<td>4</td>
<td>-35.3674</td>
</tr>
<tr>
<td>5</td>
<td>0.0009</td>
</tr>
<tr>
<td>6</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td>9.7197</td>
</tr>
<tr>
<td>8</td>
<td>-0.3665</td>
</tr>
<tr>
<td>9</td>
<td>0.0002</td>
</tr>
<tr>
<td>10</td>
<td>0.0007</td>
</tr>
<tr>
<td>11</td>
<td>0.0000</td>
</tr>
<tr>
<td>12</td>
<td>0.0001</td>
</tr>
<tr>
<td>13</td>
<td>8.613</td>
</tr>
<tr>
<td>14</td>
<td>0.6312</td>
</tr>
<tr>
<td>15</td>
<td>6.9213</td>
</tr>
<tr>
<td>16</td>
<td>1.9329</td>
</tr>
<tr>
<td>17</td>
<td>-0.0001</td>
</tr>
<tr>
<td>18</td>
<td>10.5777</td>
</tr>
<tr>
<td>Constant</td>
<td>187.2848</td>
</tr>
</tbody>
</table>

Note: The numbers of each of these variables agree with Table II
**TABLE XXI**

Regression Equation for Original Demographic Variables for Pizza Hut Stores

\[
\begin{align*}
\text{Multiple } R^2 &= 0.3394 \\
\text{Multiple } R^2 &= 0.1152 \\
\text{Sum of Squares of Residuals} &= 439461.13
\end{align*}
\]

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.0026</td>
</tr>
<tr>
<td>8</td>
<td>-0.7316</td>
</tr>
<tr>
<td>9</td>
<td>-0.0004</td>
</tr>
<tr>
<td>10</td>
<td>0.0023</td>
</tr>
<tr>
<td>11</td>
<td>-0.0005</td>
</tr>
<tr>
<td>Constant</td>
<td>187.2848</td>
</tr>
</tbody>
</table>

**TABLE XXII**

\[
F = \frac{R^2(n-k-1)}{(1-R^2)(k)} = \frac{0.1152(153-6-1)}{(1-0.1152)(6)} = 3.17^* \\
\text{Significant: } \alpha = 0.01
\]

**TABLE XXIII**

Regression Equation for Original Traffic Pattern Variables for Pizza Hut Stores

\[
\begin{align*}
\text{Multiple } R^2 &= 0.2982 \\
\text{Multiple } R^2 &= 0.0889 \\
\text{Sum of Squares of Residuals} &= 452526.38
\end{align*}
\]

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>11.7328</td>
</tr>
<tr>
<td>7</td>
<td>1.9608</td>
</tr>
<tr>
<td>12</td>
<td>-0.0001</td>
</tr>
<tr>
<td>13</td>
<td>1.1554</td>
</tr>
<tr>
<td>14</td>
<td>0.7601</td>
</tr>
<tr>
<td>Constant</td>
<td>83.3087</td>
</tr>
</tbody>
</table>
TABLE XXIV

\[ F = \frac{R^2(n-k-1)}{(1-R^2)(k)} = \frac{.889(153-6-1)}{(1-.0889)(6)} = 2.37^* \]

*Significant: \( \alpha = .05 \)

Tables XXV and XXVI present regression equations for the full array of variables excluding demographic variables and pattern variables respectively.

**TABLE XXV**

Pizza Hut Regression Equation for the Original Variable of Pizza Hut Stores Excluding Demographic Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-.0023</td>
</tr>
<tr>
<td>2</td>
<td>-.1765</td>
</tr>
<tr>
<td>3</td>
<td>103.5515</td>
</tr>
<tr>
<td>4</td>
<td>-36.6083</td>
</tr>
<tr>
<td>6</td>
<td>.0000</td>
</tr>
<tr>
<td>7</td>
<td>12.5416</td>
</tr>
<tr>
<td>12</td>
<td>.0001</td>
</tr>
<tr>
<td>13</td>
<td>.9477</td>
</tr>
<tr>
<td>14</td>
<td>.6891</td>
</tr>
<tr>
<td>15</td>
<td>7.3712</td>
</tr>
<tr>
<td>16</td>
<td>2.0113</td>
</tr>
<tr>
<td>17</td>
<td>.0000</td>
</tr>
<tr>
<td>18</td>
<td>10.4895</td>
</tr>
<tr>
<td>Constant</td>
<td>52.9140</td>
</tr>
</tbody>
</table>
### TABLE XXVI

Regression Equation for Original Variable of Pizza Hut Stores with Traffic Pattern Variable Excluded

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.0024</td>
</tr>
<tr>
<td>2</td>
<td>1.8271</td>
</tr>
<tr>
<td>4</td>
<td>-30.2538</td>
</tr>
<tr>
<td>5</td>
<td>0.0003</td>
</tr>
<tr>
<td>6</td>
<td>0.0000</td>
</tr>
<tr>
<td>8</td>
<td>0.2845</td>
</tr>
<tr>
<td>9</td>
<td>0.0005</td>
</tr>
<tr>
<td>10</td>
<td>0.0007</td>
</tr>
<tr>
<td>11</td>
<td>-0.0003</td>
</tr>
<tr>
<td>15</td>
<td>6.9534</td>
</tr>
<tr>
<td>16</td>
<td>1.9986</td>
</tr>
<tr>
<td>17</td>
<td>-0.0001</td>
</tr>
<tr>
<td>18</td>
<td>20.2243</td>
</tr>
<tr>
<td>Constant</td>
<td>69.2252</td>
</tr>
</tbody>
</table>

Tables XXVII and XXVIII show the results of analysis of variance tests that compared the predictive power of the two prior equations to the equation that included all of the original variables (Table XVII). The results of the two analysis of variance tests show that compared to all of the variables the subtraction of either demographic variables or traffic pattern variables makes no statistical difference in the predictive power of the model.
TABLE XXVII

Significance Test for the Semi-Partial Correlation Between the Full Model Regression Equation and the Equation that Excludes the Traffic Pattern Variables

<table>
<thead>
<tr>
<th>Equation</th>
<th>Degrees of Freedom</th>
<th>R</th>
<th>Residual</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Model</td>
<td>6</td>
<td>.3333</td>
<td>331159.35</td>
<td>1.23</td>
</tr>
<tr>
<td>Traffic Patterns Exc.</td>
<td>134</td>
<td>.2965</td>
<td>349403.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.0367</td>
<td>18243.77</td>
<td></td>
</tr>
</tbody>
</table>

TABLE XXVIII

<table>
<thead>
<tr>
<th>Equation</th>
<th>Degrees of Freedom</th>
<th>R</th>
<th>Residual</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic Variables Excluded</td>
<td>6</td>
<td>.3333</td>
<td>331159.35</td>
<td>0.59</td>
</tr>
<tr>
<td>Full Model</td>
<td>134</td>
<td>.3158</td>
<td>339854.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.0175</td>
<td>8694.72</td>
<td></td>
</tr>
</tbody>
</table>

While no direct comparison can be made between the demographic predictors and traffic pattern predictors, it is fairly apparent that traffic pattern analysis is not a better predictor than demographic variables. In fact, the correlation between the demographic variable and sales was higher than a similar comparison for traffic pattern variables. Also, excluding both traffic pattern variable and demographic variable was found to make statistically insignificant changes in the predictive ability of the equation in both cases.
CHAPTER V

SUMMARY, IMPLICATIONS, DELIMITATIONS, AND RECOMMENDATIONS

Summary

Samples of forty-two stores and seventeen stores were drawn from two regions within the Pizza Hut chain of stores. Several of the stores had more than one year of sales data, giving total data points from the two samples of 108 and 45, respectively. A third sample of 92 stores was drawn from Zale Corporation, a chain of jewelry stores. Site specific material and sales data were furnished by Pizza Hut for each of the stores. Included among the data provided were traffic pattern analysis around each store and demographic material pertinent to the area surrounding the site. Material relative to the physical characteristics of the shopping centers that Zale's stores occupy was furnished by the chain. Demographic material was garnered from the U. S. Census data.

Through the use of an analysis of variance tests each of the three equations were found to have a statistically significant linear relationship. Statistically significant differences were found between the importance of variance in the prediction of sales between regions by using Fisher's Z' Transformations to test the difference between correlation coefficients. Three of the four similar variables that were available for use for both Pizza Hut and Zale were found to
be statistically different using the above technique. Eight of the eighteen variables in the Pizza Hut study were found to be statistically different between the Texas and Louisiana regions. Additionally, analysis of variance tests were used to indicate that traffic pattern variables were not better predictors of sales than demographic variables.

Implications

On the basis of this study, it appears that regression analysis is probably very effective for predicting the sales of a store at any prospective site. Certainly the ability of this technique to explain the majority of the variance among the sales of stores is better than using no systematic site selection. Additionally, the face validity of regression analysis is greater than the other methods that are commonly used due to the inclusion of more variables that seem to be pertinent to the decision.

Yet, the key is whether regression analysis is the answer to the retail site selection problem. The response to this inquiry has to be that regression analysis isn't the answer. However, it is highly doubtful that any quantitative technique will ever be the answer to the objective approach to site selection. Around each site there will always exist unique, individual features that can never be objectively placed into a quantitative model. For example, the presence of a median in the street in front of the site might impede the access of automobiles to the future store, but an objective measurement
of this factor can never be determined until the store is opened (and perhaps not even at that time). True, it could be possible to develop a checklist of site specific characteristics that could effect the store performance and temper the regression forecast accordingly. But that method can't be any better than the subjective evaluation of the site done by the executive in charge of choosing the potential locations. Therefore, regression analysis and other quantitative methods can never be more than a tool to aid the executive in decision making. While it is true that the quantitative approach is a very strong tool, the final inputs and decisions have to be made by the executives who view the site and have the knowledge of the performance characteristics of the firm.

The above paragraph is not included to discount the findings of the study. Actually the contrary is true, the above is included merely to temper the positive results. This study found regression analysis to be a strong tool for prediction of sales for more than a single type of retail store. The results show that regression analysis is a viable method for predicting sales for stores in an array of different types of sites, including both free-standing (Pizza Hut) and in shopping centers (Zale Corporation). In contrast, other techniques seem to be limited to either one type of store or to one type of location.

Additionally, contrary to prior analysis traffic patterns were not found to be better predictors of sales than demographic
variables. In fact, demographic variables were found to have a higher correlation with sales than traffic patterns. This result is certainly in agreement with the pervading marketing philosophies dealing with market segmentation, which stresses the importance of demographics as a segmentation variable.

**Delimitations**

The use of only two chains with one chain having two samples is not sufficient to infer that regression analysis is a viable technique for all types of stores. Certainly, a larger sample of different types of retail chains would have added some more validity to the study. However, in defense of this point, finding stores willing to cooperate in the study was not an easy matter.

Another possible problem with the study is the size of the samples. Each of the samples was rather small, certainly a sample of 200 for each chain would have been more comfortable. However, due to the regional approach of the study larger samples makes the regression analysis approach of questionable use to the small company that has only one store. It would be hard to infer that all pizza stores could successfully use the equation developed for Pizza Hut. In fact, based on the findings of the study, it would be highly unlikely that a decision such as this would be valid.

The possibility of autocorrelation was ignored. The decision to ignore autocorrelation could have some negative ramifications on the Pizza Hut study where several years of data
were used on several sites.

Another questionable area for the study was the source of data for both of the chains. For Pizza Hut, the source was company records compiled mostly by real estate agents hoping to sell a piece of property to the corporation. Obviously, the real estate agents have a vested interest in the property and could be biased. The problem with the Zale Corporation data was twofold. First, the demographic material was extracted from the 1970 U. S. Census data. Eight years of change hardly builds confidence in the accuracy or validity of this data. The second problem with the Zale data was on management ratings for each of the store managers. The ratings were established by the regional manager, who was asked to evaluate each manager on a scale of one to nine with one being the lowest, nine being the highest and five being average. It is probable that a significant amount of subjectivity entered the evaluations. In fact, it could be possible to suggest that stores with significant deviations between predicted sales and actual sales, might have a store manager who is not rated correctly.

Additionally, sales for both chains were predicted on the basis of sales data that occurred at stores that had been opened for several years. Consequently, the predictions have to be assumed to be more valid for established stores than new stores. Yet, most stores that fail do so within the first couple of years of business.
Recommendations

More research needs to be done on the use of regression analysis in the prediction of sales in a variety of retail industries. It would also be interesting to take a deeper look at the variables that are most important in the selection of sites for stores within the same industry. Also, more research should be done on the use of regression analysis to determine the differences between variables that are important in selecting retail sites for stores in different industries and between different regions of the same company.
APPENDIX 1

Table 4-2
Nelson's Checklist for Site Evaluation

Use an "E" for excellent, a "G" for good, a "F" for fair and a "P" for poor

<table>
<thead>
<tr>
<th>Item</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E</td>
</tr>
</tbody>
</table>

I. Trading area potential
   A. Public utility connections (residential)
   B. Residential building permits issued
   C. School enrollment
   D. New bank accounts opened
   E. Advertising lineage in local newspapers
   F. Retail sales volume
   G. Sales tax receipts
   H. Employment-specific
   I. Employment-general

II. Accessibility
   A. Public transportation (serving site)
   B. Private transportation (serving site)
   C. Parking facilities
   D. Long-range trends (transportation facilities)

III. Growth potential
   A. Zoning pattern
   B. Zoning changes
   C. Zoning potential
   D. Utilities trend
   E. Vacant-land market (land zoned for residential use)
   F. Land use pattern (in areas zoned for other than residential)
   G. Retail-business land use trend
   H. Retail-building trend (building permits issued for new retail business construction)
Table 4-2 continued:

<table>
<thead>
<tr>
<th>Item</th>
<th>E</th>
<th>G</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Retail-improvement trend (permits issued for remodeling, expansion, etc., in existing properties)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. Retail-location trend (changes in occupancy of retail-business locations)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K. Income trend for average family unit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. Plant and equipment expenditure trend</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. Payroll trend</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV. Business Interception</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Location pattern-competitive business</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(between site and trade area (served by and sharing traffic arteries with site)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V. Cumulative-attraction potential</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Neighboring business survey</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VI. Compatibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Compatibility factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VII. Competitive-hazard survey</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Competitive pattern-competitors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(within 1 mile of site (non-intercepting)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Competitive pattern-potential competitive sites</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VIII. Site economics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Cost and return analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Site efficiency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Natural description</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Adjacent amenities (for both vacant-land and existing building sites)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY

Books


Articles


