A GEOGRAPHIC MODEL OF AN URBAN AUTOMOBILE MARKET

Theodore E. Hlavac, Jr.

John D. C. Little

180-66

A GEOGRAPHIC MODEL OF AN URBAN AUTOMOBILE MARKET

Theodore E. Hlavac, Jr.

John D. C. Little

180-66

April 1966

Abstract

A person chooses a store to shop at partly on the basis of the difficulty of getting there. An understanding of the relationships involved is important in marketing strategy, particularly for site selection purposes. We study an urban new-car market. As a measure of shopping difficulty, the distance from a person's residence to the car dealer is used. Assumptions about buyer behavior lead to a model of competitive interaction among dealers and car makes. The model is fitted to three months of sales data for metropolitan Chicago. An interactive computer system has been programmed to make the model available for on-line investigation of a variety of site selection and inter-brand competition questions. For example, a user can add, eliminate, or move a dealer and then ask for the model-predicted changes in the sales and penetration of any dealer or make.

A GEOGRAPHIC MODEL OF AN URBAN AUTOMOBILE MARKET

Theodore E. Hlavac, Jr.
Cambridge Research Institute

John D. C. Little
Sloan School of Management
Massachusetts Institute of Technology

1. Introduction

When a person buys a car, he takes into account either explicitly or implicitly, the distance he must travel to the prospective place of purchase. He is likely to be attracted to a nearby dealer because of easy accessibility for shopping trips and future service visits. He is less likely to be attracted to a distant dealer, but some attraction will exist, especially in the case of a well-advertised dealer having an image of low price and high throughput. Within an urban market, where there is considerable choice of dealers, the probability of purchase from a given dealer can be expected to fall off with distance; and, in fact, such a relationship is easily established empirically.

In the present paper we develop a model in which a customer's probability of purchase at a given dealer is affected by dealer location and customer make preference, as well as the locations and strengths of all other dealers. Aggregation of the customer model gives a dealer market share (penetration) model, which may also be viewed as a model of competitive interaction. Such a model is fit to data for metropolitan Chicago. After fitting, the model permits estimation of the sales of a dealership with specified strength and location.

The most obvious practical use of the model relates to market strategy for new dealerships in the automobile industry, but the model appears to be adaptable to site location problems in other fields as well.

This work was supported in part by Project MAC, an MIT research program sponsored by the Advanced Research Projects Agency, Department of Defense, under Office of Naval Research Contract Number Nonr-4102(01). Part of the research was done at the MIT Computation Center.

2. Model

2.1 Dealer Pull. To some extent all dealers are attractive to a person planning to buy a car. The attractiveness of a specified dealer presumably is a function of dealer characteristics (such as make of car sold, extent of advertising), buyer characteristics (such as make preference), and distance from dealer to buyer. The attractiveness of a dealer will be called his "pull". Pull will not be a directly observable quantity but will be used to develop expressions that are.

We shall assume that the number of car purchases is fixed for the time period under consideration. We have not modeled the effect of a dealer in creating sales that would not have occurred in his absence. Our data do not seem to lend themselves to the estimation of this effect, and perhaps it is small in today's well developed markets.

Buyers will be separated into market segments. The pull of a dealer on a buyer in a given segment will be broken into two parts: (1) an intrinsic pull independent of the make sold by the dealer and (2) the make preference of the buyer. Let

$g(i,j)$ = the pull of dealer j on a buyer in market segment i. $i = 1,\ldots,S$ $j = 1,\ldots,D$.

$h(i,j)$ = the intrinsic pull of dealer j on a buyer in segment i.

$q(i,m)$ = the make preference of a buyer in segment i for make m. $m = 1,\ldots,M$. We specify that $q(i,m) \geq 0$ and $\Sigma_{m=1}^{M} q(i,m) = 1$.

Let m(j) denote the make sold by dealer j. We stipulate that the above quantities be related by

$$g(i,j) = h(i,j)\, q(i, m(j)) \tag{1}$$

Thus, the pull of a dealer on a buyer is the dealer's intrinsic pull weighted by the buyer's brand preference.

2.2 Purchase probability. The probability that a buyer purchases at a given dealer will be taken as the pull of that dealer on the buyer divided by the total pull on the buyer. Let

$p(i,j)$ = the probability that a buyer in market segment i purchases at dealer j.

$$p(i,j) = g(i,j) \,/\, \Sigma_{k=1}^{S} g(i,k) \tag{2}$$

We note that make preference can be interpreted as the probability of purchase of the make under the conditions that the sum of the intrinsic pulls on the buyer is the same for each make. This result can be deduced from (1) and (2).

2.3 Geographic effect. We hypothesize that pull falls off exponentially with the distance between dealer and buyer. This is only one of many possible relations but it appears to work well. Let

$x(i,j)$ = distance of the buyers in market segment i to dealer j.

$$h(i,j) = a_j e^{-b_j x(i,j)} \quad (3)$$

Here a_j and b_j are constants specific to dealer j. The constant a_j expresses the dealer's strength in his own immediate neighborhood. The constant b_j tells how fast his sales fall off with distance. Using (1), (2), and (3), we get

$$p(i,j) = \frac{q(i,m(j))\ a_j\ e^{-b_j x(i,j)}}{\Sigma_{k=1}^{D}\ q(i,\ m(k))\ a_k\ e^{-b_k\ x(i,k)}} \quad (4)$$

2.4 Dealer sales and penetration. Let

$N(i)$ = number of buyers in market segment i (called the potential of the segment) in a given time period.

$s(j)$ = expected sales of dealer j in the given time period.

$\pi(j)$ = expected penetration of dealer j in the whole city.

Then

$$s(j) = \Sigma_{i=1}^{S} N(i)\ p(i,j) \quad (5)$$

$$\pi(j) = s(j) / \Sigma_{i=1}^{S} N(i) \quad (6)$$

3. Fitting the Model to Data

3.1 Data. The model has been fit to R.L.Polk new car registrations for April, May, and June 1963 in Cook County, Illinois (fleet sales eliminated). In order to maintain reasonable sample sizes with three months of data, the analysis has been confined to 8 major makes (Ford, Chevrolet, Rambler, Pontiac, Plymouth, Buick, Oldsmobile, and Dodge) and to dealers selling 80 or more cars in the time period. This involves 47,670 cars and 184 dealers, or 91.5% of the cars and 62% of the dealers of the makes

considered. The county has been divided into 140 marketing areas or cells. The rationale for this is discussed by James [1]. The location of each dealer has been determined as has been the sales of each dealer in each cell.

The market has been segmented by geographic area, i.e., by cell. For make preference we have used market share. Since the city is laid out rectangularly, distance has been measured rectangularly instead of diagonally. Some alternatives to these choices will be suggested below.

3.2 Fitting procedure. The model has been fit to the data by a modified maximum likelihood procedure. We observe that, if the denominator of (4) is regarded as a known constant, each dealer becomes a separate estimation problem involving two parameters, a_j and b_j, and 140 data points, the dealer's sales in each cell. Given all the a_j and b_j the denominator of (4) can be calculated. This suggests an iterative procedure. Suppressing the dealer index j, let

n_i = sales of dealer j in cell i.

$$w_i = q(i, m(j))/\Sigma_{k=1}^{D} q(i, m(k))\, a_k\, e^{-b_k x(i,k)} \qquad (7)$$

$p_i = w_i\, a\, e^{-bx_i}$ = probability a buyer in cell i purchases at dealer j.

Then assuming independence of purchase, the likelihood function for the observed sales figures is

$$L = \pi_{i=1}^{S} \binom{N(i)}{n_i} p_i^{n_i}(1-p_i)^{N(i)-n_i} \qquad (8)$$

We start with a trial set of w_i. Values of a and b are then chosen to maximize L. This is done by setting the derivatives of log L with respect to a and b to zero and solving the resulting equations by Newton's method. The a and b are calculated for each dealer. These are then used to recompute the w_i from (7). The process is repeated until the values of the a's and b's converge.

3.3 Results. Figure 1 shows the difference between actual sales and model-predicted sales for the 184 dealers. The two are very close. This is of course desirable, but it is only a moderately good measure of model fit. Although the fitting is done on cell sales rather than dealer sales, and there is no requirement that dealer sales fit exactly, it would be expected from the nature of the calculation that the fit would be close. Figure 2 compares actual and predicted penetration by cell for one dealer. The discrepancies are considerable but, again, they are only a fair indication of fit. Most of these discrepancies are random fluctuations resulting from the small sample sizes involved in any given cell.

Figure 1.

Error (Predicted Sales - Actual Sales)
vs. Dealer Size (Actual Sales)

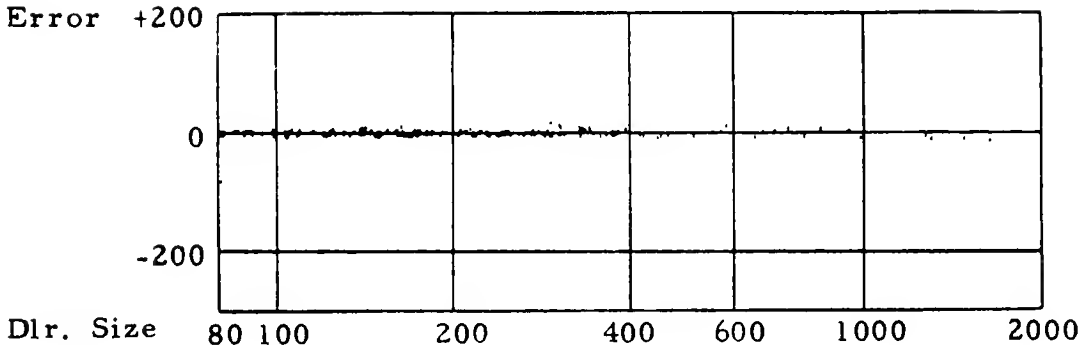
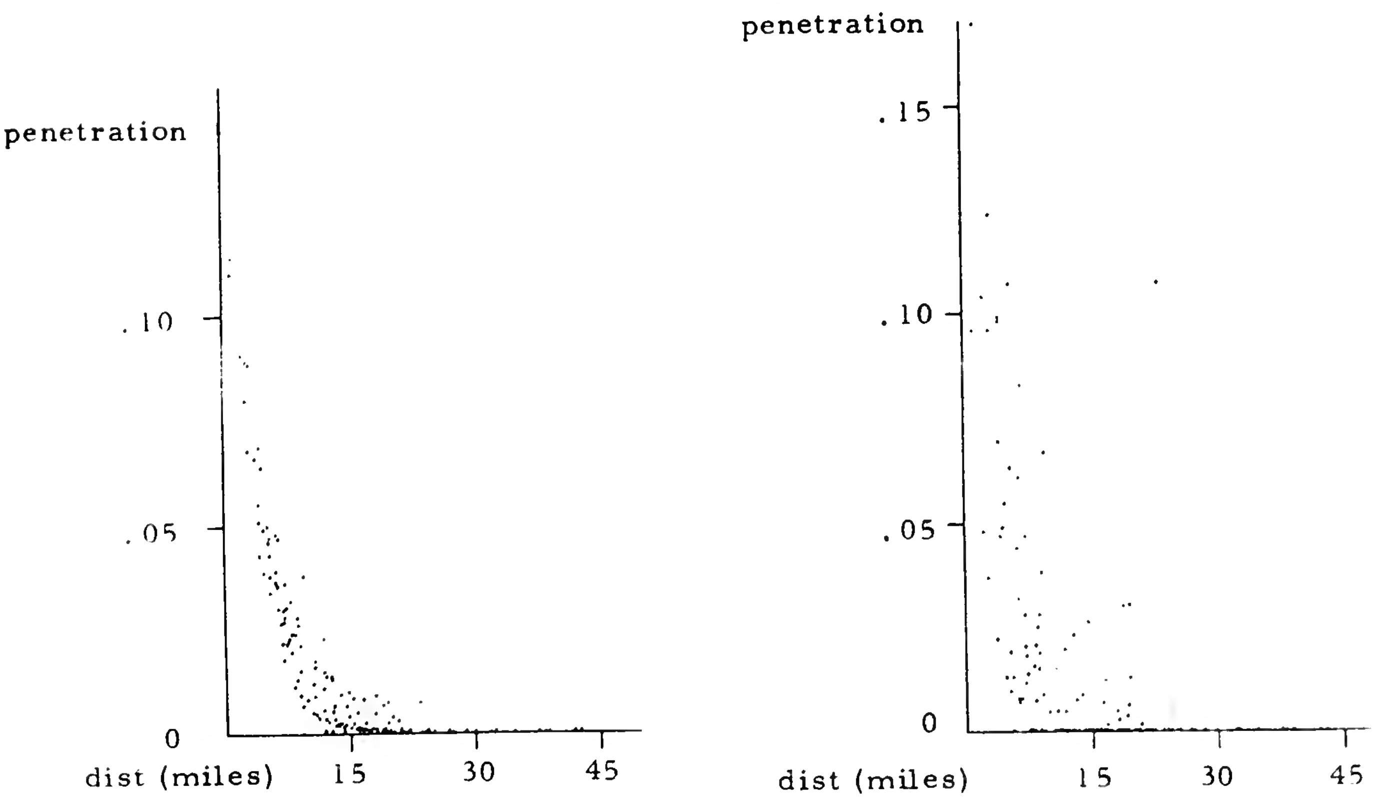


Figure 2.

Penetration vs. Distance from Dealer
for a Large Chevrolet Dealer



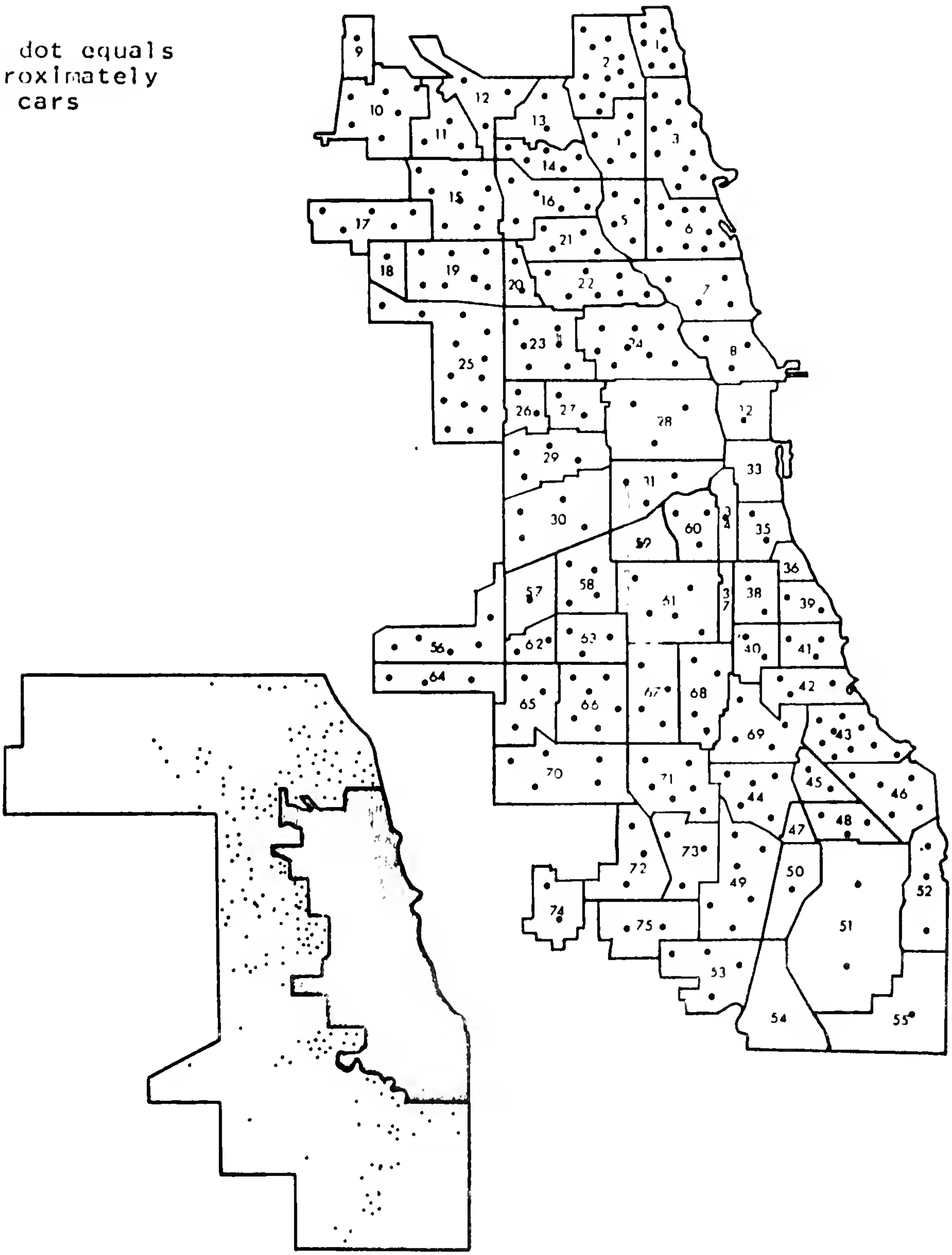
2a. Predicted Penetration
vs. Distance

2b. Actual Penetration
vs. Distance

Figure 3.

POTENTIAL DENSITY

one dot equals
approximately
100 cars



For a better appraisal of fit, we have hypothesized that, for a given dealer, the probability of purchase in cell i is not p_i but is p_i' where

$$p_i' = p_i(1 + \epsilon_i)$$

and ϵ_i is a random error term. The standard deviation, σ of ϵ_i may be viewed as a type of a coefficient of variation. The value of σ has been estimated for each dealer and has a median value of .32. This seems quite good, considering how few factors are included in the model.

4. An Interactive Computer System

Conceivably, one could use the model to work out a mathematically optimal pattern of dealers over the city or, more modestly, the optimal location of a new dealer. Such an optimization is probably sterile. A decision on a dealership involves many factors not included in the model: the availability of property, financing, the micro-geography of the location, etc. Perhaps some of these factors can be modeled but as of now they are not. Yet we do not want to wait to take advantage of what we can learn about macro-geography and competitive interaction. What is needed is a convenient way to make the information available to a person working on dealership problems.

To demonstrate how this can be done, an interactive computer system was programmed for the model on the Project MAC time-shared computer at MIT. With the system a user can sit at a remote console of the computer, make hypothetical changes in dealerships, and learn immediately the model's prediction of the effects. See Exhibit I for an example of the system in action. Changes that a user can make include adding a dealer, moving a dealer, eliminating a dealer, changing a dealer's a and b parameters, and changing the potential of a cell.

5. Possible Improvements

Experimentation with the functional form of the distance relations would be of interest. Similarly, different measures of "distance" might be tried. Travel time has strong intuitive appeal. Much better market segmentation is possible. The natural one would be to break down the cell population by make-model year of car owned. Then make preference could be related to the buying rates of each segment for each make. Another subject for investigation is the effect of clustering of dealers. Perhaps there is a special advantage to being in an "automobile row" because the row itself attracts customers. If so, the model might be modified to account for this. A highly desirable line of research would be to investigate how the a and b parameters are related to various observable characteristics of the dealer.

Reference

1. J. W. James, "An Analysis of the Optimum Market Representation Policies Relating to a Large Metropolitan Market," MS Thesis, MIT, 1964.

EXHIBIT I
A TYPICAL CONSOLE SESSION

AN INTERACTIVE MODEL OF THE CHICAGO AUTOMOBILE MARKET T. E. HLAVAC, JR. J. D. C. LITTLE	-the computer prints a title (User commands are always in lower case, computer responses in upper case)
BEGIN PLEASE	-a signal that initialization is complete and the user may begin
move dlr 222 to cell 131	-the user moves a dealer from his present location
DODG DEALER NUMBER 222 HAS BEEN MOVED TO CELL 131.	-the system executes the instruction,
CONTINUE	and requests the user to continue
change for dlr 222	-the user wishes to know how this move affects the dealer's performance
DODG DLR 222 — SALES PENETRATION FORMER SYSTEM 82 .17126 PRESENT SYSTEM 73 .15391 NET LOSS 9 .01735	-the dealer suffered a net loss by the move,
CONTINUE	
begin	-so the user returns the system to its initial state
CONTINUE	-ready for a fresh start
potential of cell 54, pot = 600	-the user wishes to increase the potential of a cell, possibly because of an anticipated new housing development
THE POTENTIAL OF CELL 54 HAS BEEN CHANGED FROM 8 UNITS TO 600 UNITS. THE TOTAL POTENTIAL OF THE SYSTEM IS NOW 48262 UNITS.	-the result of the change
CONTINUE	
change in dodg sales	-how did the increase in potential affect Dodge sales?
8 DODG DEALERS — SALES PENETRATION FORMER SYSTEM 1119 2.34671 PRESENT SYSTEM 1128 2.33794 NET GAIN 9 -.00877	-Dodge dealers obtained 9 of the 592 new units
CONTINUE	
add dodg in cell 54, a = 4.0, b = -.32	-an additional Dodge dealer is added in the newly developed cell
THIS DEALER HAS BEEN ASSIGNED NUMBER 226. THERE ARE NOW 9 DODG DEALERS.	-the result
CONTINUE	

EXHIBIT I

(continued)

sales of dlr 226

-the user wishes to know the sales of the dealer he has just added

SALES OF DODG DEALER NUMBER 226 ARE 337 UNITS
FOR A PENETRATION OF .69870 PERCENT.

-the answer

CONTINUE

change in dodg sales

-the user realizes that some of the 337 units represent sales formerly belonging to other Dodge dealers. How did Dodge fare on the whole?

9 DODG DEALERS	SALES	PENETRATION
FORMER SYSTEM	1119	2.34671
PRESENT SYSTEM	1459	3.02305
NET GAIN	340	.67634

-the net gain of 340 units comprises 9 units gained by an increase in potential and 331 units gained by the addition of a new dealer. Other Dodge dealers lost a total of 6 units to the new dealer

CONTINUE

report

-what has been the effect on each of the other makes?

MAKE	--------SALES--------			--------PENETRATION--------			DEALERSHIPS			SALES PER DLR		
	FORMER	PRESENT	CHANGE	FORMER	PRESENT	CHANGE	FMR	PRS	CHG	FMR	PRS	CHG
FORD	8137	8164	27	17.07010	16.91545	-.15465	38	38	0	214	215	1
CHEV	20907	21016	109	43.85713	43.54614	-.31099	47	47	0	445	447	2
RAMB	1010	1020	10	2.11928	2.11266	-.00662	8	8	0	126	127	1
PONT	7307	7355	48	15.32770	15.24008	-.08762	28	28	0	261	263	2
PLYM	500	501	1	1.04939	1.03752	-.01187	5	5	0	100	100	0
BUIC	4862	4893	31	10.19915	10.13804	-.06111	24	24	0	203	204	1
OLDS	3828	3854	26	8.02921	7.98574	-.04347	26	26	0	147	148	1
DODG	1119	1459	340	2.34671	3.02305	.67634	8	9	1	140	162	22
TOTAL	47670	48262	592				184	185	1	259	261	2

CONTINUE

quit

-the user terminates the program

RUN TERMINATED. GOOD-BYE.
R 27.550+11.133

-an exit message from the program. The session has used 28 seconds of running time and 11 seconds of swap time. Cost at $10/minute comes to about $6.50