COWLES FOUNDATION DISCUSSION PAPER NO. 36



# Consumer Response to Innovation: Television*

Thomas F. Dernburg

July 2, 1957

---

* This paper is a shortened version of a dissertation presented for the degree of Doctor of Philosophy at Yale University. I would like to express my sincerest thanks to the members of my advisory committee--Professors James Tobin, Harold Guthrie, and Charles Berry--for their many suggestions and their painstaking advice.

---

## Consumer Response to Innovation: Television

### I. Introduction

This is a study of consumer purchases of television receivers between the years 1946 and 1950. As such it fits into an almost totally neglected area of the economics of consumption. Little is known about the reaction of consumers to the appearance of a new produce.[1] Despite this fact, it is clear that a

---

1 It is not necessary to cite the many well known works on the pure theory of consumer behavior: consumer surplus; statistical demand studies; and aggregative consumption. Few of the studies except J. S. Duesenberry, Income, Saving, and the Theory of Consumer Behavior, (Cambridge, 1952); C. F. Roos and V. von Szeliski, "Factors Governing Changes in Domestic Automobile Demand," in The Dynamics of Automobile Demand, (New York, 1939); and P. de Wolff, "The Demand for Passenger Cars in the United States," Econometrica, Vol. 6, No. 1., are relevant to the question of consumer response to innovation.

While there have been a number of efforts to analyze TV demand in a number of cities, the analyses were, for the most part, conducted for purposes other than studying the time rate of consumer absorption of the product. Of interest are L. Saxon Graham, Selection and Social Stratification, Ph.D. Thesis, Sterling Memorial Library, Yale University (1951); "TV Today: Report II," further data from the NBC-Hofstra Study, Television Today: its Impact on People and Products (New York, 1952); C. A. Siepman, Radio, Television and Society (New York, 1950).

---

number of important economic problems and well known economic theories depend significantly on the ability to deal with the question of the response of consumers to innovation. A few of these problems and theories may be suggested.

It was Schumpeter's view that both economic growth and cyclical fluctuations are the result and manifestation of innovation.[2] The process of "creative destruction"

---

2 J. A. Schumpeter, The Theory of Economic Development (Cambridge, 1949).

revolves about the introduction of new innovations of a type which both increase the efficiency of producing existing goods and displace old products in capital and consumption markets. Within the scope of this theory, the rate of economic progress, as well as the severity of cyclical disturbances necessarily depend upon the nature of the innovations and the rapidity of the spread of acceptance. It follows that a partial explanation of the problem of economic fluctuations, the rate of growth of the economic system, and the rate of obsolescence of old methods and old commodities, may depend to a large extent upon the rate at which consumers are willing to adjust their preference fields. Lack of knowledge as to how this process comes about is a serious hiatus in economic knowledge.

The fact that the consumption function in the United States has exhibited a steady secular upward drift is well known. One hypothesis which has been suggested as an explanation of this drift is that it is attributable to the steady introduction of new commodities.[3] To give operational meaning to this

---

3 Duesenberry, op. cit., pp. 58-61.

---

hypothesis it is necessary to determine to what extent different new goods are additions to aggregate demand. Verification of the hypothesis, and prediction of the effects on aggregate demand, depend upon the study of a number of new consumer goods and, particularly, on the rate of absorption of the new goods.

In a full employment economy, if a new consumer good is introduced, resources will be diverted from other uses into the production of the new good if the resources required are greater than those released by the rate of growth of productivity.

As the new commodity gains acceptance its subsitutes may lose ground, if not absolutely, then relatively. The speed of this acceptance and the nature of the substitutes are important factors in determining changes in the pattern of resource use.

Finally, it is suggested that economists can perform a valuable service if their studies of new products help remove some of the uncertainties associated with prospective innovations and some of the maladjustments associated with innovation and the relocation of productive resources.

A basic premise of this inquiry is that an examination of the various characteristics of new commodities and the consumer groups who purchase them will assist in predicting consumer response to similar new commodities. A knowledge of the sales level which can be expected at different points in time presents the means whereby the effects of new commodities on industrial growth, on economic stability, on the distribution of income, and on the other issues which concern economists, can be analyzed.

## II. Hypotheses

For statistical reasons discussed in Appendix C, the dependent variable used for the purpose of fitting regressions is a linear transformation of the logistic function:

$$\hat{p} = \frac{1}{1 + e^{-(a + \sum_{i=1}^{n} b_i x_i)}}$$

where $\hat{p}$ is percentage TV ownership and $x_i$ is the i-th of a set of $n$ independent variables. The general shape of the function is that of the cumulative normal curve. As shown by Roos and von Szeliski and also by de Wolff[4] in their respective

4 Roos and von Szeliski, op. cit., and de Wolff, op. cit.

studies of automobile demand, a function with the same general shape as the logistic function has economic meaning when applied to the problem of the growth of demand for a new consumer durable good. In the view of Roos and von Szeliski, automobile sales $dC/dt$ depend in part on the potential market, $(M-C)$, where $M$ is the saturation level and $C$ is the present stock of automobiles owned by the public, and also upon $C$ independently. When $C$ is very low, potential sales are high but do not materialize at a rapid rate because the product is in its experimental stages and therefore few consumers are bold enough to be "innovators." Gradually, however, consumer resistance is overcome as it becomes evident that purchases are being made by other persons. The rate of increase of sales with respect to the stock of ownership rises until one-half the potential market is covered. Subsequently, the rate of growth of ownership declines as the decline of the potential market, $(M-C)$, overcomes the positive effect of increasing general acceptance.[5]

---

5 Let, $S = aC(M-C)$

where $S = dC/dt$, the rate of sales, and let $a$ be an arbitrary constant greater than zero.

$$dS/dC = a(M-2C) .$$

Consequently, if

$$dS/dC > 0 ,$$

$$C < .5M .$$

If $C$ is expressed as a percent of $M$, and if $M$ is 100% ownership, the rate of increase of ownership will be at a maximum at the ownership level of 50%.

Applied to TV ownership, the Rcos and von Szeliski hypothesis suggests a function of percentage TV ownership with respect to time and other independent variables the shape of which is closely approximated by the logistic function. Consequently, if hypotheses are stated as functions of a linear transformation of the actual growth curve, they must be interpreted to imply that the relationships depend upon the level of TV ownership, relative to saturation, which already obtains. The linear transformation, X, of the logistic function is $\ln(\hat{p}/(1-\hat{p}))$, where $\hat{p}$ is percentage TV ownership in an area. The preliminary hypotheses to follow are all stated in terms of X, the "logit" of $\hat{p}$, and are to be interpreted as partial relationships.

1. Up to a certain income level TV ownership in an area is expected to increase as income increases. However, TV is expected to be an inferior good and, as a consequence, ownership among middle income groups will be higher than among high income groups because TV provides middle income groups with the means with which to save on other entertainment expenditures. Symbolically,

$$X = a_o + a_1 Y$$

where Y is median personal income received by families and unrelated individuals in an area. To this expression the interaction hypothesis:

$$a_1 = a_2 - a_3 Y$$

is added. The relationship between X and Y will therefore be the parabolic function:

$$X = a_o + a_2 Y - a_3 Y^2 .$$

2. TV ownership in an area will increase as the length of time the area has been exposed to TV coverage increases and as the number of available signals increases.

Accordingly,

$$X = b_o + b_1 Ta + b_2 Ts$$

where Ta is the age of the oldest station in the area, and Ts is the number of available singals. Because X is stated as a linear relationship with respect to Ta and Ts, a saturation level of 100% ownership is implied. It is reasonable to suppose, however, that "relative" saturation levels short of 100% may be reached. For example, where television coverage is poor, there may be no combination of other factors which will bring a portion of the potential market over the threshold between non-ownership and ownership. Similarly, the n-th signal established in an area may be expected to have a smaller impetus to increased ownership than the introduction of the preceding signal. These possibilities may be summarized by:

$$b_1 = b_3 - b_4 Ta$$

and,

$$b_2 = b_5 - b_6 Ts \ .$$

Combining the expressions:

$$X = b_o + b_3 Ta - b_4 Ta^2 + b_5 Ts - b_6 Ts^2 \ .$$

The proviso that dX/dTa and dX/dTs remain greater than or equal to zero over the range covered by the data must be added because it is not reasonable to suppose that decreasing values of X could possibly be associated with increasing values of Ta and Ts .

3. TV ownership will be inversely related to the educational level E attained by persons on the grounds that:

   a. It has always been, and still remains, a common complaint among highly educated persons that TV offers little in the way of serious high level entertainment.

b. Highly educated persons fear the effects of TV on their children.

c. Highly educated persons are apt to prefer activities in which they can participate actively to "canned" entertainment.

It is reasonable to suppose, however, that the objections to TV exhibited by highly educated persons may break down over time, particularly as program choice, reflected by an increase in the number of singals, increases. Symbolically,

$$X = c_o - c_1 E$$

and,

$$c_1 = c_2 - c_3 Ts$$

which combine to yield:

$$X = c_o - c_2 E + c_3 ETs \ .$$

4. In areas where median incomes are high, TV ownership will be associated in an inverse way with the dispersion of income. The reverse will be the case where median incomes are low. These hypotheses are offered on the presumption that:

a. The lower the dispersion of income in high median income areas, the more persons, relative to the total in the area, will have the means with which to purchase TV.

b. The lower the dispersion of income, the greater will be the degree of social homogeneity in the area and consequently the greater the social pressure to emulate neighbors.

c. Where median income in an area is low, a higher dispersion of income indicates that more persons in the area will have incomes great enough to afford set purchases than if the dispersion is low.

Symbolically, these hypotheses may be written:

$$X = d_o - d_1 Y_q$$

where $Y_q$ is the relative quartile deviation of income in an area, and:

$$d_1 = d_2 - d_3 Y .$$

Consequently,

$$X = d_o - d_2 Y_q + d_3 Y Y_q$$

5. Urban dwellers will own fewer sets than suburban dwellers because alternative forms of entertainment will be more easily accessible. Accordingly;

$$X(A) > X(U)$$

where $X(A)$ and $X(U)$ are the "logits" of percentage set ownership in "adjacent" and "urban" areas respectively.

## III. Data and Variables.

The study was made possible by the happy circumstance that in 1950 the Bureau of the Census included in its Tract Statistics[6] information on television ownership.

---

6 U.S. Department of Commerce, Bureau of the Census, 1950 Population Census Report, Vol. III, "Census Tract Statistics."

---

A Census Tract is a contiguous geographic area, varying greatly both in area and in population. The number of dwelling units in a tract varies from less than five-hundred to more than ten thousand. Tracts are classified by the Census into "urban" and "adjacent" areas. These categories make it possible to make urban-suburban comparisons. Three tables of information are reported for each tract. Tables I and II present breakdowns of the sex, race, and national origin characteristics as well as income and education distributions. Table III reports household characteristics including television ownership. With respect to TV ownership, households were classified as "yes," "no," or "no response." In 1950 there were approximately 38

cities for which data on TV reception were available.

The project derives much of its interest from the fact that in April 1950, the date of enumeration, different areas had been exposed to different lengths of time and intensities of TV coverage. For example, Philadelphis, Pennsylvania was covered by three stations which, among them, had accounted for a total of 108 station months. On the other hand, Toledo, Ohio had only one station, which had been in operation for 21 months. If tracts in Toledo and Philadelphia that are otherwise similar can be found, the differences in ownership in April 1950 caused by differences in TV broadcasting histories can be observed. As a consequence, the combination of the Census Tract statistics with the history of TV coverage in different cities makes it possible to study the introduction and spread of TV as well as the factors governing the demand for TV at any one time.

A. Stratification

Tract statistics were published for sixty-two cities. Of these, twenty-four are not used either because they had no TV signals in April 1950 or because the published data were not available when the study was undertaken. The remaining thirty-eight cities are divided into three television coverage classes. Class I cities had only local signals in April, 1950. Class II cities had only unambiguous-i.e., originating within a fifty mile radius of the city-"foreign" signals. Class III cities had at least one local signal and in addition had one or more ambiguous (fifty to seventy-five mile radius) or unambiguous foreign signal.

For reasons of computation feasibility, the sample was arbitrarily limited to 3,000 of the 7,419 eligible tracts and was stratified in three ways. The first stratification was the division of cities according to the three classes of TV coverage. The second principle of stratification was to over-represent "adjacent" areas.

The sample ratio of "urban" to "adjacent" tracts is two to one, compared with a more than three to one ratio for all eligible tracts. The final stratification was by median income. To facilitate comparison between the various strata, an effort was made to represent them as evenly as possible. In strata where selection was required, the method used was to choose every n-th tract, in each city, following the Census tract number. This procedure was modified to avoid, in so far as possible, retaining geographically adjoining tracts.

B. Variables.

A complete listing of the dates of station introduction in all tracted areas was gathered from Television: The Business Magazine of the Industry.[7] This

---

7 Television: The Business Magazine of the Industry, vols. 3-7, 1946-1950.

---

information made it possible to define four variables which describe the time shape and intensity of TV coverage in different cities. The age of the oldest station Ta and the number of signals Ts as of April, 1950 give a fair measure of the coverage of a city. However, set ownership will probably also depend on other dimensions of the history of coverage. For example, set ownership may differ in two cities which had equal values for Ta and Ts if, in one city, most of the signals had existed for some time, while in the second city, all but the first station had been recently introduced. This possibility can be taken into account either by Tt, the average age per station, or by T, the total number of television station months.

The remaining variables were obtained from the Tract Statistics. Summary measures such as median income, median education, and median number of persons in dwelling units could be obtained directly from the published data. In other cases, minor preliminary computations were necessary. Some information, for example television ownership, is reported by the Census as number reporting TV ownership, and number reporting in the affirmative. It is thus necessary to divide the number owning TV by the number reporting ownership to obtain the desirable variable of percentage TV ownership. In still other cases, as for example the relative quartile deviation of income, it was necessary to calculate the first and third quartile points from the frequency distribution of income which was reported.[8]

---

8 The methods of calculating quartile points and estimating points in open-end classes are discussed in Appendix: A.

---

It is useful to have a complete list of the variables used in the study and to identify them symbolically. Accordingly, we define,

Ta - the age of the oldest station, local or unambiguous foreign, in months dating backwards from April, 1950.

Ts - the number of signals, local or unambiguous foreign, available in April, 1950.

T - the total number of local or unambiguous foreign television months dating backwards from April, 1950.

Tt - the average station age as of April, 1950. $Tt = T/Ts$ .

E - median education of persons 25 years old and over in years.

Y - median personal income in 1949 received by families and unrelated individuals.

$Y_q$ - the relative quartile deviation of income.

A - median age of the male population between the ages of 14-64.

Ac - the percent of the male population under 14 years of age.

Aa - the percent of the male population over 64 years of age.

N - the percent of the total population, non-white.

M - the percent of the male population, married.

F - the percent of the total population, female.

Flf-the percent of the female population over 14 years of age in the labor force.

OO - the percent of occupied dwelling units occupied by their owners.

C - the percent of those reporting who have more than 1.01 persons living in one room.

DU - the median number of persons in dwelling units.

$\hat{p}$ - the percent reporting TV ownership who own sets.

X = $\ln(\hat{p}/(1-\hat{p}))$, the "logit" of percentage TV ownership.

## IV. Testing of Hypothesis: Television Coverage and Income.

Multiple regression analysis is used to test the income and coverage hypotheses. With respect to the TV coverage variables multiple regression analysis is the most useful technique to employ because different combinations of several variables may be tried in order to obtain the best possible fit. In the case of income, the regressions employed in testing the hypotheses of this section will later be used to calculate residuals which in turn serve as the dependent variable for the analysis of subsequent sections. In this way the effects of the high correlations between income and other independent variables are removed. The analysis is, for the time being, confined to Classes I and III. Class II is a small class of only four cities, where, because a particular value of one coverage variable is associated with a unique value for all other coverage variables, it is obvious that meaningful regressions, containing more than one of the coverage variables, cannot be calculated.

### A. TV Coverage Hypotheses

In addition to simple testing of the coverage hypotheses, it is desirable to find the appropriate combination of independent variables for the estimating equations of

Table I

Class I Regression Coefficients, Explained Variations ($S^2$), Coefficients of Multiple Co-Determination ($R^2$), and t ratios (bracketed values) for Regressions Relating Income and Television Coverage to X, the "logit" of Television Ownership*

| Code | Independent Variables | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Constant | Y | $Y^2$ | Ts | Ta | $T_t$ | T | $Ts^2$ | $Ta^2$ | $Tt^2$ | $S^2$ | $R^2$ |
| I:a | -4.537 | .6818 (20.72**) | -.0444 (12.58**) | .0384 (4.61**) | .0298 (14.87**) | .0032 (1.07) | x | x | x | x | 1,342,283 | .6951 |
| I:b | -4.501 | .6772 (20.76**) | -.0441 (12.55**) | .0366 (4.49**) | .0314 (11.47**) | x | x | x | x | x | 1,341,789 | .6948 |
| I:c | -4.467 | .6830 | -.0046 | .0013 | .0307 | x | .0011 | x | x | x | 1,329,113 | .6882 |
| I:d | -4.010 | .8176 | -.0558 | x | x | x | .0053 | x | x | x | 1,128,171 | .6463 |
| I:e | --5.058 | .6802 ( 20.60**) | -.0441 (12.43**) | .2028 (2.10*) | .0059 (0.31) | .0463 (2.94**) | x | -.0162 (1.63) | .0001 (0.65) | -.0009 (2.60**) | 1,347,132 | .6976 |

* A single asterisk accompanying a t-ratio denotes significance at the 5 percent level. Double asterisks denote significance at the one percent level. An x indicates that the particular variable was not included in the regression. The total number of independent observations is 1,373. The total variation about the sample mean of X is 1,931,184.

Table II

Class III Regression Coefficients, Explained Variations ($S^2$), Coefficients of Multiple Co-Determination ($R^2$), and t-ratios (bracketed values) for Regressions Relating Income and Television Coverage to X, the "logit" of Television Ownership*

| Code | Independent Variables | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Constant | Y | $Y^2$ | Ts | Ta | Tt | T | $Ts^2$ | $Ta^2$ | $Tt^2$ | $S^2$ | $R^2$ |
| III:a | -3.712 | .7406 (17.58**) | -.0592 (11.71**) | .0428 (3.00**) | .0071 (1.64) | .0214 (4.00**) | x | x | x | x | 614,549 | .4946 |
| III:b | -3.612 | .7239 | -.0580 | -.0036 | .0256 | x | x | x | x | x | 603,342 | .4856 |
| III:c | -3.699 | .7118 | -.0570 | .0291 | .0303 | x | -.0018 | x | x | x | 603,020 | .4845 |
| III:d | -3.273 | .7726 | -.0609 | x | x | x | .0034 | x | x | x | 553,178 | .4452 |
| III:e | -5.652 | .7201 (19.13**) | -.0555 (12.28**) | .3671 (7.54**) | .0481 (4.35**) | .1259 (6.68**) | x | -.0120 (2.48*) | -.0018 (2.67**) | -.0015 (2.83**) | 737,372 | .5935 |

*A single asterisk accompanying a t-ratio denotes significance at the five percent level. Double asterisks denote significance at the one percent level. An x indicates that the particular variable was not included in the regression. The total number of independent observations is 1,4000. The total variation about the sample mean of X is 1,242,431.

Classes I and III. The process of finding these variables is broken down into two stages. The first step is to try different combinations of the linear forms to determine if all four coverage variables are independently significant. Non-linear forms are then added to test the relative saturation hypotheses and to see if they yield an improvement in fit. Because the correlation between the coverage variables is high, finding the appropriate variables requires a rather lengthy process of trial and error in which many different combinations of independent variables were tried. The complete process need not be described. A brief review of the links in the chain is, however, instructive.

T is a composite variable which reflects the length of time TV has been available, the intensity of coverage, and the time shape of coverage. Ts and Ta may be assumed to be independent of each other,[9] the first describing the intensity

---

9 This is, strictly speaking, not true. In areas where Ts is large, Ta also tends to be large.

---

of coverage in April, 1950 and the latter describing the length of time TV has been available to the area. Because Ts and Ta are viewed as being complementary, the question is to determine whether the combination of Ts and Ta yields a better explanation than T alone, and also whether T makes a significant addition to Ta and Ts . The regressions I:b, I:c, I:d, III:b, III:c, and III:d of Tables I and II were computed in order to answer these questions. Comparison of the $S^2$ values of the b and d regressions in both classes clearly shows that Ta and Ts yield better results than T alone. Testing for the significance of the addition of T to Ta and Ts by subjecting the differences $S^2_c - S^2_b$ to F tests, yields insignificant F ratios of 1.10 and 3.75 in Classes I and III respectively. As a conse-

quence of these tests it seems clear that T can be dropped from the analysis.

The combination of Ts and Ta do not give a complete description of the time shape of TV coverage. For example, a difference in percentage ownership would undoubtedly exist between two areas were Ta is 25 months and Ts is equal to 3, if in one area all three stations were introduced 25 months ago while in the other area all but the first station are one month old. As a consequence of this possibility Tt was tried together with Ta and Ts . In the resulting regressions (I:a and III:a) Tt is not significant in Class I but is significant in Class III where, however, the addition of Tt has made the coefficient of Ta insignificant. From these results it can be concluded that, with the data at hand, the added time shape dimension does not improve the explanation of TV ownership.

Regressions I:a and III:a were used to compute the residuals which are later employed as the dependent variable in variance and regression analysis conducted for the purpose of testing remaining hypotheses and supplementing the results with analysis of additional independent variables.

The final step in the process of estimating TV ownership in terms of the coverage variables is to add the non-linear forms $Ts^2$, $Ta^2$, and $Tt^2$ in an effort to test the relative saturation hypothesis. The results of the additions are the regressions I:e and III:e. In Class III the addition of the three higher order terms improves the fit markedly. The additional variation explained by adding these variables is 122,823 which amounts to 9.8 percent of the total variation in Class III. $R^2$ is raised from .495 to .593. The improvement in fit is such that all independent variables are significant at the 5% level and only $Ts^2$ is not significant at the 1% level. In Class I however, there is almost no improvement. The additional variation explained by adding the three new variables $(S^2e - S^2a)$ yields a barely significant F ratio of 3.82 and lifts the value of $R^2$ by about one-fourth of one percent. As a consequence

it can be concluded that the relative saturation hypothesis cannot be supported for Class I.

Interpretation of the results is facilitated by graphic presentation. For purposes of illustration four mythical tracts, all assumed to have median income values of $4,000, are chosen and four different broadcasting histories[10] are ascribed

10 The particular histories are those of actual cities. A may be identified with St. Louis, Missouri; B with New York, N.Y.; C with New Haven, Connecticut; and D with Detroit, Michigan.

to these tracts. It will simplify the presentation if the particular broadcasting histories are labeled in accordance with the code of Table III. Comparison of the time paths of absorption beteen A and C tracts show differences between Class I

Table III

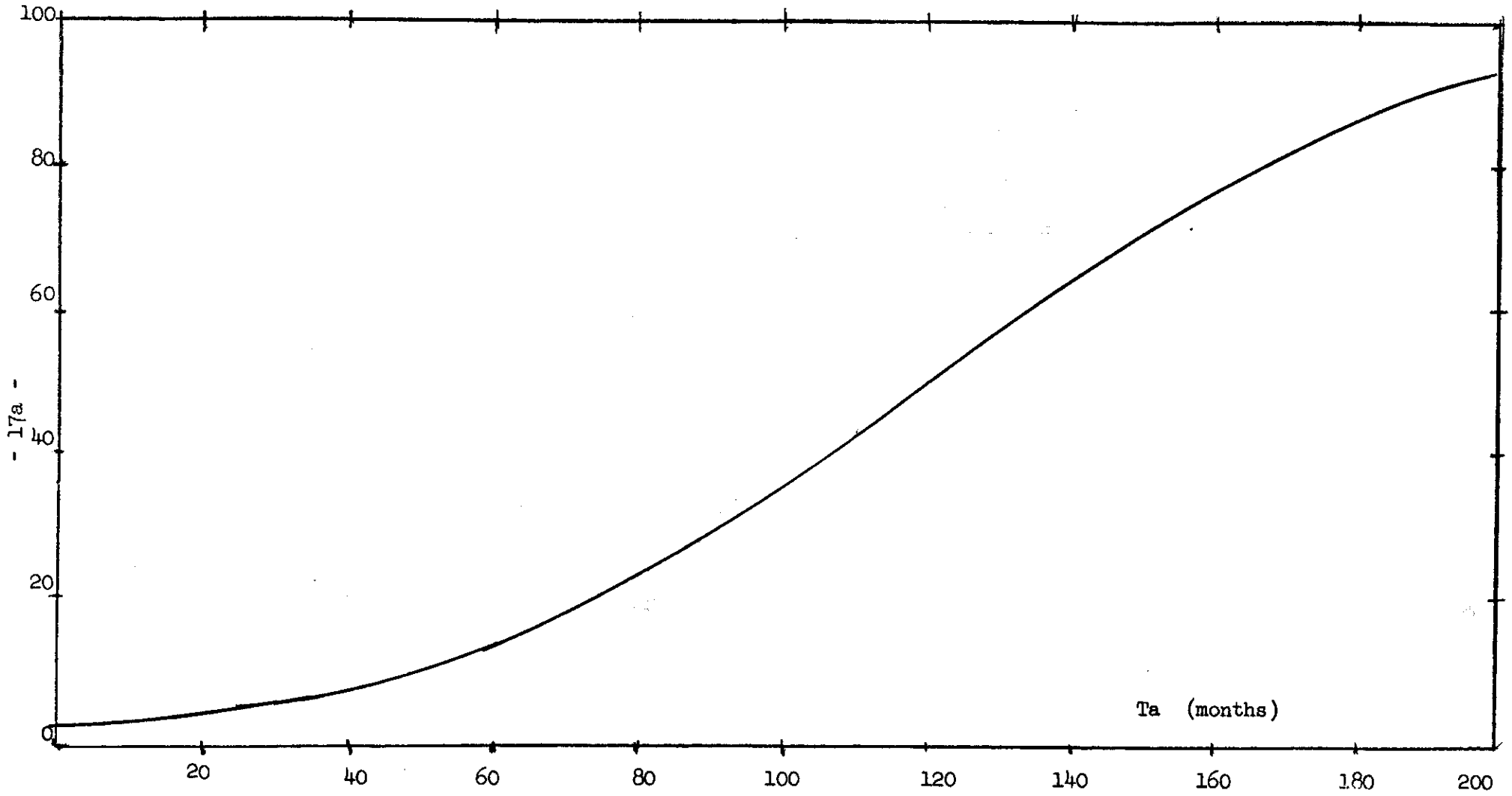
| Code | Television Class | Number of Signals |
|---|---|---|
| A | I | one |
| B | I | several |
| C | III | one |
| D | III | several |

and III areas with one signal. Comparison of B with D shows differences between Class I and III tracts with several signal, and comparison of A with B and C with D shows differences due to different levels of coverage in cities of the same coverage class. The time path of absorption of TV for A is plotted in Chart 1, for C in Chart 2, and for B and D in Chart 3.

A single signal is assumed to have been introduced at t(0) in tract A. Subsequent increases in percentage ownership are associated with the mere passage of time. Because the non linear forms of the coverage variables were not found to be

Chart 1.

Growth of Percentage TV Ownership for a Hypothetical Tract with Coverage A.*





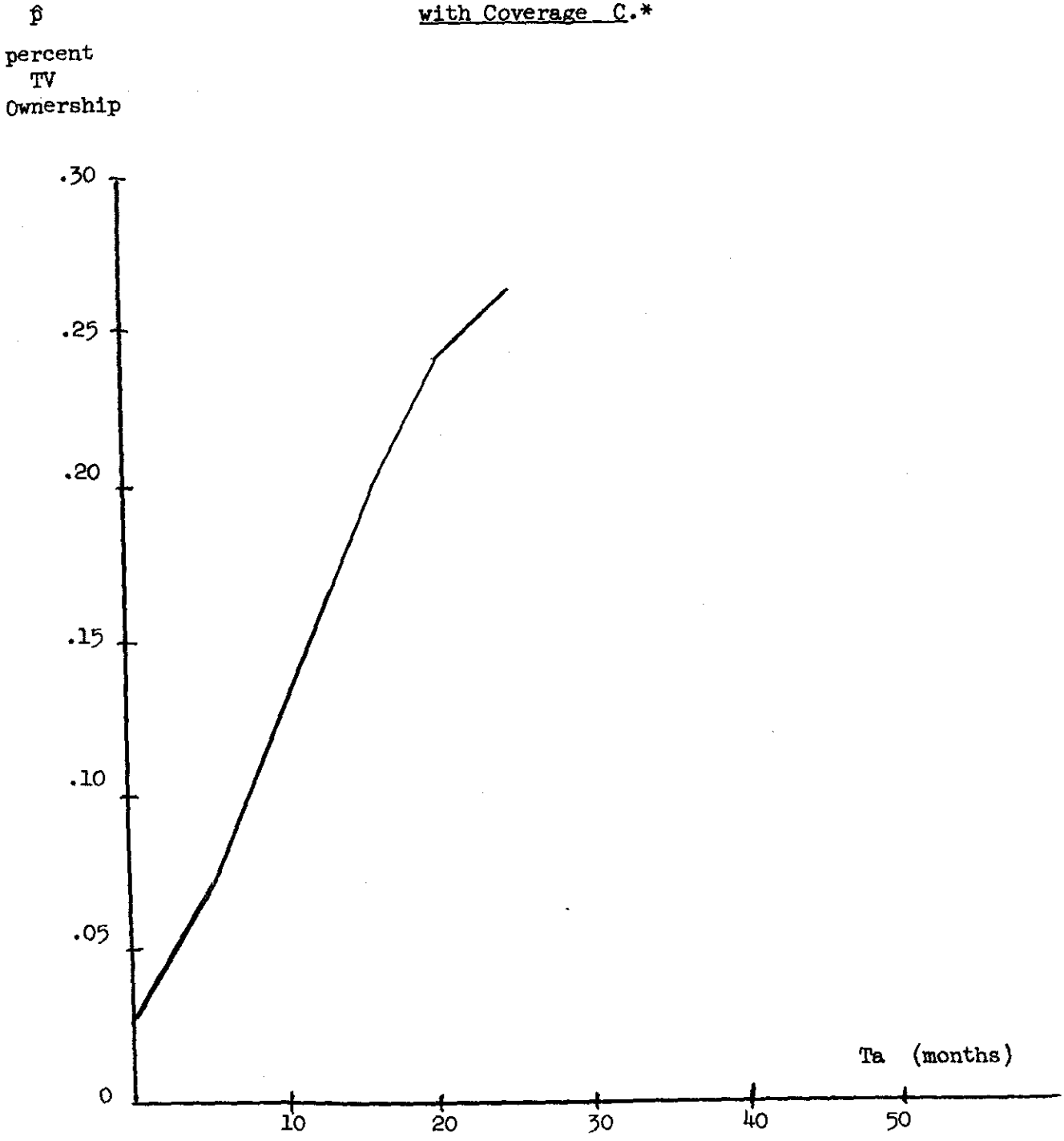

* Y = $4,000, Ts = 1

Chart 2.

Growth of Percentage TV Ownership

for a Hypothetical Tract,

with Coverage C.*



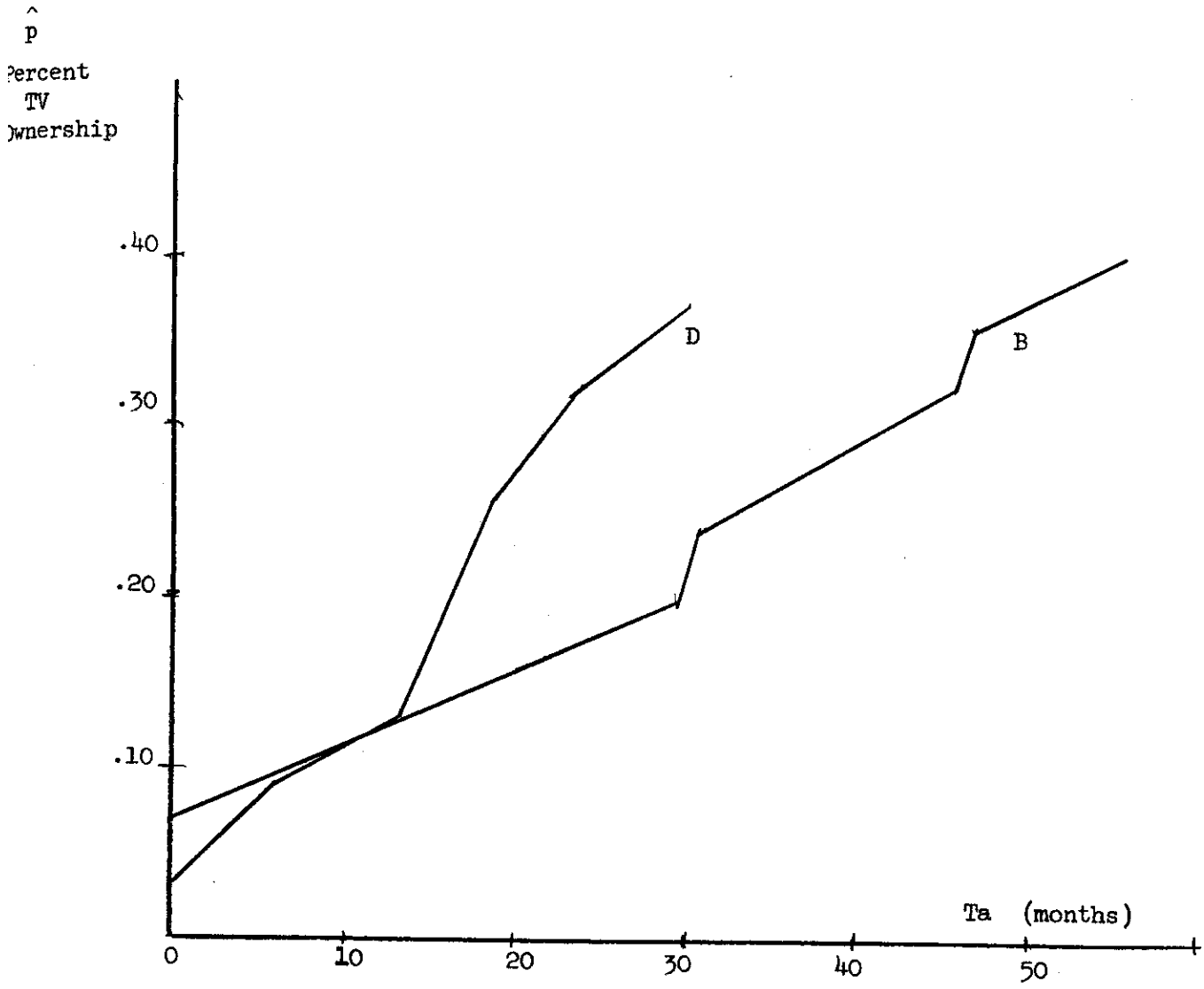

* Y = $4,000, $Ts_1 = 1$

Chart 3.

Growth of Percent TV Ownership

for Two Hypothetical Tracts with

Coverage B and D



significant in Class I cities, the curve which is traced out is a simple continuous logistic with 100% saturation expected in approximately 180 months.

In tract C (Chart 2) a single signal is introduced at t(0) and as in tract A subsequent increases in ownership are associated only with the passage of time. However, in this Class III tract the non-linear coverage variables were significant and thus, in accordance with regression III:e relative saturation is reached and the rate of increase of percentage TV ownership beings to decline after the eighteenth month. Because, in Class III, curves fitted to the data would show declining ownership as time passes beyond thirty months, it is not useful to extrapolate curve C beyond the range of time covered by the data.

Curve B (Chart 3) is the curve appropriate for the Class I tract with several signals, while curve D is the appropriate curve for a comparable Class III tract. Assuming the existence of three signals at the outset in B, estimated percentage ownership at t(0) is .080. Absorption increases to approximately 20% over the next twenty-nine months. In the twenty-ninth month a new signal becomes available. Another new signal is introduced in the thirtieth month and a third in the thirty-second month. The effect of these new signals is to produce a more rapid increase in ownership over the three month period. Subsequently, absorption proceeds at the old rate until the forty-eight month at which time the seventh, and last, signal is introduced. Again there is an increase in the rate of absorption and a subsequent return to the old rate.

The Class III tract with coverage D begins its career with one signal and consequently less ownership at t(0) than tract B. There is, however, a rapid increase in ownership and by the sixth month percentage ownership in D has caught up with ownership in B. Absorption continues at a more rapid rate than in B. However, the diminishing slope of the curve after t(6) indicates that a mild degree of relative saturation has begun to set in. In the thirteenth month a new station is established

in a nearby city and in the sixteenth month time more local stations are established. The effect of this station introduction is to increase the subsequent time rate of absorption to a higher level than was obtained with one signal. By the twenty-fourth month, however, the rate has begun to decline as saturation, relative to the number of signals, begins to set in.

Comparison between Class I and Class III areas is instructive. In general, Class I is dominated by large cities such as New York and Chicago where a steady introduction of new signals helps to shift relative saturation levels up. Because the regressions utilized to estimate absorption in small Class I cities also included the New York and Chicago data, it quite naturally appears that relative saturation is not reached in any Class I city. Initially, TV appears to be greeted much more avidly in Class III cities. This fact is probably due to the circumstance that the average Class III city is smaller than the average Class I city and therefore has less to offer in the way of alternative entertainment opportunities. However, the average number of signals available in the average Class III city is less than in the average Class I city. In addition some of the coverage may be poor because it is external. As a consequence of this relatively poorer coverage, it is not surprising to find relative saturation levels in Class III.

B. Income Hypothesis

That television is an inferior good with respect to income is clearly supported by the available evidence. In the final estimating equations, I:b and III:e it is evident that the coefficients of income and income squared are highly significant. Increasing income is associated with increasing $X$ and $p$ values until income levels of $7,676 and $6,487 are reached in Classes I and III respectively. Beyond

these income levels, additional income is associated with declining percentage ownership. The estimated relationships of income on $p$ for the two classes are traced out in Chart 4.[11]

---

11 The values of the intercepts were obtained by substituting the mean values of the coverage variables in the respective classes. This substitution yields the regressions:

$$I{:}b'X = x - 2.955 + .6772Y - .0441Y^2$$

and,

$$III{:}e'X = -2.684 + .7201Y - .0555Y^2$$

from which the curves in Chart 3 are traced.

---

Before concluding that television is an inferior good with respect to income, the possibility that alternative hypotheses might yield equally good fits must be considered. The most likely hypothesis of this nature would be that increasing percentage ownership is associated with increasing income but at a constantly diminishing rate. In this case, percentage ownership would approach a maximum assymptotically as income increases but would never exhibit an actual decline no matter how great income becomes. The logical function to use in testing this hypothesis would be a logarithmic curve or a hyperbola. It is not necessary actually to fit such curves to compare the differences, because a simple calculation of the mean values of the residuals, $V$, in different income classes, indicates that the regressions I:a and III:a slightly overestimate $X$ for very low and very high income levels and slightly underestimate $X$ for intermediate levels. Consequently, an improvement in fit might be obtained by adding the third order term, $Y^3$, to the existing regression, while substitution of a logarithmic or hyperbolic function would unquestionably harm it.

C. Interactions Between Income and Coverage Variables.

It is evident from Chart 4 that considerable difference exists between the income regressions of Classes I and III. In Class I, ownership is lower than in Class III for income levels up to a level of $7,150 when estimated Class I ownership begins to exceed Class III ownership. As income increases the spread between the classes grows wider. Because cities were classified by the type of TV coverage they had, it is reasonable to expect the differences in income regressions to be traceable to differences in coverage. Such a difference, or interaction, could be taken into account in the regressions by terms of the form YTs and YTa . Analysis of the interaction proceeds by dividing the data into the two coverage classes and conducting analysis of variance[12] with the

12 For a discussion of the variance analysis model used in the analysis see Appendix D.

data in each coverage class divided into a) five income classes and five Ts classes, and b) five income classes and four classes of Ta . The results are summarized in Tables IV and VI.

Table IV.

Analysis of Variance on V for Five Income Classes and Five Classes of Signals. Ts.*

| Source | Variation | | Degrees of Freedom | | F Ratio |
|---|---|---|---|---|---|
| | Explained | Unexplained | Numerator | Denominator | |
| Class I | | | | | |
| Mean | — | 588,901 | — | 1,373 | — |
| 1) Ts | 11,700 | 577,201 | 5 | 1,368 | 5.45** |
| 2) Y | 32,313 | 565,588 | 5 | 1,368 | 16.05** |
| 3) YTs | 55,091 | 533,810 | 25 | 1,348 | 5.57** |
| 4) (3)-(2)-(1) | 20,046 | 533,810 | 15 | 1,348 | 3.40** |

Residual

Chart 5.*

Means and Differences for Two Signal Classes by Income, Class for Coverage Class I.

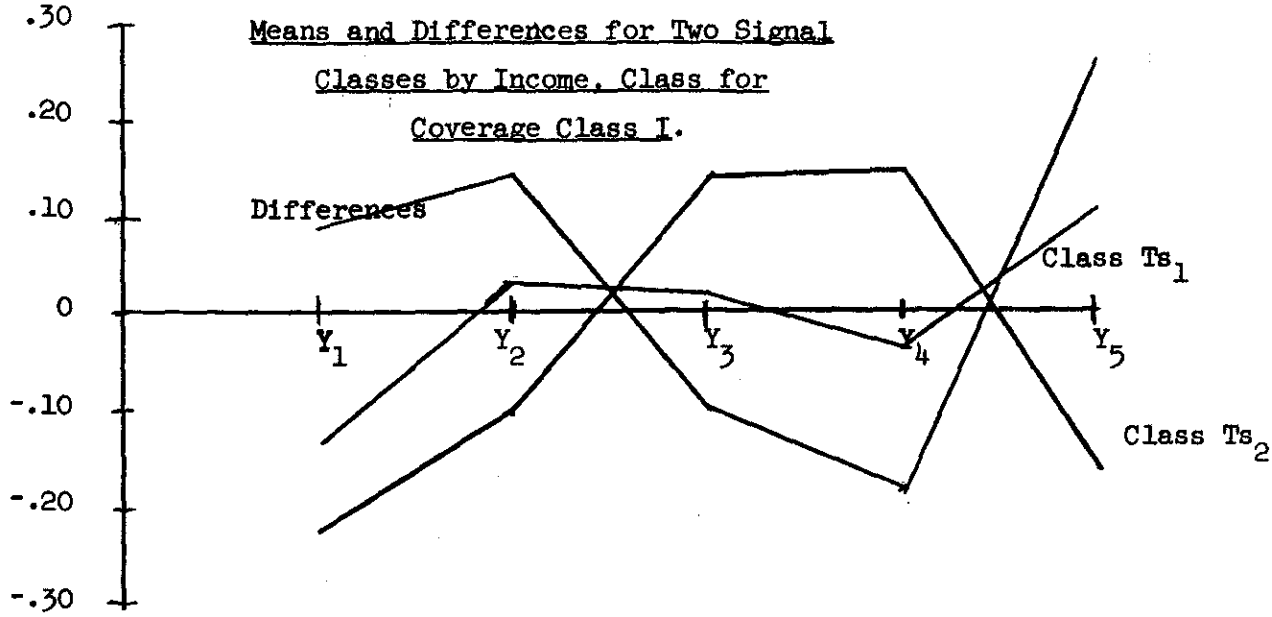
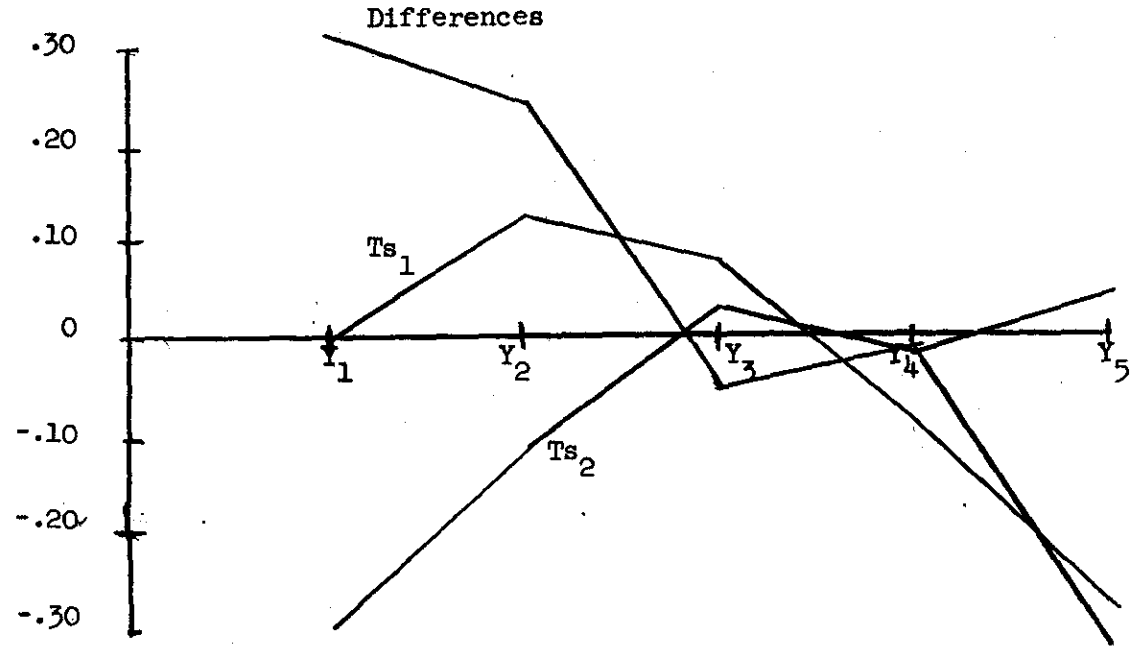



Chart 6.

Means and Differences for Two Signal Classes by Income Class for Coverage Class III

Residual




* $Ts_1$: 1-3 Signals

$Ts_2$: 4-7 Signals.

$Y_1$ 0 - 1,999 Dollars
$Y_2$: 2,000 - 2,999
$Y_3$: 3,000 - 3,999
$Y_4$: 4,000 - 4,999
$Y_5$: 5,000 and over.

Table IV. Continues

Analysis of Variance on V for Five Income Classes and Five Classes of Signals. Ts.*

| Source | Variation Explained | Variation Unexplained | Degrees of Freedom Numerator | Degrees of Freedom Denominator | F Ratio |
|---|---|---|---|---|---|
| Class III | | | | | |
| Mean | — | 627,882 | — | 1,396 | — |
| 1) Ts | 118,635 | 509,247 | 5 | 1,391 | 64.83** |
| 2) Y | 8,235 | 619,651 | 5 | 1,391 | 3.70** |
| 3) YTs | 143,960 | 483,922 | 25 | 1,371 | 16.31** |
| 4) (3)-(2)-(1) | 17,091 | 483,922 | 15 | 1,371 | 3.23** |

* The identical format as the above is used in subsequent analysis of variance tables. The criterion of acceptance used is the one percent level of significance denoted by double asterisks. A single asterisk denotes significance at the five percent level.

The variation explained by Ts (row 1) is high in Class III but quite low in Class I. This result is in line with expectations because the non-linear forms of the coverage variables added little to the explanation in Class I but did add significant amounts to the Class III explanation. The interaction terms are significant in both classes. To simplify the interpretation of the interaction terms the five Ts classes are collapsed into two classes where the first class includes one to three signals and the second class includes four to seven signals. The means of these classes and the differences between these means are computed and entered in Table V and are plotted in Charts 5 and 6.

Table V

Means of Residuals for Five Income and Two Signal Classes.

| Class I | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ | $Y_5$ |
|---|---|---|---|---|---|
| 1) $Ts_1$ | -.1327 | .0365 | .0246 | -.0356 | .1052 |
| 2) $Ts_2$ | -.2284 | -.1036 | .1421 | .1479 | -.1600 |
| 3) Difference | .0957 | .1401 | -.1175 | -.1835 | .2652 |

Table V Continues

Means of Residuals for Five Income and Two Signal Classes.

| Class III | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ | $Y_5$ |
|---|---|---|---|---|---|
| 1) $Ts_1$ | -.0018 | .1232 | .0815 | -.0872 | -.2949 |
| 2) $Ts_2$ | -.3129 | -.1239 | .0256 | -.0020 | .0179 |
| 3) Differences | .3111 | .2471 | -.0559 | -.0052 | -.3128 |

If no interaction exists, the entries in row three would all be of nearly equal value and with the same sign. This is obviously not the case. In Class I the residuals are positive for high income classes which indicates that regression I:a underestimates ownership among high income classes for cities with few signals and overestimates ownership for cities with numerous signals. This difference may arise from the fact that the cities with more signals provide better entertainment alternatives than the cities with only a few signals. This suggests that the degree to which TV is inferior with respect to income depends upon the degree to which alternatives exist. Hot dogs can be inferior to steak only if steak is a possible alternative. If steak is unknown, and if no other meats exist, hot dogs would undoubtedly be considered a choice delicacy. As the alternatives improve, therefore, the more inferior can TV be expected to become.

The upshot of the argument is as follows: Had the interaction between Ts and Y been taken into account in the Class I regression, the coefficient of YTs would have been negative and thus peak ownership would have been reached at a lower level of income and the whole curve would have been rotated slightly downwards about

the intercept.[13]

---

13 When $X = a_o + a_1Y - a_2Y^2$ maximum ownership is reached when $Y = \frac{a_1}{2a_2}$.

Assuming that the effect on the coefficients of $Y$ and $Y^2$ of adding the interaction term is negligible, we have:

$$X = a_o + a_1Y - a_2Y^2 + a_3YTs .$$

Maximum ownership is now reached at,

$$Y = \frac{a_1 + a_3Ts}{2a_2} .$$

If

$$\frac{a_1}{2a_2} > \frac{a_1 + a_3Ta}{2a_2}$$

then,

$$0 > a_3Ts .$$

Consequently, if $a_3$ is negative maximum ownership occurs at a lower income level and if it is positive maximum ownership will occur at a higher income level.

---

It is evident from Table V, that in Class III the interaction effect is the opposite of that obtained in Class I. Regression III:a underestimates ownership for high incomes with few signals and slightly overestimates ownership for high incomes with many signals. It is plausible to suppose that in Class I cities the ratio of TV coverage to alternative entertainments, if such an index can be imagined, decreases as the number of signals increases while in Class III this ratio increases. As a consequence, the coefficients of the interaction term in Class III will be positive--i.e., there will be an opposing tendency to the inferior good effect resulting from better coverage relative to alternatives-and therefore the regression line will rotate slightly upwards thus increasing the income level at which peak ownership occurs.

The combined effects of the interactions in the two classes suggest that when the interaction between Y and Ts is taken into account, the spread between the income regression lines narrows.

The possibility of the existence of an interaction between the passage of time from the date of introduction of the first station, Ta, and income, is considered below. Table VI summarizes analysis of variance for the two coverage classes, in five income classes and four classes of Ta .

Table VI.

Analysis of Variance on V for Five Income Classes and Four Classes of Age of Oldest Station. Ta .

| Source | Variation | | Degrees of Freedom | | F Ratio |
|---|---|---|---|---|---|
| | Explained | Unexplained | Numerator | Denominator | |
| Class I | | | | | |
| Mean | — | 588,901 | — | 1,373 | — |
| 1) Ta | 14,620 | 574,281 | 4 | 1,369 | 6.33** |
| 2) YTa | 50,352 | 538,349 | 20 | 1,353 | 32.04** |
| 3) (YTa)-Ta-Y | 3,419 | 538,349 | 10 | 1,353 | 0.86 |
| Class III | | | | | |
| Mean | — | 627,882 | — | 1,396 | — |
| 1) Ta | 62,417 | 565,465 | 4 | 1,392 | 38,42** |
| 2) YTa | 95,674 | 532,208 | 20 | 1,376 | 12.36** |
| 3) (YTa)-Ta-Y | 25,026 | 532,208 | 10 | 1,376 | 6.47** |

The variation explained by Ta in Class I is quite small while in Class III it is quite large. As with Ts this result is to be expected from the fact that non-linear forms of the coverage variables improved the fit significantly in Class III but did not do so in Class I.

The interaction term is not significant in Class I. On the basis of the present analysis it cannot therefore be asserted that the passage of time will have any effect on the shape of the income regression. In Class III, however, the

interaction term is significant. The nature of the term can be explored by employing the procedure of collapsing the four Ta classes into two classes. The mean and differences of V in each class are entered in Table VII and are posted in Chart 7.

Table VII

Means of Residuals for Five Income and Two Ta Classes

| Class III | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ | $Y_5$ |
|---|---|---|---|---|---|
| 1) $Ts_1$ | -.0846 | .0758 | .0526 | -.0043 | .0155 |
| 2) $Ts_2$ | -.3088 | -.1299 | .1109 | -.1402 | -.1554 |
| 3) Difference | .2242 | .2057 | -.0543 | .1359 | .1709 |

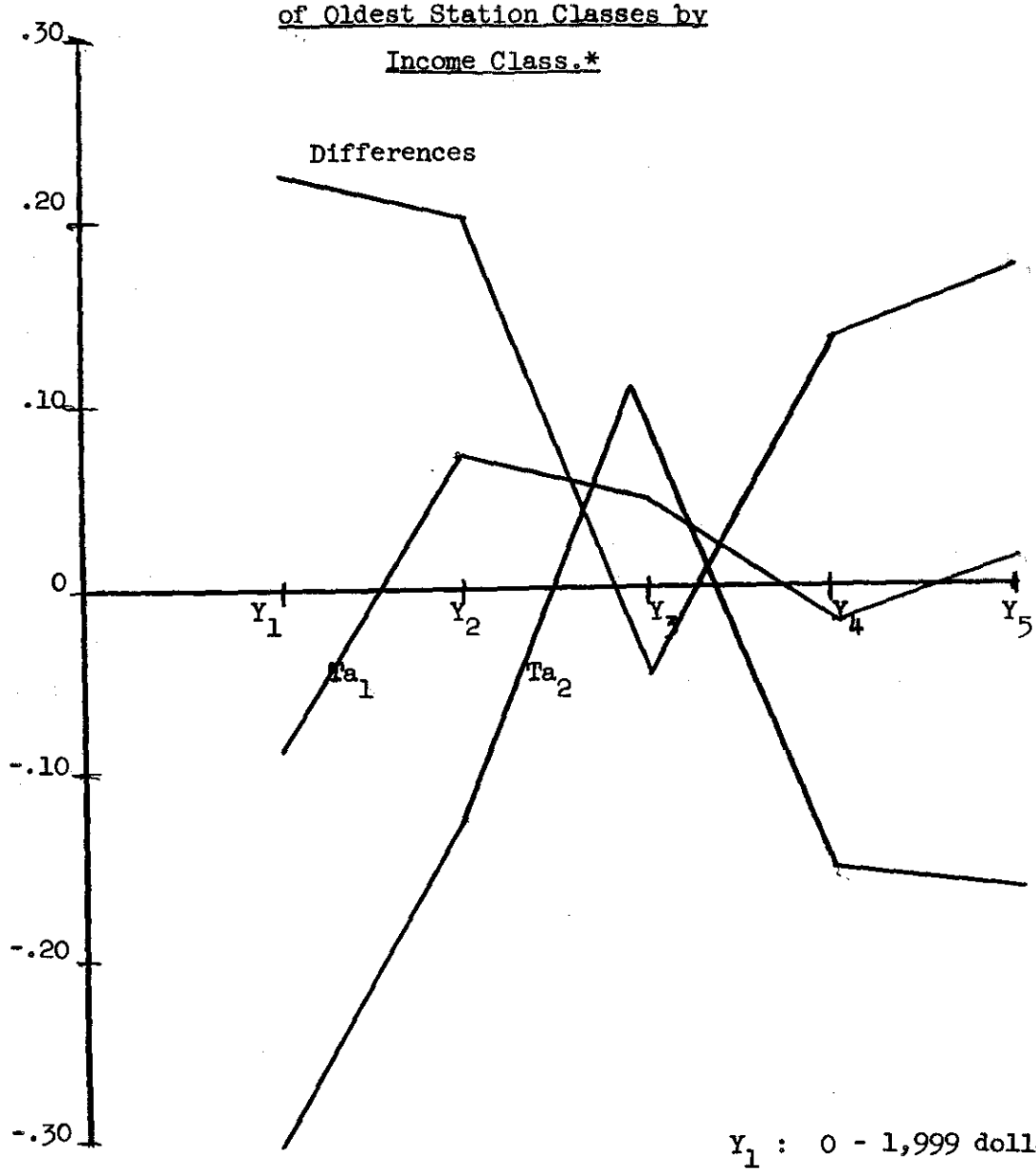
X, in Class III, is overestimated by regression III:a for high income groups when Ta in large. Consequently, had an interaction variable YTa been included in the regression the coefficients would have been negative with the consequence that the regression line of income on X would have been rotated downward about the intercept. This interaction term would have widened the spread between the Class I and III regression lines.

Taking both the YTs and YTa interactions into account, the net effect on the regression line of Class III would have been negligible because the coefficients of YTs and YTa would have had opposite signs and thus the effects would probably cancel each other. The Class I regression would, however, clearly be rotated downwards as the coefficient of YTs would be negative while the coefficient of YTa would not be significantly different from zero. The net effect, therefore, would be to narrow the gap between the Class I and Class III regression lines.

Chart 7.

Residual

Means and Differences for Two Age of Oldest Station Classes by Income Class.*




* $Ta_1$ : 1 - 30 months.
$Ta_2$ : 30 months and over.

$Y_1$ : 0 - 1,999 dollars
$Y_2$ : 2,000 - 2,999
$Y_3$ : 3,000 - 3,999
$Y_4$ : 4,000 - 4,999
$Y_5$ : 5,000 and over.

D. Analysis of Class II

It is not possible to analyze Class II with the methods used to treat Classes I and III. There are only four cities in Class II. Each value of Ts is associated with a unique value for the other coverage variables. Consequently, a meaningful regression containing more than one coverage variable cannot be computed. Although this difficulty makes it impossible to conduct as thorough-going an analysis in Class II as in the other classes, some interesting comparisons may nevertheless be made.

Intuitively, TV ownership might be expected to be highest in Class I, second highest in Class III and lowest in Class II where only foreign coverage is available. However, the weighted mean values of X in the three classes show this suspicion to be false. X equals -1.1186, -.9033, and -1.7037 in Classes I, II, and III respectively Percentage TV ownership is consequently highest in Class II and lowest in Class III. The high level of ownership in Class II is probably due to the fact that two of the four cities in Class II-Bridgeport and Paterson-receive New York's seven signals. In no other Class do half the cities receive this many signals.

A second comparison might be made between Class I and III cities with cities in Class II which have comparable numbers of signals. The result of such a comparison are entered in Table VIII. The entries are the weighted mean values of X .

Table VIII

Mean Values of X for three TV Coverage Classes by Signal Class

| | $Ts_1$ | $Ts_3$ | $Ts_7$ |
|---|---|---|---|
| Class I | - 2.0238 | - 2.0167 | - 0.8116 |
| Class II | - 3.2004 | - 1.2925 | - 0.6330 |
| Class III | - 1.2259 | - 1.0801 | - 0.7748 |

Where there is only one available signal X is lowest in Class II. With three signal, however, Class II moves up to an intermediate position, and with seven signals X is highest in Class II cities. The low ownership in Class II cities with one signal is probably due to the fact that the one foreign signal which is received will inevitably be higher than a local signal. On the other hand, the cities with no home signal and seven external signals, Bridgeport and Paterson, are fairly small cities with limited alternative entertainment facilities and thus comparison with the Class I city from which the seven signals originate (New York) shows that percentage ownership is higher in the Class II cities than in the Class I city.

Finally, it is interesting to compare the relationship between income and X in the three classes. Table IX summarizes the income regressions for the classes. In Class II a $Y^2$ term is not significant, a result which may be interpreted as

Table IX

Regressions and Coefficients of Multiple Determination ($R^2$) of Income on X for Three Coverage Classes.

| Class | Regression | R | $R^2$ |
|---|---|---|---|
| I | $X = -3.285 + .8058Y - .0541Y^2$ | | .3179 |
| II | $X = -3.195 + .6701Y$ | | .1833 |
| III | $X = -2.987 + .7543Y - .0552Y^2$ | | .3363 |

supporting the conclusion reached before--namely, that in the absence of alternative entertainment opportunities, TV becomes less of an inferior good. In addition, the low value of $R^2$ in Class II suggests that, again, because alternatives are not as readly available as in most Class I and Class III cities, TV comes closer to becoming

a household necessity and consequently income makes a smaller contribution to the explanation of the level of ownership.

## V. Testing of Hypotheses Concluded

In this section the remaining preliminary hypotheses which were set forth in Section II are subjected to statistical tests and the results of the tests are interpreted. Whereas regression analysis was used in testing the income and coverage hypotheses the results of the present section are obtained by means of variance analysis on the residual, V, of regressions I:a and III:a. The results of the section may therefore be interpreted as being free from interference with possible correlations with income and the linear coverage variables.[14]

---

14 It should be observed that independent variables which are significant on V will also be significant on X. However, if the independent variables are not significant on V, they may nevertheless become so when taken on X together with the variables originally used to calculate V. In the analysis to follow no adjustments are made to take account of this possibility, because variables which are not significant on V can only become significant on X by "stealing" some of the explanation from other variables and not by adding anything to the overall explanation.

---

### A. Education and Signals

Table X summarizes the results of analysis of variance on E, Ts, and, to make sure that the education effects are not correlated, on Y.

Table X

Analysis of Variance on V for Six Education, Five Signal, and Five Income Classes.

| Source | Variation | | Degrees of Freedom | | F Ratio |
|---|---|---|---|---|---|
| | Explained | Unexplained | Numerator | Denominator | |
| Mean | — | 1,216,783 | — | 2,767 | — |
| 1) E | 51,814 | 1,164,969 | 6 | 2,761 | 20.46** |
| 2) E+Ts | 106,420 | 1,123,891 | 10 | 2,757 | 26.08** |
| 3) ETs | 139,347 | 1,077,436 | 30 | 2,737 | 11.79** |

Table X Continues

Analysis of Variance on V for Six Education, Five Signal, and Five Income Classes.

| Source | Variation Explained | Variation Unexplained | Degrees of Freedom Numerator | Degrees of Freedom Denominator | F Ratio |
|---|---|---|---|---|---|
| Mean | | | | | |
| 4) (3)-(2) | 33,591 | 1,077,436 | 20 | 2,737 | 4.26** |
| 5) Y+E | 73,866 | 1,142,917 | 10 | 2,757 | 17.80** |
| 6) (Y+E)-Y | 49,425 | 1,167,358 | 5 | 2,757 | 23.37** |
| 7) (Y+E)-E | 2,389 | 1,164,969 | 4 | 2,757 | 5.66** |

The explained variation is row one is the variation explained by grouping the data into six education classes. The resulting variation of 51,814 is high relative to the degrees of freedom used up and thus yields a highly significant F ratio of 20.46. A regression equation with the residual, V, as dependent variable and the six education classes as dummy independent variables yields:

$$V = -.1476E_1 + .0629E_2 + .0600E_3 + .0890E_4 + .0083E_5 - .1430E_6.$$

The coefficients of this equation are the class means of the dependent variable, V . If, for example, E were twelve years, $E_5$ would equal unity, and all other E's would be zero. V would consequently equal .0083 which is simply the mean value of the residuals in class 5.

It is evident that the relationship between education and TV ownership is non-linear. For very high and very low educational levels the residuals are negative, while for intermediate values they are positive. Consequently, persons who have almost, but not quite, completed high school are the largest owners. The result, for higher education levels, is as originally expected.[15]

---

15 The possibility that the variation explained by the education classes is partially due to correlation with the remaining effects of income gives rise to the entries in row 5, 6 and 7 of Table X. A glance at row 7 shows that very little of the relationship between E and the residuals is due to correlation with the remaining income effects.

The interaction hypothesis-- that the effect produced by education will depend on the number of available signals--is also supported by the fact that a significant interaction (row 4 of Table X) exists. The direction of the interaction is, however, the exact opposite of the expected direction. The analysis of the interaction is simplified in the same way as before--namely, by collapsing the five classes of signals into two classes, each being subdivided by the standard six education classes.

In Table XI the respective cell means are posted in the first two rows. The third row entries are the differences between the values in the first two rows. These means and differences are posted in Chart 8.

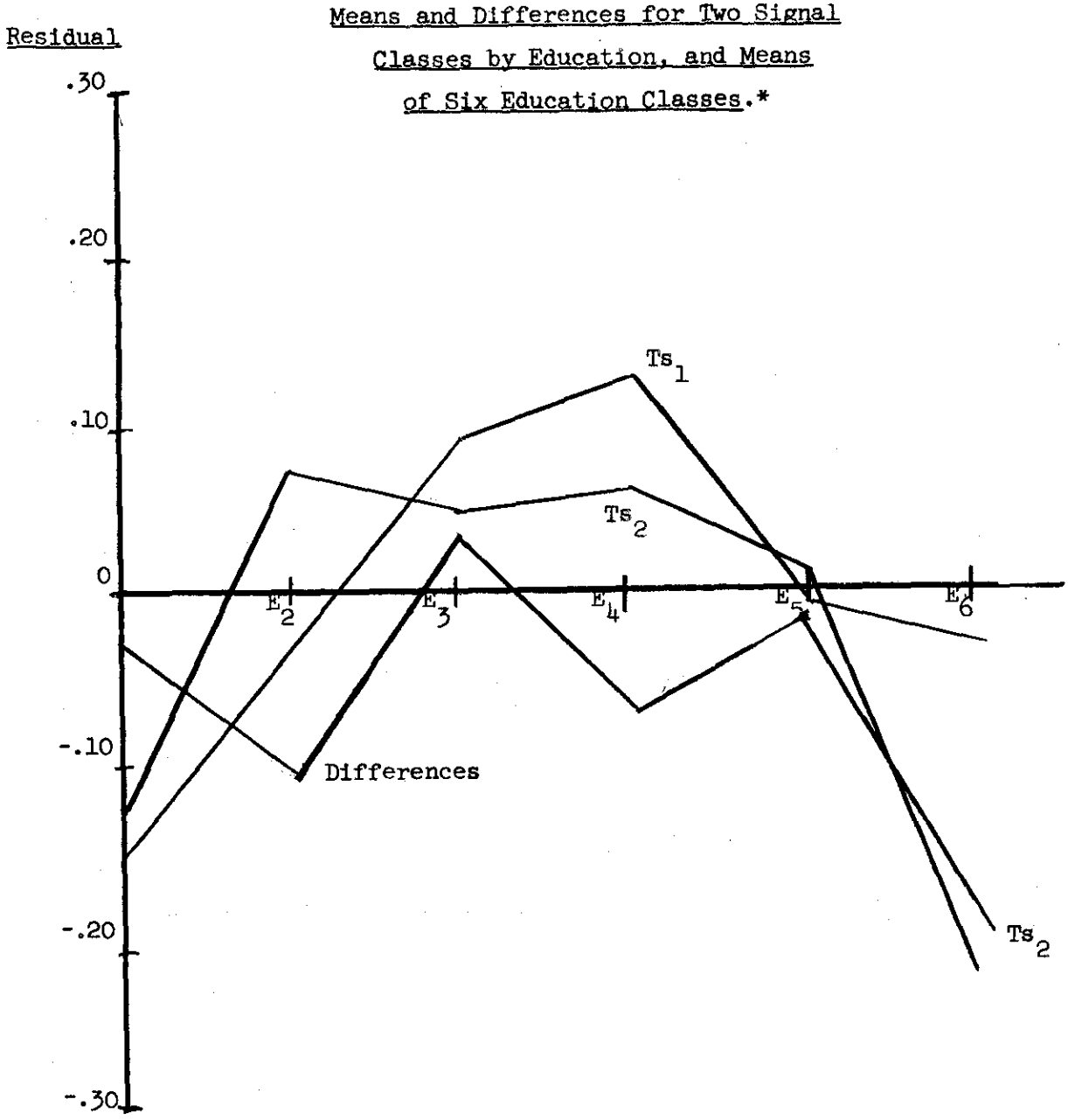
Table XI
Means of Residuals for Five Income and
Two Signal Classes

| Class I | $E_1$ | $E_2$ | $E_3$ | $E_4$ | $E_5$ | $E_6$ |
|---|---|---|---|---|---|---|
| 1) $Ts_1$ | - .1533 | - .0302 | .0866 | .1298 | .0007 | - .0164 |
| 2) $Ts_2$ | - .1303 | .0757 | .0500 | .0609 | .0132 | - .2262 |
| 3) Differences | - .0230 | - .1059 | .0366 | - .0689 | - .0125 | - .2098 |

In the absence of interaction, the entries in row 3 would be of nearly equal values with the same sign. This is clearly not the case. Two parabolas fitted with the variables $E$ and $E^2$ for the two signal classes would cross in two places. The curvature (the coefficient of $E^2$) would be greater with more signals than with less and this is contrary to the interaction hypothesis which holds that the coefficient of $E^2$ should decline as Ts increases. In fact for the very high educational levels a lower V value is observed where the number of available signals is high. This may be explained as follows: A large number of signals

Chart 8.

Means and Differences for Two Signal Classes by Education, and Means of Six Education Classes.*




$Ts_1$ : 1 - 3 Signals

$Ts_2$ : 4 - 7 Signals.

$E_1$ : 0 - 8.4 years
$E_2$ : 8.5 - 8.9 years
$E_3$ : 9.0 - 9.9 years
$E_4$ : 10.0 - 11.9 years
$E_5$ : 12.0 - 12.4 years
$E_6$ : 12.5 years

exists only in the larger cities. The larger the city the greater are the alternative entertainment opportunities. New York City, for example, had seven signals in 1950, but it also had a vast number of theaters, art galleries, opera houses and other entertainment facilities which exist in other cities to a much smaller extent if, indeed, they exist at all. The persons most likely to take advantage of these entertainment facilities are the most highly educated person and consequently these persons may be expected to substitute other forms of entertainment in place of television whenever other forms are available.

Is the relationship between education and the mere passage of time similar to the relationship between education and the number of signals? Table XII summarizes the information needed to answer this question.

Table XII

Analysis of Variance on V for Six Education and Four Ta Classes

| Source | Variation Explained | Variation Unexplained | Degrees of Freedom Numerator | Degrees of Freedom Denominator | F Ratio |
|---|---|---|---|---|---|
| Mean | —— | 1,216,783 | —— | 2,767 | —— |
| 1) Ta | 16,458 | 1,150,325 | 4 | 2,763 | —— |
| 2) E | 45,967 | 1,170,816 | 6 | 2,761 | —— |
| 3) ETa | 130,095 | 1,086,688 | 24 | 2,743 | —— |
| 4) (3)-(2)-(1) | 17,670 | 1,086,688 | 14 | 2,743 | 3.19** |

The interaction term is significant. Compressing the four Ta classes into two classes and computing the means and differences of V in the various classes yields the entries of Table XIII.

Table XIII

Means of Residuals for Six Education and Two Signal Classes

| | $E_1$ | $E_2$ | $E_3$ | $E_4$ | $E_5$ | $E_6$ |
|---|---|---|---|---|---|---|
| $Ts_1$ | - .1906 | - .0189 | - .0272 | .0695 | .0258 | - .1498 |
| $Ts_2$ | - .0719 | .1084 | .1093 | .0466 | - .0222 | - .2804 |
| Differences | - .1187 | - .1273 | - .1365 | .0229 | .0480 | .1306 |

Beyond $E_4$ differences between the mean residuals of the two Ta classes increase with higher education. If the passage of time is associated with a breakdown of snobbism, such an effect is not discernible. A counteracting tendency which may be attributable to the same set of circumstances that made the education Ta relationship the opposite of what was expected, seems to prevail.

B. Income Dispersion

The hypothesis that the more closely knit a neighborhood, the greater will be the degree to which the consumption habits of neighbors will be imitated, can be tested by assuming that the dispersion of income in a tract, as measured by the relative quartile deviation of income, $Y_q$, is a fair measure of social homogeneity.

Table XIV summarizes the results of variance analysis conducted simultaneously with Y, $Y_q$ and also the urban-adjacent classification, U . U is included because it is possible that differences between urban and adjacent tracts may be due to differences in income dispersion. Because it is desirable to analyze the effects of the urban-adjacent classification subsequently, it is convenient to include it in the present classification. Income is included to make certain that all possible income effects are removed.

The explained variation in row 6 of Table XIV is the variation explained by $Y_q$. Although the size of the variation is not large relative to what was explained by Y and E, it is nevertheless significant. Row 10 gives the variation, on V, added by $Y_q$, after the effects of Y and U are taken into account. As is evident from row 13, most of the loss of taking $Y_q$ along with Y and U is due to correlation with Y rather than with U.

Table XIV

Analysis of Variance on V for Five Median Income Classes, Four $Y_q$ Classes, and the Urban-Adjacent Classification.

| Source | Variation Explained | Variation Unexplained | Degrees of Freedom Numerator | Degrees of Freedom Denominator | F Ratios |
|---|---|---|---|---|---|
| Mean | — | 1,216,783 | — | 2,767 | — |
| 1) $Y + Y_q + U$ | 76,326 | 1,140,457 | 9 | 2,758 | 20.49** |
| 2) $Y + Y_q$ | 52,440 | 1,164,343 | 8 | 2,750 | 15.53** |
| 3) $Y + U$ | 54,365 | 1,162,418 | 6 | 2,761 | 21.52** |
| 4) $Y_q + U$ | 79,848 | 1,136,935 | 5 | 2,762 | 38.76** |
| 5) Y | 26,971 | 1,189,812 | 5 | 2,762 | 12.52** |
| 6) $Y_q$ | 35,894 | 1,189,812 | 4 | 2,763 | 21.02** |
| 7) U | 28,355 | 1,188,428 | 2 | 2,765 | 32.97** |
| 8) $YY_q$ | 69,495 | 1,147,288 | 20 | 2,747 | 8.31** |
| 9) (1)-(2) | 23,886 | 1,140,457 | 1 | 2,758 | 57.70** |
| 10) (1)-(3) | 21,961 | 1,140,457 | 3 | 2,758 | 10.77** |
| 11) (1)-(4) | 17,403 | 1,140,457 | 4 | 2,758 | 10.05** |
| 12) (8)-(2) | 17,055 | 1,140,457 | 12 | 2,747 | 3.40** |
| 13) (2)-(5) | 25,469 | 1,164,343 | 3 | 2,759 | 20.00** |

The relationship between the residual, V, and $Y_q$ is illustrated by the regression:

$$V = .0688Y_q1 - .0126Y_q2 - .1550Y_q3 - .0466Y_q4 .$$

For tracts with very low dispersion of income ($Y_{q1}$), the residuals are positive. For other values of $Y_q$ the residuals are, as expected, negative. Low dispersion of income indicates a greater degree of social homogeneity and consequent greater pressure to conform. However, the higher V value for $Y_{q4}$ than $Y_{q3}$ indicates that the relationship is not a very precise one.

The interaction between Y and $Y_q$, as is evident from row 12 of Table XIV is significant. The analysis of the interaction is again facilitated by collapsing the four $Y_q$ classes into two classes. The means and differences of the remaining classes are enterted in Table XV and are plotted in Chart 9.

Table XV

Means of Residuals for Five Income and Two Income Dispersion Classes.

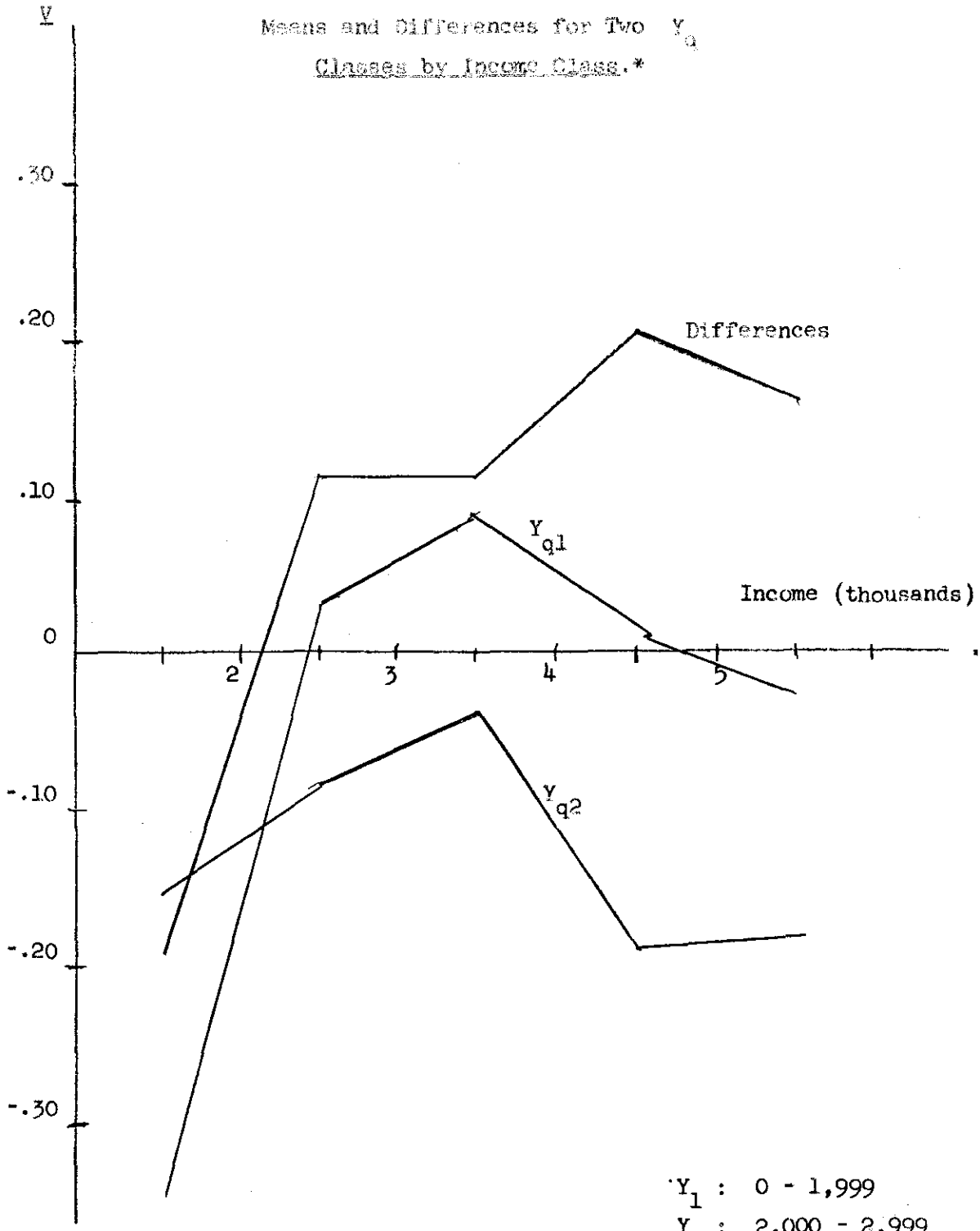
| | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ | $Y_5$ |
|---|---|---|---|---|---|
| 1) $Y_{q1}$ | - .3400 | .0331 | .0868 | .0190 | - .0177 |
| 2) $Y_{q2}$ | - .1508 | - .0815 | - .0301 | - .1822 | - .1797 |
| 3) (1)-(2) | - .1892 | .1146 | .1169 | .2012 | .1620 |

The entries lend support to the social homogeneity hypothesis. For very low incomes TV ownership is highest in tracts where the dispersion of income is greatest. For incomes greater than $2,000, however, the opposite is the case. TV ownership is greater where the dispersion of income is lowest. This suggests that the homogeneity of the neighborhood is an important factor in explaining ownership.

Residual

Chart 9.

Means and Differences for Two $Y_q$ Classes by Income Class.*




* $Y_{q1}$ : 0 - .99
$Y_{q2}$ : 1.0 - and over.

$Y_1$ : 0 - 1,999
$Y_2$ : 2,000 - 2,999
$Y_3$ : 3,000 - 3,999
$Y_4$ : 4,000 - 4,999
$Y_5$ : 5,000 and over.

C. Urban vs. Adjacent

Reference, once again, to Table XIV indicates that the final hypothesis--that TV ownership will be greater in suburban areas than in urban area--is supported by the evidence. The explained variation in row 7 is the variation explained by dividing the data into urban and adjacent categories. The variation is reduced somewhat by the inclusion of $Y_q$ and $Y_t$ (row 9) but the effects of U remain significant. The mean value of the residuals for urban areas is -.0506 and for adjacent areas the mean value is .0919. These values suggest that there is a considerable difference in TV ownership between urban and adjacent households.

The explanation for this difference may be that city dwellers are closer than suburban dwellers to the amusement centers of the city. Suburbanites, finding it too bothersome to travel to the center of town, may be content to take their entertainment at home from the TV screen. On the other hand, it might be supposed that there are good reasons why persons live in the suburbs rather than in town and these reasons may give a truer explanation than the explanation that the trouble of going to the center of town varies in direct proportion to the distance from the front door to Broadway. Common sense suggests that in most cities and for most persons it would probably be correct to say that the choice is not between the TV set in Long Island and the Belasco Theater on Broadway but between the TV set and the neighborhood theater or between watching the ball game on the TV set and going to the local sand lot. Consequently, the differences between set ownership among urban and adjacent persons may be due not so much to the distance to the center of town but to true differences in the characteristics of households in urban and adjacent areas, such as differences in family size, degree of crowding in the home, home ownership and other factors. Some of these possibilities are explored below.

Table XVI summarizes the results of analysis of variance on V for the urban-adjacent classification, six classes of median number of persons in dwelling units, DU, and four classes of percent with 1.01 or more persons per room, C .

Table XVI

Analysis of Variance on V for the Urban-Adjacent Classification, Six Classes of Median Number of Persons in Dwelling Units, and Four Classes of Percent 1.01 or More Per Room.

| Source | Variation | | Degrees of Freedom | | F Ratio |
|---|---|---|---|---|---|
| | Explained | Unexplained | Numerator | Denominator | |
| Mean | — | 1,216,783 | — | 2,767 | — |
| 1) C + DU + U | 217,364 | 999,419 | 10 | 2,757 | 59.88** |
| 2) C + DU | 215,998 | 1,000,785 | 9 | 2,758 | 65.73** |
| 3) DU + U | 196,099 | 1,020,684 | 7 | 2,760 | 75.71** |
| 4) C + U | 49,644 | 1,167,139 | 5 | 2,762 | 23.47** |
| 5) C | 23,878 | 1,192,905 | 4 | 2,763 | 13.82** |
| 6) U | 28,355 | 1,188,428 | 2 | 2,765 | 32.97** |
| 7) DU | 194,077 | 1,022,706 | 6 | 2,761 | 87.42** |
| 8) (1)-(2) | 1,366 | 999,419 | 1 | 2,757 | 3.74 |
| 9) (1)-(3) | 21,265 | 999,419 | 3 | 2,757 | 19.42** |
| 10) (1)-(4) | 167,720 | 999,419 | 5 | 2,757 | 91.90** |
| 11) (3)-(7) | 2,022 | 999,419 | 1 | 2,760 | 6.53** |

The explained variation in row 8 is what remains of the variation explained by the urban-adjacent classification after DU and C are taken into account. This remainder is not significant even at the 5% level. When DU and C are taken into account it is not worth the loss of another degree of freedom to obtain the contribution of the urban-adjacent distinction.

Rows 9 and 10 show the explanations attributable to C after DU and U are taken into account, and DU after C and U are taken into account. The 167,720 explained variation in row 10 is particularly high and indicates that family size ranks in importance with income and the extent of TV coverage in explaining TV ownership.

Further consideration of the importance of family size is deferred to Section VI.

C, the percent of households with 1.01 or more persons per room is clearly a measure of the degree of crowding in the home. From Table XVI (row 3 minus row 7) it is evident that the inclusion of C was not necessary as far as the reduction to insignificance of the variation explained by the urban-adjacent classification was concerned. Although it is reasonable to expect homes to be more crowded in urban than in adjacent areas, the difference in family size seems to compensate for this difference. C, moreover, is probably not a very meaningful measure of crowding because there is a large difference between the size of rooms. The explained variation due to C is probably a per capita income effect. When households are very crowded, per capita income will be low. Consequently, it is not surprising to find that the class means of the residuals of the C classes yield the same U shape as the mean value of the residuals of the income classes.

## VI: Supplementary Analysis

The analysis so far has concentrated rather heavily on the testing of the formal hypotheses and the interpretation of the results. In this section additional variables, particularly those concerning household characteristics, are taken up. The method of analysis employed is to screen the remaining independent variables listed at the end of Section III for significance by means of uni-variate variance analysis. Possible correlations between variables are then analyzed by a series of regressions on V .

### A. Analysis of the Variables

Reference to the list of independent variables posted in Section III shows that the variables not yet considered are:

A - Median age of the male population between the ages of 14-64.

Ac - The percent of the male population under 14 years of age.

Aa - The percent of the male population over 64 years of age.

N - The percent of the total population, non-white.

M - The percent of the male population, married.

F - The percent of the total population, female.

Flf-The percent of the female population over 14 years of age in the labor force.

OO - The percent of occupied dwelling units occupied by their owners.

DU - The median number of persons in dwelling units.

Table XVII summarizes analysis of variance on V for these variables. The format of the table is the same as previous variance analysis tables with the exception of an additional column in which the signs of the coefficients of the independent variables are entered.

Table XVII

Analysis of Variance on V for Classes of A, Ac, Aa, N, M, F, Flf, DU, and OO.*

| Source | Variation | | Degrees of Freedom | | F Ratio | Sign |
|---|---|---|---|---|---|---|
| | Explained | Unexplained | Numerator | Denominator | | |
| Mean | --- | 1,216,783 | --- | 2,773 | --- | |
| A | 40,354 | 1,176,429 | 5 | 2,768 | 18.99** | - |
| Ac | 78,299 | 1,138,484 | 4 | 2,769 | 47.63** | + |
| Aa | 33,840 | 1,182,943 | 4 | 2,769 | 19.13** | - |
| N | 42,539 | 1,174,244 | 4 | 2,769 | 25.09** | - |
| M | 7,379 | 1,209,404 | 4 | 2,769 | 4.22** | + |
| F | 52,530 | 1,164,433 | 5 | 2,768 | 24.87** | - |

Table XVII Continues

Analysis of Variance on V for Classes of A, Ac, Aa, N, M, F, Flf, DU, and OO .*

| Source | Variation Explained | Variation Unexplained | Degrees of Freedom Numerator | Degrees of Freedom Denominator | F Ratio | Sign |
|---|---|---|---|---|---|---|
| DU | 194,077 | 1,022,706 | 6 | 2,767 | 87.42** | + |
| OO | 2,101 | 1,020,605 | 2 | 2,765 | 2.85 | + |
| Flf | 84,082 | 1,132,701 | 4 | 2,769 | 51.55** | - |

* The variation explained by OO is a remainder after the effect of DU was taken out.

Table XVII suggests that OO may be dropped on the grounds that it is not significant. Similarly, the explained variation of M is so small that it does not seem worthwhile to consider it further. The remaining variables are all significant when taken by themselves. Some of the signs, however, are the opposite of what might be expected. It is not surprising to find that increased participation in the labor force by females should be associated with less set ownership, but that an increase in the relative number of females in a tract is associated with less ownership is difficult to understand. Similarly, it is strange to find that an increase in the relative number of aged persons is associated with declining TV ownership.

These results suggest that there are undoubtedly some correlations between some of the variables and that when these are taken into account the signs of some of the coefficients may turn out to be more in line with expectations. In order to test this possibility some of the more likely joint relationships are analyzed

with the aim of determining whether combinations of the variables yield different results than univariate treatment. The analysis proceeds by means of regression analysis conducted with a sub-sample of 554 observations. The dependent variable in each case is the residual, V . The regressions are reported in Table XVIII.

Table XVIII

Regression Equations on V and Corresponding t Ratios (bracketed values) for Selected Combinations of Independent Variables.*

| Code | Regression Equation |
|---|---|
| 6.1 | V = .0099 + .0771F - .1350F1f<br>(17.59**) (19.20**) |
| 6.2 | V = -.0483 + .0027F1f + .0447N<br>(1.35) (16.65**) |
| 6.3 | V = .0041 - .0851Aa + .0083A<br>(3.92**) (2.04*) |
| 6.4 | V = -.0182 + .2719DU - .3933Aa - .2568Ac<br>(38.19**) (38.54**) (33.13**) |

* Double asterisks accompanying a t ratio indicate significance at the 1% level of significance, while a single asterisk indicates significance at the 5% level. The criterion of acceptance is the one percent level.

Regression 6.1 combines the effects of F and F1f . Whereas the sign of the coefficient of F, when taken independently, is negative, it now becomes, as originally expected, positive. The coefficient of F1f remains negative, a result which again is in line with common sense expectations. In addition, the coefficient of F1f falls from the -.0166 it would have been had F1f been taken alone, to a value of -.1350

The result suggests that the coefficient of F, when taken alone, is negative because in tracts where the relative size of the female population is large, a larger proportion of the females will participate in the labor force. Labor force participation, in turn, reduces TV ownership. However, when both F and Flf are taken together, housewives are, in effect, segregated from working women and thus, as expected, V rises as the relative number of housewives increases.

Regression 6.2 estimates V as a function of Flf and N . The two variables are combined on the expectation that participation in the labor force is higher for non-white females than for whites. The coefficients of both independent variables are positive. Taken independently, the coefficient of Flf would have been -.0166 and the coefficient of N would have been -.0464. However, the coefficient of Flf is insignificant, and this indicates that N and Flf are so highly correlated that, in so far as their relationship to TV ownership is concerned, both are practically the same variable.

The foregoing suggests that the relationship between F, Flf, and N is as follows: TV ownership is inversely related to the relative size of the non-white population. Once the non-white population has been taken into account, additional information regarding female participation in the labor force adds very little to the explanation of TV ownership. There is, however, a strong correlation between female labor force participation and the relative number of females in a tract. If the number of labor force participating females is large, the proportion of those who are housewives is apt to be small and TV ownership will be small.

Further speculation on the causes of low ownership in non-white areas suggests that perhaps TV programs are designed to appeal to a white audience rather than to colored audiences. Except for sports events this supposition is undoubtedly true. Perhaps the TV program makers have overestimated the importance of income as an enabling factor in owning TV and have therefore guided their program making towards a rather narrow audience. On the other hand, it may very well be true that there is a very real difference between white and non-whites which, apart from income and program design, makes for a lower level of ownership among non-whites.

Regression 6.3 brings together median age of persons 14-64 years of age, A, and the relative number of aged persons, Aa, in a tract. Independently, the coefficient of A is -.0062 and the coefficient of Aa is -.0444. The negative values of Aa are unexpected because common sense suggests that old persons might welcome TV as a source of relatively effortless entertainment and as a means of passing their leisure hours. In regression 6.3, however, the addition of A to Aa has not changed the coefficient of Aa to the expected positive value. The coefficient of A, moreover, is not significant at the one percent level. Because both Aa and A are independently fairly significant it is evident that where Aa is high A is also high (the correlation coefficient between the two variables is .916) and thus when Aa is included together with A, the additional variation explained by A is unimportant. Consequently, if there are relatively more aged persons in one tract than in another, median age of persons 14-64 will also be higher in the first tract, and higher age is associated with less TV ownership. Provisionally, it may be inferred that age makes for conservatism in the adoption of a new product such as TV.

Finally, the combined effects on V of family size, children, and aged persons are taken into account in regression 6.4. The sign of the coefficient of Ac now becomes negative whereas it was positive when taken independently. The coefficient of Aa is also negative and absolutely greater than the coefficient of Ac, a result which is to be expected because the coefficients of Ac and Aa are positive and negative respectively when the two are taken independently.

Regression 6.4 must be interpreted with care. The addition of a child or an aged person to a family necessarily involves an increase in family size. When family size increases, however, the increase is not necessarily reflected in an increase in the number of children or old persons. The regression therefore suggests that as family size increases, the probability that a family will own TV increases but that the increase will be smaller if the new family member is a child instead of an adult. The negative coefficient of Aa when taken independently suggests that an additional aged person may actually overcome the family size effect and may reduce a family's probability of owning a set.

B. The Inferior Good Hypothesis and Family Size and Composition

The coefficients of regression 6.4 are so highly significant and the results so extremely suggestive that further discussion is in order.

At the outset it should be observed that the commonly held notion, that the motivation behind the decision to purchase TV is provided by pressure from children, is only a half truth. Adults may justify the purchase of a set to their snobbish friends in these terms but the present analysis indicates that, in fact, adults themselves provide more motivation towards set purchase than do children. The question of why it is that the probability that a family will own a set is greater if the family is composed wholly of adults rather than partially of adults and children, nevertheless, remains. An answer to this question, it is submitted, can be found in the inferior

good hypothesis analyzed in Section IV and in the peculiar nature of television as an entertainment economy.

Because income, as reported by the Census, is personal family income, and because the effects of family income have been removed from these data through calculation of the residual, V, an increase in family size, family income remaining constant, must necessarily be associated with a decline in per capita income. The fact that as family size increases, TV ownership increases, suggests that TV becomes more and more important as per capita income, and therefore the ability to purchase alternatives, decreases. The great advantage of television appears to be that, once the initial cost of purchasing a set has been met, the marginal cost of viewing one more program is negligible, and the marginal cost of allowing one more family member to view a program which is already being watched by others, is zero. On the other hand, the marginal cost of taking one more person to the movies, or to a baseball game, is the price of admission and perhaps also bus fare. It is clear, therefore, that there are considerable "economies of scale" to owning TV.

The next question to be answered is: For what age groups are these economies the greatest? Casual observation suggests that because the majority of aged persons are not habitual movie goers, TV is less of a money saving device, and therefore less of an inferior good for families composed partly or wholly of aged persons. In the case of children two factors are important. The admission price to motion pictures is lower for children than for adults, and children, probably, do not attend movies as frequently as adults. Consequently, there will be a lower saving on alternative expenditures if the family is partly composed of children. However, a counteracting factor needs to be cited--namely, that the effective admission price for adults to a motion picture, theater, or baseball game is raised

if there are young children in the family because the cost of baby sitters must be included. If, however, there are also aged persons who serve as baby sitters in the family, this cost is not encountered and consequently the need for TV is lessened.

In conclusion, the results of regression 6.4 quite clearly seem to support the inferior good hypothesis. Both on statistical and theoretical grounds there is considerable reason to suppose that the probability of ownership of TV increases more if an adult is added to a family than if a child is added.

## VII: Summary

Common sense suggests that the two factors which will be most important in explaining TV ownership in a tract are the length of time and intensity of television coverage and the income characteristics of the tract. As the analysis of Section IV showed, these presumptions were entirely correct. Income and television coverage explain about 65% of the total variation about the mean of the dependent variable. Moreover, the shapes of the relationships are of a sort which lend support to the original hypotheses. Television ownership increases as the length of time television has been available to a region increases and as the extent of TV coverage increases. In Class III, but not in Class I, the relative saturation hypothesis--that a plateau of ownership may be reached at less than 100%--is supported by the findings.

With respect to income, increasing percentage ownership is associated with increasing income to median income levels of $7,676 and $6,487 in Classes I and III respectively, after which increasing income is associated with declining percentage ownership. For high income tracts in Class I, TV ownership increases by much less when the number of signals increases, than it does for other income groups. This result corroborates the inferior good hypothesis. It leads to the interesting

possibility that, because a larger number of signals is associated with a large number of alternative entertainment opportunities, the degree to which TV is an inferior good with respect to income depends upon the degree to which alternative entertainment facilities are available.

The hypothesis that TV is an inferior good not only with respect to income but also with respect to education is supported by the evidence at hand. TV ownership is highest in tracts where the median educational level is from 10.0 to 11.9 years. For higher education levels, and also for lower levels, ownership is lower. The interaction hypothesis--that additional signals will increase ownership relatively more among highly educated groups than among other groups is not borne out by the analysis. Indeed, quite the opposite is the case. The interaction term, although significant, shows that set ownership among highly educated groups is overestimated to a greater degree by regressions I:a and III:a where the number of signals is large than where the number of signals is small. This rather peculiar result can be rationalized by reference to the same argument which was employed to interpret the income-signals interaction.

The social homogeneity hypothesis--that for areas with equal median incomes, percentage TV ownership will be inversely related to the dispersion of income except in very low median income areas--is borne out by the analysis. For tracts with median income in excess of $2,000 a decrease in the dispersion of income is associated with greater percentage ownership, while the opposite is true for tracts with median incomes less than $2,000.

Division of data into urban and adjacent categories shows that, as hypothesized, ownership of TV receivers in suburban areas is significantly greater than in urban tracts. Extension of the analysis to include family size shows that the urban-adjacent

difference is in reality a difference caused by family size.

Although there is a positive association between percentage set ownership and family size, the degree to which this is true depends upon family composition. For a given family size the probability that the family will own TV is greatest where the family is composed of adults from the ages of 14-64. The probability of ownership will decline if a child is substituted in place of an adult, and declines still further if an aged person is substituted. These findings, it was argued, support the inferior good hypothesis and help to shed light on the peculiar nature of TV as an expenditure saving device.

Finally, it was found that an increase in the relative size of the non-white population is associated with declining TV ownership. This may be due to the nature of TV programs but also to the fact that an increase in the relative size of the non-white population is associated with an increase in the percent of the female population which participates in the labor force. There is a positive association between TV ownership and an increase in the relative numbef of housewives but a negative association between ownership and an increase in the percent of females participating in the labor force.

## Appendix A

## Quartile Estimation

The measure of income dispersion used in the analysis is the relative quartile deviation:

$$Y_q = \frac{Y_{q3} - Y_{q1}}{Y}$$

where $Y$ is median income and $Y_{q3}$ and $Y_{q1}$ are the third and first income quartile points respectively. The relative quartile deviation is used because it is a well known fact that the quartile deviation of income is highly correlated with the median level of income. $Y_{q3}$ was computed from the income frequency distribution reported by the Census by the expression:

$$Y_{q3} = M_{q3} - (c_3/f_3)(F_3 - .75F)$$

where $M_{q3}$ is the upper class mark of the third quartile class, $c_3$ is the class interval, $f_3$ the frequency in the third quartile class, $F_3$ the cumulative frequency through $M_{q3}$, and $F$ the total frequency of those reporting income in the tract.

In 121 tracts third quartile income points fell in the open-end interval of $10,000 or more. To estimate quartile points in these cases, a sub-sample of the fifty-one highest median income tracts where third quartile incomes were known was used to compute a linear regression of $Y_{q3}$ on $f_o$, the percentage of the total frequency falling in the open-end class. Accordingly,

$$Y_{q3} = a + bf_o$$

subject to the restriction that,

$$\$10,000 = a + .25b.$$

The estimated value of $b$ was 5543. For each percent that the total frequency in the open-end class exceeds 25%, $Y_{q3}$ was therefore estimated to exceed $10,000 by $55.43.

## Appendix B

## Weighting

Since there are different numbers of persons in different tracts, each observation must be weighted proportionately to the number of respondents in the tract. The number of respondents is generally different for different questions and these differences vary from tract to tract. However, the lack of proportionality between numbers answering different questions from tract to tract does not appear great. The differences are usually consistent in direction. For example, in a high income tract a higher proportion of the total population will report income than in a low income tract, while the proportions reporting do not differ. Evidently, people are less reluctant to let others know that they are well to do than to let them know that they are poor. The consistency found here suggests that no loss of significance will be caused by this difference in the percent reporting.

## Appendix C

## Transformation of Dependent Variable

An individual either owns a TV set or he does not. It is impossible to define a continuous function relating a person's TV ownership to other characteristics the way it might be possible to establish such a link between his characteristics and, for example, his food expenditures. It is, however, possible to establish a threshold index as a function of his characteristics. If the index exceeds his critical value he will own a set and if it does not he will not be an owner. For a large group of people with a given value of the index (equal incomes, education, etc.) a proportion, p, of the persons will have thresholds less than the index (owners) and (1-p) will have

thresholds greater than the index (non-owners). If the thresholds of the individuals are assumed to be normally distributed, the percentage of ownership out of a large number of individuals varies, with respect to the index, in a way which approaches the cumulative normal curve.

If $p$ varies in the way suggested above, it is obvious that linear regression with $p$ as the dependent variable would be poor procedure. Approximation of the integrated normal curve, however, requires iteration and it is desirable to avoid this if possible. Berkson[16] has suggested a

16. J. Berkson, "Application of the Logistic Function to Bio-Assay", Journal of the American Statistical Association, Vol. 39, #227, pp. 357-365.

procedure for doing this. He assumes the probability, $p$, of a positive response to be given by the logistic function which has the same general shape as the cumulative normal curve. Stated as a function of $n$ independent variables:

$$p = \frac{1}{1 + e^{-(a + \sum_{i=1}^{n} b_i x_i)}}$$

and has the linear transformation:

$$\ln(p/(1-p)) = a + \sum_{i=1}^{n} b_i x_i$$

$p$ is the true probability rather than the observed proportion. $p$ can, however, be approximated, without iteration, by the observed proportion, $\hat{p}$, provided that $\hat{p}$ is based on a large number of cases and that each observation is weighted by $n\hat{p}(1-\hat{p})$ where $n$ is the number reporting TV ownership in a tract. The component $\hat{p}(1-\hat{p})$ is included as part of the weight

because at very high or very low values of $\hat{p}$ the approximation is least good with the consequence that observations at the extremes are weighted less heavily than observations in intermediate regions.

Following Berkson's procedure, the natural logarithm of the ratio $p/(1-p)$ is used as the dependent variable in linear regressions appropriately weighted.

## Appendix D

## Statistical Models

Variance and multiple regression analysis are employed as complementary tools of analysis in this study.

Regression analysis estimates the value of a dependent variable as a function of a set of independent variables by statistically evaluating the parameters of a function in a way which minimizes the sum of squared residuals. In the general case, the parameters of the function will be such that:

$$\sum_{i=1}^{n} [x_i - G(y_{1i}, y_{2i}, \cdots, y_{ni})]^2$$

where $x_i$ is the observed value of the dependent variable of the i-th observation and where the $y$'s are the $n$ independent variables, will be at a minimum.

Variance analysis indicates whether, if observations are grouped in accordance with different values of one or more independent variable, there is a difference in the mean values of the dependent variables. Use of variance analysis as a preliminary to regression analysis serves the purpose of indicating whether the variables under analysis are significant, and, because variance analysis focuses attention on the means of the dependent variable associated

with particular classes of the independent variable, indicates whether the relationship between the variables is linear or non-linear. Moreover, where it is fairly evident that independent variables are not correlated, the difference between the partial relationships gotten through multi-variate regression analysis and uni-variate variance analysis will be slight. Because variance analysis is a more economical technique than regression analysis, it is utilized wherever it is appropriate.

The great virtue of regression analysis is that correlations between independent variables are taken into account. Consequently, multiple regression analysis reduces the risk of giving the same explanation twice.

The advantages of regression analysis can be combined with the economy of variance analysis by proper construction of a variance analysis model.[17]

---

17. I am indebted to Professor James Tobin for illustrating this model.

---

Because the model is not among the more common variance analysis models and because it is used extensively in the present investigation it is instructive to consider its properties briefly.

In a two way classification according to classes of the independent variables E, education, and Y, income, the total sum of squared residuals about the sample mean of the dependent variable X, can be written:

$$S^2 = I^2_{ey} + S^2_{e+y} + S^2_u$$

where $S^2$ is the over all sum of squares, $I^2_{ey}$ is the portion of the total due to interaction between E and Y; $S^2_{e+y}$ is the sum of squares explained by the Y classification together with the explanation added by E; and $S^2_u$ is the within cell, or unexplained, sum of squares. If E and Y are entirely independent of each other, the Y (row) classification will be

entirely independent of the E (column) classification and as a consequence:

$$S^2_{e+y} = S^2_y \neq S^2_e$$

where $S^2_y$ and $S^2_e$ are the row and column sums of squares respectively. In many examples of variance analysis, substitution of $S^2_y$ and $S^2_e$ for $S^2_{e+y}$ is appropriate because the experiment can be designed or the classifications can be "rigged" to eliminate dependence of row on column classifications by making the number of observations in each cell of one column proportional or equal to the number of observations in the corresponding cells of other columns. This technique simplifies computation because $S^2_y$ is simply equal to the sum of the products of the row means and the sum of the dependent variable in the row classification. $S^2_e$ is the column counterpart.

The procedure outlined above is difficult to follow if data are not obtained from controlled experiments and is practically impossible to use with non unit weight data. Consequently, the problem is to evaluate $S^2_{e+y}$, the row sum of squares plus the column sum of squares with the component due to correlation between E and Y eliminated.

A return to the fundamentals of regression analysis shows how this can be accomplished. If each row and column class is defined as a "dummy" variable with either a zero or one value, the problem can be solved by the familiar technique with which the parameters of a regression and the sum of squares explained by the regression are calculated.

As an illustration, assume that there are three education classes, $E_1$, $E_2$, and $E_3$ and three income classes $Y_4$, $Y_5$, and $Y_6$. Knowledge of the sum of

weights $\Sigma w_{ij}$ in each cell, the sums of the weighted dependent variable $\Sigma w_i X_i$ in the rows and columns, makes it possible to set up the familiar "moments matrix" of regression analysis. Because the sum of the weights and the sum of the weighted dependent variable in one class will be determined once the same values for the other five classes are known, any one of the new "dummy" variables must be dropped out.

| | $E_1$ | $E_2$ | $E_3$ | $Y_4$ | $Y_5$ | $X$ |
|---|---|---|---|---|---|---|
| $E_1$ | $\Sigma w_1$ | | | | | |
| $E_2$ | 0 | $\Sigma w_2$ | | | | |
| $E_3$ | 0 | 0 | $\Sigma w_3$ | | | |
| $Y_4$ | $\Sigma w_{14}$ | $\Sigma w_{24}$ | $\Sigma w_{34}$ | $\Sigma w_4$ | | |
| $Y_5$ | $\Sigma w_{15}$ | $\Sigma w_{25}$ | $\Sigma w_{35}$ | 0 | $\Sigma w_5$ | |
| $X$ | $\Sigma w_1 X_1$ | $\Sigma w_2 X_2$ | $\Sigma w_3 X_3$ | $\Sigma w_4 X_4$ | $\Sigma w_5 X_5$ | $\Sigma w X^2$ |

In general, the weighted moment between any two variables, expressed as deviations from their sample means, is $\Sigma w x_i x_j$. When $E$ is $E_1$, it cannot also be $E_2$ or $E_3$. Consequently, all $E_2$ and $E_3$ are zero and $E_1$ equals unity. The moments of $E_1 E_2$ and $E_1 E_3$ will therefore be zero, the moment of $E_1 Y_4$ will equal the sum of the weights in that cell, and the moment of $E_1 X_1$ will be $\Sigma w_1 X_1$.

Given the moments, the regression can be solved and the sum of squares explained by the regression can be found. The sum of squares is the sum of the variation explained by $Y$, a deduction for correlation between E and Y having been made.

Any two or more independent variables can be analyzed by a combination of

three variance analysis models. Continuing the income and education sample, the first and simplest model is:

a. $$X = \sum_{i=1}^{n} a_i Y_i$$

or,

b. $$X = \sum_{j=1}^{m} b_j E_j$$

where n and m are the number of classes of Y and E respectively. Model (a) states that X is a function of Y, and model (b) states that X is a function of E. The models do not take into account the fact that income and education may be substitutes for each other as explanations of X. Correlation between E and Y is taken into account in the second model:

$$X = \sum_{i=1}^{n} a_i Y_i + \sum_{j=2}^{m} b_j E_j .$$

The model states that X is a function of Y and also of E, the effect of Y having been removed. The third model:

$$X = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} Y_i E_j$$

states that X depends upon both Y and E and that the way in which Y affects X depends on the value of E. This third model therefore takes into account all possible effects of the classification according to Y and E.

By combining these three models, all manner of tests can be made. Let the explained variation of the third model be $S^2_{ye}$, of the second model be $S^2_{e+y}$, and of the first model be $S^2_y$ and $S^2_e$. Subtracting $S^2_{e+y}$ from $S^2_{ey}$ leaves the interaction effect, while $S^2_{e+y} - Se^2$ gives the explanation provided

by Y, E having been taken into account.

The significance of the relationships can be tested by means of F tests in which the variation explained per degree of freedom lost is compared with the remaining variation per degree of freedom.