DOCUMENT RESUME

ED 097 956 PS 007 164

AUTHOR	Cohen, Allan S.; And Others
TITLE	Determining a Sufficient Sample Size for Analysis of Mail Questionnaire Data. Center for Human Services Development, Report No. 33.
INSTITUTION	Pennsylvania State Univ., University Park. Coll. of Human Development.
SPONS AGENCY	Pennsylvania State Dept. of Public Welfare, Harrisburg.
REPORT NO	PSU-CHSD-R-33
PUB DATE	[73]
NOTE	111p.; For other reports in this series, see PS 007 154 - PS 007 164
EDRS PRICE	MF-$0.75 HC-$5.40 PLUS POSTAGE
DESCRIPTORS	Certification; Data Analysis; *Data Collection; *Day Care Programs; Enrollment; Ethnic Distribution; *Mathematical Models; Negro Youth; Prediction; *Questionnaires; *Trend Analysis
IDENTIFIERS	Pennsylvania

ABSTRACT

This paper presents the description and several applications of a model which can be used to determine how long a researcher must wait for the return of completed mail questionnaires to be sure the data collected reflect the true values of the parameters of interest. This model proposes to fit two linear time trend lines to a set of mailed questionnaire data. Application of the model provides an estimate of the intersection point during the time period at which the estimate of the parameter has stabilized. It is at the stabilization point that the researcher can be relatively confident that his estimates do not differ significantly from those which he would make were he to wait for additional returns to be received. To illustrate the application of this model, questionnaires were sent to 1,120 child care centers in Pennsylvania. Nine weekly subsamples of the 545 returned questionnaires were analyzed according to the model to estimate the average maximum allowed enrollment, the proportion of centers either licensed or approved, and the proportion of centers reporting Black children in attendance. Also, the appendices contain analyses of the data from each of twenty categories of information contained in the questionnaire. (SDH)

INSTITUTE FOR THE STUDY OF HUMAN DEVELOPMENT

CENTER FOR HUMAN SERVICES DEVELOPMENT

DETERMINING A SUFFICIENT SAMPLE SIZE FOR ANALYSIS OF MAIL QUESTIONNAIRE DATA

Allan S. Cohen, Thomas Hettmansperger, Paul D. Mowery

CHSD Report No. 33

COLLEGE OF HUMAN DEVELOPMENT
PENNSYLVANIA STATE UNIVERSITY
UNIVERSITY PARK, PA. 16802

DETERMINING A SUFFICIENT SAMPLE SIZE FOR ANALYSIS OF MAIL QUESTIONNAIRE DATA

By: Allan S. Cohen
Thomas F. Hettmansperger
Paul D. Mowery

Center for Human Services Development

College of Human Development

The Pennsylvania State University

University Park, Pa. 16802

This report was prepared under contract with the Pennsylvania Department of Public Welfare. The opinions and recommendations expressed are those of the authors and do not necessarily reflect the official position of the sponsoring agency.

ABSTRACT

One of the major problems facing researchers using a mailed questionnaire to collect data is the length of time they should wait for recipients to return completed questionnaires. Since this decision can affect the degree to which the data reflect the true values of the parameters of interest, it is important that care be taken in allowing enough time for data to be collected and yet not unnecessarily extending this waiting period. A model is described in this paper which proposes to fit two linear time trend lines to a set of mailed questionnaire data. If the slope of the second trend line is not significantly different from 0, the point of intersection provides an estimate of the point during the time period at which the estimate of the parameter has stabilized. At this stabilization point, in other words, the researcher can feel relatively confident that his estimates do not differ significantly from those which he would make were he to wait for additional returns to be received. Explanation of the model and examples are provided to show applications for both means and proportions.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd.)

TABLE OF TABLES

TABLE OF TABLES (Cont'd.)

TABLE OF TABLES (Cont'd.)

TABLE OF TABLES (Cont'd.)

TABLE OF FIGURES

DETERMINING A SUFFICIENT SAMPLE SIZE FOR ANALYSIS OF MAIL QUESTIONNAIRE DATA

Allan S. Cohen Thomas P. Hettmansperger

Paul D. Mowery

Introduction

The mailed questionnaire is frequently used as a means of data collection because of its economic attractiveness and facility over other types of surveys such as telephone or personal interviews. That is, mailed questionnaires require relatively little expense and/or effort to collect information from each individual in a sample and can consequently be used to collect information from very large numbers of people. This type of survey technique, therefore, has a high degree of appeal to researchers who seek to obtain a set of data on some (perhaps large) population but who are constrained from using more comprehensive information gathering procedures.

Questionnaires of this type, consequently, can be used to reach large numbers of people and do not suffer from the effects of subject inavailability due to other commitments in their daily lives. For example, persons who work irregular hours or who do not have telephones, can, however, generally be contacted by mail.

It is typical in most mail questionnaire surveys for completed data to be received over an uncontrolled period of time. Whereas the preference of the researcher may be to collect as much data as possible before analyzing them, the time allowed for data collection is usually limited. It is obvious that the amount of data collected can be an important factor determining the stability of the parameter estimates made on this data. It is, however, fully as important that these estimates reflect the true parameters. The choice of a stopping point, therefore, in the data collection time period is a crucial one that can affect the utility or effectiveness of a set of data. The purpose of this paper, therefore, is to present the description and several applications of a model which can be used to identify a stopping point with respect to time for the collection of returned questionnaires. This model attempts to predict how quickly various parameter estimates can be made for future surveys based on the results of a previous survey.

A mail questionnaire (described in a later section of this paper) was designed as part of a sampling plan to estimate several parameters of the population of child care centers in Pennsylvania. No estimates were made from the data until the end of the tenth week following the mailing of the questionnaire. However, because knowledge regarding some of the parameters decreases in value with time, it would have been desirable to make estimates of these parameters based on less than nine weeks of returns (returns were grouped on a weekly basis where subsample 1 was formed at the end of the second week following the mailing). Furthermore, it was observed by plotting the cumulative estimates based on one, two,..., nine weeks of returns versus the week number in which the cumulative estimate was made, that after an initial period of instability, the cumulative estimates for some parameters remained very close to the cumulative estimate made at the end of the ninth week in the sampling time period.

The model discussed in this paper attempts to statistically determine the week in which this stabilization of the cumulative estimate for a given parameter occurs. With this information, based on the assumption that future populations of respondents will return their questionnaires in the same time sequence as did the respondents in the current survey, similar surveys in the future which utilize the same sampling scheme can then report the estimate for this parameter at the end of the stabilization week.

The postulated model functionally relates the cumulative point estimate, the dependent variable, to the number of the week in which this point estimate was made, the independent variable. In particular, the cumulative point estimates are assumed to be linearly related to week number by two different straight lines. This relationship can be stated as:

$$E[\overline{Y}_i] = \begin{cases} \beta_1 + \beta_2 i, \; i = 1, \ldots, K-1 & \text{(first line)} \\ \beta_1 + \beta_2 K + \beta_3(i-K) + \beta_4; \; i = K, \ldots p & \text{(second line)} \end{cases}$$

where $\overline{Y}_i$ is the cumulative point estimate (e.g., sample mean) at the end of the i^{th} week, K is the week number at which the second line enters the model, and p is the last week for which questionnaires were collected (p equals 9 for this study). Note that in the equation of the second line the quantity $(\beta_1 + \beta_2 K + \beta_4)$ is a constant and that β_4 is the vertical distance between the first and second lines at week K. β_4 is positive if the second line lies above the first at week K, zero if the two lines intersect at week K, and negative if the second line lies below the first at week K. Thus, the use of K in the model allows the investigator to fit the model without knowing at what week the two lines intersect.

By defining $\overline{Y}_i$ as the cumulative point estimate made at the end of week i, it is possible to define the expected value of $\overline{Y}_i$ as the expected value of the parameter computed from the population of all centers who would respond by the end of the i^{th} week. This model tests for a difference between the expected values of the two populations of respondents defined to be those who would respond by the end of the i^{th} week and those who would respond by the end of the p^{th} (or last) week (note that the former population is a subset of the latter). Whether or not any $\overline{Y}_i$ is a good estimate of the expected value for the population defined as the entire collection of child care centers depends upon the differences between these populations and the population of non-respondents, which is not considered by this model.

The Model

The model used herein is basically that described in Draper and Smith (1966, p. 140) to fit two distinct linear time trends to a set of data. If the slope of the second line is not significantly different from 0, the intersection of the two lines provides an estimate of the point at which the mean value of the dependent variable has stabilized. (With appropriate modifications, the model used here can also be utilized to estimate proportions.)

Assume that the data is received in p equally spaced time frames (e.g. weeks).

Let $Y_{i1}, Y_{i2}, \ldots, Y_{in_i}$ be a random sample from a distribution with unknown mean and unknown variance, σ_i^2, representing the n_i data values received in the i^{th} time frame.

Let $N_i = \sum_{j=1}^{i} n_j$ represent the cumulative sample size at the end of the i^{th} time period.

Then:

$$\bar{Y}_i = \frac{\sum_{j=1}^{i}\sum_{k=1}^{nj} Y_{jk}}{N_i} = \frac{N_{i-1}\bar{Y}_{i-1} + \sum_{j=1}^{n_i} Y_{ij}}{N_i}$$

is the sample mean based on all the data received through the first i time periods.

An estimate of the population variance σ^2 at the end of p time frames is

$$S^2 = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n_i} (Y_{ij}-\bar{Y}_p)^2}{N_p-1}$$

The variance of $\bar{Y}_i$ is

$$V(\bar{Y}_i) = \frac{\sigma^2}{N_i}$$

The covariance of $\overline{Y}_i$ and $\overline{Y}_j$ $(i < j)$ is

$$\mathrm{Cov}(\overline{Y}_i, \overline{Y}_j) = \mathrm{Cov}\left[\overline{Y}_i, \frac{N_i\overline{Y}_i + \sum_{k=i+1}^{j}\sum_{m=1}^{n_k} Y_{km}}{N_j}\right]$$

$$= \frac{\sigma^2}{N_j}$$

Thus, the variance-covariance matrix of the $\{\overline{Y}_i;\ i = 1, \ldots, p\}$ is

$$V = \sigma^2 \begin{bmatrix} \frac{1}{N_1} & \frac{1}{N_2} & \frac{1}{N_3} & \cdots & \frac{1}{N_p} \\ \frac{1}{N_2} & \frac{1}{N_2} & & & \\ & & & & \\ \frac{1}{N_p} & \cdot & \cdot & \cdots & \frac{1}{N_p} \end{bmatrix}$$

The linear model used to fit the p cumulative means is

$$\overline{Y}_i = \beta_1 + \beta_2 X_{1i} + \beta_3 X_{2i} + \beta_4 X_{3i} + e_i;\ i = 1, \ldots, p$$

where $e_1, \ldots, e_p \sim MVN(0,V)$.

Let K be a guess as to where the two lines intersect. Then

$$X_{1i} = \begin{cases} i; & i = 1, \ldots, K \\ K; & i = K+1, \ldots, p \end{cases}$$

$$X_{2i} = \begin{cases} 0; & i = 1, \ldots, K \\ i-K; & i = K+1, \ldots, p \end{cases}$$

$$X_{3i} = \begin{cases} 0; & i = 1, \ldots, K-1 \\ 1; & i = K, \ldots, p \end{cases}$$

Defining the model in this manner produces the following regression parameters

β_1 = intercept of line 1

β_2 = slope of line 1

β_3 = slope of line 2

β_4 = vertical distance from line 1 to line 2 at the time point K

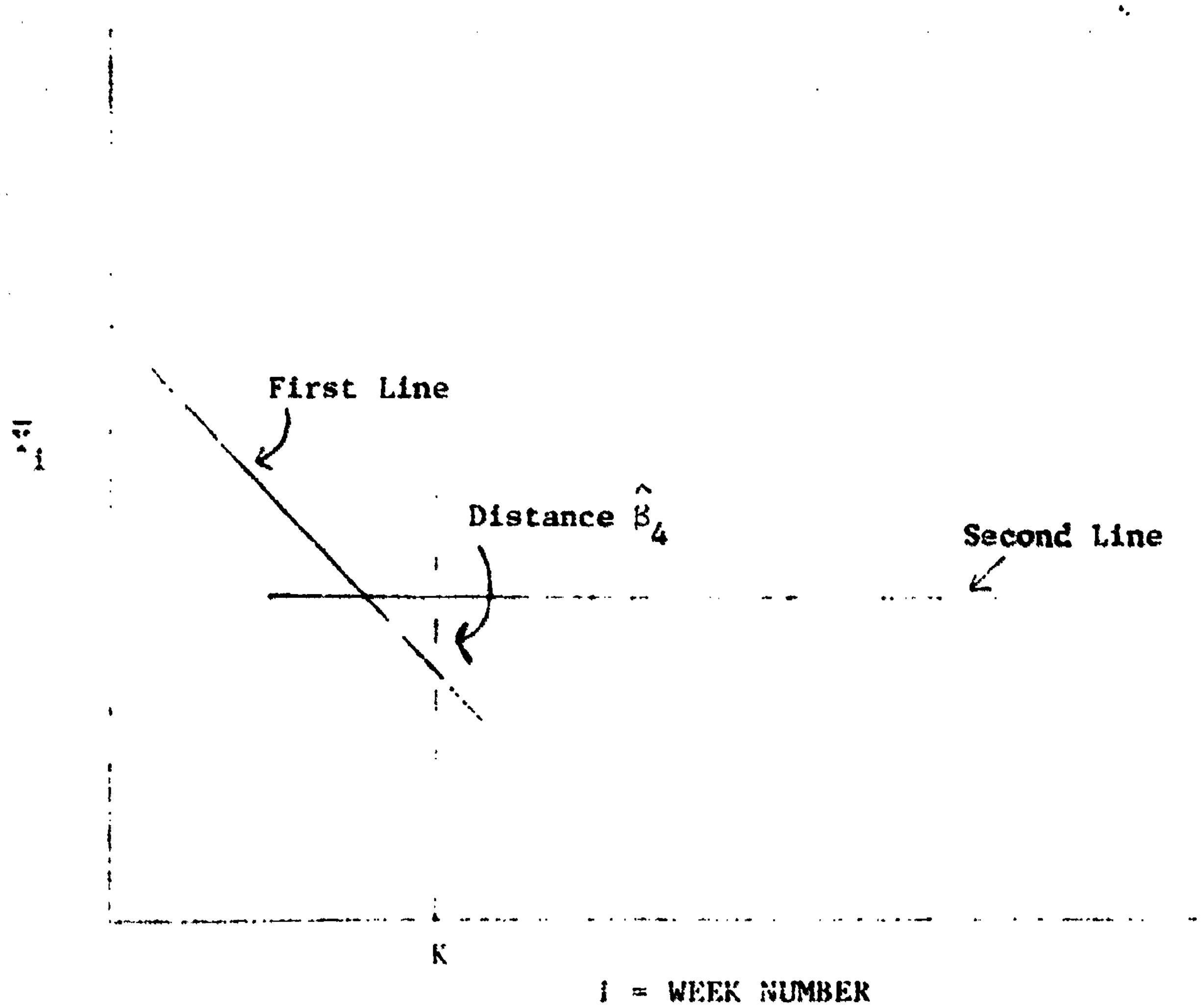

FIGURE 1. Illustration Of The Plot Of The Trend Lines.

Let $$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$$

$$X = \begin{bmatrix} X_{11} & X_{21} & X_{31} \\ X_{12} & X_{22} & X_{32} \\ \vdots & \vdots & \vdots \\ X_{1p} & X_{2p} & X_{3p} \end{bmatrix}$$

$$Y = \begin{bmatrix} \overline{Y}_1 \\ \vdots \\ \overline{Y}_p \end{bmatrix}$$

and

$$e = \begin{bmatrix} e_1 \\ \vdots \\ e_p \end{bmatrix}$$

then the model can be written as

$$Y = X\beta + e$$

Estimates of the regression coefficients are provided by

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y \quad .$$

The variance-covariance matrix of the estimated regression parameters is

$$V(\hat{\beta}) = (X'V^{-1}X)^{-1}$$

where σ^2 will have to be estimated by s^2. A test of H_0: $\beta_3 = 0$ is based on

$$t = \frac{\hat{\beta}_3}{\sqrt{V(\hat{\beta}_3)}}$$

Which in this case has a t distribution with $N_p - 1$ df.

If β_3 is not significantly different from 0, the intersection of the two lines will represent a stabilization point. This point, say Z, is the point where

$$\hat{\beta}_1 + \hat{\beta}_2 Z = \hat{\beta}_1 + \hat{\beta}_2 K + \hat{\beta}_3 (Z-K) + \hat{\beta}_4$$

Solution of this equation for Z yields

$$Z = K + \frac{\hat{\beta}_4}{\hat{\beta}_2 - \hat{\beta}_3}$$

As noted previously, the same technique can be applied to estimate a stopping point using proportions.

The dependent variable used is

$$p_i = \frac{\sum_{k=1}^{i} \sum_{j=1}^{n_k} Y_{kj}}{N_i} \qquad \text{where } Y_{kj} \sim \text{binomial } (1,p) \quad .$$

The variance of any Y_{kj} is estimated as

$$\hat{p}_p(1-\hat{p}_p) \quad .$$

Procedures

Sampling

In January, 1973, 1,120 child care centers had been identified in Pennsylvania. All programs were mailed a questionnaire packet containing a letter describing the questionnaire and its purpose, a four page questionnaire, and a stamped envelope for return of the completed questionnaire. Four weeks after this first mailing, a second packet was sent to programs from whom no response had yet been received. This second packet contained a polite reminder letter requesting that the questionnaire be completed and returned promptly, a copy of the questionnaire, and a stamped envelope for return of the completed questionnaire. Four weeks after this second mailing an attempt was made to telephone each of the remaining non-respondent centers. When a center was reached by telephone, if possible the information on the questionnaire was taken over the phone. Otherwise, a request was made to return the completed questionnaire as soon as possible. (By this time, none of the contacted non-respondent centers reported that they had not received a copy of the questionnaire.)

The resulting sample of completed questionnaires was obtained by accepting only those which had been received by the end of the tenth week after the first mailing. Subsamples of returned questionnaires were formed on a weekly basis beginning with the week after the first mailing. In all, nine subsamples were formed consisting of a total of 545 returned questionnaires. As can be seen from the data presented in Table 1, this was approximately 48.7% of the total number of questionnaires originally mailed.

Table 1 - Weekly Returns Of Mailed Questionnaires

Weekly Subsamples	Frequency	Frequency Of Respondents (Percent)	Cumulative Frequency Of Total Identified Programs (Percent)
1	39	7.2	3.5
2	131	31.2	15.2
3	56	41.5	20.2
4	49	50.5	24.6
5	39	57.6	28.0
6	20	61.3	29.8
7	79	75.8	36.9
8	35	82.2	40.0
9	97	100.0	48.7
Total	545		

Questionnaire Description

The questionnaire consisted of two pages printed on front and back. Information requested on each page was considered to be basic and easily obtainable by the director of the center. The categories of information contained in the questionnaire (aside from address information) are listed in Table 2. (Note: Categories in Table 2 which are not designated with either * or ** were not included in the analysis reported in this study.)

Table 2 - Categories Of Information
Requested On Mailed Questionnaire

Category	Description
1**	Is program a Head Start center?
2**	Is program a nursery school?
3**	Open during summer?
4**	How is program licensed?
5**	Are exceptional children provided with care?
6**	Is program profit or non-profit?
7*	Maximum number of children allowed present at one time.
8*	Number of weeks program operates during year.
9*	Total number of child-care-staff hours in a week
10*	Total enrollment.
11	Number of children attending part-time, half-time, and full-time.
12*	Number of hours per week considered part-time, half-time, and full-time.
13**	Number of children enrolled who are in one of six ethnic categories.
14**	Number of children enrolled who are in one of seven age categories.
15**	Daily hours of operation
16*	Average daily attendance.
17	Number of meals served each month.
18	Number of children provided with transportation to and from program.
19*	Number of children enrolled who have working mothers.
20	Type of fee charged per week for full-time child care.

*Mean computed on this information
**Proportion computed on this information

Results

The model described earlier was used to analyze the information in the categories presented in Table 2. A value of K = 5 (where K is the "guess" at the intersection of the two lines) was used for the analyses reported in this paper. In this section, three examples are presented in order to illustrate the manner in which the results of the analysis can be interpreted. (Results for all information categories are presented in the appendices.) The first example is an estimation of the intersection point on the time period axis for a mean, the average maximum allowed enrollment (information category 7). The second example is an estimation of the intersection point for a proportion, the proportion of centers either licensed or approved (information category 4). The final example is an unsuccessful estimation of an intersection point, the proportion of centers reporting Black children in attendance.

Example 1: The Average Maximum Allowed Enrollment

In order to provide estimates of the allowable maximum enrollment, each center director was asked to indicate the maximum capacity of his or her center program. Of the 545 completed questionnaires, 542 responded with complete data to this information category. The subsample (input) data for this example are given in Table 3.

Table 3 - Example 1 Input Data

Subsample	Frequency	Mean*
1	39	49.92
2	131	38.61
3	56	39.53
4	49	37.58
5	37	38.22
6	20	37.79
7	78	35.13
8	35	34.23
9	9[illegible]	36.53
Total	542	
Standard Deviation = 28.58		

* Figures in this column are cumulative estimates. Therefore, the total mean for all 542 returns is that given for subsample 9.

The data presented in Table 3 were used to obtain a solution for this example. The resulting regression coefficient estimates are presented in Table 4.

Table 4 - Regression Coefficient Estimates For Example 1

Coefficient	Estimate	Standard Deviation	t-ratio for H_0: $\beta_i=0$
β_1	42.3048	3.0448	13.8939*
β_2	- 1.2042	0.6383	- 1.8867
β_3	- 0.3995	0.2363	- 1.6903**
β_4	1.8442	0.8716	2.1140*

*$p<.05$, df = 541

**$p>.05$, retain H_0: $\beta_3 = 0$

Table 5 - Mean Estimates Predicted For Example 1

Subsample	Predicted Means	Residual (From Input Data)
1	41.1006	8.8194
2	39.8963	-1.2863
3	38.6291	.8379
4	37.4879	.0921
5	38.1279	.0921
6	37.7284	.0616
7	37.3289	-2.1989
8	36.9295	-2.6995
9	36.5300	0.0

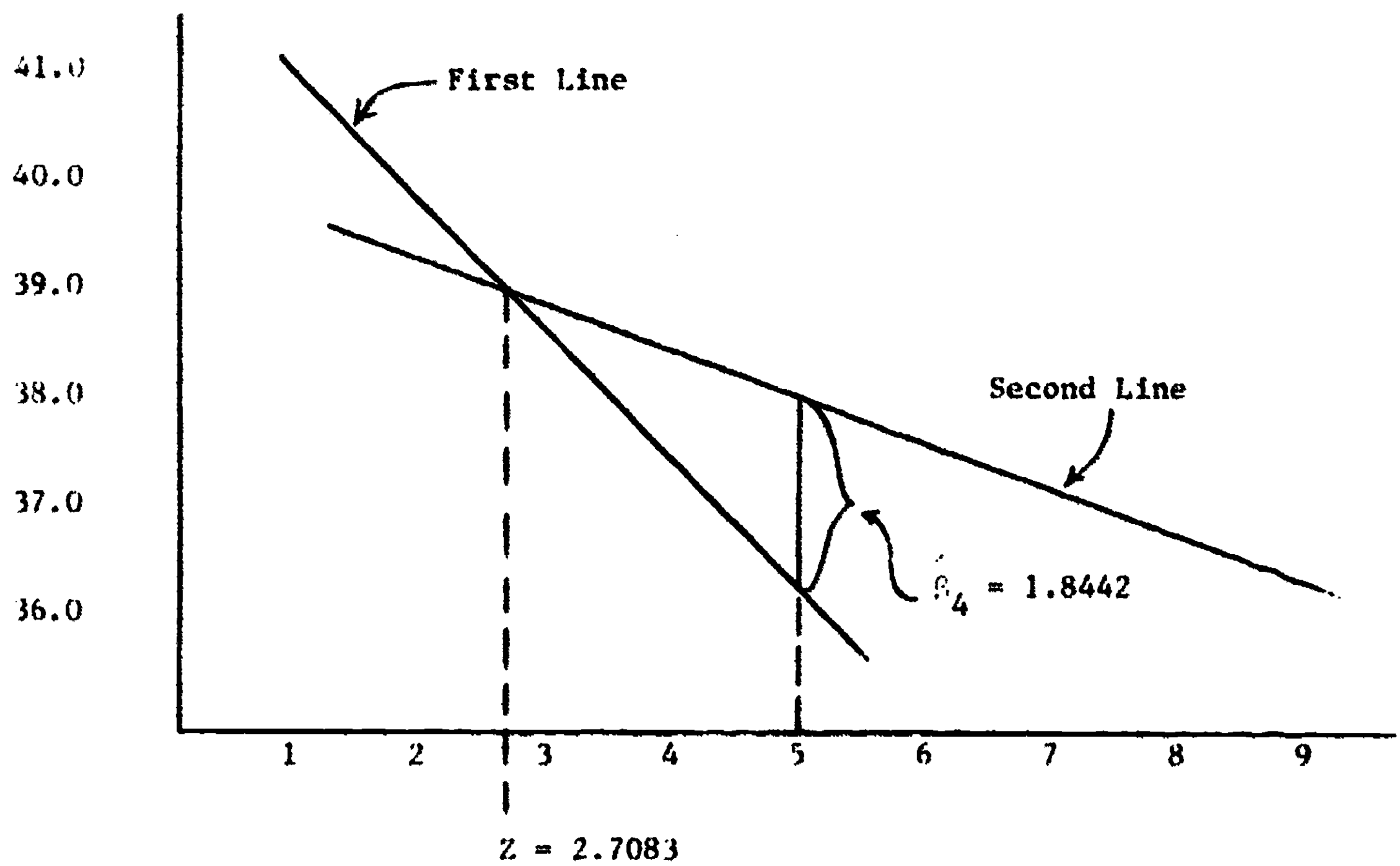

FIGURE 2. Plot Of The Trend Lines For Example 1.

Using the information presented in Table 4, the estimate of the intersection point (or the stopping point) can be obtained by:

$$Z = K + \frac{\hat{\beta}_4}{\hat{\beta}_2 - \hat{\beta}_3}$$

$$= 5 + \frac{1.8442}{-1.2042 - (-0.3995)}$$

$$= 2.7083$$

That is, it would have been possible to collect data only up to the end of the third weekly subsample of returns, (note: between 2.1 and 3.0 is interpreted as being during the third weeks of the sampling time period). The estimate of the average maximum allowable enrollment would have been essentially the same as that obtained from the cumulative estimate at the end of the ninth week in the sampling time period.

Figure 2 is a plot of the two trend lines predicted by the model. The sampling time period (i.e., week number) is represented along the horizontal axis; the predicted values of the average maximum capacity are plotted against the horizontal axis. $\hat{\beta}_4$ is a step change (positive in this case) and indicates the distance the second trend line lies above the first line at the value of K.

Conclusion

Results of the analysis suggest that the cumulative estimate of the mean was essentially the same at (approximately) the end of week three as it was at the end of week nine in the sampling time period. For purposes of estimation of this mean, therefore, it would have been possible to stop collecting further data after this time and to base decisions made regarding the mean maximum capacity upon data accumulated to the end of week three.

Example 2: The Proportion of Centers Either Licensed or Approved

Since licensing categories are oftentimes very confusing, especially in the case of Pennsylvania day care, it was necessary to collect information from each center director regarding the licensing status of the center. The director of each center was asked to fill in the appropriate licensing status and the appropriate licensing agency(ies). For purposes of this analysis, this information was dichotomized into two subcategories: (1) licensed or approved, and (2) not licensed and/or not approved. 490 centers responded to this information category with complete information. The subsample (or input) data for this example are given in Table 6.

Table 6 - Example 2 Input Data

Subsample	Frequency	Proportion*
1	35	.94286
2	116	.92715
3	50	.93035
4	44	.93061
5	31	.93479
6	16	.93151
7	76	.93479
8	31	.93985
9	91	.95102
Total	490	
Standard Deviation	.2158	

*Figures in this column are cumulative estimates. Therefore, the total proportion for all 490 returns is that given for subsample 9.

The resulting regression coefficient estimates, using the data presented in Table 6, are presented in Table 7.

Table 7 - Regression Coefficient Estimates For Example 2

Coefficient	Estimate	Standard Deviation	t-ratio for H_0: $\beta_i = 0$
β_1	.9325	.02435	38.2987*
β_2	.00086	.00513	.1671
β_3	.00271	.00187	1.4477**
β_4	.00332	.00690	.4815

* $p < .05$, df = 489
** $p > .05$, retain H_0: $\beta_3=0$

It is evident that the slope of the second line, β_3, is not significantly different from 0. This is interpreted to mean that the intersection of the two trend lines will occur at a point, Z, at which the estimate of the proportion, $\hat{p}_3$, will be essentially the same as the cumulative estimate obtained at the end of the ninth week of the sampling time period.

The predicted subsample proportions, $\hat{p}_j$, obtained from the analysis are presented in Table 8.

Table 8 - Proportion Estimates Predicted For Example 2

Subsample	Predicted Proportion	Residual (From Input Data
1	.93345	.00941
2	.93430	.00715
3	.93516	.00481
4	.93602	.00541
5	.94020	.00541
6	.94290	.01139
7	.94561	.01082
8	.94831	.00846
9	.95102	0.0

Using the information presented in Table 7, the estimate of the point of intersection can be obtained by:

$$Z = K + \frac{\hat{\beta}_4}{\hat{\beta}_2 - \hat{\beta}_3}$$

$$= 5 + \frac{.003323}{.000857 - .002706}$$

$$= 3.19403$$

It would have been possible, therefore, to collect data only up to about the end of the third weekly subsample of returns. The estimate of the proportion of licensed or approved programs would have been essentially the same (i.e., not significantly different) as that estimate obtained at the end of the ninth week in the sampling time period.

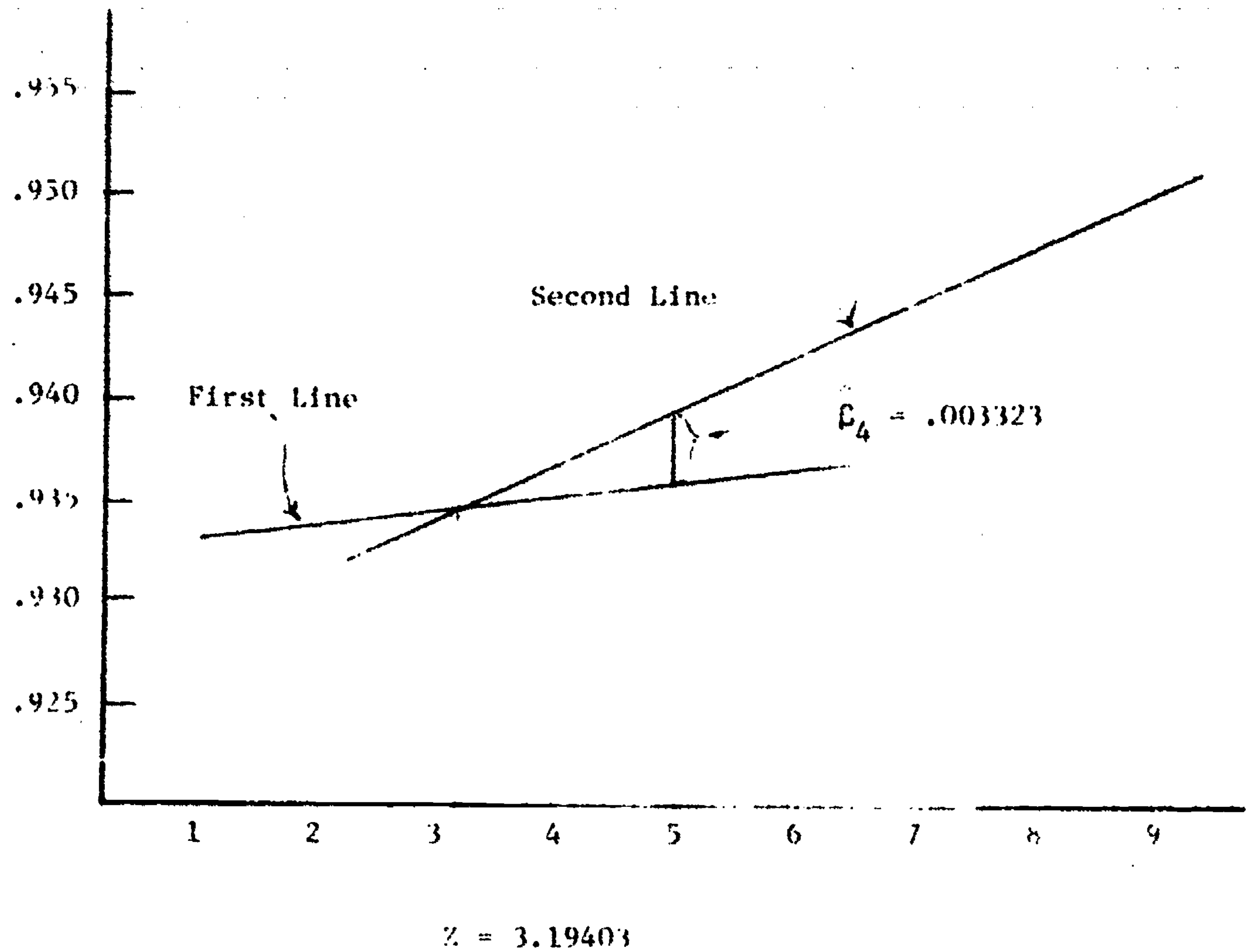

Figure 3. Plot Of The Trend Lines For Example 2.

Figure 3 is the plot of the two fitted (trend) lines. It is clear from this picture th t little real differences existed between any of the nine predicted subsample points. This was also noted in the analysis (presented prior to the plot).

Conclusion

Results of the analysis suggest that the cumulative estimate of the proportion was essentially the same at about the end of the third week as it was at the end of the ninth week. In fact, if the variability (or, more specifically, the proportioned estimate of the variability used by the model) had been smaller sooner in the sampling time period, the intersection could very likely have occurred as early as the end of the first week. The real difference, in other words, between $\hat{P}_1$ and $\hat{P}_9$ was small. However, for purposes of obtaining a stable estimate of the proportion of licensed or approved programs, it appears better to wait for at least the first three or possibly four subsamples of data to accumulate.

Example 3: The Proportion of Black Children of Children Attending Centers

Category 13 information (see Table 2) was felt to be important for several reasons. An intervention program such as Title IV-A Day Care must be made available to and used by the people for whom it was intended. In addition, these people must want to use this service or else the service is not benefiting the subpopulations who most need it. Finally, the program must be responsive to those for whom it is intended. While these reasons do not exhaust the rationale underlying Title IV-A Day Care, they do suggest that in order for the program to be functioning as it was intended it must include specific groups of people as part of its clientele.

In order to provide a partial description of the existing clientele of the day care programs, the centers were asked to supply (on the mailed questionnaire) the number of children in each of six ethnic categories who were attending their programs. As a preliminary estimate of the ethnic characteristics of day care recipients, therefore, the proportions of day care children in each of these categories were estimated. Since not all categories were represented in a majority of programs, the proportion described above was felt to be a better index than a mean per program in each category. The subsample data for this example are given in Table 9.

Table 9 - Example 3 Input Data

Subsample	Frequency	Proportion
1	1691	.06742
2	4162	.15360
3	2133	.21350
4	1276	.24595
5	1312	.25601
6	628	.26103
7	1926	.29426
8	786	.29180
9	4242	.40841
Total	18,156	
Standard Deviation	.49155	

*Figures in this column are total number of children attending programs in each subsample

**Figures in this column are cumulative estimates. Therefore, the total proportion for all 18,156 children attending programs in the respondent programs is that given for subsample 9.

The resulting regression coefficient estimates, using the data presented in Table 9, are given in Table 10.

Table 10 - Regression Coefficient Estimates For Example 3

Coefficient	Estimate	Standard Deviation	t-ratio for H_0: $\beta_i = 0$
β_1	.16229	.00840	19.3305*
β_2	.04132	.00171	24.1084*
β_3	.01770	.00066	26.8684*
β_4	-.03126	.00248	- 12.5823*

* $p < .05$, df = 18155 [Reject $H_0:\beta_3 = 0$.]

It is immediately apparent that the null hypothesis, $H_0:\beta_3 = 0$, is rejected. This suggests that the intersection point may not be either (1) within the sampling time period or (2) useful for estimating the proportion of Black children of children attending centers. The input data for subsample 9 indicated that the final week of the sampling period was not typical of the other eight weeks. To some degree, therefore, it is not surprising that the analysis suggests the subsamples may not be uniform on this attribute.

It is also evident from the regression coefficient estimates that the estimate of the slope of the first trendline, $\hat{\beta}_2$, is greater than that for the second trend line, $\hat{\beta}_3$. This is reflected in the sign of $\hat{\beta}_4$, which is negative and indicates the second line lies below the first at K = 5. It is further apparent from the predicted subsample proportions obtained for this example (see Table 11) that the model may not provide useful estimates.

Table 11 - Proportion Estimates Predicted For Example 3

Subsample	Predicted Proportion	Residual (From Input Data)
1	.20361	- .13619
2	.24493	- .09133
3	.28624	- .07274
4	.32756	- .08161
5	.33762	- .08161
6	.35531	- .09429
7	.37301	- .07875
8	.39071	- .09897
9	.40841	0.0

The size of the residuals $(\overline{P}_j - \hat{P}_j)$ suggests that the model is predicting larger proportions at each subsample point than occurred in the data. Although this is desirable from a predictive viewpoint, it does suggest that the data may not conform closely enough to the model to be usefully analyzed in this manner.

Using the information given in Table 10, the estimate of Z can be obtained by:

$$Z = K + \frac{\hat{\beta}_4}{\hat{\beta}_2 - \hat{\beta}_3}$$

$$= 5 + \frac{-.03126}{.04132 - .01770}$$

$$= 3.67655$$

According to this estimate of the intersection of the two fitted (trend) lines, the sampling could have ceased at about the end of the fourth week. However, since β_4 cannot be said to be zero, it is questionable whether the true proportions at the end of the fourth and ninth week of the sampling period are the same.

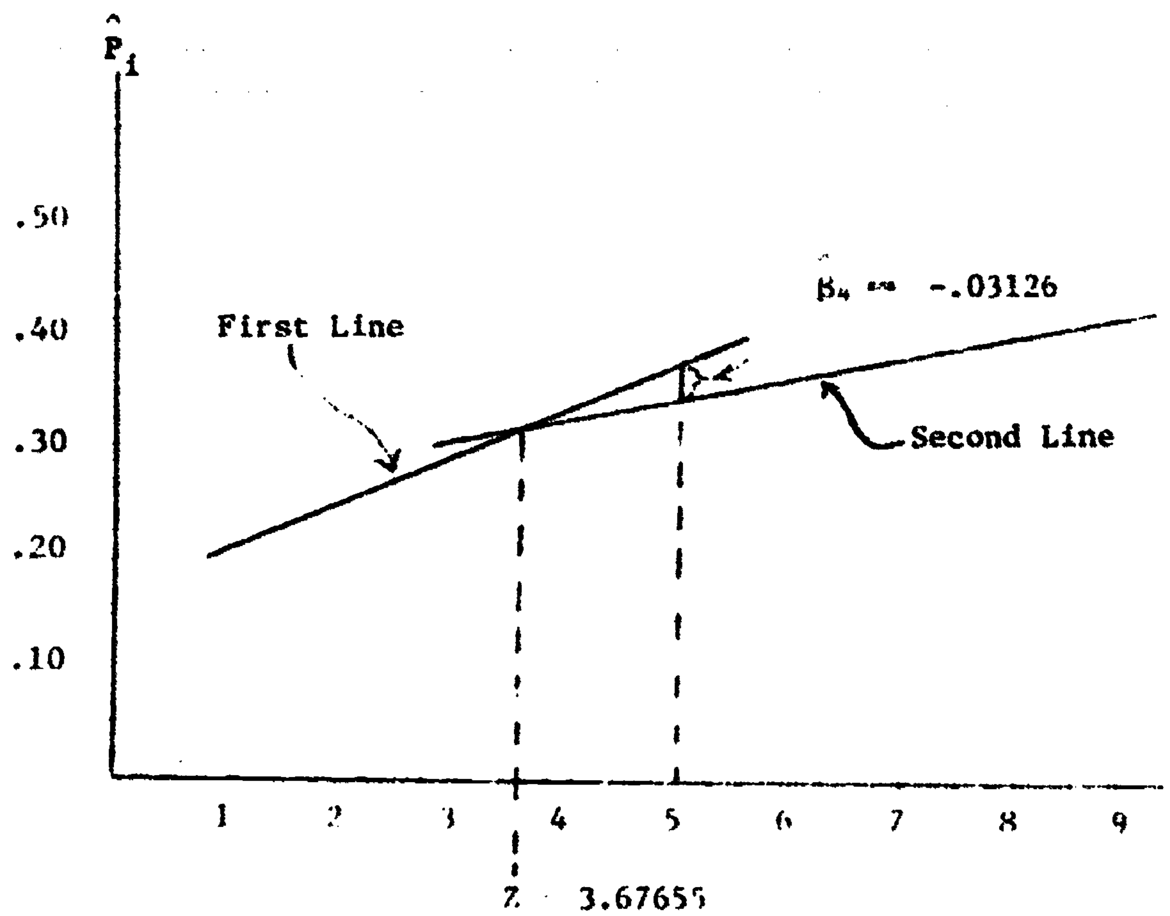

Figure 4. Plot Of The Trend Lines For Example 3.

Figure 4 is the plot of the two fitted lines. It is evident that, although the second line does pass through (or predict) the same value of $P_9 = \hat{P}_9 = .40841$, the second line is not horizontal (i.e., $\beta_3 \neq 0$). Therefore, the estimated point of intersection does not represent a stopping point which is the same as P_9.

Conclusion

Results of this analysis have been presented to show an "unsuccessful" application. In fact, however, the model was successful with respect to indicating that there were differences among the subsample cumulative proportions. Accepting $Z = 3.67655$ (or approximately the end of the fourth weekly subsample) as a suitable stopping point will depend upon the concern of the investigator regarding the importance of the difference of $(\hat{P}_4 = .40841) - (\hat{P}_9 = .24595) = .16246$.

Summary

One of the problems facing researchers who use a mailed questionnaire to collect data is the relative lack of investigator control over recipients returning their completed questionnaires. The point at which the investigator decides to cease waiting for further returns, however, can have important consequences in terms of the utility of his data. The model described in this paper was applied to this situation in order to estimate a stopping point in the waiting period at which the final parameter estimate and the estimate at the stopping point were not significantly different.

Examples discussed within the report indicated that useful information was provided regarding the variables analyzed. In the case of the two "successful" applications, a stopping point was found at which the parameter estimates were not significantly different than the final estimates. In the case of the "unsuccessful" application, it was noted that data collected on this variable did not provide a stabilized parameter estimate even though a stopping point prior to the end of the sampling time period was obtained. Such a failure as this suggests that the trend was not linear and that estimation of this parameter should be deferred until at least the ninth week following the mailing of the questionnaires.

Information of this type is useful to the user of mailed questionnaires because it provides him with an estimate of the length of time he must wait to obtain stable estimates on certain parameters. Such information should also prove to be useful in the planning, analysis, and reporting of data collected in this manner.

REFERENCES

Draper, N. R. and Smith, H. *Applied Regression Analysis.* New York: John Wiley & Sons, Inc., 1966.

APPENDICES

Results Of Analyses Of The
Data In Each Of Twenty Categories
Of Information Contained In The
Mailed Questionnaire

APPENDIX A

Category 1: Proportion Of Centers That Are Head Start Programs

Table A.1 Input Data For Category 1

Subsample	Frequency	Cummulative Proportion
1	38	.18421
2	131	.18343
3	55	.17411
4	49	.22344
5	38	.22187
6	19	.22121
7	79	.24695
8	35	.24775
9	97	.22366
Total	541	
Standard Deviation	.41669	

Table A.2 Regression Coefficients For Category 1

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	.11826	.04457	2.65375*
β_2	.02776	.00935	2.96858*
β_3	-.00102	.00343	- .29700**
β_4	-.02934	.01285	-2.28293

* $p < .05$, df = 540
** $p > .05$, retain H_0: $\beta_3 = 0$

Table A.3 Predicted Proportions For Category 1

Subsample	Predicted Proportion	Residual (From Input Data)
1	.14603	.03818
2	.17388	.00964
3	.20155	-.02744
4	.22931	-.00587
5	.22773	-.00587
6	.22672	-.00550
7	.22570	.02125
8	.22468	.02307
9	.22366	0.0

Intersection point is at week 3.98055

APPENDIX B

Category 2: Proportion Of Programs That Are Nursery Schools

Table B.1 Input Data For Category 2

Subsample	Frequency	Cummulative Proportion
1	38	.18421
2	131	.24852
3	55	.23214
4	49	.20147
5	39	.19872
6	18	.20303
7	72	.18408
8	35	.17163
9	96	.14822
Total	533	
Standard Deviation	.35531	

Table B.2 Regression Coefficients For Category 2

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	.27900	.03803	7.33526*
β_2	-.02338	.00797	-2.93137*
β_3	-.00863	.00290	-2.97635*
β_4	.02063	.01102	1.87228

* $p < .05$, df = 532 [reject H_0: $\beta_3 = 0$]

Table B.3 Predicted Proportions For Category 2

Subsample	Predicted Proportions	Residual (From Input Data)
1	.25562	-.07141
2	.23225	.01628
3	.20887	.02327
4	.18549	.01597
5	.18275	.01597
6	.17412	.02892
7	.16548	.01860
8	.15685	.01477
9	.14822	0.0

Intersection point is at week 3.60136

APPENDIX C

Category 3: Proportion Of Programs Open During Summer

Table C.1 Input Data For Category 3

Subsample	Frequency	Cummulative Proportion
1	37	.81081
2	131	.68453
3	55	.72646
4	49	.70221
5	39	.69775
6	20	.69487
7	78	.68460
8	35	.69820
9	96	.75000
Total	540	
Standard Deviation	.43302	

Table C.2 Regression Coefficients For Category 3

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	.72368	.04653	15.5526*
β_2	-.00379	.00976	- .03879
β_3	.01148	.00359	3.2005*
β_4	-.00067	.01348	- .0496

* $p < .05$, df = 539 [reject H_0: $\beta_3 = 0$]

Table C.3 Predicted Proportions For Category 3

Subsample	Predicted Proportions	Residual (From Input Data)
1	.71990	.09092
2	.71611	-.03158
3	.71232	.01414
4	.70853	-.00632
5	.70408	-.00632
6	.71556	-.02069
7	.72704	-.04244
8	.73852	-.04032
9	.75000	0.0

Intersection point is at week 5.04382

APPENDIX D

Category 4: Proportion Of Centers Either Licensed Or Approved

Table D.1 Input Data For Category 4

Subsample	Frequency	Cummulative Proportion
1	35	.94286
2	116	.92715
3	50	.93035
4	44	.93061
5	31	.93479
6	16	.93151
7	76	.93479
8	31	.93985
9	91	.95102
Total	490	
Standard Deviation	.2158	

Table D.2 Regression Coefficients For Category 4

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	.93251	.02435	38.2987*
β_2	.00086	.00513	.1671
β_3	.00271	.00187	1.4477**
β_4	.00332	.00690	.4815

* $p < .05$, df = 489
** $p > .05$, retain H_0: $\beta_3 = 0$

Table D.3 Predicted Proportions For Category 4

Subsample	Predicted Proportions	Residual (From Input Data)
1	.93345	.00941
2	.93430	- .00715
3	.93516	- .00481
4	.93602	- .00541
5	.94020	- .00541
6	.94290	- .00139
7	.94561	- .01082
8	.94831	- .00846
9	.95102	0.0

Intersection point is at week 3.19403

APPENDIX E

Category 5: Proportion Of Programs With Exceptional Children Attending

Table E.1 Input Data For Category 5

Subsample	Frequency	Cummulative Proportion
1	38	.21053
2	130	.14881
3	56	.19197
4	49	.19414
5	38	.19293
6	20	.20242
7	79	.18049
8	35	.18202
9	96	.15527
Total	541	
Standard Deviation	.36215	

Table E.2 Regression Coefficients For Category 5

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	.11136	.03887	2.86501*
β_2	.01457	.00816	1.78486
β_3	-.00329	.00300	-1.09555**
β_4	-.01578	.01119	-1.40965

* $p < .05$, df = 540

** $p > .05$, retain $H_0: \beta_3 = 0$

Table F.3 Predicted Proportions For Category 5

Subsample	Predicted Proportions	Residual (From Input Data)
1	.12592	.08461
2	.14049	.00832
3	.15506	.03691
4	.16962	.02452
5	.16841	.02452
6	.16512	.03729
7	.16184	.01865
8	.15855	.02347
9	.15527	0.0

Intersection point is at week 4.11608

APPENDIX F

Category 6: Proportion Of Centers That Are Run For Profit

Table F. 1 Input Data For Category 6

Subsample	Frequency	Cummulative Proportion
1	38	.18421
2	130	.08929
3	56	.08036
4	49	.07326
5	39	.06410
6	20	.06024
7	79	.05839
8	34	.05618
9	96	.04806
Total	541	
Standard Deviation	.21389	

Table F.2 Regression Coefficients For Category 6

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	.11186	.02295	4.87407*
β_2	-.00991	.00482	-2.05647*
β_3	-.00375	.00176	-2.13209*
β_4	.00076	.00665	.11367

* $p < .05$, df = 540 [reject H_0: $\beta_3 = 0$]

Table F.3 Predicted Proportions For Category 6

Subsample	Predicted Proportion	Residual (From Input Data)
1	.10195	.08226
2	.09204	- .00275
3	.08213	- .00177
4	.07221	.00105
5	.06306	.00105
6	.05931	.00093
7	.05556	.00284
8	.05181	.00437
9	.04806	0.0

Intersection point is at week 4.87741

APPENDIX G

Category 7: Average Maximum Capacity

Table G.1 Input Data For Category 7

Subsample	Frequency	Cummulative Mean
1	39	49.92
2	131	38.61
3	56	39.53
4	49	37.58
5	37	38.22
6	20	37.79
7	78	35.13
8	35	34.23
9	97	36.53
Total	542	
Standard Deviation	28.58	

Table G.2 Regression Coefficients For Category 7

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	42.3048	3.0448	13.8939*
β_2	- 1.2042	.6383	- 1.8867
β_3	- .3995	.2363	- 1.6903**
β_4	1.8442	.8716	2.1160*

* $p < .05$, df = 541
** $p > .05$, retain H_0: $\beta_3 = 0$

Table G.3 Predicted Means For Category 7

Subsample	Predicted Mean	Residual (From Input Data)
1	41.1006	8.8194
2	39.8963	-1.2863
3	38.6291	.8379
4	37.4879	.0921
5	38.1279	.0921
6	37.7284	.0616
7	37.3289	-2.1989
8	36.9295	-2.6995
9	36.5300	0.0

Intersection point is at week 2.7083

APPENDIX H

Category 8: Average Number Of Weeks Program Is Open During Year

Table H.1 Input Data For Category 8

Subsample	Frequency	Cummulative Mean
1	39	46.26
2	131	45.26
3	56	45.91
4	48	45.59
5	38	45.72
6	19	45.76
7	78	45.78
8	34	46.08
9	97	46.99
Total	540	
Standard Deviation	7.85001	

Table H.2 Regression Coefficients For Category 8

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	45.81530	.83298	55.00160*
β_2	− .00863	.17437	− .04949
β_3	.26980	.06417	4.20455*
β_4	.13863	.24041	.57664

* $p < .05$, df = 539 [reject H_0: $\beta_3 = 0$]

Table H.3 Predicted Means For Category 8

Subsample	Predicted Mean	Residual (From Input Data)
1	45.8067	.45329
2	45.7981	- .53808
3	45.7895	.12055
4	45.7808	- .19082
5	45.9108	- .19082
6	46.1806	- .42062
7	46.4504	- .67041
8	46.7202	- .66021
9	46.9900	0.0

Intersection point is at week 4.50210

APPENDIX I

Category 9: Average Child Care Staff Hours Per Week

Table I.1 Input Data For Category 9

Subsample	Frequency	Cummulative Mean
1	39	181.28
2	130	164.69
3	55	186.03
4	48	181.40
5	39	192.01
6	20	190.88
7	79	196.66
8	34	192.36
9	91	206.63
Total	535	
Standard Deviation	306.9799	

Table I.2 Regression Coefficients For Category 9

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	175.19900	32.74190	5.35091*
β_2	3.96848	6.85004	.57934
β_3	1.23672	2.52086	.49059**
β_4	6.64152	9.50629	.69864

* $p < .05$, $df = 534$
** $p > .05$, retain H_0: $\beta_3 = 0$

Table I.3 Predicted Means For Category 2

Subsample	Predicted Mean	Residual (From Input Data)
1	179.168	2.11232
2	183.136	-18.44620
3	187.105	- 1.07465
4	191.073	- 9.67313
5	201.683	- 9.67313
6	202.920	-12.03990
7	204.157	- 7.49657
8	205.393	-13.03330
9	206.630	0.0

Intersection point is at week 7.43122

APPENDIX J

Category 10: Average Total Enrollment

Table J.1 Input Data For Category 10

Subsample	Frequency	Cummulative Mean
1	39	43.74
2	128	36.11
3	56	36.62
4	49	34.72
5	39	35.08
6	20	34.87
7	79	32.85
8	33	32.21
9	97	34.32
Total	540	
Standard Deviation	26.0000	

Table J.2 Regression Coefficients For Category 10

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	39.64950	2.79795	14.17090*
β_2	- 1.22892	.58821	- 2.08924*
β_3	- .19345	.21347	- .90622**
β_4	1.58892	.81096	1.95931

* $p < .05$, df = 539
** $p > .05$, retain H_0: $\beta_3 = 0$

Table J.3 Predicted Means For Category 10

Subsample	Predicted Mean	Residual (From Input Data)
1	38.4206	5.31944
2	37.1916	-1.08164
3	35.9627	.65729
4	34.7338	- .01379
5	35.0938	- .01379
6	34.9003	- .03034
7	34.7069	-1.85690
8	34.5134	2.30345
9	34.3200	0.0

Intersection point is at week 3.46551

APPENDIX K

Categories 11 And 12 (Combined): Average Total Child Hours Per Week

Table K.1 Input Data For Categories 11 And 12 (Combined)

Subsample	Frequency	Cummulative Mean
1	37	1246.95
2	130	1009.92
3	55	1026.40
4	49	995.44
5	39	1004.79
6	20	995.88
7	79	944.73
8	33	926.59
9	95	1050.47
Total	537	
Standard Deviation	966.0502	

Table K.2 Regression Coefficients For Categories 11 And 12 (Combined)

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	1098.58000	104.1150	10.55170*
β_2	- 19.58040	21.8801	- .89490
β_3	5.21437	7.9409	.65665**
β_4	28.93040	30.1991	.95800

* $p < .05$, df = 536
** $p > .05$, retain H_0: $\beta_3 = 0$

Table K.3 Predicted Means For Categories 11 And 12 (Combined)

Subsample	Predicted Mean	Residual (From Input Data)
1	1079.00	167.9460
2	1059.42	- 49.5034
3	1039.84	- 13.4429
4	1020.26	- 24.8225
5	1029.61	- 24.8225
6	1034.83	- 38.9469
7	1040.04	- 95.3113
8	1045.26	-118.6660
9	1050.47	0.0

Intersection point is at week 3.83321

APPENDIX L

Category 13: Proportion Of Black Children Of Children Attending Centers

Table L.1 Input Data For Category 13

Subsample	Frequency	Cummulative Proportion
1	1691	.06742
2	4162	.15360
3	2133	.21350
4	1276	.24595
5	1312	.25601
6	628	.26103
7	1926	.29426
8	786	.29180
9	4242	.40841
Total	18,156	
Standard Deviation	.49155	

Table L.2 Regression Coefficients For Category 13

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	.16229	.00840	19.3306*
β_2	.04132	.00171	24.1084*
β_3	.01770	.00066	26.8684*
β_4	-.03126	.00248	-12.5823*

* $p < .05$, df = 18,155 [reject H_0: $\beta_3 = 0$]

Table L.3 Predicted Proportions For Category 13

Subsample	Predicted Proportion	Residual (From Input Data)
1	.20361	- .13619
2	.24493	- .09133
3	.28624	- .07274
4	.32756	· 08161
5	.33762	- .08161
6	.35531	- .09429
7	.37301	- .07875
8	.39071	- .09892
9	.40841	0.0

Intersection point is at week 3.67655

APPENDIX M

Category 13: Proportion Of Spanish American Children Of Children Attending Centers

Table M.1 Input Data For Category 13

Subsample	Frequency	Cummulative Proportion
1	1691	.00887
2	4162	.01418
3	2133	.01678
4	1276	.01566
5	1312	.01863
6	628	.01803
7	1926	.01722
8	786	.01703
9	4242	.02165
Total	18,156	
Standard Deviation	.14552	

Table M.2 Regression Coefficients For Category 13

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	.01805	.00249	7.26053*
β_2	.00005	.00051	.10558
β_3	.00010	.00020	.52540**
β_4	.00292	.00074	3.97272*

* $p < .05$, df = 18,155
** $p > .05$, retain H_0: $\beta_3 = 0$

Table M.3 Predicted Proportions For Category 13

Subsample	Predicted Proportions	Residual (From Input Data)
1	.01820	-.00923
2	.01815	-.00397
3	.01821	-.00143
4	.01826	-.00261
5	.02124	-.00261
6	.02134	-.00331
7	.02144	-.00423
8	.02154	.00451
9	.02165	0.0

Intersection point is at week -53.40000

APPENDIX N

Category 13: Proportion Of White Children Of Children Attending Centers

Table N.1 Input Data For Category 13

Subsample	Frequency	Cummulative Proportion
1	1691	.91721
2	4162	.82641
3	2133	.76384
4	1276	.73267
5	1312	.71893
6	628	.71452
7	1926	.68259
8	786	.68535
9	4242	.56163
Total	18,156	
Standard Deviation	.49613	

Table N.2 Regression Coefficients For Category 13

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	.81220	.00847	95.84990*
β_2	-.04126	.00173	-23.85080*
β_3	-.01795	.00066	-26.99890*
β_4	.02752	.00251	10.97310*

* $p < .05$, df = 18,155 [reject H_0: $\beta_3 = 0$]

Table N.3 Predicted Proportions For Category 13

Subsample	Predicted Proportion	Residual (From Input Data)
1	.77094	.14627
2	.72968	.09673
3	.68843	.07541
4	.64717	.08550
5	.63343	.08550
6	.61548	.09903
7	.59753	.08506
8	.57958	.10577
9	.56163	0.0

Intersection point is at week 3.81933

APPENDIX O

Category 14: Proportion Of Children
Attending Centers Who Are Less Than Two Years Old

Table 0.1 Input Data For Category 14

Subsample	Frequency	Cummulative Proportion
1	1673	.00060
2	4319	.00801
3	2131	.00665
4	1271	.01128
5	1446	.01347
6	630	.01290
7	1944	.01178
8	785	.01254
9	4240	.01020
Total	18,439	
Standard Deviation	.10048	

Table 0.2 Regression Coefficients For Category 14

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	-.00329	.00169	-1.94401
β_2	.00310	.00034	9.01481*
β_3	-.00028	.00013	-2.10856*
β_4	-.00092	.00051	-1.79464

* $p < .05$, df = 18,438 [reject H_0: $\beta_3 = 0$]

Table 0.3 Predicted Proportions For Category 14

Subsample	Predicted Proportion	Residual (From Input Data)
1	-.00019	.00079
2	.00292	.00509
3	.00602	.00063
4	.00912	.00216
5	.01131	.00216
6	.01103	.00187
7	.01075	.00103
8	.01047	.00206
9	.01020	0.0

Intersection point is at week 4.72839

APPENDIX P

Category 14: Proportion Of Children Attending Centers Who Are Two Years Old

Table P.1 Input Data For Category 14

Subsample	Frequency	Cummulative Proportion
1	1673	.06874
2	4319	.03722
3	2131	.03275
4	1271	.03332
5	1446	.03303
6	630	.03130
7	1944	.03019
8	785	.02923
9	4240	.02321
Total	18,439	
Standard Deviation	.15059	

Table P.2 Regression Coefficients For Category 14

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	.03724	.00254	14.68010*
β_2	-.00165	.00052	- 3.20547*
β_3	-.00178	.00020	- 9.00906*
β_4	.00136	.00077	1.77428

* $p < .05$, df = 18,438 [reject H_0: $\beta_3 = 0$]

Table P.3 Predicted Proportions For Category 14

Subsample	Predicted Proportion	Residual (From Input Data)
1	.03559	.03315
2	.03393	.00328
3	.03228	.00047
4	.03063	.00269
5	.03033	.00269
6	.02855	.00275
7	.02677	.00342
8	.02500	.00423
9	.02321	0.0

Intersection point is at week 15.77440

APPENDIX Q

Category 14: Proportion Of Children Attending Centers Who Are Three Years Old

Table Q.1 Input Data For Category 14

Subsample	Frequency	Cummulative Proportion
1	1673	.26778
2	4319	.23632
3	2131	.22676
4	1271	.22163
5	1446	.22140
6	630	.22424
7	1944	.23513
8	785	.23621
9	4240	.24378
Total	18,439	
Standard Deviation	.42937	

Table Q.2 Regression Coefficients For Category 14

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	.25732	.00723	35.57650*
β_2	-.00703	.00147	- 4.78202*
β_3	.00370	.00056	6.57654*
β_4	.00681	.00219	3.11214*

* $p < .05$, df = 18,438 [reject H_0: $\beta_3 = 0$]

Table Q.3 Predicted Proportions For Category 14

Subsample	Predicted Proportion	Residual (From Input Data)
1	.25029	.01749
2	.24326	- .00694
3	.23622	- .00946
4	.22919	- .00756
5	.22896	- .00756
6	.23266	- .00843
7	.23637	- .00124
8	.24007	- .00386
9	.24378	0.0

Intersection point is at week 4.36631

APPENDIX R

Category 14: Proportion Of Children Attending Centers Who Are Four Years Old

Table R.1 Input Data For Category 14

Subsample	Frequency	Cummulative Proportion
1	1673	.39689
2	4319	.43975
3	2131	.40687
4	1271	.41729
5	1446	.41347
6	630	.41038
7	1944	.41300
8	785	.41172
9	4240	.42692
Total	18,439	
Standard Deviation	.49468	

Table R.2 Regression Coefficients For Category 14

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	.43042	.00833	51.65050*
β_2	-.00031	.00169	- .18058
β_3	.00039	.00065	.59642**
β_4	-.00351	.00252	- 1.39439

* $p < .05$, df = 18,438
** $p > .05$, retain H_0: $\beta_3 = 0$

Table R.3 Predicted Proportions For Category 14

Subsample	Predicted Proportion	Residual (From Input Data)
1	.43011	- .03322
2	.42980	.00995
3	.42950	- .02263
4	.42919	- .01190
5	.42537	- .01190
6	.42576	- .01538
7	.42615	.01315
8	.42653	- .01481
9	.42692	0.0

Intersection point is at week 10.06810

APPENDIX S

Category 14: Proportion Of Children Attending Centers Who Are Five Years Old

Table S.1 Input Data For Category 14

Subsample	Frequency	Cummulative Proportion
1	1673	.19785
2	4319	.19509
3	2131	.19993
4	1271	.19502
5	1446	.18939
6	630	.19911
7	1944	.19085
8	785	.19346
9	4240	.18944
Total	18,439	
Standard Deviation	.39189	

Table S.2 Regression Coefficients For Category 14

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	.19916	.00660	30.16860*
β_2	-.00224	.00134	- 1.66686
β_3	.00121	.00051	2.35901*
β_4	-.00339	.00200	- 1.69806

* $p < .05$, df = 18,438 [reject $H_0: \beta_3 = 0$]

Table S.3 Predicted Proportions For Category 14

Subsample	Predicted Proportions	Residual (From Input Data)
1	.19692	.00092
2	.19469	.00041
3	.19245	.00748
4	.19021	.00481
5	.18458	.00481
6	.18580	.00531
7	.18701	.00384
8	.18822	.00524
9	.18944	0.0

Intersection point is at week 5.98208

APPENDIX T

Category 14: Proportion Of Children Attending Centers Who Are Six To Twelve Years Old

Table T.1 Input Data For Category 14

Subsample	Frequency	Cummulative Proportion
1	1673	.04782
2	4319	.06726
3	2131	.10144
4	1271	.09708
5	1446	.10166
6	630	.10227
7	1944	.09490
8	785	.09205
9	4240	.08737
Total	18,439	
Standard Deviation	.28231	

Table T.2 Regression Coefficients For Category 14

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	.06624	.00476	13.92930*
β_2	.00662	.00097	6.84825*
β_3	-.00249	.00037	- 6.71193*
β_4	-.00205	.00144	- 1.42338

* $p < .05$, df = 18,438 [reject $H_0: \beta_3 = 0$]

Table T.3 Predicted Proportions For Category 14

Subsample	Predicted Proportion	Residual (from Input Data)
1	.07287	- .02505
2	.07949	- .01223
3	.08611	.01533
4	.09274	.00435
5	.09731	.00435
6	.09483	.00744
7	.09234	.00256
8	.08986	.00219
9	.08737	0.0

Intersection point is at week 4.77535

APPENDIX U

Category 14: Proportion Of Children Attending Centers Who Are Twelve Years Old Or Older

Table U.1 Input Data For Category 14

Subsample	Frequency	Cummulative Proportion
1	1673	.0203
2	4319	.0163
3	2131	.0256
4	1271	.0243
5	1446	.0275
6	630	.0278
7	1944	.0241
8	785	.0247
9	4240	.0190
Total	18,439	
Standard Deviation	.13579	

Table U.2 Regression Coefficients For Category 14

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	.01296	.00229	5.66456*
β_2	.00147	.0047	3.15584*
β_3	-.00076	.00018	-4.25168*
β_4	.00173	.00069	2.50428*

* $p < .05$, df = 18,438 [reject H_0: $\beta_3 = 0$]

Table U.3 Predicted Proportions For Category 14

Subsample	Predicted Proportion	Residual (From Input Data)
1	.01443	.00587
2	.01589	.00041
3	.01736	.00824
4	.01883	.00547
5	.02203	.00547
6	.02127	.00653
7	.02052	.00358
8	.01976	.00494
9	.01900	0.0

Intersection point is at week 5.77814

APPENDIX V

Category 15: Average Hours Open Daily (Monday - Friday)

Table V.1 Input Data For Category 15

Subsample	Frequency	Cummulative Mean
1	38	8.33
2	128	7.88
3	54	7.99
4	48	7.92
5	38	7.94
6	19	7.90
7	79	7.96
8	32	7.96
9	96	8.34
Total	532	
Standard Deviation	2.73000	

Table V.2 Regression Coefficients For Category 15

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	8.17086	.29418	27.75530*
β_2	- .01728	.06176	- .27982
β_3	.05457	.02243	2.43297*
β_4	.03728	.08525	.43732

* $p < .05$, df = 531 [reject H_0: $\beta_3 = 0$]

Table V.3 Predicted Means For Category 15

Subsample	Predicted Mean	Residual (From Input Data)
1	8.15358	.17642
2	8.13630	- .25630
3	8.11902	- .12902
4	8.10174	- .18174
5	8.12174	- .18174
6	8.17630	- .27630
7	8.23087	- .27087
8	8.28543	- .32543
9	8.34000	0.0

Intersection point is at week 4.48114

APPENDIX W

Category 16: Average Daily
Attendance At Centers (Monday--Sunday) For Days Open

Table W.1 Input Data For Category 16

Subsample	Frequency	Cummulative Mean
1	37	34.49
2	127	30.76
3	53	30.47
4	35	29.52
5	32	29.70
6	18	29.54
7	77	27.28
8	30	28.96
9	92	30.23
Total	501	
Standard Deviation	44.60000	

Table W.2 Regression Coefficients For Category 16

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	31.14040	4.59794	6.77269*
β_2	- .81398	.93371	- .87178
β_3	.54138	.38180	1.41796**
β_4	.99398	1.32711	.74898

* $p < .05$, df = 500
** $p > .05$, retain H_0: $\beta_3 = 0$

Table W.3 Predicted Means For Category 16

Subsample	Predicted Mean	Residual (From Input Data)
1	30.3264	4.16356
2	29.5125	1.24754
3	28.6985	1.77153
4	27.8845	1.63551
5	28.0645	1.63551
6	28.6059	.93413
7	29.1472	-1.86725
8	29.6886	- .72862
9	39.2300	0.0

Intersection point is at week 4.26663

APPENDIX X

Category 19: Average Number Of
Working Mothers Per Center Whose Child(ren) Attends The Program

Table X.1 Input Data For Category 19

Subsample	Frequency	Cummulative Mean
1	39	24.08
2	128	17.32
3	56	17.14
4	49	15.69
5	39	15.83
6	20	15.80
7	79	15.23
8	33	14.73
9	97	16.90
Total	540	
Standard Deviation	22.77000	

Table X.2 Regression Coefficients For Category 19

Coefficient	Estimate	Standard Deviation	t-ratio
β_1	20.98880	2.45036	8.56558*
β_2	- 1.15601	.51514	-2.24408*
β_3	.09882	.18695	.52862**
β_4	1.29601	.71021	1.82482

* $p < .05$, df = 539
** $p > .05$, retain H_0: $\beta_3 = 0$

Table X.3 Predicted Means For Category 19

Subsample	Predicted Mean	Residual (From Input Data)
1	19.8327	4.24726
2	18.6767	-1.35673
3	17.5207	- .38071
4	16.3647	- .67470
5	16.5047	- .67470
6	16.6035	- .80353
7	16.7024	-1.47235
8	16.8012	-2.07118
9	16.9000	0.0

Intersection point is at week 3.96718