CHAPTER 3

AUTOMATING A MANUAL TELEPHONE PROCESS

Mary Batcher, Kevin Cecco, and Dennis Lin

This case study illustrates the use of statistics to evaluate a new technology that will be implemented nationally in over 30 locations if it proves successful in a pilot study. Specifically, the case study is of an interactive telephone application that will let certain types of calls to the IRS be handled without the intervention of a staff person.

When introducing new technology, it is important to consider the human interaction with that technology. The technology may function perfectly but the ability or willingness of people to use it may not be there. It is thus very important to pilot test new systems. The interactive telephone system was pilot tested and evaluated in terms of cost, customer satisfaction, and ease of use. This case study focuses on the assessment of cost, in terms of time to complete a transaction, and ease of use, in terms of the percent of users who successfully completed their transaction without requiring the assistance of IRS staff.[1] The case study illustrates the use of hypothesis testing in decision making and the use of basic time series statistics and plots to examine periodic fluctuation over time.

---

[1] The customer satisfaction measure was an automated survey of a sample of callers. The caller was asked, at the end of the transaction, to participate in a short survey. The general satisfaction level from the survey was 93.6 percent and the survey response rate was 64.4 percent.

Introduction

The problem to be solved is to assess the effectiveness of a new interactive telephone system by examining its performance in a pilot test. The goals of the statistical analysis are twofold. The first goal is to determine whether a new interactive telephone system is more efficient than the manual process it replaces, as measured by the time needed to complete a transaction. The second is to examine ease of use and the effect of extraneous occurrences in the pilot test by tracking successful completion rates over time to identify any problematic occurrences or any periodicity in the data.

Background Information

Interactive telephone systems allow us to exchange limited information using the push buttons on our touch-tone telephones. The most basic of these systems allows us to route ourselves to a recorded message or to the best person to handle our issue. More complex systems allow a complete exchange of information, with problems fully resolved in the automated setting. Interactive telephone systems are increasingly common. We use them to check bank balances, order merchandise, inquire about service, etc. They offer opportunities to decrease costs while maintaining or increasing accessibility. The IRS is developing interactive systems for some of their telephone work. When people are unable to pay all of the taxes they owe, they are able to establish installment agreements with the IRS to pay off the balance over time. One of the interactive telephone systems developed by the IRS allows people to request a payment extension or establish an installment agreement. (Additional examples of the use of telephone pushbuttons for the automated exchange of information can be found in [Nicholls & Appel, 1994], [Rosen, Clayton, and Pivetz, 1994], and [Werking, Clayton, and Harrell, 1996].)

The IRS has developed the Voice Balance Due (VBD) fully automated telephone system. It permits people who call the IRS for the purpose of requesting a payment extension or establishing a monthly payment plan to do so by using the push buttons on their telephone. The system is fully automated for eligible callers and does not require any contact with an IRS representative. The VBD system automates a process that is currently manual.

During the pilot test of the VBD system, taxpayers receiving IRS balance due notices were also notified that they might be eligible to use their touch-tone telephone to request an extension of time to pay their taxes or to set up a schedule to pay their taxes in installments. Callers to the VBD system had to enter their social security number and a unique caller identification number printed on the notice. There were checks to screen out ineligible callers based on several criteria. Once into the system, the caller was given the option to select a payment extension or a monthly payment plan. Those who selected a payment plan were asked to enter the amount they could afford to pay each month and the day of the month they wished to establish as their monthly due date. Callers who established a monthly installment plan or a payment extension received a confirmation letter from the IRS. The confirmation letter included the terms and conditions of the arrangement.

Some of the anticipated advantages to the VBD system are that it can provide callers better access to the IRS in terms of increased capacity to handle calls. Twenty-four–hour access was not available because throughout the call, callers were given the option of exiting the automated system to speak with an IRS representative. The establishment of fully automated telephone systems also provides the IRS the opportunity to redirect staff to deal with more complex issues. A goal for the VBD system is to provide the same or better service to callers with greater efficiency than the manual process.

Questions of Interest

New systems and procedures must be evaluated not only in terms of their cost savings or operating efficiencies but also in terms of their ability to provide service that is at least as easy to use as the system they are replacing. The automated installment agreement application of the VBD system replaces a manual process for the less complex installment agreements. One of the ways that we can measure ease of use is by examining the percentage of callers who succeed in establishing installment agreements through the VBD system out of those eligible callers who attempt to set up agreements. Callers can default out of the automated process at any time to an IRS employee. Therefore, the percent of callers who successfully complete an automated agreement is an indicator that the system is easy for them to use. Ease of use is affected by many factors, including the clarity of the information mailed to taxpayers informing them that they owe a balance and inviting them to use the VBD system, the clarity of the telephone script and on-line prompts, and the motivation and general capability of the caller. We might speculate that there are differences in the ability of callers to use the VBD system related to day of the week, time of day, etc. During the pilot test, there was some variability in the material mailed to taxpayers; this might also account for some differences in ease of use of the system. Beginning in June 1995, material was added to the mailing to taxpayers informing them that they might be eligible to use the VBD system to establish an installment agreement and providing some information about using the system. This additional material was discontinued in December. There is also some clustering of the mailings of notices.

To assess ease of use, we can explore the following questions: To what extent have callers been served by the VBD application; i.e., what percent of eligible callers successfully completed the application, and did that percent differ over the period of the pilot test? Were there any periodic fluctuations in completion rates?

Major goals of automated systems are to increase access and decrease costs. Cost can be measured in terms of efficiency. Automated systems are valuable primarily to the extent that they handle calls more efficiently than manual processes.

To assess system efficiency we would like to know the following: Is it more efficient for the IRS to have taxpayers set up installment agreements manually or through the use of the VBD automated system?

Note: A limited study in a single district found that, in 60 calls handled manually, the average time needed to complete an installment agreement was 15 minutes and 38 seconds. The standard error associated with this estimate of mean time was 4 minutes and 58 seconds. This is our best information about the time needed to complete a manual installment agreement of the type handled by the VBD system.

Data

Many more variables are captured by the VBD system than are presented in this case study. We selected a subset of variables that have direct bearing on the questions posed above. They include the date (Monday through Friday only), daily counts of the number of callers requesting an installment agreement, the number of callers successfully completing an installment agreement, and the day's average call length for a completed agreement.

Data collected by the automated telephone system are free from many of the measurement issues that exist in data collected manually. However, there are still issues about the refinement of the measure. Our data are presented as daily counts rather than as specific information about each call. This precludes us from tracing the success of particular types of calls as they progress through the system and limits the types of analysis that can be done.

Although the pilot test period was from April 24 through June 23, 1995, we have included additional data through December 8, 1995.

Variable Name	Description
Date	mm/dd/yy
Request	Callers Requesting an Installment Agreement Plan
Complete	Callers Successfully Completing an Installment Agreement
Length	Average Call Length, in seconds, for a Completed Payment Agreement

Date	Request	Complete	Length
042495	25	22	206
042595	15	11	172
.	.	.	.
.	.	.	.
.	.	.	.
120895	41	34	194

Analysis

Two questions were identified in Questions of Interest as the key issues to be addressed in the statistical analysis. They are listed below with suggested analytic approaches.

Question 1

To what extent have callers been served by the VBD application; i.e., what percent of eligible callers successfully completed the application, and did that percent differ over the period of the pilot test? Were there any periodic fluctuations in completion rates?

Question 1 can be answered by the following statistical techniques: (1) Convert raw data into a meaningful picture of completion rates on a daily and weekly basis. (2) Create graphs of the daily and weekly rates and examine the data over time. (See Figures 1 and 2.) (3) Calculate and examine the autocorrelates to determine whether or not there appears to be any significant fluctuation in time (e.g., from week to week or day to day) among the completion rates. Statistical methods we recommend are time series plot for both weekly and daily data, as well as autocorrelation plots. (See Figures 3 and 4.) (Useful time series references include [Box, Jenkins, and Reinsel, 1994] and [Wei, 1990].) The Pearson correlation coefficient is probably the most common measure for the dependency between two random variables. When there is only one random variable in a time sequence, we can evaluate the autocorrelation to measure the dependency among all the observations along the time sequence.

For example, given a time series $y = (y_t, y_{(t-1)}, \ldots, y_2, y_1)$, we can artificially generate a lag – 1 series $y_{-1} = (y_{(t-1)}, y_{(t-2)}, \ldots, y_2, y_1)$ and then evaluate the correlation between the first (t –1) observations from y and y_{-1}. This is called the autocorrelation of lag –1. If such an autocorrelation is significant, we know that the observations are not independent, and in fact, each observation is correlated to the previous observation. Similarly, we can define autocorrelations of lag –2, lag –3, and so on to investigate the dependency among all observations.

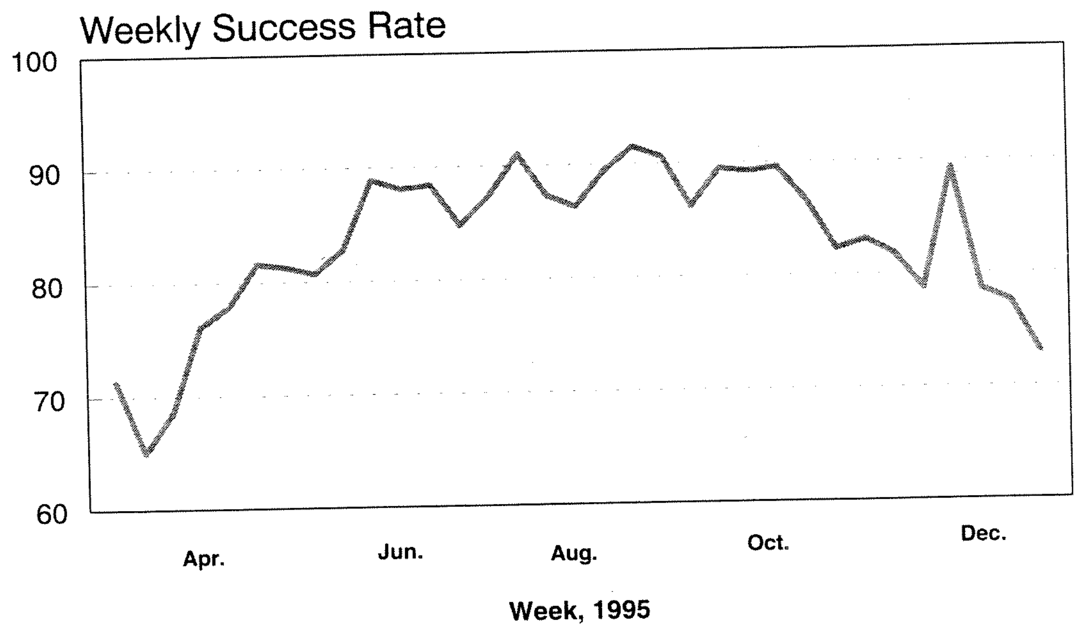

Fig. 1. *Weekly percentage of successfully completed installment agreements.*

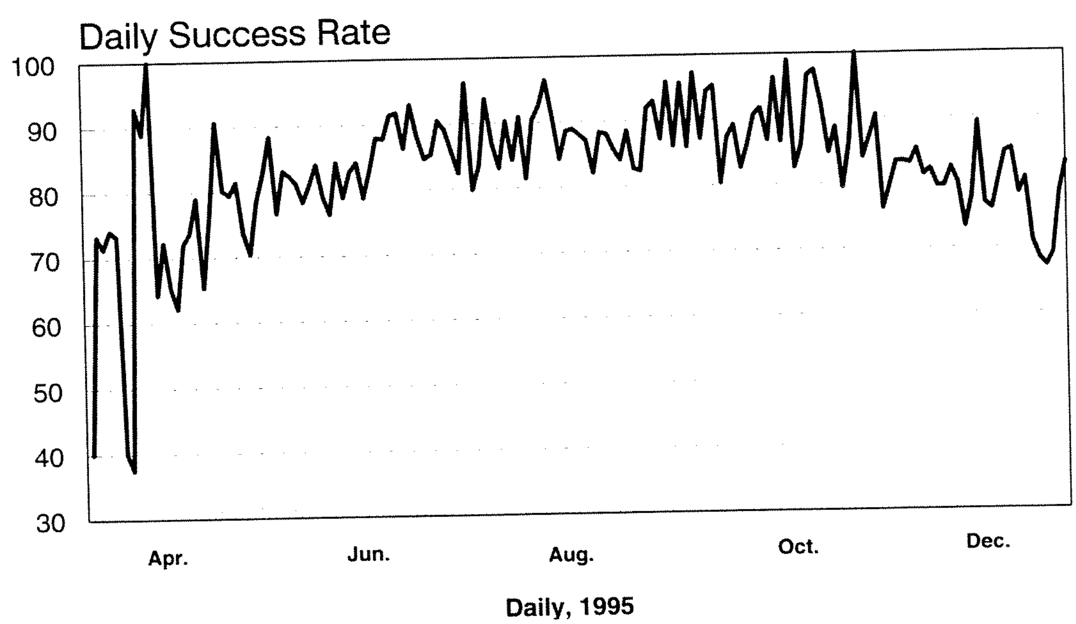

Fig. 2. *Daily percentage of successfully completed installment agreements.*

Figure 3 (autocorrelates for lags 1 through 10 for VBD data summarized on a weekly basis) shows a significant fluctuation in the success rates during the first two weeks of the pilot. This would be indicative of correcting unexpected problems that quite frequently occur in the earliest stages of a pilot test. The remaining eight weeks (lags) shown on Figure 3 reveal no significant change from one week to the next in the completion rates.

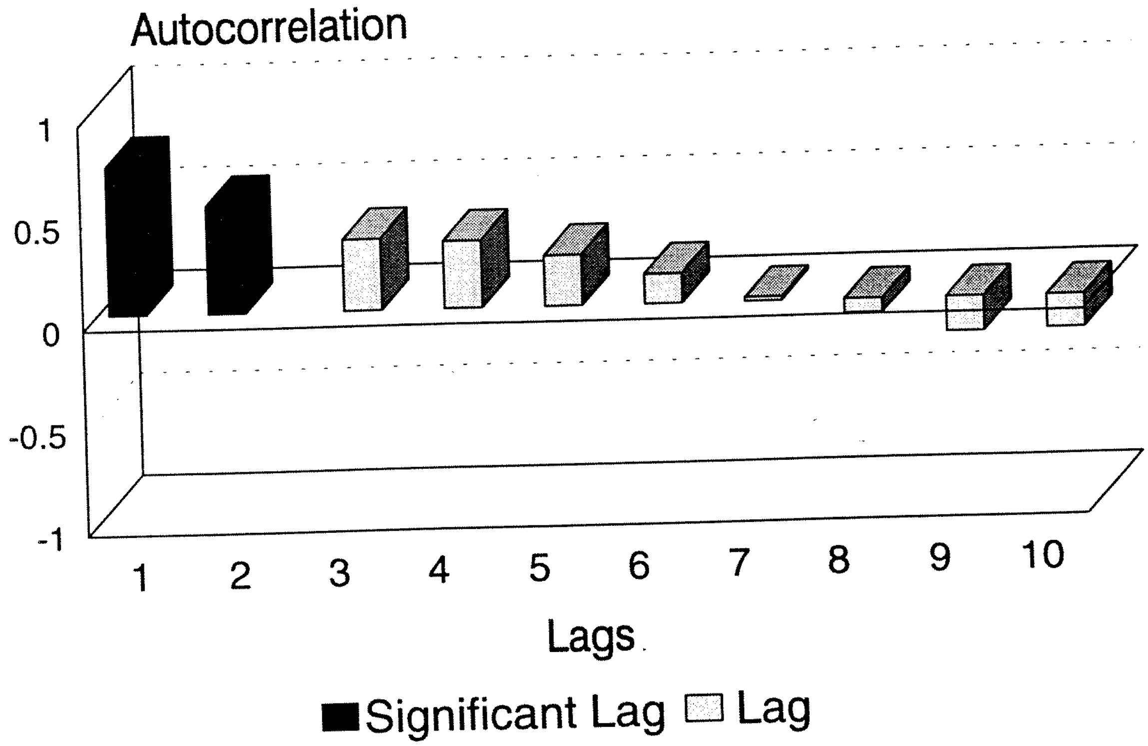

Fig. 3. *Autocorrelations for lags* 1 *through* 10 *for VBD data summarized on a weekly basis.*

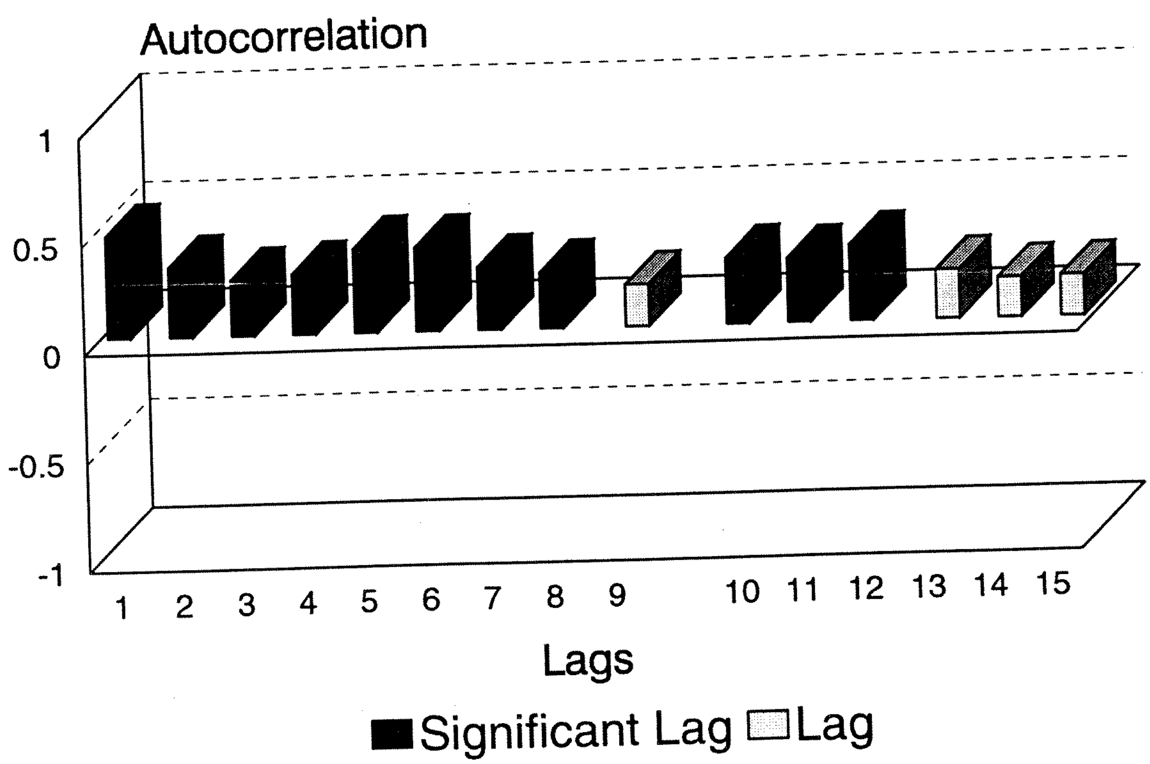

Fig. 4. *Autocorrelations for lags* 1 *through* 15 *for VBD data summarized on a daily basis.*

Question 2

Is it more efficient for the IRS to have taxpayers set up installment agreements manually or through the use of the VBD automated system?

Question 2 can be answered by the following statistical techniques: (1) Compute the average automated time to complete an automated installment agreement and determine the variability (standard error) associated with the estimate. (2) Conduct a hypothesis test where the null hypothesis is that the average length of time to complete a manual installment agreement is the same as the length of time to complete an automated installment agreement and the alternative hypothesis is that the average length of time to complete a manual installment agreement is greater than the length of time to complete an automated installment agreement. Namely, test the hypothesis Ho: $\mu_{Lm} = \mu_{Li}$ vs. Ha: $\mu_{Lm} > \mu_{Li}$, where μ_{Lm} = average time to complete a manual installment agreement and μ_{Li} = average length of time to complete an automated installment agreement. Evaluate the decision with some level of confidence. Statistical methods we recommend here are descriptive statistics, hypothesis testing, two-sample t test, and the F test to test the equality of variance assumption. We have found [Berenson & Levine, 1993], [Keller and Warrack, 1997], [Siegel, 1994], and [Vardeman, 1994] to be useful references for introductory statistical methods.

INSTRUCTIONS FOR PRESENTATION OF RESULTS

For Question 1, results should be presented with a brief written description of the findings, accompanied by graphic displays. Some statement about the apparent ease of use should be included, as well as a recommendation for any additional data collection and analysis.

The "trend patterns" can easily be seen from the time series plot for weekly data, while the "seasonal patterns" are most easily shown in the time series plot of the daily data. The trend pattern can be captured by the simple linear regression, and the seasonal pattern can be evaluated by the autocorrelation plot (as shown in Figures 3 and 4). In contrast to most introductory statistical content, the "independent" assumption is in general not valid in time series data.

For Question 2, results should be summarized in a written report, accompanied by a recommendation about the implementation of the VBD system, based on system efficiency.

Here we have used some descriptive statistics to illustrate the basic comparison. A statistical test of the hypotheses Ho: $\mu_{Lm} = \mu_{Li}$ vs. Ha: $\mu_{Lm} > \mu_{Li}$ will be sufficient to answer Question 2. We also suggest the simple comparison based on confidence intervals. This seems to be an easier approach to helping undergraduate students grasp the idea of making comparisons in the presence of uncertainty.

A report should be generated that addresses the findings. It should include graphics and a discussion of the limitations of the study and recommendations for further analysis. It should also include a recommendation about adopting the VBD system based on system efficiency and ease of use. There should be a discussion about the effectiveness of these measures for their stated purpose.

REFERENCES

Berenson, Mark L. and Levine, David M. (1993), *Statistics for Business and Economics*, Prentice–Hall: Englewood Cliffs.

Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. (1994), *Time Series Analysis Forecasting and Control, Third Edition*, Prentice–Hall: Englewood Cliffs.

Keller, Gerald and Warrack, Brian (1997), *Statistics for Management and Economics*, Duxbury Press: Belmont.

Nicholls II, William L. and Appel, Martin V. (1994), "New CASIC Technologies at the U.S. Bureau of the Census," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 757–762.

Rosen, Richard J., Clayton, Richard L., & Pivetz, Lynda L. (1994), "Converting Mail Reporters to Touchtone Data Entry," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 763–768.

Siegel, Andrew F. (1994), *Practical Business Statistics*, Irwin: Boston.

Vardeman, Stephan B. (1994), *Statistics for Engineering Problem Solving*, PWS Publishing Company: Boston.

Wei, William W.S. (1990), *Time Series Analysis*, Addison–Wesley: Redwood City.

Werking, George S., Clayton, Richard L., and Harrell Jr., Louis J. (1996), "TDE and Beyond: Feasibility Test of E-Mail/World Wide Web for Survey Data Collection," *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

NOTES TO THE INSTRUCTOR

Level of Case

The case can be used in teaching basic descriptive statistics in an introductory applied class. It can also be used to teach inferential statistics in the comparison of two populations. The graphical time series analysis can be used in an intermediate or introductory classroom depending on the level of presentation. The case is primarily appropriate for business statistics.

Necessary Tools

Understanding of statistical graphs, basic knowledge of inferential statistics and correlation.

Objectives

To use real-world data in decision making and to recognize the strengths and limitations of statistical analysis in informing business decisions. The statistical teaching goals are to use real data for hypothesis testing and to learn to use statistical graphics and limited time series analysis.

Comments and Suggestions for the Instructor

This case study can be used for several important subjects in undergraduate statistics courses:

- Descriptive statistics (mean, variance, percentage, and their physical meanings)
- Comparing two populations (confidence intervals and hypothesis testing)
- Time series plots and explanations from the plot

Specifically, the descriptive statistics can be used to estimate the means and variances, etc. of the number of completed and attempted installment agreements and mean completion times. Since the length of call data are presented as daily averages, they must be weighted by daily volume of completed agreements to estimate an overall mean length of time. This is useful in teaching the need to weight linear combinations of estimates.

Comparing two populations can be used for the choice of automated vs. manual processes in terms of efficiency. This is useful for teaching confidence intervals and hypothesis testing.

The time series plots can be used for teaching basic concepts about graphical analysis of data, trend over time, periodic patterns in the data, and attempting to understand the patterns.

The basic features of the data can be displayed by a time series plot. Note that the time series plot for weekly percentages is more appropriate to illustrate the trend pattern, as compared to the plot for daily percentages. Thus the time series plot for weekly percentages is recommended for the introductory course when the case is presented.

Moreover, as shown in the case study, we have introduced an intermediate-level technique here: the autocorrelation function. An autocorrelation function is a set of correlations between the original time series and the time series created by time shifting the original series. It has been proven to be an important tool in understanding the trend pattern but does not appear in most introductory texts. A student with exposure to the correlation function can take one further step to the autocorrelation function to understand the fluctuations in the time series.

One of the data limitations the students may identify is the use of percent completed agreements as a measure of ease of use/customer satisfaction. This is a limitation and, for the study from which the case is drawn, a customer satisfaction survey of callers completing automated agreements was part of the evaluation. The survey had a 64 percent response rate, based on the number completing the survey of those callers who successfully completed either an installment agreement or payment extension. Survey results show that 97 percent of respondents felt that the VBD system was either very easy (88%) or somewhat easy (9%) to use. Results for customer satisfaction indicate that 94 percent were either very satisfied (75%) or somewhat satisfied (18%).

Typical Results

Question 1

To what extent have callers been served by the VBD application; i.e., what percent of eligible callers successfully completed the application, and did that percent differ over the period of the pilot test? Were there any periodic fluctuations in completion rates?

1. Calculate daily and weekly completion rates.
2. Develop daily and weekly graphs using any standard statistical or graphic software package.
3. Examine daily and weekly graphs for any trends.
4. Compute autocorrelates using standard statistical software. (We used SAS Proc Arima.)
5. Develop a graph of autocorrelates using a graphic software package.

Results for Question 1

For parts (1) through (3) of Question 1, refer to Figures 1 and 2. Daily and weekly graphs show a significant increase in completion rates during the early stages of the pilot test. This was probably due to correcting unexpected problems that occurred in the early stages of the pilot test. The completion rate stabilized during the summer months of 1995. The sharp peak and valley in the completion rate during the fall of 1995 may have been attributed to system problems caused by the congressional shutdown of government. For parts (4) and (5), see Figures 3 and 4. Figure 3 indicates a significant autocorrelation for lag –1 and lag –2 weekly data, while Figure 4 suggests a five-day (weekly) seasonal pattern.

Question 2

Is it more efficient for the IRS to have taxpayers set up installment agreements manually or through the use of the VBD automated system?

Conduct a hypothesis test with null hypothesis Ho: $\mu_{Lm} = \mu_{Li}$, where
L_m = Average length of time to complete a manual installment agreement,
L_i = Average length of time to complete an automated installment agreement
against the alternative hypothesis test Ha: $\mu_{Lm} > \mu_{Li}$.
$\bar{x}_{Lm}$ = 15.64 mins, $\bar{x}_{Li}$ = 3.15 mins,

s^2_{Lm} = 24.75 mins, s^2_{Li} = 1.11 mins,
$n_1 = 44$, $n_2 = 146$.

We plan to use a t test for the equality of group means to test our hypotheses. Our first check is to determine if we have equal variances. A comparison between the calculated F and the critical F will tell us if the variances are equal.

$$F = s^2_{Lm}/s^2_{Li} = 24.75 / 1.11 = 22.23.$$

Critical $F_{(1-.05),\ 43,\ 145} = 1.58$. Therefore, since the calculated F value is greater than the critical F, we know that the manual and automated variances are not equal. In addition to the large difference in variances, the sample sizes are very different in the two groups as well. The t test that does not use pooled variances is

$$t^* = [\,\bar{x}_{Lm} - \bar{x}_{Li}\,] / [\text{sqrt}\,(\,s^2 Lm / n_1\,) + \text{sqrt}(\,s^2 Li / n_2\,)]$$

$$= (15.64 - 3.15) / [\text{sqrt}(24{,}75/44) + \text{sqrt}(1.11/146)] = 16.64.$$

(Alternatively, a nonparametric test could be used.)

The critical t value with 43 degrees of freedom at the 90th percentile is 1.30. The calculated t^* of 16.64 exceeds the critical value. Therefore, we reject the null hypothesis that mean time to complete a manual installment agreement is the same as the mean time to complete an automated installment agreement. In other words, we conclude that the IRS should continue to allow taxpayers the opportunity to create an installment agreement in an automated setting.

An alternative method of analysis would be to compare the confidence intervals around the two estimates of mean time to complete an installment agreement. This has intuitive appeal to undergraduate students and is frequently easier for them to grasp than the logic of hypothesis testing.

The 90 percent confidence interval around the estimate of mean time to complete a manual installment agreement is

$$\bar{x}_{Lm} \pm s_{Lm}\, t_{.95,\,42},$$

$$15.64 \text{ "} (4.97)(1.68),$$

$$8.35 \text{ minutes} \le \mu_{Lm} \le 23.99 \text{ minutes}.$$

The corresponding confidence interval around the estimate of mean time to complete an automated agreement, using $t_{.95,\ 145}$, is 1.42 minutes $\le \mu_{Li} \le$ 4.88 minutes.

The two confidence intervals do not overlap. From this we can also conclude that the average time to complete a manual installment agreement is not equal to the average time needed to complete an automated agreement.

BIOGRAPHIES

Mary Batcher manages a statistical consulting group at the Internal Revenue Service (IRS). The group consults on a number of internal projects such as testing new systems, designing surveys, and designing samples. Before coming to the IRS, she was at the National Center for Education Statistics.

Kevin Cecco is a consulting statistician at the IRS, working on the development of new automated systems for handling some telephone calls and for the submission of simple tax returns. He came to the IRS from the Bureau of the Census.

Dennis Lin is an associate professor at Penn State University, where he holds a joint appointment in the Departments of Statistics and Management Science. Prior to his appointment at Penn State, he was an associate professor at the University of Tennessee.