2002

An analysis of presentation rates to a paediatric emergency department

Andrew Hiskins
Edith Cowan University

An Analysis of Presentation Rates to a Paediatric Emergency Department

BY

ANDREW HISKINS

A thesis submitted in partial fulfilment of the requirements for the degree of Bachelor of Science (Mathematics) Honours, to the Faculty of Communications, Health and Science, Edith Cowan University, Perth Western Australia.

October 2002

USE OF THESIS

The Use of Thesis statement is not included in this version of the thesis.

ABSTRACT

The health care industry is a very expensive one, constituting a significant proportion of the government budgets. Princess Margaret Hospital (PMH) for children is the only tertiary paediatric centre in Western Australia. PMH has over 40,000 children aged 0 – 16 years of age who present to the emergency department each year. PMH is one of many hospitals funded from government sources. The emergency department is a high cost area and an area with limited ability to curtail services due to financial constraints. A busy hospital will over a period of time have a constantly changing number of people presenting to the emergency department for treatment. This may depend upon the time of year, the day of the week, the weather of the day or the presence of a holiday period. An ability to accurately predict the daily, weekly patient flow, together with seasonal fluctuations, could enable more informed decisions to be made regarding support services needed. This would in turn result in cost savings and improved medical care. The study used Time Series Analysis techniques, Regression Analysis and other statistical techniques to model the presentation rates to the PMH emergency department. The analysis also examined the factors that have the greater influences upon these rates. The resulting prediction models can be used for quantitative short-term forecasts of future ED volumes. An overall comparison was made between the differing techniques used in the analysis. Two cycles were clearly identified within the ED data series used. The dynamic methods proved to be the better at modelling and providing forecast values for the series analysed. It is hoped these types of models may provide formal criteria to assist staff in earlier recognition of extreme periods.

DECLARATION

I certify that this thesis does not, to the best of my knowledge and belief:

(i) incorporate without acknowledgment any material previously submitted for a degree or diploma in any institution of higher education;

(ii) contain any material previously published or written by another person except where due reference is made in the text; or

(iii) contain any defamatory material.

Signature.........................

Date....29/12/02........

ACKNOWLEDGMENTS

I would like to thank Princess Margaret Hospital for Children, Subiaco, for making the raw data (daily presentation number by triage code) for the emergency department available. In particular, I would like to thank Dr. Gary Geelhoed, Director of Emergency Medicine for giving permission for the data to be gathered.

Of course, special thanks must be directed towards Associate Professor James Cross for his initial encouragement, tireless and ongoing support.

Most importantly, I would like to thank my wife, Diana for her continued support over the past year. Without this, the project would not have been completed.

List of Tables

Page

Table 5.1 Holt-Winters parameter values 38

Table 5.2 Forecasted values using 1999 model 41

Table 5.3 Modified Box-Piece (Ljung-Box) Chi-Squared statistic 42

Table 5.4 Mean Squared Error (MSE) values for Holt-Winters & Box-Jenkins models 47

Table 5.5 Forecasts for December 1999 using regression model with day variables 48

Table 5.6 Forecasted values using 1999 model 54

Table 5.7 ASTSA statistics and Box-Piece statistics: 1999 total presentation residuals series 55

Table 7.1 Mean Squared Error value for each model developed. Best in blue, worst in red 66

Table 7.2 R-squared adjusted value for each regression model developed... 66

List of Figures

		Page
Figure 1.1	Monthly presentation for 2000 ($\log_{10}$)	4
Figure 1.2	Presentation by day of the week for 1999	5
Figure 2.1	Application of single smoothing to June – December 1998 Triage series $\alpha = 0.9$	11
Figure 2.2	1998 Triage series	12
Figure 2.3	Smoothing parameter values post modelling	14
Figure 2.4	1998 Triage series versus Holt-Winters model	14
Figure 3.1	ACF plot of presentations January – December 1998	17
Figure 3.2	PACF for time series May – October 1998	18
Figure 3.3	Plot of 1998 time series and plot of first difference of the series	19
Figure 3.4	Polynomial trend line for the 2000 series	20
Figure 3.5	SARIMA model (yellow) over raw data from 2000 triage Series	23
Figure 3.6	Spectral density plots. (a) Triage data series the year 1998. (b) detrended series using first differencing method	25
Figure 3.7	Spectral density plot, post trend and seasonal component Removal	26
Figure 3.8	ASTSA statistics for 1998/1999 Box-Jenkins Model..............	26
Figure 4.1	Correlation matrix for 1998 triage data series	28
Figure 4.2	Boxplot graph of 1998 triage data by day of presentation	28
Figure 4.3	Scatterplot of 1998 triage data, day of week, type of day Coding.	30
Figure 5.1	(a) ED daily presentations by triage category 1999. (b) Total daily presentation versus Category 4	34

Figure 5.2	ED presentations January 1998 – December 2000 with 6th polynomial trend line	35
Figure 5.3	ED daily presentations March – June 1999	36
Figure 5.4	Spectral density plots: (a) 3 year data series. (b) 1999 data series	37
Figure 5.5	Holt-Winters models: (a) & (b) 3 year series. (c) & (d) 1999 data series	39
Figure 5.6	Residual Analysis: 1999 Total presentations model	40
Figure 5.7	Autocorrelation function and partial autocorrelation function; 1999 calendar series	42
Figure 5.8	ACF and PACF; May '99 to April '00 series	43
Figure 5.9	Box-Jenkins Model: (a) 3 year data series. (b) 1999 data series	45
Figure 5.10	Model comparisons: May – December 1999 series (a) Total presentations, (b) Triage category 4.	46
Figure 5.11	Comparison of 95% confidence intervals for 30 forecasts using 1999 models	47
Figure 5.12	1999 ED total daily presentations by day of week	48
Figure 5.13	1999 Regression model using weekday variables	49
Figure 5.14	1999 regression model using weekday variables + Time^2 variable	49
Figure 5.15	1999 monthly presentations by time group (Log_{10})	50
Figure 5.16	Best subsets output for 1999 ED series	51
Figure 5.17	(a) 1999 Regression model; (b) 3 year regression model; each Using weekday, month and weather variables	52
Figure 5.18	Residual Analysis: 1999 Total presentation model	53
Figure 5.19	Spectral analysis; residual series from 1999 regression model	54
Figure 5.20	Autocorrelation function and Partial autocorrelation function; 1999 Regression model residual series	55
Figure 5.21	1999 Regression residual model	56

Figure 5.22	3 year Regression residual model	56
Figure 6.1	Residual plots for each of the 4 models developed; 1999 & 3 year Total presentation	59
Figure 6.2	ACF and PACF correlograms of Box-Jenkins model residuals 1999 & 3 year total series	60
Figure 6.3	Histogram of residuals, 1999 & 3 year total series	60
Figure 7.1	1999 time series of triage category 4 as percentage of total presentations	63
Figure 7.2	1999 May to September complete series, compared to all four model fits	64

TABLE OF CONTENTS

Abstract ii

Declaration iii

Acknowledgments iv

List of Tables v

List of Figures vi

Chapter 1 – Introduction 1

1.1 About this Chapter 1
1.2 Background to Study 1
1.3 Importance of Research 5
1.4 Thesis Aims 6
1.5 Data Sets 7
1.6 Computer Software 8

Chapter 2 – Exponential Smoothing 10

2.1 About this Chapter 10
2.2 Single Exponential Smoothing 10
2.3 Holt-Winters Method 11
2.3.1 Initialisation 13

Chapter 3 – Stochastic Models 15

3.1 About this Chapter 15
3.2 Types of Stochastic Models 15
3.3 Stationarity 15
3.4 Autocorrelation 16
3.5 Partial Autocorrelation 17
3.6 Non-Stationarity 18
3.7 Notation 20

3.8 Autoregressive models 21

3.9 Moving Average Models 21

3.10 Autoregressive Moving Average models 22

3.11 Box-Jenkins Autoregressive Moving Average Models ... 22

3.12 Spectral Properties of Stationary Models 24

3.13 Akaike's Information Criterion 26

Chapter 4 – Regression Analysis **27**

4.1 About this Chapter 27

4.2 Correlation 27

4.3 Simple Regression 29

4.4 Multiple Regression 29

4.5 Time-Related Regression 31

4.6 Regression with ARIMA 31

4.6.1 Method of applying 32

4.7 Best Subsets Regression 33

4.8 Cross Correlation Function 33

Chapter 5 – Data Analysis **34**

5.1 About this Chapter 34

5.2 The time plot 35

5.3 Spectral Analysis 36

5.4 Holt-Winters method 37

5.5 Box-Jenkins 41

5.6 Holt-Winters v Box-Jenkins 46

5.7 Regression Modelling 47

Chapter 6 – Residual Analysis **57**

6.1 About this Chapter 57

6.2 The Method Used 57

6.3 The Residual Findings 58

Chapter 7 – Discussion and Conclusion **61**

7.1 About this Chapter 61

7.2 Summary of Analysis 61

7.3 Comparison of Models 64

7.4 Future Research Directions 67

7.5 Conclusions 68

References: **69**

Appendix A Data values used in thesis 71

Appendix B Model parameter values 76

Appendix C Holt-Winters Models 78

Appendix D Box-Jenkins Models 87

Appendix E Regression Models 97

Appendix F Approval letter for data collection 120

1 INTRODUCTION

1.1 ABOUT THIS CHAPTER

This section of the thesis provides a general background to the study, the importance of continued research within the health sector and the aims of the study. The data and its source are also discussed.

1.2 BACKGROUND TO STUDY

Princess Margaret Hospital for Children (PMH) is the only tertiary paediatric medical centre within Western Australia. It is the reference centre for paediatric illness and injury for the state. Although the catchment zone may potentially be the entire state, it will not see all children requiring hospital treatment in any one year. Many are treated at regional hospitals. On average, approximately 43,000 children will present to PMH seeking medical assistance from the hospital's emergency department each year. It is this average number that the WA Health Department uses for its budgetary forecasts.

The health care industry is a very expensive one. The new techniques and medications continue to increase in cost. It could be argued that many countries around the world are unable to improve their standards of medical care, due to these prohibitive costs. PMH is one of many hospitals funded directly from government sources, and there will always be a budgeted limit to the funding available. Once allocated, responsibility for dividing funds between departments falls to individual hospital management boards. This task may be easier if the numbers of patients seen within the departments could be predicted. The emergency department (ED) is a high cost area and one department without the ability to curtail services in responce to financial constraints.

Employees within an ED will be aware that the number presenting on each day will vary. It is common for one or two days to be noted as the "busiest of the week" on a regular basis. For a busy metropolitan hospital, Saturday or Sunday may often be busiest, where a suburban or country hospital may see Monday providing the peak in

presentations. As a member of an ED I noted a similar pattern. The presence of a weekly cycle in presentation numbers seemed likely. But are there other cycles in the daily numbers? It is common for more patients to arrive in August than in January. But does this mean that the numbers each month will be similar to the previous year's?

These findings led me to consider what may influence the number of patients who will present to the department each day. In discussions with my co-workers, influences included the day of the week, the presence of holidays and weather variables. Christoffel (1984), and Attia & Edward (1996) both concluded that a common belief among emergency workers is that weather conditions will significantly influence these daily numbers and increase the triage acuity. Both studies were undertaken to ascertain if this belief was a reality or not. Christoffel (1984) used data from 3 diversely different months of the year, then categorised each day into one of 7 measures of extreme weather. These ranged from extreme heat and cold, to atypical heat and cold, to the presence of rain, storms or snow. The statistical analysis, two-tailed, that was undertaken compared the proportion of days in each category of weather with the proportion of presentations on days with those categories. Only the severe extreme weather was shown to be statistically significant in its impact upon presentation numbers. The study by Attia and Edward (1996) over a decade later re-examined the impact of weather. The influence of the milder changes was examined, accepting that severe weather did impact on department numbers. The data length was expanded to 1 year to avoid seasonal impacts and grouped into 3 hour'y intervals. This time division enabled the clustering of presentations before and after weather events, such as heavy rainfall, to be included. Analysis was undertaken using a chi-squared (χ^2) or Fischer's exact tests, with Yate's* correction employed where appropriate. The conclusion they draw from this study, was that no influence upon presentation rates could be attributed to less severe changes in prevailing weather conditions.

Diehl et al (1981) noted that based upon empirical data, calendar factors (day of week or month of year) and meteorological conditions strongly influenced the number of

* Yates was a researcher at Romstead who developed a tabular technique for deriving the factorial effects when performing analysis of variance tests. The technique attempts to provide a better estimate of significance levels.

unscheduled visits. Despite this, they noted weather conditions added little predictive value. By matching daily presentation numbers and weather data over a 4-year period, they obtained a predictive equation through use of stepwise linear regression. The study also examined the impact of a day's status, as a public holiday or the day after the public holiday. However, they were unable to determine if the high use post public holiday was due to the desire not to disturb holiday events and gatherings or that holiday excesses were the cause of illness and injury.

A study by Saez et al (1995) into the effects of temperature change upon death rates concluded that three or more days of above average temperature increased patient morbidity. This study, while based upon several medical conditions not seen in children, did show that transfer function models (ARIMA) could be applied to health data series. They concluded that most predictive models had an autoregressive component of 1, indicating that the current presentation numbers were dependant upon the preceding value. Severe weather events do impact upon presentation numbers.

Several studies showed a 5% to 20% reduction often occurs during a severe weather event. Attia (1996), while analysing the impact of a snowstorm, noted a 78% reduction during the storm. She compared the 36 hours of a storm with the 36 hours before by use of a χ^2 test. However, as she points out, this significant impact is more likely due to perceived transportation problems associated with the impending storm than the storm itself. Each of these studies relates to the Northern Hemisphere. While weather conditions differ in Australia, it cannot be assumed they would not impact upon presentation rates, as their impact has not been fully analysed.

There is a belief by some emergency workers that a full moon increases presentations. A study by Reno (1996) looked at this topic by examining if the lunar phases directly affected human behaviour and therefore presentation rates to hospital emergency departments. He noted a past study (Lieber & Shenn, 1972) that stated as humans had the same elements in roughly the same proportion as the earth (80% water to 20% organic and non-organic minerals) that the same gravitational effects upon the earth would occur to humans. Reno examined presentation data on tidal and non-tidal days and found no significant increase in presentations associated with the presence or not of

a full moon. Interestingly, a questionnaire he distributed to emergency department room staff about their belief in this theory showed an illusory correlation between lunar phases and emergency presentations. The saying "it must be a full moon" is commonly heard within departments on a busier than expected day.

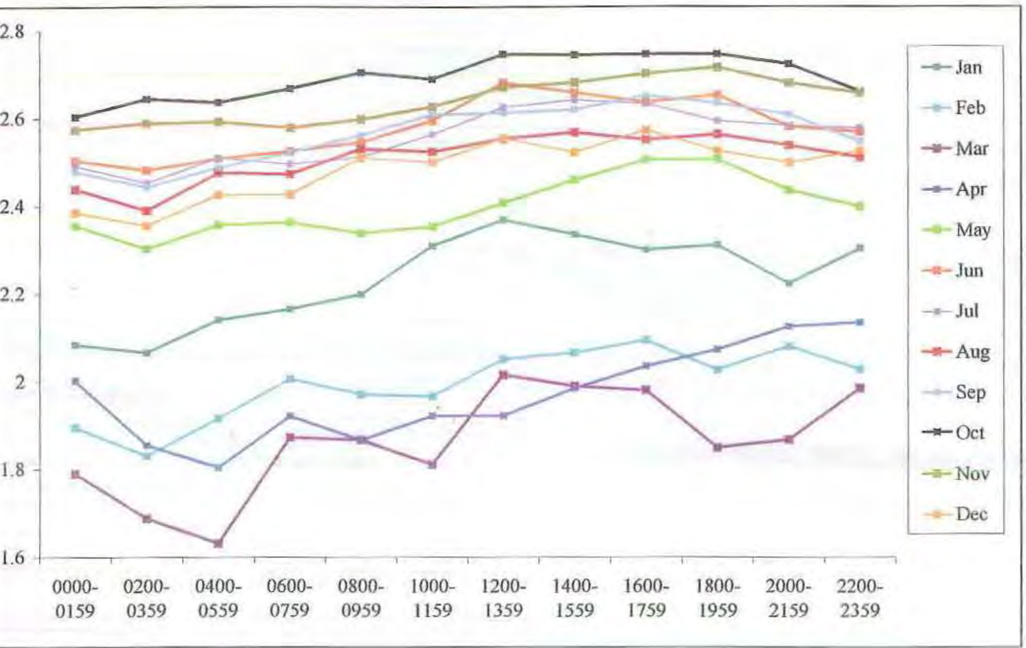


Figure 1.1: PMH Monthly presentations for 2000 ($\log_{10}$)

Environmental factors – pollen, pollution, dust, etc – have been shown to directly influence presentation rates in many studies. Emergency departments often see high admission rates in springtime due to the influence of these factors, as displayed in Figure 1.1. Goldsmith et al (1983) linked these factors with temperature, concluding a higher temperature increases a factors' impact. They undertook a basic study of correlation and regression between the factors, basing the choice of models upon plausibility of pair-wise dependence, strength of correlations and examination of the partial correlation matrix. This study will not examine these environmental factors, as their significant influence is well known, particularly in the pre summer months.

The presentation rates to PMH emergency regularly peak near 7pm each evening. Figure 1.2 shows the average daily variation in presentation numbers for the year 1999. A difference in the presentation pattern is clear between weekdays and weekend. It is conjectured, from anecdotal evidence from staff and observation, that family

circumstances have an influence upon the presentation time. These factors may include availability of transportation and babysitters, and work commitments of either parent. However no studies were identified that investigated these apparent factors. The perceived unavailability of general practitioners is often quoted as a possible factor associated with increased weekend presentations.

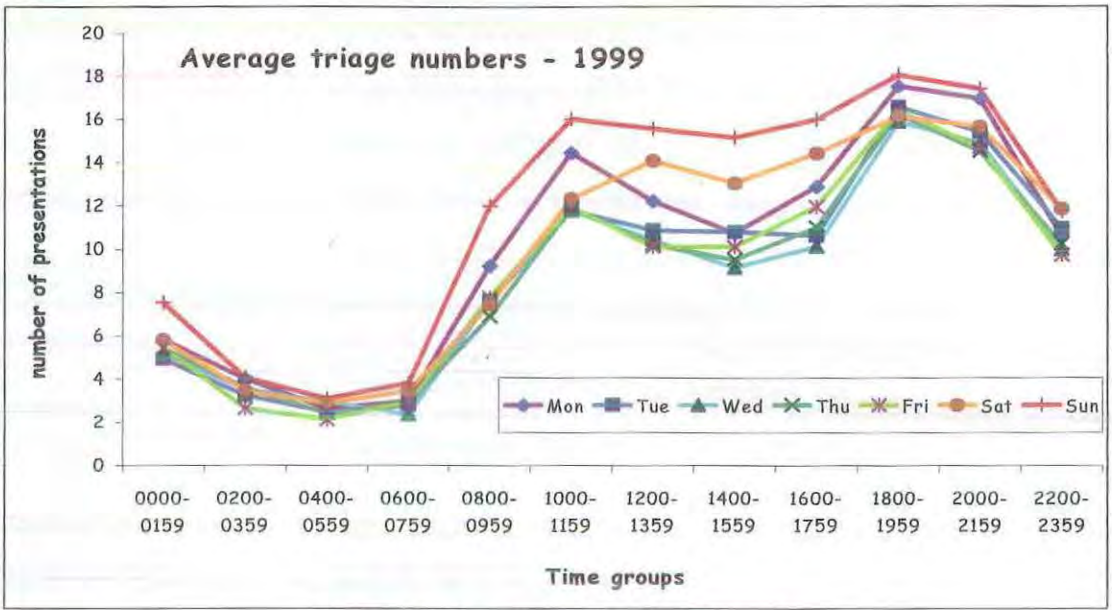


Figure 1.2: Presentations by day of the week

1.3 IMPORTANCE OF RESEARCH

The emergency departments of hospitals are high cost sectors. It is not uncommon for an ED to pass its allocated budget months before the close of a financial year. This results in extra funds having to be found. Reduction in services and the non-replacement of equipment can also be implemented. The closure of an ED is not viewed as desirable due to the community impact. However, the temporary closures witnessed at Swan Districts Hospital in winter 2001 indicate this can happen.

A basic understanding of the daily numbers, together with the ability to predict coming levels may enable these problems to be avoided. At the least, it should allow prior warning of an upcoming shortage. Diehl et al (1981) noted that an ability to accurately predict the weekly patient flow, together with any seasonal fluctuations, could enable more informed decisions to be made. This would in turn result in appropriate cost savings and improved patient care. They further noted "the more efficient use of

salaried services allows an overall reduction in staff, yielding financial benefits without the sacrifice of quality medical care". Although their analysis was completed 20 years ago its relevance is apt given recent upheavals within this state's health sector.

Efficiency and effectiveness of modern business depend on the skill level of its staff together with the level of computer power available. A hospital is no different. Databases are the current method of choice for all hospital patient details. Tandberg & Qualls (1993) indicated that the increasing availability of these databases as well as ever improving computer hardware and software has brought sophisticated quantitative analysis into the realms of practicality. It is with this thought in mind that the use of Time Series Analysis techniques would be appropriate in this setting. Their analysis used such a database of over 42,000 patients presenting over two sequential years. All patients seen had time of arrival, discharge and level of acuity recorded. Five differing model types were then attained, ranging from raw data and simple moving average through to ARIMA, to determine if forecasting was possible. The resulting models provided at best a 42% explanation of presentation variation over the 2-year data range. Tandberg & Qualls concluded, that this type of analysis can provide powerful, quantitative short-term forecasts of future ED volumes. An appropriate model could be used to develop formal criteria for calling additional staff to overcome sudden increases in patient numbers. These types of decisions could then be made consistently, and more importantly, the criteria would allow earlier recognition of extreme periods. The emergency departments of a hospital are high-intensity users of ancillary resources, such as clinical laboratories, radiology and house keeping. The ability to accurately provide short-term forecasts of ED patient volumes would also prove beneficial to these services within a hospital.

1.4 THESIS AIMS

The aims of this thesis are as follows:

a) To investigate the possibility of a cyclic pattern occurring within daily presentation rates to PMH
b) To investigate factors that may influence the pattern of presentation

c) Apply 4 differing mathematical analysis tools to this data over a range of time intervals
d) To develop appropriate mathematical prediction models for this data through use of several computer packages
e) To compare each of these methods for suitability

1.5 DATA SET

Princess Margaret Hospital uses a computer-based database, Emergency Department Information System (EDIS), to collate and record all patient details of children presenting the hospital's ED. This system has been in operation since January 1998 and is subject to quality assurance by the injury surveillance officer on a daily basis, thus accuracy and integrity of the data will not be a concern. The EDIS database is accessed from one of 5 remote computer terminals within the department, the database itself being networked. All patients presenting to the department are first seen by the triage nurse. Their basic demographic details and clinical information are recorded, then each child is given a triage code, an indication of the level of "emergency", based upon their reason for presentation to the department. The codes range from 1 (resuscitation) to 5 (non-urgent) and are treated on the basis of their triage code. I have included those who chose not to wait for their treatment, as they do impact upon the treatment of others and upon staff levels required. The times when the patient is seen by medical staff, their final diagnosis, tests and time of discharge from the department are also recorded within the EDIS system. This latter information I shall not use within the scope of this project.

The database enabled me to obtain data for each child who presented to the emergency department for the four-year period 1998 – 2001 inclusive - a sufficient volume of data to look for any underlying seasonal effects.

The following data fields were obtained from the system for each patient who presented to the emergency department at PMH during the 4-year period.

- Triage date
- Triage time
- Triage code

The data was then imported into a Microsoft Access database for initial storage and analysis. Data fields for day of the week, Monday 1 to Sunday 7, and "type of day", normal (1), public holiday (2), school holiday (3) or both (4), were then added for the purpose of this analysis.

The National Climate Centre located in Canberra kindly supplied the following meteorological data for the same time period; daily minimum, maximum and rainfall recordings for Perth. This data will be used during this analysis, but with some reservation as outlined by previous studies. The recordings are a generalisation of the weather experienced in the Perth metropolitan area for each day. These recordings are for the Meteorological Bureau's recording station located in the Perth suburb of Mount Lawley. We are all aware that rainfall in one area may not occur in all areas of a city. More importantly, the weather experienced by the patients presenting to PMH on any given day will depend entirely on where they reside at the time of illness or injury. Any subsequently discovered weather impact must be viewed with this knowledge.

Finally, missing data is not a concern for this study. Both the EDIS and the National Climate data are complete with no missing data entries.

1.6 Computer Software

Commercial Application

The Emergency Department Information System (EDIS) version 8.54.006. This is a database provided by HAS Solutions (Hospital Administrative Software Solutions) that contains all details pertaining to a patient's presentation to Princess Margaret Hospital for Children's emergency department. Approval to use the EDIS data was obtained prior to data acquisition. (Refer to appendix for approval letter).

Other Applications

Microsoft Access: for storage of EDIS data and initial data analysis

Microsoft Excel:

Minitab 12 for majority of analysis

ASTSA time series package

Microsoft Word: compilation of thesis.

SPSS version 9.0 for Windows, student version: Statistical compilations

All these software packages were used within a Microsoft Windows 98 environment on a Pentium III microcomputer.

2 EXPONENTIAL SMOOTHING

2.1 ABOUT THIS CHAPTER

Time series may be analysed using relatively simple forecasting and smoothing methods that model the components within a series that are often identifiable from a time series plot. The data is decomposed into these components and then estimates the components into the future for the purpose of forecasting. Static (trend analysis and decomposition) and dynamic methods (moving average, single and double exponential smoothing) may be employed. These methods have a major advantage in that they can produce quite reliable short-term forecast relatively quickly for a large collection of time series. This chapter examines dynamic methods, in particular the Holt-Winters method. A detailed examination of this method, its purpose, strengths and weaknesses shall be presented, together with an analysis of a triage data series using the method.

Exponential smoothing is an extension of the moving average transformation method in that it forecasts values for a time series by use of weighted averages. The moving average method implies an equal weight to the k past values, that is a weight of $1/k$. However, the exponential method makes the assumption that the more recent data values have a greater influence upon future values. This implies that the weights applied to each previous data value must decrease with age, in other words these weights exponentially decrease with time.

2.2 SINGLE EXPONENTIAL SMOOTHING

This basic smoothing method provides a forecast value by taking the previous periods forecast and adjusting it by the previous forecast error. This may be defined by $F_{t+1} = F_t + \alpha(Y_t - F_t)$ with α a constant between 0 and 1. This implies that the next forecast is adjusted by the negative error of the past value. Due to this, the forecasts will tend to trail any trend or seasonal pattern present. The initial value of α is taken to be Y_t for convenience. All exponential smoothing methods must have initialisation.

When α approaches 1 the initial value rapidly becomes irrelevant. Conversely, its influence is significant has α nears 0. However the small values tend to provide a smoother series, approximating a smoothed average.

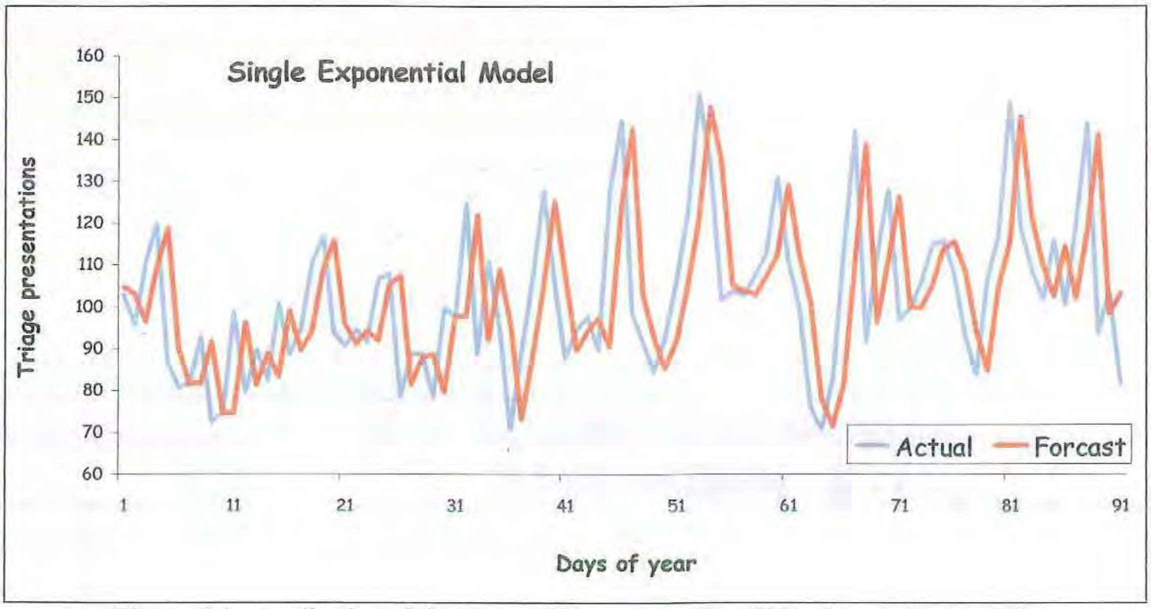


Figure 2.1: Application of single smoothing to Jun – Dec 1998 triage series, $\alpha = 0.9$

Exponential smoothing is best suited to data that does not have underlying trend, seasonality or other patterns identifiable within the time series. The forecast horizon is always assumed for one period only. Beyond this and a "flat" series will ensue.

2.3 HOLT-WINTERS METHOD

This model generalises the single exponential method to enable analysis of a data series containing trend and seasonality. Winters extended the algorithm developed by Holt in 1960 to enable the analysis of seasonal time series. This involves the inclusion of an extra equation used to adjust forecasted values to reflect a seasonal pattern (Chatfield, 1996). The method is based upon three smoothing parameters; trend, level and seasonality. Through the use of each parameter, the Holt-Winters method seeks to explain the variations occurring within a given time series. The basic equations used depend upon whether the cyclical components of the series being analysed are additive or multiplicative. A series that exhibits a trend and the size of the seasonal effect

appears to increase in relation to the mean, then the series is considered multiplicative. A seasonal effect is said to be additive when this variation is constant. An examination of the time series plot enables identification of the appropriate seasonal effect. Figure 2.2 indicates that the triage data series has a multiplicative seasonal effect. The multiplicative and additive models are often both trialled and compared to determine the better model.

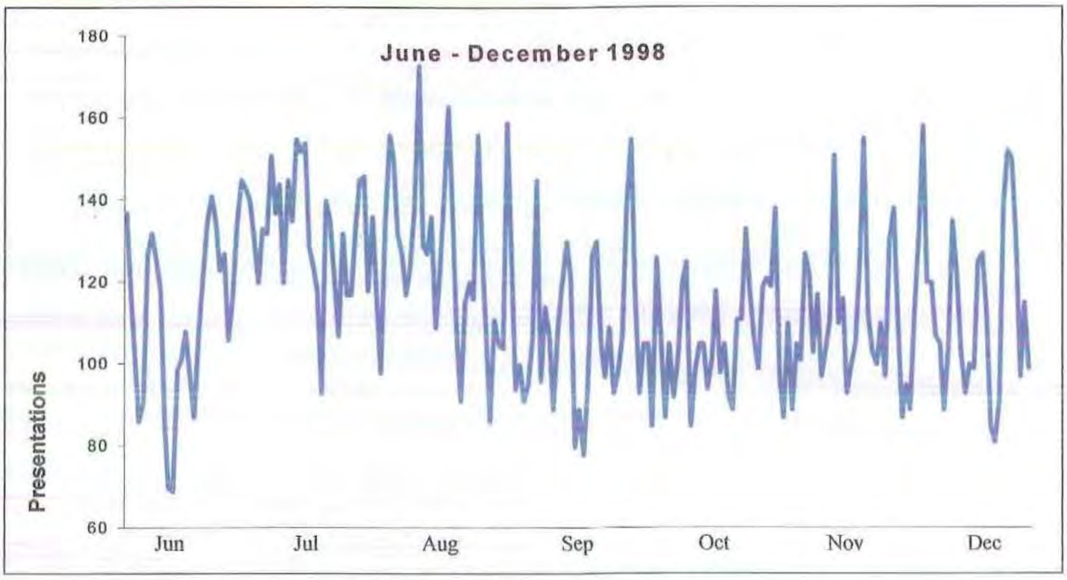


Figure 2.2: 1998 Triage series

Let the estimation of the level, trend and seasonal components be denoted by L_t, b_t and S_t, the length of the season is s, the past data values by Y_t and the smoothing parameters by α, β and γ. These parameters are usually given a value between 0 and 1. The model's algorithm makes use of new observations to update the previous estimates by use of weighted averages in a recurrence form. The three basic equations for the multiplicative model are:

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + b_{t-1})$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$$

$$S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma)S_{t-s}$$

The forecast values are realized by the equation $F_{t=m} = (L_t + b_{tm})S_{t+m}$ for m periods ahead. This method realizes L_t as a smoothed average value of the series that does not include seasonality. That is, Y_t/S_{t-s} provides the deseasoned data values.

2.3.1 Initialisation

All exponential smoothing methods require initialisation of the smoothing indices. Initialisation of the smoothing parameters may be done arbitrarily or through use of subjective assessment methods. The seasonal index s, requires a complete seasons data. The trend and level are initialised at period s. The trend makes use of 2 complete seasons of data values. A moving average of order s is used to initialise the level.

$$L_s = \frac{1}{s}(Y_1 + Y_2 + ... + Y_s)$$

The initial trend parameter is determined by the following equation. The initial value for b_s is an average of s such terms.

$$b_s = \frac{1}{s}\left[\frac{Y_{s+1} - Y_1}{s} + \frac{Y_{s+2} - Y_2}{s} + ... + \frac{Y_{s+s} - Y_s}{s}\right]$$

The initialisation of the seasonal index is done by use of a ratio of the first few data values to the first years mean such that

$$S_1 = \frac{Y_1}{L_s}, S_2 = \frac{Y_2}{L_s}, \ldots\ S_s = \frac{Y_s}{L_s}.$$

The final values chosen for the smoothing parameters α, β and γ may be arbitrary, with default values of 0.2 not being uncommon. An alternative approach involves the use of non-linear optimisation algorithms. The increasing speed of computers makes the use of these algorithms more feasible. The Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) are two such methods that provide such parameter estimates.

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2 \qquad\qquad \text{MAPE} = \frac{1}{n}\sum_{t=1}^{n} |PE_t|$$

The use of large weights for smoothing parameters results in a more rapid change in forecasts, conversely smaller weights ensure a less rapid change. If change in the level

dominates short-term variations, then a small α value would be appropriate. Conversely small α values are appropriate for prominent short-term variations over level changes.

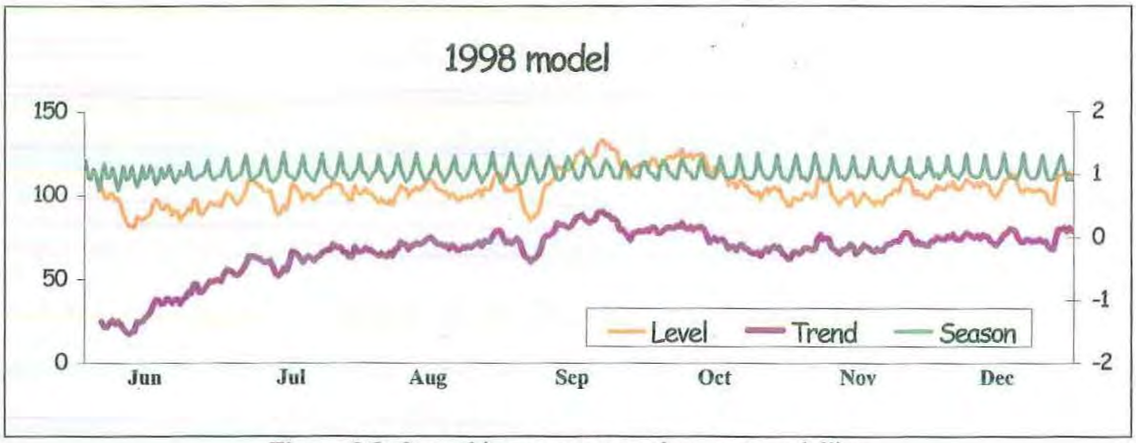


Figure 2.3: Smoothing parameter values post modelling

Figure 2.2 displays the daily triage data for the last 7 months of 1998. There appears a clear cyclical pattern over a short time frame. A longer cyclical pattern may exist over a complete year. In Figure 2.3 the level, trend and seasonal patterns are shown resulting from the application of the Holt Winter's method to this time series. The cyclical nature of this series is clearly displayed, however the trend for the series would be taken to approach linear. This becomes more apparent after the initial values. Figure 2.4 displays the 1998 series together with the Holt-winters forecasts. Clearly, the forecasts are unable to replicate accurately when consecutive data values differ by a significantly large amount.

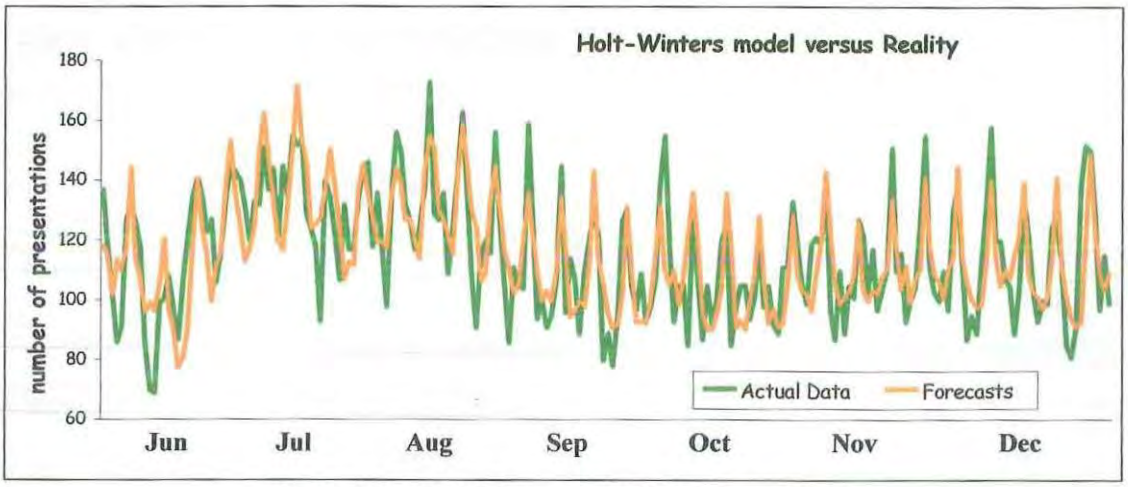


Figure 2.4: 1998 Triage series v Holt Winters Model

3 STOCHASTIC MODELS

3.1 ABOUT THIS CHAPTER

Time series analysis aims to develop mathematical models to describe sample data in a plausible manner. This chapter examines the Box–Jenkins Autoregressive Integrated Moving Average (ARIMA) models. Although this technique was originally developed in the 1930's, Box and Jenkins published a detailed description of it in their 1970 book that led to it becoming known as the Box-Jenkins methodology. The concepts of stationary and non-stationary time series are defined. The methods of attaining stationarity of a time series are discussed.

3.2 TYPES OF STOCHASTIC MODELS

A time series is a collection of observations made sequentially in time (Chatfield, 1996) that depicts a stochastic process. This is a process portrayed by a stochastic or probability model. A principal of time series analysis is that future values may be predicted based upon past observations. If this prediction can be exact, it is termed deterministic. However most such time series are stochastic, with the future partly determined by past values. Exact predictions are impossible, thus the future values are represented by a probability distribution conditioned by the knowledge of the past values. There are two important classes of stochastic models, stationary and non-stationary (Box and Jenkins, 1970).

3.3 STATIONARITY

A stochastic model may have its usefulness affected by the assumption of stationarity. A time series is considered to be stationary if the underlying generating process is based upon a constant mean and variance. The statistical properties of a time series are independent of the time period of observation (Makridakis et al, 1998).

Strict stationarity occurs when the probabilistic behaviour of $X(t_1),...,X(t_k)$ is the same as $X(t_{1+h}),...,X(t_{k+h})$ for all t_1, ..., t_k, for any k = 1, 2, ..., or shift h = 0, ± 1,... . Shifting of the time origin does not effect the joint distributions, only the interval between t_1, t_2, ... t_n. This definition in practice is found to be too strong and a more lenient one is required. A weakly stationary time series imposes conditions upon only the first two moments of the series and has a constant mean and variance.

3.4 AUTOCORRELATION

This refers to the correlation between values within the same time series at different time periods or lags. The autocorrelation function (ACF) is one of the major tools in time series analysis and is used for identifying features within a time series. It has similarities with the correlation, but relates a series for different time lags. Analysis in the time domain refers to inferences based upon this function. The autocorrelation between X_t and X_{t+k} is define as

$$r_k = Corr(X, X_{t+k}) = \frac{\mathrm{cov}(X_t, X_{t+k})}{\sqrt{[\mathrm{var}(X_t), \mathrm{var}(X_{t+k})]}}$$

Consider a simple random model $Y_t = c + e_t$ where e_t is a random error component that is uncorrelated between time periods. This model is often know as a "white noise process", where no pattern is discernable in a given time series. Theoretically the coefficients of such a series of random numbers would be zero. However, as all time series are finite in length, the sample ACF will not be exactly zero. (Makridakis et al, 1998). The autocorrelation coefficients of the white noise process have a distribution approximated by a normal curve with a mean of zero and standard error $1/\sqrt{n}$ where n represents the number of observations. This provides the value that 95% of all sample autocorrelation coefficients should lie within $\pm 1.96/\sqrt{n}$, referred to as a 95% confidence interval. It should be noted that it is common to approximate the limits to $\pm 2/\sqrt{n}$. The 95% confidence interval for the triage data series becomes ± 0.105.

The plot of the ACF against the lag is known as a *Correlogram*, with r_k plotted against the lag k. The correlogram is a useful tool often used to identify seasonality within a time series, appropriate time series models and whether a series is stationary.

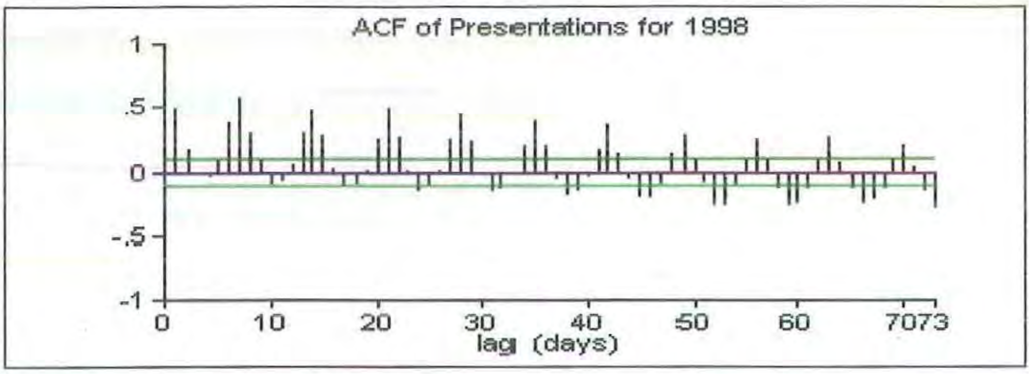


Figure 3.1: ACF plot of presentations January – December 1998.

The correlogram displayed in Figure 3.1 clearly indicates that this time series is not stationary and the "roller coaster" pattern is evidence of seasonality within the series.

Chatfield (1990, p31) describes the ACF of a stationary model as having 3 significant characteristics:

1) the ACF is an even function such that $\rho_{t-k} = \rho_{k-t}$
2) the value of the ACF is one at log zero.
3) lack of uniqueness.

If a given stochastic process does not have a unique covariance structure, then in most cases the converse is false.

3.5 PARTIAL AUTOCORRELATION

This is a second measure of correlation used to identify the relationship between current values with earlier values of the same variable while keeping the effects of all other lags constant (Makridakis et al, 1998). The partial autocorrelation function (PACF) is another significant analytical tool. The values of the PACF are obtained by substituting sample ACF's r_j as

$$r_j = \phi_{k1} r_{j-1} + \phi_{k2} r_{j-2} + \ldots + \phi_{kk} r_{j-k}$$

where $j = 1, 2, \ldots, k$

and solving the resulting sample PACF coefficients. These coefficients, ϕ_{kk} are assumed to be normally distributed and have a mean of zero and standard error $\mathrm{var}[\phi_{kk}] \cong 1/\sqrt{n}$ for $k \geq p+1$ (Box et al., 1976, p. 64-65)

Knowledge of the PACF for a time series provides a simple method for identifying pure autoregressive behaviour of any order. The PACF is graphed in the same manner that for the ACF, with ϕ_{kk} versus lag k and is characterised by an abrupt cut-off at values of k higher than the true autoregressive order p. The partial autocorrelations should all be close to zero for a white noise series. The same critical value, the 95% confidence interval, is used to identify model parameters on a graph of the partial autocorrelation values. Figure 3.2 displays the PACF for the January – December period in 1998. A sharp decline is seen after the first lag. Subsequent values above the 95% confidence interval are likely evidence of seasonal values. Unlike the plot of the correlogram in Figure 3.1, non-stationarity of the series is not as clearly defined.

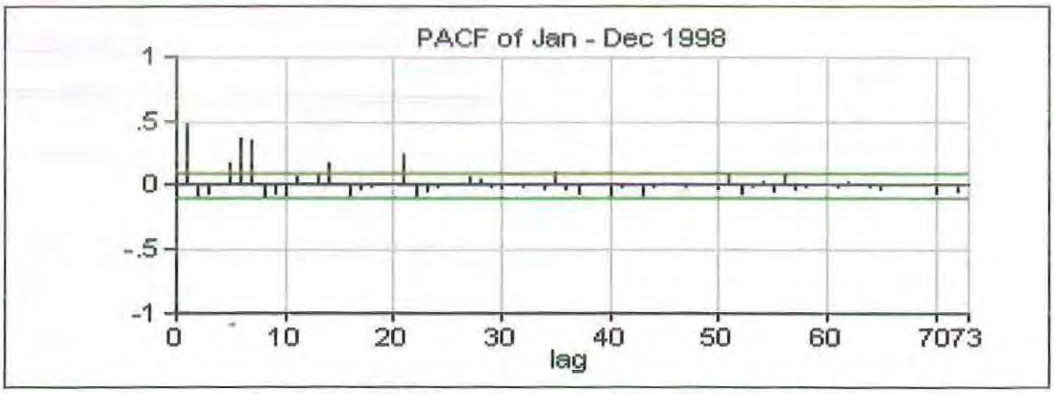


Figure 3.2: PACF for time series May – October 1998

3.6 NON-STATIONARITY

A time series that is non-stationary defines a process that does not have a constant mean or variance. A graphical representation of the time series can help identify the presence of either or both of these conditions, together with use of the ACF. Periodic fluctuations, termed seasonal behaviour or seasonality, may also be identifiable.

Time series modelling techniques are based upon stationarity. Thus a non-stationary series must be transformed to stabilize the variance, turn a seasonal effect additive and attain a normally distributed data series. There are two main methods used to attain this

stationary state. When a series does not have a constant mean, then it may be transformed through *differencing*. This creates a new series of successive differences by applying successive operations of the form $\nabla x_t = x_t - x_{t-1}$. A transformation by this method aims to remove any underlying trend present within the time series. When a time series does not have a constant mean and variance it will often exhibit exponential growth. A logarithmic transformation is required to remove the multiplicative nature underlying this type of time series. Figure 3.3 displays the effect of taking a first difference to the time series of presentations for 1998.

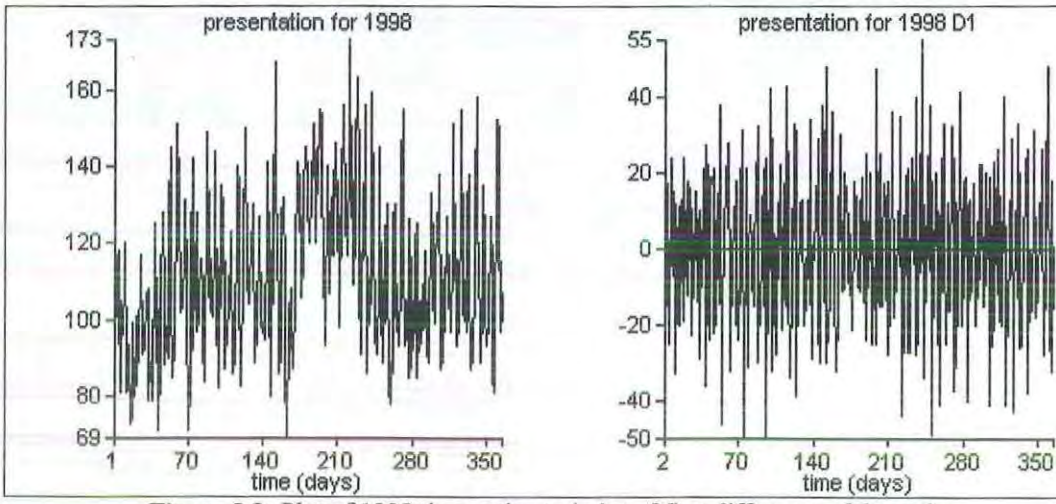


Figure 3.3: Plot of 1998 time series and plot of first difference of the series.

A third method of transformation for a time series that contains a trend pattern is curve fitting. A variety of curves are used including exponential, logistic, polynomial and Gompertz. The fitting of a polynomial curve is a recommended transformation method when a significant length of historical data is available. While this method is commonly used for non-seasonal data, it can be appropriate for the removal of an underlying trend pattern. A second transformation method may then be employed to remove any seasonal pattern present within the de-trended time series. Many statistical packages today will add differing trend lines to a given time series plot. A polynomial trend line may defined as:

$$y = b + c_1 x + c_2 x^2 + c_3 x^3 + ... + c_k x^k + e$$

where b and $c_1 ... c_k$ are constants and e the residuals.

The result of fitting a 6^{th} degree polynomial trend line to the series representing presentations for the year 2000 is shown in Figure 3.4. This type of function provides a measure of trend and the residuals an estimate of local fluctuations

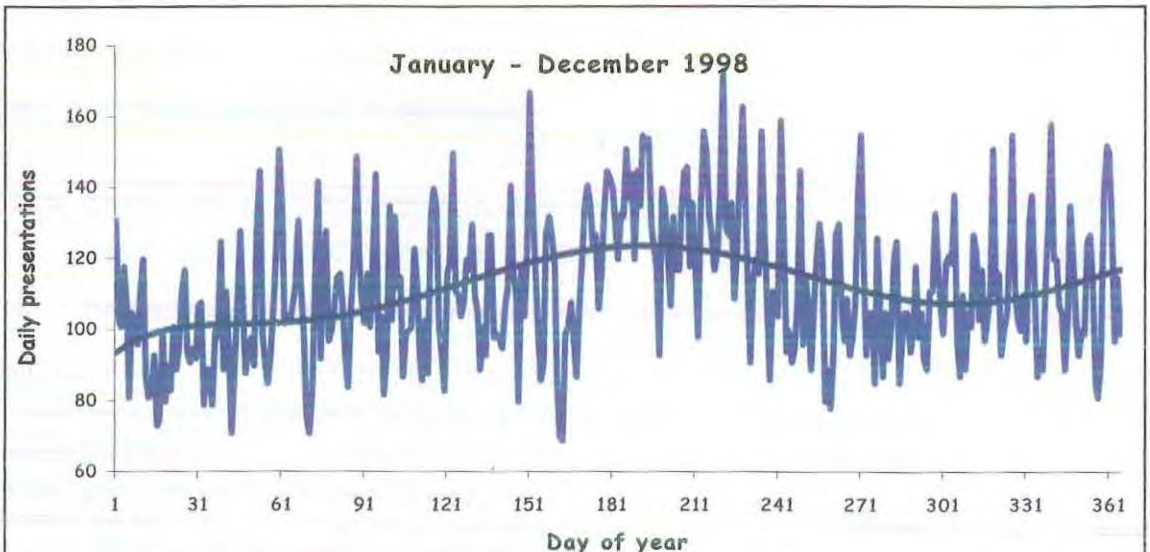


Figure 3.4: Polynomial trend line for the 2000 series.

3.7 NOTATION

The study of stochastic models, like most mathematical study, benefits from the use of simplified notation for the understanding of model equations that are generated. Two operators commonly made use of are the *difference operator* and the *backward shift operator*.

The first difference of a time series, as displayed previously may be expressed by $\nabla x_t = x_t - x_{t-1}$. A second consecutive difference of the original series may be written as $\nabla^2 x_t = \nabla(x_t - x_{t-1}) = \nabla(\nabla x_t)$. Thus in general terms, the d^{th} consecutive difference becomes $\nabla^d x_t$.

The backward shift operator is the main such operator preferred by many authors including Box et al, Makridakis et al and Chatfield. It is may be defined as $BX_t = X_{t-1}$. That is B operating on X_t effectively moves the data back one period. The

backshift operator therefore provides a convenient method for describing the differencing process. Thus $\nabla X_t = X_t - X_{t-1} = X_t - BX_t = (1-B)X_t$. Applying to the general case, a d^{th} difference may be written as $(1-\boldsymbol{B})^d \boldsymbol{X}_t$.

3.8 AUTOREGRESSIVE MODELS

These models are extremely useful in providing description of non-seasonal business and economic time series (Newbold et al, 1990). The model is a form of regression in that it expresses a forecast as a function of previous values of the time series together with an error or white noise component. The model may be represented by

$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} + Z_t$ where Z_t represents the error term,

or using the backshift operator

$$\boldsymbol{Z}_t = (1 - \alpha_1 \boldsymbol{B} - \alpha_2 \boldsymbol{B}^2 - ... - \alpha_p \boldsymbol{B}^p)\boldsymbol{X}_t$$

These models are termed an autoregressive (AR) process of order p and abbreviated to AR(p). (Chatfield, 1990, p. 35)

3.9 MOVING AVERAGE MODELS

The autoregressive model expresses the error Z_t as a finite weighted sum of p previous errors plus a random shock or white noise. In a moving average model, the value of the time series at time t is influenced by a current error term and a finite number q of weighted error terms in the past. Thus a moving average (MA) model of order q is represented by $X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + ... + \beta_q Z_{t-q}$ or alternatively $X_t = \theta(B)Z_t$ where $\theta(B)$ is a polynomial of order q in B. Moving average models are often denoted by MA(q). These models are commonly used in econometrics, where random events, such as material shortages, government decisions and strikes, will perceivably affect not only current economic indicators but also several subsequent time periods.

3.10 AUTOREGRESSIVE MOVING AVERAGE MODELS

This class of model is a coupling of autoregressive (AR) models with moving average (MA) models. These models provide a better representation of an actual stationary time series as they include the predictive ability of both prior models. An autoregressive moving average (ARMA) model of order (p,q) is represented by

$$X_t = \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-1} + Z_t + \beta_1 Z_{t-1} + \ldots + \beta_q Z_{t-q}$$

or, alternatively using the backshift operator

$$\phi(B)X_t = \theta(B)Z_t$$

where $\phi(B) = 1 - \alpha_1 B - \ldots - \alpha_p B^p$ and $\theta(B) = 1 + \beta_1 B + \ldots + \beta_q B^q$

3.11 BOX JENKINS AUTOREGRESSIVE MOVING AVERAGE MODELS (ARIMA)

The ARMA model may be expanded to an autoregressive integrated moving average (ARIMA) model by allowing for differencing of the given time series. These are forecasting models popularised by George Box and Gwilym Jenkins in the early 1970's and commonly referred to as the Box and Jenkins methodology. The forecast series is expressed as a function of both the previous values (autoregressive terms) and the previous error values (moving average terms) from the forecasting producing an ARMA model. By using the notation applied to differenced time series data together with that for the ARMA model, the ARIMA model is represented by

$$W_t = \alpha_1 W_{t-1} + \ldots + \alpha_p W_{t-p} + \ldots + \beta_q Z_{t-q}$$

or using the backshift operator

$$\phi(B)(1-B)^d X_t = \theta(B)Z_t$$

Note differing authors use their own symbols for identifying autoregressive and moving average terms; W_t used by Box and Jenkins (1976) equates to X_t by Chatfield (1996).

This last type of model is often applied with the addition of a seasonal component. Many real world time series contain periodic components repeating at a regular time period. Company sales figures, the usage of public transport system and maximum

temperature recordings will vary each year in a similar pattern to the previous years. Box and Jenkins generalised the ARIMA model to incorporate a seasonal component, the generalized multiplicative seasonal ARIMA commonly referred to as SARIMA, and defined it by

$$\phi_p(B)\Phi_P(B^s)W_t = \theta_q(B)\Theta_Q(B^s)Z_t$$

where $\phi_p, \Phi_P, \theta_q, \Theta_Q$ are polynomials of order p, q, P and Q respectively, Z_t represents white noise and $W_t = \nabla^d \nabla_s^D X_t$ with dth difference and D seasonal difference (Chatfield, 1990, p.60).

Software packages will often return a predictor equation for a SARIMA in the following expanded format of the above equation.

$$(1-\alpha B)(1-\alpha B^s)(1-B)(1-B^S)X_t = (1-\theta B)(1-\theta B^S)Z_t$$

It must be remembered that only those parameters identified in a model will be presented. Figure 3.5 displays a model fitted to a 245 day time series for triage data from 2000.

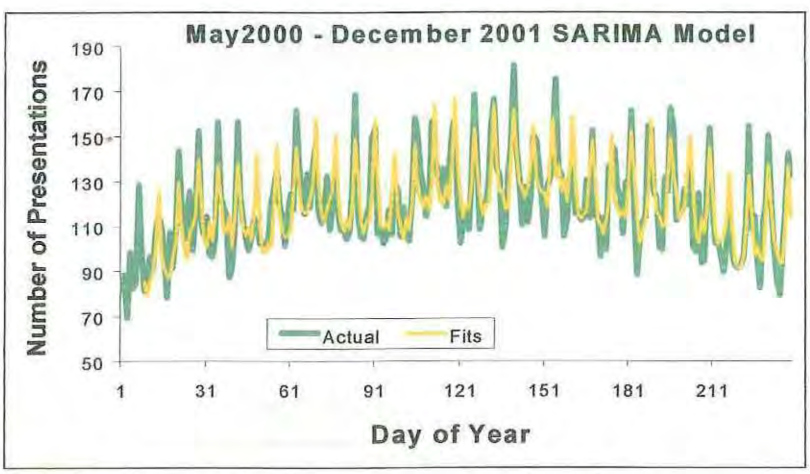


Figure 3.5: SARIMA model (yellow) over raw data from 2000 triage series.

The prediction equation for the SARIMA model for this time series is as follows

$(1 - 0.26B)(1 + 0.89B^7)\nabla_1 \nabla_7 x(t) = (1 - 0.92B)(1 - 0.05 B^7)(1 - 0.83 B^{14}) w(t)$

This model under predicts future triage values when they differ significantly from the previous data value.

3.12 SPECTRAL PROPERTIES OF STATIONARY MODELS

An alternative method of analysing a time series is to assume that it consists of sine and cosine waves of different frequencies. The periodogram, introduced by Schuster in 1898, was the first to use this idea. Models may be developed to include a combination of sinusoidal components at differing frequencies. A stationary time series may be obtained by allowing the amplitude to vary according to a stochastic process and this may be expressed as $X_t = \sum\left(a_j \cos\omega_j t + b_j \sin\omega_j t\right) + Z_t$. All frequencies are in the range 0 to π, with the upper frequency termed the Nyquist frequency (Chatfield, 1990). A link between the frequency and the time domain is provided by the Weiner-Khintchine theorem and for a stationary stochastic process is

$$\gamma(k) = \int \cos\omega t \; d\; F(\omega)$$

where $F(\omega)$ is called the spectral distribution function. It should be noted that $F(\omega)$ represents the contribution to the variance of the frequencies in the range 0 to ω. When $F(\omega)$ is continuous, the equation becomes

$$\gamma(k) = \int \cos\omega t \; d\; f(\omega)\; d\omega$$

where $f(\omega)$ is termed the spectrum. It is this measure that allows possible identification of the frequencies that contribute towards the variation within the data. Should one such frequency dominate, this could indicate a cycle that should be removed prior to modelling. Through use of this technique, the Box-Jenkins ARIMA model can be expanded to accommodate a seasonal component, a model termed a SARIMA as discussed in the previous section.

The spectral function provides an analytical tool allowing conclusions about data variations to be drawn from another source. The software package TSA – 32 is used to plot the density function. The x-axis of the resulting graph is the frequency measured in radians. The total area under the curve equates to the variance of the process. A spike or peak in the curve indicates an important contribution to the variance at that interval.

In Figure 3.6(a) the spectral density function has been plotted for the triage data January to December 1998. The variance is clearly concentrated at two significant levels. The

first indicates the presence of a trend component. This was removed, with the resulting density function displayed by Figure 3.6 (b). Here the peak corresponds to ω = 0.143. The other two smaller peaks, at $\omega = 0.287$ and $\omega = 0.430$ are termed harmonics as they occur at multiples of ω. In general these extra peaks simply indicate the non-sinusoidal character of the main seasonal component (Chatfield, 1996). If we refer to the frequency by $f = \omega / 2\pi$, number of cycles per unit of time, we have a much easier to interpret form for the frequency (Chatfield, 1996). The period or wavelength becomes $1/f$ and is calculated by $2\pi/\omega$. The detrended series corresponds to a cycle or season length of 7 days.

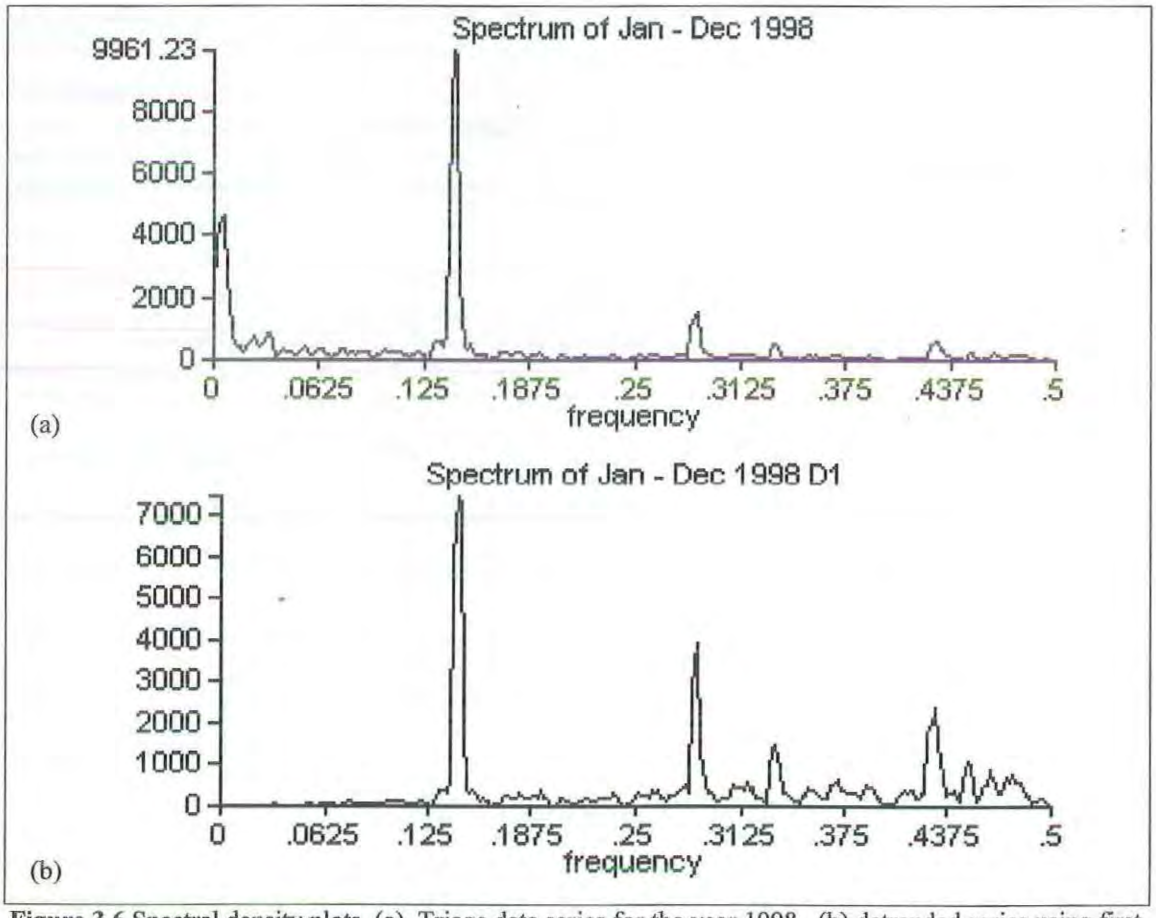


Figure 3.6 Spectral density plots. (a). Triage data series for the year 1998. (b) detrended series using first differencing method within TSA-32 software.

Figure 3.7 displays the final spectral density plot post removal of this seasonal component. The dominant frequencies now equate to the harmonics of the season.

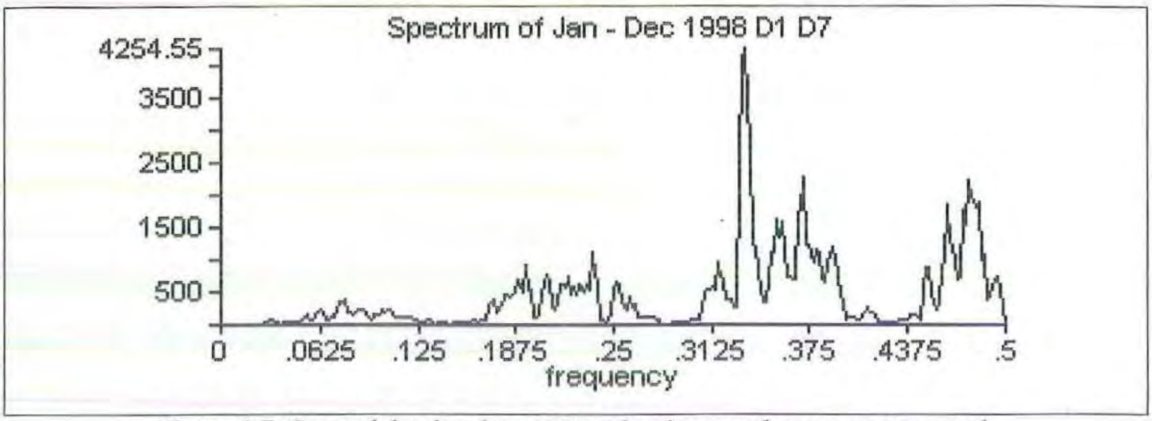


Figure 3.7: Spectral density plot, post trend and seasonal component removal.

3.13 AKAIKE'S INFORMATION CRITERION

Often more than one plausible model is identified for any given time series. The majority of computer based analysis will produce models that optimise by use of a chosen function of the data. This is analogous to a regression model being chosen based upon the adjusted value of R^2. Similarly for ARIMA models, the likelihood of a model is penalised for each additional term. The Akaike's Information Criterion (AIC) is the most common of these procedures and the one utilised by the TSA-32 software package and may be defined by $-2LogL + 2m$, where L denotes the likelihood, and $m = p+q+P+Q$. It should be noted that while not all computer packages provide the AIC, all will produce a variance value (σ^2) and using this an approximation of the AIC can be attained by $n(1+(2\pi)+n\log\sigma^2+2m)$, n being the number of observations in the series. Figure 2.6 displays the results of a model selection for the 1998/99 triage series, based upon the AIC value.

ARIMA search on May 1998 - Dec 1999

p	*d*	*q*	*P*	*D*	*Q*	*s*	*AIC*	
0	1	1	1	1	1	7	6.2849	
0	1	1	2	1	1	7	6.2780	
1	1	1	1	1	1	7	6.2370	
1	**1**	**1**	**2**	**1**	**1**	**7**	**6.2138**	**Best Model**
2	1	1	3	1	1	7	6.2447	
2	1	1	3	1	2	7	6.2507	
2	1	1	3	1	3	7	6.2129	
2	1	1	3	1	4	7	6.2285	

Figure 3.8: ASTSA statistics for 1998 / 1999 Total presentation Box-Jenkins Model

4 REGRESSION ANALYSIS

4.1 ABOUT THIS CHAPTER

Regression analysis allows the influence of a number of factors in a data set to be analysed. This chapter provides information on different regression models available, together with the any assumptions underlying each model.

Regression analysis is a commonly used statistical technique that permits the modelling of a relationship between a set of variables. The general tendencies of the relationship between variables are summarized by the mathematical curve and a measure of the variation in data around that curve (Cryer & Miller, 1994, p.176). A regression model is used to relate a dependent variable to one or more explanatory, independent variables. These are then used for prediction of future values of the dependant variable from a given data series.

4.2 CORRELATION

The strength of any regression model is measured by the correlation coefficient, a measure of the relationship or mutual dependence between two variables. This plays a significant role in multivariate, multiple variable, data analysis. Unlike the covariance, the correlation provides a pure, scale free measure of the strength of a linear relationship and may be defined by:

$$\rho = \frac{C_{xy}}{\sigma_x \sigma_y} \qquad \text{with, } -1 \le \rho \le 1.$$

A large absolute value of ρ, represents a strong linear association, while no association has a correlation value of zero. While this simple test of data association is widely used, it should be used with caution. It is a measure of linear association. Variables associated non-linearly will provide a near zero correlation coefficient. Secondly, a small sample size will impact upon the correlation and could produce a spurious value. The final word of caution in using a correlation coefficient concerns the presence of

extreme values. The presence of the outlier may significantly skew the overall data weights to produce a new correlation value. A scatterplot of a data series will alert the analyst to the possible occurrence of each of these errors.

Correlations

		TRIAGE98	HOLCOD98	DAYTYP98
TRIAGE98	Pearson Correlation	1.000	.063	.330**
	Sig. (2-tailed)	.	.230	.000
	N	365	365	365
HOLCOD98	Pearson Correlation	.063	1.000	.059
	Sig. (2-tailed)	.230	.	.265
	N	365	365	365
DAYTYP98	Pearson Correlation	.330**	.059	1.000
	Sig. (2-tailed)	.000	.265	.
	N	365	365	365

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 4.1: Correlation matrix for 1998 triage data series

The correlation matrix for the January to December 1998 triage series, Figure 4.1, indicates a significant association between the daily presentation value and day of the week. The boxplots for the day of the week, Figure 4.2, support this association with a significant difference between each day's presentations.

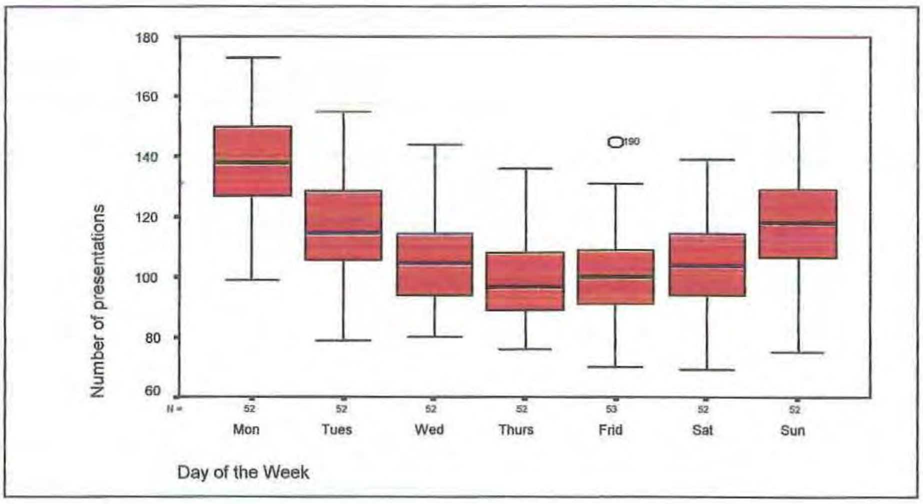


Figure 4.2: Boxplot graph of 1998 triage data by day of presentation

4.3 SIMPLE REGRESSION

Simple regression refers to the fitting of a straight line to the scatter plot of a data series to explain the relationship between a predictor variable, X and its response variable, Y. The equation for this straight line is generally chosen by the *method of least squares*. The regression equation has the form: $\boldsymbol{Y} = \boldsymbol{b}_0 + \boldsymbol{b}_1 \boldsymbol{X}$ where b_1 is the slope of the line and b_0 the Y intercept, commonly termed the constant.

The correlation between dependent and independent variables ρ, designated R in regression analysis, provides a measure of their relationship. It is common to present the R^2 statistic, known as *the coefficient of determination*, to illustrate the adequacy of a regression model. It may be defined by

$$R^2 = r_{y\hat{y}}^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{SSR}{SST}$$

This value can be viewed as the proportion of variance that is explained by the fitted model.

4.4 MULTIPLE REGRESSION

This may be thought of as an extension of simple regression, such that there is one variable to be predicted by two or more explanatory variables. The general form of multiple regression is

$$Y_i = b_0 + b_1 X_{1,i} + b_2 X_{2,i} + \ldots + b_k X_{k,i} + e_i$$

where $b_0, b_1, \ldots, b_k$ represent fixed but unknown parameters, $Y_i, X_{1,i}, \ldots, X_{k,i}$ are the ith observations at each variable $Y, X_1, \ldots, X_k$, and e_i represents white noise.

A scatterplot matrix enables visualisation of relationships between variables. Perhaps of more importance, the matrix helps identify collinearity, the presence of a linear

relationship between variables. Figure 4.3 indicates a relationship is present

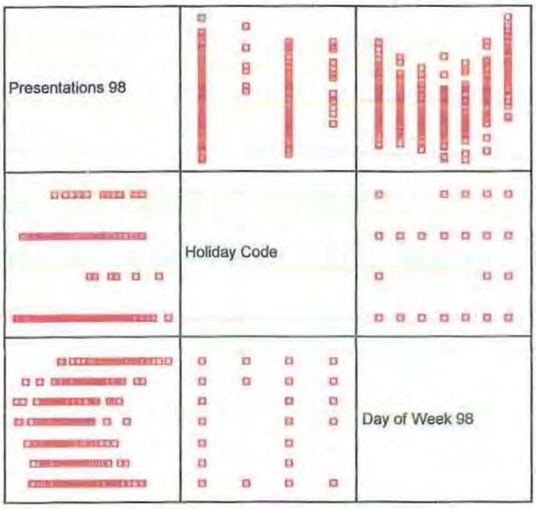


Figure 4.3: Scatterplot of 1998 triage data, day of week, type of day coding.

between the daily presentations and both the day of the week and the type of day it may be. The relationship between type of holiday and the day of the week is of little concern. The regression coefficients of those explanatory variables will be highly unstable in the presence of collinearity resulting in computational errors.

All regression is based upon several basic assumptions.

- The Y_t dependent observations are statistically independent of one another.
- The mean value of Y_t for each combination of the explanatory variables, $X_{1t}, X_{2t}, ..., X_{kt}$, is a linear function of those variables.
- The variance of Y_t is the same for any fixed combination of the explanatory variables, $X_{1t}, X_{2t}, ..., X_{kt}$. a condition termed homoscedasticity.
- The explanatory variables $X_{1t}, X_{2t}, ..., X_{kt}$. have values that are both random and uncorrelated with the error term e_i, or are measured without error and fixed. Their values must differ from one variable to the next.
- The residuals or error terms, e_i, are uncorrelated, have a zero mean, variance σ_s^2 and are normally distributed. That is they represent white noise. These residuals support the use of the ordinary least squares method for estimation of regression coefficients.

4.5 TIME-RELATED REGRESSION

A multiple regression model can be expanded to include variables that allow for time related features within a series. Each of these extra variables, termed binary predictor variables, equates to a new explanatory variable and can explain the effects of seasonality, the impact of holidays or one off events such as new safety campaigns. The general rule applied when adding seasonal binary variables is to use 1 less than the number of periods. For the emergency data series an appropriate model would be

$$Y_t = b_0 + b_1 Mon_t + b_2 Tue_t + \ldots + b_6 Sat_t + e_t$$

where the binary seasonal indicators are defined as

$$Mon_t = \begin{cases} 1 \text{ if } t = Monday \\ 0 \text{ if } t \neq Monday \end{cases}, \; Tue_t = \begin{cases} 1 \text{ if } t = Tuesday \\ 0 \text{ if } t \neq Tuesday \end{cases}, \ldots, \; Sat_t = \begin{cases} 1 \text{ if } t = Saturday \\ 0 \text{ if } t \neq Saturday \end{cases}$$

The resulting model coefficients represent the average difference between daily forecast values for each of the days and the omitted day. To avoid multicollinearity problems, the number of indicator variables to use must be one less than the seasonal divisions being used. Hence, above there is no indicator variable for Sunday.

The addition of variables to a regression model should be done with caution. Each addition requires another regression coefficient to be estimated, and a reduction in the degree of freedom for the error term. It must be noted that while extra variables always increase the R^2 value, this alone does not justify their inclusion.

4.6 REGRESSION WITH ARIMA

The ARIMA models, discussed in chapter three, enable analysis of a time series but allow no input of other information. More often than not, variations within a given time series are influenced by other factors. The ED data is influenced by the day of the week, season of the year and the occurrence of holidays. Regression models allow such

input, but often the resulting model is inaccurate due to the presence of autocorrelations within the error term of each model. A combination of these two models would allow the advantages of regression to be coupled with the power of ARIMA. It must be noted by the reader, that a more complicated model such as this may not produce a better result. Simpler models often produce adequate data modelling with acceptable forecasts.

A regression model assumes that the resulting error term e_t is uncorrelated, that is "white" noise. However, autocorrelations often exist within the error term and as such it would be appropriate to fit an ARMA model to those errors. These autocorrelated errors are referred to by N_t. It should be noted that an error term will still remain after such a model is fitted, the "white" noise. Applying ordinary least squares estimation often fits regression models to a time series. The application of an ARMA model to the regression model error terms results in two common problems. There is no account of time relationships between data values and the resulting error terms will be too small in the presence of autocorrelations. The latter is a far more serious error. The small errors lead to misinformation; explanatory variables become significant when they are not, an occurrence known as "spurious regression".

Thus, an autocorrelated error series necessitates the use of maximum likelihood estimation or generalised least squares estimation. The latter estimates are obtained by minimising

$$G = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j N_i N_j$$

where w_i and w_j are weights bases upon autocorrelations pattern.

4.6.1 REGRESSION WITH ARIMA

Residual analysis, discussed in Chapter 6, will help identify if autocorrelations are present in the resulting error term of a regression model. Once identified, an initial proxy model such as an AR(1) or AR(2) It may be necessary to difference the error

series, if the application of a proxy model does not result in stationarity. The stationary series may then have an appropriate ARMA model applied. The completed model, regression with ARMA parameters, is then reapplied to the time series under estimation. The resulting residual series should be represented by e_t, "white" noise.

4.7 BEST SUBSETS REGRESSION

The addition of each variable to a regression equation may improve a given model's prediction ability. However, these additions come with an added cost and complexity of model. The best subsets regression technique determines the best linear regression models using the maximum R^2 criterion (coefficient of determination). The technique examines all linear regression models with 1, 2, 3, ... supplementary variables, determines the best 2 and then displays summary statistics for each. These statistics include the R^2, R^2 adjusted and Mallows C-p statistic. A researcher will usually adopt the model that optimises the balance between simplicity and fit. The model with the smallest Mallows C-p statistic is that model (Minitab, version 12.2).

4.8 CROSS CORRELATION FUNCTION

The Cross Correlation Function (CCF) is a standardized measure of the association between one series Y_t and examining past, present or future values of a second series X_s. It is commonly utilised to check for correlation between two time series or to determine if one series can provide a measure of predictability for a second such time series.
It may be defined by:

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sqrt{\gamma_{xx}(0)\gamma_{yy}(0)}} \qquad \text{(Chatfield, 1996)}$$

5 DATA ANALYSIS

5.1 ABOUT THIS CHAPTER

This section of the thesis describes how the data series were analysed using the methods outlined. Each method was used to study different aspects of the data to address the aims of the thesis. Modelling was undertaken using total daily presentation numbers and those with a triage category 4. It was discovered during the initial examination of raw data that this triage category represents up to 80% of a day's total presentations. Modelling and subsequent forecasting of this group could provide an indication of total expected presentations and thus resource allocation requirements. Modelling with other triage categories or combinations of them was considered unviable due to insufficient numbers and irregular patterns of presentation, as displayed in Figure 5.1. It should be noted that the 1999 and complete 3-year series are included within this chapter. All analysis and models, including 1998 and 1999 are to be found within the appendices.

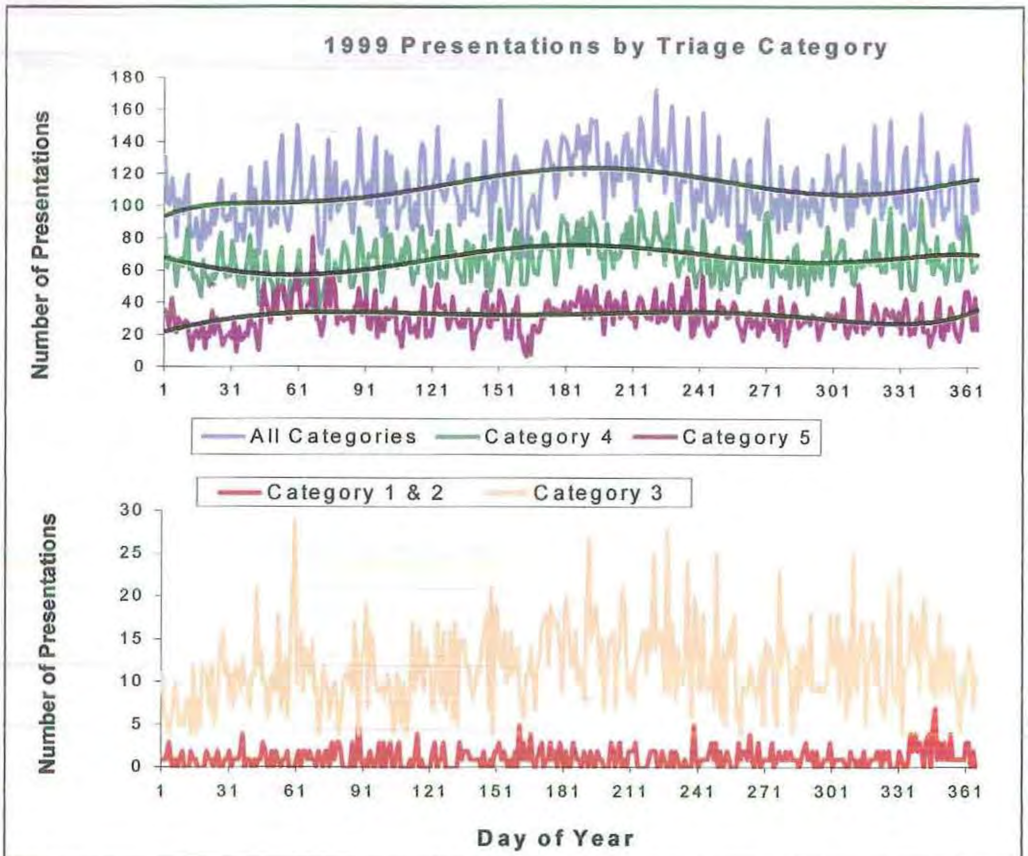


Figure 5.1: (a) ED daily presentations by triage category 1999. (b) Total daily presentations versus category 4

5.2 THE TIME PLOT

In time series analysis the first step is to plot the given data series against time. This plot may immediately reveal systemic features such as a trend, seasonal or cyclic patterns, the presence of outliers and any discontinuities within the given series.

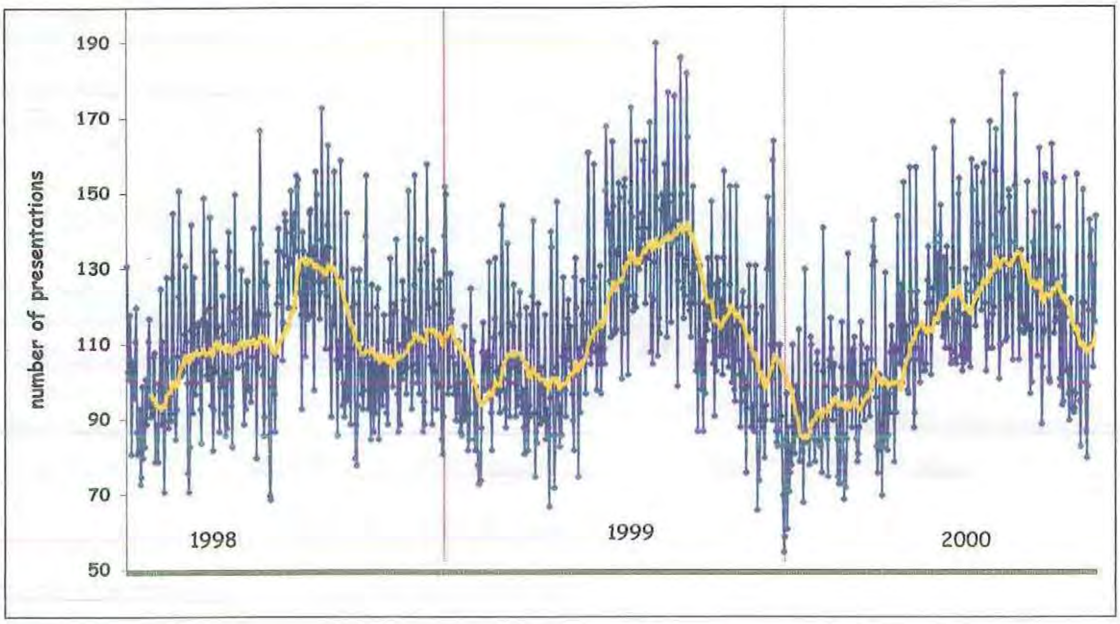


Figure 5.2: ED presentations January 1998 – December 2000 with 6^{th} polynomial trend line.

The time plot of the complete 3 years of the emergency data (ED) series is displayed in Figure 5.2. Clearly there is a seasonal cycle, as indicated by the 6^{th} polynomial trend line placed over the series, which equates to approximately 12 months in length. Discontinuity is not an issue with this series. However, the occasional data value appears significantly higher or lower than those nearby it. These data values could be classed as outliers. As a past worker within an emergency department, these sudden changes in presentation rates were often related to one-off events or public holidays. The reader should note that the New Year holidays in 1999 and 2000 fell upon a weekend. This resulted in a significantly higher presentation rate than would have been normally been expected for a Saturday or Sunday.

The length of season was expected given the nature of the ED data series. Hospital emergency departments have come to expect higher daily numbers in the winter and spring months. Christoffel (1984), Attia & Edward (1996), and Diehl et al (1981) all

concluded that the prevailing season of the year strongly influenced the daily presentation numbers. To investigate the presence of other seasonal cycles within the data, the time scale was reduced to 4 months. Figure 5.3 displays such a series from the 1999 data where a much shorter seasonal pattern of approximately 7 days becomes apparent.

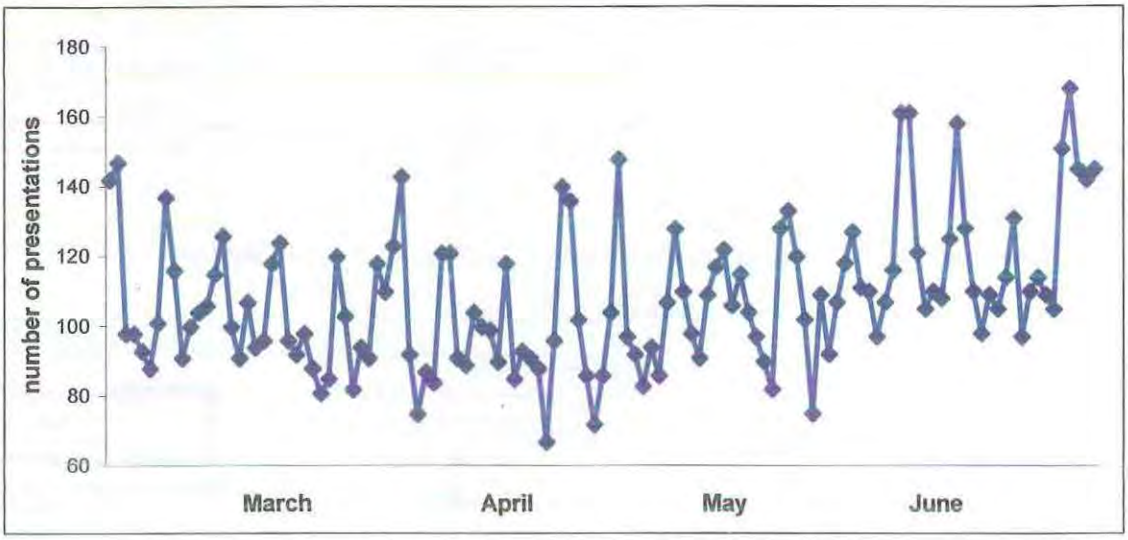


Figure 5.3: ED daily presentations March – June 1999

This would concur with the findings of previous studies undertaken in the northern hemisphere. While the pattern cannot be considered constant, it is strong enough to enable progress to the next stage of analysis.

5.3 SPECTRAL ANALYSIS

The spectral density function discussed in Chapter 3 provides the mathematical method of verifying what these time plots display. In Figure 5.4 both plots are remarkably similar, each having a very high peak at a value near zero and a second significant peak at 0.143. The initial peak would be indicative of a probable strong trend component, but is more likely evidence of the 12-month cycle underlying the ED data series. The period of the second peak, calculated from the inverse of the frequency, equates to 7 days. Both of these findings equate to those observable from the time series graphs.

Once the components of a time series are identified reliable short-term forecasts can be produced using either dynamic or stochastic modelling techniques. The two techniques

applicable to this time series are the Holt-Winters and the Box-Jenkins. Both of these models express a forecast based upon previous values within a time series.

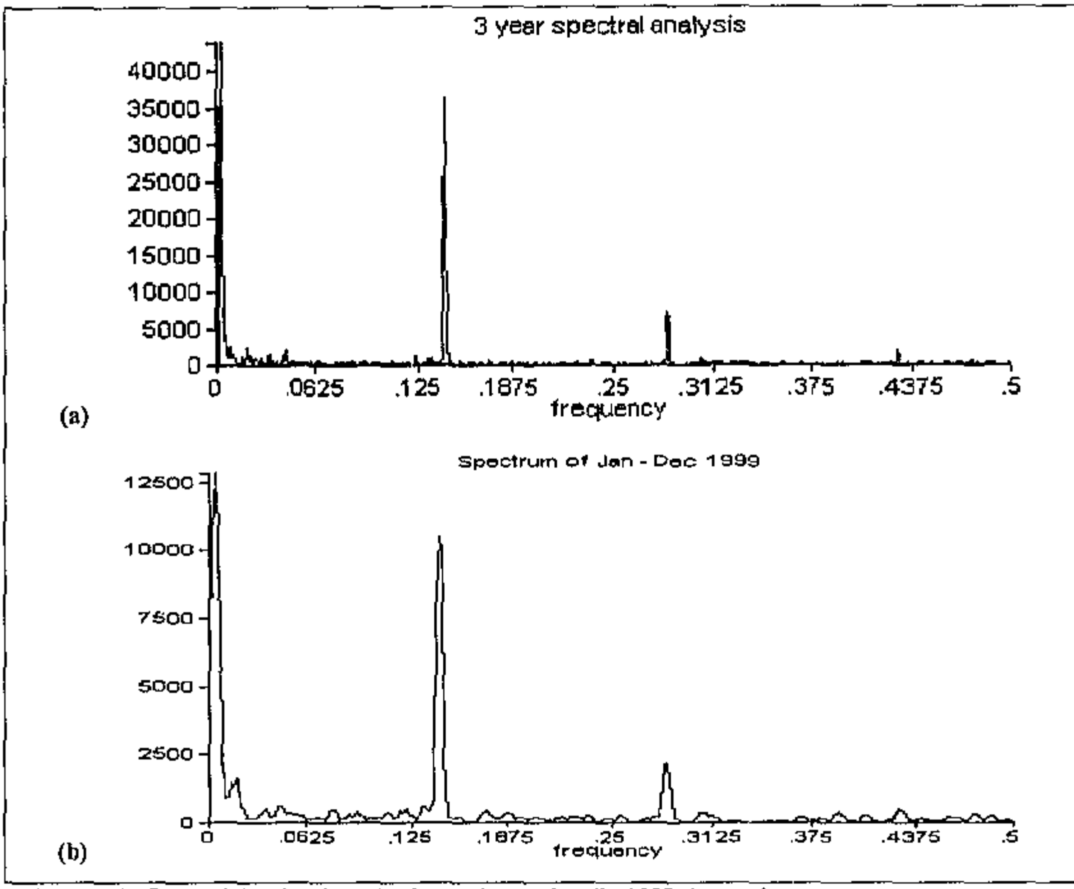


Figure 5.4: Spectral density plots. (a) 3-year data series. (b) 1999 data series.

5.4 THE HOLT-WINTERS METHOD

This was first choice to analyse the ED data. As indicated in Chapter two, this generalisation of the single exponential model enables a data series containing both trend and seasonality to be analysed. The first decision that needed to be taken was to determine if the ED data series could be considered multiplicative or additive. The time series plots clearly indicate a marked variation in season and thus could not be considered additive. This means that the higher values will have a greater variation in the level of trend than the lower values. A template was established using an Excel spreadsheet to undertaken the necessary calculations using the three basic equations given in Chapter two.

All Holt-Winters models need to have the three smoothing parameters α, β and γ initialised, with values usually chosen between 0 and 1. While the default values of 0.2 are often an adequate starting position, the Mean Squared Error (MSE) method provides an appropriate estimation and removes the guesswork. The Excel spreadsheet coupled with use of the solver algorithm enabled appropriate parameter values to be determined that minimised the MSE. A second Holt-Winters model, using these 3 parameter values, was developed through use of the Minitab program. This led to comparative models being obtained for each year of the years 1998, 1999, 2000 and the 3-years combined series using the total daily presentations and the triage category 4 data values. The parameter values for the 3-year series and 1999 model are given in Table 5.1.

	α	β	γ
1999 series	0.24	0.03	0.15
3 year series	0.25	0.01	0.14

Table 5.1: Holt-Winter parameter values.

The similarity in the parameter values for all the Holt-Winters models, given in Appendix B, would enable all models to be developed with a single set of parameter values, with minimal impact upon models or forecast values.

Figures 5.5 displays the prediction model versus the actual data values for the complete 3-years of data and the 1999 series model. These models can be considered a good fit. They are able to replicate the short term 7 day cycle and the longer annual cycle, though they do appear to often miss peaks and troughs in the cycles. This was a common fault with the other Holt-Winters models included in Appendix C. The two models using the 2000 data suffered initially due to the significant difference in presentation numbers in the first weeks of January. However, the residual analysis for each model does indicate appropriate models were developed, with the residuals approximating a normal distribution.

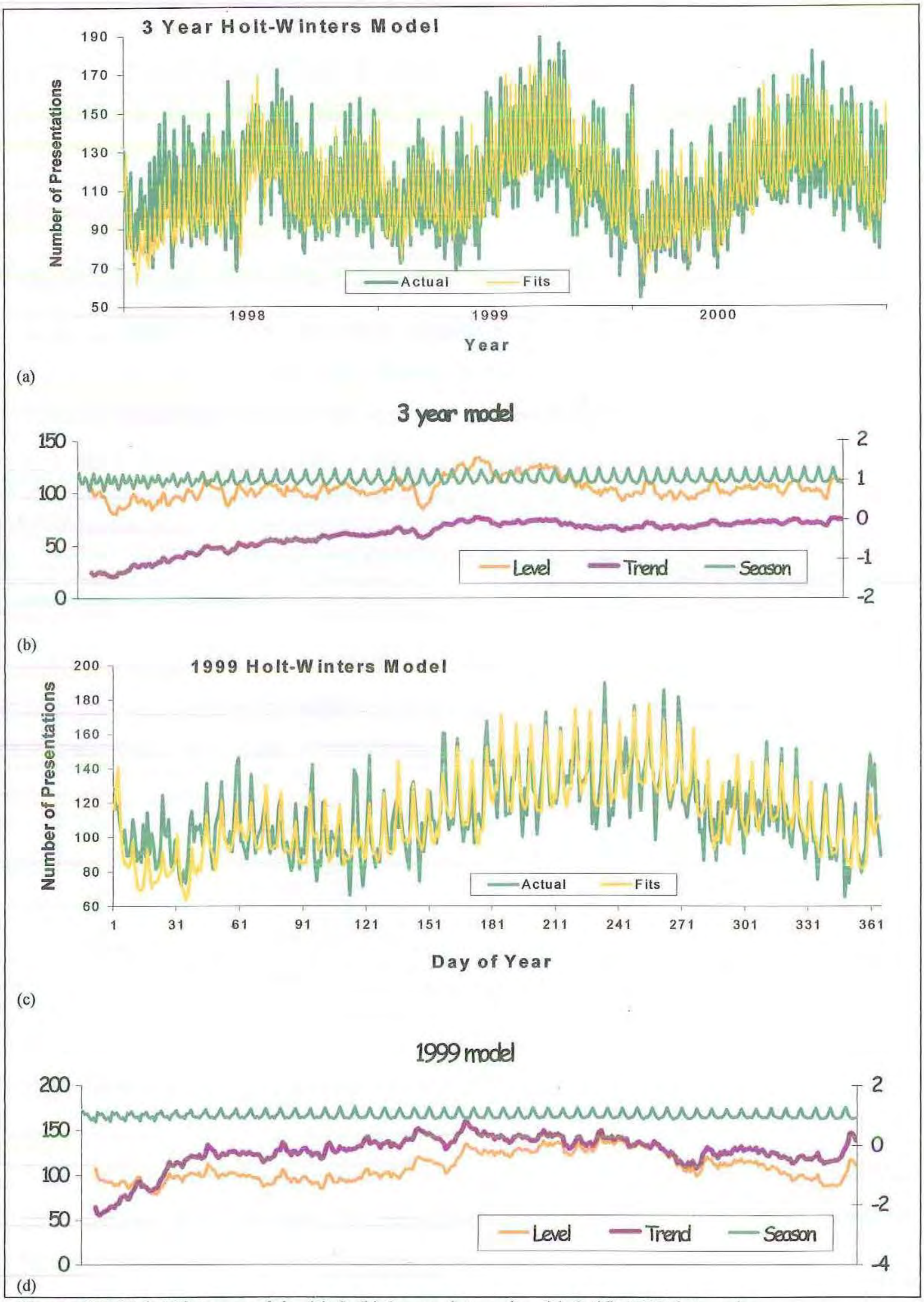


Figure 5.5: Holt-Winters models: (a) & (b) 3-year data series. (c) & (d) 1999 data series.

However, as shown in Figure 5.6, the histograms do indicate the presence of some outlying values. The dot plot of residuals versus fitted values does not indicate a widening as data values increase thus indicating these outlying values do not significantly impact on a model's predictive ability.

The Holt-Winters method decomposes the data into the components level, trend and cyclic or season indices, with these results shown in Figure 5.4 (b) and (d). This dynamic method allows the cyclic indices to change over time. However due to this it provides conservative fits for the peaks and troughs of a series and will generally not fit well when such a series displays rapid data variation over a short time interval. The majority of the parameter values are also noted to be small, with the majority of β and γ values close to zero. When this occurs the resulting models will tend to have a smoothed average effect.

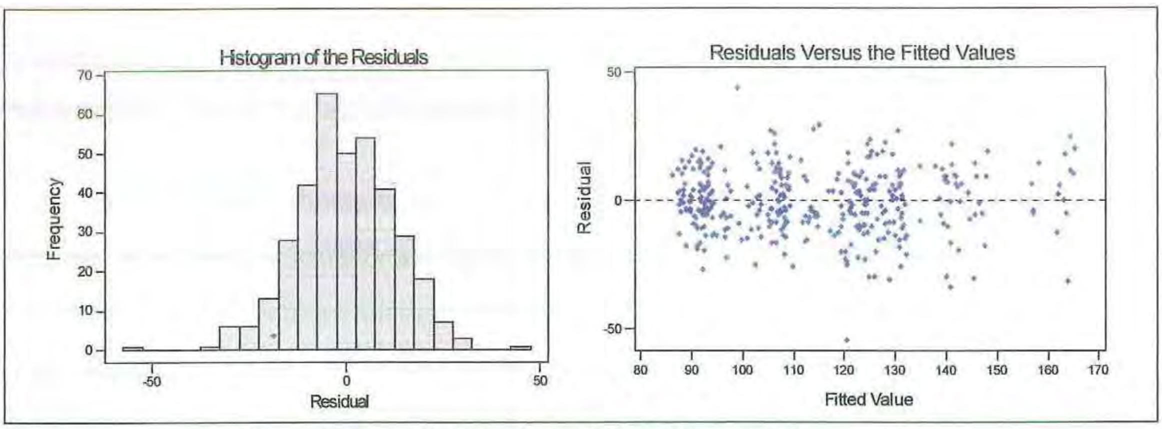


Figure 5.6: Residual Analysis: 1999 Total presentation model.

These facts are clearly illustrated in each of the Holt-Winters model fits to the actual data. The models deteriorate when consecutive days presentations differ significantly from previous days. The forecast values are similarly influenced. The models are each for a calendar year, therefore the initial forecasts include estimates for the month of January, which includes several public holidays. The Holt-Winters model is unable to take such factors into consideration with the result that subsequent forecasts are likely to be affected. Table 5.2 provides the first 20 forecasts for the 2000 year using the 1999 model. The very large presentations, highlighted, on the New Year's Day weekend have influenced the forecasted values. Despite this, the forecasts given by this

technique model the underlying variation in the data and the majority of actual presentations were within the given 95% confidence intervals.

In an attempt to overcome this impact, differing 12-month spans were trialled. However, the impact seen above occurred with each of the other attempts. A shorter time span, perhaps over a 3 or 4-month time interval may be able to provide more accurate forecasts.

Time Period	Actual	Forecast	95% CI Lower	Upper	Time Period	Actual	Forecast	95% CI Lower	Upper
361	**130**	107	78.7	134.4	371	**83**	94	55.7	133.2
362	**143**	96	67.5	124.8	372	**87**	95	55.2	135.4
363	**110**	93	63.9	123.0	373	**110**	106	64.5	147.4
364	**101**	96	65.3	126.4	374	**110**	96	52.8	138.4
365	**90**	98	66.4	129.5	375	**91**	93	48.7	137.1
366	**159**	117	83.9	149.2	376	**83**	95	49.7	140.9
367	**164**	133	99.3	166.9	377	**70**	97	50.4	144.4
368	**110**	109	73.7	143.7	378	**55**	116	67.4	164.3
369	**87**	100	63.7	136.1	379	**59**	132	82.4	182.3
370	**84**	93	55.8	130.7	380	**85**	106	54.2	157.0

Table 5.2: Forecasted values using 1999 model.

5.5 BOX-JENKINS MODEL

This was the second choice of model to analyse the ED series. Both the Holt-Winters and the Box-Jenkins models are able to analyse time series that are cyclic in nature. However, the Box-Jenkins method benefits from the moving average (MA) component. This enables modelling of a time series influenced by random events, such as public holidays or severe weather conditions, that could potentially affect future forecasts. However it could be argued that a disadvantage of the Box-Jenkins technique is it requires more of an analytical background to understand and use of specific software packages to implement than the Holt-Winters. Despite these reservations, it is an appropriate method of analysis for the ED data series.

The initial steps involve the use of the spectral analysis, as previously outlined, to determine the length of cycles or seasons within a series. Unlike the Holt-Winters method, when these cycles are identified it infers that the series is not stationary. The Box-Jenkins models must be made stationary for further analysis, by removing the first difference. The majority of the ED time series required a yearly or trend differencing

and the subsequent removal of the 7-day cycle by taking a further 7th difference. This left time series that were now stationary.

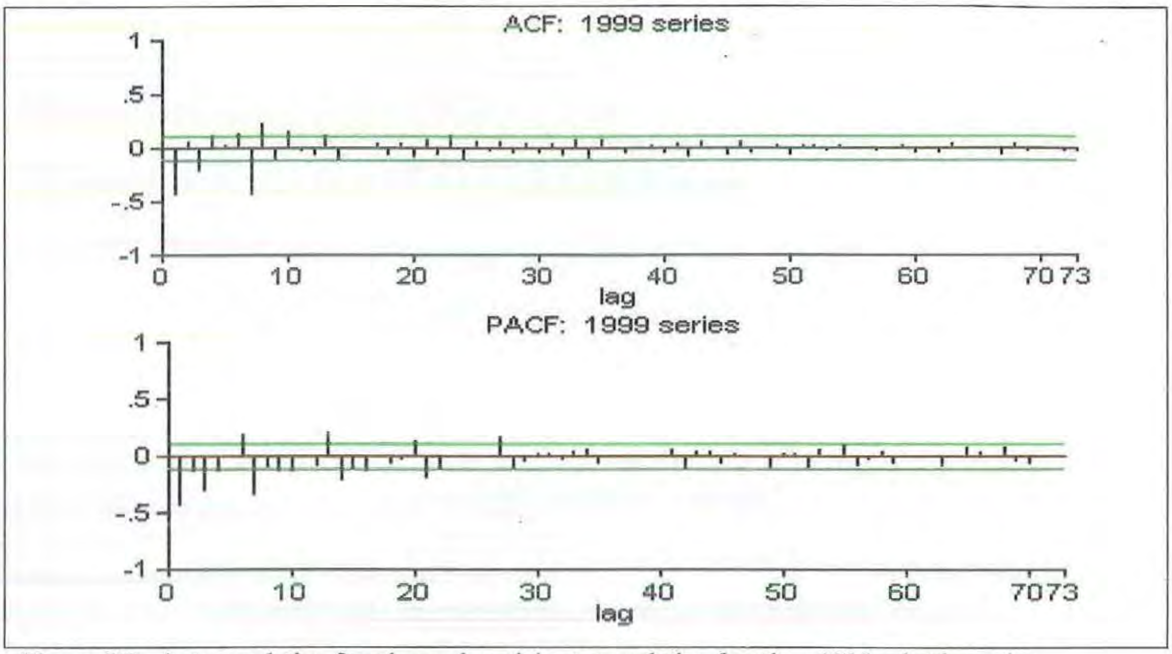


Figure 5.7: Autocorrelation function and partial autocorrelation function; 1999 calendar series

The ASTSA program was used to analye each of the time series in this study. Once stationarity was achieved, the correlograms representing the ACF and PACF for each time series were attained. The correlograms enable identification of seasonality and therefore appropriate parameter values for each model. Figure 5.7 shows the correlograms for the 1999 total presentation series. This time frame proved inappropriate for this model type. Unlike the Holt-Winters method that still produced a "good" model despite the stated difficulties, the Box-Jenkins method did not produce anything approaching a good model. As indicated, the month of January does have a number of public holidays and the New Year holiday period in 1999 and 2000 was given "special" significance. Higher presentations were seen at these times that may have caused modelling problems with the months of January and February.

Lag	12	24	36	48
Chi-Square	11.4	19.7	27.8	52.9
DF	6	18	30	42
P-Value	0.076	0.351	0.581	0.121

Table 5.3: Modified Box-Pierce (Ljung-Box) Chi-Square statistic

In an attempt to resolve this problem, differing 12-month periods were tested. After much experimentation, the period May to April proved to be the most successful. The resulting correlograms, Figure 5.8, are more easily interpreted. This time of the year is devoid of public or school holidays, a fact that may have led to the improved ability to model the series with this technique.

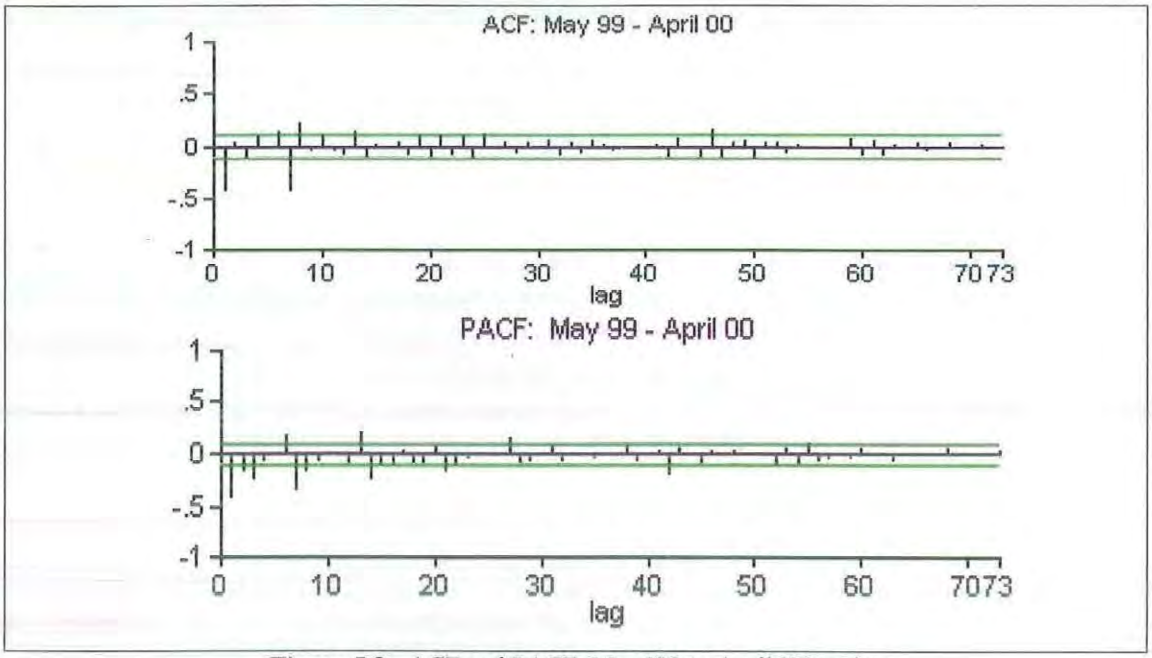


Figure 5.8: ACF and PACF; May '99 to April '00 series

The final stage of modelling can in some circumstances come down to subjective modelling. The overall aim in any modelling is to attain the model that provides the best estimate of the data. The ASTSA program incorporates an ARIMA search function that removes a significant portion of the guesswork. It provides an AIC value for each permutation for the given parameters. The best 2 or 3 models were then identified and their parameter values used to develop comparative models using the Minitab program. The results for the May 99 to April 00 series are given in Table 5.3. A model is deemed to be a good fit when the Ljung-Box statistics have p values greater than 0.05 at each lag. Once appropriate parameters have been identified, a SARIMA model and mathematical prediction equation is developed using the Minitab and ASTSA program.

The final estimation parameters developed for the 1999 complete data series are:

AR (1)	0.2446	MA (1)	0.8400
SAR (1)	- 0.0299	SMA (1)	0.9598
SAR (2)	- 0.1327		

producing the mathematical prediction model:

$$(1-0.24\boldsymbol{B})(1+0.03\boldsymbol{B}^{7}+0.13\boldsymbol{B}^{14})\nabla_{1}\nabla_{14}\boldsymbol{X}_{t}=(1-0.84\boldsymbol{B})(1-0.96\boldsymbol{B}^{7})\boldsymbol{W}_{t}$$

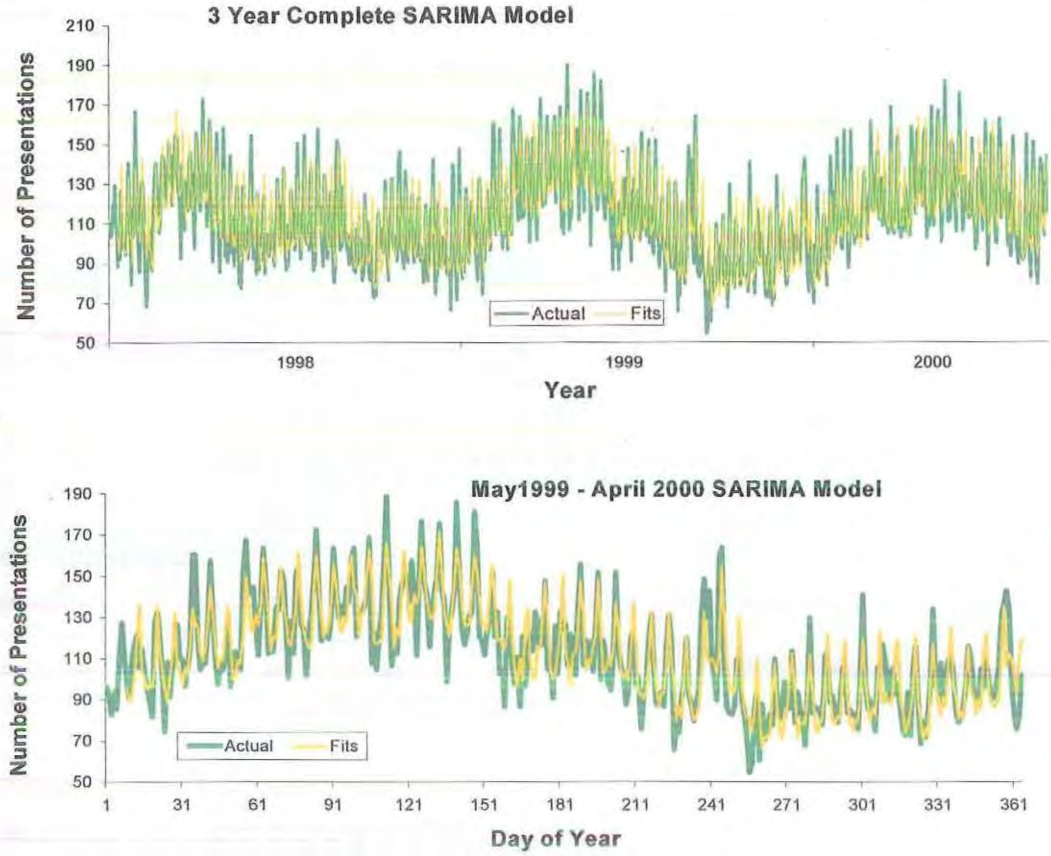
and the final estimation parameters developed for the 3 year complete data series are:

AR (1)	0.2597	MA (1)	0.8732
SAR (1)	- 0.0469	SMA (1)	0.9778
SAR (2)	- 0.0734		

producing the mathematical prediction model:

$$(1-0.26\boldsymbol{B})(1+0.05\boldsymbol{B}^{7}+0.07\boldsymbol{B}^{14})\nabla_{1}\nabla_{14}\boldsymbol{X}_{t}=(1-0.87\boldsymbol{B})(1-0.98\boldsymbol{B}^{7})\boldsymbol{W}_{t}$$

Appendix D contains the analysis for each Box-Jenkins model. The parameters as indicated above for each these models were noted to be similar. The exceptions found were for the 1999 and 2000 triage category 4 models, each requiring a significant AR (1) parameter value. The two 1999 models were noted to need a second seasonal parameter. Figure 5.9 displays the prediction model versus the actual presentations for the complete 3-years of data and the 1999 series. Once again, each model is able to replicate in general the two underlying cycles within the data. However, similar to the Holt-Winters method, the predictive models do not accurately fit the time series when consecutive daily presentations differ significantly, but unlike a Holt-Winters model, it "recovers" quickly. The residual analysis of all the models produces a similar result to the Holt-Winters, with each analysis indicating an appropriate model, despite the presence of outlying values in each residual histogram. The moving average (MA) component of a Box-Jenkins model expresses the residual error term Z_t as a finite weighted sum of the previous errors plus a random component. For any given model this is restricted to the current error term and q weighted error terms in the past.



(a)

(b)

Figure 5.9: Box-Jenkins Model: (a) 3-year data series. (b) 1999 data series.

From personal experience it could be argued that these peaks and troughs would be the result of the one-off factors like public holidays, sporting events or weather conditions previously discussed. There is often a noticeable decline in presentations during events like the Australia Day fireworks display. However, these types of impacts are not easily identified with the data set used for this thesis.

5.6 HOLT-WINTERS V BOX-JENKINS

These two models are both suited to the type of time series the ED data represents. The presence of the two seasons within the data are identifiable and in the case of the Box-Jenkins able to be incorporated within a given model. The analysis has shown that both provide models that are a good fit to the data and able to provide adequate forecasts though each model's fit is somewhat conservative. Figures 5.10(a) and 5.10(b) indicate

that the two models are very similar. Appendix D shows that these two models are representative of the Box-Jenkins models developed for each series. In particular, those developed for the triage 4 category, in particular would not be seen as strong. Neither technique is able to provide an accurate model when the daily numbers are significantly different to a previous day's numbers.

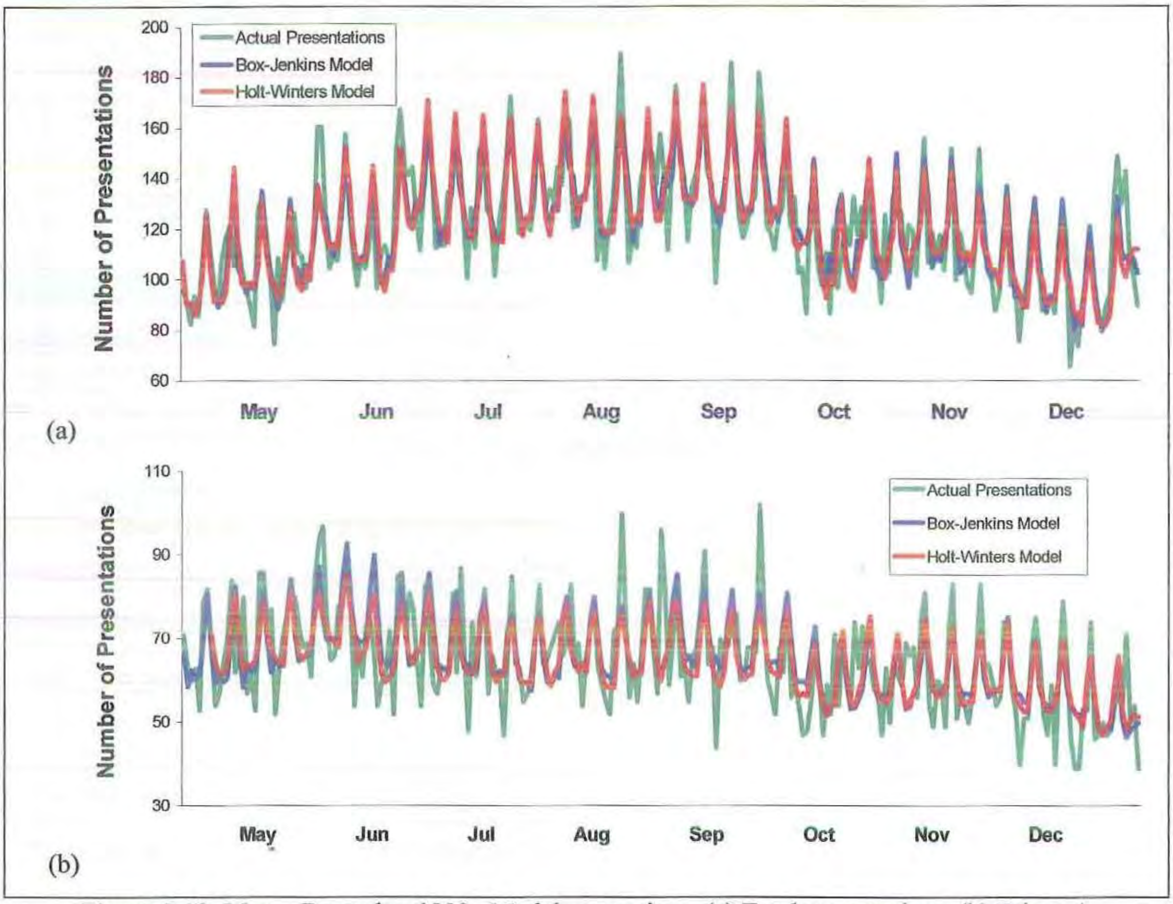


Figure 5.10: May – December 1999. Model comparison: (a) Total presentations, (b) Triage 4.

These two modelling techniques can also be compared through use of either the Mean Absolute Difference (MAD) or Mean Squared Error (MSE). Table 5.4 displays the MSE values for each of the 1999 and 3 year models constructed taken from the complete table in Appendix B. Clearly the Box-Jenkins method provides the better overall fit, and would be expected to provide the more accurate forecasts. This is displayed in Figure 5.12, with only 3 early actual presentation values falling beyond the confidence interval.

Model	Holt-Winters	Box-Jenkins
1999 – total presentations	203.08	177.78
3 Year	188.71	165.66
1999 – Category 4	96.3	70.44
3 Year	96.20	82.03

Table 5.4: Mean Squared Error (MSE) values for Holt-Winters & Box-Jenkins models

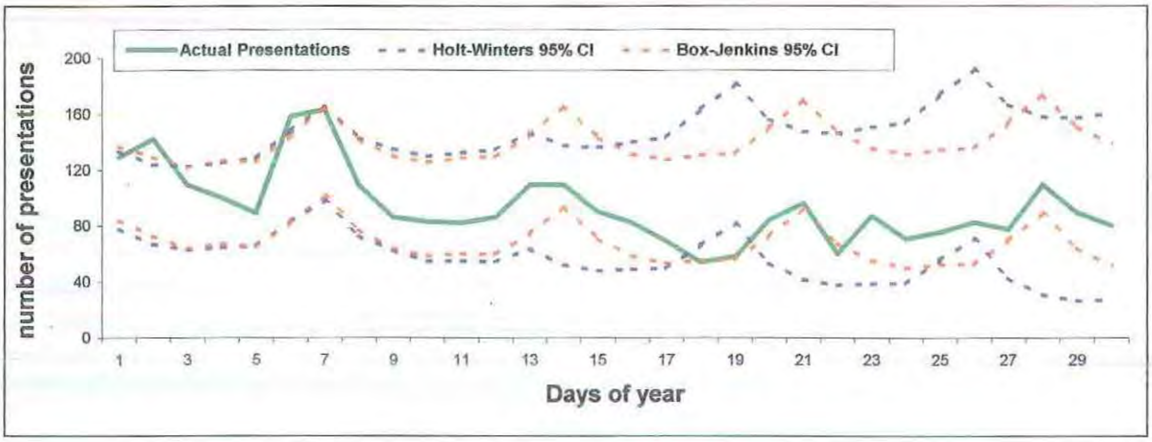


Figure 5.11: Comparison of 95% confidence intervals, for 30 forecasts using 1999 models

5.7 REGRESSION MODEL

The two modelling techniques used above are appropriate when a general modelling of seasonal data is required. Each clearly indicates that the ED series have both an annual and a 7-day cycle. However, neither of these techniques makes allowance for other factors that can influence a given data series. Regression analysis is a technique that models the relationship between variables and then provides a prediction model for a dependent variable. Figure 5.12 displays the presentations for the year 1999 by the day of the week. The 7-day cycle is clearly evident, with Sunday the busiest day and either Tuesday or Wednesday the quietest.

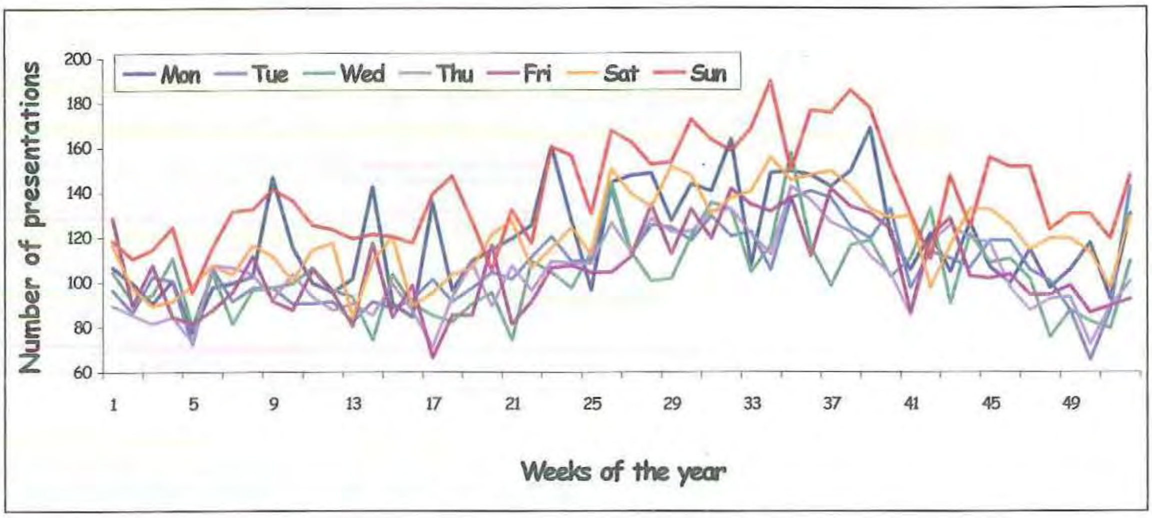


Figure 5.12: 1999 ED total daily presentations by day of week

A multiple regression model using binary indicators was established using both an Excel spreadsheet and Minitab. To avoid collinearity concerns, as discussed in Chapter 4, binary variables were not used for every day of the week. Wednesday was chosen as the day to exclude due to it consistently being the low point in the weekly presentation cycle. A binary variable was used to identify public holidays and associated weekends. The resulting prediction models, as displayed in Table 5.5, however becomes inadequate due to the induced trend from this type of regression model.

Actual	Forecast	Actual	Forecast	Actual	Forecast
106	150	114	132	74	115
88	127	131	151	90	116
88	117	118	128	99	132
94	114	66	118	120	151
87	115	83	114	93	128

Table 5.5: Forecasts for December 1999 using regression model with day variables.

The trend evident in Figure 5.13 is due to the time variable included in this type of model. Appendix E contains the initial model for each series. Clearly as the ED series does not continually increase, this type of regression model fails.

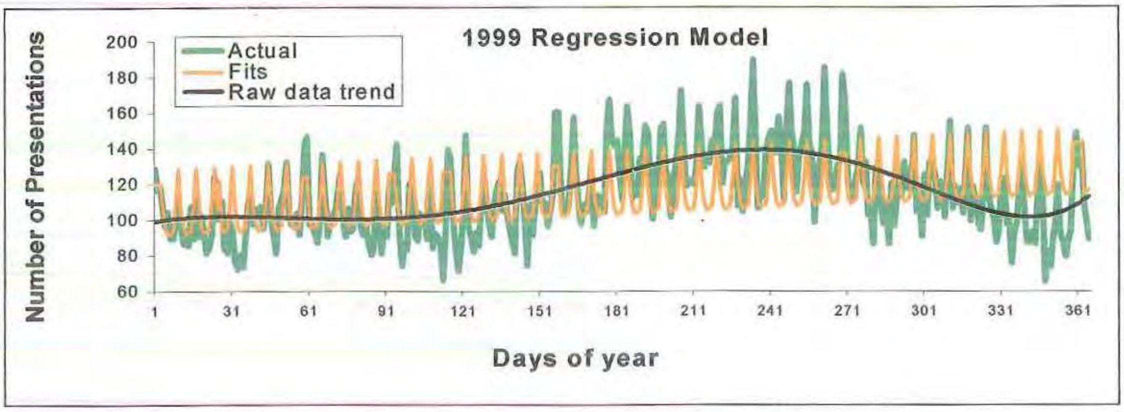


Figure 5.13: 1999 regression model using weekday variables

This particular model would need to be restricted to a shorter time span with a positive trend component. However the trend line of the data, a 6^{th} polynomial, shown in Figure 5.13 does indicates the possibility of expanding this regression model. While outside the initial scope of this thesis a brief analysis of the addition of $Time^2$, $Time^3$, ... , $Time^6$ was investigated. These regression models, shown in Appendix E, did prove impressive when compared to those above. Figure 5.14 clearly displays a far better fit that effectively follows the raw data trend line. It was of interest to note the adjusted R-squared value improved from 0.36 to 0.69, a significant improvement with obvious effect upon forecasting ability.

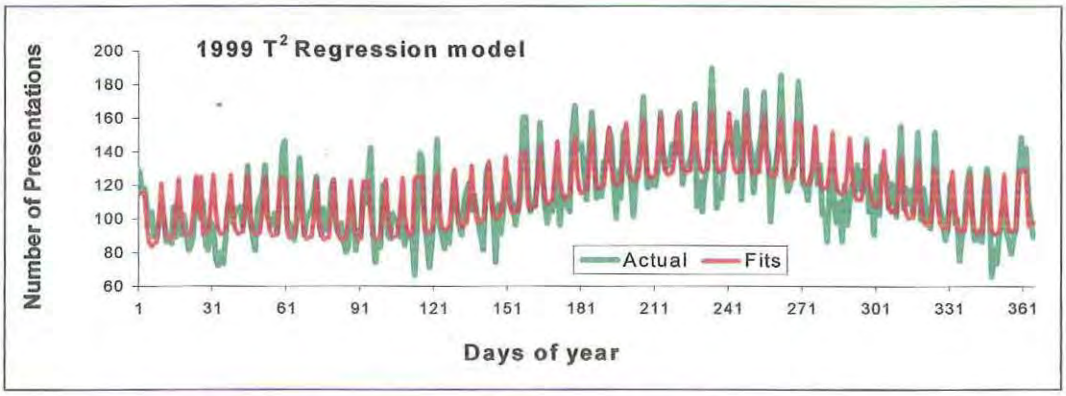


Figure 5.14: 1999 regression model using weekday variables + $Time^2$ variable

The use of the extra time variables did produce good models. However, this left the challenge to find a binary regression model that could incorporate a time series with the 2 known cycles. The annual cycle, as seen in Figure 5.15, shows the influence of each month of the year. The impact of season on presentation rates to emergency

departments is well documented. The obvious step was to add a variable for each month of the year. However, this immediately highlights collinearity problems and the inherent modelling conflicts in using such a model. Each data value would now have two binary variables – one for day of the week or holiday and one for the month of the year. Despite these concerns, it was felt that the inclusion of the extra variables resulted in more appropriate modelling of the ED data series.

The last factor to examine for this study was the impact upon presentation rates of the daily weather. The impact of severe weather conditions and environmental hazards, like pollen or smoke, are well documented. However the effect of the daily variations in weather conditions has only occasionally be examined. Many workers within emergency departments view a day of rain as a blessing due to the perception that fewer people will present with what are loosely termed "minor" problems. Similarly, a very cold or hot day appears to reduce the numbers seen on that particular day, but increase the numbers when weather conditions improve. The weather variables considered in this study; total daily rainfall, maximum and minimum temperature, form three time series. These variables are not easy to include within a binary regression model. While rainfall could be recorded as a yes or no, it is the amount of rain and the actual temperatures that are considered to impact upon presentation rates. The decision was taken to include the actual raw data value in the model development.

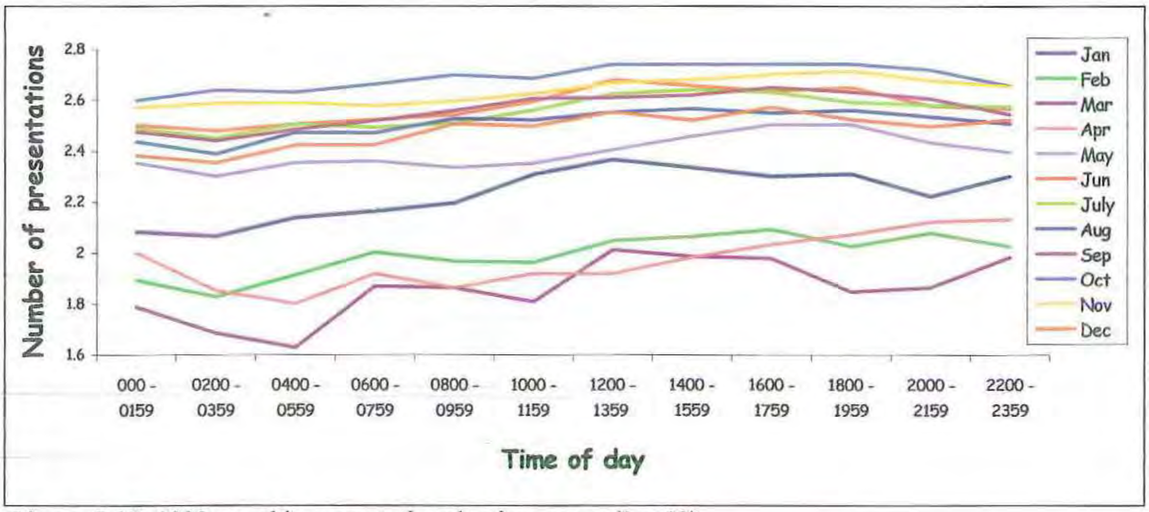


Figure 5.15: 1999 monthly presentations by time group (Log10).

The cross correlation function (CCF) is a method for estimating the degree to which two time series are correlated. Using the CCF, these 3 series were compared with both the total daily presentation and triage category 4 series for each year and the total 3-year period. The CCF analysis was undertaken by two different methods. The initial analysis compared the raw time series of a weather variable with that of an ED series. Next, an ARIMA model was fitted to both the ED and weather variable series with a CCF analysis being performed on the two residuals time series. The overall result of this analysis failed to identify any correlation between the ED series and the weather variables. As a worker within this area this was a surprise finding, but may indicate what is commonly believed is just a belief. However, as both Christoffel (1984) and Attia (1996) indicated in their research, the impact of the weather conditions may be more significant when the data is analysed over shorter time intervals within each day.

Vars	R-Sq	Adj. R-Sq	C-p	s	Time	Mon	Sat	Sun	Holiday	Apr	Jun	Jul	Aug	Sep	Dec	Maximum	Rainfall
8	64.7	63.9	78.1	13.841		X	X	X	X		X	X	X	X			
8	64.2	63.4	83.5	13.929	X		X	X	X		X	X	X	X			
9	67.1	66.3	50.3	13.366	X	X	X	X	X		X	X	X	X			
9	66.9	66.0	53.5	13.420	X		X	X	X		X	X	X	X	X		
10	69.6	68.8	22.5	12.870	X	X	X	X	X		X	X	X	X	X		
10	67.6	66.7	47.0	13.295	X	X	X	X	X	X	X	X	X	X			
11	70.2	69.3	17.1	12.757	X	X	X	X	X	X	X	X	X	X	X		
11	70.0	69.1	20.0	12.808	X	X	X	X	X		X	X	X	X	X	X	
12	70.6	69.6	14.5	12.693	X	X	X	X	X	X	X	X	X	X	X	X	
12	70.6	69.6	14.7	12.697	X	X	X	X	X	X	X	X	X	X	X		X
13	70.8	69.7	14.0	12.666	X	X	X	X	X	X	X	X	X	X	X	X	X

Figure 5.16: Best Subsets output for 1999 ED series

This development phase resulted in a model that potentially had 24 explanatory variables. Although this would obviously increase the R^2 value it would make such a model far too complex and costly to implement and its usefulness would be justifiably questioned. A simple model is usually the better model. The Best Subsets regression technique was used to determine the importance of each of these variables with the result that only 14 are significant. Although CCF analysis did not indicate the three weather variables had any significant impact upon presentation rates, they were included in the final Best Subsets analysis. The model with the smallest Mallows C-p statistic for each calendar year was then chosen. The 1999 total presentation model,

Figure 5.16, included the maximum temperature and the rainfall figure. However as the R^2 adjusted statistic improved only 0.4% with both these variables included, an easy argument could be made for their removal. This type of reasoning is not uncommon with regression modelling.

The final prediction model chosen for the complete 1999 series was

80.7 + (0.0638 Time) + (11.3 Mon) + (15.7 Sat) + (33.9 Sun) + (31.6 Holiday) - (7.19 Apr) + (17.3 Jun) + (25.0 Jul) + (29.4 Aug) + (28.3 Sep) – (19.4 Dec) + (0.276 Maximum) – (0.203 Rainfall)

and for the 3 year complete triage series:

93.07 + (0.0087 Time) + (9.93 Mon) + (10.55 Tue) + (21.10 Thu) + (15.79 Fri) + (12.45 Sat) + (29.45 Sun) + (20.76 Holiday) - (11.69 Apr) - (7.71 Jun) + (9.70 Jul) + (24.96 Aug) + (20.43 Sep) – (0.35 Rainfall)

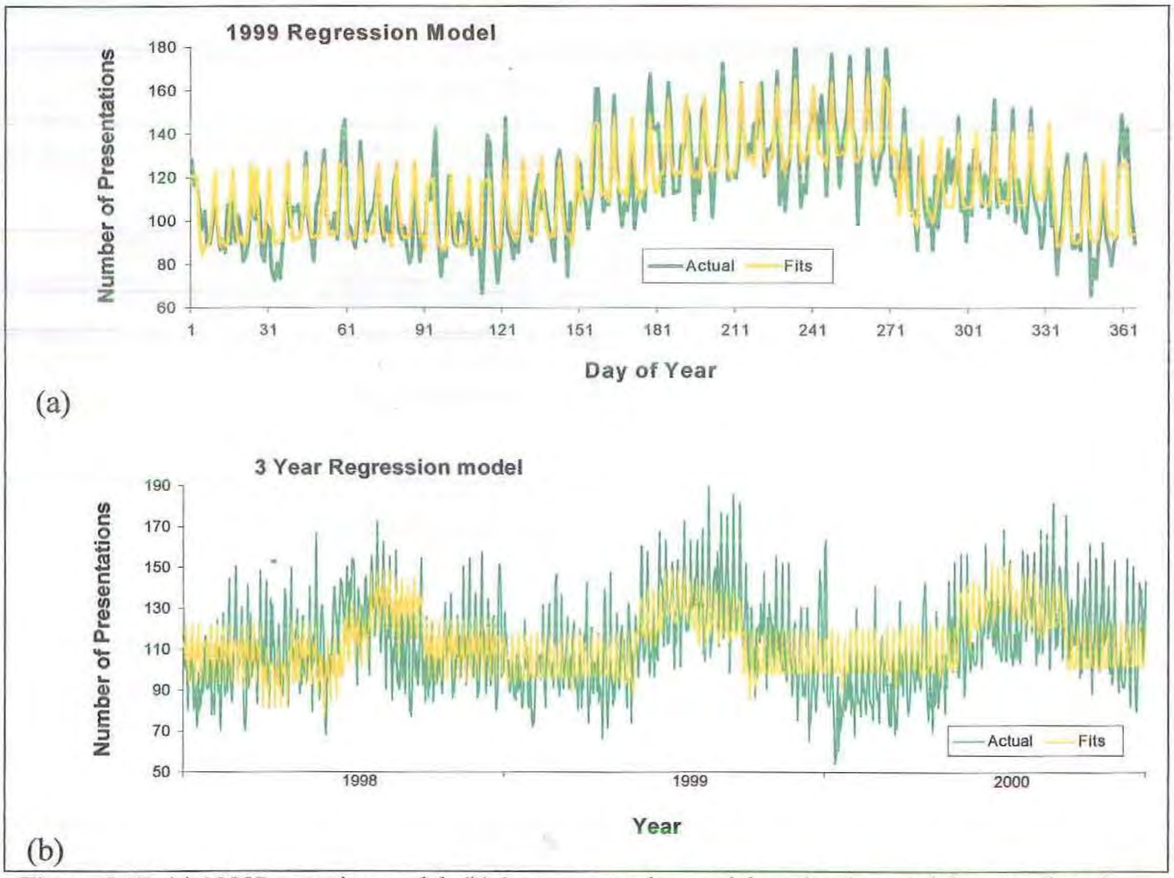


Figure 5.17: (a) 1999Regression model; (b) 3-year regression model; each using weekday, month and weather variables

The revised regression models in Appendix E are a significantly better fit than the models that ignored the annual impact. Figure 5.17, shows the marked improvement for the 1999 total presentation model. The use of the month variables has enabled the

annual trend to be modelled together with the weekly cycle. By comparison the model the complete 3 years is a relatively poor fit. The R-squared adjusted value is only 0.38 compared to the 1999 model at 0.70. The 3 year model in particular clearly indicates this modelling technique is probably not suited to the long-term data series. Table 5.6 provides the first 20 forecasts for the year 2000 using the 1999 model. The very large presentations, highlighted, on the New Year's Day weekend have influenced the forecast values. It's notable that the regression model did predict this sudden increase. However, the forecasts beyond the first 7 days must be considered inaccurate, with the majority of actual values falling outside the 95% confidence intervals. These forecasts were significantly higher than reality and were found to be representative of those given by each regression model.

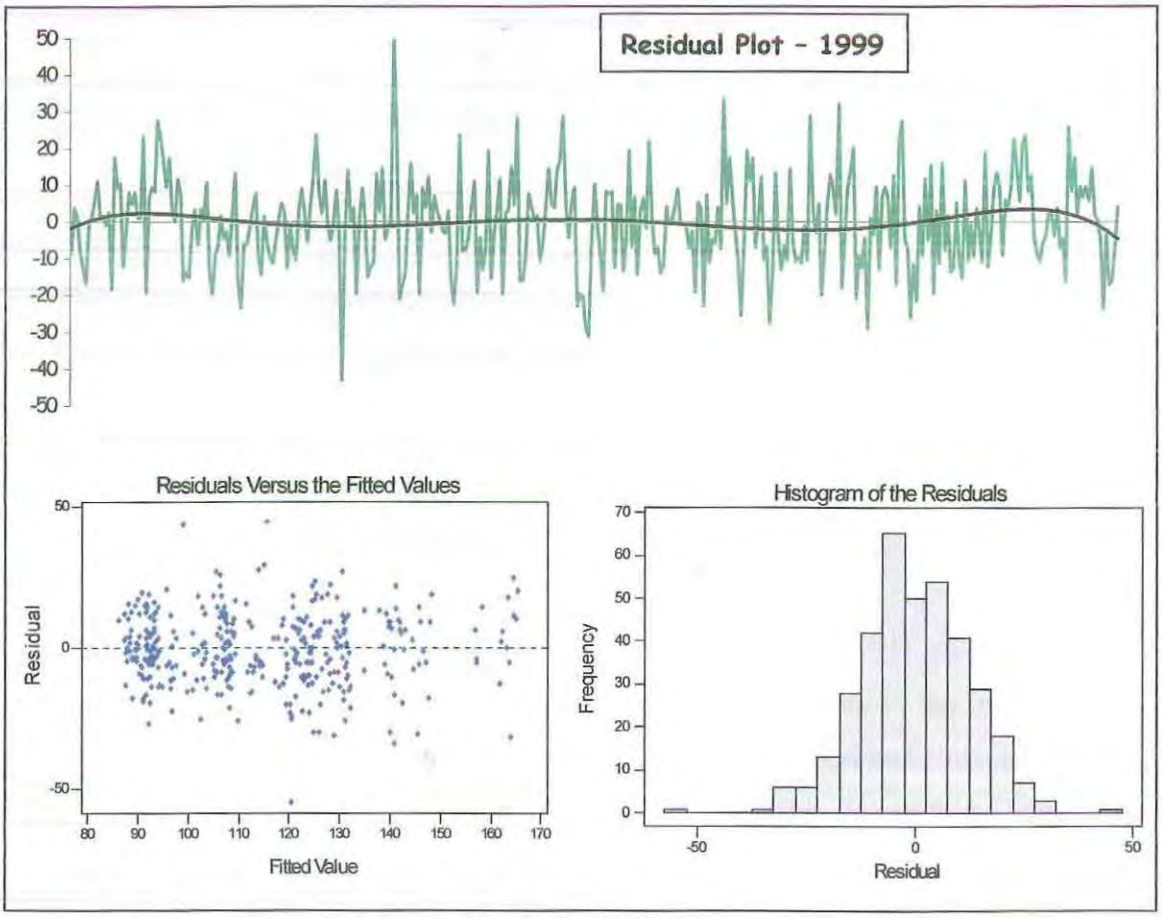


Figure 5.18: Residual Analysis: 1999 Total presentation model

The residual analysis for each model, included with the models in Appendix E, supports this with the residuals approximating a normal distribution. However, as shown in Figure 5.18, the histogram indicates the presence of some outlying values. The dot plot

of residuals versus fitted values does not indicate a widening as data values increase. This indicates these outlying values are not significant. The time series plot of the residuals does not represent "white noise" and would not be considered stationary.

Time Period	Actual	Forecast	95% CI Lower	Upper	Time Period	Actual	Forecast	95% CI Lower	Upper
361	**130**	127	85.0	169.1	371	**83**	114	85.5	142.7
362	**143**	126	85.2	167.1	372	**87**	114	85.6	142.2
363	**110**	94	59.3	128.9	373	**110**	129	97.4	160.7
364	**101**	94	59.4	128.4	374	**110**	147	115.6	179.3
365	**90**	94	59.4	129.3	375	**91**	124	93.3	154.7
366	**159**	146	111.3	180.1	376	**83**	112	86.2	138.4
367	**164**	146	111.3	180.6	377	**70**	112	85.9	138.2
368	**110**	146	111.4	180.2	378	**55**	113	86.1	140.4
369	**87**	115	85.4	143.7	379	**59**	111	86.5	136.4
370	**84**	115	85.4	143.7	380	**85**	125	92.2	157.3

Table 5.6: Forecasted values using 1999 model.

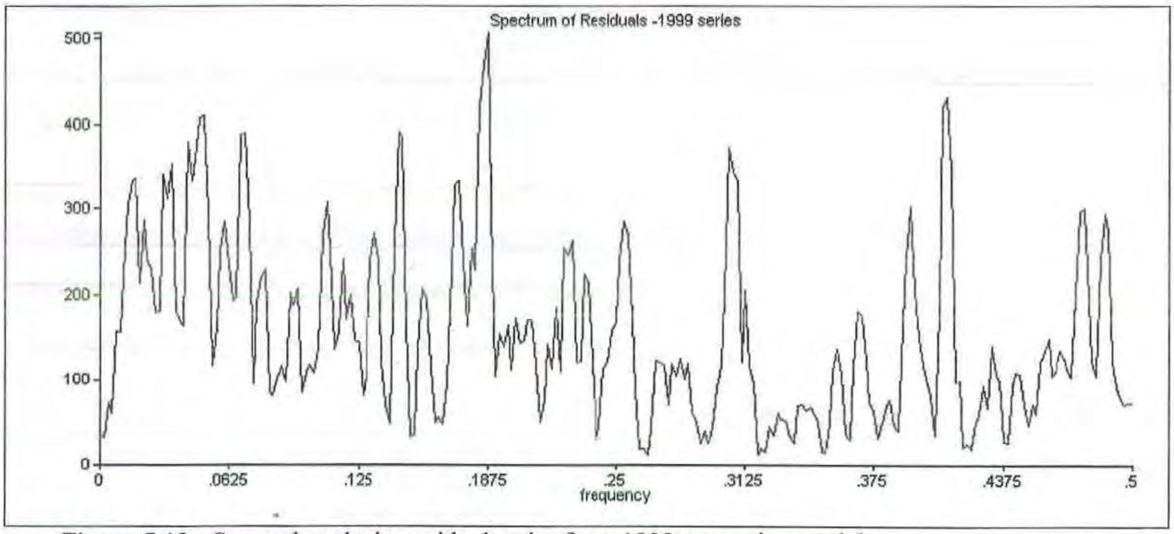


Figure 5.19: Spectral analysis; residual series from 1999 regression model

This non-stationarity of the residual plots from the regression models led to discussions about expanding the regression model. Consideration was given to the inclusion of a dynamic model to explain any pattern remaining within the residual time series. The choice was to apply a Box-Jenkins ARIMA model to the residuals and then combine the forecasts from this model with the original regression forecasts. This choice proved difficult to implement with any confidence. While the residuals from each regression model were not strictly stationary, there did not appear to be any easily identifiable cycle or trend present within these residual series. Figure 5.19 displays the spectral analysis of the 1999 residual series. A peak is noted at a frequency 0.1875 that equates to a cycle of approximately 5 days. However the removal of such a cycle would be

inappropriate, due to the multitude of other peaks within the series. This was not the expected outcome given the results from the regression models previously. The ACF and PACF for the 1999 residual series are displayed in Figure 5.20. Neither correlogram indicated the presence of an underlying trend therefore an ARMA model was applied to the majority of residual time series. The ASTSA and Minitab programs were utilised to determine the final parameters for each model.

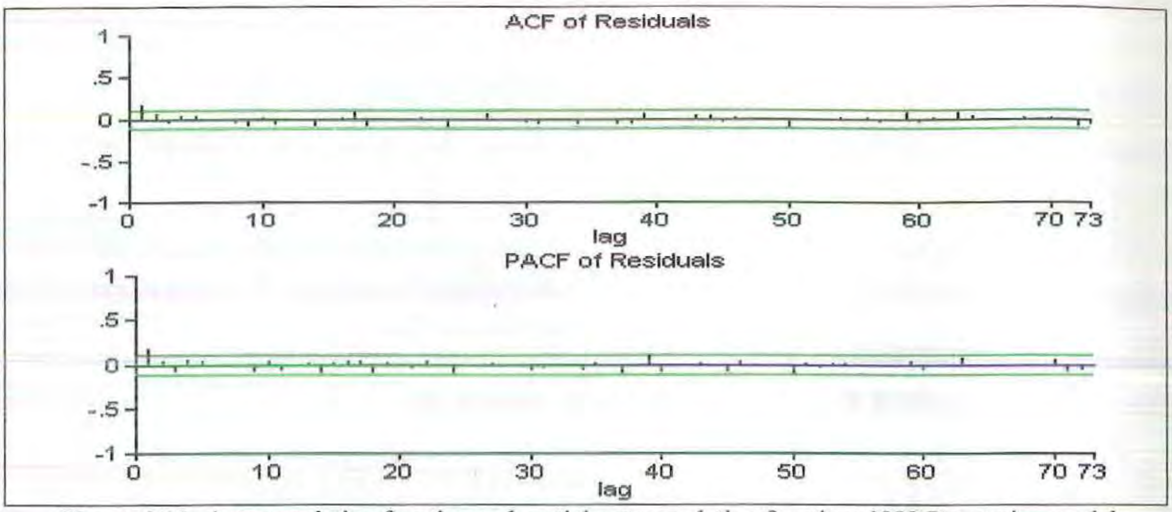


Figure 5.20: Autocorrelation function and partial autocorrelation function; 1999 Regression model residual series

While both correlograms indicated an AR and MA parameter were likely, the simpler AR(1) model provided the stronger AIC and Box-Pierce statistics, as shown in Table 5.7. Models of this nature are a form of regression where forecasts are a function of previous values plus an error component. Figure 5.21 indicates that the AR (1) model could be seen as a poor fit to the underlying residual model, but must be considered to have removed the remaining non-randomness from the series. This is typical of the resulting models given in Appendix E for the regression residuals. However these models provide minimal extra information to the original regression forecasts. After the initial few forecasts there is no apparent improvement to the original regression models.

ASTSA Statistics

p	d	q	P	D	Q	s	AIC	
0	0	0	0	0	0	0	6.0549	
0	0	1	0	0	0	0	6.0322	
0	1	0	0	0	0	0	6.5514	
0	1	1	0	0	0	0	6.1026	
1	0	0	0	0	0	0	6.0297	best model
1	0	1	0	0	0	0	6.0342	
1	1	0	0	0	0	0	6.3439	
1	1	1	0	0	0	0	6.063	

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	12.8	29.6	38.3	52.1
DF	10	22	34	46
P-Value	0.236	0.129	0.281	0.248

Table 5.7 Forecasted values using 1999 model

The exception to this general conclusion concerns the 3 year models. The regression models, as indicated in Figure 5.17 (b) were a poor fit to the underlying series. The modelling of these residuals, Figure 5.22, resulted in SARIMA models being fitted. These SARIMA models provided an improved fit to the underlying ED series with the corresponding better forecasting ability.

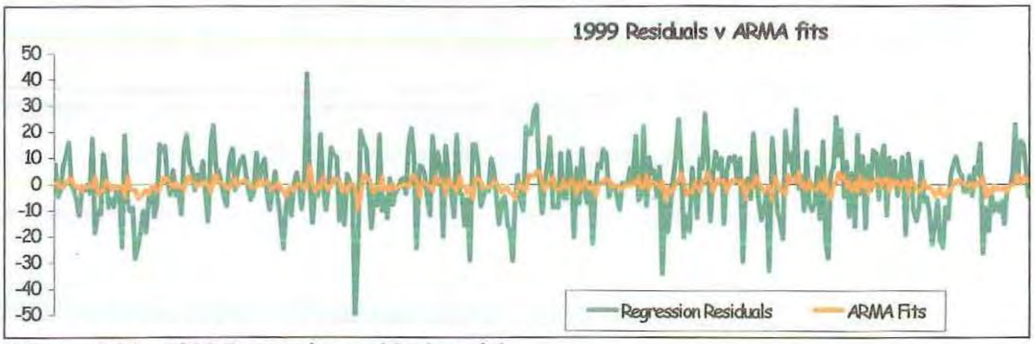


Figure 5.21: 1999 Regression residual model.

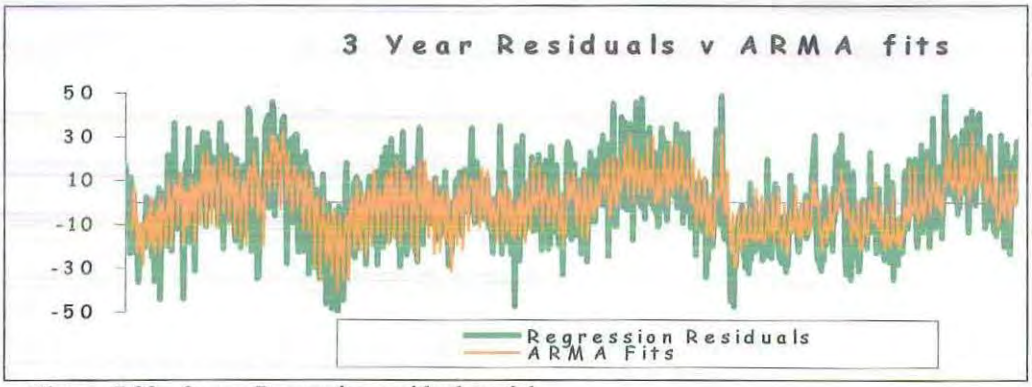


Figure 5.22: 3 year Regression residual model.

6 RESIDUAL ANALYSIS

6.1 ABOUT THIS CHAPTER

Residual analysis is a common statistical method employed to check the viability of a time series model, once the model has been developed. It is a method applicable when knowledge of the kind of discrepancies expected is unavailable (Box and Jenkins, 1970). It could be argued, that virtually all statistical modelling can be defined by the equation:

Residual = Observed value – fitted value.

The fitted value attempts to explain the general pattern of a time series while the residuals record how much the observed or actual value deviates from this general pattern. Models with residual values large in magnitude represent a poor explanation of the underlying time series. A benefit of analysing the residuals from a time series is that they are time ordered, and as such can be treated like any time series data. These residuals or errors (Z_t) are assumed to represent independent random variables that fit a Normal Distribution, N $(0,\sigma)$. This chapter examines the methods used to check each model used for this thesis and the results of the residual analysis.

6.2 THE METHODS USED

An obvious first step in the analysis is to plot the residuals time series. A "good" model can be expected to display a random fluctuation and one that is close to zero. Non random patterns are indicative of a poor model that needs review. Box and Jenkins (1970) clearly indicate that the visual inspection of a plot of the residuals is an indispensable first step in the analysis process. This plot will reveal the presence of outlying values and any cyclic effects or autocorrelations. A second plot, residuals versus fits, can similarly be used. A correlogram of the residuals, however, enables any such cyclic behaviour to be further examined. The correlogram of a "good" model

should approximate that of white noise. This type of residual analysis is appropriate when modelling with Box-Jenkins ARIMA models. Two other graphical techniques that can be used to determine the goodness of fit for a given model are the histogram and Q-Q plot. The Normal Quantile-Quantile or Q-Q plot enables the quantiles of the residuals to be plotted against the corresponding quantiles of a normal distribution. A near straight-line relationship on the Q-Q plot indicates the residuals have a normal distribution and the model a good fit. Similarly, the histogram should approximate the bell shape of the normal distribution.

6.3 THE RESIDUAL FINDINGS

The residual plot, as the first step in the analysis, provides a quick and effective method of checking a model's viability. Figure 6.1 displays the residual plots for the 1999 and 3 year total triage series by each of the four modelling techniques. The Holt-Winters method clearly displays an initial under-fitting. This was apparent in the majority of the models developed by this technique. The high numbers of presentations experienced in early January appear to significantly impact upon subsequent forecasts until this impact is reduced. The majority of the residual plots provide evidence that the chosen models are appropriate. The exception is those for the regression models, most displaying a non-random pattern still evident within the residual series. The 3 year residual plot in Figure 6.1 clearly displays such a pattern. However, the application of an ARMA model to these regression residuals appears to have resulted in appropriate, random residuals.

The ACF and PACF correlograms of the residuals provide an alternative check for the Box-Jenkins developed models. Those displayed in figure 6.2 indicate that the final model choice for the 1999 and 3 year total presentation models is appropriate with no significant lags. While this was the finding with the final choice of model for each of the 8 series examined, the histograms of most did indicate the presence of outlying values.

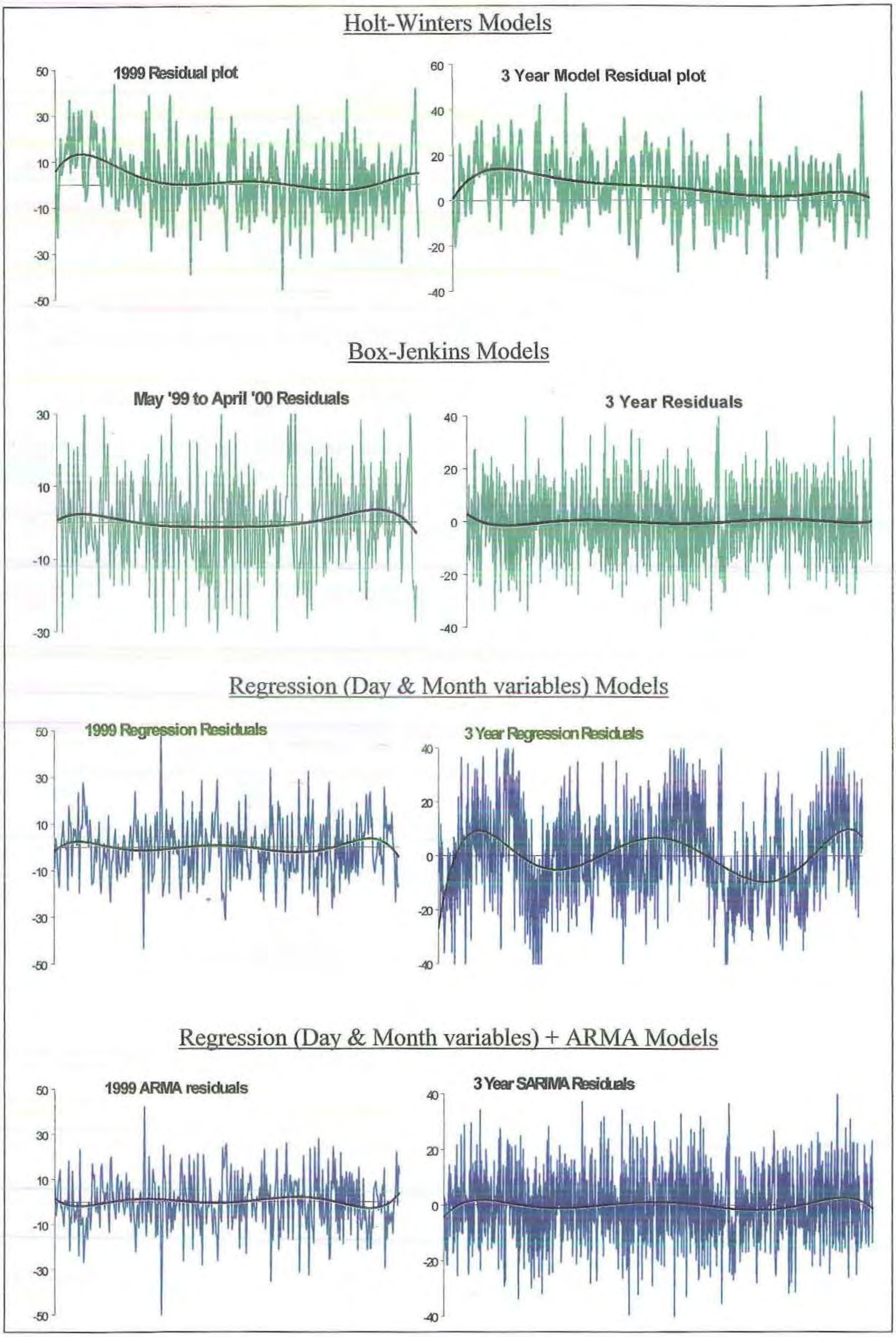


Figure 6.1: Residual plots for each of the 4 models developed; 1999 & 3 year Total presentation Models

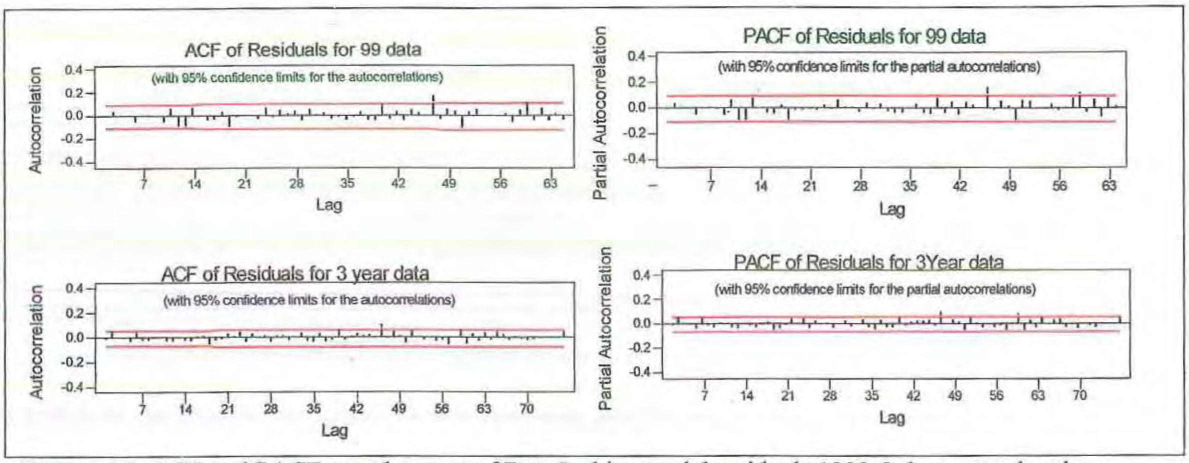


Figure 6.2: ACF and PACF correlograms of Box-Jenkins model residuals 1999 & 3 year total series.

The histogram, similar to the residual plot, provides a pictorial indication of the spread of the residuals for a given time series. Figure 6.3 displays the typical histogram for the models presented in this thesis. While each histogram approaches that expected for a normal distribution, there are clearly outlying data points in each histogram.

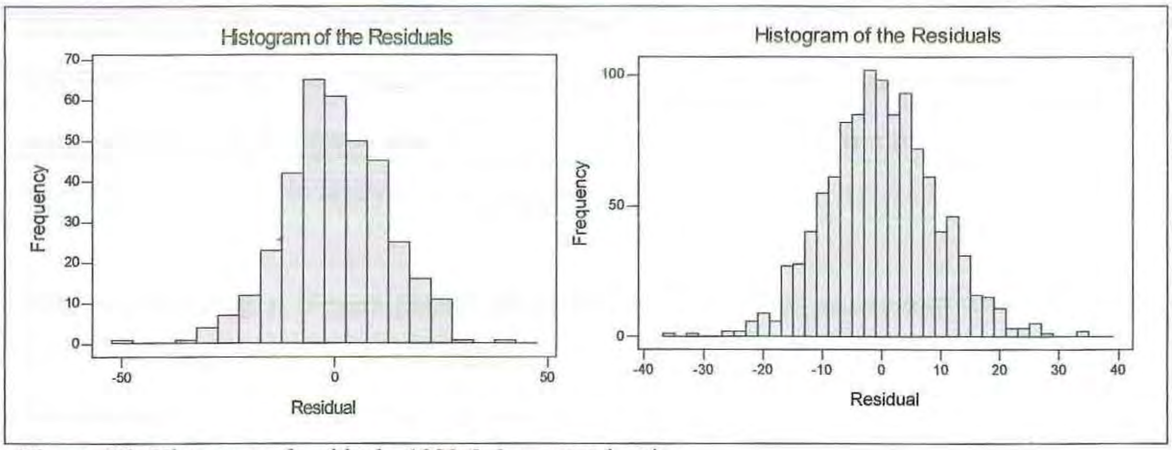


Figure 6.3: Histogram of residuals, 1999 & 3 year total series

The models presented in this thesis have each been examined by an appropriate residual analysis method applicable to the modelling method used. While the Holt-Winters model does have initial problems, all the developed models can be taken to provide an adequate representation of the behaviour of the dependant variable.

7 DISCUSSION AND CONCLUSION

7.1 ABOUT THIS CHAPTER

The aim of this thesis was to use time series analysis techniques and regression analysis in an attempt to model the presentation rates to the emergency department (ED) of Princess Margaret Hospital for Children (PMH). The resulting mathematical prediction models could be applied to provide short-term forecasts of future patient numbers. This study is viewed as an initial one in an area where little study has occurred, which may lead to more advanced analysis in the future.

7.2 SUMMARY OF ANALYSIS

The ideas behind this study came from employment within the emergency department of a busy metropolitan hospital. Over a period of time you realise that the number of patients seen each day varies, that these patients present at varying rates throughout the day and that springtime usually has higher presentation rates than autumn. However each of these impressions had little factual analysis to back it up. We decided to undertake this study to analyse these long held beliefs.

Data representing a 3-year period was obtained from the emergency department at PMH. A time series over 100 data points is sufficient to enable appropriate model development. The data was categorised by the triage code, a number system representing a patient's level of acuity. It was at this early stage of analysis that the patients represented by triage category 4 were noted to represent up to 80% of a day's total presentations. We decided to develop separate models for this group, to determine if the presentation rates matched those of the total population. Prior to beginning analysis the data was divided by calendar year, into weekly and monthly total presentation numbers, and combined to represent the complete 3-year time series.

Initial analysis centred on determining the length of cycle(s) within the ED data series. Both time series graphs and spectral analysis confirmed the presence of two cycles, the first one week in length and the second an annual cycle. The shorter 7 day cycle could not be regarded as either additive or multiplicative, but perhaps a combination of both. The pattern could best be described as irregularly regular. During the subsequent analysis this led to models of a questionable nature.

The first models developed used the dynamic techniques of Holt-Winters and Box-Jenkins, each of which was applied to the various time series. These techniques are appropriate to use on time series that contain seasonal data. The Holt-Winters three smoothing parameters were initialised using the Mean Squared Error (MSE). The Box-Jenkins models were developed through use of the ASTSA and Minitab programs, together with appropriate analysis of the correlograms for each series to determine the parameter values required.

These two models clearly identified that the ED time series could be modelled, but the results obtained were far from perfect. Neither series is able to account for factors that may influence the series. As stated before, the time of the year, prevailing weather conditions, and day of the week are all considered to impact upon the number of patients presenting each day. Regression analysis was used to investigate if any of these perceived factors had an influence. Binary regression was used to enable the isolation of factors. However, this immediately brought about conflict in model design. While modelling of the dominant 7-day cycle could be attained, the resulting models proved inappropriate due to the presence of the annual cycle. One solution to this problem was to introduce a binary variable for the months of the year. This immediately results in collinearity, correlation between explanatory variables of a model. However, after discussions, it was felt that despite this obvious problem, the resulting regression models better reflected each time series.

The impact of a day's weather upon presentation rates remains a debatable point. The inclusion of such variables introduces further collinearity into the models. Use of the Cross Correlation Function (CCF) did not indicate any such correlation. Severe weather is well known to impact on presentation rates, though its effect is to delay

rather than stop presentation. The impact of day-to-day variations in temperature and rainfall is not so clearly defined. Emergency department workers commonly believe that a day's weather condition does affect presentation rates. In the end, we chose to include these variables in the model development. The method of inclusion should perhaps be questioned. While the raw data value was chosen, it can easily be argued that for a true binary regression models the weather variables should be categorised to enable their inclusion as binary variables.

The final regression models were developed using best subsets regression to identify those variables of significance. These models provided significantly better fits to each series, though were still unable to replicate them to any great extent. While the inclusion of extra binary variables may have resulted in improved models this was considered inappropriate. Such inclusion leads to increased complexity for little statistical benefit. The majority of models returned R-squared adjusted percentages approaching 70%, thus accounting for the majority of presentation variation.

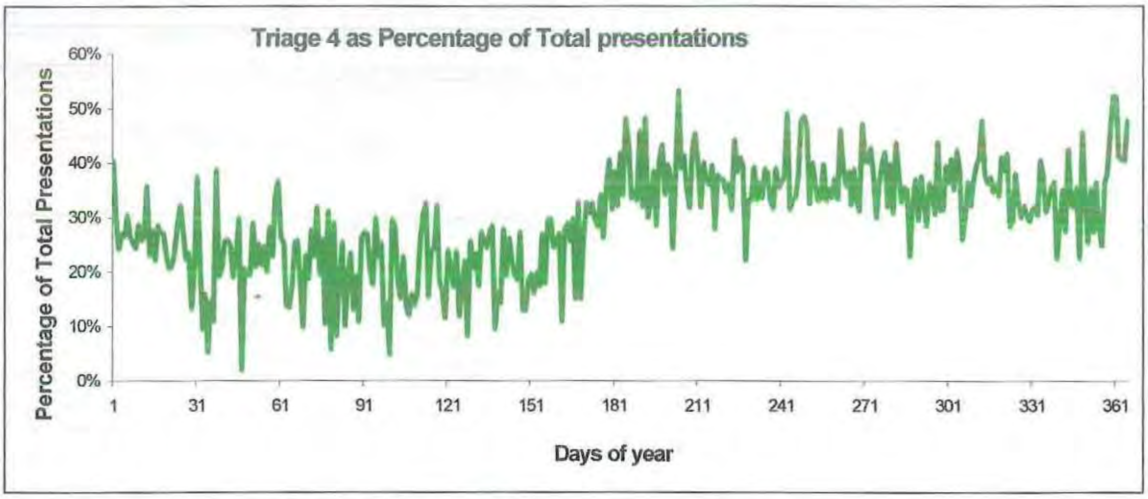


Figure 7.1: 1999 time series of triage category 4 as percentage of total presentations

The application of each of these methods to the triage category 4 time series did not prove overall successful. While this category does represent up to 80% of presentations it falls as low as 2% on occasion. Figure 7.1 clearly displays the variation in its percentage of total presentations. Therefore, despite initial thoughts, the use of this or other single triage categories is unlikely to provide any useful modelling information. The combination of such categories may prove more useful in predicting workloads.

All models were verified by use of the residual analysis techniques discussed in Chapter 6 of the thesis. Each of the final models developed did have residuals approaching that expected for a normal distribution. However the initial regression models, with explanatory variables only for day of the week must be seen as the weakest in predictive ability.

7.3 COMPARISON OF MODELS

The Box-Jenkins and Holt-Winters techniques both produce models suited to the time series of the ED data. Figure 5.10 showed that these models provide similar fits to the underlying time series, though each is of a conservative nature. The forecasts produced are affected by the underlying irregular pattern of presentations. The Box-Jenkins technique in particular needed to be applied to a different twelve month period to enable the development of usable models. Despite this initial difficulty, the resulting Box-Jenkins models were able to provide more accurate forecasts and 95% confidence intervals. Application of these techniques to the triage category 4 presentation time series produced similar results.

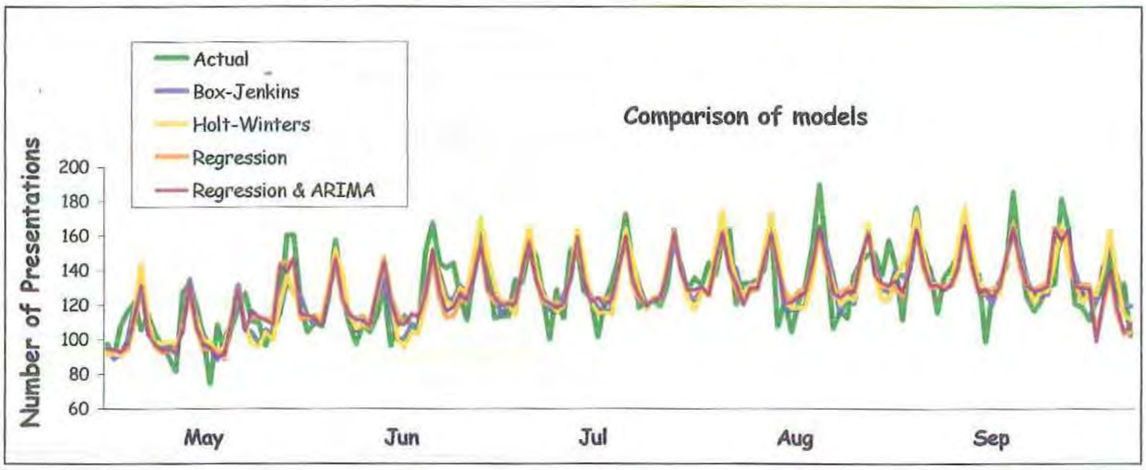


Figure 7.2: 1999 May to September complete series, comparing all four model fits

The regression model provides a method of incorporating those factors commonly believed to impact upon the presentation rates. Unlike the dynamic methods, which allow only for seasonal cycles, the regression model enables other influences upon a time series to be accommodated within a model. The initial binary regression models excluded variables for an annual cycle. The results were ineffectual models with an in-built trend component. These models were suited only to the shortest of time spans. The raw trend noted in Figure 5.13 led to the inclusion of extra time variables, $Time^2$ ... $Time^6$. While analysis indicated the need for only the $Time^2$, this type of regression model overcame the induced trend problem with the resulting models having significantly improved predictive power. These models were also freed of the time period constraints placed upon the initial regression models developed. However, they were not fully explored as considered outside the boundaries set for this thesis.

Subsequent regression modelling resulted in the inclusion of binary variables for the months of the year in an attempt to model the annual cycle. This produced better models than the initial ones and did represent the underlying series and also overcame the time induced trend problem. However, the inclusion of the months led to obvious concerns with collinearity. Similarly the inclusion of variables to model the impact of the weather does further complicate these concerns. The decision to use the weather variables in their raw data form could be easily questioned. The impact of a days weather conditions may be more definable if the data was divided into groupings. This would enable the binary regression technique to be applied in a more pure manner. Christoffel (1984) divided the weather data into 7 different categories - for the purpose of this study it was considered unnecessary.

A final adjustment to the regression models was the addition of a Box-Jenkins ARMA model to attempt to explain the residuals from each model. This was undertaken due to the apparent non-stationary nature of the residual time series from these regression models. This was very evident in both the 3 year time series. However, the results from this indicated that only minimal improvement occurred to each model, to the extent that the forecasts from the combined model were almost indistinguishable from those of the original regression forecasts.

Figure 7.2 displays the fit from each model for a 5 month period of the 1999 complete triage presentation series. All the models provide good fits, though none is able to provide an excellent fit to this series.

Model	Holt-Winters	Box-Jenkins	Regression (Day Variables)	Regression (Day & Time^2)	Regression (Day & Month)	Regression & Box-Jenkins
1998	198.64	163.55	224.45	173.11	156.60	137.69
1998 Category 4	111.54	96.46	112.07	87.26	90.50	95.54
1999	203.08	177.78	331.07	165.01	154.94	149.87
1999 Category 4	96.3	70.44	98.29	81.45	83.07	83.19
2000	209.72	156.03	232.49	153.65	157.28	143.45
2000 Category 4	84.09	78.57	74.93	61.85	89.69	89.85
3 Year	188.71	165.66	322.72	267.30	299.62	168.27
3 Year Category 4	96.2	82.03	116.32	95.34	123.63	85.92

Table 7.1: Mean Squared Error value for each model developed. Best in blue, worst in red.

The overall aim of any model comparison is to determine which model may be considered to have the best fit and therefore the better forecasting ability. Table 7.1 below display the MSE values for the models developed for this thesis. This value has the advantage of being relatively easier to use mathematically than other measures and can is used in statistical optimisation. It is not surprising to see from the table that the initial regression model, with day variables only, is by far the worst fitting model. This poor fit is further emphasised in Table 7.2 with the R-squared adjusted values being below 30% for the triage category 4 models for 1999 and the 3 year combined model. The best fitting, and consequently better forecasting, models are in the main the Box-Jenkins and when combined with the regression model can provide an even better fit. However it should be noted from the two tables that the regression model with the polynomial fits does remarkably improve in both fit and forecasting ability. This type of regression model is easily utilized and understood by those without a mathematical background.

Model	Regression (Day Variables)	Regression (Day & Time^2)	Regression (Day & Month)
1998	0.4245	0.5498	0.5961
1998 Category 4	0.3562	0.4802	0.4669
1999	0.3599	0.6764	0.6975
1999 Category 4	0.2993	0.4110	0.4069
2000	0.5479	0.6970	0.6916
2000 Category 4	0.4533	0.5423	0.3516
3 Year	0.3323	0.4444	0.3762
3 Year Category 4	0.2923	0.4173	0.2379

Table 7.2: R-squared adjusted value for each regression model developed.

7.4 FUTURE RESEARCH DIRECTIONS

This research study examined the possibility that the number of children presenting to an urban paediatric emergency centre each day is predictable. We did this through the use of 2 dynamic modelling techniques and binary regression. Other models that could be considered include state spaced models, dynamic regression models or multiple autoregressive models. In particular, state spaced modelling allows time series models to be studied within a common mathematical framework, permitting common coding to apply to different model types.

The majority of models formulated in this study were for daily presentation rates. Tranberg and Qualls (1993) in their analysis indicated that examining the presentation rates in 2 hourly groupings could provide powerful short term forecasts. This type of grouping may also enable the impact of both weather and environmental factors upon the presentation rates to be studied. Emergency department managers make staffing level decisions based upon the expected number of presentations. It is common practice to reduce staff numbers over the summer period, however it is often an educated guess based on past experience.

All patients are given a triage code based upon their level of acuity at the time of presentation. While there are fewer patients with a triage code of one, each patient requires a greater staff involvement than those with a higher triage coding. The allocation of resources within an emergency department may benefit from models based upon triage coding.

The variation in presentation rates is linked to the time of year. Environmental factors are known to directly influence presentation rates. It could also be stated that the cause of a patient's presentation could provide useful modelling information and allowing better resource allocation.

We examined common variables that may influence presentation rates by use of the regression technique. While we noted the impact of one off events, eg the Australia Day fireworks, recent years have seen media reports have dramatic short-term impacts

upon presentation rates to emergency departments. This type of event could make future research more challenging.

7.5 CONCLUSION

This thesis has examined the daily presentation rates to a metropolitan paediatric emergency department with the aim of developing mathematical models to predict future patient volume. Factors of interest were those influencing the patterns of presentation – these were found to include day of the week (Sunday being generally the busiest day, and Tuesday or Wednesday the quietest), season (spring and winter being busier), and the prevailing weather conditions to a lesser extent (extreme weather conditions tend to delay presentations). One-off events such as public holidays and media reports were also noted to affected presentation rates (rates drop during televising of major sporting events and media reports of hospital mistakes, and increase after stories current diseases), but do not appear to cause cyclic patterns.

Analysis of presentation rates by four modelling techniques indicated the presence of two seasonal patterns – a weekly cycle and an annual cycle. Each model provided an adequate fit, but the dynamic models gave the better forecasts. In particular the Box-Jenkins model provided the best fit and forecast values. However, the Box-Jenkins model requires the use of an analytical program to find its initial parameter values, while Holt-Winters modelling and regression analysis techniques can be undertaken using most spreadsheet applications.

REFERENCES

Attia, M.W. (1998). The blizzard of 1996: A pediatric emergency department. Prehospital Emergency Care, 2(4), 285-288.

Attia, M.W., & Edward, R. (1996). Effect of weather on the number and the nature of visits to a pediatric ED. American Journal of Emergency Medicine, 1998; 16:374-30

Box, G.E.P., Jenkins, G.M. (1976). Time Series Analysis forecasting and control. California: Holden Day Inc.

Chatfield, C (1996). The Analysis of time series: an introduction. London: Chapman & Hall

Cryer, J.D., Robert, B.M. (1994). Statistics for business: Data analysis and modelling. California: Wadsworth Publishing Company.

Diehl, A.K., Morris, M.D., & Mannis, S.A. (1981). Use of calendar and weather data to predict walk-in attendance. Southern Medical Journal, 74(6), 709-712.

Goldsmith, J.R., Griffith, H.L., Detels, R., Beeser, S. & Neumann, L. (1983). Emergency room admissions, meteorologic variables, and air pollutants: a path analysis. American Journal of Epidemiology, 118(5), 759-78.

Gottman, J.M. (1984). Time Series Analysis: A comprehensive introduction for social scientists. New York: Press Syndicate of University of Cambridge.

Jorgensen, B., Lundbye-Christensen, S., Xue-Kun, S. & Sun, L. (1996). A longitudinal study of emergency room visits and air pollution for Prince George, British Columbia. Statistics in Medicine, 15, 823-836.

Lindsay, J.K. ((1995). Introductory Statistics: A modelling Approach. New York: Oxford University Press Inc.

Makridakis, S., Wheelwright, S.C., McGee, V.E. (1983). Forecasting: Methods and applications, 2nd Edition. Canada: John Wiley & Sons Inc.

Makridakis, S., Wheelwright, S.C. Hyndman, R.J. (1998). Forecasting: Methods and application, 3rd Edition. New York: John Wiley & Sons Inc.

Newbold, P, Bos, T. (1990). Introductory Business Statistics. Ohio: South-Western Publishing Co.

Saez, M., Sunyer, J., Castellsagué, J. Murillo, C., & Antó, J.M. (1995). Relationship between weather temperature and mortality: A time series analysis approach in Barcelona. International Journal of Epidemiology, 24(3), 576-582.

Shumway, R.H., Stoffer, D.S. (2000). Time Series Analysis and its Applications. New York: Springer-Verlag New York Inc.

Tandberg, D. & Qualls, C. (1993). Time series forecasts of emergency department patient volume, length of stay and acuity. Annals of Emergency Medicine, 23(2), 299-306.

Walpole, R.E., Myers, R.H. (1993). Probability and Statistics for Engineering and Scientists, 5th Edition. New York: Macmillan Publishing Company.

Xu, X., Li, B. & Huang, H. (1995). Air pollution and unscheduled hospital outpatient and emergency room visits. Environmental Health Perspectives, 103(3), 286-289.

Yamamoto, L.G., Iwamoto, L.M., Yamamoto, K.S. & Worthley, R.G. (1993). Effect of environmental conditions on emergency department use by wheezing children. Annals of Emergency Medicine, 22(3), 523-529.

APPENDIX A: Data values used in the analysis

Note: Data for each year is in date order from top left down and across the page

1998 Total daily presentations

131	83	111	134	115	103	116	95	86	140	122	173	104	94	92	99	116	120	150
103	101	95	102	116	132	119	107	70	131	107	129	159	127	101	119	93	120	129
101	89	71	104	108	112	150	112	69	120	132	127	128	130	120	121	100	107	97
118	95	90	103	93	115	111	141	99	133	117	136	94	105	125	119	105	105	115
106	111	107	108	84	87	110	116	101	132	117	109	100	97	85	138	126	89	99
81	117	128	113	107	99	104	110	108	151	134	123	91	109	98	97	155	104	
105	94	106	131	117	100	108	80	97	137	145	142	95	93	105	87	112	135	
103	91	88	111	149	101	120	118	87	144	146	163	107	98	105	110	103	121	
96	95	95	100	119	123	117	104	107	120	118	136	145	107	94	89	100	103	
111	92	98	76	109	113	130	119	121	145	136	109	96	139	101	105	110	93	
120	107	90	71	102	92	110	167	135	135	118	91	114	155	118	101	97	100	
87	108	128	83	116	86	105	137	141	155	98	115	107	118	98	127	130	99	
81	79	145	114	101	103	89	118	135	152	134	120	89	93	105	122	138	125	
82	89	99	142	120	88	101	102	123	154	156	116	109	105	92	103	112	127	
93	89	92	92	144	131	93	86	127	129	150	156	120	105	89	117	87	113	
73	79	85	113	94	140	127	91	106	124	132	128	130	85	111	97	95	85	
75	100	93	128	104	135	127	127	116	118	127	106	121	126	111	104	89	81	
99	98	107	97	82	101	98	132	134	93	117	86	80	107	133	111	113	91	
80	125	123	100	93	94	99	126	145	140	124	111	89	87	118	151	132	139	
90	89	151	106	135	83	98	118	143	135	138	105	78	105	106	110	158	152	

1999 Total daily presentations

129	82	105	98	96	121	104	82	110	145	125	164	150	143	87	126	111	88	130
117	85	108	98	92	91	148	128	108	127	102	121	150	186	130	103	98	88	143
119	92	96	93	98	89	97	133	125	112	123	133	138	150	131	133	95	94	110
107	125	104	88	88	104	92	120	158	140	133	133	158	125	106	125	115	87	101
97	102	132	101	81	100	83	102	128	164	148	135	143	117	98	127	152	114	90
105	101	100	137	85	99	94	75	110	147	173	141	112	123	111	107	115	131	
90	111	92	116	120	90	86	109	98	113	144	169	148	131	87	122	106	118	
90	85	82	91	103	118	107	92	109	114	119	108	177	133	121	120	102	66	
98	82	107	100	82	85	128	107	105	114	121	123	148	182	97	102	88	83	
111	99	112	104	94	93	110	118	114	135	123	105	141	165	111	133	95	74	
100	96	117	106	91	91	98	127	131	134	120	122	116	121	122	156	120	90	
87	78	133	115	118	88	91	111	97	153	131	132	137	119	114	107	124	99	
92	73	103	126	110	67	109	110	110	149	164	156	142	112	133	119	98	120	
86	81	98	100	123	96	117	97	114	126	141	190	150	124	118	110	101	93	
108	74	97	91	143	140	122	107	109	101	130	149	176	129	129	118	76	88	
90	88	104	107	92	136	106	116	105	129	136	107	143	152	117	104	93	80	
115	108	92	94	75	102	115	161	151	113	132	117	138	121	148	126	99	90	
91	116	113	96	87	86	104	161	168	152	145	113	99	133	105	152	120	94	
103	98	142	118	84	72	97	121	145	154	138	138	127	103	111	100	131	130	
95	107	147	124	121	86	90	105	142	128	159	146	134	105	91	119	106	149	

2000 Total daily presentations

159	83	86	90	69	110	102	109	92	116	109	106	121	182	132	130	117	92	140
164	78	78	85	80	97	82	144	109	125	105	119	103	146	135	130	127	93	110
110	110	112	75	72	94	89	123	157	118	108	109	112	131	116	162	120	97	104
87	90	110	106	85	90	70	105	124	162	130	104	116	111	117	118	141	113	131
84	81	89	104	99	105	99	118	113	137	169	124	109	128	113	89	101	155	116
83	88	92	117	134	96	83	126	106	128	116	159	131	112	114	104	99	109	144
87	99	85	109	98	109	85	100	100	116	106	151	169	123	131	113	125	115	
110	87	79	90	93	98	129	116	106	134	105	134	139	151	114	114	94	95	
110	86	89	104	108	95	101	153	114	119	116	128	109	149	153	155	95	83	
91	114	105	105	88	91	82	113	113	139	119	115	118	137	115	154	128	112	
83	79	108	93	102	98	86	107	103	147	150	122	120	122	115	126	154	121	
70	94	84	98	109	106	97	115	103	118	154	157	129	106	97	123	117	151	
55	86	85	85	112	131	92	98	103	112	107	142	156	128	114	101	103	121	
59	82	83	75	100	136	108	97	105	117	110	134	167	132	100	100	103	100	
85	68	83	73	85	143	115	106	121	133	103	126	136	152	137	132	97	87	
97	93	76	94	79	132	110	157	127	109	105	136	133	176	133	126	90	80	
61	130	103	73	81	96	99	124	136	118	118	119	101	133	145	163	104	99	
88	84	141	104	88	83	79	119	121	136	107	135	110	134	127	153	122	119	
71	87	99	116	99	76	108	116	114	117	130	153	133	106	124	113	98	143	
76	85	86	86	116	82	92	88	102	109	126	158	145	114	107	116	93	133	

1998 – triage category 4 daily presentations

83	51	73	73	78	64	76	60	52	90	70	97	59	57	50	62	72	79	88
62	58	66	51	75	83	67	69	52	76	62	85	92	84	58	71	60	77	73
69	55	39	54	67	68	85	68	53	63	72	80	80	85	75	73	63	66	60
68	59	47	57	50	67	56	82	67	81	68	79	59	64	77	80	60	74	62
67	77	41	52	50	62	60	51	81	79	71	64	70	57	52	79	82	59	64
51	84	72	60	57	64	56	66	64	91	83	82	51	69	57	61	100	62	
63	66	58	37	62	74	56	48	59	76	77	74	54	64	67	62	63	82	
68	56	54	47	87	67	89	57	58	93	90	102	72	57	60	68	63	64	
63	63	52	60	76	81	78	70	70	78	75	81	75	70	46	53	65	62	
74	57	38	37	59	77	70	68	75	87	94	65	49	96	63	52	66	65	
87	79	46	45	44	64	72	99	84	71	71	57	68	97	66	65	50	53	
61	69	66	48	71	52	60	77	83	97	63	66	57	78	55	66	90	60	
62	61	75	57	70	60	59	66	74	92	79	65	52	51	65	63	84	77	
55	51	43	73	69	59	53	66	77	86	99	70	62	66	63	64	67	78	
57	60	40	43	80	82	53	52	68	77	94	86	67	66	56	73	53	72	
50	50	41	51	64	88	79	62	52	68	81	61	76	61	73	67	49	55	
44	61	54	66	61	72	77	74	61	71	80	75	66	57	70	65	49	55	
67	68	47	59	47	73	54	86	82	56	71	58	47	61	88	66	65	61	
56	82	55	54	60	57	76	68	95	90	77	50	55	56	72	93	85	88	
57	55	62	68	87	51	69	85	91	77	81	59	56	69	52	64	105	95	

1999 – triage category 4 daily presentations

58	48	66	60	74	71	56	53	66	77	66	83	82	69	48	63	54	52	54
64	56	62	63	58	51	100	86	78	64	62	64	68	74	55	50	50	47	71
77	58	61	62	55	55	71	86	78	54	61	69	57	76	67	65	58	59	50
69	72	66	55	66	77	63	74	93	83	47	54	96	60	62	61	65	40	54
56	53	80	74	57	63	60	77	82	72	67	66	82	61	47	59	83	66	39
64	62	53	83	58	77	63	52	54	59	85	72	59	68	59	69	64	79	
56	69	69	71	83	73	53	63	70	57	64	77	70	68	52	66	64	64	
60	60	57	55	65	88	79	71	61	63	63	63	81	67	71	68	60	46	
62	57	70	74	54	65	82	76	73	63	55	59	61	102	54	56	54	39	
61	72	80	62	56	58	68	80	74	63	57	55	62	77	70	75	56	39	
61	49	64	67	57	50	54	80	67	81	62	52	55	60	62	81	74	53	
57	49	90	71	75	52	57	73	54	62	72	72	66	56	55	54	71	53	
50	58	59	84	75	34	64	68	60	87	83	73	68	52	74	49	58	74	
62	59	61	56	81	72	70	69	61	67	64	100	78	68	63	60	52	46	
65	56	58	60	90	91	84	61	72	48	65	73	91	62	73	59	40	47	
62	68	71	64	53	82	62	85	52	74	68	56	64	81	64	49	51	50	
76	87	56	72	40	57	64	94	85	61	71	64	68	56	65	72	51	47	
49	64	68	55	63	56	80	97	86	73	74	55	44	59	58	83	68	48	
63	64	82	96	61	54	57	74	70	82	66	67	70	53	56	51	75	54	
56	69	76	81	106	66	61	65	81	57	70	82	66	47	47	58	67	56	

2000 – triage category 4 daily presentations

64	46	44	48	35	60	51	60	42	61	56	60	49	98	81	70	61	55	71
74	43	40	49	47	45	43	82	54	53	53	63	55	71	59	73	75	53	65
48	60	70	47	42	49	47	57	83	65	44	51	51	67	62	79	52	52	60
47	47	65	58	46	46	36	59	55	74	67	52	60	49	66	65	81	61	67
48	37	46	52	60	53	55	60	68	58	84	67	56	66	60	52	48	77	56
40	43	59	60	64	47	48	73	60	63	60	70	71	54	62	65	54	71	75
50	42	52	52	50	54	47	52	66	63	51	71	75	67	71	65	63	64	
47	44	46	50	49	51	71	58	69	69	65	65	67	61	49	52	43	42	
60	43	52	54	49	48	46	76	57	66	50	71	48	86	78	69	50	52	
47	68	56	56	48	42	46	55	63	72	69	66	63	68	59	77	79	58	
41	42	59	56	54	47	40	54	52	77	84	61	60	62	69	53	80	64	
35	46	52	52	53	52	47	58	46	62	72	82	53	52	53	50	65	69	
30	48	46	50	73	71	54	59	61	58	54	83	78	74	62	50	54	67	
36	33	46	43	62	68	51	46	59	67	68	59	81	80	51	54	57	43	
42	35	59	43	46	76	75	62	71	72	57	63	72	70	80	76	54	50	
53	54	40	48	44	68	64	83	62	75	56	66	63	89	69	69	38	43	
35	70	55	45	53	59	57	56	72	63	57	61	45	77	67	82	55	63	
38	50	77	61	51	48	44	57	60	73	56	68	67	67	72	81	55	62	
46	53	47	64	59	41	62	65	53	58	72	85	71	55	58	63	56	77	
43	50	56	50	57	42	55	43	51	55	67	85	78	59	62	61	54	74	

Weather Data used for the Regression Modelling

1998 Maximum Temperature (degrees Celsius)

29.4	26.8	34.9	30.9	24.3	23.8	24.2	22.8	17.0	14.1	18.0	24.6	20.4	17.8	23.5	23.4	28.3	35.1	33.2
30.4	25.0	31.3	32.7	25.6	23.9	25.9	19.3	15.8	13.3	18.4	17.9	19.7	20.4	23.9	25.0	24.7	29.2	32.5
31.5	28.3	37.7	32.8	30.7	24.7	31.3	25.0	18.9	17.7	17.5	20.1	21.3	19.8	23.6	30.7	30.2	31.5	35.7
35.4	34.9	38.8	38.1	33.2	26.4	30.4	26.1	19.0	19.3	21.4	18.4	17.7	15.6	20.5	34.1	33.5	27.4	38.5
38.8	39.4	36.8	40.8	34.9	22.8	25.4	27.9	19.3	17.2	17.5	19.5	18.3	15.8	20.6	21.3	36.5	27.8	28.6
41.0	40.3	32.7	35.8	33.0	20.7	27.7	19.3	21.0	15.5	18.6	18.0	17.6	18.4	21.1	21.7	24.7	30.1	
43.2	36.8	35.0	30.6	31.4	23.6	29.0	21.0	19.9	13.9	13.4	22.1	16.6	20.8	21.2	20.2	24.8	35.4	
30.8	36.0	30.9	25.5	34.0	23.0	31.8	19.5	15.5	15.1	14.7	20.9	17.8	19.3	23.1	27.0	27.3	40.4	
33.6	36.0	28.5	22.9	33.9	24.1	22.9	20.1	17.5	17.4	16.2	19.4	20.9	21.0	21.4	22.1	22.5	33.4	
36.8	31.4	35.6	23.6	33.7	22.9	22.5	20.6	18.1	17.7	16.8	18.7	23.6	20.8	17.8	21.0	21.8	27.8	
31.6	30.3	35.9	25.1	34.1	26.3	21.1	24.9	16.2	18.7	17.4	18.8	23.7	21.5	17.5	25.5	27.9	30.3	
29.3	26.0	33.0	26.4	33.8	23.8	22.0	23.6	16.1	16.7	15.8	19.1	24.0	20.0	20.8	28.8	33.5	20.7	
27.3	29.7	35.8	31.5	36.2	23.0	22.9	20.4	18.9	17.1	18.0	16.9	19.9	15.9	24.4	32.7	28.0	27.2	
28.3	31.7	30.4	32.7	33.7	24.2	21.4	21.6	19.1	16.4	17.3	19.1	20.5	19.0	27.5	35.0	26.5	28.2	
28.7	29.4	29.0	35.2	28.3	27.2	22.7	18.4	18.6	18.9	18.9	21.3	20.6	20.0	29.0	38.3	20.1	33.4	
24.5	30.8	[illegible]	37.8	25.9	26.6	26.1	18.7	16.8	17.6	21.9	19.0	18.7	23.4	26.5	29.0	22.8	26.5	
28.9	30.8	28.2	38.7	21.1	26.2	29.4	19.6	17.2	18.8	17.0	20.6	19.8	21.5	24.8	18.4	29.2	24.7	
31.2	31.3	31.9	33.0	21.1	24.3	19.6	25.7	18.4	18.4	19.1	28.3	22.0	19.7	23.6	22.2	35.1	24.5	
36.3	37.2	41.6	30.0	28.4	23.1	22.2	19.7	16.6	17.5	20.9	16.8	24.9	20.6	27.2	28.8	31.2	26.7	
29.6	41.6	40.0	22.8	25.7	23.0	23.7	21.5	14.3	17.0	22.4	19.0	18.9	22.4	25.7	29.1	29.3	30.1	

1998 Minimum Temperature (degrees Celsius)

15.3	15.9	14.7	19.7	12.6	10.2	5.8	15.4	12.1	6.0	5.0	10.2	13.6	3.2	11.4	13.4	16.9	18.6	16.1
15.8	12.6	11.2	17.6	12.6	10.1	13.3	10.4	8.2	8.9	9.4	14.3	13.6	4.4	12.2	11.1	16.2	20.2	16.8
15.0	10.8	16.2	18.2	14.4	14.5	16.7	7.1	9.0	11.0	4.6	11.8	10.6	4.1	14.0	11.0	15.3	17.1	16.0
16.2	17.0	21.5	19.6	17.3	15.5	17.7	9.0	9.6	8.9	5.6	10.3	13.7	8.6	9.4	19.1	17.3	19.9	20.8
19.6	19.4	20.9	25.7	14.9	18.9	13.6	13.2	8.3	12.3	10.1	12.6	11.1	5.5	6.7	12.5	20.5	13.3	20.3
23.0	17.9	18.1	21.9	15.1	15.0	10.6	16.3	7.7	3.8	5.9	11.3	13.0	2.3	12.1	10.2	19.1	13.5	
25.8	19.5	17.2	22.8	16.5	16.4	15.3	12.9	9.8	8.0	7.0	12.7	2.2	6.8	9.8	6.9	13.2	15.7	
21.8	17.9	17.3	20.7	16.2	13.2	20.8	14.2	12.1	0.3	-0.2	10.6	5.0	12.4	8.0	8.2	10.8	22.3	
18.5	16.8	14.4	18.0	19.0	10.2	17.9	8.6	7.9	4.7	3.3	10.1	8.4	9.3	13.1	7.8	10.6	21.6	
18.3	16.3	14.7	16.6	17.4	11.6	11.8	8.7	8.7	3.8	2.4	13.0	10.2	13.0	12.6	11.0	7.5	17.9	
17.1	13.1	16.9	11.2	17.3	11.8	11.4	12.9	6.5	8.0	5.4	12.6	9.8	17.2	5.5	8.3	11.0	16.9	
16.6	11.9	17.6	13.6	15.8	10.5	8.9	12.1	4.1	11.5	10.5	7.4	4.2	14.3	4.5	9.5	11.4	15.9	
18.9	11.3	17.3	15.8	16.5	11.8	9.2	14.9	9.6	8.2	5.5	7.1	7.6	13.4	9.8	14.6	12.5	12.7	
18.8	13.1	14.8	14.1	11.4	13.3	9.8	11.3	8.3	10.6	9.3	5.0	9.3	7.1	10.3	15.3	12.1	13.2	
16.5	15.4	19.6	19.0	16.5	13.6	10.5	13.5	5.8	5.7	10.1	6.2	8.5	11.9	10.0	16.6	15.3	12.2	
18.9	15.3	14.3	16.7	11.4	16.8	11.9	9.9	3.6	3.5	5.7	12.1	5.7	5.5	11.6	17.6	8.6	10.8	
11.5	13.3	15.1	15.3	17.9	9.6	11.1	10.1	8.6	2.7	12.7	5.7	7.0	9.2	9.8	16.7	9.8	16.9	
15.1	13.0	15.9	14.8	16.2	9.0	14.6	11.2	6.3	4.2	7.9	11.4	6.2	9.2	9.1	11.9	14.7	11.4	
18.6	16.2	20.1	20.9	16.5	12.5	6.8	12.9	1.9	8.9	13.6	15.0	10.0	9.9	12.6	10.5	21.1	11.0	
18.1	19.1	22.1	13.8	16.4	9.7	12.7	11.4	4.6	9.0	12.3	13.1	11.5	16.0	17.6	13.5	18.1	13.5	

1998 Rainfall (in millimetres)

0	0	0	0.4	0.2	0	0	0.4	34.2	0	0	0	35.8	1.4	0	0	0.8	0	0
0	0	0	0	0	0	0	2.2	10.8	5.2	0.4	0	12	0	0	0	3.6	0	0
0	0	0	0	0	0	0	0	0.8	0	5.8	6.4	0.6	0	0	0	0	0	0
0	0	0	0	0	0	0	0	4.4	0.8	0	1.6	14.2	10.8	0	0	0	0.2	0
0	0	0	0	0	0	0	0	2.2	7.2	0	0	24	23.6	0.4	0	0	0	0
0	0	0	0	0	7.2	1	23.6	0	5	0	0	0.8	0	4.8	0.2	0	0	
0	0	0	0	0	0	0	2	0	21.6	32.4	0	3.2	0	0.2	0	0	0	
0	0	0	0	0	3.8	0	7	3.8	1.2	0	0.4	0	9.8	0	0	0	0	
0	0	0	13.8	0	0.6	33.4	0	2	0	0	0.8	0	0.6	7.2	0	0	4.6	
0	0	0	8.8	0	0	2.6	0	1.8	0	0	9.6	0	12.6	4.4	1	0	0	
0	0	0	1	0	0	0	0	0.2	0	0	0.2	0	0.2	2.8	0	0	0	
0	0	0	0	0	0	0	0	0	0.2	1.2	0	0	2	0	0	0	0	
0	0	0	0	0	0	0	0.8	0	0.4	9.4	3.6	0	0.6	0	0	0	0	
1.2	0	0	0	0	0.8	1	0.6	0	0	2.4	0	0	14.4	0	0	0	0	
0	0	0	0	0	0	0	0.8	0	0	3.4	0	0	1.8	0	0	1.8	0	
1	0	0	0	0	0	0	21.2	0	0	0	5	10.2	0	0	0	0.4	0	
0.4	0	0	0	0	0	0	0	6.4	0	6.8	0	0	0	0	1.2	0	0.2	
0	0	0	0	0	0	2.2	0	1.6	0	6.8	0	0	0.4	0	6.4	0	1	
0	0	0	1	0	0	0	30.8	5.2	14.6	5.6	22	0	0	0	0	0	0	
0	0	0	4.6	0	0	3.4	0.8	0	7	0	23.2	0.4	0	0	0	1.4	0	

1999 Maximum Temperature (degrees Celsius)

30.7	29.7	30.7	27.9	29.5	28.4	27.6	20.4	22.6	16.5	18.7	21.1	19.3	24.5	17.4	25.3	24.0	27.7	40.4
31.8	29.5	35.3	27.4	30.9	33.1	24.9	18.1	17.1	17.2	20.8	18.1	18.2	24.3	18.1	23.2	22.1	26.6	37.0
29.4	32.6	37.8	27.5	24.9	34.3	25.2	19.4	15.7	18.8	19.0	15.4	21.7	22.2	20.3	23.0	26.6	27.0	35.2
26.5	37.1	30.8	27.1	24.7	33.9	27.9	22.6	19.7	22.3	21.9	17.6	22.1	17.2	20.4	26.5	26.7	29.0	34.2
23.1	35.5	31.4	30.5	25.7	31.0	25.4	18.0	20.7	18.3	25.0	18.3	18.7	20.2	22.1	29.5	28.6	29.7	35.6
26.7	38.7	32.1	33.0	26.3	29.2	16.7	19.3	20.9	19.4	23.8	18.2	14.2	21.2	19.4	31.7	30.8	35.3	
30.2	34.3	33.8	24.2	28.1	27.4	16.9	24.2	20.3	16.4	19.0	17.6	16.6	25.9	17.5	30.6	32.9	29.7	
37.8	32.6	33.7	27.9	29.0	27.5	20.8	20.2	20.9	16.9	18.8	23.4	19.6	28.2	19.8	28.5	36.9	30.9	
25.2	33.3	36.9	35.1	26.4	29.8	22.4	19.8	19.0	17.1	15.0	20.2	20.3	24.3	19.0	25.5	32.7	29.5	
25.9	33.3	32.1	35.8	24.7	29.7	27.6	21.6	20.9	17.8	17.1	18.2	21.0	18.5	20.9	22.9	30.2	29.1	
30.4	26.6	32.4	35.3	25.2	33.9	22.1	22.4	21.2	19.6	18.3	16.5	20.5	19.0	24.9	23.3	30.5	32.9	
27.4	25.6	31.8	32.7	25.5	28.4	18.7	22.1	20.4	22.6	19.1	17.1	24.5	22.1	28.6	24.3	35.5	34.9	
36.0	26.6	34.3	27.7	28.7	24.0	19.0	21.6	21.7	24.0	20.7	18.4	22.9	21.8	28.4	28.7	30.7	34.2	
27.3	29.1	35.4	31.6	31.7	24.0	24.3	20.0	19.7	17.2	19.5	21.2	18.7	19.9	26.7	32.0	27.6	31.6	
30.6	32.3	37.8	30.9	33.9	24.3	22.5	16.6	20.2	18.6	19.2	24.1	21.0	22.7	29.9	28.8	22.9	36.8	
38.5	31.7	38.8	33.3	35.5	24.5	24.0	18.4	19.7	20.0	19.5	18.0	20.2	27.6	22.7	25.9	23.8	39.3	
32.6	35.6	38.4	29.3	34.8	26.0	24.2	20.5	19.6	18.3	20.9	18.1	17.7	26.4	20.3	26.7	30.3	29.9	
31.5	39.7	38.4	25.4	28.3	26.3	26.5	20.4	20.8	18.9	16.7	19.0	16.6	21.0	21.7	23.9	39.3	38.0	
33.5	39.7	36.2	25.4	30.5	26.0	22.6	18.8	18.6	18.8	15.5	21.5	16.9	21.7	23.8	24.5	32.0	35.1	
40.4	29.4	28.5	28.0	26.4	27.4	22.5	20.1	17.5	19.3	18.3	21.2	21.2	19.2	25.2	24.0	29.7	36.5	

1999 Minimum Temperature (degrees Celsius)

19.7	18.4	15.9	16.2	19.8	19.0	9.9	13.7	10.5	2.2	9.3	8.6	13.5	10.9	10.3	10.7	16.3	16.4	21.6
18.4	19.2	16.4	14.4	21.8	19.8	14.1	7.4	13.1	5.9	11.7	11.7	7.1	5.5	10.6	12.8	7.9	13.7	22.5
13.9	19.1	22.7	11.6	15.4	20.4	12.5	7.4	5.9	9.9	8.4	8.9	4.6	9.1	8.8	10.8	12.9	14.9	20.3
14.5	21.2	19.3	10.0	13.5	23.9	11.2	9.7	8.4	5.0	7.0	4.2	7.0	13.0	5.9	12.8	13.5	14.0	21.5
11.2	21.6	15.3	10.5	8.9	19.2	14.9	13.2	12.9	9.2	6.8	3.9	11.3	12.4	11.7	13.1	14.0	14.6	21.4
11.5	21.9	15.5	13.6	14.9	19.7	14.7	14.8	10.4	12.7	7.1	8.8	9.5	6.3	11.8	16.5	13.7	18.4	
14.7	26.3	14.0	15.2	14.1	16.0	13.6	16.4	10.6	9.0	11.5	11.3	1.8	6.8	10.4	17.1	15.3	17.2	
17.5	20.5	17.8	8.4	13.0	13.2	15.0	15.8	9.7	3.3	6.2	11.4	2.9	10.5	10.2	17.6	15.3	15.6	
20.2	22.3	19.2	12.9	15.5	13.9	16.5	8.2	11.8	1.9	11.7	15.8	5.1	14.9	13.2	17.4	15.3	18.6	
12.7	18.9	16.9	21.3	15.9	17.0	16.5	8.0	11.6	4.1	2.5	12.6	5.1	12.9	6.8	13.7	14.5	14.2	
13.0	20.4	17.9	21.1	14.7	18.3	15.3	8.9	9.5	9.6	6.5	5.7	8.9	6.8	9.2	9.8	12.4	14.9	
13.2	17.5	18.8	21.1	9.0	15.7	14.3	9.4	7.0	12.6	4.4	3.3	8.9	12.0	12.7	7.1	15.9	19.8	
13.7	12.8	19.0	16.4	13.1	14.4	15.2	9.5	10.2	12.9	4.8	5.1	10.1	7.7	10.0	12.4	16.2	19.2	
18.7	15.4	18.4	20.0	11.2	9.4	11.0	10.2	11.4	12.4	3.6	6.5	12.4	5.0	14.8	14.0	17.0	18.2	
16.3	15.8	21.8	21.1	12.6	6.2	15.1	3.9	7.6	13.5	5.4	10.7	10.0	8.5	12.3	9.2	16.0	17.7	
18.2	16.4	24.7	18.2	17.1	11.1	12.2	3.4	11.3	12.3	5.7	12.0	9.9	8.7	15.5	10.2	10.9	20.3	
19.4	17.8	22.9	21.4	16.3	7.7	10.9	6.9	9.9	12.3	7.4	10.9	11.2	8.7	12.5	11.8	9.5	16.5	
17.4	20.7	21.3	19.4	15.1	8.6	10.4	6.3	14.1	12.4	10.0	7.1	7.1	11.6	9.2	12.2	14.7	17.3	
15.5	20.1	15.2	20.6	17.3	11.0	15.7	8.1	12.2	12.8	2.5	12.2	3.4	12.9	7.7	14.2	19.6	18.6	
19.4	15.2	19.0	19.7	18.6	10.1	13.6	8.0	5.3	8.8	2.9	9.8	6.7	14.4	7.8	14.8	16.2	17.6	

1999 Rainfall (in millimetres)

0	0	0	0	0	0.2	0	9.4	0	0	0	0	11.4	0	32.8	0	0	0	0
0	0	0	0	0	0	3.2	6.2	18.4	0	0	3.2	9.8	0	9.4	0	0	0	0
0	32.6	0	0	0	0	0.8	0	6.4	0	0	31.4	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	4	0	2.2	0	0	0	0	0
12.2	0	0	0	0	2.8	0	0.2	14.8	0	0	0	0	7	0	0	0	0	0
0	0	0	0	0	0	9.2	3.8	1.2	27.6	0	0	15.2	0	0	0	0	0.4	
0	0	0	4.8	0	0.6	0	14	6.8	2.8	15.8	6.4	7.2	0	19.6	0	0	3.6	
0	0	0	0.4	0	4.6	0	31.4	0	1.6	3.6	3.4	0	0	1.2	1.8	0	0	
0	0	0	0	0	0	0	4.2	23.2	0	13.4	10.2	0	1.4	4.4	0	0	0	
0	0	0	0	0	0	0.2	0	8.6	0	1.2	34.4	0	7.2	0	0	0	0	
0	0	0	0	0	0	0	0	2.4	0	0	2.8	0	3.2	0	0	0	0	
0	0	0	0	0	0	0	0.2	0	0.4	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	1.8	0	0	0	0	0	0	0	0	0	0	0	
0.8	0	0	0	0	0	4.4	4.8	22	4.6	0	0	26.4	1.2	0	0	0	4.2	
0	0	0	0	0	0	2.4	1.6	0.2	13	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	6.2	0	16.8	0	0	0	0	0.2	0	
0	0	0.6	0	0	0	0	0	9.2	16.2	0	8.4	12.6	0	0	0	0	0	
0	0	0	19.6	0	0	0	0	1.8	6.2	12.2	0	9.4	0	0	0	0	0	
0	0	0	2.6	0	0	14	4.2	4	1.8	2.2	0	0	0	0	0	0	0	
0	0	0	0	5.6	0	6.2	3.8	4.2	1.6	0	0	0	10.2	0	3	0	0	

2000 Maximum Temperature (degrees Celsius)

36.5	26.2	28.3	26.7	25.8	31.2	20.2	21.9	20.3	18.0	18.9	16.6	19.3	22.5	20.0	31.0	28.8	33.5	38.3
37.2	20.6	29.6	30.2	23.8	27.4	22.2	21.8	17.8	20.2	18.1	16.7	16.9	20.1	20.7	34.4	25.2	33.1	41.1
36.3	24.6	30.5	31.1	25.7	24.3	24.0	22.8	23.1	17.2	18.7	17.0	19.4	18.5	21.9	24.1	29.6	33.1	29.9
37.4	27.8	33.3	31.8	28.5	23.6	24.9	22.2	26.1	19.0	19.4	17.7	15.4	21.3	27.0	22.4	28.9	30.3	24.6
37.2	28.5	35.4	32.4	29.6	24.2	24.3	22.4	23.2	19.5	17.6	21.1	17.7	25.1	21.3	23.3	27.7	31.5	28.1
35.4	29.1	36.1	34.9	27.0	23.5	24.4	21.0	19.4	17.9	19.3	20.9	18.1	22.2	19.8	25.4	29.5	34.6	30.0
34.5	30.8	36.6	37.0	25.2	25.0	24.1	18.5	18.5	18.3	16.2	20.7	19.8	23.6	23.3	23.4	34.3	35.0	
32.2	33.1	37.7	38.4	26.4	25.7	25.6	18.5	19.7	19.6	16.0	18.5	21.4	23.8	26.9	23.8	27.2	33.3	
32.7	37.1	32.3	35.9	28.7	29.3	26.3	18.8	19.7	18.5	19.7	18.8	24.3	24.9	29.5	28.1	24.5	35.1	
29.4	37.5	27.9	31.7	29.1	30.2	27.6	22.5	21.2	18.5	15.7	18.4	18.4	28.0	24.8	29.8	27.9	36.2	
27.6	29.9	30.4	28.2	28.8	25.7	24.4	23.5	17.3	18.6	18.1	19.2	19.1	23.0	23.5	23.3	23.0	30.2	
27.1	27.9	34.6	26.7	29.2	28.5	17.0	22.5	19.3	19.4	18.8	20.8	20.7	19.7	27.0	21.6	21.5	32.2	
30.7	27.7	26.3	27.7	26.3	29.5	17.7	19.6	20.4	18.3	18.1	19.3	19.9	19.3	33.0	20.1	23.5	32.3	
23.9	27.5	28.3	30.1	23.5	30.2	19.4	21.1	19.3	17.8	18.6	17.6	21.7	21.4	23.7	22.5	25.1	29.6	
24.5	30.2	28.8	30.0	23.9	30.1	20.8	19.7	20.7	17.9	14.3	17.9	28.1	26.3	20.3	24.7	30.1	24.4	
25.2	31.9	30.2	25.6	24.6	26.4	22.3	20.9	18.9	19.5	17.2	23.0	22.4	27.9	22.4	32.7	28.0	21.9	
25.7	35.9	31.2	25.6	27.5	26.3	23.1	20.6	20.7	19.3	19.9	17.3	19.5	27.6	24.6	34.8	26.0	23.6	
29.1	38.5	35.1	29.0	27.9	20.9	23.6	21.3	16.6	21.2	14.1	17.6	19.5	23.3	29.1	33.3	31.7	29.4	
30.9	41.1	33.2	28.4	29.7	21.7	23.1	19.8	16.8	18.5	15.6	17.4	19.8	22.6	23.1	32.6	30.1	35.7	
33.0	31.6	36.0	34.2	30.4	19.1	23.1	21.3	19.6	16.3	17.8	16.9	20.0	19.2	26.4	36.9	31.4	38.7	

2000 Minimum Temperature (degrees Celsius)

20.7	17.1	14.9	20.6	21.0	20.8	9.3	8.4	9.5	10.6	8.7	5.3	8.3	5.7	7.2	12.8	20.5	15.8	20.9
21.0	19.2	16.6	15.1	10.6	20.7	10.6	9.2	11.6	6.3	11.1	9.9	8.0	6.4	5.2	13.9	16.4	15.6	24.0
18.3	16.6	15.9	15.9	12.8	10.8	13.6	9.8	10.3	10.7	10.0	3.6	7.4	7.5	4.2	15.2	14.8	17.2	24.2
19.7	17.0	19.1	16.6	12.6	15.2	9.7	4.6	12.3	12.1	9.3	3.6	11.3	6.0	5.1	14.5	14.3	18.2	13.4
22.8	16.6	16.3	16.9	13.7	9.7	12.7	7.5	16.2	11.5	12.2	7.8	7.9	6.7	8.7	8.2	17.9	15.7	10.1
22.4	16.2	21.3	18.7	17.9	7.9	11.1	7.3	12.9	12.4	8.6	9.1	12.7	8.0	4.6	10.0	14.4	17.4	14.2
16.1	17.7	20.3	22.6	15.1	7.4	9.2	4.9	8.2	7.0	6.5	6.6	12.0	12.3	4.6	8.0	16.7	18.0	
14.6	18.4	19.8	25.1	14.1	5.6	8.3	3.8	5.5	9.2	5.8	11.1	7.9	10.7	10.3	7.7	13.6	17.4	
16.5	22.9	17.4	25.2	11.2	13.2	12.0	7.3	5.3	11.1	7.2	12.0	9.2	8.5	13.7	7.9	14.5	19.0	
17.0	22.0	14.4	23.5	11.4	12.6	10.0	12.3	5.6	6.4	11.3	8.1	11.0	10.2	10.0	11.2	12.4	21.0	
14.4	22.6	15.5	21.4	16.5	17.8	10.0	14.5	10.3	5.9	9.0	12.1	10.1	15.6	9.2	9.8	11.9	19.0	
17.1	16.9	20.0	18.8	16.8	16.9	13.0	15.8	9.2	6.9	9.2	7.0	10.1	15.4	9.5	7.7	14.3	15.6	
15.6	13.0	17.7	13.8	12.8	16.9	6.4	8.4	11.9	11.2	6.4	8.8	7.3	10.8	11.6	8.4	10.4	13.3	
18.8	12.7	10.9	13.0	16.8	16.1	10.6	7.5	11.9	6.9	8.0	0.4	7.5	11.3	11.5	10.6	8.4	15.4	
17.2	15.5	12.9	10.5	12.8	15.6	4.8	9.3	13.6	11.8	11.4	2.6	7.8	8.4	9.4	10.5	13.0	16.0	
16.3	16.9	14.1	10.2	15.3	15.1	6.8	10.4	13.1	9.3	2.4	6.1	7.5	8.3	4.2	11.6	9.9	6.7	
15.2	19.5	16.2	8.2	8.8	12.6	8.1	9.4	11.2	7.8	4.4	11.6	8.3	9.3	6.4	17.8	9.9	5.6	
11.6	23.2	20.0	9.7	16.8	15.1	6.0	8.1	2.8	11.7	8.7	8.1	12.6	11.2	8.8	17.4	11.1	12.9	
13.3	24.9	22.1	13.1	9.4	12.5	7.4	8.5	8.7	14.0	8.0	7.8	11.1	9.6	10.2	17.0	12.0	15.3	
16.2	21.3	25.3	20.3	17.1	4.2	8.9	8.2	9.4	11.4	10.1	5.6	9.2	11.4	9.2	20.9	16.1	21.2	

2000 Rainfall (in millimetres)

0	0	0	0	0	0	0	0	0	0	1.2	0.2	0.2	0	0	0	0	0	0
0	60.8	0	0	24	0	0	0	3.2	0	2.8	0.4	9.6	0	0	0	0.5	0	0
0	14.4	0	0	0	0	0	0	2.8	32.4	2.8	0	2.2	0	0	0	0	0	0
0	0	0	0	0	1.2	0	0	0	39.8	2	0	29	0	0	0.8	0	0	0.6
0	0	0	0	0	0	0	0	0	5.2	2.8	0	4.2	0	0	0	0	0	0
0	0	0	0	0	0	0	0	33.2	20.2	0.4	1	5.2	0	0	0	0	0	0
0	0	0	0	0	0	0	0	15	8.2	0	0.4	1.8	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	1.8	0	0	0.6	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	3	0	16	0	0	0	0	0	0	
0	0	0	5.6	0	0	0	0	0	0	0	0.2	18.6	0	0	0	0	0	
0	3.8	0	0	0	0	0	0	9.8	0	5.2	0	1	1.2	0	0.8	0	0	
0.8	0	0	0	0	14.4	11.8	0	3	2.8	0	0	1.2	1.8	0	0	10.4	0	
0	0	0.2	0	0	0	8	4.4	2.4	30.2	0	0.2	0	0.4	0	2.6	0	0	
0	0	0	0	0	0	1	0	1.4	0	0.6	0.2	0	1.6	0	4.2	0	0	
13	0	0	0	0	0	0	0	9.8	15.8	5.2	0	0	0	0.4	0.4	0	0	
5	0	0	0	0	0	0	0	2.4	3	2.4	0	0	0	0	0	0	0	
4.6	0	0	0.4	0	6.4	0	0	3.4	6.8	0	12.8	3	0.2	0	0	0	0	
0	0	0	0	0	5.8	0	0	3	7.8	6.2	0	2.6	0	0	0	0	0	
0	0	0	0	0	3	0	0	0.8	13.6	20.6	12.4	1.4	0.2	0	0	0	0	
0	0	0	0	0	10.2	0	0	0.8	9.6	10.2	0	0.8	3.8	0	0	0	0	

APPENDIX B: Model Parameter Values

Box-Jenkins model parameters (May to April):

Year	Model	Constant	AR(1)	SAR(7)	SAR(14)	MA(1)	SMA(7)	SMA(14)
1998	$(1,1,1)(1,1,1)_7$	-0.0020	0.3065	0.0560	-	0.9018	0.9613	-
1998 cat 4	$(1,1,1)(1,1,1)_7$	-0.0017	0.1434	- 0.0394	-	0.9081	0.9579	-
1999	$(1,1,1)(2,1,1)_7$	-0.0090	0.2446	- 0.0299	- 0.1327	0.8400	0.9598	-
1999 cat 4	$(1,0,1)(2,1,2)_7$	-0.0620	0.9187	- 0.5643	- 0.0334	0.7703	0.3802	0.5752
2000	$(1,1,1)(1,1,2)_7$	-0.0320	0.2553	- 0.8898	-	0.9164	0.0527	0.8335
2000 cat 4	$(1,0,1)(1,1,1)_7$	0.0281	0.8173	0.0394	-	0.6467	0.9334	-
3 year	$(1,1,1)(2,1,1)_7$	-0.0002	0.2597	0.0469	-0.0734	0.8732	0.9778	-
3 year cat 4	$(1,1,1)(2,1,1)_7$	-0.0001	0.1177	0.0051	-	0.9127	0.9802	-

Regression model parameters

Variable	1998	'98 Cat 4	1999	'99 Cat 4	2000	'00 Cat 4	3 year	3 year cat 4
Constant	80.0740	54.8302	80.6741	60.9704	61.6010	32.9612	93.0704	60.5918
Time	0.0436	0.0247	0.0638	-0.0090	0.1269	0.0633	0.0087	-0.0069
Monday	10.1962	4.1894	11.3191	3.8763	8.7375	2.7417	9.9270	3.2529
Tuesday	-	-	-	-	-	-	10.5504	6.6864
Thursday	-	-	-	-	-	-	21.1047	9.5097
Friday	-	-	-	-	-	-	15.7943	6.0775
Saturday	14.4614	10.0201	15.6884	11.5804	14.0822	7.2822	12.4050	8.0747
Sunday	34.3614	19.1719	33.9110	17.2850	34.3086	16.2850	29.4512	15.3324
Holiday	31.1801	18.8657	31.6219	13.1074	37.0315	15.3641	20.7594	8.3535
April	-	-	-7.1919	-	-	-	-11.6870	-4.3239
June	11.3078	7.6315	17.2999	8.6167	10.3642	5.0943	-7.7129	-2.7501
July	28.0912	14.4615	25.0196		18.3990	8.4758	9.7014	3.4609
August	21.3373	11.2637	29.4333	4.6998	17.0654	7.3999	24.9643	9.6551
September	-	-	28.3437	5.8152	16.7259	6.0762	20.4292	7.2590
December	-4.7913	-	-19.4434	-8.5045	-22.7835	-11.2300	-	-
Max temp	0.3930	-	0.2758	-	0.5589	0.3387	-	-
Rainfall	-0.2513	-0.2658	-0.2026			0.0056	-0.3465	-0.1653

Box-Jenkins parameters – Regression residual models

Year	Model	Constant	AR(1)	MA(1)	MA(2)	SMA(7)
1998	(1,1)	0.0062	0.7850	0.5553	-	-
1998 cat 4	(1,1)	0.0062	0.7784	0.5967	-	-
1999	(1,0)	0.0019	0.1817	-	-	-
1999 cat 4	(1,0)	-0.0042	0.0583	-	-	-
2000	(1,0)	0.0385	0.2971	-	-	-
2000 cat 4	(1,1)	0.0041	0.6998	0.5886	-	-
3 year	$(1,0,2)(0,0,1)_7$	0.0015	0.9561	0.6306	0.1301	0.9472
3 year cat 4	$(1,0,2)(0,0,1)_7$	0.0004	0.9769	0.7840	0.0746	0.9714
Weekly	$(1,1,1)(0,1,0)_7$					

Holt-Winters model parameters

Model	α	β	γ
1998	0.23	0.02	0.24
1998 Category 4	0.14	0.05	0.19
1999	0.24	0.03	0.15
1999 Category 4	0.10	0.04	0.06
2000	0.27	0.03	0.24
2000 Category 4	0.21	0.03	0.14
3 years Combined	0.25	0.01	0.14
3 years Category 4	0.15	0.05	0.14
Monthly	0.89	0.01	0.50

Mean Squared Error (MSE) Values for each model

- Best model highlighted in blue, worst in red.

Model	Holt-Winters	Box-Jenkins	Regression (Day Variables)	Regression (Day & Time^2)	Regression (Day & Month)	Regression & Box-Jenkins
1998	198.64	163.55	224.45	173.11	156.60	137.69
1998 Category 4	111.54	96.46	112.07	87.26	90.50	95.54
1999	203.08	177.78	331.07	165.01	154.94	149.87
1999 Category 4	96.3	70.44	98.29	81.45	83.07	83.19
2000	209.72	156.03	232.49	153.65	157.28	143.45
2000 Category 4	84.09	78.57	74.93	61.85	89.69	89.85
3 Year	188.71	165.66	322.72	267.30	299.62	168.27
3 Year Category 4	96.2	82.03	116.32	95.34	123.63	85.92

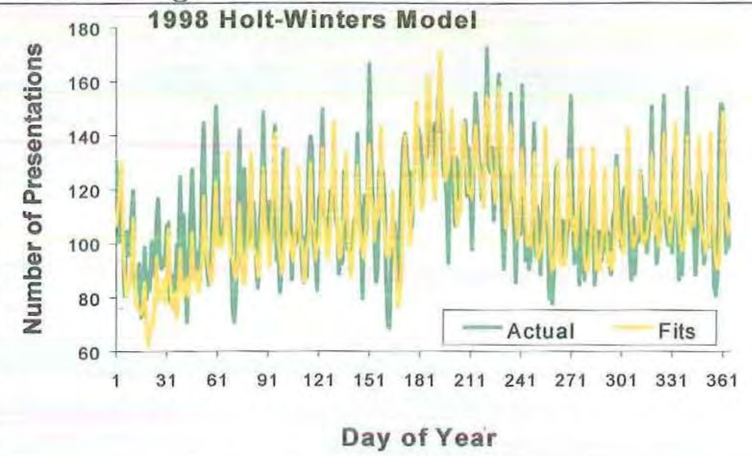
R-squared Adjusted Values for each Regression model

- Best model highlighted in blue, worst in red.

Model	Regression (Day Variables)	Regression (Day & Time^2)	Regression (Day & Month)
1998	0.4245	0.5498	0.5961
1998 Category 4	0.3562	0.4802	0.4669
1999	0.3599	0.6764	0.6975
1999 Category 4	0.2993	0.4110	0.4069
2000	0.5479	0.6970	0.6916
2000 Category 4	0.4533	0.5423	0.3516
3 Year	0.3323	0.4444	0.3762
3 Year Category 4	0.2923	0.4173	0.2379

APPENDIX C: Holt-Winters Models

1998 All Triage model



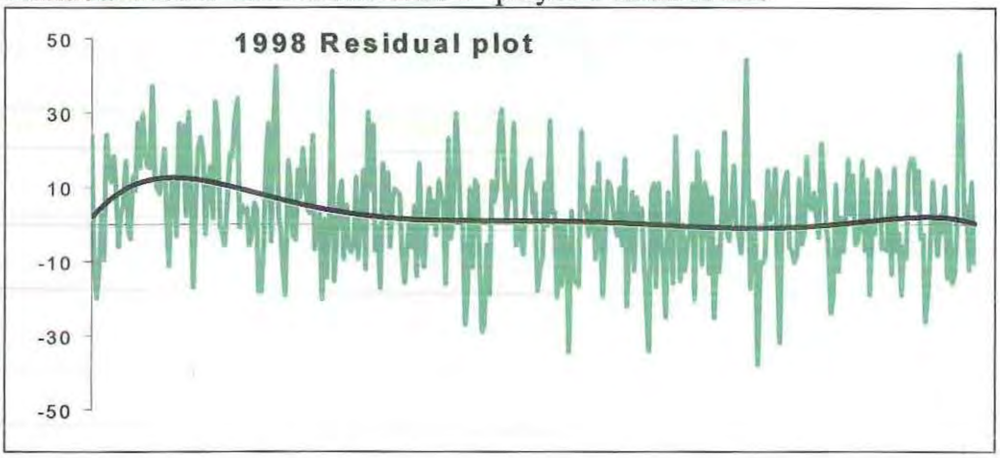
Smoothing Constants
Alpha (level): 0.23
Gamma (trend): 0.02
Delta (seasonal): 0.24

Accuracy Measures
MAPE: 10.448
MSE: 198.64

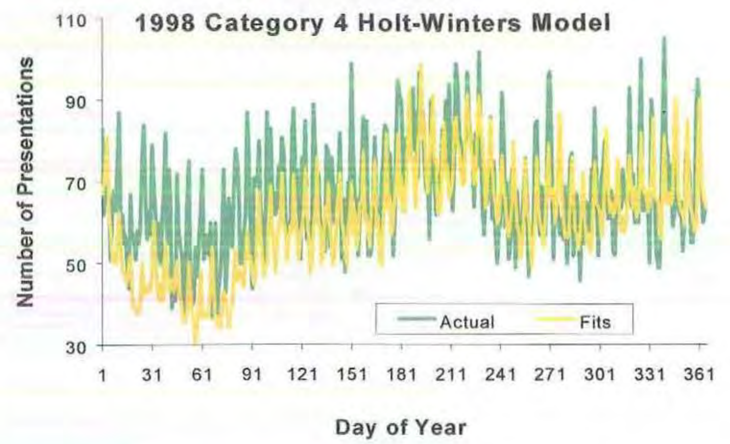
Forecasts for month of January 1999 using above model

Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper
150	149	120.9	177.1	**105**	108	68.9	147.2	**105**	108	68.9	147.2
129	123	94.4	152.3	**90**	107	66.6	147.6	**90**	107	66.6	147.6
97	108	77.8	137.5	**90**	151	109.6	193.3	**90**	151	109.6	193.3
115	105	74.1	135.7	**98**	125	82.2	168.6	**98**	125	82.2	168.6
99	108	76.1	139.9	**111**	109	64.8	154.1	**111**	109	64.8	154.1
129	120	86.8	152.7	**100**	107	60.6	152.7	**100**	107	60.6	152.7
117	135	100.6	168.9	**87**	110	62.3	157.3	**87**	110	62.3	157.3
119	151	115.3	185.9	**92**	122	72.7	170.7	**92**	122	72.7	170.7
107	126	89	162.1	**86**	137	86.5	187.4	**86**	137	86.5	187.4
97	106	68.6	144.3	**108**	153	100.9	204.8	**108**	153	100.9	204.8

Residual Plot of 1998 model with 6^{th} polynomial trend line



1998 Triage category 4 model



Smoothing Constants
Alpha (level): 0.14
Gamma (trend): 0.05
Delta (seasonal): 0.19

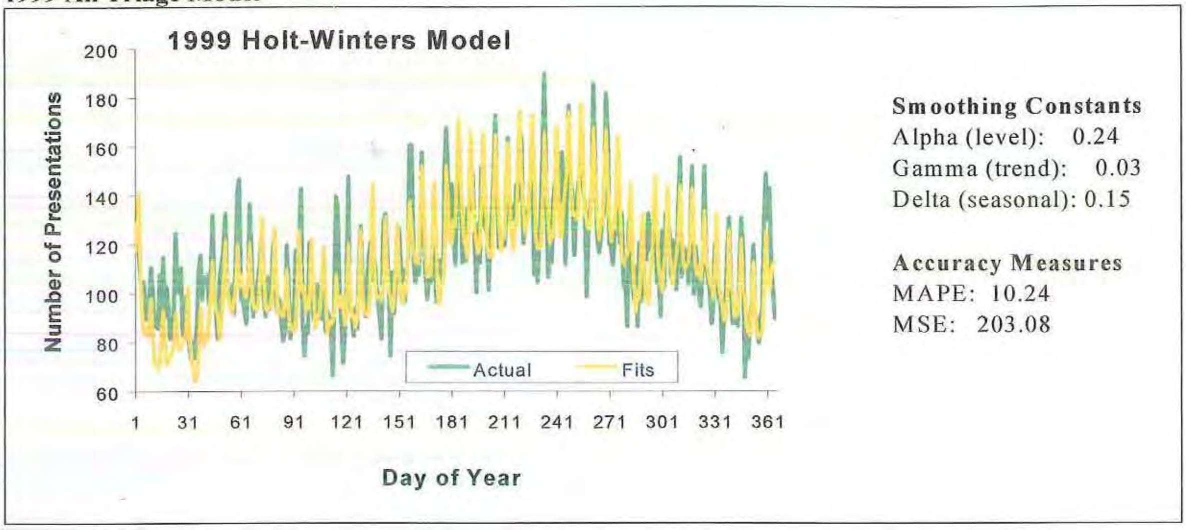
Accuracy Measures
MAPE: 13.092
MSE: 111.54

Forecasts for month of January 1999 using above model

Actual	Forecast	95% CI		Actual	Forecast	95% CI		Actual	Forecast	95% CI	
		Lower	Upper			Lower	Upper			Lower	Upper
88	90	64.6	115.8	**64**	66	33.7	97.3	**62**	72	32.1	112.1
73	72	45.9	98.1	**56**	65	32.4	97.5	**76**	66	25.5	107.3
60	66	39.7	92.9	**60**	90	57.0	123.6	**49**	66	24.3	107.9
62	66	38.8	93.1	**62**	72	37.9	106.1	**63**	65	22.2	107.7
64	65	37.1	92.6	**61**	66	31.4	101.3	**56**	69	25.9	113.1
58	69	41.0	97.7	**61**	66	30.3	101.8	**48**	83	38.3	127.4
64	83	53.8	111.7	**57**	65	28.3	101.5	**56**	90	44.9	135.9
77	90	60.2	119.5	**50**	69	32.0	106.9	**58**	72	25.7	118.6
69	72	41.9	102.6	**62**	83	44.6	121.1	**72**	66	19.1	113.8
56	65	34.2	96.3	**65**	90	51.2	129.5	**53**	66	17.8	114.4

Residual Plot of 1998 Category 4 model with 6^{th} polynomial trend line

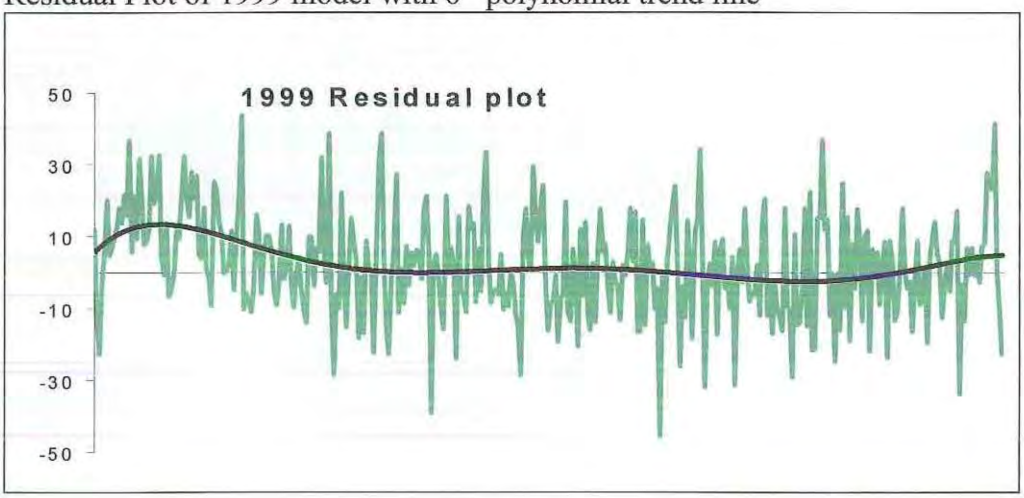
1999 All Triage Model



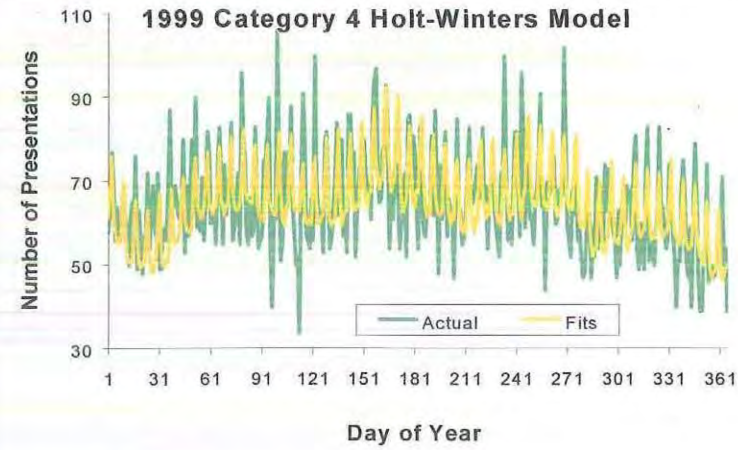
Forecasts for month of January 2000 using above model

Actual	Forecast	95% CI		Actual	Forecast	95% CI		Actual	Forecast	95% CI	
		Lower	Upper			Lower	Upper			Lower	Upper
130	107	78.75	134.36	**83**	94	55.67	133.20	**97**	95	42.37	148.21
143	96	67.51	124.80	**87**	95	55.22	135.40	**61**	93	38.18	147.03
110	93	63.89	123.00	**110**	106	64.51	147.38	**88**	95	39.08	150.95
101	96	65.34	126.41	**110**	96	52.80	138.42	**71**	97	39.62	154.53
90	98	66.39	129.53	**91**	93	48.72	137.12	**76**	115	56.48	174.45
159	117	83.85	149.18	**83**	95	49.72	140.95	**83**	132	71.42	192.47
164	133	99.34	166.95	**70**	97	50.36	144.45	**78**	105	43.17	167.31
110	109	73.74	143.72	**55**	116	67.36	164.35	**110**	95	31.35	158.60
87	100	63.71	136.13	**59**	132	82.43	182.34	**90**	92	27.12	157.48
84	93	55.79	130.74	**85**	106	54.16	157.03	**81**	95	27.95	161.44

Residual Plot of 1999 model with 6th polynomial trend line



1999 Triage Category 4 Model



Smoothing Constants
Alpha (level): 0.10
Gamma (trend): 0.04
Delta (seasonal): 0.06

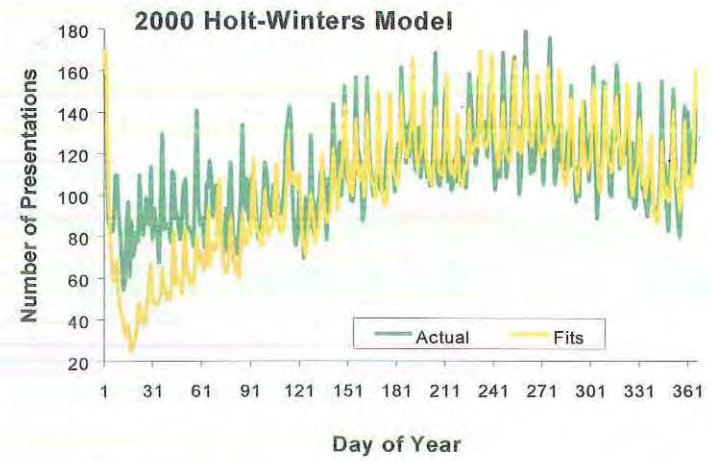
Accuracy Measures
MAPE: 11.877
MSE: 96.3

Forecasts for month of January 2000 using above model

Actual	Forecast	95% CI		Actual	Forecast	95% CI		Actual	Forecast	95% CI	
		Lower	Upper			Lower	Upper			Lower	Upper
130	50	68.82	31.99	83	44	63.72	24.65	97	41	62.23	20.19
143	46	64.50	27.48	87	43	63.15	23.81	61	40	61.38	19.01
110	45	63.49	26.27	110	47	66.93	27.31	88	41	62.26	19.56
101	46	64.42	27.00	110	43	62.92	23.02	71	41	62.47	19.43
90	46	64.60	26.97	91	42	62.01	21.83	76	48	69.41	26.02
159	53	72.31	34.46	83	43	62.92	22.44	83	53	75.24	31.50
164	60	78.79	40.71	70	43	63.12	22.35	78	43	65.33	21.24
110	49	67.83	29.51	55	50	70.35	29.26	110	39	61.68	17.23
87	45	64.72	26.15	59	56	76.42	35.03	90	38	60.88	16.06
84	43	62.67	23.86	85	45	66.06	24.35	81	39	61.74	16.55

Residual Plot of 1999 Category 4 model with 6^{th} polynomial trend line

2000 All Triage Model



Smoothing Constants
Alpha (level): 0.27
Gamma (trend): 0.03
Delta (seasonal): 0.24

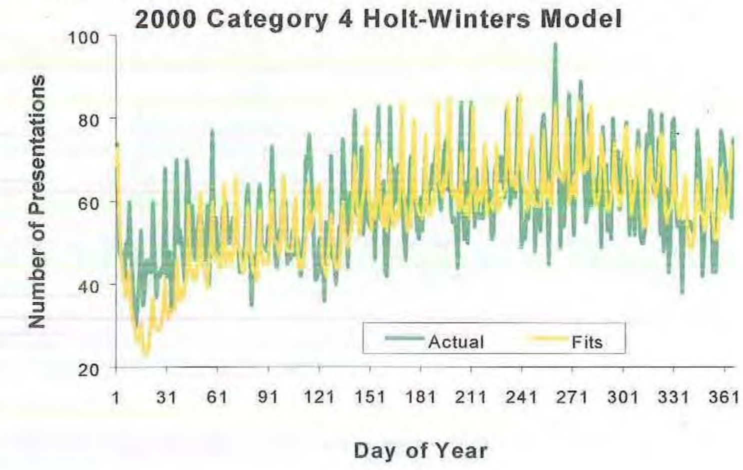
Accuracy Measures
MAPE: 15.113
MSE: 209.72

Forecasts for month of December 2000 using above model

Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper
154	137	96.85	177.00	**92**	101	41.21	161.61	**121**	113	27.52	198.83
117	116	74.10	157.32	**93**	95	32.72	157.93	**151**	135	47.08	223.74
103	103	59.30	145.88	**97**	101	35.81	165.90	**121**	107	16.12	198.15
103	104	58.94	149.13	**113**	114	46.76	181.82	**100**	99	5.52	192.94
97	98	50.58	144.60	**155**	133	63.08	203.16	**87**	96	-0.64	192.20
90	106	56.94	154.98	**109**	110	36.96	182.13	**80**	90	-8.87	189.41
104	119	67.40	169.65	**115**	99	23.43	173.75	**99**	100	-2.02	201.73
122	137	83.80	190.40	**95**	99	21.60	177.10	**119**	112	7.08	216.31
98	113	57.46	168.54	**83**	95	14.29	175.02	**143**	134	26.95	241.68
93	100	42.30	157.99	**112**	100	16.55	182.56	**133**	106	-4.04	216.21

Residual Plot of 2000 model with 6th polynomial trend line

2000 Triage Category 4 Model



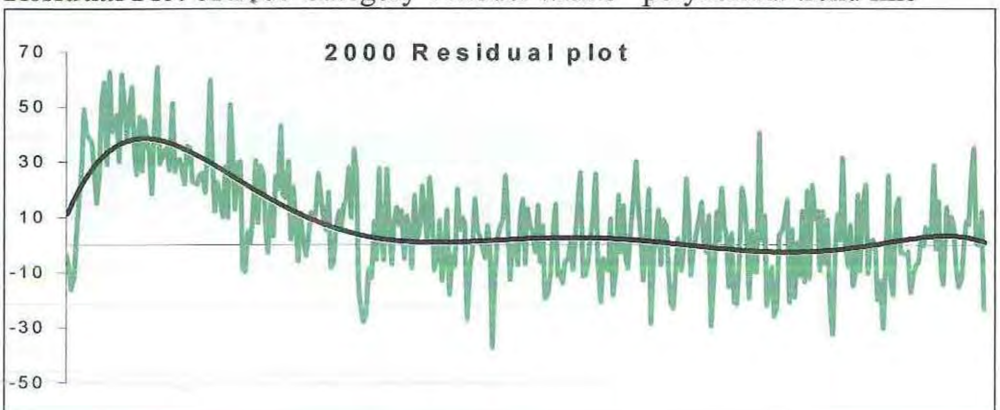
Smoothing Constants
Alpha (level): 0.21
Gamma (trend): 0.03
Delta (seasonal): 0.14

Accuracy Measures
MAPE: 13.912
MSE: 84.09

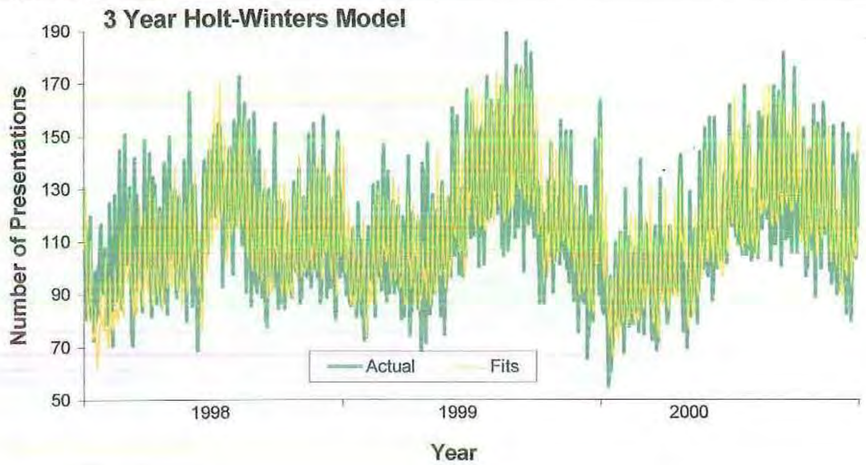
Forecasts for month of December 2000 using above model

Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper
80	72	52.03	91.18	**55**	55	29.15	79.97	**64**	63	29.90	96.10
65	60	40.08	80.13	**53**	53	27.31	79.56	**69**	70	35.83	103.69
54	57	36.21	77.23	**52**	55	28.20	81.89	**67**	60	25.20	94.71
57	55	34.17	76.23	**61**	64	36.09	91.27	**43**	55	19.75	90.94
54	54	32.72	75.89	**77**	69	41.01	97.69	**50**	52	15.66	88.54
38	58	36.11	80.43	**71**	59	30.22	88.44	**43**	52	15.07	89.66
55	65	42.51	88.04	**64**	55	25.42	85.19	**63**	54	15.95	92.25
55	72	48.19	94.98	**42**	55	23.84	85.19	**62**	62	23.21	101.24
56	60	35.65	83.74	**52**	53	21.68	84.63	**77**	69	28.80	108.56
54	55	30.75	80.19	**58**	54	22.09	86.66	**74**	60	19.11	100.62

Residual Plot of 2000 Category 4 model with 6th polynomial trend line



3 Year All Triage Model



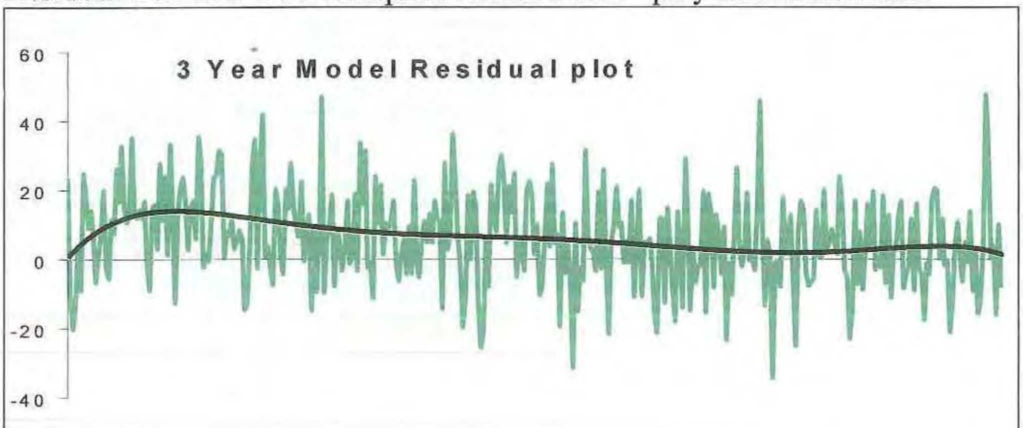
Smoothing Constants
Alpha (level): 0.25
Gamma (trend): 0.01
Delta (seasonal): 0.14

Accuracy Measures
MAPE: 10.02
MSE: 188.71

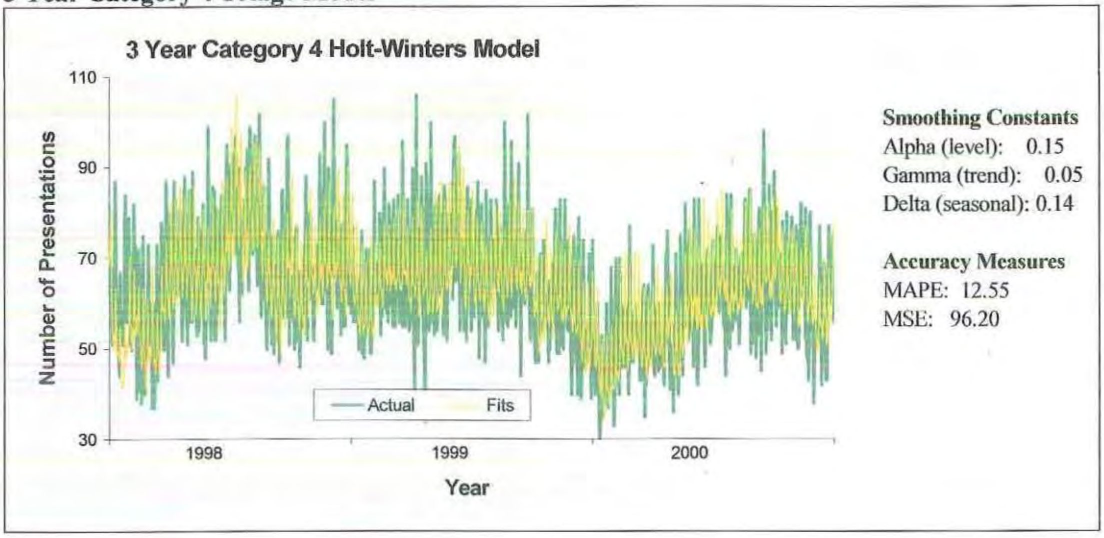
Forecasts for month of December 2000 using above model

Actual	Forecast	95% CI		Actual	Forecast	95% CI		Actual	Forecast	95% CI	
		Lower	Upper			Lower	Upper			Lower	Upper
90	107	80.17	134.54	**109**	115	76.91	152.71	**80**	97	45.51	148.99
104	120	91.79	147.80	**115**	103	63.91	142.29	**99**	106	52.44	158.85
122	139	109.73	167.52	**95**	103	62.09	143.11	**119**	118	63.53	172.90
98	116	86.59	146.29	**83**	99	57.48	141.18	**143**	140	84.20	196.54
93	104	72.76	134.49	**112**	105	61.51	147.93	**133**	114	56.45	171.78
92	103	71.24	135.10	**121**	118	73.54	162.73	**140**	103	43.94	162.28
93	99	66.28	132.37	**151**	140	93.64	185.62	**110**	99	38.53	159.89
97	105	70.76	139.17	**121**	114	66.62	161.44	**104**	95	32.97	157.37
113	118	82.61	153.41	**100**	104	55.44	153.12	**131**	105	41.47	168.91
155	137	100.46	173.73	**87**	101	51.03	151.59	**116**	119	53.37	183.87

Residual Plot of 3 Year Complete Model with 6th polynomial trend line



3 Year Category 4 Triage Model

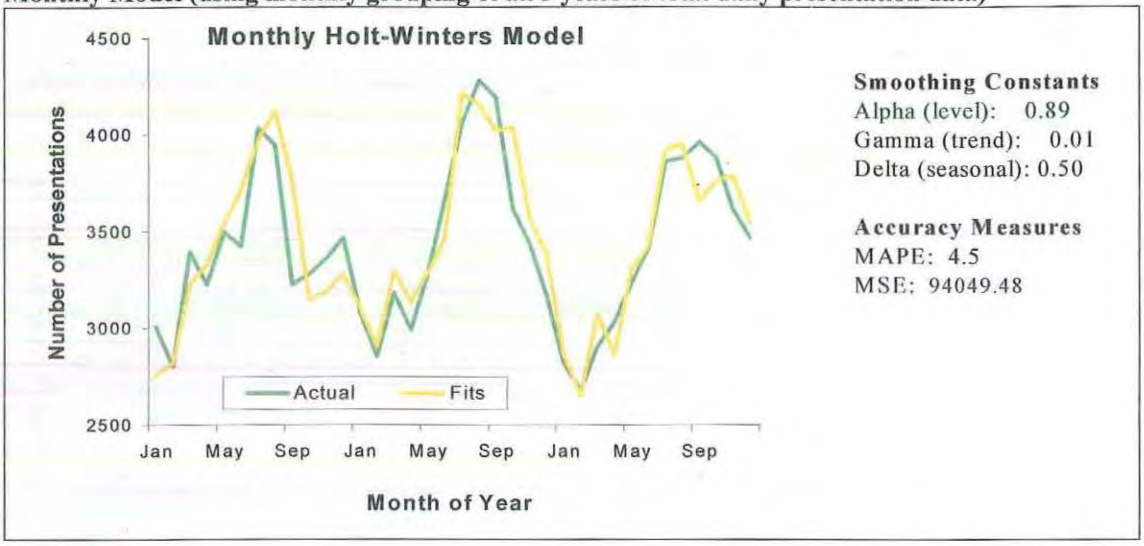


Forecasts for month of December 2000 using above model

Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper
38	59	40.25	78.25	**71**	60	38.44	81.97	**43**	53	27.65	78.63
55	66	47.16	85.60	**64**	56	33.98	78.18	**63**	55	29.05	80.86
55	73	53.42	92.33	**42**	55	32.90	77.80	**62**	63	36.84	89.47
56	61	41.10	80.50	**52**	54	31.11	76.71	**77**	70	43.06	96.54
54	56	36.40	76.32	**58**	55	32.00	78.33	**74**	61	33.88	88.21
55	55	35.20	75.66	**64**	64	40.35	87.41	**71**	54	26.35	81.54
53	54	33.64	74.68	**69**	71	46.97	94.79	**65**	52	24.02	80.09
52	56	35.00	76.63	**67**	61	36.79	85.38	**60**	51	22.91	79.85
61	65	43.40	85.64	**43**	56	31.68	81.06	**67**	56	26.66	84.49
77	70	48.85	91.72	**50**	53	27.85	78.02	**56**	62	33.00	91.73

Residual Plot of 3 Year Category 4 Model with 6th polynomial trend line

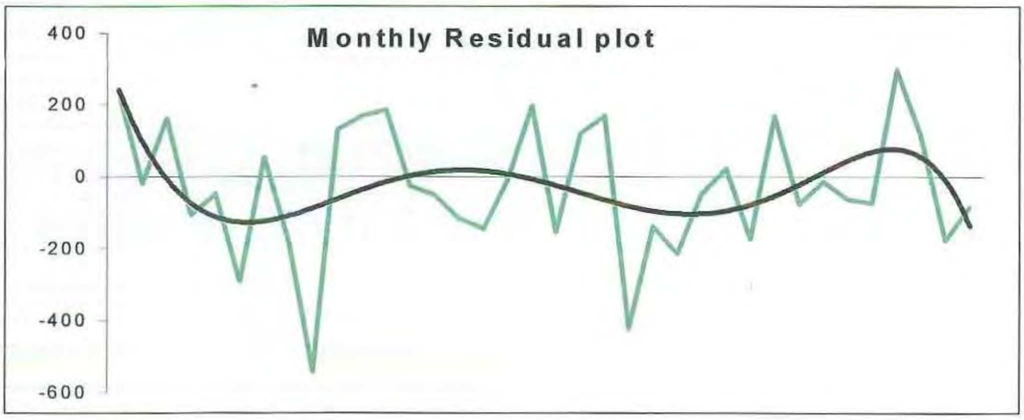
Monthly Model (using monthly grouping of all 3 years of total daily presentation data)



Forecasts for month of September - December 2000 using above model

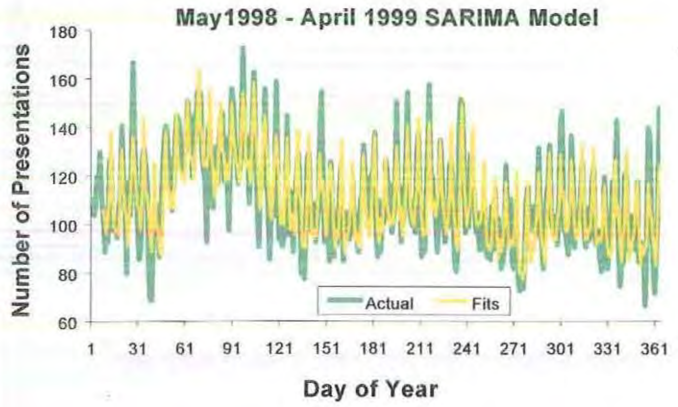
Actual	Forecast	Lower CI	Upper CI
3963	3663	3317.74	4008.27
3883	3509	3018.17	3999.29
3614	3434	2789.03	4079.2
3470	3357	2553.67	4160.73

Residual Plot of Monthly Model with 6^{th} polynomial trend line



APPENDIX D: Box-Jenkins Models

1998/ 1999 Complete Triage model



ASTSA Statistics

p	d	q	P	D	Q	s	AIC	
0	1	1	1	1	1	7	6.2849	
0	1	1	2	1	1	7	6.2780	
1	1	1	1	1	1	7	6.2370	
1	1	1	2	1	1	7	6.2138	best model

Minitab: Final Estimates of Parameters

Type	Coef	StDev	T	P
AR 1	0.3065	0.0610	5.02	0.000
SAR 7	0.0560	0.0584	0.96	0.338
MA 1	0.9018	0.0278	32.43	0.000
SMA 7	0.9613	0.0221	43.56	0.000
Constant	-0.002004	0.003982	-0.50	0.615

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

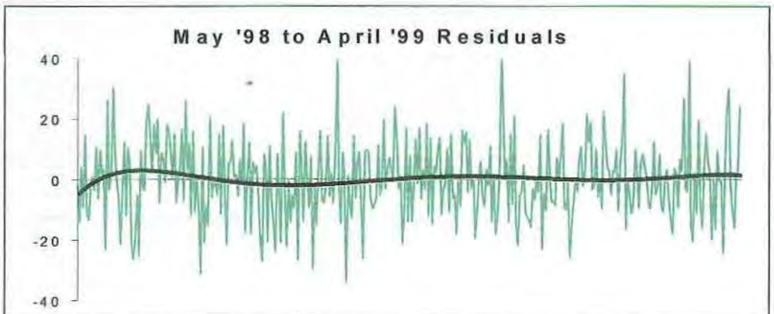
Lag	12	24	36	48
Chi-Square	7.6	16.1	22.6	34.3
DF	7	19	31	43
P-Value	0.371	0.649	0.863	0.825

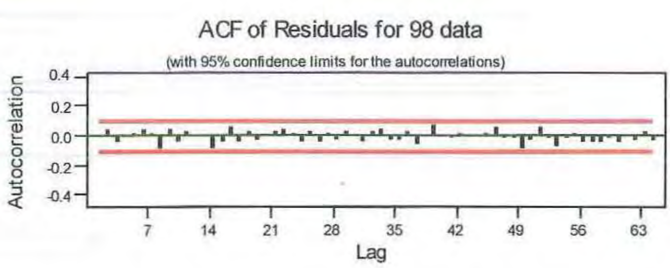
Prediction Model: $(1 - 0.31B)(1 - 0.056B^7)\nabla_1 \nabla_7 x(t) = (1 - 0.89B)(1 - 0.96 B^7) w(t)$

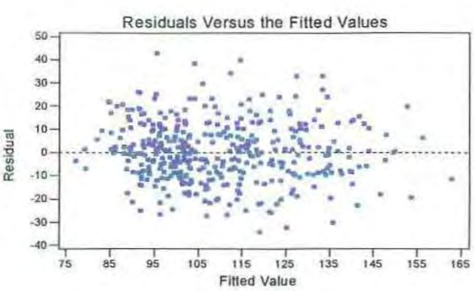
Forecasts for month of May 1999 using above model

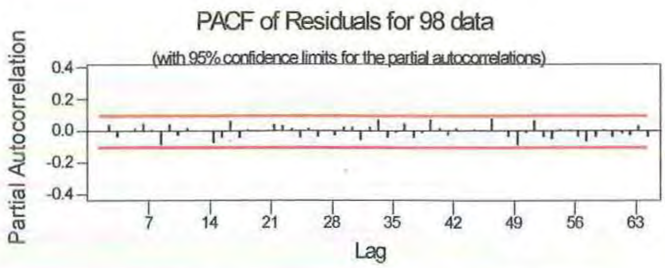
Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper
72	88	62.36	112.86	**128**	123	92.97	153.64	**97**	89	56.35	122.52
86	90	62.89	117.37	**110**	104	73.29	134.47	**90**	87	53.67	120.62
104	104	75.83	131.45	**98**	91	60.03	121.71	**82**	90	55.99	123.55
148	125	96.70	152.95	**91**	90	59.12	121.30	**128**	102	67.91	136.01
97	106	77.88	134.63	**109**	88	56.45	119.43	**133**	122	87.35	155.98
92	92	63.54	120.75	**117**	91	58.77	122.36	**120**	102	67.61	136.76
83	91	61.85	119.51	**122**	103	70.70	134.82	**102**	89	54.40	124.06
94	89	59.17	118.06	**106**	122	90.15	154.80	**75**	89	53.53	123.70
86	91	61.45	121.03	**115**	103	70.42	135.58	**109**	86	50.85	121.80
107	104	73.43	133.58	**104**	90	57.21	122.88	**92**	89	53.17	124.73

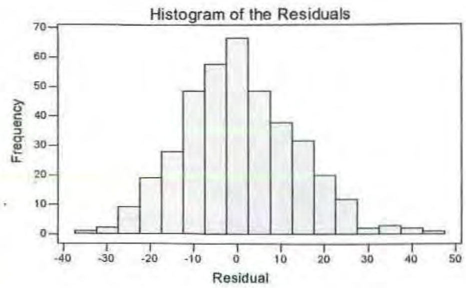
Residual Analysis for 1998/ 1998 model



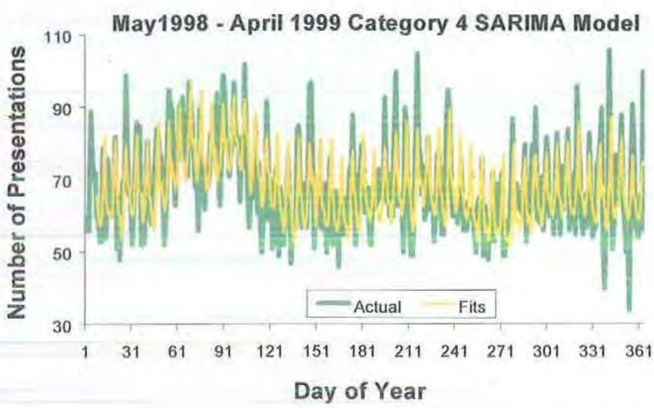








1998/ 1999 Triage Category 4 model



ASTSA Statistics

p	d	q	P	D	Q	s	AIC	
1	1	1	1	1	1	7	5.7773	
1	1	1	1	1	2	7	5.7602	
1	1	1	1	1	3	7	5.7624	
2	1	1	1	1	1	7	5.7420	
2	1	1	1	1	2	7	5.7302	best model
2	1	1	1	1	3	7	5.7336	

Minitab: Final Estimates of Parameters

Type	Coef	StDev	T	P
AR 1	0.1434	0.0601	2.39	0.018
SAR 7	-0.0394	0.0574	-0.69	0.493
MA 1	0.9081	0.0243	37.31	0.000
SMA 7	0.9579	0.0236	40.55	0.000
Constant	-0.001732	0.003067	-0.56	0.573

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

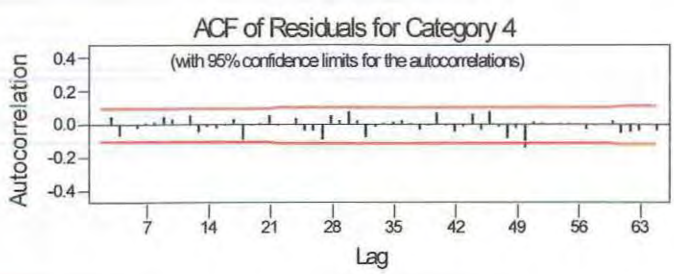
Lag	12	24	36	48
Chi-Square	8.5	14.1	25.3	36.8
DF	7	19	31	43
P-Value	0.292	0.775	0.752	0.735

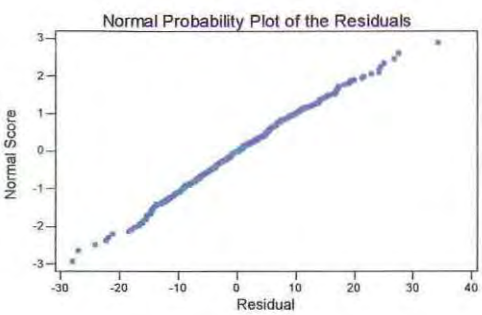
Prediction Model: $(1 - 0.14B)(1 + 0.04B^7)\nabla_1 \nabla_7 x(t) = (1 - 0.91B)(1 - 0.96 B^7) w(t)$

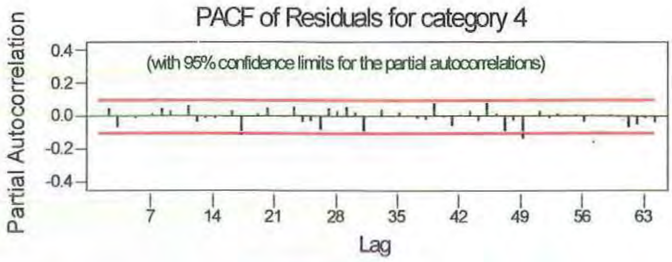
Forecasts for month of May 1999 using above model

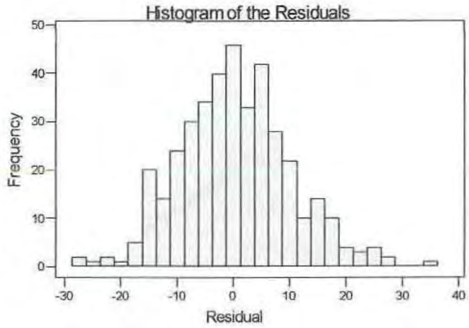
Actual	Forecast	95% CI		Actual	Forecast	95% CI		Actual	Forecast	95% CI	
		Lower	Upper			Lower	Upper			Lower	Upper
54	59	39.50	78.29	**82**	79	57.95	99.80	**57**	59	36.41	80.60
66	61	41.30	81.14	**68**	65	43.55	85.61	**61**	58	35.79	80.37
56	72	52.30	92.44	**54**	59	38.37	80.63	**53**	59	36.74	81.57
100	79	58.58	98.94	**57**	59	37.63	80.11	**86**	71	48.72	93.79
71	64	43.96	84.54	**64**	58	37.02	79.89	**86**	78	55.46	100.75
63	60	39.59	80.38	**70**	60	37.98	81.09	**74**	64	41.05	86.56
60	59	38.88	79.88	**84**	72	49.97	93.30	**77**	59	35.88	81.61
63	59	38.22	79.44	**62**	78	56.71	100.26	**52**	58	35.14	81.09
53	60	39.14	80.58	**64**	64	42.31	86.07	**63**	58	34.51	80.87
79	72	51.18	92.82	**80**	59	37.14	81.12	**71**	59	35.46	82.07

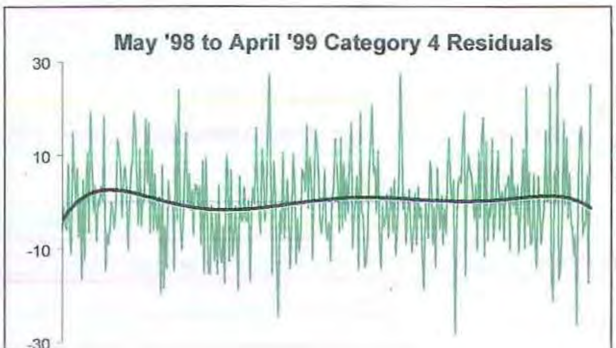
Residual Analysis for 1998 /1999 model

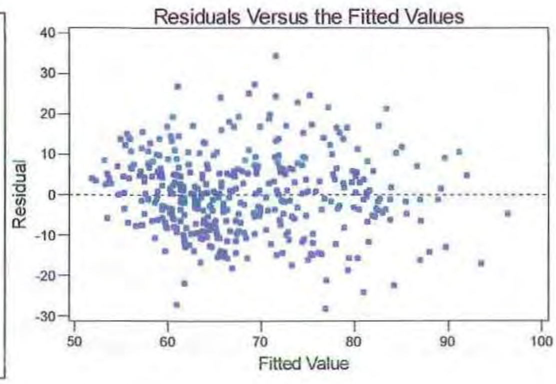




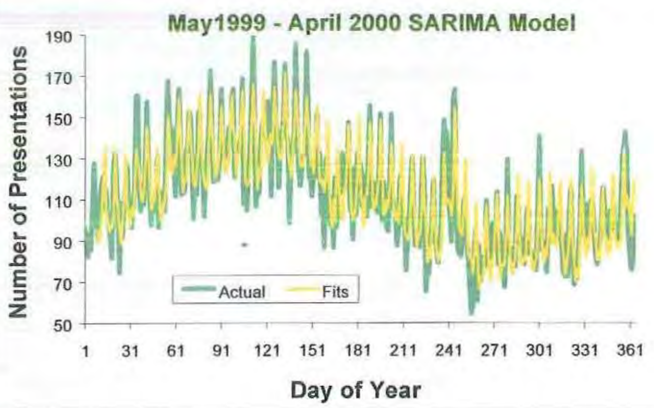








1999 / 2000 Complete model



ASTSA Statistics

p	d	q	P	D	Q	s	AIC
1	1	1	1	1	1	7	6.3653
1	1	1	1	1	2	7	6.2936
1	1	1	2	1	1	7	6.2973
1	1	1	2	1	2	7	6.2821 best model

Minitab: Final Estimates of Parameters

Type	Coef	StDev	T	P
AR 1	0.2446	0.0698	3.50	0.001
SAR 7	-0.0299	0.0582	-0.51	0.607
SAR 14	-0.1327	0.0564	-2.35	0.019
MA 1	0.8400	0.0401	20.95	0.000
SMA 7	0.9598	0.0206	46.49	0.000
Constant	-0.009021	0.006645	-1.36	0.176

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

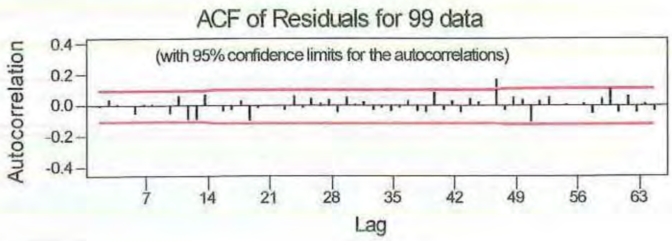
Lag	12	24	36	48
Chi-Square	11.4	19.7	27.8	52.9
DF	6	18	30	42
P-Value	0.076	0.351	0.581	0.121

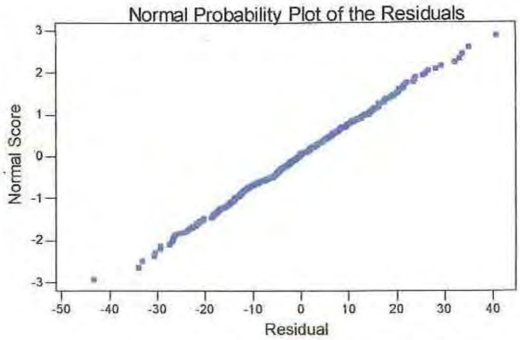
Prediction Model: $(1 - 0.24B)(1 + 0.03B^7)(1 + 0.13B^{14})\nabla_1 \nabla_7 x(t) = (1 - 0.84B)(1 - 0.96 B^7) w(t)$

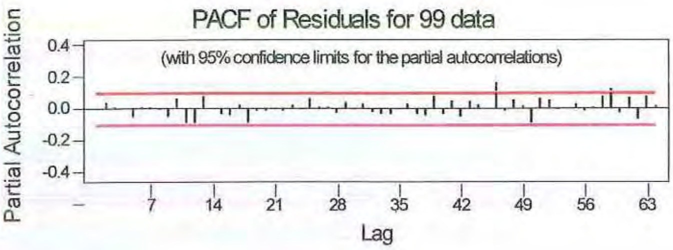
Forecasts for month of May 2000 using above model

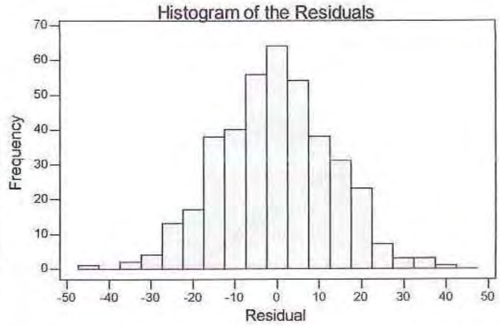
Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper
102	119	92.74	145.46	**86**	84	50.69	117.49	**109**	101	63.61	137.70
82	100	71.11	127.98	**97**	86	52.53	120.27	**144**	118	80.48	155.62
89	91	62.14	120.63	**92**	88	53.55	122.23	**123**	97	58.75	134.70
70	86	56.52	116.19	**108**	105	70.67	140.27	**105**	88	49.69	126.41
99	86	56.12	116.85	**115**	120	85.35	155.24	**118**	82	42.82	120.29
83	86	55.33	117.09	**110**	98	63.03	133.53	**126**	82	42.69	120.90
85	102	70.18	132.94	**99**	90	53.97	125.19	**100**	82	42.64	121.58
129	120	88.19	152.04	**79**	83	47.41	119.35	**116**	98	58.27	137.93
101	94	61.97	126.82	**108**	84	47.53	120.20	**153**	116	75.50	156.29
82	86	53.17	119.00	**92**	84	47.71	121.09	**113**	94	53.23	134.87

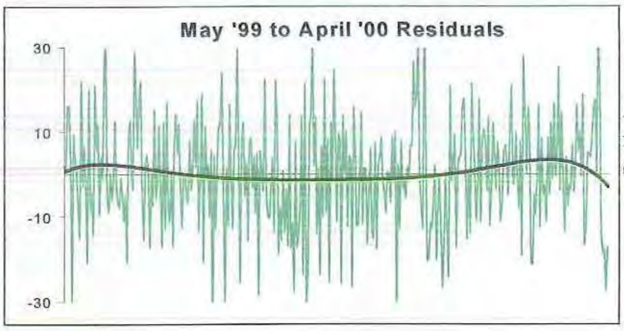
Residual Analysis for 1999 / 2000 model

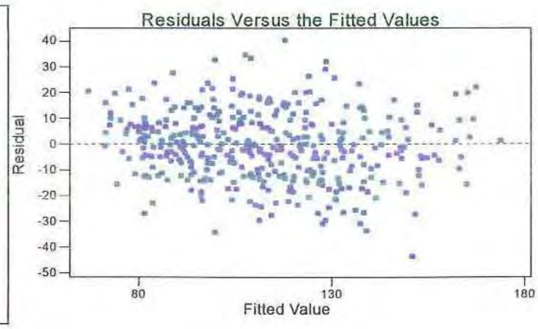




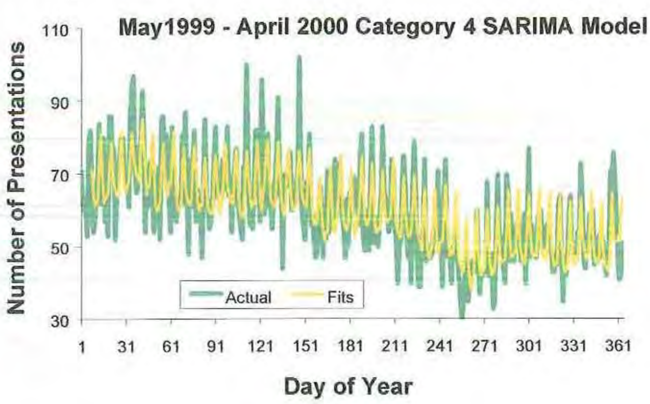








1999 / 2000 Triage Category 4 model



ASTSA Statistics

p	d	q	P	D	Q	s	AIC
1	0	1	1	1	1	7	5.4509
1	0	1	1	1	2	7	5.3861
1	0	1	1	1	3	7	5.3918
1	0	1	2	1	1	7	5.4188
1	0	1	2	1	2	7	5.3849
1	0	1	2	1	3	7	5.3718 best model

Minitab: Final Estimates of Parameters

Type	Coef	StDev	T	P
AR 1	0.9187	0.0460	19.98	0.000
SAR 7	-0.5643	0.5729	-0.99	0.325
SAR 14	-0.0334	0.0701	-0.48	0.634
MA 1	0.7703	0.0735	10.48	0.000
SMA 7	0.3802	0.5707	0.67	0.506
SMA 14	0.5752	0.5571	1.03	0.303
Constant	-0.062035	0.007785	-7.97	0.000

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

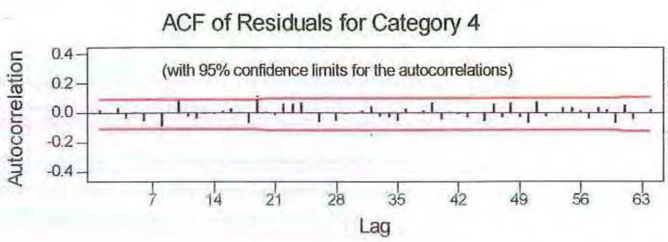
Lag	12	24	36	48
Chi-Square	6.4	23.1	30.1	40.3
DF	5	17	29	41
P-Value	0.266	0.145	0.407	0.503

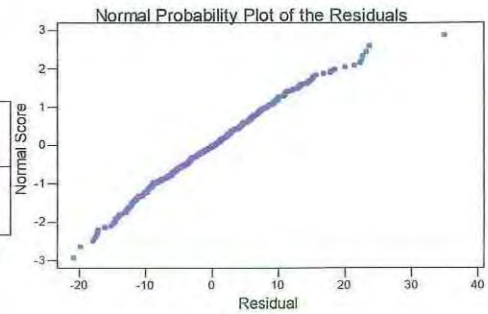
Prediction Model: $(1 - 0.92B)(1 + 0.56B^7)(1 + 0.03B^{14})\nabla_7 x(t) = (1 - 0.77B)(1 - 0.38 B^7)(1 + 0.57B^{14}) w(t)$

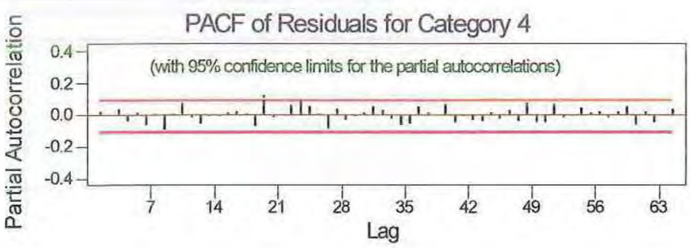
Forecasts for month of May 2000 using above model

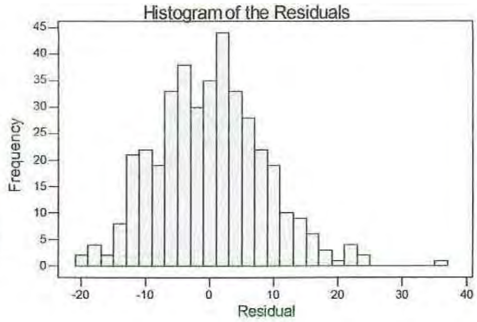
Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper
51	63	46.34	79.58	**40**	42	24.66	60.01	**60**	51	33.19	68.90
43	52	35.28	68.88	**47**	43	25.48	60.92	**82**	59	41.06	76.87
47	47	30.02	63.92	**54**	44	26.03	61.53	**57**	48	29.62	65.45
36	45	28.22	62.38	**51**	54	36.07	71.62	**59**	43	24.87	60.72
55	45	27.32	61.69	**75**	60	42.23	77.80	**60**	41	23.28	59.14
48	45	27.25	61.80	**64**	49	31.51	67.11	**73**	42	23.61	59.48
47	52	34.97	69.67	**57**	44	26.58	62.21	**52**	42	24.01	59.89
71	61	43.21	78.24	**44**	43	24.81	60.46	**58**	51	33.41	69.30
46	49	30.96	66.12	**62**	42	24.47	60.15	**76**	58	40.45	76.35
46	44	26.23	61.50	**55**	42	24.57	60.27	**55**	47	29.41	65.32

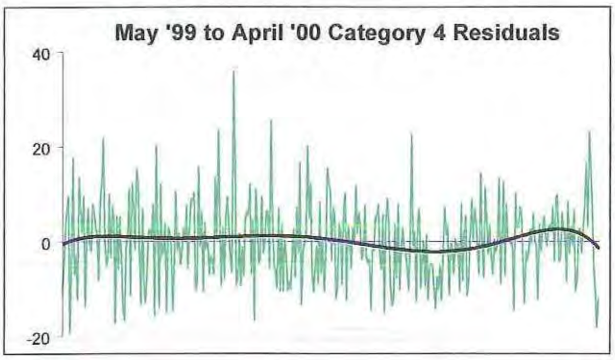
Residual Analysis for 1999 / 2000 model

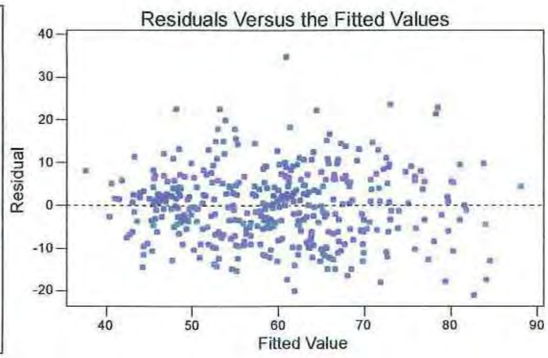




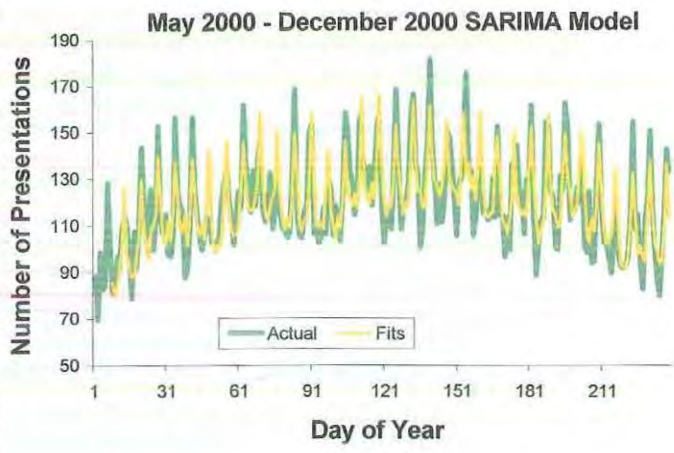








May 2000 / December 200 Complete Triage model



ASTSA Statistics

p	d	q	P	D	Q	s	AIC	
1	0	1	1	1	1	7	6.1889	
1	0	1	1	1	2	7	6.1705	best model
1	1	1	1	1	1	7	6.2596	
1	1	1	1	1	2	7	6.1893	

Minitab: Final Estimates of Parameters

Type	Coef	StDev	T	P
AR 1	0.2553	0.0749	3.41	0.001
SAR 7	-0.8898	2.8896	-0.31	0.758
MA 1	0.9164	0.0341	26.85	0.000
SMA 7	0.0527	2.9059	0.02	0.986
SMA 14	0.8335	2.7443	0.30	0.762
Constant	-0.03200	0.01237	-2.59	0.010

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

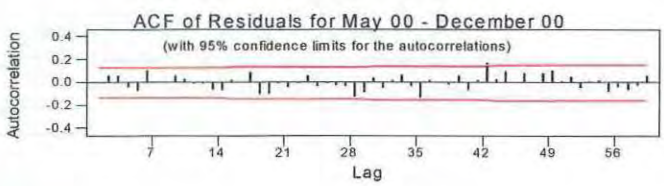
Lag	12	24	36	48
Chi-Square	6.9	18.9	35.5	52.7
DF	6	18	30	42
P-Value	0.330	0.399	0.225	0.125

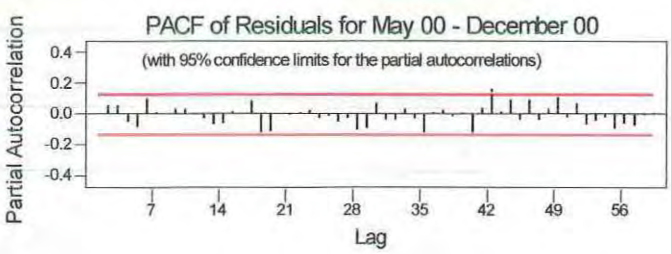
Prediction Model: $(1 - 0.26B)(1 + 0.89B^7)\nabla_1 \nabla_7 x(t) = (1 - 0.92B)(1 - 0.05 B^7)(1 - 0.83 B^{14}) w(t)$

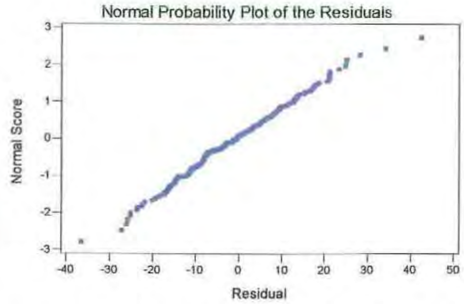
Forecasts for month of December 2000 using above model

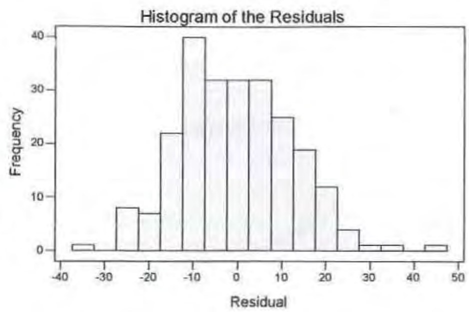
Actual	Forecast	95% CI		Actual	Forecast	95% CI		Actual	Forecast	95% CI	
		Lower	Upper			Lower	Upper			Lower	Upper
104	115	90.64	140.24	**115**	102	73.92	129.92	**99**	99	69.34	129.06
122	140	113.86	166.23	**95**	99	71.06	127.36	**119**	113	82.87	143.28
98	117	90.40	143.45	**83**	97	68.81	125.42	**143**	135	104.65	165.49
93	104	77.24	130.66	**112**	101	72.78	129.70	**133**	111	80.47	141.68
92	101	74.38	128.11	**121**	115	86.54	144.16	**140**	98	66.75	128.32
93	99	71.93	125.96	**151**	137	108.05	166.09	**110**	95	63.82	125.73
97	103	76.28	130.59	**121**	113	83.97	142.36	**104**	93	61.48	123.74
113	117	89.84	144.80	**100**	100	70.55	129.29	**131**	97	65.42	128.03
155	139	111.73	167.08	**87**	97	67.60	126.66	**116**	111	79.09	142.44
109	115	87.61	143.29	**80**	95	65.10	124.49	**144**	132	100.58	164.37

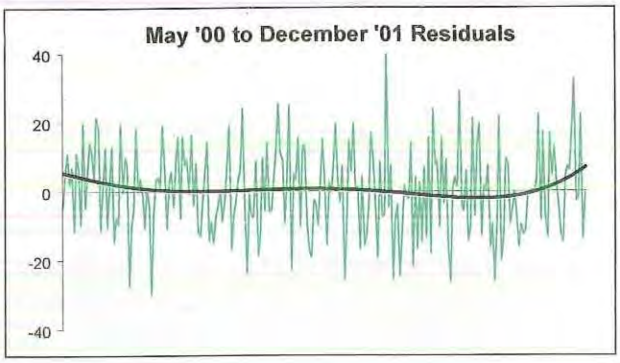
Residual Analysis for May 2000 / December 2000 model

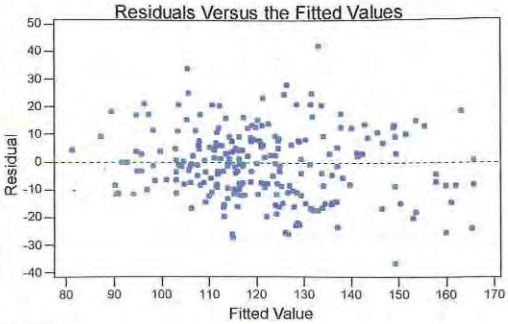




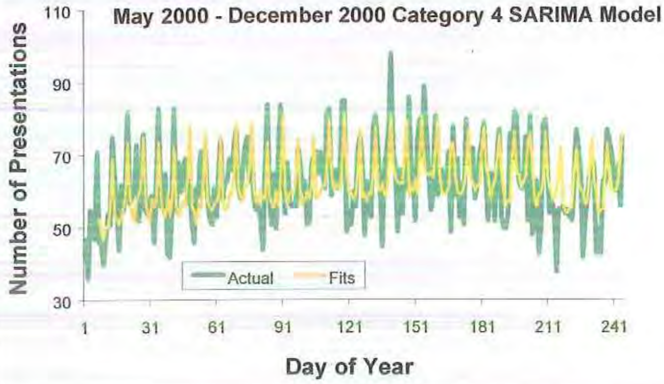








May 2000 / December 2000 Triage Category 4 Model



ASTSA Statistics

p	d	q	P	D	Q	s	AIC
0	0	1	1	1	1	7	5.5420
0	0	1	1	1	2	7	5.5506
1	0	0	1	1	1	7	5.5223
1	0	0	1	1	2	7	5.5246
1	0	1	1	1	1	7	5.4665
1	0	1	1	1	2	7	5.4664 best model

Minitab: Final Estimates of Parameters

Type	Coef	StDev	T	P
AR 1	0.8173	0.1121	7.29	0.000
SAR 7	0.0394	0.0750	0.53	0.600
MA 1	0.6467	0.1468	4.40	0.000
SMA 7	0.9334	0.0341	27.41	0.000
Constant	0.02809	0.01921	1.46	0.145

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

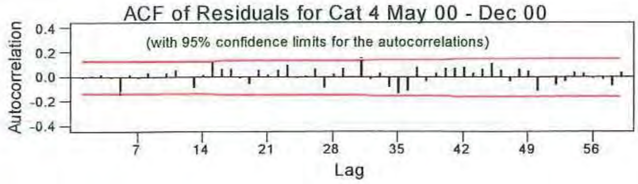
Lag	12	24	36	48
Chi-Square	6.9	19.6	42.6	56.2
DF	7	19	31	43
P-Value	0.442	0.416	0.081	0.086

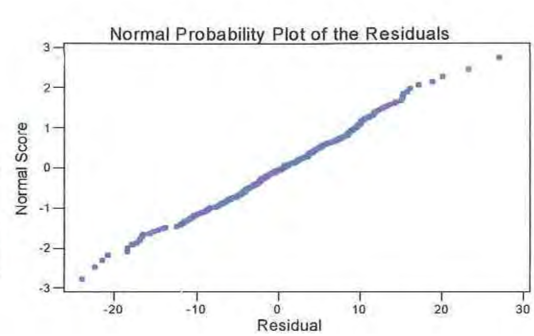
Prediction Model: $(1 - 0.82B)(1 - 0.39B^7)\nabla_7\, x(t) = (1 - 0.65B)(1 - 0.93\, B^7)\, w(t)$

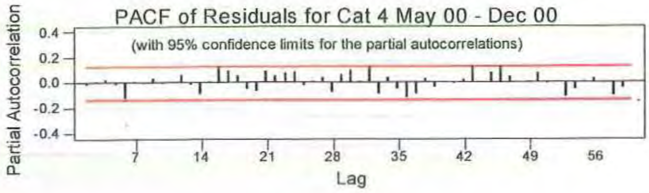
Forecasts for month of December 2001 using above model

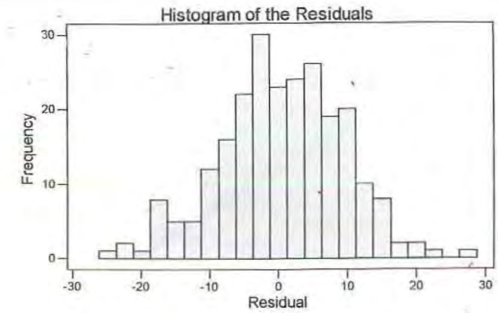
Actual	Forecast	95% CI		Actual	Forecast	95% CI		Actual	Forecast	95% CI	
		Lower	Upper			Lower	Upper			Lower	Upper
55	65	47.78	82.91	**64**	60	41.84	78.88	**63**	61	42.65	79.89
55	74	55.91	91.54	**42**	59	40.64	77.70	**62**	70	51.57	88.91
56	62	43.61	79.58	**52**	59	40.83	77.90	**77**	78	59.34	96.69
54	58	39.94	76.13	**58**	61	42.23	79.30	**74**	65	46.58	83.93
55	57	39.19	75.53	**64**	70	51.22	88.42	**71**	61	42.72	80.07
53	58	39.51	75.94	**69**	78	59.03	96.25	**65**	60	41.40	78.75
52	59	40.45	76.95	**67**	65	46.31	83.54	**60**	60	41.49	78.84
61	69	50.18	87.10	**43**	61	42.48	79.71	**67**	62	42.83	80.18
77	77	58.16	95.13	**50**	60	41.18	78.42	**56**	70	51.74	89.19
71	64	45.57	82.58	**43**	60	41.29	78.53	**75**	78	59.50	96.95

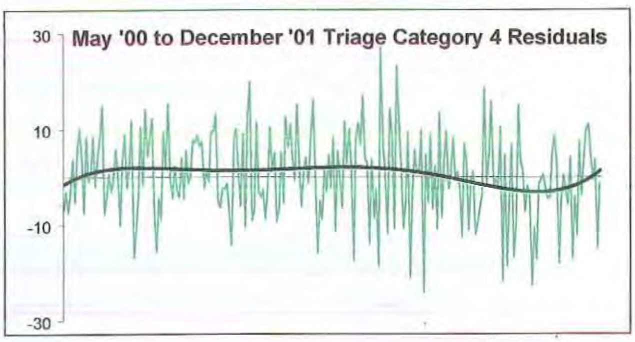
Residual Analysis for 2000 / 2001 model

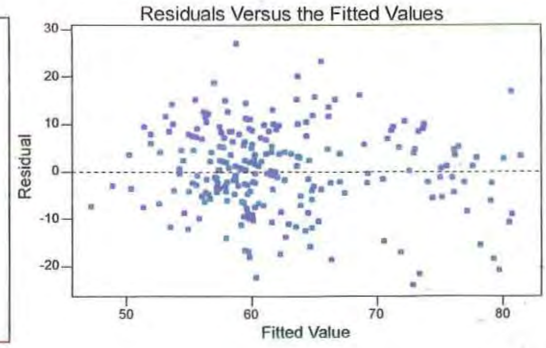




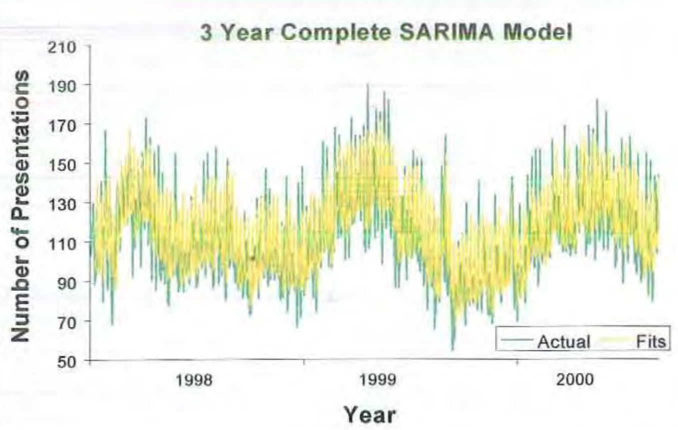








3 Year Complete model



ASTSA Statistics

p	d	q	P	D	Q	s	AIC
1	1	1	1	1	1	7	6.1930
1	1	1	1	1	2	7	6.1698
1	1	1	2	1	1	7	6.1694
1	1	1	2	1	2	7	6.1520 best model

Minitab: Final Estimates of Parameters

Type	Coef	StDev	T	P
AR 1	0.2597	0.0392	6.62	0.000
SAR 7	0.0469	0.0336	1.39	0.164
SAR 14	-0.0734	0.0325	-2.26	0.024
MA 1	0.8732	0.0201	43.51	0.000
SMA 7	0.9778	0.0037	264.12	0.000
Constant	-0.000191	0.001698	-0.11	0.910

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

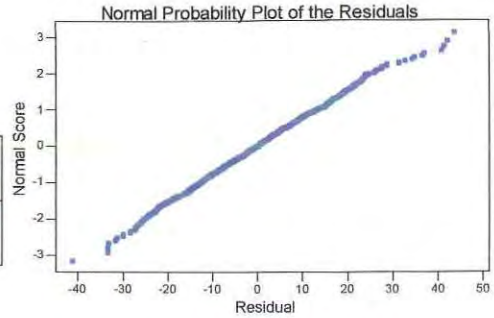
Lag	12	24	36	48
Chi-Square	10.9	21.8	30.2	52.6
DF	6	18	30	42
P-Value	0.091	0.241	0.456	0.127

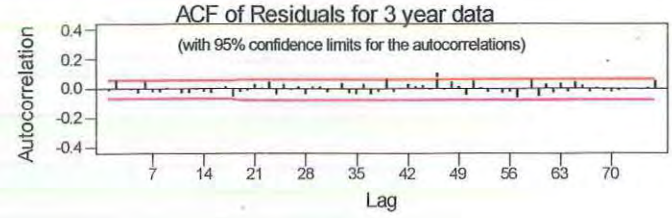
Prediction Model: $(1 - 0.26B)(1 - 0.05B^7)(1 + 0.07B^{14})\nabla_1\nabla_7\, x(t) = (1 - 0.87B)(1 - 0.98\, B^7)w(t)$

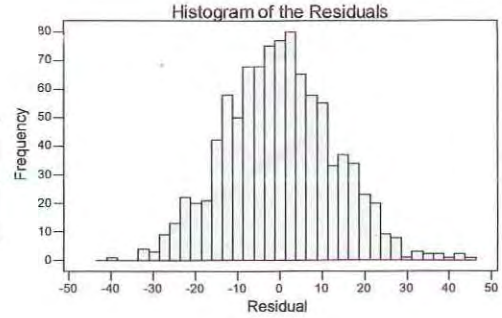
Forecasts for month of December 2000 using above model

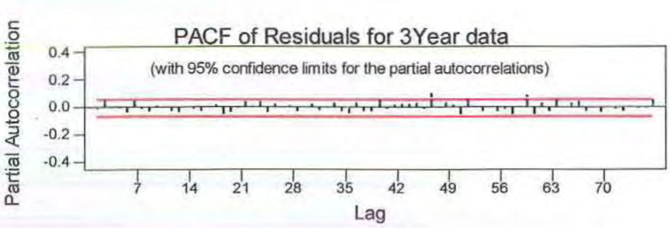
Actual	Forecast	95% CI		Actual	Forecast	95% CI		Actual	Forecast	95% CI	
		Lower	Upper			Lower	Upper			Lower	Upper
55	65	47.78	82.91	**64**	60	41.84	78.88	**63**	61	42.65	79.89
55	74	55.91	91.54	**42**	59	40.64	77.70	**62**	70	51.57	88.91
56	62	43.61	79.58	**52**	59	40.83	77.90	**77**	78	59.34	96.69
54	58	39.94	76.13	**58**	61	42.23	79.30	**74**	65	46.58	83.93
55	57	39.19	75.53	**64**	70	51.22	88.42	**71**	61	42.72	80.07
53	58	39.51	75.94	**69**	78	59.03	96.25	**65**	60	41.40	78.75
52	59	40.45	76.95	**67**	65	46.31	83.54	**60**	60	41.49	78.84
61	69	50.18	87.10	**43**	61	42.48	79.71	**67**	62	42.83	80.18
77	77	58.16	95.13	**50**	60	41.18	78.42	**56**	70	51.74	89.19
71	64	45.57	82.58	**43**	60	41.29	78.53	**75**	78	59.50	96.95

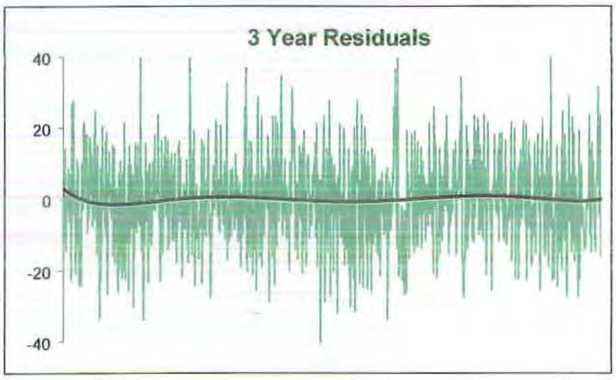
Residual Analysis for 3 Year complete model

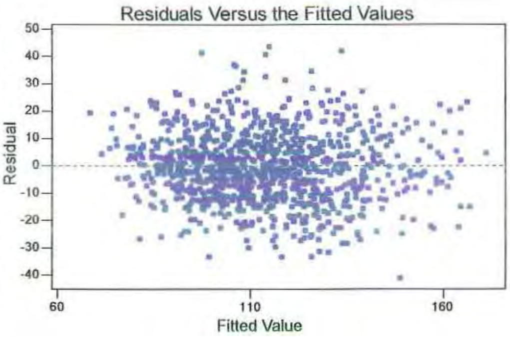




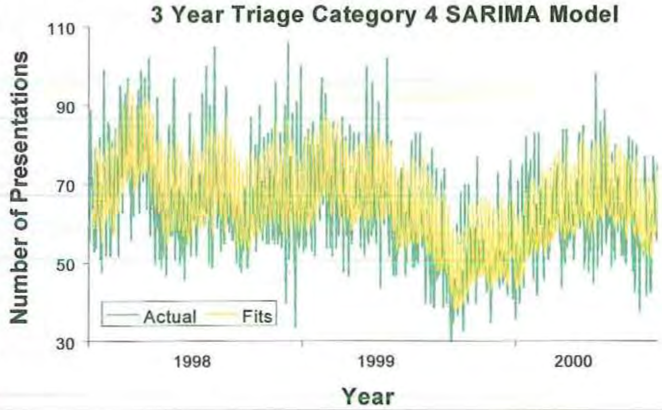








3 Year Triage Category 4 model



ASTSA Statistics

p	d	q	P	D	Q	s	AIC	
1	1	1	1	1	1	7	5.5536	
1	1	1	1	1	2	7	5.5283	
1	1	1	2	1	1	7	5.5209	
1	1	1	2	1	2	7	5.4804	best model

Minitab: Final Estimates of Parameters

Type	Coef	StDev	T	P
AR 1	0.1177	0.0356	3.30	0.001
SAR 7	0.0051	0.0333	0.15	0.878
MA 1	0.9127	0.0140	64.96	0.000
SMA 7	0.9802	0.0023	427.46	0.000
Constant	-0.0000643	0.0007715	-0.08	0.934

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

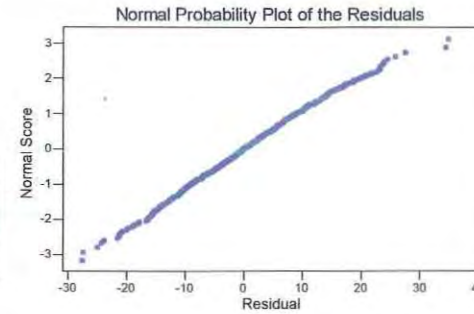
Lag	12	24	36	48
Chi-Square	6.7	18.0	26.7	37.3
DF	7	19	31	43
P-Value	0.457	0.524	0.685	0.715

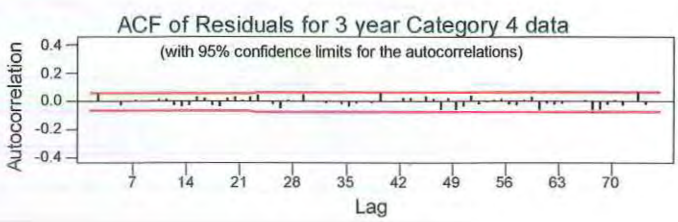
Prediction Model: $(1 - 0.12B)(1 - 0.01B^7)\nabla_1\nabla_7\, x(t) = (1 - 0.91B)(1 - 0.98\, B^7)\, w(t)$

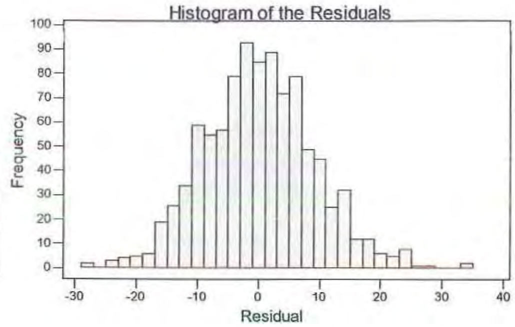
Forecasts for month of December 2000 using above model

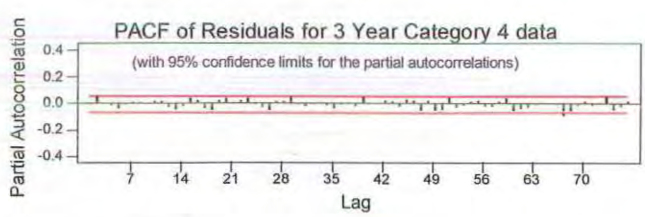
Actual	Forecast	95% CI		Actual	Forecast	95% CI		Actual	Forecast	95% CI	
		Lower	Upper			Lower	Upper			Lower	Upper
55	65	47.78	82.91	**64**	60	41.84	78.88	**63**	61	42.65	79.89
55	74	55.91	91.54	**42**	59	40.64	77.70	**62**	70	51.57	88.91
56	62	43.61	79.58	**52**	59	40.83	77.90	**77**	78	59.34	96.69
54	58	39.94	76.13	**58**	61	42.23	79.30	**74**	65	46.58	83.93
55	57	39.19	75.53	**64**	70	51.22	88.42	**71**	61	42.72	80.07
53	58	39.51	75.94	**69**	78	59.03	96.25	**65**	60	41.40	78.75
52	59	40.45	76.95	**67**	65	46.31	83.54	**60**	60	41.49	78.84
61	69	50.18	87.10	**43**	61	42.48	79.71	**67**	62	42.83	80.18
77	77	58.16	95.13	**50**	60	41.18	78.42	**56**	70	51.74	89.19
71	64	45.57	82.58	**43**	60	41.29	78.53	**75**	78	59.50	96.95

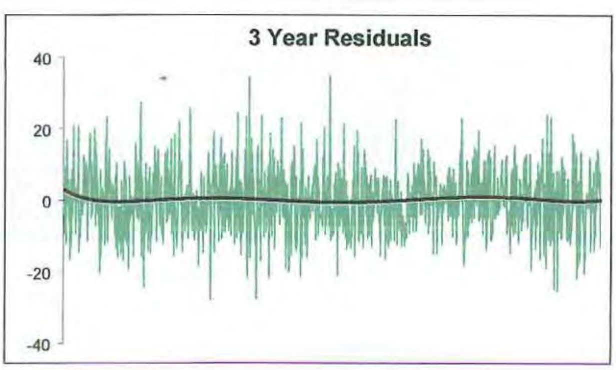
Residual Analysis for 3 year Category 4 model

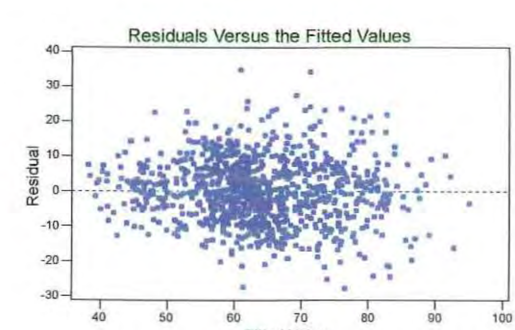






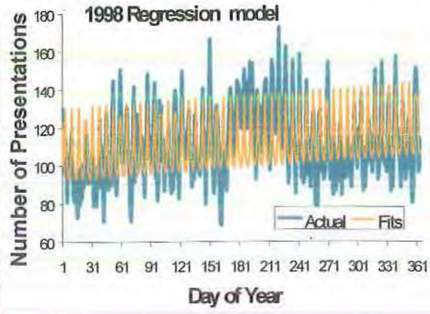






APPENDIX E: Regression Models

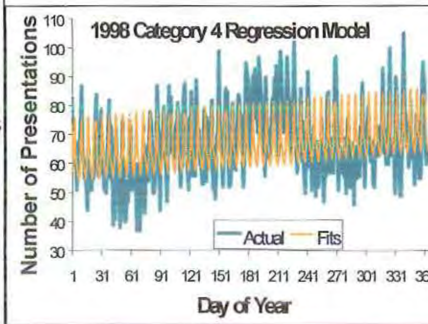
Regression Models with Day variable only



Accuracy Measures

Multiple R: 0.6612
R Square: 0.4372
Adjusted R Square: 0.4245
Standard Error: 15.1699
Observations: 365

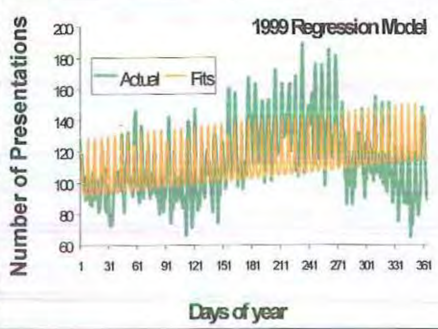
MSE: 224.45



Accuracy Measures

Multiple R: 0.5969
R Square: 0.3562
Adjusted R Square: 0.3418
Standard Error: 10.7192
Observations: 365

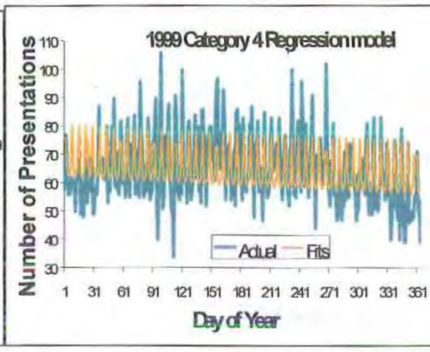
MSE: 112.07



Accuracy Measures

Multiple R: 0.6155
R Square: 0.3740
Adjusted R Square: 0.3599
Standard Error: 18.4239
Observations: 365

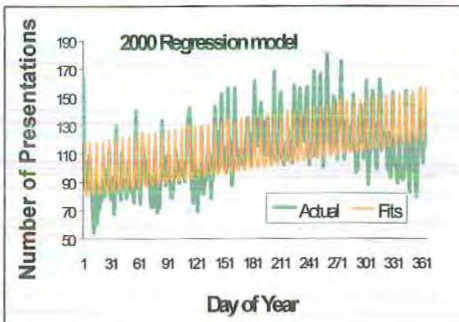
MSE: 331.07



Accuracy Measures

Multiple R: 0.5610
R Square: 0.3147
Adjusted R Square: 0.2993
Standard Error: 10.0386
Observations: 365

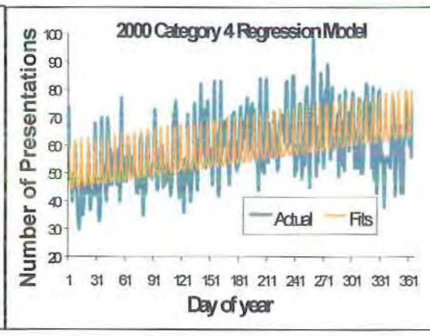
MSE: 98.29



Accuracy Measures

Multiple R: 0.7469
R Square: 0.5578
Adjusted R Square: 0.5479
Standard Error: 15.4387
Observations: 366

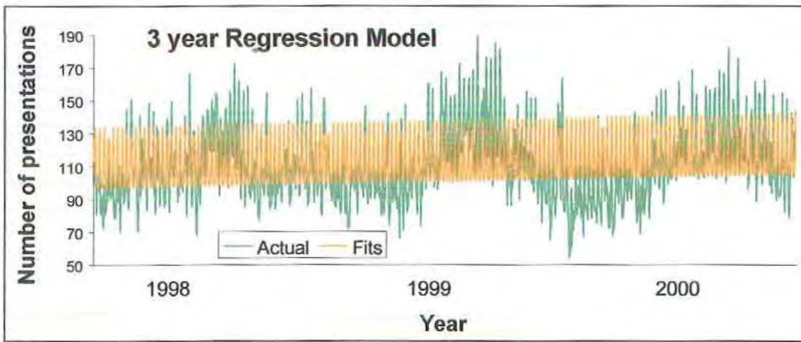
MSE: 232.49



Accuracy Measures

Multiple R: 0.6821
R Square: 0.4653
Adjusted R Square: 0.4633
Standard Error: 8.7648
Observations: 366

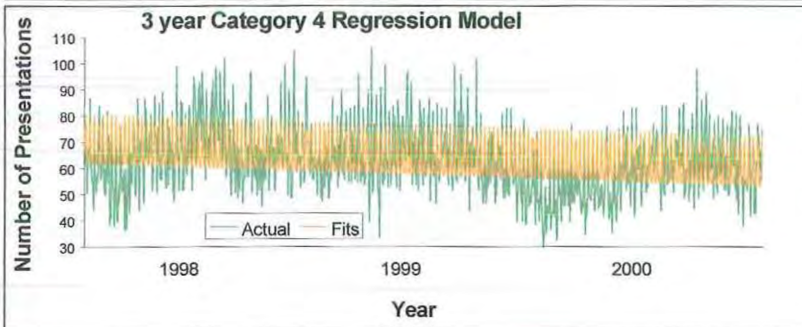
MSE: 74.93



Accuracy Measures

Multiple R: 0.5807
R Square: 0.3372
Adjusted R Square: 0.3323
Standard Error: 18.0386
Observations: 1096

MSE: 322.72

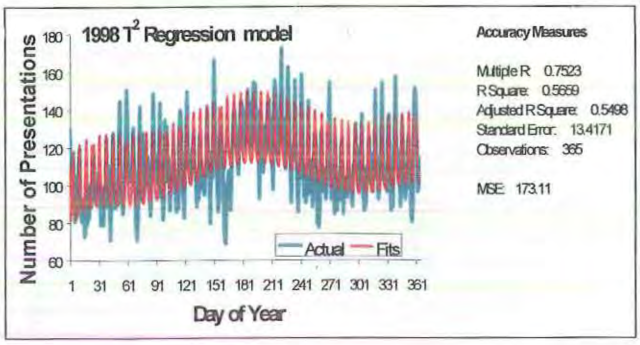


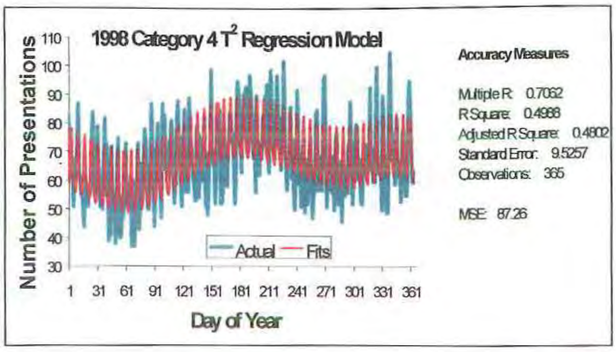
Accuracy Measures

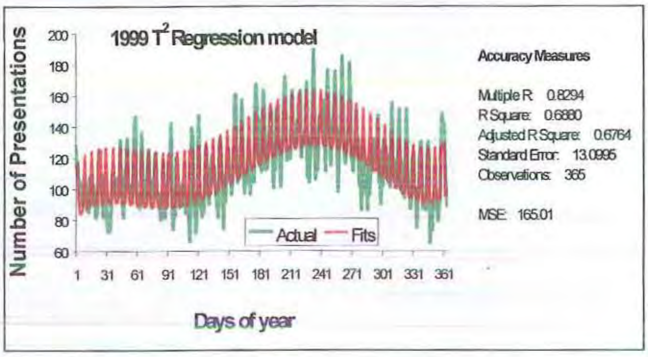
Multiple R: 0.5454
R Square: 0.2975
Adjusted R Square: 0.2943
Standard Error: 10.8299
Observations: 1096

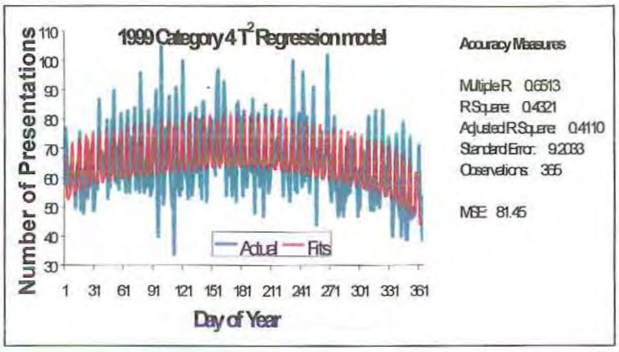
MSE: 116.3246

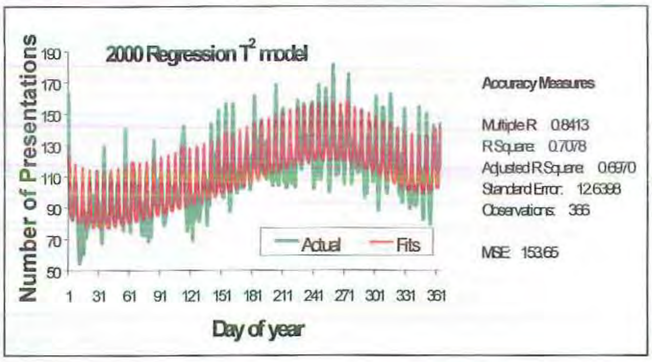
Regression models with T^2 variable

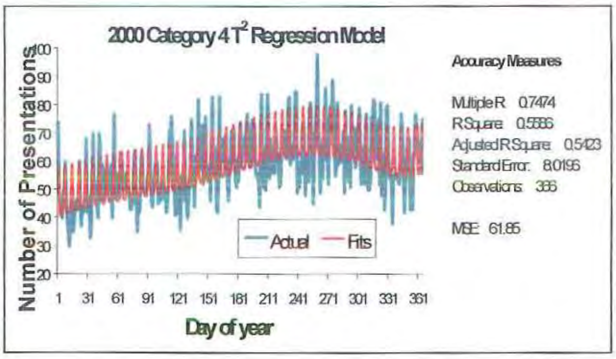


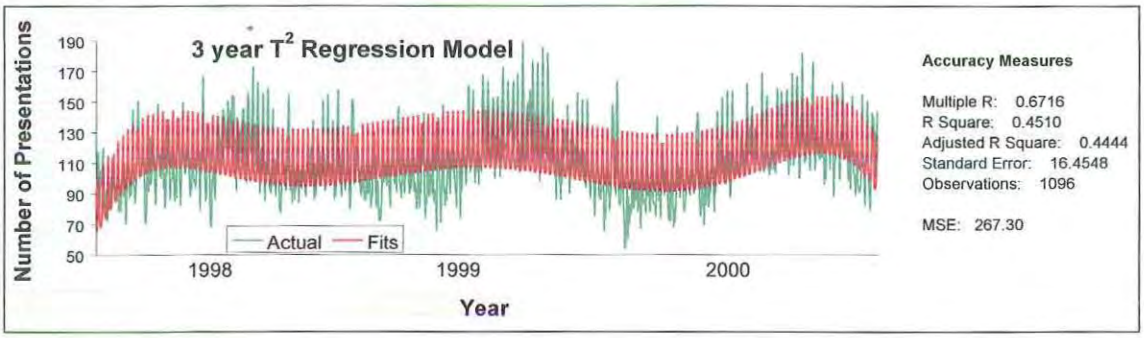


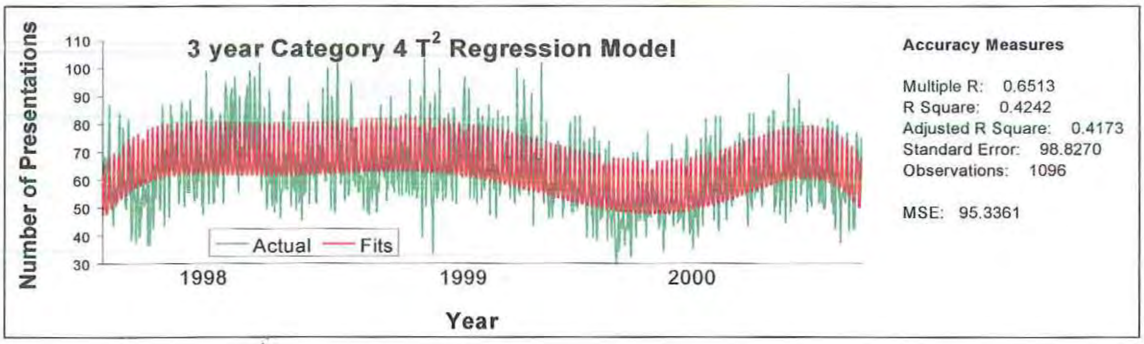




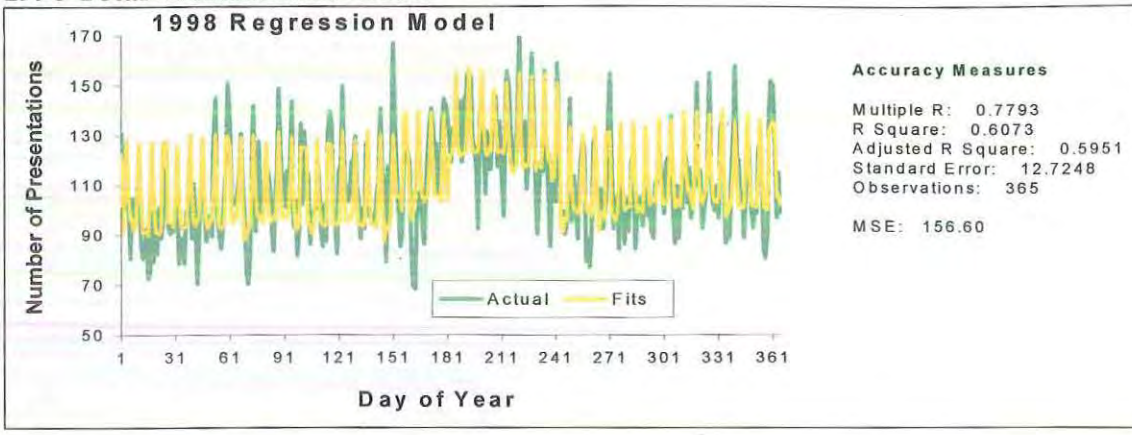








1998 Total Presentations model



Best Subsets Regression

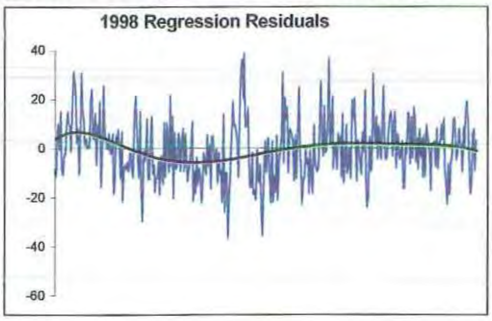
Vars	R-Sq	Adj. R-Sq	C-p	s	Time	Mon	Sat	Sun	Holiday	Apr	Jun	Jul	Aug	Sep	Dec	Maximum	Rainfall
8	59.2	58.3	18.4	12.911	X	X	X	X	X		X	X	X				
8	58.7	57.7	23.4	12.999	X	X	X	X	X			X	X				X
9	60.1	59.0	13.0	12.798	X	X	X	X	X		X	X	X			X	
9	60.0	58.9	13.8	12.813	X	X	X	X	X		X	X	X				X
10	60.5	59.4	11.2	12.749	X	X	X	X	X		X	X	X			X	X
10	60.3	59.2	12.5	12.771	X	X	X	X	X		X	X	X		X	X	
11	60.7	59.5	10.9	12.725	X	X	X	X	X		X	X	X		X	X	X
11	60.5	59.3	12.6	12.756	X	X	X	X	X	X	X	X	X			X	X
12	60.8	59.5	12.2	12.730	X	X	X	X	X	X	X	X	X		X	X	X
12	60.7	59.4	12.8	12.742	X	X	X	X	X		X	X	X	X	X	X	X
13	60.8	59.4	14.0	12.745	X	X	X	X	X	X	X	X	X	X	X	X	X

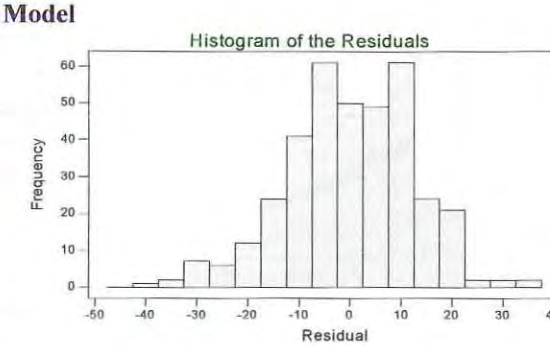
Prediction Model: Triage Number = 80.1 + 0.0436 Time + 10.2 Mon + 14.5 Sat + 34.4 Sun + 31.2 Holiday + 11.3 Jun + 28.1 Jul + 21.3 Aug - 4.79 Dec + 0.393 Maximum - 0.251 Rainfall + Z(t)

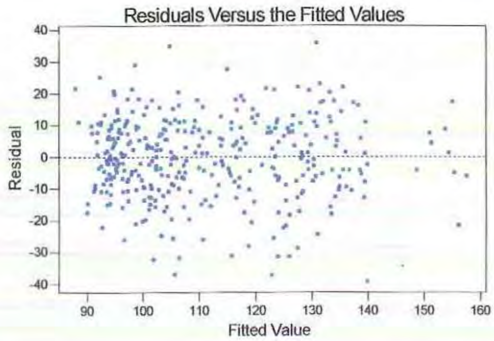
Forecasts for month of January 1999 using above model

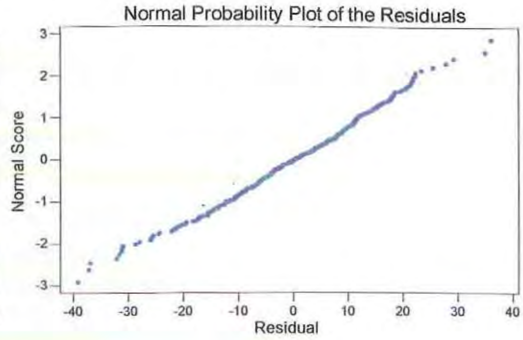
Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper
150	135	97.6	172.9	**105**	91	72.8	108.8	**90**	110	84.5	136.2
129	135	97.6	172.5	**90**	92	73.2	111.3	**115**	128	103.9	152.1
97	105	72.2	138.1	**90**	95	73.9	116.6	**91**	103	79.7	127.2
115	106	72.5	140.1	**98**	105	83.1	126.6	**103**	94	73.8	114.3
99	102	71.6	133.3	**111**	125	103.1	147.0	**95**	97	74.5	119.2
129	123	98.8	148.0	**100**	103	79.4	126.0	**82**	93	73.5	111.8
117	124	98.9	148.8	**87**	91	73.0	109.7	**85**	93	73.5	111.7
119	123	98.7	147.2	**92**	95	73.9	115.7	**92**	100	67.4	132.9
107	101	78.8	122.9	**86**	91	72.7	109.7	**125**	130	104.5	155.6
97	86	66.2	106.4	**108**	93	73.4	112.1	**102**	105	80.2	130.4

Residual Analysis of 1998 Total Presentations Model

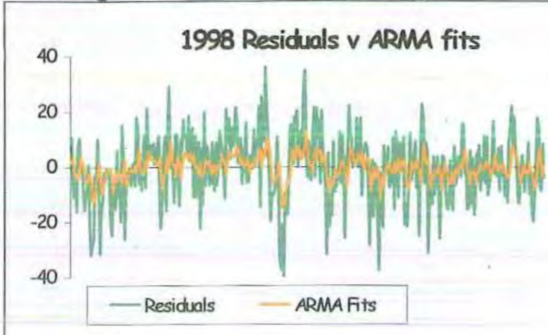








1998 Regression Residuals Modelling



Minitab: Final Estimates of Parameters

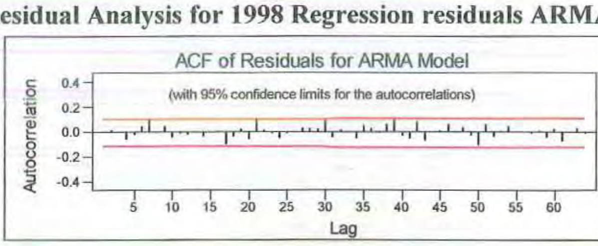
Type	Coef	StDev	T	P
AR 1	0.7850	0.0804	9.77	0.000
MA 1	0.5553	0.1079	5.15	0.000
Constant	0.0062	0.2742	0.02	0.982

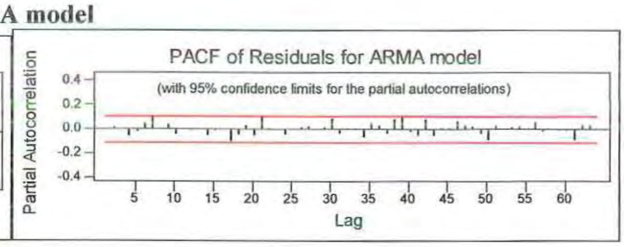
Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

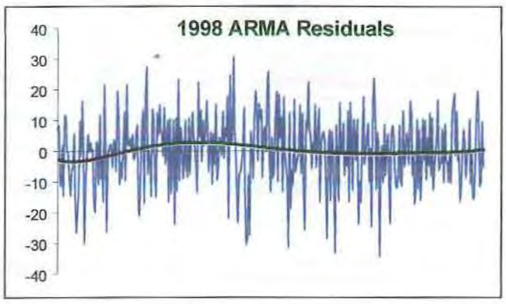
Lag	12	24	36	48
Chi-Square	8.3	22.4	30.1	44.3
DF	9	21	33	45
P-Value	0.507	0.379	0.613	0.500

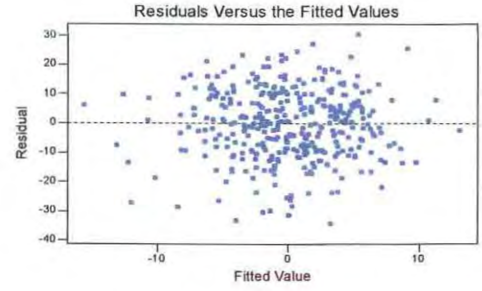
Prediction Model: Z(t) $(1 - 0.78B)x(t) = (1 - 0.56B)w(t)$

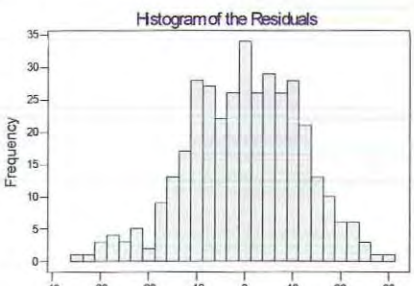
Residual Analysis for 1998 Regression residuals ARMA model

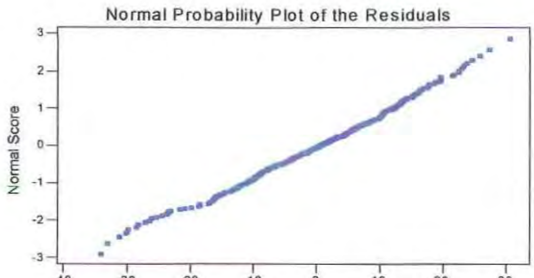




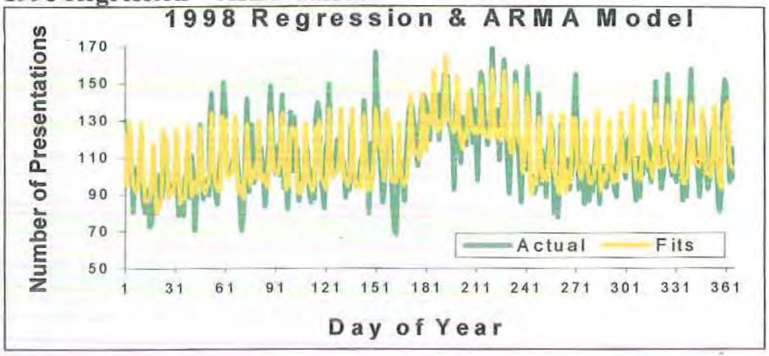








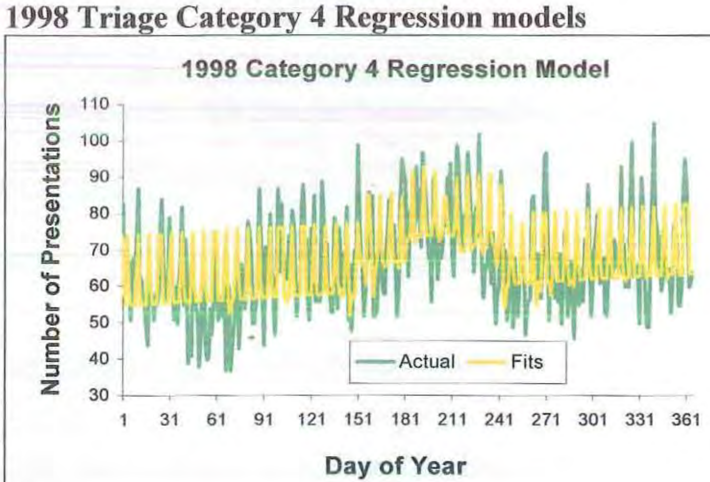
1998 Regression + ARMA model



Forecasts for month of January 1999 using Regression + ARMA model

Actual	Forecast	95% CI Lower	Upper	Actual	Forecast	95% CI Lower	Upper	Actual	Forecast	95% CI Lower	Upper
150	138	77.65	199.13	**105**	91	48.50	133.76	**90**	110	59.94	160.90
129	137	76.34	198.64	**90**	92	48.80	136.19	**115**	128	79.32	176.76
97	107	50.09	164.09	**90**	95	49.49	141.47	**91**	103	55.06	151.90
115	108	49.75	165.90	**98**	105	58.62	151.37	**103**	94	49.21	139.01
99	104	48.39	158.92	**111**	125	78.59	171.78	**95**	97	49.88	143.84
129	124	75.21	173.42	**100**	103	54.84	150.78	**82**	93	48.90	136.49
117	125	75.09	174.10	**87**	91	48.51	134.41	**85**	93	48.91	136.41
119	124	74.70	172.37	**92**	95	49.33	140.41	**92**	100	42.84	157.54
107	101	54.68	147.99	**86**	91	48.12	134.45	**125**	130	79.91	180.28
97	87	41.97	131.40	**108**	93	48.85	136.77	**102**	105	55.61	155.08

1998 Triage Category 4 Regression models



Accuracy Measures

Multiple R: 0.6929
R Square: 0.4801
Adjusted R Square: 0.4669
Standard Error: 9.6464
Observations: 365

MSE: 90.50

Best Subsets Regression

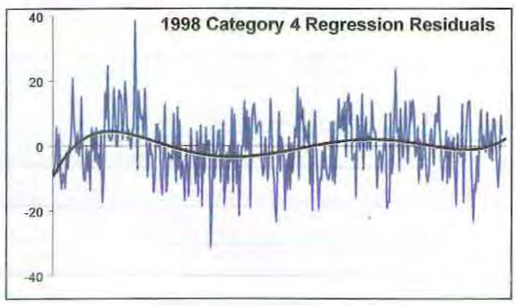
```
                                                              R
                                              H             M a
                                              o             a i
                                              l             x n
                                    T         i             i f
                                    i M F S S d A J J A S D m a
                  Adj.              m o r a u a p u u u e e u l
Vars   R-Sq   R-Sq    C-p        s  e n i t n y r n l g p c m l

   8   47.0   45.8   14.4   9.7264  X     X X X   X X X       X
   8   46.9   45.7   14.9   9.7335  X X   X X X   X X X
   9   48.0   46.7    9.5   9.6464  X X   X X X   X X X       X
   9   47.6   46.3   12.2   9.6832  X X   X X X X X X X
  10   48.6   47.2    7.2   9.6009  X X   X X X X X X X       X
  10   48.0   46.6   11.4   9.6583  X X   X X X   X X X     X X
  11   48.7   47.1    9.1   9.6138  X X   X X X X X X X X     X
  11   48.7   47.1    9.1   9.6139  X X X X X X X X X X       X
  12   48.7   46.9   11.0   9.6267  X X   X X X X X X X X   X X
  12   48.7   46.9   11.1   9.6268  X X X X X X X X X X X     X
  13   48.7   46.8   13.0   9.6397  X X X X X X X X X X X   X X
  13   48.7   46.8   13.0   9.6404  X X X X X X X X X X X X   X
  14   48.7   46.6   15.0   9.6535  X X X X X X X X X X X X X X
```

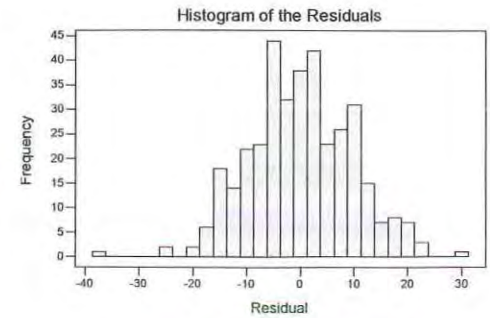
Prediction Model: Triage Number = 54.8 + 0.0247 Time + 4.19 Mon + 10.0 Sat + 19.2 Sun + 18.9 Holiday + 7.63 Jun + 14.5 Jul + 11.3 Aug - 0.266 Rainfall + Z(t)

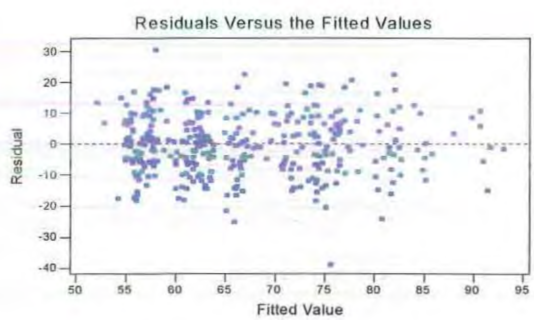
Forecasts for month of January 1999 using above Category 4 model

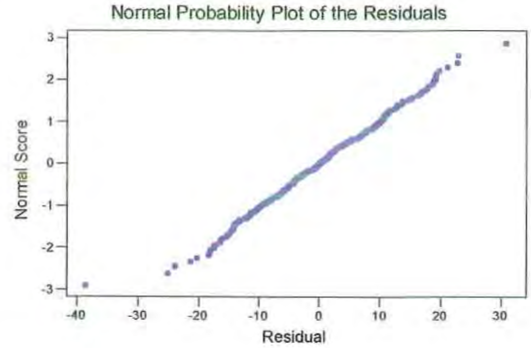
Actual	Forecast	95% CI Lower	Upper	Actual	Forecast	95% CI Lower	Upper	Actual	Forecast	95% CI Lower	Upper
88	83	72.8	92.4	**64**	55	52.7	57.3	**62**	65	59.7	70.8
95	83	72.8	92.4	**56**	55	52.7	57.3	**76**	74	68.9	80.0
88	83	72.8	92.4	**60**	55	52.7	57.3	**49**	59	53.9	65.0
73	83	72.8	92.5	**62**	65	59.6	70.5	**63**	55	52.9	57.7
60	64	58.1	69.5	**61**	74	68.8	79.7	**56**	55	52.9	57.8
62	64	58.1	69.5	**61**	59	53.8	64.8	**48**	55	52.9	57.8
64	64	58.1	69.6	**57**	55	52.8	57.5	**56**	55	52.9	57.8
58	74	67.3	80.1	**50**	55	52.8	57.5	**58**	57	44.9	68.6
64	74	67.4	80.1	**62**	55	52.4	57.5	**72**	75	69.0	80.2
77	74	67.4	80.2	**65**	55	52.8	57.6	**53**	60	54.0	65.2

Residual Analysis of 1998 Category 4 Triage Model









1998 Triage Category 4 Regression Residuals Modelling

Minitab: Final Estimates of Parameters

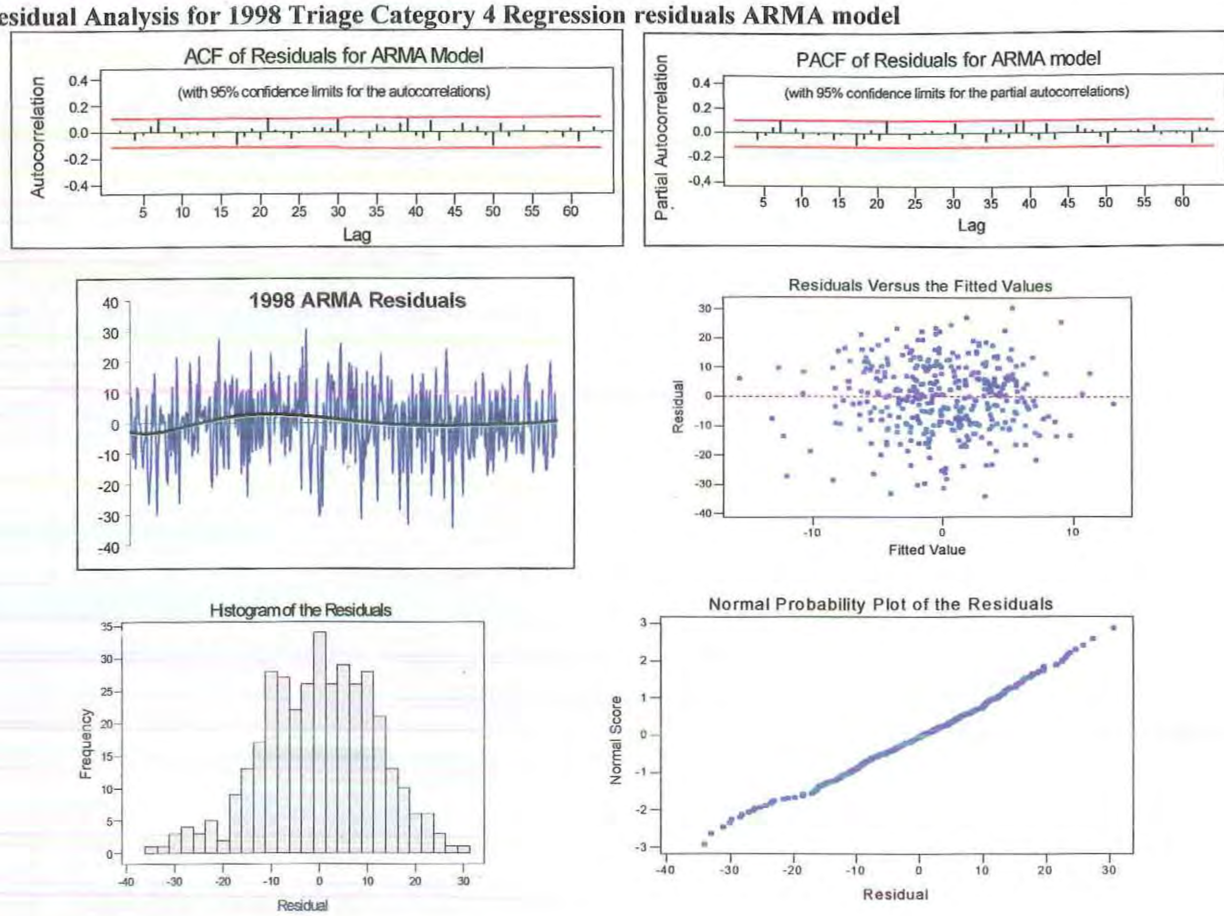
Type	Coef	StDev	T	P
AR 1	0.7784	0.0981	7.94	0.000
MA 1	0.5967	0.1253	4.76	0.000
Constant	0.0062	0.1937	0.03	0.974

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic Statistic

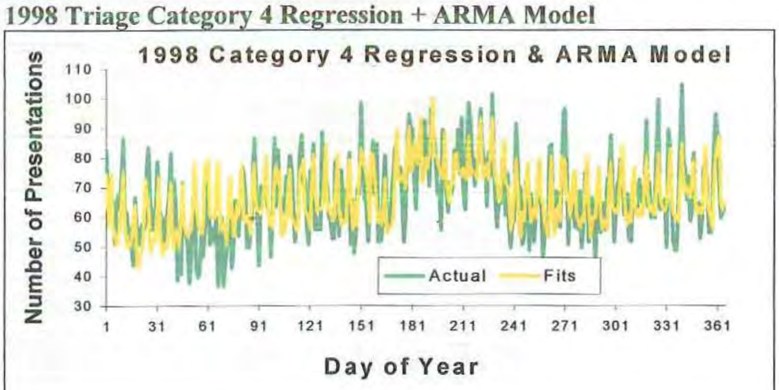
Lag	12	24	36	48
Chi-Square	9.7	16.8	25.4	36.9
DF	9	21	33	45
P-Value	0.375	0.726	0.824	0.800

Prediction Model: Z(t) $(1 - 0.78B)x(t) = (1 - 0.56B)w(t)$

Residual Analysis for 1998 Triage Category 4 Regression residuals ARMA model



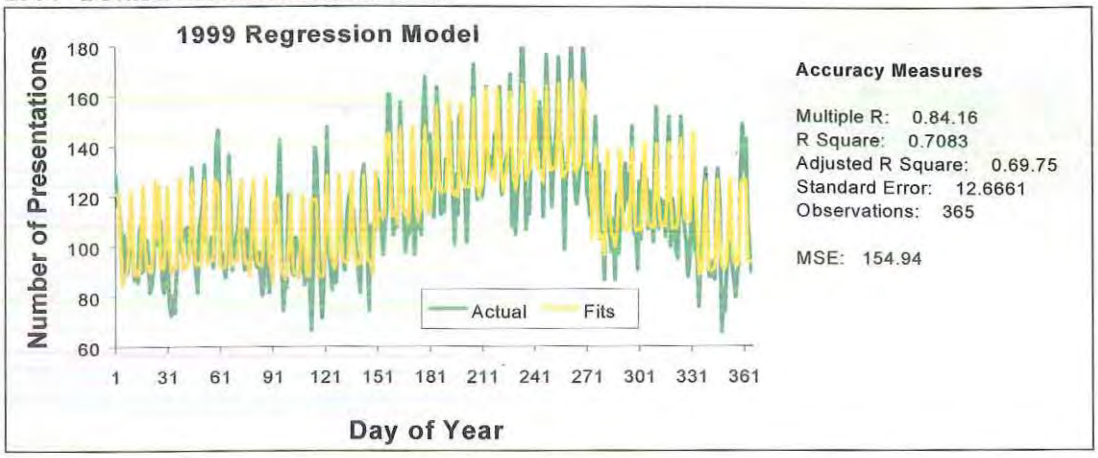
1998 Triage Category 4 Regression + ARMA Model



Forecasts for month of January 1999 using Regression + ARMA model

Actual	Forecast	95% CI		Actual	Forecast	95% CI		Actual	Forecast	95% CI	
		Lower	Upper			Lower	Upper			Lower	Upper
150	138	77.65	199.13	**105**	91	48.50	133.76	**90**	110	59.94	160.90
129	137	76.34	198.64	**90**	92	48.80	136.19	**115**	128	79.32	176.76
97	107	50.09	164.09	**90**	95	49.49	141.47	**91**	103	55.06	151.90
115	108	49.75	165.90	**98**	105	58.62	151.37	**103**	94	49.21	139.01
99	104	48.39	158.92	**111**	125	78.59	171.78	**95**	97	49.88	143.84
129	124	75.21	173.42	**100**	103	54.84	150.78	**82**	93	48.90	136.49
117	125	75.09	174.10	**87**	91	48.51	134.41	**85**	93	48.91	136.41
119	124	74.70	172.37	**92**	95	49.33	140.41	**92**	100	42.84	157.54
107	101	54.68	147.99	**86**	91	48.12	134.45	**125**	130	79.91	180.28
97	87	41.97	131.40	**108**	93	48.85	136.77	**102**	105	55.61	155.08

1999 Total Presentations models



Best Subsets Regression

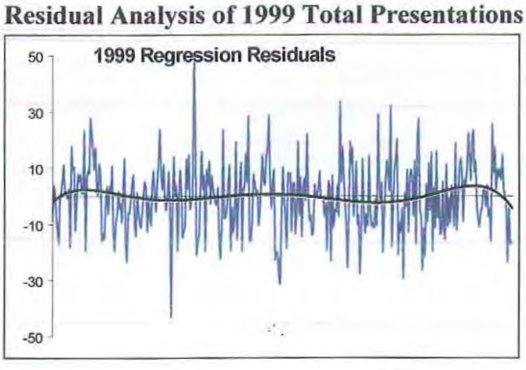
Vars	R-Sq	Adj. R-Sq	C-p	s	Time	Mon	Sat	Sun	Holiday	Apr	Jun	Jul	Aug	Sep	Dec	Maximum	Rainfall
8	64.7	63.9	78.1	13.841		X	X	X	X		X	X	X	X			
8	64.2	63.4	83.5	13.929	X		X	X	X		X	X	X	X			
9	67.1	66.3	50.3	13.366	X	X	X	X	X		X	X	X	X			
9	66.9	66.0	53.5	13.420	X		X	X	X		X	X	X	X	X		
10	69.6	68.8	22.5	12.870	X	X	X	X	X		X	X	X	X	X		
10	67.6	66.7	47.0	13.295	X	X	X	X	X	X	X	X	X	X			
11	70.2	69.3	17.1	12.757	X	X	X	X	X	X	X	X	X	X	X		
11	70.0	69.1	20.0	12.808	X	X	X	X	X		X	X	X	X	X	X	
12	70.6	69.6	14.5	12.693	X	X	X	X	X	X	X	X	X	X	X	X	
12	70.6	69.6	14.7	12.697	X	X	X	X	X	X	X	X	X	X	X		X
13	70.8	69.7	14.0	12.666	X	X	X	X	X	X	X	X	X	X	X	X	X

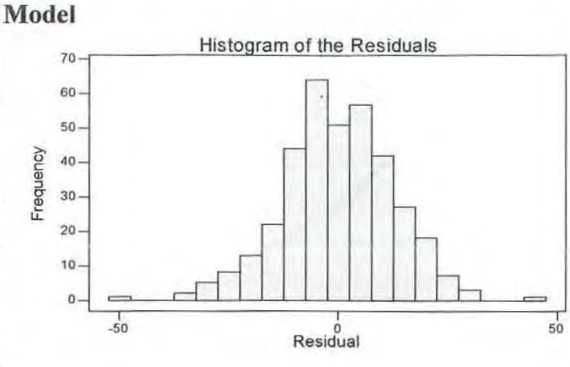
Prediction Model: Triage Number = 80.7 + 0.0638 Time + 11.3 Mon + 15.7 Sat + 33.9 Sun + 31.6 Holiday - 7.19 Apr + 17.3 Jun + 25.0 Jul + 29.4 Aug + 28.3 Sep – 19.4 Dec + 0.276 Maximum - 0.203 Rainfall + Z(t)

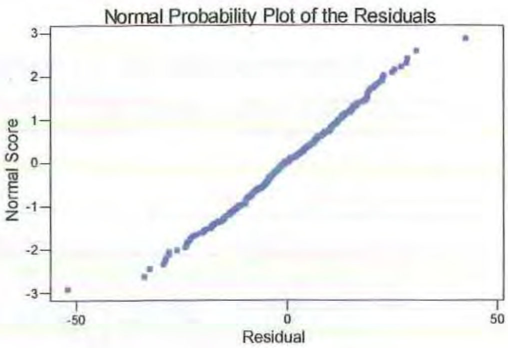
Forecasts for month of January 2000 using above Total Presentations model

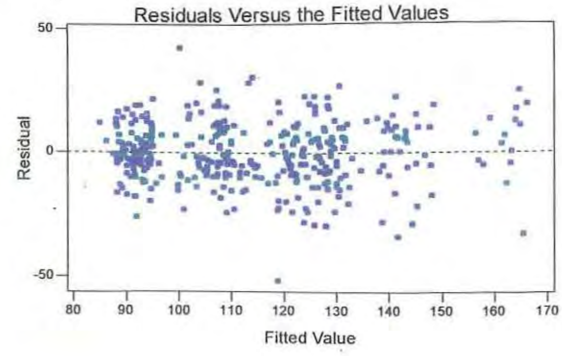
Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper
130	127	85.0	169.1	**83**	91	68.5	113.2	**97**	122	97.0	146.1
143	126	85.2	167.1	**87**	91	68.6	112.7	**61**	99	74.7	123.8
110	94	59.3	128.9	**110**	106	80.3	131.2	**88**	90	69.4	110.3
101	94	59.4	128.4	**110**	124	98.5	149.8	**71**	90	69.3	111.5
90	94	59.4	129.3	**91**	101	76.2	125.2	**76**	91	69.3	112.9
159	122	94.2	150.6	**83**	89	69.1	108.9	**83**	89	69.7	108.8
164	123	94.2	151.1	**70**	89	68.8	108.7	**78**	91	53.9	128.3
110	122	94.3	150.7	**55**	90	69.0	110.9	**110**	120	93.1	146.8
87	91	68.3	114.2	**59**	88	69.4	106.9	**90**	101	77.0	125.4
84	91	68.3	114.2	**85**	101	75.1	127.8	**81**	90	69.7	110.5

Residual Analysis of 1999 Total Presentations Model

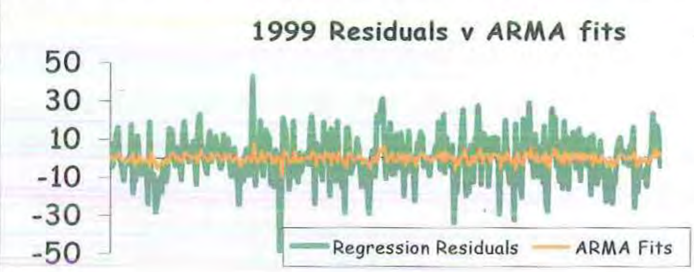








1999 Total Presentations Regression Residuals Modelling



Minitab: Final Estimates of Parameters

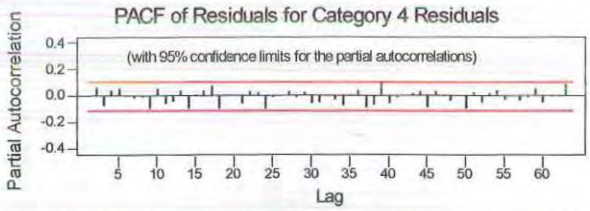
Type	Coef	StDev	T	P
AR 1	0.1817	0.0516	3.52	0.000
Constant	0.0019	0.6411	0.00	0.998

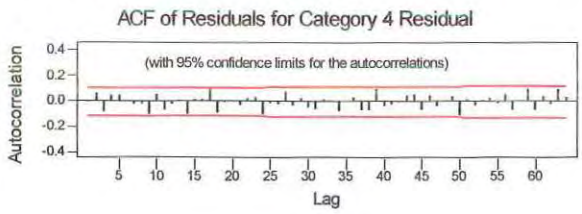
Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

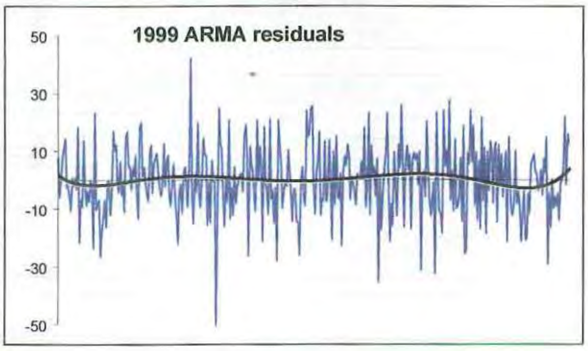
Lag	12	24	36	48
Chi-Square	12.8	29.6	38.3	52.1
DF	10	22	34	46
P-Value	0.236	0.129	0.281	0.248

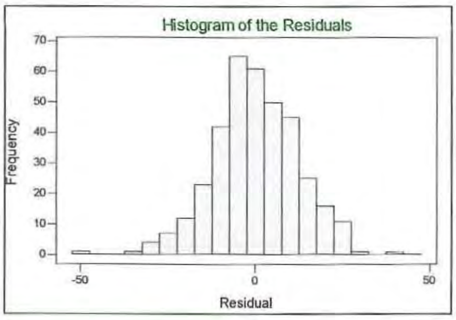
Prediction Model: Z(t) $(1 - 0.18B)x(t) = w(t)$

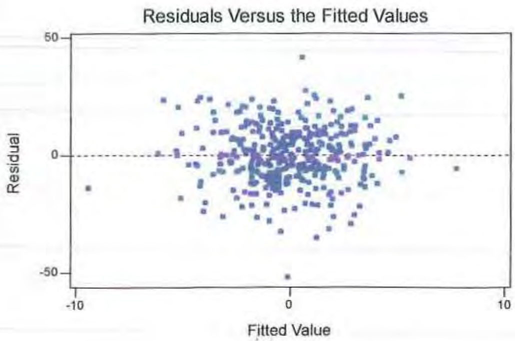
Residual Analysis for 1999 Total Presentations Regression residuals ARMA model

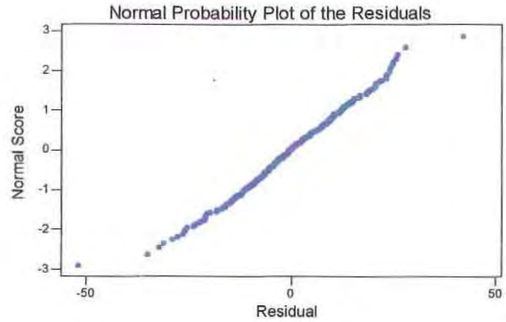




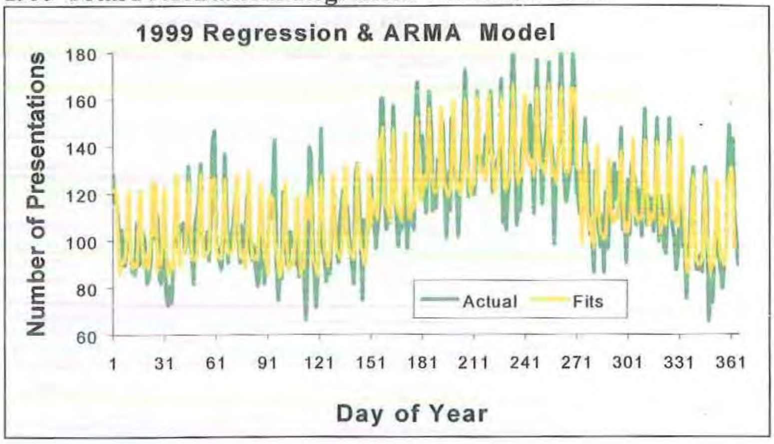








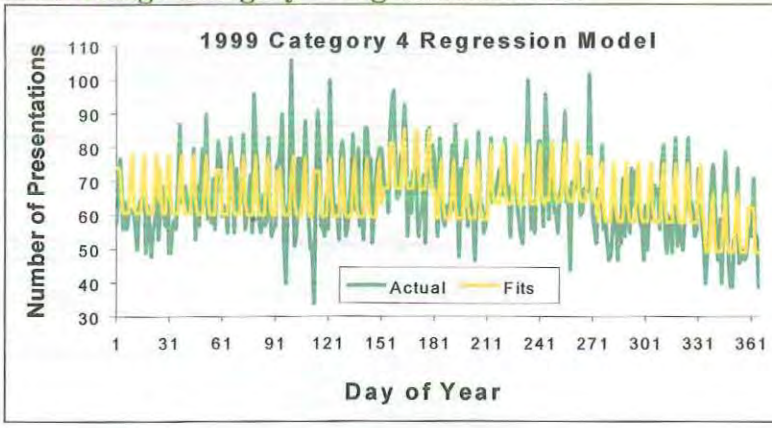
1999 Total Presentations Regression + ARMA Model



Forecasts for month of January 1999 using Regression + ARMA model

Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper
150	131	65.15	197.32	**105**	91	44.06	137.59	**90**	122	72.55	170.54
129	127	61.54	192.30	**90**	91	44.15	137.13	**115**	99	50.26	148.22
97	94	35.01	153.48	**90**	106	55.89	155.62	**91**	90	44.95	134.76
115	94	35.00	152.84	**98**	124	74.11	174.25	**103**	90	44.90	135.92
99	94	34.95	153.75	**111**	101	51.83	149.66	**95**	91	44.84	137.27
129	122	69.80	175.06	**100**	89	44.70	133.29	**82**	89	45.24	133.25
117	123	69.81	175.56	**87**	89	44.41	133.10	**85**	91	29.51	152.75
119	123	69.91	175.10	**92**	90	44.63	135.32	**92**	120	68.64	171.21
107	91	43.86	138.64	**86**	88	45.03	131.29	**125**	101	52.57	149.82
97	91	43.92	138.60	**108**	101	50.71	152.18	**102**	90	45.31	134.96

1999 Triage Category 4 Regression models



Accuracy Measures

Multiple R: 0.6493
R Square: 0.4216
Adjusted R Square: 0.4069
Standard Error: 9.2356
Observations: 365

MSE: 83.07

Best Subsets Regression

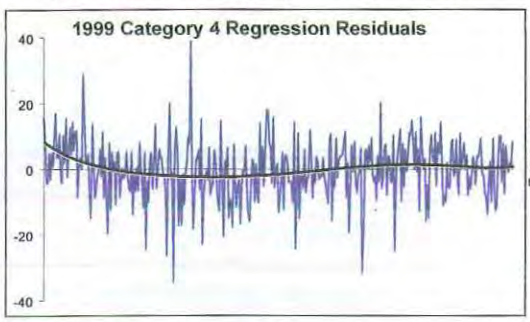
Vars	R-Sq	Adj. R-Sq	C-p	s	Time	Mon	Sat	Sun	Holiday	Apr	Jun	Jul	Aug	Sep	Dec	Maximum	Rainfall
7	40.9	39.8	12.5	9.3084		X	X	X	X		X			X	X		
7	40.7	39.5	13.9	9.3275			X	X	X		X		X	X	X		
8	41.7	40.4	9.3	9.2550		X	X	X	X		X		X	X	X		
8	41.1	39.7	13.5	9.3090	X	X	X	X	X		X			X	X		
9	42.2	40.7	8.8	9.2356	X	X	X	X	X		X		X	X	X		
9	41.9	40.4	10.5	9.2577		X	X	X	X		X		X	X	X		X
10	42.4	40.7	9.5	9.2319	X	X	X	X	X		X	X	X	X	X		
10	42.3	40.6	10.1	9.2396	X	X	X	X	X		X		X	X	X		X
11	42.5	40.7	10.6	9.2327	X	X	X	X	X		X	X	X	X	X		X
11	42.5	40.7	10.9	9.2362	X	X	X	X	X	X	X	X	X	X	X		
12	42.6	40.7	12.0	9.2378	X	X	X	X	X	X	X	X	X	X	X		X
12	42.5	40.6	12.6	9.2457	X	X	X	X	X		X	X	X	X	X	X	X
13	42.6	40.5	14.0	9.2508	X	X	X	X	X	X	X	X	X	X	X	X	X

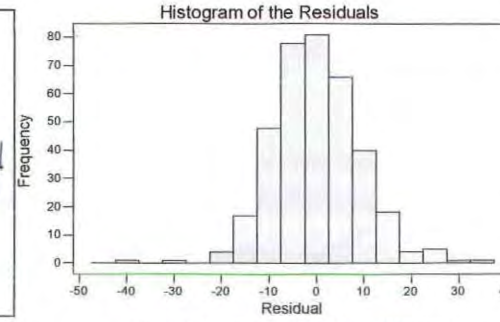
Prediction Model: Triage Number = 61.0 - 0.00904 Time + 3.88 Mon + 11.6 Sat + 17.3 Sun + 13.1 Holiday + 8.62 Jun + 4.70 Aug + 5.82 Sep - 8.50 Dec + Z(t)

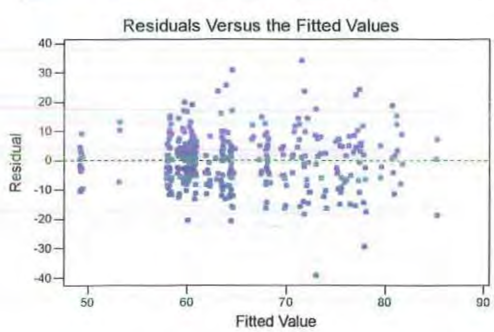
Forecasts for month of January 1999 using above Category 4 model

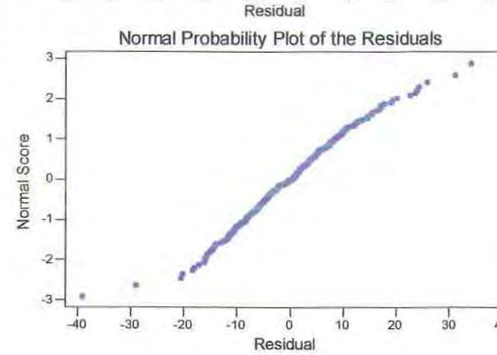
Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper
54	62	60.5	81.1	**40**	61	58.6	63.2	**53**	78	72.7	83.5
71	62	60.5	81.1	**50**	61	58.6	63.2	**35**	65	59.3	70.1
50	49	51.4	64.0	**47**	72	67.2	77.8	**38**	61	58.4	63.2
54	49	51.4	64.0	**60**	78	72.9	83.5	**46**	61	58.4	63.2
39	49	51.3	64.0	**47**	65	59.5	70.0	**43**	61	58.3	63.2
64	74	67.9	80.3	**41**	61	58.5	63.2	**46**	61	58.3	63.2
74	74	67.8	80.3	**35**	61	58.5	63.2	**43**	72	66.9	77.8
48	74	67.8	80.3	**30**	61	58.5	63.2	**60**	78	72.6	83.5
47	61	58.7	63.2	**36**	61	58.5	63.2	**47**	65	59.2	70.1
48	61	58.6	63.2	**42**	72	67.0	77.8	**37**	61	58.2	63.2

Residual Analysis of 1999 Category 4 Triage Model









1999 Triage Category 4 Regression Residuals Modelling

Minitab: Final Estimates of Parameters

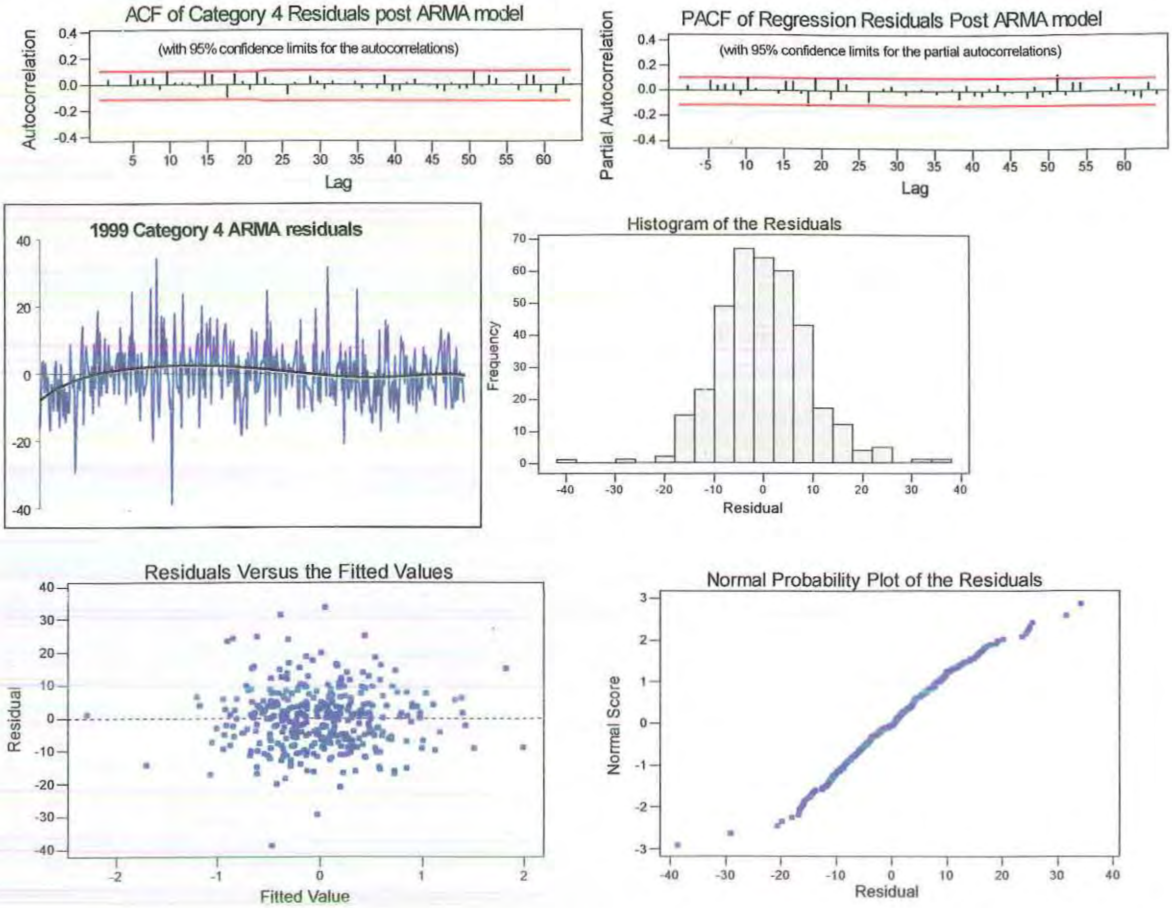
Type	Coef	StDev	T	P
AR 1	0.0583	0.0525	1.11	0.267
Constant	-0.0042	0.4773	-0.01	0.993

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

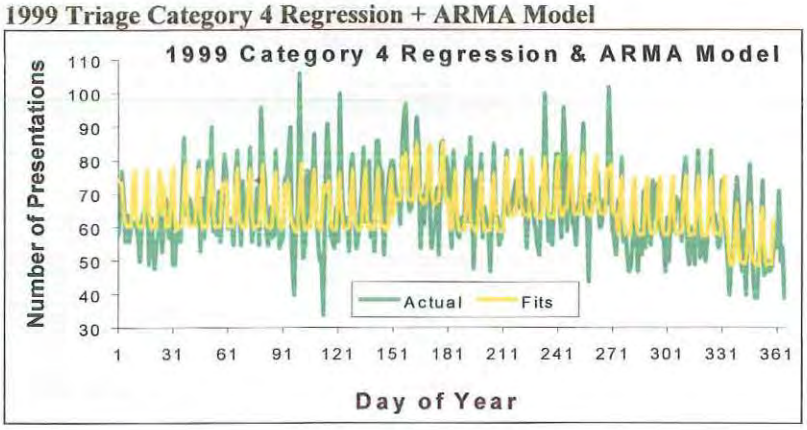
Lag	12	24	36	48
Chi-Square	11.0	28.9	34.5	41.3
DF	10	22	34	46
P-Value	0.361	0.147	0.443	0.668

Prediction Model: Z(t) $(1 - 0.58B)x(t) = w(t)$

Residual Analysis for 1999 Triage Category 4 Regression residuals ARMA model



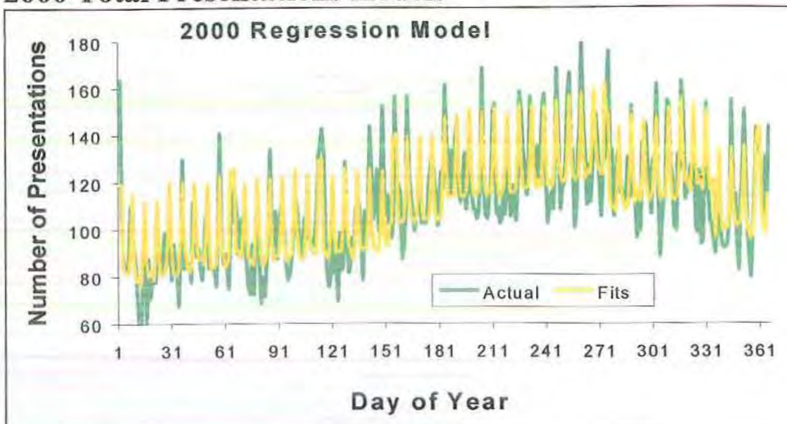
1999 Triage Category 4 Regression + ARMA Model



Forecasts for month of January 2000 using Category 4 Regression + ARMA model

Actual	Forecast	95% CI		Actual	Forecast	95% CI		Actual	Forecast	95% CI	
		Lower	Upper			Lower	Upper			Lower	Upper
54	62	42	99	**40**	61	41	81	**53**	78	55	101
71	62	43	99	**50**	61	41	81	**35**	65	41	88
50	49	33	82	**47**	72	49	96	**38**	61	40	81
54	49	33	82	**60**	78	55	101	**46**	61	40	81
39	49	33	82	**47**	65	42	88	**43**	61	40	81
64	74	50	98	**41**	61	41	81	**46**	61	40	81
74	74	50	98	**35**	61	41	81	**43**	72	49	96
48	74	50	98	**30**	61	41	81	**60**	78	55	101
47	61	41	81	**36**	61	41	81	**47**	65	41	88
48	61	41	81	**42**	72	49	96	**37**	61	40	81

2000 Total Presentations models



Accuracy Measures

Multiple R: 0.8372
R Square: 0.7009
Adjusted R Square: 0.6916
Standard Error: 12.7519
Observations: 366

MSE: 157.28

Best Subsets Regression

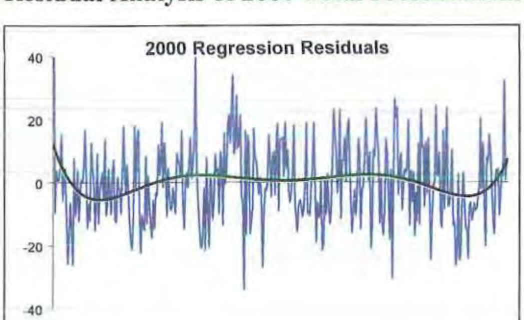
Vars	R-Sq	Adj. R-Sq	C-p	s	Time	Mon	Sat	Sun	Holiday	Apr	Jun	Jul	Aug	Sep	Dec	Max Temp	Rain
7	65.3	64.6	58.8	13.653	X		X	X	X			X		X	X		
7	65.3	64.6	59.1	13.658	X	X	X	X	X			X			X		
8	67.1	66.4	39.4	13.308	X		X	X	X			X	X	X	X		
8	66.9	66.1	42.8	13.367	X	X	X	X	X			X		X	X		
9	68.7	67.9	23.2	13.010	X	X	X	X	X			X	X	X	X		
9	67.6	66.8	35.6	13.227	X		X	X	X		X	X	X	X	X		
10	69.2	68.4	18.9	12.917	X	X	X	X	X		X	X	X	X	X		
10	68.9	68.1	22.3	12.977	X	X	X	X	X			X	X	X	X	X	
11	70.1	69.2	10.7	12.752	X	X	X	X	X		X	X	X	X	X	X	
11	69.4	68.4	18.9	12.900	X	X	X	X	X		X	X	X	X	X		X
12	70.1	69.1	12.1	12.759	X	X	X	X	X		X	X	X	X	X	X	X
12	70.1	69.1	12.6	12.769	X	X	X	X	X	X	X	X	X	X	X	X	
13	70.1	69.0	14.0	12.776	X	X	X	X	X	X	X	X	X	X	X	X	X

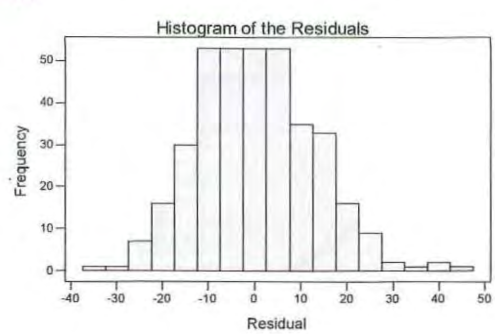
Prediction Model: Triage Number = 61.6 + 0.127 Time + 8.74 Mon + 14.1 Sat + 34.3 Sun + 37.0 Holiday + 10.4 Jun + 18.4 Jul + 17.1 Aug + 16.7 Sep - 22.8 Dec + 0.559 Max Temp + Z(t)

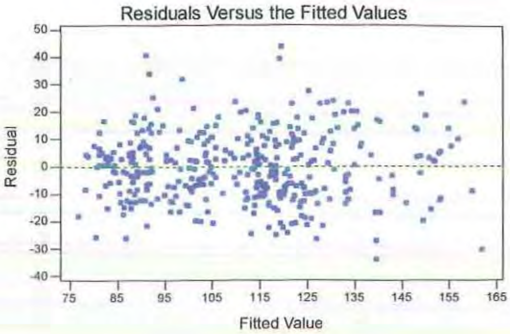
Forecasts for month of December 2000 using above Total Presentations model

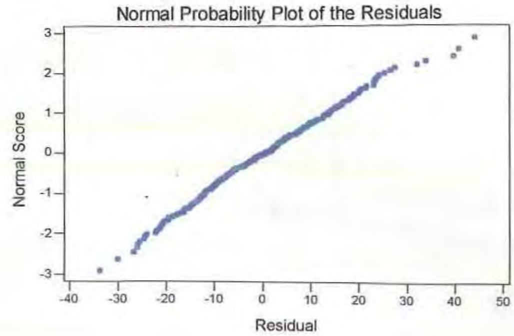
Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper
95	117	92.7	141.4	**98**	107	70.5	144.3	**83**	103	68.1	137.4
128	133	103.6	162.7	**93**	100	66.3	132.7	**112**	103	68.4	138.5
154	151	122.8	178.7	**92**	101	66.8	134.8	**121**	114	77.3	151.4
117	124	97.0	152.0	**93**	101	66.9	134.6	**151**	136	98.1	173.5
103	117	92.9	141.1	**97**	101	67.0	134.7	**121**	110	72.5	148.3
103	118	93.3	142.7	**113**	113	76.5	150.4	**100**	100	67.4	133.1
97	121	94.5	147.3	**155**	135	97.1	171.9	**87**	98	66.4	128.6
90	97	65.1	129.1	**109**	111	72.3	149.3	**80**	96	66.0	126.5
104	110	74.8	145.6	**115**	102	67.8	137.0	**99**	97	66.5	128.1
122	134	96.4	171.1	**95**	102	67.6	135.6	**119**	115	77.9	151.6

Residual Analysis of 2000 Total Presentations Model

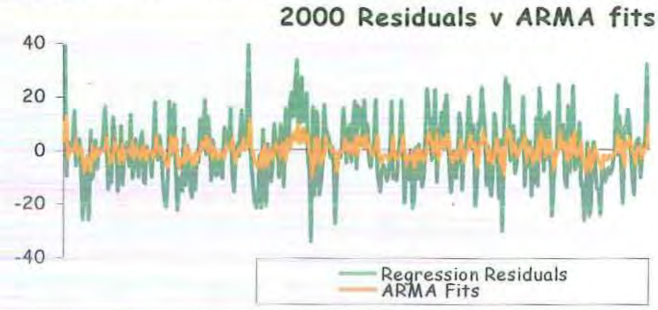








2000 Total Presentations Regression Residuals Modelling



Minitab: Final Estimates of Parameters

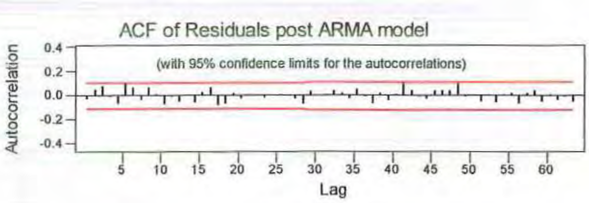
Type	Coef	StDev	T	P
AR 1	0.2971	0.0500	5.94	0.000
Constant	0.0385	0.6278	0.06	0.951

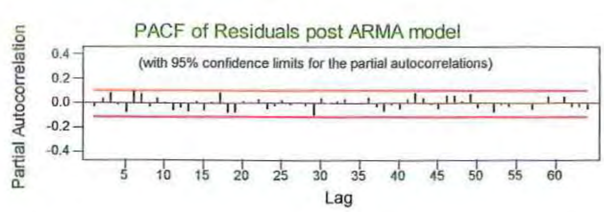
Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

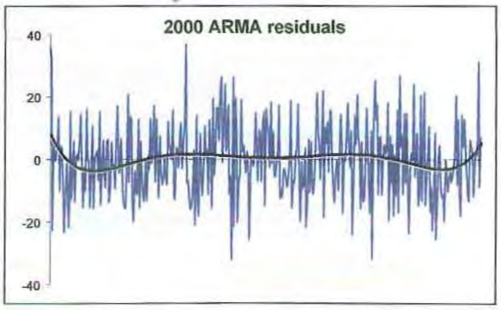
Lag	12	24	36	48
Chi-Square	15.8	25.7	31.2	41.8
DF	10	22	34	46
P-Value	0.106	0.265	0.608	0.650

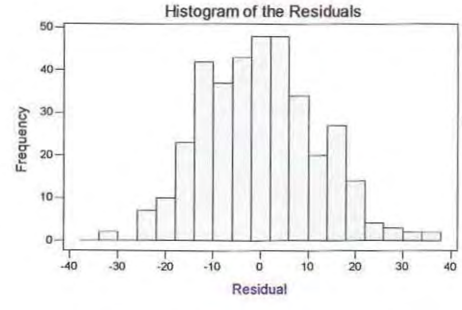
Prediction Model: Z(t) $(1 - 0.30B)x(t) = w(t)$

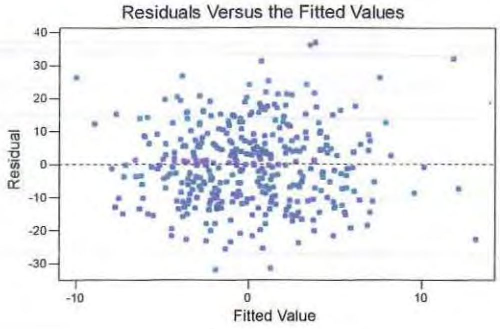
Residual Analysis for 2000 Total Presentations Regression residuals ARMA model

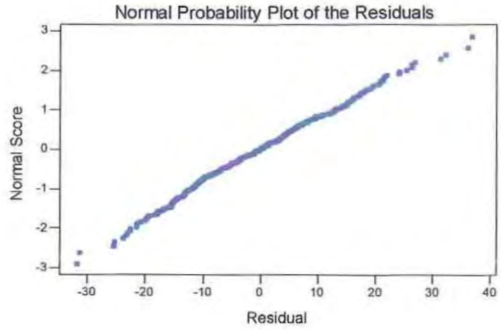






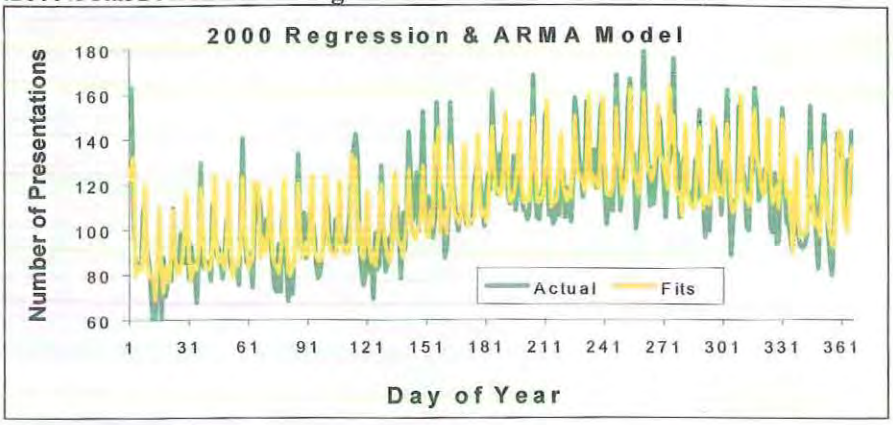






00000

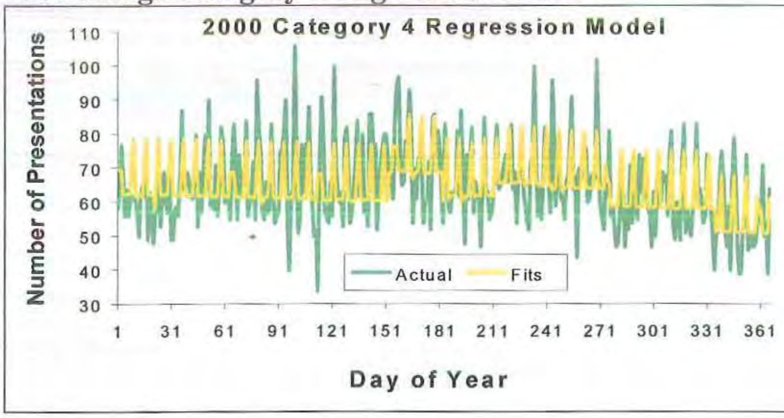
.2000 Total Presentations Regression + ARMA Model



Forecasts for month of December 2000 using Regression + ARMA model

Actual	Forecast	95% CI Lower	Upper	Actual	Forecast	95% CI Lower	Upper	Actual	Forecast	95% CI Lower	Upper
95	116	67.6	163.5	**98**	107	45.9	169.0	**83**	103	43.5	162.1
128	133	78.6	186.9	**93**	100	41.7	157.5	**112**	104	43.8	163.2
154	151	98.1	203.3	**92**	101	42.2	159.5	**121**	114	52.7	176.1
117	124	72.3	176.7	**93**	101	42.3	159.3	**151**	136	73.4	198.2
103	117	68.3	165.8	**97**	101	42.4	159.4	**121**	110	47.9	173.0
103	118	68.7	167.4	**113**	114	51.9	175.1	**100**	100	42.8	157.8
97	121	69.9	172.0	**155**	135	72.5	196.6	**87**	98	41.8	153.3
90	97	40.5	153.8	**109**	111	47.7	174.0	**80**	96	41.4	151.2
104	110	50.2	170.3	**115**	102	43.2	161.7	**99**	97	41.9	152.9
122	134	71.8	195.8	**95**	102	43.0	160.3	**119**	115	53.3	176.3

2000 Triage Category 4 Regression models



Accuracy Measures

Multiple R: 0.6107
R Square: 0.3730
Adjusted R Square: 0.3516
Standard Error: 9.6431
Observations: 366

MSE: 89.69

Best Subsets Regression

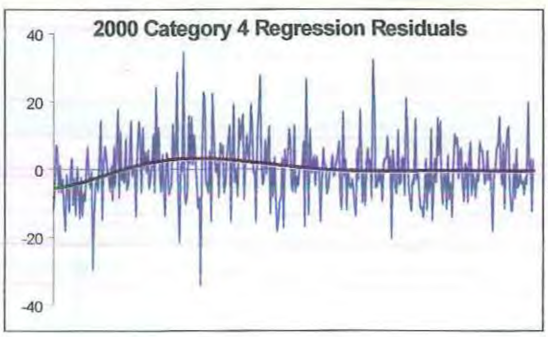
Vars	R-Sq	Adj. R-Sq	C-p	s	Time	Mon	Sat	Sun	Holiday	Apr	Jun	Jul	Aug	Sep	Dec	Max Temp	Rain
7	53.6	52.7	26.1	8.1554	X		X	X	X			X	X		X		
7	53.5	52.6	26.6	8.1608	X	X	X	X	X			X			X		
8	54.4	53.4	21.3	8.0929	X		X	X	X			X	X	X	X		
8	54.2	53.2	23.2	8.1135	X	X	X	X	X			X	X		X		
9	55.0	53.9	18.5	8.0511	X	X	X	X	X			X	X	X	X		
9	55.0	53.8	18.9	8.0559	X		X	X	X			X	X	X	X	X	
10	56.0	54.8	12.5	7.9735	X		X	X	X		X	X	X	X	X	X	
10	55.5	54.2	16.5	8.0183	X	X	X	X	X			X	X	X	X	X	
11	56.5	55.2	10.1	7.9348	X	X	X	X	X		X	X	X	X	X	X	
11	56.0	54.6	14.5	7.9841	X		X	X	X	X	X	X	X	X	X	X	
12	56.6	55.1	12.0	7.9453	X	X	X	X	X	X	X	X	X	X	X	X	
12	56.5	55.1	12.1	7.9460	X	X	X	X	X		X	X	X	X	X	X	X
13	56.6	54.9	14.0	7.9565	X	X	X	X	X	X	X	X	X	X	X	X	X

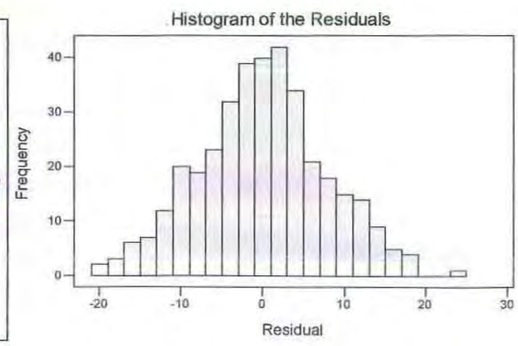
Prediction Model: Triage Number = 33.0 + 0.0633 Time + 2.74 Mon + 7.28 Sat + 16.3 Sun + 15.4 Holiday + 5.09 Jun + 8.48 Jul + 7.40 Aug + 6.08 Sep - 11.2 Dec + 0.339 Max Temp + 0.0056 Rain + Z(t)

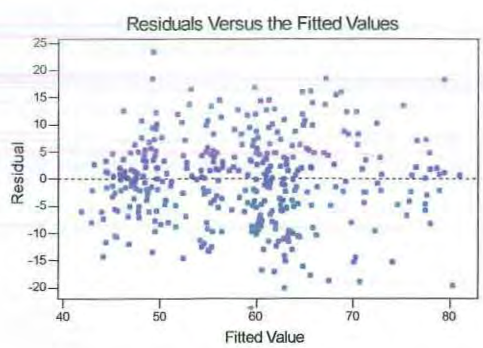
Forecasts for month of December 2000 using above Category 4 model

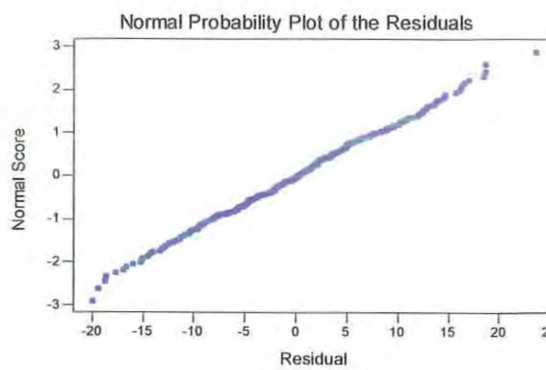
Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper
54	58	39.4	77.0	**40**	75	39.6	96.2	**39**	51	24.6	77.9
56	58	35.6	81.1	**50**	67	25.8	76.9	**39**	51	24.3	78.2
74	68	46.9	89.9	**47**	52	25.2	77.5	**53**	51	22.9	79.8
71	74	51.0	97.1	**60**	47	25.3	77.4	**53**	61	32.5	90.4
58	58	39.5	76.7	**47**	59	25.2	77.4	**74**	68	38.7	96.8
52	58	39.1	77.2	**41**	40	23.1	79.8	**46**	51	25.9	76.4
40	58	37.8	78.5	**35**	66	32.8	90.2	**47**	51	27.2	75.0
51	51	26.8	76.0	**30**	79	38.3	97.4	**50**	51	27.8	74.3
51	52	24.4	78.7	**36**	64	24.7	77.9	**47**	51	27.4	74.8
68	62	33.0	90.3	**42**	46	25.1	77.4	**48**	51	22.9	79.6

Residual Analysis of 2000 Category 4 Triage Model

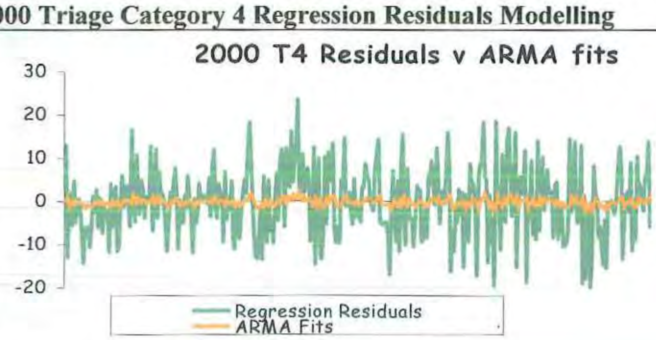








2000 Triage Category 4 Regression Residuals Modelling



Minitab: Final Estimates of Parameters

Type	Coef	StDev	T	P
AR 1	0.6998	0.2020	3.46	0.001
MA 1	0.5886	0.2283	2.58	0.010
Constant	0.0041	0.1665	0.02	0.980

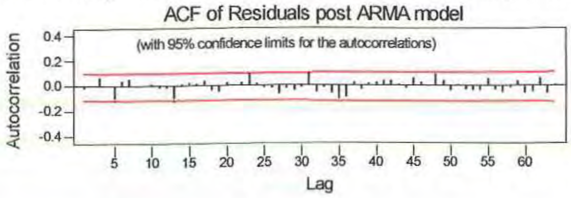
Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-Square	9.5	23.2	41.6	49.4
DF	9	21	33	45
P-Value	0.391	0.333	0.146	0.303

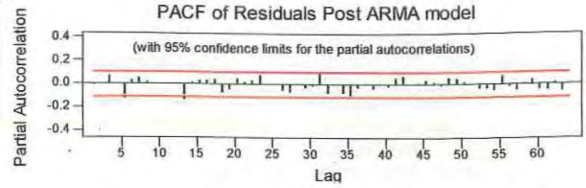
Prediction Model: Z(t) (1 - 0.70B)x(t) = (1 - 0.59B)w(t)

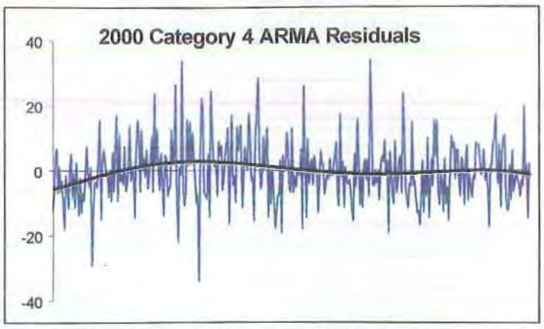
Residual Analysis for 2000 Triage Category 4 Regression residuals ARMA model

ACF of Residuals post ARMA model

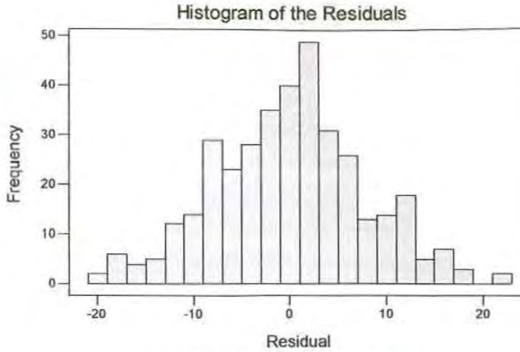


PACF of Residuals Post ARMA model

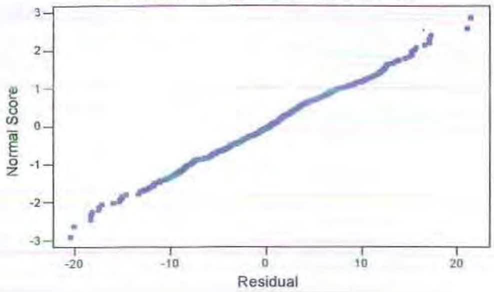




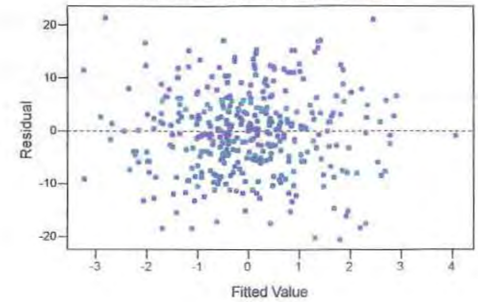
Histogram of the Residuals



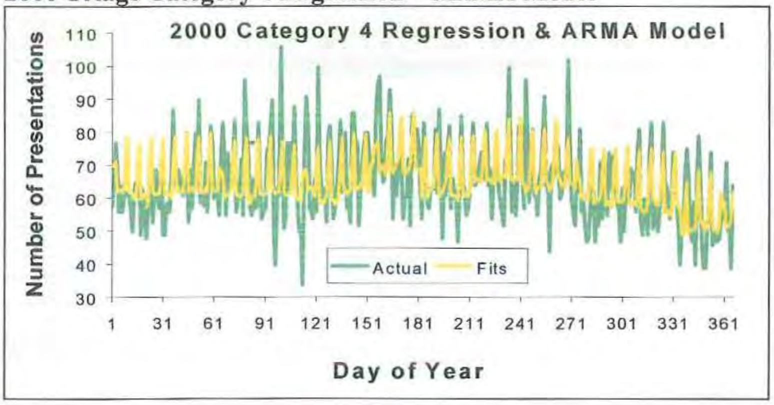
Normal Probability Plot of the Residuals



Residuals Versus the Fitted Values



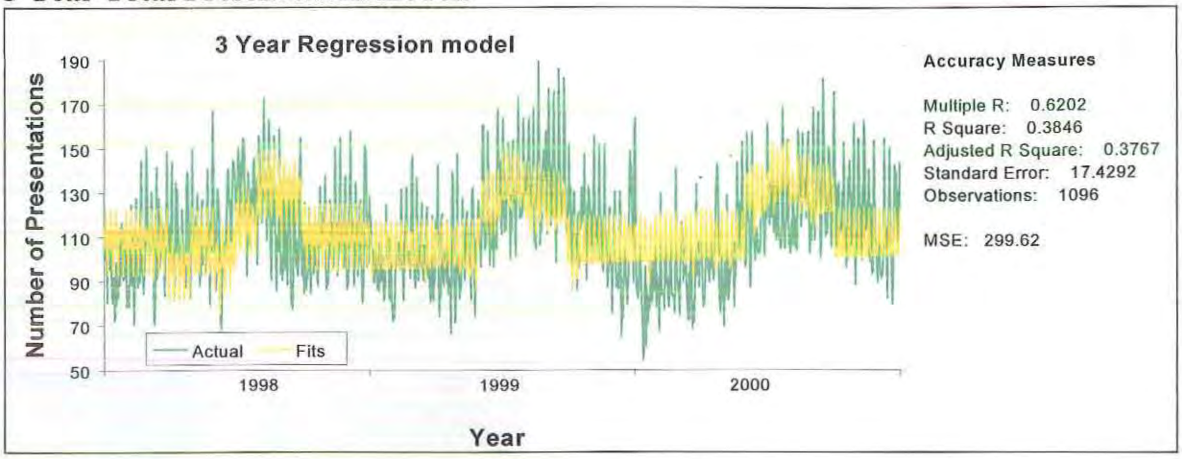
2000 Triage Category 4 Regression + ARMA Model



Forecasts for month of December 2000 using Category 4 Regression + ARMA model

		95% CI				95% CI				95% CI	
Actual	Forecast	Lower	Upper	**Actual**	Forecast	Lower	Upper	**Actual**	Forecast	Lower	Upper
56	57	19.4	95.4	**67**	51	10.4	92.3	**39**	51	8.9	93.6
74	68	31.0	104.5	**52**	51	9.8	92.8	**53**	51	7.6	95.2
71	74	35.2	111.9	**47**	51	9.9	92.7	**53**	61	17.1	105.8
58	58	23.9	91.8	**59**	51	9.9	92.7	**74**	68	23.3	112.2
52	58	23.5	92.3	**40**	51	7.7	95.2	**46**	51	10.5	91.8
40	58	22.2	93.7	**66**	62	17.5	105.6	**47**	51	11.9	90.4
51	51	11.3	91.3	**79**	68	22.9	112.8	**50**	51	12.5	89.7
51	51	8.9	94.0	**64**	51	9.3	93.2	**47**	51	12.0	90.2
68	62	17.5	105.6	**46**	51	9.7	92.8	**48**	51	7.6	95.0
75	68	24.2	111.6	**39**	51	9.2	93.3	**54**	61	16.0	106.7

3 Year Total Presentations models



Best Subsets Regression

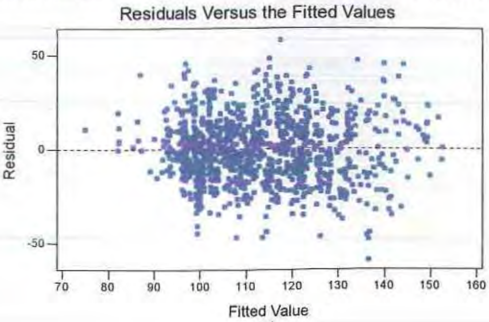
Vars	R-Sq	Adj. R-Sq	C-p	s	Time	Mon	Tue	Thu	Fri	Sat	Sun	Holiday	Apr	Jun	Jul	Aug	Sep	Dec	Maxim	Rainfall
8	33.0	32.5	96.8	18.136		X	X	X	X		X	X				X	X			
8	33.0	32.5	97.2	18.139	X			X	X		X	X			X	X	X			
9	34.3	33.8	75.5	17.963		X	X	X	X		X	X			X	X	X			
9	34.3	33.7	76.9	17.974	X	X	X	X	X		X	X				X	X			
10	35.6	35.0	56.0	17.803	X	X	X	X	X		X	X			X	X	X			
10	35.1	34.5	63.4	17.861		X	X	X	X		X	X			X	X	X			X
11	36.3	35.7	44.2	17.702	X	X	X	X	X		X	X			X	X	X			X
11	36.2	35.5	47.1	17.725	X	X	X	X	X	X	X	X	X			X	X			
12	37.1	36.4	32.6	17.602	X	X	X	X	X	X	X	X	X	X		X	X			
12	37.0	36.3	34.1	17.614	X	X	X	X	X	X	X	X	X		X	X	X			
13	37.9	37.2	20.4	17.496	X	X	X	X	X	X	X	X	X		X	X	X			X
13	37.7	37.0	23.8	17.523	X	X	X	X	X	X	X	X	X	X	X	X	X			
14	38.5	37.7	13.1	17.429	X	X	X	X	X	X	X	X	X	X	X	X	X			X
14	38.0	37.2	20.9	17.492	X	X	X	X	X	X	X	X	X		X	X	X		X	X
15	38.5	37.6	15.0	17.437	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X
15	38.5	37.6	15.1	17.437	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X
16	38.5	37.6	17.0	17.445	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

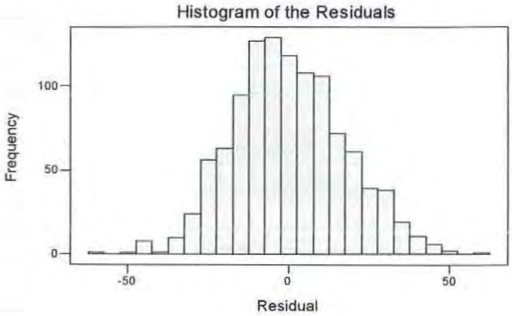
Prediction Model: Triage Number = 93.1 + 0.00871 Time + 9.93 Mon + 10.6 Tue + 21.1 Thu + 15.8 Fri + 12.4 Sat + 29.5 Sun + 20.8 Holiday - 11.7 Apr - 7.71 Jun + 9.70 Jul + 25.0 Aug + 20.4 Sep - 0.346 Rainfall + Z(t)

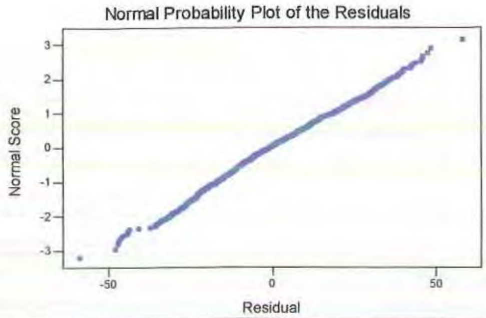
Forecasts for month of December 2000 using above Total Presentations model

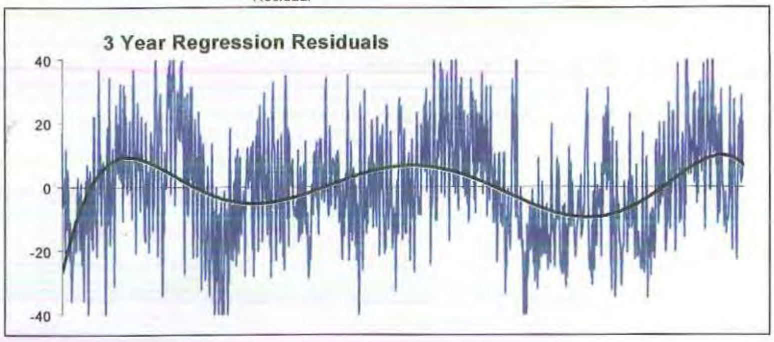
Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper
128	113	103.3	122.4	**93**	102	96.1	108.7	**112**	102	96.2	108.8
154	123	113.9	132.9	**92**	102	96.1	108.7	**121**	113	103.5	122.6
117	109	97.1	120.2	**93**	102	96.1	108.7	**151**	124	114.0	133.2
103	102	96.1	108.6	**97**	102	96.1	108.7	**121**	112	102.8	122.1
103	102	96.1	108.6	**113**	113	103.4	122.5	**100**	103	96.2	108.8
97	102	96.1	108.6	**155**	124	114.0	133.1	**87**	103	96.2	108.8
90	102	96.1	108.6	**109**	112	102.8	122.0	**80**	103	96.2	108.9
104	113	103.4	122.4	**115**	102	96.2	108.7	**99**	103	96.2	108.9
122	123	113.9	133.0	**95**	102	96.2	108.8	**119**	113	103.5	122.7
98	112	102.7	121.9	**83**	102	96.2	108.8	**143**	124	114.1	133.3

Residual Analysis of 3 Year Total Presentations Model

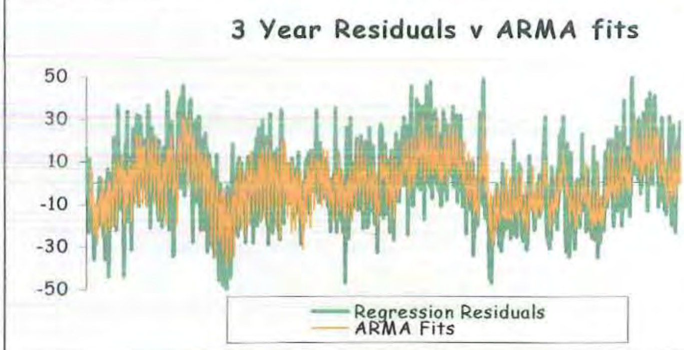








3 year Total Presentations Regression Residuals Modelling



Minitab: Final Estimates of Parameters

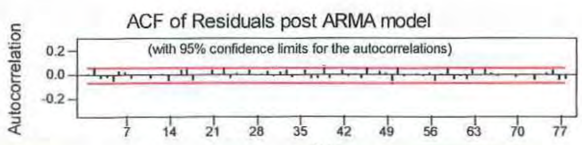
Type	Coef	StDev	T	P
AR 1	0.9561	0.0139	68.65	0.000
MA 1	0.6306	0.0338	18.64	0.000
MA 2	0.1301	0.0321	4.06	0.000
SMA 7	0.9472	0.0078	120.89	0.000
Constant	0.001482	0.005262	0.28	0.778

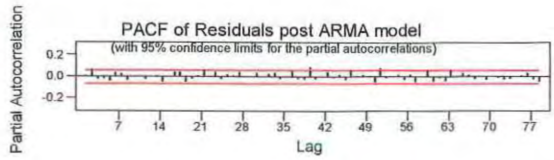
Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

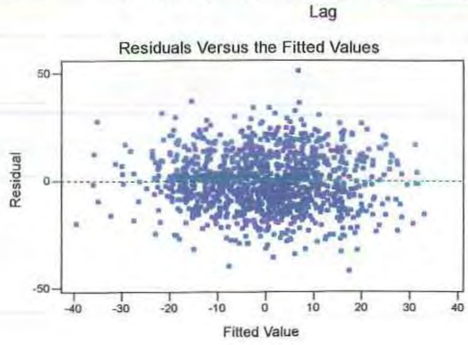
Lag	12	24	36	48
Chi-Square	11.5	27.6	35.6	55.8
DF	7	19	31	43
P-Value	0.118	0.092	0.262	0.090

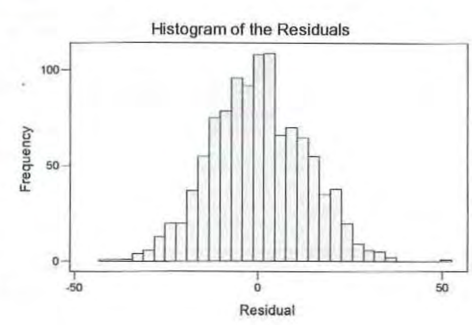
Prediction Model: Z(t) $(1 - 0.96B)\nabla_7 x(t) = (1 - 0.63B)(1 - 0.13B^2)(1 - 0.95 B^7) w(t)$

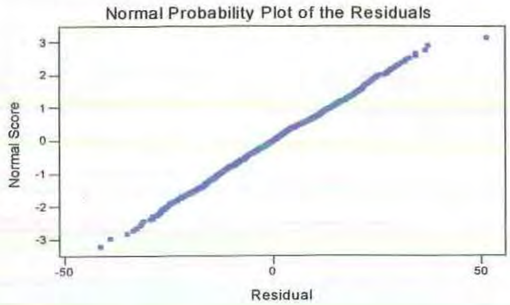
Residual Analysis for 3 Year Total Presentations Regression residuals ARMA model

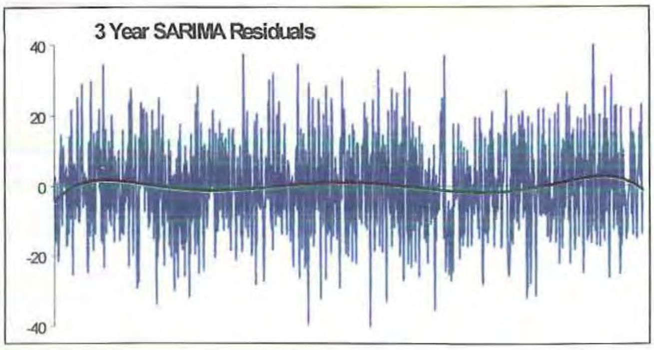




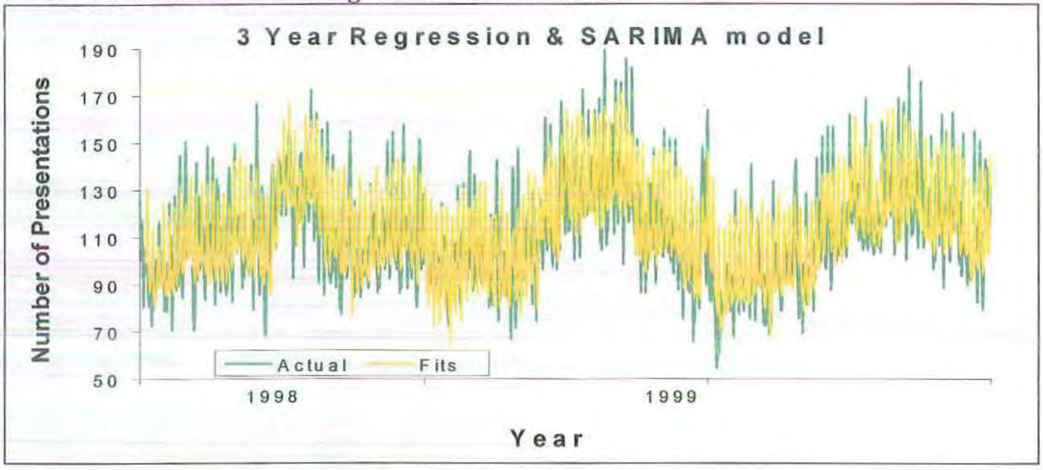








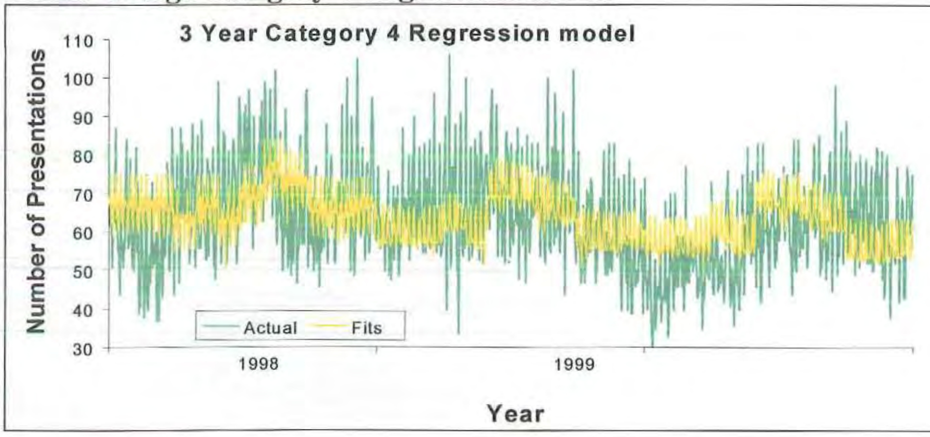
3 Year Total Presentations Regression + ARMA Model



Forecasts for month of December 2000 using Regression + ARMA model

Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper
128	133	98.4	168.3	93	104	68.4	140.1	112	109	71.9	146.7
154	131	94.2	166.8	92	102	66.1	138.2	121	130	89.5	171.3
117	114	75.0	152.5	93	106	70.1	142.6	151	129	88.0	170.0
103	105	71.1	138.7	97	110	73.3	146.0	121	116	74.8	157.1
103	103	68.6	136.9	113	131	90.8	170.7	100	104	65.6	141.4
97	107	72.5	141.3	155	129	89.3	169.5	87	101	63.5	139.5
90	110	75.5	144.9	109	116	76.0	156.5	80	106	67.6	143.8
104	131	92.9	169.7	115	104	66.7	140.9	99	109	70.9	147.2
122	130	91.2	168.5	95	102	64.5	139.0	119	130	88.6	171.8
98	117	77.8	155.7	83	106	68.6	143.3	143	129	87.2	170.5

3 Year Triage Category 4 Regression models



Accuracy Measures

Multiple R: 0.4991
R Square: 0.2491
Adjusted R Square: 0.2379
Standard Error: 11.2383
Observations: 1096

MSE: 123.63

Best Subsets Regression

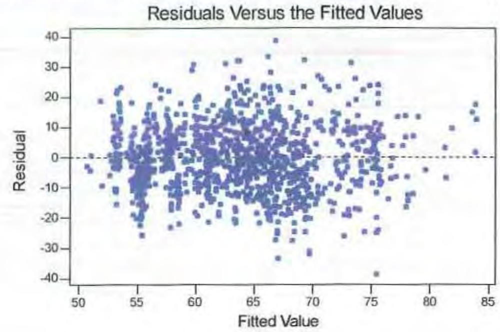
Vars	R-Sq	Adj. R-Sq	C-p	s	Time	Mon	Tue	Thu	Fri	Sat	Sun	Holiday	Apr	Jun	Jul	Aug	Sep	Dec	Maximal	Rainfall
8	21.4	20.8	51.6	11.457	X		X	X		X	X	X				X	X			
8	21.2	20.6	54.3	11.470	X		X	X	X		X	X				X	X			
9	22.5	21.8	37.8	11.381	X		X	X	X	X	X	X				X	X			
9	22.0	21.3	44.8	11.417	X		X	X	X	X	X				X	X	X			
10	23.1	22.4	31.0	11.342	X	X	X	X	X	X	X	X				X	X			
10	23.0	22.3	32.3	11.348	X		X	X	X	X	X	X			X	X	X			
11	23.6	22.9	25.2	11.306	X	X	X	X	X	X	X	X	X			X	X			
11	23.6	22.8	26.2	11.312	X		X	X	X	X	X	X			X	X	X			X
12	24.1	23.3	20.1	11.275	X	X	X	X	X	X	X	X	X			X	X			X
12	24.1	23.2	20.9	11.279	X	X	X	X	X	X	X	X			X	X	X			X
13	24.5	23.6	16.6	11.252	X	X	X	X	X	X	X	X	X		X	X	X			X
13	24.4	23.5	17.9	11.258	X	X	X	X	X	X	X	X	X	X		X	X			X
14	24.7	23.7	15.8	11.242	X	X	X	X	X	X	X	X	X	X	X	X	X			X
14	24.7	23.7	16.1	11.244	X	X	X	X	X	X	X	X	X	X		X	X		X	X
15	24.8	23.8	16.4	11.241	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X
15	24.8	23.7	16.7	11.242	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X
16	24.9	23.8	17.0	11.238	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

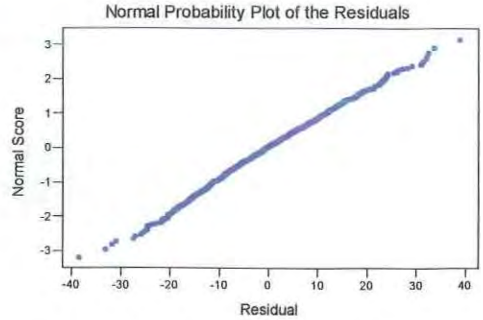
Prediction Model: Triage Number = 60.6 - 0.00690 Time + 3.25 Mon + 6.69 Tue + 9.51 Thu + 6.08 Fri + 8.07 Sat + 15.3 Sun + 8.35 Holiday - 4.32 Apr - 2.75 Jun + 3.46 Jul + 9.66 Aug + 7.26 Sep - 0.165 Rainfall + Z(t)

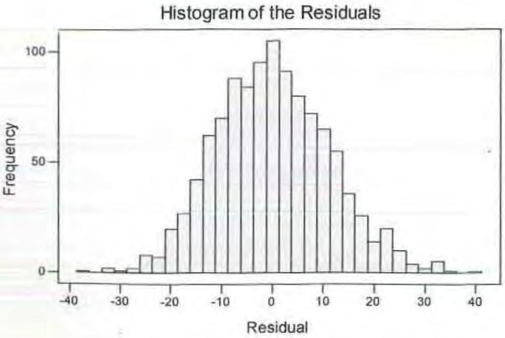
Forecasts for month of December 2000 using above Category 4 model

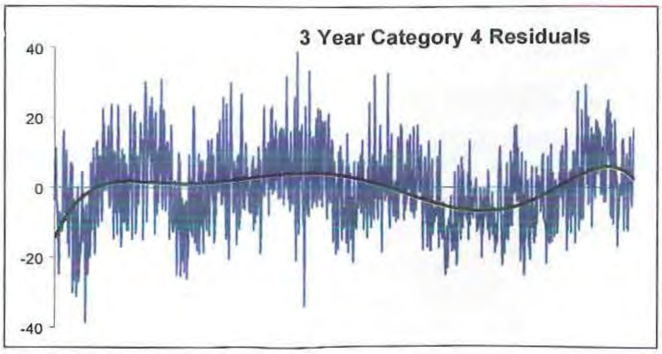
Actual	Forecast	95% CI		Actual	Forecast	95% CI		Actual	Forecast	95% CI	
		Lower	Upper			Lower	Upper			Lower	Upper
79	60	46.1	73.1	**54**	54	39.5	68.6	**58**	53	38.2	68.8
80	63	50.2	75.7	**55**	54	39.0	68.7	**64**	61	44.3	77.2
65	55	41.2	68.9	**53**	54	39.1	68.7	**69**	63	46.6	80.2
54	53	42.7	64.1	**52**	54	39.1	68.7	**67**	57	40.4	74.0
57	53	42.2	64.2	**61**	61	44.3	77.3	**43**	54	39.9	68.4
54	53	40.9	64.4	**77**	64	46.9	80.2	**50**	55	41.2	68.2
38	54	40.5	68.5	**71**	57	39.8	74.2	**43**	55	41.9	68.1
55	61	45.5	77.1	**64**	54	38.5	68.7	**63**	55	41.4	68.1
55	64	46.9	80.2	**42**	54	39.0	68.6	**62**	61	44.4	77.2
56	58	41.1	74.0	**52**	54	38.5	68.7	**77**	63	45.6	80.3

Residual Analysis of 2000 Category 4 Triage Model

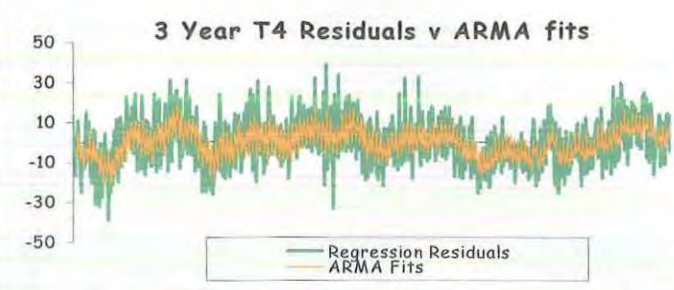








3 Year Triage Category 4 Regression Residuals Modelling



Minitab: Final Estimates of Parameters

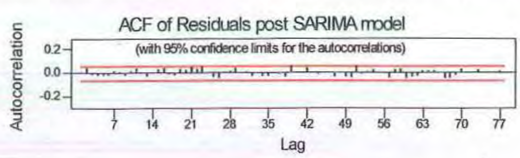
Type	Coef	StDev	T	P
AR 1	0.9769	0.0098	100.00	0.000
MA 1	0.7840	0.0326	24.08	0.000
MA 2	0.0746	0.0312	2.39	0.017
SMA 7	0.9714	0.0055	176.73	0.000
Constant	0.000417	0.001311	0.32	0.750

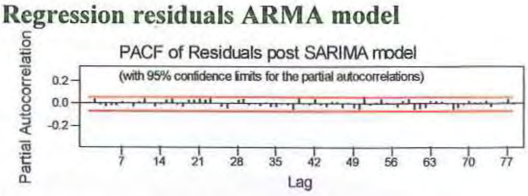
Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

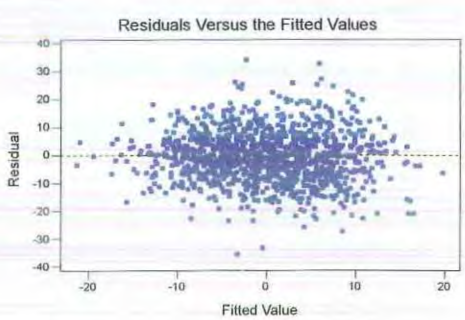
Lag	12	24	36	48
Chi-Square	6.9	19.4	28.7	38.4
DF	7	19	31	43
P-Value	0.436	0.430	0.584	0.670

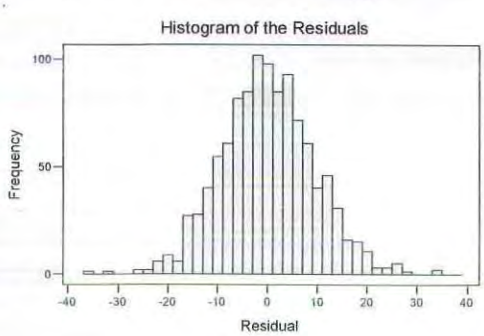
Prediction Model: Z(t) $(1 - 0.98B)\nabla_7 x(t) = (1 - 0.78B)(1 - 0.07B^2)(1 - 0.97B^7)\, w(t)$

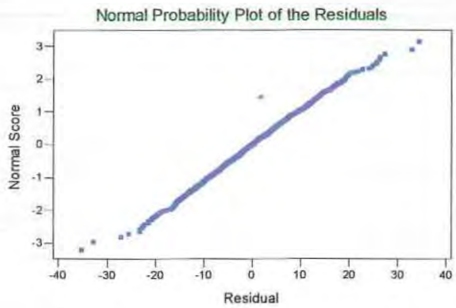
Residual Analysis for 3 Year Triage Category 4 Regression residuals ARMA model

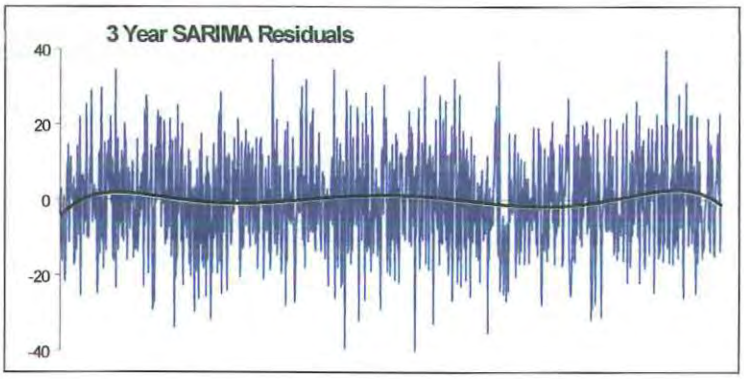




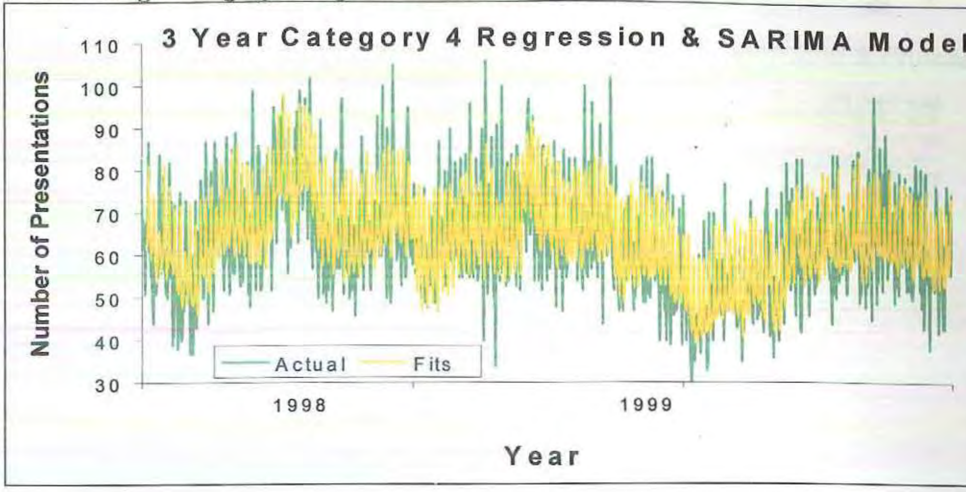








3 Year Triage Category 4 Regression + ARMA Model



Forecasts for month of December 2000 using Category 4 Regression + ARMA model

Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper	Actual	Forecast	95% CI Lower	95% CI Upper
79	72	40.6	103.8	**54**	57	22.7	90.5	**58**	59	23.3	94.0
80	68	36.3	98.7	**55**	56	21.7	90.3	**64**	71	34.2	107.4
65	59	26.8	91.7	**53**	57	22.9	91.6	**69**	67	29.6	103.6
54	56	27.1	85.9	**52**	59	25.0	93.8	**67**	60	22.9	97.1
57	56	26.1	85.6	**61**	71	35.1	107.4	**43**	56	21.3	90.5
54	57	25.9	87.2	**77**	67	30.6	103.5	**50**	56	22.2	89.9
38	60	27.5	93.5	**71**	60	23.2	97.2	**43**	58	24.1	91.1
55	72	37.3	107.1	**64**	56	20.8	90.8	**63**	60	25.8	93.4
55	68	31.7	103.5	**42**	56	20.7	90.3	**62**	70	33.6	107.4
56	61	25.4	96.9	**52**	57	21.4	91.7	**77**	66	27.9	103.7