NDSU UPPER GREAT PLAINS TRANSPORTATION INSTITUTE

Department Publication No. 283
August 2015

Quality Assurance of Emergency Management Operation Processes through Statistical Process Control

Prepared by:

Lara Piesch
Department of Industrial and Manufacturing Engineering
North Dakota State University

EunSu Lee, PhD
Upper Great Plains Transportation Institute
North Dakota State University

Acknowlegement

The authors thank Dr. Kim Vachal and Lindsey Narloch for providing the data and comments regarding public policy. The authors also thank Dr. Om P. Yadav for his support in statistical quality control and Sherm Syverson, Ron Lawler, and Jerome Differding for their expert comments regarding ambulance management and operations. The research reported was partially supported by the Mountain-Plains Consortium, which is sponsored by the U.S. Department of Transportation through its University Transportation Centers Program.

Disclaimer

This research was supported by the North Dakota Department of Transportation. The contents presented in this report are the sole responsibility of the Upper Great Plains Transportation Institute and the authors.

North Dakota State University does not discriminate on the basis of age, color, disability, gender expression/identity, genetic information, marital status, national origin, public assistance status, sex, sexual orientation, status as a U.S. veteran, race or religion. Direct inquiries to the Vice President for Equity, Diversity and Global Outreach, 205 Old Main, (701) 231-7708.

ABSTRACT

Ambulance services continually strive for improvement of their emergency response operations to provide better service to the community. The difference of a couple minutes can be a crucial factor in the survival of a patient. Therefore, it is of great importance for ambulance services to have reliable and stable response times. The study applies three phases of statistical process control to assess the process and determine the current performance. By implementing an exploratory analysis, process behavior analysis, and predictive analysis patterns, trends and process stability are analyzed. The research showed a process with very high variability while having a stable low average response time. With a percentage of 9.96% of response time operations being above the limit of nine minutes established by the North Dakota Legislature, the process calls for further improvement. To this end, several factors and characteristics, which can be considered in future management decisions, were detected.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Healthcare providers continually strive to improve their services for better treatment of their patients. In health care, monitoring quality to detect deterioration or improvement after the implementation of new procedures is crucial to overall performance. In health care, many target values can be categorized as higher-the-better or lower-the-better values and can therefore be monitored by statistical process control (SPC). The understanding of process variation through control charts is the foundation for successfully eliminating undesirable variation and creating procedures that result in reliable output.

A control chart is a chronological time series plot of measurements of important variables. Any quantities of interest, such as averages, proportions, or rates, can be plotted. Furthermore, upper and lower reference thresholds can also plotted. These so-called control limits are calculated using process data and define the natural range of variation in which all data points naturally should fall. Points outside of the control limits indicate an abnormal process variation and can be analyzed for cause and effect relationships. In addition, observing trends through control charts can reveal quality improvement or quality deterioration, especially when values cross control limits.

In this study, medical emergency response times are analyzed for quality and consistency. When people call 911 in a medical emergency, they want help immediately. Therefore, emergency management response time data clearly consist of lower-the-better values. Response times for the state of North Dakota are mandated by the North Dakota Legislature, with target values being nine minutes in urban areas, 20 minutes in rural areas, and 30 minutes in frontier areas (North Dakota Legislature, 2010). By analyzing response time values through control charts, extraneous process variation will be identified and then further investigated. The analysis of the data will provide insight to possible environmental condition and process characteristics that affect the outcome of emergency response time performance.

Furthermore, statistical process control analysis will provide guidance for healthcare providers and ambulance services for making improvements to emergency management processes. Faster response times are beneficial, not just to the ambulance service, but to the overall community by providing the best possible care to patients.

1.1 Scope of Study

Different environmental conditions and characteristics can affect emergency response time performance. The goal is to detect abnormal process variation and behavior to assess the reasons and initiate further improvement and process stabilization. To this end, the following correlations were analyzed for response time performance:

- Seasonal variation (summer/winter)
- Incident location and ZIP code impact (highway/residence/public place/acute care facility)
- Peak time performance (performance during peak time of incoming emergency calls)

The monitoring of response rates over time in a medical context has been studied by many researchers. However, these studies focused mainly on issues directly related to health such as mortality rates, rates of disease, or rates of congenital malfunctions. The monitoring of emergency response times through statistical process control is mostly uncharted. Many studies modeling emergency response times merely investigated the optimal location and routing settings for ambulance services under different constraints. These studies have covered many different aspects, such as rural areas (Lee, et al., 2013), priority constraints (Oran, et al., September 4-6, 2012), or time window constraints (Solomon, 1987).

It must be noted that the use of control charts in healthcare-related applications can differ from industrial practice. In general, public-health-related monitoring puts less emphasis on sampling only a portion of the output of a process than industrial SPC does. Thus, many sampling-based approaches used in industrial SPC are not fully applicable in healthcare-related fields (Woodall, 2006). This is especially true for emergency response times, as it is impossible to re-measure a certain operation. Every measurement can only be obtained once, so every operation is considered and no sampling is conducted. Furthermore, goal setting for emergency response times is somewhat difficult as the optimal value would be 0 minutes, a value that is impossible to attain.

1.2 Medical Emergencies

An average of 70% of 911 calls are medical emergencies (Olympia, 2014). While these calls can vary in type (location, incident, number of casualties, etc.) and cause, response time is crucial for all of them. When responding to medical emergencies, the probability of patient survival is directly related to the elapsed time until treatment is first received.

Victims of cardiac arrest are in special need of urgent treatment because the body is no longer sufficiently supplied with oxygen. Thus the brain is in danger of damage due to a lack of oxygen. The American Heart Association established standard response times for victims of cardiac arrest based on the probability of survival curve. After four minutes, basic life support must be undertaken; after eight minutes, advanced life support must be introduced (McIntyre, 1991). A delay in life support can have severe consequences for the patient, ranging from a permanent reduction in quality of life to death.

Additionally, studies have shown that response time is also linked to the survival of severe trauma victims. A medical research study showed that survival of major trauma victims was linked to a short response time (average 3.5 minutes) and patients with less severe traumas did not survive if the response time was considerably longer (average 5.9 minutes) (Feero, et al., 1994; Valenzuela, et al., 1997).

1.3 Emergency Response Time Allocation

Overall, emergency response time, defined as pre-hospital time, can be segmented into four components.

a. Activation interval
b. Response interval
c. On-scene interval
d. Transportation interval

The first component is the activation interval, which is defined as the time between when a 911 call has been received and when the emergency vehicle is enroute to its patient. Performance related to activation time is merely determined through internal processes and procedures by the ambulance. In the state of North Dakota, this time must not exceed 10 minutes (North Dakota Legislature, 2010). The second component is the response interval, which is traveling time of the ambulance to the scene of the incident. Response time is subject to environmental conditions and external forces not controlled by the ambulance combined with internal operations management. However, it is possible to detect characteristics (e.g., weather, road conditions, and traffic) that can influence the time performance and implement adaptations to secure a stable process time. The on-scene interval takes place while emergency personnel are at the scene of the incident. This interval is the combined set-up time and treatment time at the scene of the incident with the goal of stabilizing the patient so that transportation to a treatment facility is possible.

The pre-hospital time is concluded by the transportation interval, which is when the patient is transported to the hospital. When deciding on a transportation route, many factors are considered. Although time is a dominant factor, safety for the patient and emergency personnel, road conditions, and accessibility of location play a role in the decision (Patel, et al., 2012; Lee, 2014).

This study focuses solely on the response interval of the overall pre-hospital time. The measured time was taken as the interval between time of dispatch and arrival at the accident scene.

2. DATA ANALYSIS

2.1 Data Source

The data analyzed in this study are composed of emergency response activities by F-M Ambulance Service Inc. (911 response) in 2014 (January-October) within the city limits of Fargo, ND, and West Fargo, ND. Therefore, a nine-minute control limit can be applied to all operations. Furthermore, all ambulance services in North Dakota are required by state law to meet response time standards 90% of the time (North Dakota Legislature, 2010). Analysis showed 9,627 emergency response operations by F-M Ambulance Service Inc. in Fargo and West Fargo for the first 10 months of 2014. The average response time was 5.54 minutes with a median of 5 minutes while ranging from 0 minutes to 84 minutes. Out of 9,627 emergency responses, 8,668 operations had a response time of less than or equal to 9 minutes. The response time standards set by the North Dakota Legislature were met 90.04% of the time.

The dataset described above showed an excellent average response time, but the nearly 10% of calls that exceeded the nine-minute response time were outliers that could be subjected to closer analysis of the variations of the process. Those data points indicating a response time of more than 9 minutes were almost double the average response time. The variance of the overall data is 13.51 minutes with a standard deviation of 3.68 minutes. Further analysis will determine which factors are most likely to cause variance in the process resulting in response times longer than 9 minutes. In addition, time-related characteristics, which possibly increase average response times, will be explored in the next section, Exploratory Analysis.

2.2 Exploratory Analysis

2.2.1 Seasonal Variation

In environments with distinct seasons, the performance variation due to seasonal changes is of interest. To this end, all data points were segmented into either summer or winter (Figure 2.1). Summer was defined as starting in April and lasting through September, with winter defined as the months of October through March. The analysis of emergency response data for the months of April through September showed following distribution:

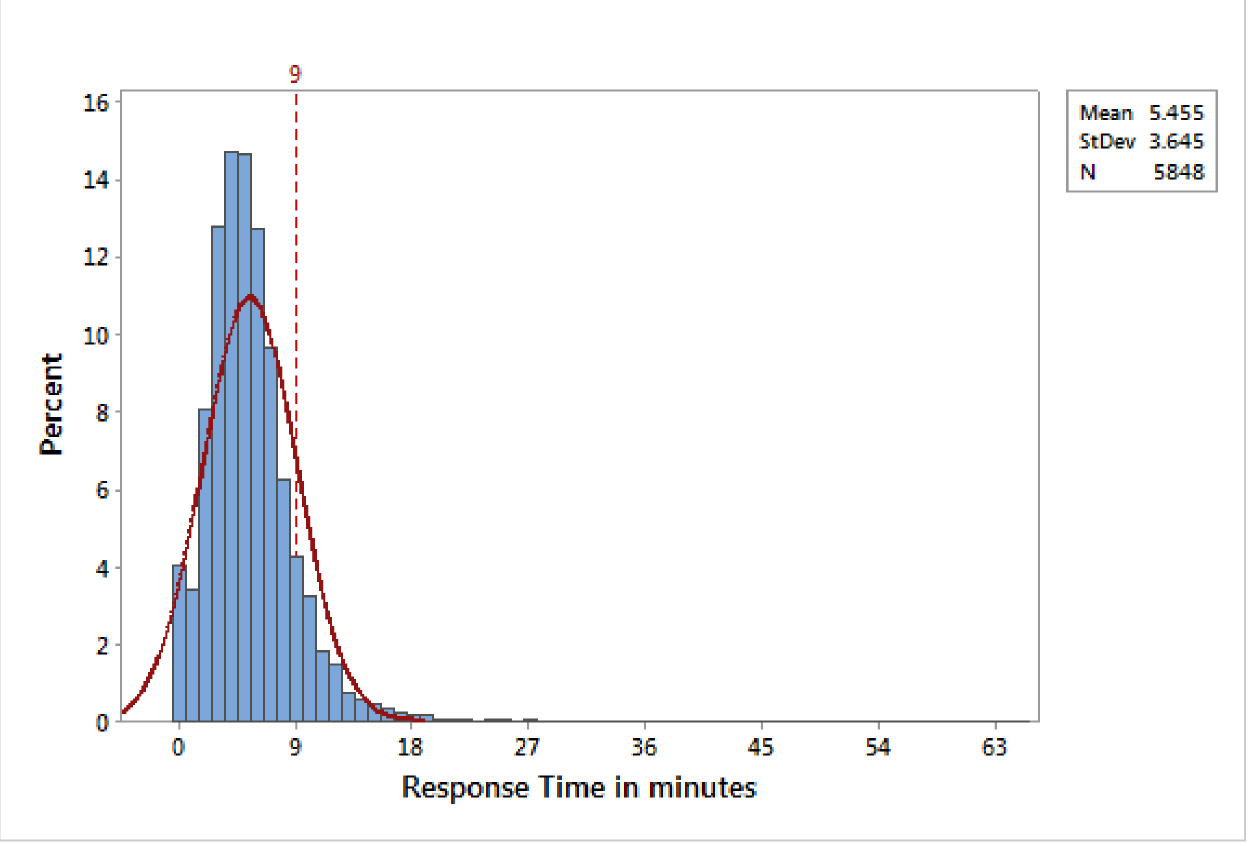

Figure 2.1 Histogram of response time during the summertime

The same histogram graph was plotted for the months of October through March. With a mean of 5.46 minutes and a variance of 13.29, the summer season has a slightly lower average response time compared with the winter season, with 5.68 minutes and a variance of 13.83. In the winter season, F-M Ambulance Service Inc. met the response time standards in only 3,384 cases, missing the projected goal of 90% slightly with 89.56%. The almost identical results can be explained by contradicting occurrences in the seasons that equalize the overall impact on the response time.

While icy roads and snow in the winter can cause delays, almost 90% of road construction in North Dakota occurs in the summer, with many speed limits being lowered and possible traffic congestion during rush hours (North Dakota Department of Transportation, December 2013). Further investigation showed that emergency response times for rural and frontier areas in North

Dakota during the summer season have, on average, higher response times because road construction is a more dominant factor in long-distance responses.

In Figure 2.2, the average monthly response time is plotted showing a strong seasonal pattern with higher average response times in the winter. A continuous decrease can be seen from January until June, which presents an abnormal trend outlier data point in this chart. The average response time then slowly increases again as the winter season approaches. The month of June should be further investigated to explain the unusual high average. An assumption here is that the outlier has a staffing-related cause, with many new employees starting in June and working over vacation periods for experienced staff.

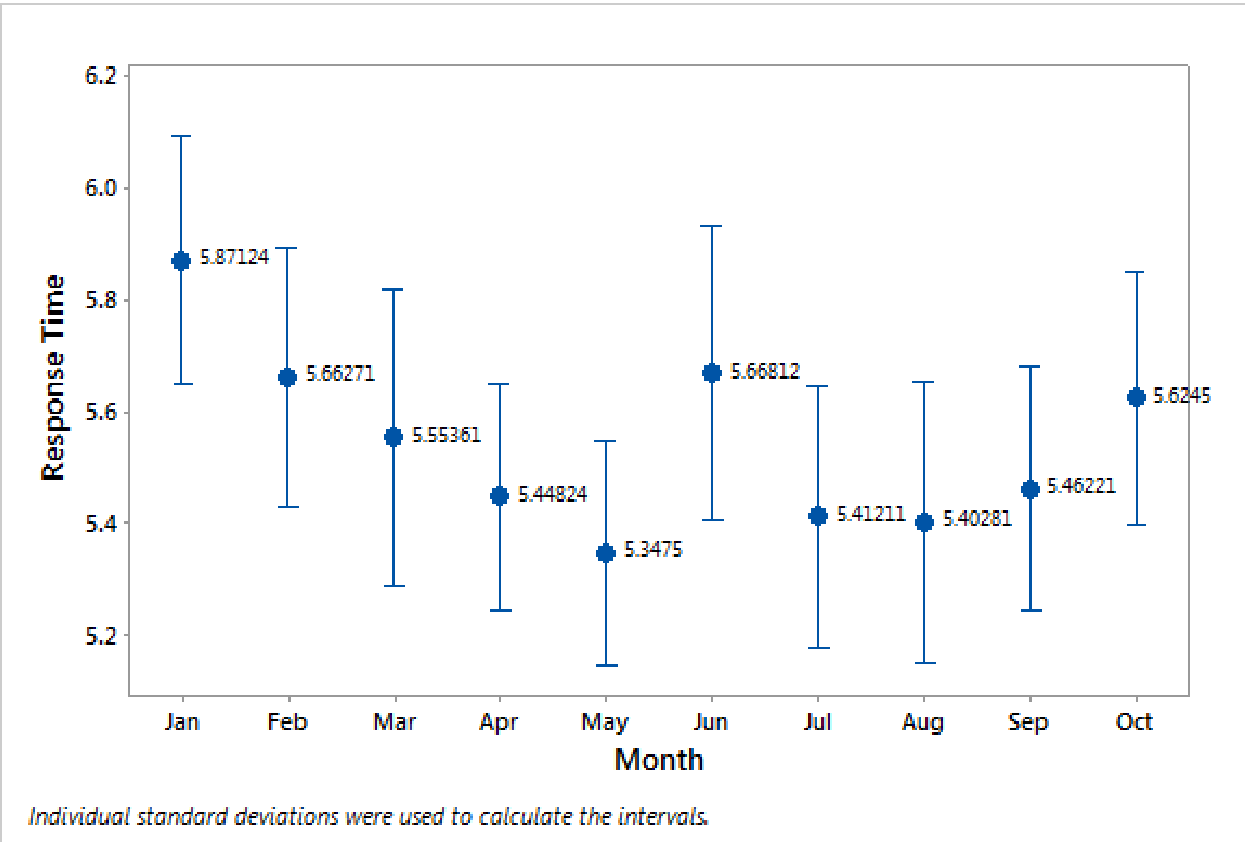

Figure 2.2 Interval plot of average response time by month

The variance for the 10 months ranges from 9.65 in May to 17.62 minutes in March and is therefore subject to great variability. The very high variance in March could be explained by March being the end of the winter season. March is subject to days with perfect weather and road conditions, but can also have large amounts of snow and icy road conditions with difficult driving.

2.2.2 Impact of Types of Incident Locations and Areas by ZIP Code

Incident locations of medical emergencies are wide-ranging and diverse. As depicted for the city of Fargo in Figure 2.3, the ambulance service was most often called to a residence (41.27%) followed by an acute care facility (24.99%), traffic way (11.29%), and bar/restaurant (11.11%). Although emergency calls to residences are most frequent, the average response time of 5.86 minutes is slightly longer than the general average of 5.54 minutes. Furthermore, note that 10.92% of residence calls have a response time greater than 9 minutes. No other major incident location shows such a high percentage of data points being outside the specified limit of 9 minutes. This issue might be related to difficulties of finding the correct private addresses, which may not be a problem of the same magnitude for public places or well-known restaurants and bars. The average response time for the other three major incident locations ranges from 5.15 minutes to 5.34 minutes and is therefore below the overall mean. However, the variance again is very high for certain incident locations, including acute care facility with 18.29 minutes and traffic ways with 15.53 minutes.

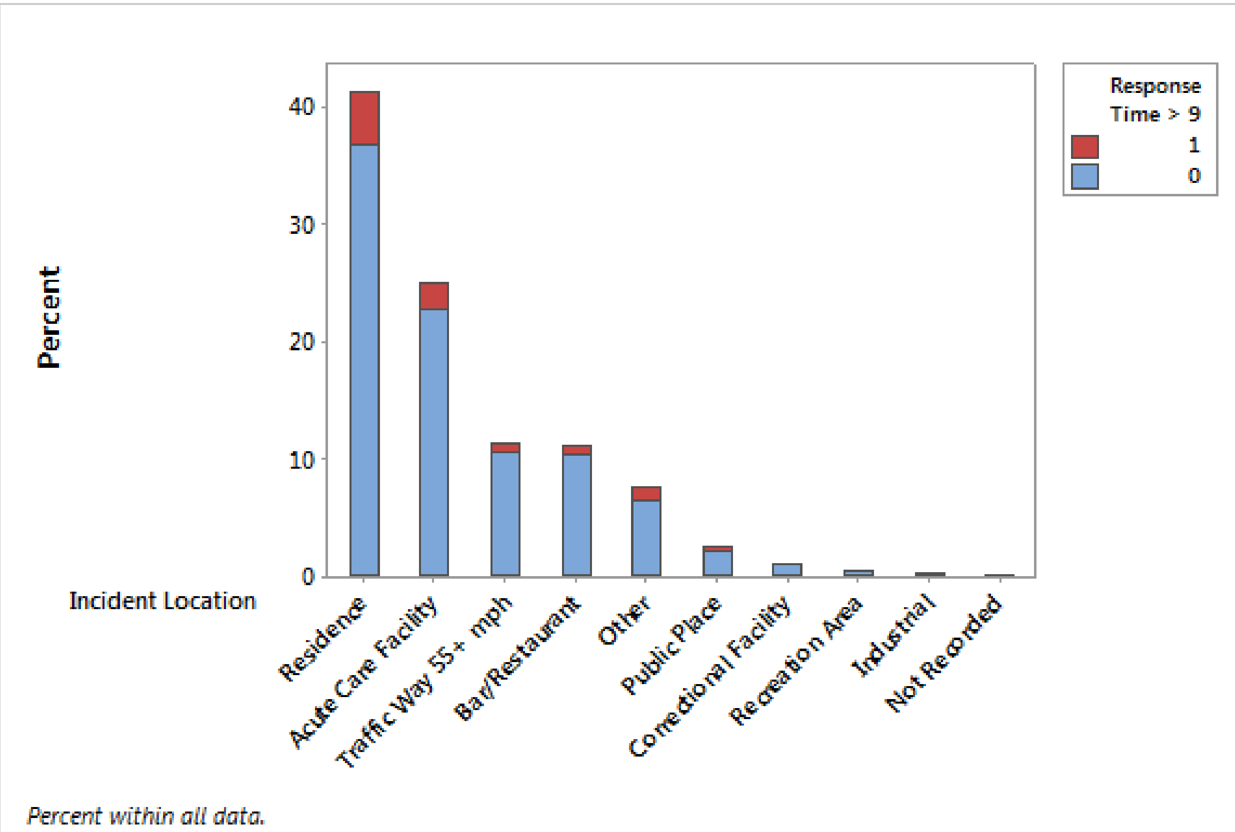

Figure 2.3 Chart of incident location distribution

Further location analysis can be performed by ZIP code zones within the cities (Figure 2.4). Fargo and West Fargo are subdivided into five ZIP code areas: 58078 (West Fargo), 58102 (north Fargo), 58103 (downtown Fargo), 58104 (south Fargo), and 58105 (North Dakota State University [NDSU] Campus). Data evaluation in Figure 2.4 showed that response time performance is dependent on the incident ZIP code. While north and downtown Fargo have low response times of around 5.2 minutes, south Fargo, West Fargo and the NDSU campus have response times that average more than 6 minutes. Variances are again wide ranged, having an inexplicable maximum in north Fargo of 17.74 minutes, while the NDSU campus only has a variance of 5.48 minutes. North Fargo shows the highest variability. However, the confidence interval is rather small compared with other ZIP codes such as the NDSU campus. This is because of the different size of the samples. While north Fargo had 3,163 calls in 2014, the NDSU campus only had 61 calls. The number of the sample size n is considered in the calculation of the confidence interval and therefore leads to highly differing output compared with the variance (1).

Confidence intervals at α=0.05 are defined as follows:

$$\bar{X} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \tag{1}$$

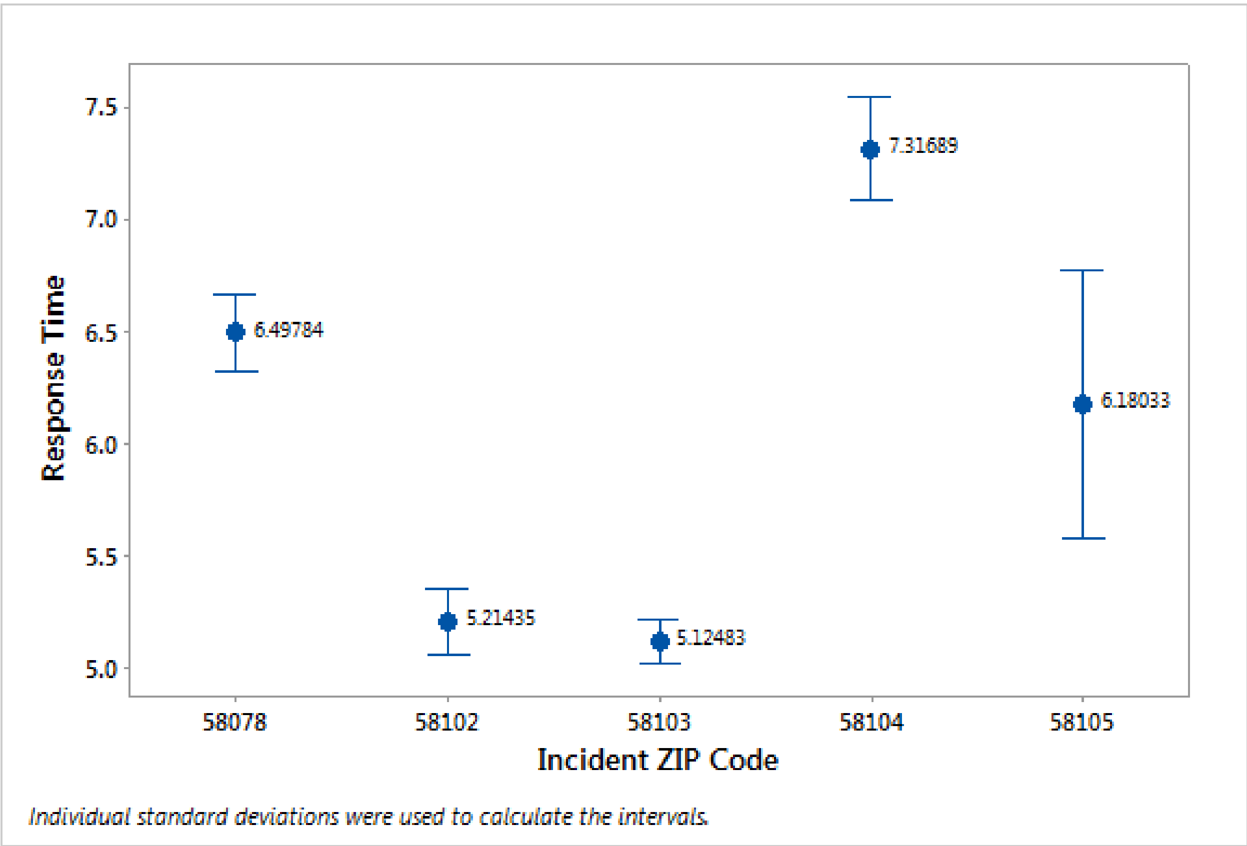

Figure 2.4 Interval plot of average response time by incident ZIP code

2.2.3 Peak Time Performance

Emergency management systems must be ready for incidents at any time. Although medical emergencies occur randomly throughout a day, they do occur more frequently during several time intervals. The distribution of time of dispatch (Figure 2.5) shows the cumulative number of calls having taken place in the time interval of 12 p.m. to 8:30 p.m.

However, the change in number of response operations is gradually distributed over time, with the lowest percentage of calls from 5 a.m. to 6 a.m. (1.76%). From 6 a.m. onward the number of calls steadily increases until noon. During the peak time of noon to 8 p.m., 49.08% of all emergency response requests are received. The period from 8 p.m. until 6 a.m. is characterized by a slowly decreasing number of calls.

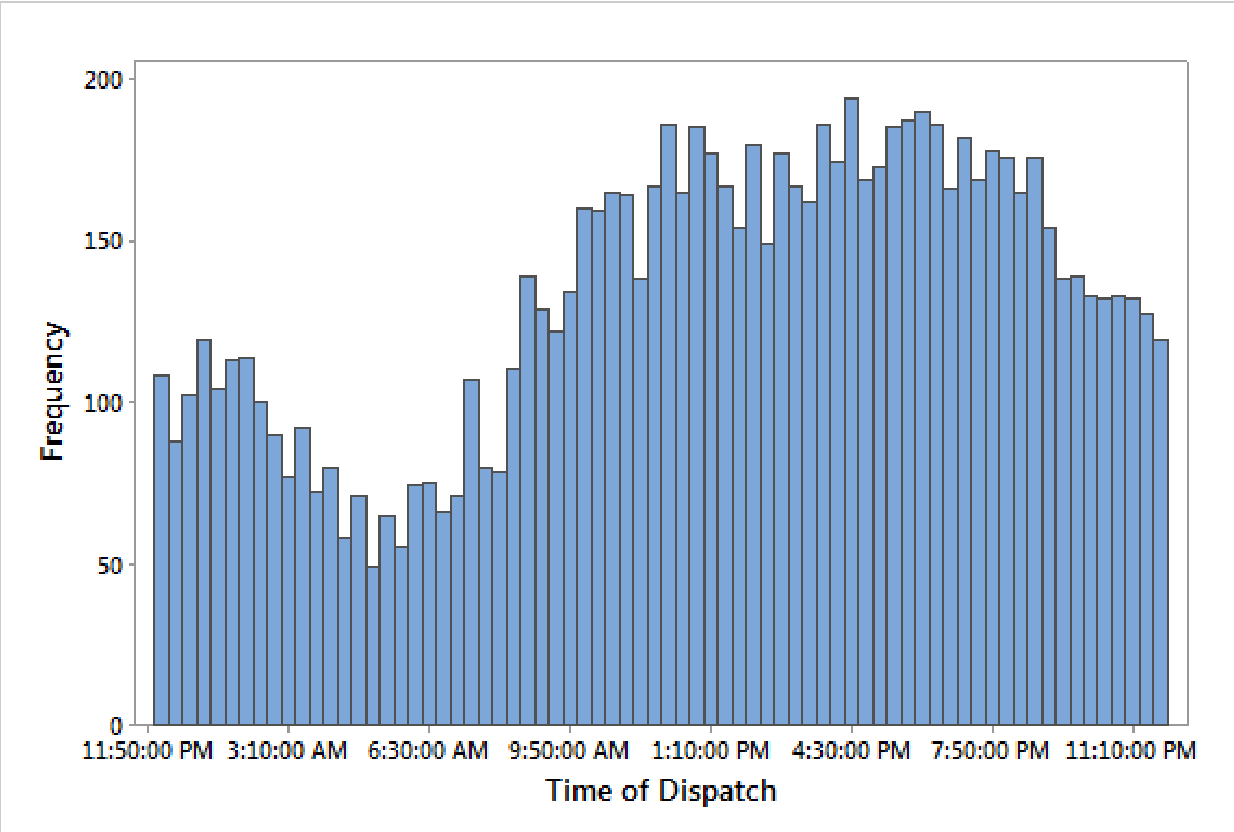

Figure 2.5 Histogram of time of dispatch

The information obtained from such histograms is crucial for healthcare providers as they make decisions regarding staffing and equipment. This is especially true when providers serve rural areas where overall pre-hospital time can be up to one hour, resulting in having resources busy with individual incidents for a very long time period.

Response time averages by hour of dispatch were also analyzed, showing unusually high values for 3 a.m. and 5 a.m. of above 6 minutes (Figure 2.6). Most of the observed data points are located around the average mean of 5.54 minutes, and very low averages can be found at 7 a.m. and 2 p.m. as well as at 10 p.m. and 11 p.m.. Decreased traffic at night would suggest lower response times after 8 p.m.. However, the major increase in average response time around 3 a.m. and 5 a.m. contradicts this theory and suggests that further time-related factors are relevant for the final response time. The short trends from 5 a.m. to 7 a.m. and 2 p.m. to 5 p.m. are very distinct and should be analyzed for management strategies that are currently applied, such as shift staffing or location strategy.

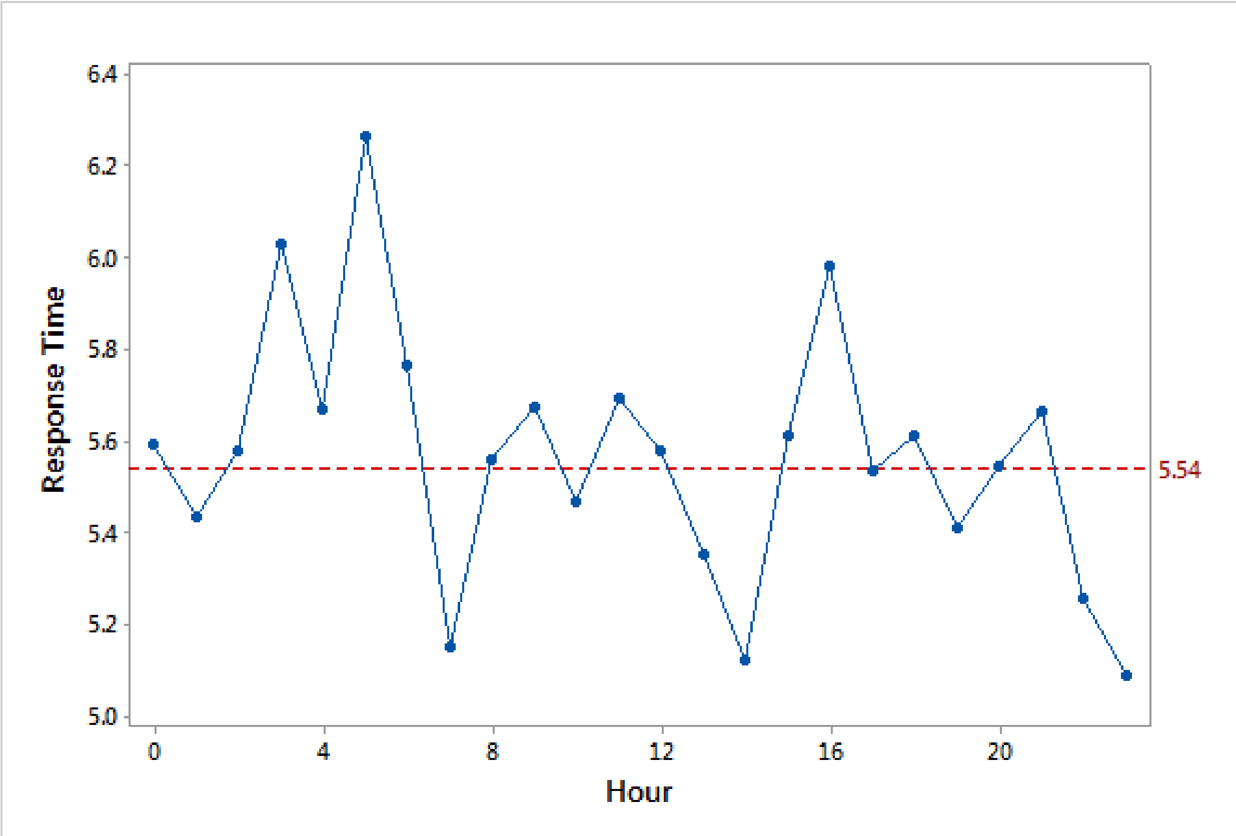

Figure 2.6 Time series plot of average response time by hour of dispatch

Closer analysis of the data points at the hours of 3 a.m. and 5 a.m. showed that a major portion (53.94% and 60.36%) of response operations were calls to residences in Fargo. No other hour of dispatch has such high percentages of calls coming from emergencies at residences. This finding might explain to some degree the unusual high response time averages; however, with the average response time for residences being 5.86 minutes, it cannot account for the average response time to be above 6 minutes.

2.2.4 Summary of Data Analysis

The data were analyzed to find characteristics that influence emergency response time. In this study, the focus was set on "seasonal variation," "hour of dispatch," and location influence such as "ZIP code" and "incident location." In all four categories some major findings were made that will help to provide guidance for process improvement options that can be implemented in the future.

The analysis showed that the average response time during the winter season is slower than during the summer. The facts recommend a better preparedness to respond to calls during winter conditions in the region. Surprisingly, the month of June breaks the seasonal trend, having a higher average than expected. Further examination for possible root causes of that change will help to eliminate the occurrence in the future. The location analysis also showed that the incident location and ZIP code have an influence on the response time and vary highly in the frequency of incidents. The ambulance service is most often called to a residence in Fargo, which has the highest percentage of response time values greater than 9 minutes (10.92%). The ZIP code allocation showed that calls to south Fargo and West Fargo, in particular, have higher response times than the average mean. When further analyzing peak time performances, the histogram showed a distinct distribution of operations over the 24-hour period of one day. This information is very useful when considering staffing for shifts. Furthermore, the period between 3 a.m. and 5 a.m. showed high average response times. This should be further analyzed by the provider to identify the initial causes. In the following chapter, these influences will be further investigated by developing predictive and causal analysis.

3. MODEL DEVELOPMENT

An important part of process control is predicting the possible process behavior. To this end, two approaches will be considered in the following sections. On one hand, the process behavior analysis will be developed through Shewhart Control Charts. On the other hand, a regression model will be set up to verify the observations made during the data analysis phase. The regression model will not only depict the significance of the different characteristics with regard to response time, but will also value the contribution of every individual aspect.

3.1 Process Behavior Analysis

The goal of process behavior analysis is to understand the behavior of a process by plotting the measured data points in different control charts. The basic charts used in this project are X-bar and S control charts, which are used to analyze the behavior of the mean and variance of the process, respectively. Especially when having excellent average response times and high process variance simultaneously, the focus should lie on the supervision and detection of abnormal process variation. In fact, it is seen as a violation of a fundamental rule when X-bar charts are shown without the associated S chart. The standard deviation chart is essential to the control chart concept (Devor, et al., 2007).

Furthermore, cumulative sum (CUSUM) charts will be implemented because CUSUM charts can detect small and moderately sized sustained changes in quality on average much more quickly than Shewhart control charts (Woodall, et al., 2012). It is therefore an optimal tool to detect deterioration in the response time characteristics. In addition to CUSUM charts, a p-chart will provide an overview of the daily percentage of response operations with a response time of more than 9 minutes. It will show if certain days or time periods are more likely to have response times greater than 9 minutes.

3.1.1 X-Bar and S Control Chart

In Figure 3.1, an X-bar and S control chart is plotted for the 2014 daily average response times. Every sample represents one day from January 1 to October 31, 2014. Because a different number of calls come in daily, the tests were performed with unequal sample sizes. As analyzed in the previous chapter, the mean is very consistent and rarely exceeds the upper control limit (UCL) or lower control limit (LCL). Only four samples exceed the control limit, three of them being above the UCL. However, the standard deviation is highly variable with many outliers throughout the year, with more than 20 data points beyond the control corridor. Note that for a period of almost three months (March 29 to June 20, or samples 91 through 172), the process appears to be stable, having no outliers above the UCL.

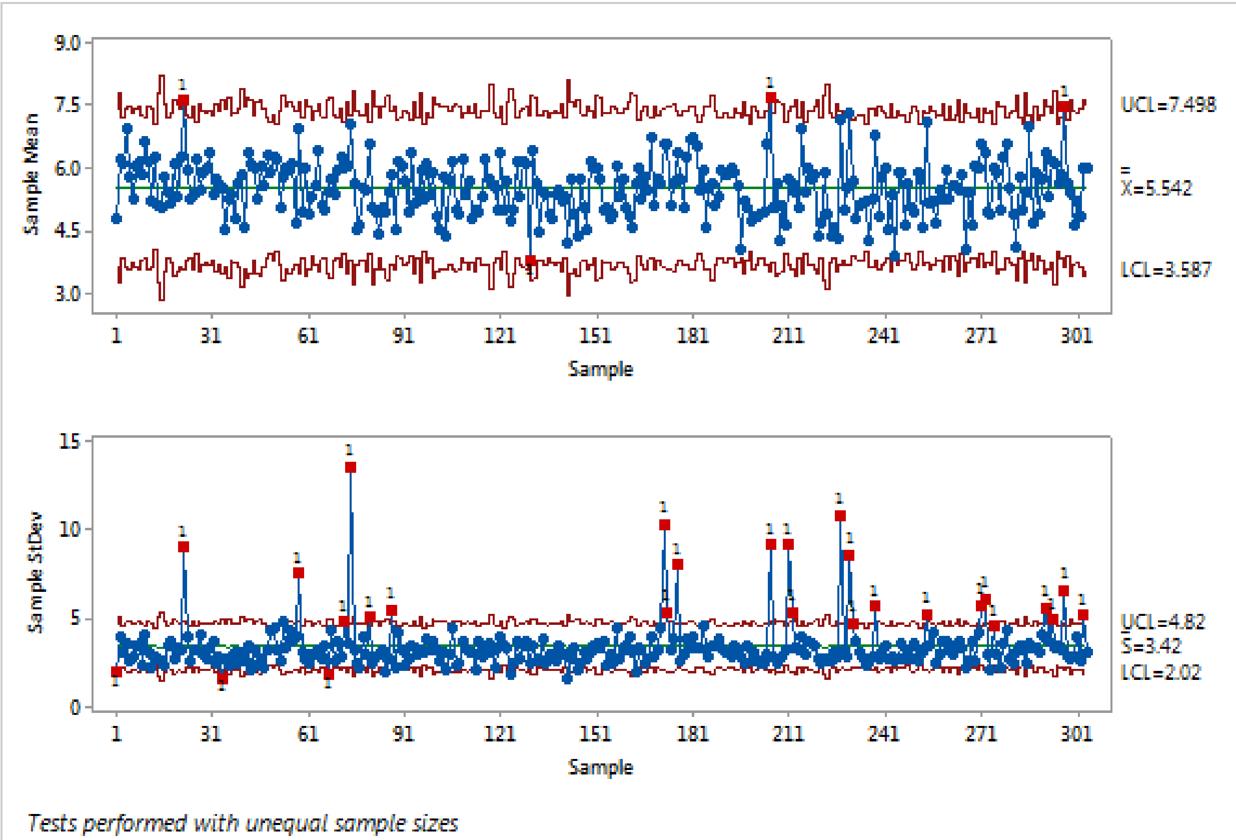

Figure 3.1 *X*-Bar and S charts of response time for 2014 sampled by days

A closer analysis of the stable time period shows that the mean is 5.43 minutes with 241 of 2,603 response operations being late (response time more than 9 minutes). That gives it an on-time response rate of 90.74%, which is 0.7% higher than the overall rate. Furthermore, the range is lower with a maximum of 27 minutes, which gives a variance of 10.13 and standard deviation of 3.28. For those response times of greater than 9 minutes, the average arrival time is 12.27 minutes, while it is 13 minutes for the overall year. Overall, all measurements of the descriptive statistic show better values.

The control limit equations (2 – 4) for the sample mean (X-bar) were implemented as follows:

$$UCL = \bar{\bar{X}} + 3\frac{\sigma}{\sqrt{n}} \tag{2}$$

$$Center\ Line = \bar{\bar{X}} \tag{3}$$

$$LCL = \bar{\bar{X}} - 3\frac{\sigma}{\sqrt{n}} \tag{4}$$

Further, the control limits for the sample standard deviation were calculated as follows (5 – 9):

$$UCL = s_n c_4 + 3c_5 s \tag{5}$$

$$\boldsymbol{Center\ Line = \overline{S}} \tag{6}$$

$$\boldsymbol{LCL = s_n c_4 - 3 c_5 s} \tag{7}$$

With c_4 for being

$$\boldsymbol{c_4 = \sqrt{\frac{2}{n-1}} \frac{\left(\frac{n}{2} - 1\right)!}{\left(\frac{n-1}{2} - 1\right)!}} \tag{8}$$

and c_5 being

$$\boldsymbol{c_5 = \sqrt{1 - c_4^2}.} \tag{9}$$

Where s is

$$\boldsymbol{s = c_4 \overline{S}} \tag{10}$$

$\bar{S}$ can be calculated by

$$\boldsymbol{\overline{S} = \frac{\sum_{i=1}^{n} h_i \frac{s_i}{c_4}}{\sum_{i=1}^{n} h_i}} \tag{11}$$

with h_i being

$$\boldsymbol{h_i = \frac{c_4^2}{1 - c_4^2}} \tag{12}$$

Because both control limits are either direct [(2) - (4)] or indirect [*c*4 (8)] dependent on the sample size *n*, the control limits are not constant over time. This is because of the above mentioned unequal sample sizes for each day.

3.1.2 CUSUM-Chart

When analyzing CUSUM (Cumulative Sum Control Chart) charts, the goal is to detect trends by having cumulative averages moving away from the target value. To this end the cumulative sum of the deviation from the target value is plotted over time. The upper part of the CUSUM chart is designed to detect increases in the mean and the lower part is designed to detect decreases in the mean (Woodall, et al., 2012).

The CUSUM chart for the year 2014 is graphed in Figure 3.2, showing a very distinct pattern. The graph clearly reflects the previously detected seasonal trend from January until the end of

June with overall decreasing values. After the end of June the process suddenly seems to be very variable with data points leaving the upper and lower control limits. Note that these phenomena are not as easily detected in the X-bar and S control charts, therefore CUSUM charts are a good addition to statistical process control. Furthermore, the addition of CUSUM charts give management the opportunity to detect and react to process deterioration much earlier.

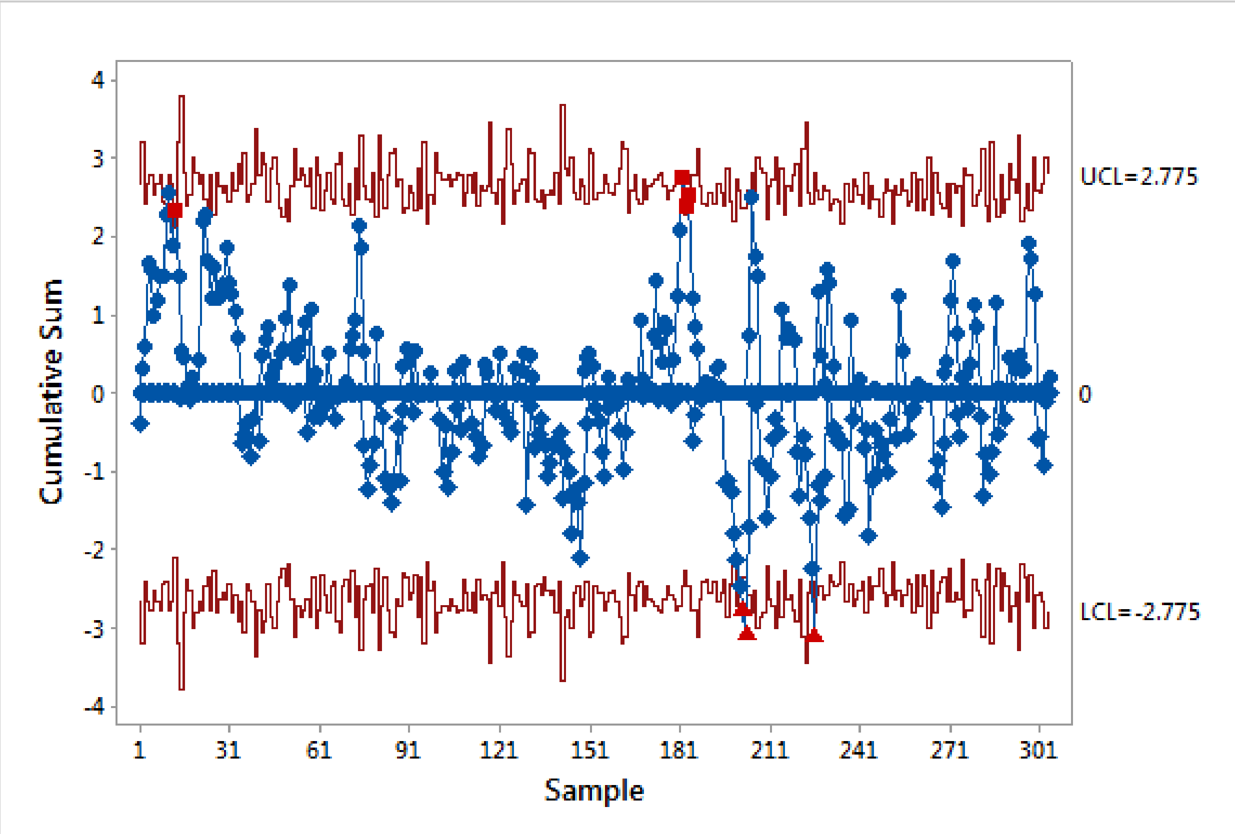

Figure 3.2 CUSUM chart of response time

3.1.3 *p*-Chart

The p-chart was developed in industry as a graphic tool to interpret and reduce sources of variability in manufacturing. Recently, it increasingly has been implemented in health care for continuous quality control and quality improvement research. The p-chart is a combination of a time series analysis and a graphical presentation of data by which successive indicators are plotted in chronological order (Duclos, 2010). For plotting the adverse events in this case study, the response operations were subdivided into response time > 9 minutes (i.e., defective) and response time ≤ 9 minutes (i.e., non-defective), with the ratio of response time > 9 minutes being graphed.

In Figure 3.3, a p-chart for late responses is graphed with every sample representing a day in 2014 (January-October). The average percentage of responses greater than 9 minutes is 9.96%; however, the chart also reveals that the ratio is highly variable. During days 61 and 151, a cumulated number of 0% late responses can be found; other than that, the chart does not show many significant observations. Only one point leaves the UCL, having a percentage of late operations greater than 27%. Thus this p-chart, showing no major deviations, trends, or abnormalities, contains very valuable information. It supports the findings from the previous chapters that no single characteristic causes delays and response times greater 9 minutes. The chart concludes that bad weather conditions cannot be a general reason for the number of late response operations since the percentage is highly variable throughout the year and not dominant in the winter. It points to the fact that an overall management strategy improvement possibly could reduce the number of response times greater than 9 minutes for the overall year.

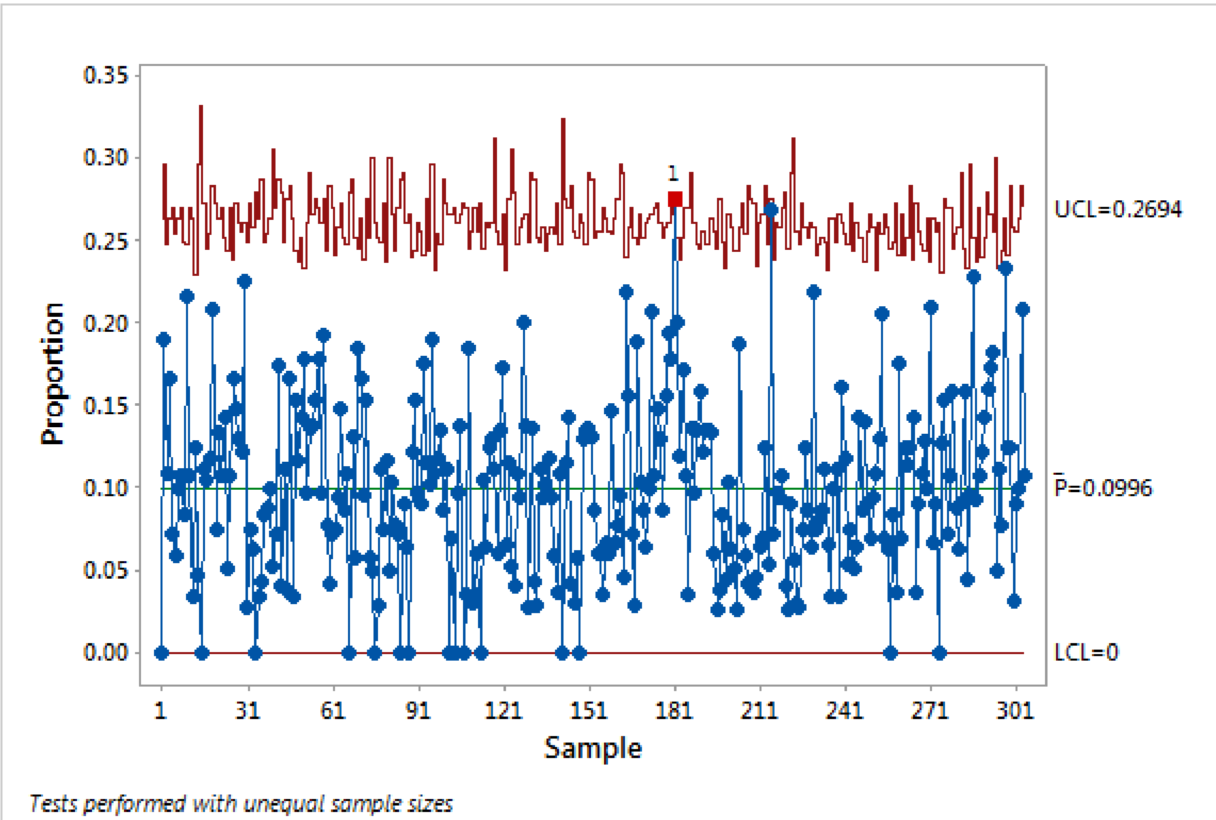

Figure 3.3 p-Chart with fraction defective values using response time greater than 9 minutes

3.2 Predictive Analysis

A regression model was used for the causal analysis of the process. The regression model is a statistical tool to investigate relationships between variables. Further, the causal effect of variables upon each other can be investigated by assembling data on the underlying variables of interest. When employing regression, the quantitative effect of the causal variables upon the influenced variable can be estimated. The regression model also assesses the statistical significance of the variables through use of a confidence level, hence the degree of confidence that the true relationship is close to the estimated relationship (Skyes, 1992).

This study constructs a model with three qualitative independent variables, the first at three levels ($ZIPCode_1$ through $ZIPCode_5$), the second at 24 levels ($HourofDispatch_1$ through $HourofDispatch_{24}$), the third at 10 levels ($IncidentLocation_1$ through $IncidentLocation_{10}$), and the fourth at 10 levels ($Month_1$ through $Month_{10}$). The regression equation to estimate expected response time (13) with main effects is stated as follows:

$$\begin{aligned} \boldsymbol{E(ResponseTime)} = \mathbf{6.652} + \boldsymbol{\beta_{i-1} IncidentZIPCode_{r,i-1}} \\ + \boldsymbol{\gamma_{j-1} HourofDispatch_{r,j-1}} + \\ \boldsymbol{\delta_{i-1} IncidentLocation_{r,k-1}} \\ + \boldsymbol{\theta_{m-1} Month_{r,m-1}} \end{aligned} \tag{13}$$

where

$\beta, \gamma, \delta,$ *and* θ: the coefficients of each independent variable
r : the number of calls (transactions)
i : the level of ZIPCode $(1 - 5)$ with the reference category of 58078
j : the level of HourofDispatch $(1 - 24)$ with the reference category of 12:00 a.m.
k : the level of IncidentLocation $(1 - 10)$ with the reference category of Acute Care Facility
m : the level of Month $(1 - 10)$ with the reference category of January

In this case study, the variable of interest is the response time (dependent variable) and the influencing variables (independent variables) are incident location, ZIP code, hour of dispatch, and month. These are the same variables analyzed in Section 3.1. By developing a regression model, the significance of these variables on the response time outcome can be proven and the average impact can be estimated. The regression analysis is summarized in Table 3.1.

Table 3.1 contains statistical measures providing model fitness and significance of the results. The p-values are very low for every causal variable. A p-value below 0.05 implies that an independent variable is significant for the outcome of the influenced dependent variable, which in this case is the response time. Therefore, all tested independent variables are significant by p-values ranging from 0.000 to 0.023, which is lower than $\alpha = 0.05$. On the other hand, the R-squared value shows a very low result of 5.35%, which means that only 5.35% of the data points can be explained by the generated model. An acceptable value when generating regression models must lie much higher than 30% in general for transportation analysis, thus this model does not seem suitable for future prediction of response times. Two possible factors could explain the low R-squared values of the regression model. The larger the sample size, the greater the likelihood of obtaining a significant relationship between the dependent variable and the independent variables (Cohen, 1983). With a sample size of 9,627 data points, it is highly possible that the sheer number of data points is a reason for the employed regression to mistakenly detect significance. Furthermore, response time may depend on many variables that the data set did not include. Foremost among these is the distance traveled to the location of emergency. Routing and human factors also play into overall performance.

Table 3.1 Testing a group of independent variables

Source	DF	Adjusted SS	Adjusted MS	F-Value	P-Value
Regression:	45	6960	154.67	12.04	0.000
Incident ZIP Code	4	5133	1283.22	99.88	0.000
Hour of Dispatch	23	541	23.51	1.83	0.009
Incident Location	9	722	80.27	6.25	0.000
Month	9	247	27.46	2.14	0.023
Model Summary	S	R-sq	R-sq(adjusted)	R-sq(pred)	
	3.58	5.35%	4.91%	4.49%	

However, the model supports the findings made in Section 3, which indicated that the analyzed variable does influence the response time (e.g., wintertime is slower than summertime). The categorical variables in Table 3.1 with multi levels are directly entered as predictor in the multiple regression model. The *p*-values and coefficients of the outputs are summarized in Table 3.2 for Incident Zip Codes, Table 3.3 for Hour of Dispatch, Table 3.4 for Incident Locations, and Table 3.5 for Incident Month.

When Hour of Dispatch, Incident Location, and Incident Month are held constant, ZIP code of 58078 (West Fargo) indicates slower response time compared with 58102, 58103, and 58105, while it shows shorter response time than 58104 (south Fargo) as shown in Table 3.2. Thus the response time in 58078 between 12:00 am and 1:00 am during January at Acute Care Facility can be estimated by E(Response Time) = 6.652 + 0.000 (58078) + 0.000 ($\text{HourofDispatch}_{j-1}$) + 0.000 ($\text{IncidentLocation}_{k-1}$) + 0.000 ($\text{IncidentMonth}_{m-1}$), resulting in 6.652 minutes. The results are slightly longer than the average response time of 6.49784 minutes in 58078.

Table 3.2 Coefficients for Incident ZIP Codes

ZIP Code	Coefficient	P-Value
58078	0.000	
58102	-1.191	0.000
58103	-1.349	0.000
58104	0.855	0.000
58105	-0.286	0.549

In a similar way as we did with Table 3.2, when Incident ZIP Codes, Incident Location, and Incident Month are held constant, the time period between 12:00 a.m. and 1:00 a.m. indicates slightly longer response time than the calls during a day; however, it shows that it is slightly shorter than the period between 1:00 a.m. and 3:00 a.m. Based on Table 3.3, the expected response time between 3:00 a.m. and 4:00 a.m. during January at Acute Care Facility can be estimated by E(Response Time) = 6.652 + 0.000($ZIPCodes_{i-1}$) + 0.302 (3:00 a.m.) + 0.000 ($IncidentLocation_{k-1}$) + 0.000 ($IncidentMonth_{m-1}$), resulting in 6.954 minutes.

Table 3.3 Coefficients for Hour of Dispatch

Hour of Dispatch	Coefficient	P-Value
12:00:00 a.m.	0.000	
1:00:00 a.m.	-0.191	0.503
2:00:00 a.m.	-0.063	0.830
3:00:00 a.m.	0.302	0.332
4:00:00 a.m.	-0.094	0.772
5:00:00 a.m.	0.485	0.161
6:00:00 a.m.	0.109	0.735
7:00:00 a.m.	-0.504	0.077
8:00:00 a.m.	-0.169	0.556
9:00:00 a.m.	0.050	0.857
10:00:00 a.m.	-0.110	0.678
11:00:00 a.m.	0.118	0.657
12:00:00 p.m.	-0.018	0.945
1:00:00 p.m.	-0.226	0.392
2:00:00 p.m.	-0.415	0.114
3:00:00 p.m.	0.051	0.845
4:00:00 p.m.	0.375	0.149
5:00:00 p.m.	-0.071	0.783
6:00:00 p.m.	0.003	0.990
7:00:00 p.m.	-0.212	0.416
8:00:00 p.m.	-0.097	0.710
9:00:00 p.m.	0.090	0.740
10:00:00 p.m.	-0.407	0.139
11:00:00 p.m.	-0.592	0.033

Table 3.5 shows the coefficients (δ) of Incident Locations in the model. When Incident ZIP Codes, Hour of Dispatch, and Incident Month are held constant, Acute Care Facility indicates slower response time than Correctional Facility and Traffic Way 55+ mph (i.e., highways), while it shows faster response time than Bar/Restraurant, Industrial, Recreation Area, and Residence. The Correctional Facility indicates a shorter reponse time than other locations since the location of the facility is well known to the repondents and the only place in the study area.

The estimated mean response time for Residence at ZIP Code 58078 between 12:00 a.m. and 1:00 a.m. during January can be estimated by by E(Response Time) = 6.652 + 0.000 ($ZIPCodes_{i-1}$) + 0.000 ($HourofDispatch_{j-1}$) + 0.388 (Residence) + 0.000 ($IncidentMonth_{m-1}$), resulting in 7.04 minutes. The results is slower than the average response time of 5.863 minutes at Residence. To lower the average response time and improve the overall service, the service company should put an effort in 58078 during January.

Table 3.4 Coefficients for Incident Location

Incident Location	Coefficient	P-Value
Acute Care Facility	0.000	
Bar/Restaurant	0.069	0.603
Correctional Facility	-1.333	0.001
Industrial	1.480	0.173
Not Recorded	3.650	0.151
Other	0.622	0.000
Public Place	0.058	0.818
Recreation Area	0.339	0.580
Residence	0.388	0.000
Traffic Way 55+ mph	-0.121	0.359

Table 3.5 provides the coefficients (θ) of the variable Month. When Incident ZIP Codes, Incident Location, and Hour of Dispatch are held constant, the average response time in January is longer than any other months. Based on Table 3.5, the expected response time between 12:00 a.m. and 1:00 a.m. during January at Acute Care Facility at 58107 is estimated by the regression model (13). The expected Response Time is equal to 6.652 minutes, which is derived from [6.652 + 0.000($ZIPCodes_{i-1}$) + 0.000 ($HourofDispatch_{j-1}$) + 0.000 ($IncidentLocation_{k-1}$) + 0.000 (January)]. The response time is slower than the average response time of 5.871 in January when the incident month of calls is controlled.

Table 3.5 Coefficients for Incident Month

Month	Coefficient	P-Value
January	0.000	
February	-0.234	0.167
March	-0.295	0.071
April	-0.452	0.006
May	-0.537	0.001
June	-0.126	0.448
July	-0.421	0.009
August	-0.478	0.003
September	-0.373	0.021
October	-0.224	0.167

As a result, the worst service can be found in ZIP Code 58104 between 5:00 a.m. and 6:00 a.m. during January at Industrial location. The expected response time under the condition is 9.477 minutes.

E(Reponse Time)=6.652 + 0.855(58104) + 0.458 (5:00 a.m.) + 1.482 (Industrial) + 0.000 (January) = 9.447

On the contrary, the shortest reponse time can be expected in ZIP Code 58103, which is Fargo-Downtown, between 11:00 p.m. and 12:00 p.m. at Correctional Facility during May with the expected response time of 2.841 minutes.

E(Reponse Time)=6.652 – 1.349(58103) - 0.592 (11:00 p.m.) - 1.333 (Coprrectional Facility) - 0.537(May) = 2.841

4. CONCLUSION

In this study, quality assurance of emergency management processes was performed through the implementation of statistical process control tools. The analysis was conducted in three phases: exploratory analysis, process behavior analysis, and predictive analysis.

With exploratory analysis, the most important characteristics were extracted from the data set and analyzed for significant characteristics. Seasonal variation, types of incident location, ZIP code, and peak time performance were examined more closely as they were viewed by using several graphing statistical process control tools. The results showed various significant factors contributing to overall response time performance. A distinct seasonal pattern was discovered with an inexplicable deviation in the month of June. Furthermore, the ZIP code and incident location analysis showed a very distinct distribution of emergency requests coming from certain locations and ZIP codes. The analysis also showed that all ZIP codes and locations have different average response times with south Fargo and West Fargo having averages of more than 6 minutes. When looking at the peak time performance, contradictory observations were made. The average response time is not highest during the peak times in the afternoon but during the hours of 3 a.m. and 5 a.m. when the fewest emergency calls were received. All these significant findings need further investigation by the ambulance company to examine the root causes and to implement possible improvements.

In the second stage, the process was examined through a predictive analysis in an attempt to detect patterns and trends to predict the process behavior in the future. By using tools such as the X-bar and S control chart, CUSUM-chart, and p-chart it was possible to subdivide the year into phases during which the process showed different signs of stability and variability. Most interestingly, a period of almost three months from March 29 until June 20 was detected during which the process showed a very stable output, being in good statistical control. The analysis of the CUSUM-chart reflected the findings made in the exploratory analysis and by the X-bar and S control chart by showing the seasonal decrease of response time until June when it is then transformed into a very unstable process with high variability as seen in the X-bar and S control chart. The p-chart showing a very equal distribution of percentages of late responses throughout the year suggests that no certain activity or characteristic is causing response times greater than 9 minutes, but an overall process malfunction may be responsible. This finding suggests the ambulance service company may be able to improve its processes.

In the last phase, a regression model was implemented for further causal analysis. An improvement on further variables and grouping strategy of the data set could improve the model, which would very likely also improve the R-squared value. Nevertheless, the model evaluated the significance of the variables analyzed in the exploratory analysis and valued their contributing time factor to the overall response time.

This research will help improve the operations of the F-M Ambulance Service Inc. in the study region by reducing response times and variability in the process. The study can be expanded to Cass County and the state of North Dakota. To apply it to the state level, the historical service data should be collected and stored in a standard format to be available for analysis.

Furthermore, this method of research is not limited to medical emergency response operations but can also be used to improve the service operations of fire department, police departments, and highway patrol.

REFERENCES

Devor, R. E., Chang, T.-h. & Sutherland, J. W., 2007. *Statistical Quality Design and Control (2nd Edition).* New York: Pearson Prentice Hall.

Duclos, A., Voirin, N. 2010. "The p-control chart: a tool for care improvement." *International Journal for Quality in Health Care*, 22(5), pp. 402-407. doi: 10.1093/intqhc/mzq037.

Cohen, J., Cohen, P., 1983. Applied Multiple Regression/Correlation Analysis for the Behavioural Sciences; Hillsdale, NJ, Lawrence Erlbaum.

Feero, S., Hedges, J., Simmons, E. & Irwin, L., 1994. "Does Out-of-Hospital EMS Time Affect Trauma Survival?" *American Journal of Emergency Medicine,* 13(2), pp. 133-135.

Lee, E., 2014. "Designing Service Coverage and Measuring Accessibility and Serviceability of Rural and Small Urban Ambulance Systems." *Systems,* 2(1), pp. 34-53. doi:10.3390/systems2010034.

Lee, E., Vachal, K. & Kayabas, P., 2013. *Phase I: Positioning Emergency Medical Services for Trauma Response for Rural Traffic Crashes: Pilot Case of Williston Basin in North Dakota,* Fargo: North Dakota State University. DP-270.

McIntyre, K. a. J. L., 1991. *Text Book of Advanced Cardiac Life Support.* Dallas: American Heart Association.

North Dakota Department of Transportation, December 2013. *Biennial Report: 2011-2013,* Bismarck, ND: North Dakota Department of Transportation.

North Dakota Legislature, 2010. *North Dakota Ground Ambulance Services Chapter 33-11-01.2,* Bismark: North Dakota Legislative Branch.

Olympia, F. D., 2014. *Fire and Emergency Medical Service Master Plan,* Olympia: Olympia Gov..

Oran, A. et al., September 4-6, 2012. *Location and Routing Models for Emergency Response Plans with Priorities.* Bonn, Germany, 7th Security Research Conference.

Patel, A. et al., 2012. A "Validation of Ground Ambulance pre-hospital Times Model using Geograpic Information Systems." *International Journal of Health Geographic,* Volume 11:42, pp. 1-10.

Skyes, A., 1992. *An Introduction to Regression Analysis,* Chicago: University of Chicago.

Solomon, M., 1987. *Algorithms for the vehicle routing and scheduling problems with time window constraints,* Operations research: vol. 35, no. 2, pp. 254-265.

Valenzuela, T. D. et al., 1997. Estimating Effectiveness of Cardiac Arrest Interventions: A Logistic Regression survival Model. *Circulation,* Volume 96, pp. 3308-3313.

Woodall, H. W., 2006. *The Use of Control Charts in Health-Care and Public Health Surveillance,* Blacksburg: Virginia Tech.

Woodall, W. H., Adams, B. M. & Benneyan, J. C., 2012. *The Use of Control Charts in Healthcare, in Statistical Methods in Healthcare.* Chichester: John Wiley & Sons, Ltd.

APPENDIX

Table A1. Data for Response Time by Month

Month	Mean	StDev	Variance	Median	Minimum	Max	Range	Count
January	5.87123	3.502602	12.26822	5	0	51	51	963
February	5.66270	3.423047	11.71725	5	0	39	39	842
March	5.55360	4.198101	17.62405	5	0	84	84	970
April	5.44823	3.155469	9.956986	5	0	21	21	937
May	5.34750	3.106850	9.652517	5	0	20	20	941
June	5.66812	4.055504	16.44711	5	0	59	59	916
July	5.41210	3.812580	14.5357	5	0	53	53	1024
August	5.40280	4.042660	16.34310	5	0	65	65	998
September	5.46220	3.556189	12.64648	5	0	35	35	1032
October	5.62450	3.661586	13.4072	5	0	33	33	1004

Table A2. Data for Response Time by ZIP Code

ZIP Code	Mean	StDev	Variance	Median	Minimu	Max	Range	Count
58078	6.497840	2.704849	7.316211	6	0	19	19	926
58102	5.214353	4.211842	17.73961	5	0	53	53	3163
58103	5.124828	3.221302	10.37679	5	0	84	84	4382
58104	7.316894	3.821788	14.60606	7	0	51	51	1095
58105	6.180327	2.341710	5.483606	6	2	13	11	61

Table A3. Data for Response Time by Incident Location

Incident Location	Mean	StDev	Variance	Median	Minimum	Max	Range	Count
Acute Care F.	5.314	4.277	18.295	5	0	53	53	2406
Bar/Restaurant	5.336	3.545	12.569	5	0	84	84	1070
Correctional F.	3.779	2.632	6.927	3	0	17	17	86
Industrial	7.273	1.954	3.818	7	5	12	7	11
Not Recorded	8.5	2.121	4.5	8.5	7	10	3	2
Other	5.651	4.114	16.924	4	0	51	51	730
Public Place	5.392	3.041	9.248	5	0	15	15	227
Recreation Area	5.686	3.879	15.045	5	0	15	15	35
Residence	5.863	3.137	9.842	5	0	39	39	3973
Traffic Way 55+	5.149	3.942	15.538	5	0	65	65	1087

Table A4. Data for Response Time by Hour of Dispatch

Hour of Dispatch	Mean	Std.	Variance	Median	Minimum	Max	Range	Count
12:00:00 a.m.	5.594	4.386	19.239	5	0	59	59	298
1:00:00 a.m.	5.432	2.981	8.885	5	0	20	20	336
2:00:00 a.m.	5.579	3.616	13.076	5	0	43	43	304
3:00:00 a.m.	6.029	3.398	11.545	6	0	22	22	241
4:00:00 a.m.	5.67	3.227	10.415	5	0	18	18	209
5:00:00 a.m.	6.266	3.144	9.887	6	0	18	18	169
6:00:00 a.m.	5.763	3.459	11.967	5	0	21	21	215
7:00:00 a.m.	5.147	2.506	6.282	5	0	13	13	258
8:00:00 a.m.	5.56	4.006	16.051	5	0	49	49	327
9:00:00 a.m.	5.673	3.642	13.268	5	0	32	32	385
10:00:00 a.m.	5.469	4.594	21.107	5	0	84	84	484
11:00:00 a.m.	5.693	3.227	10.414	5	0	39	39	469
12:00:00 p.m.	5.576	3.534	12.491	5	0	27	27	536
1:00:00 p.m.	5.351	3.442	11.846	5	0	32	32	498
2:00:00 p.m.	5.121	3.143	9.877	5	0	25	25	506
3:00:00 p.m.	5.61	3.531	12.468	5	0	26	26	515
4:00:00 p.m.	5.98	4.61	21.244	5	0	53	53	537
5:00:00 p.m.	5.534	3.68	13.54	5	0	24	24	545
6:00:00 p.m.	5.613	3.601	12.97	5	0	25	25	542
7:00:00 p.m.	5.408	4.137	17.113	5	0	65	65	529
8:00:00 p.m.	5.544	3.814	14.547	5	0	51	51	517
9:00:00 p.m.	5.666	4.133	17.079	5	0	51	51	431
10:00:00 p.m.	5.254	3.045	9.273	5	0	20	20	398
11:00:00 p.m.	5.085	2.984	8.905	5	0	15	15	378