

# Analysis of Emergency Room Waiting Time in SAS

Brent Wenerstrom, University of Louisville, Louisville, KY

## ABSTRACT

Background: Life and death may be on the line for patients visiting an emergency department (ED). The time it takes a patient to see a doctor can be critical. This is made more difficult by the fact that waiting times have been increasing over the past ten years [2]. Objective: We would like to determine what factors impact the current waiting in emergency departments through the use of Enterprise Guide®. Methods: We are using survey data from the 2006 Ambulatory Health Care Data survey [1] conducted by the National Center for Health Statistics (NCHS) containing about 36,000 data records. We used linear regression to model our data with the help of the PROC GLM function. Results: We found that self reported pain levels did not correlate with waiting time, but that ED prioritization, time of day of visit, arrival by ambulance and previous waiting times from the same emergency department correlated with waiting time.

## INTRODUCTION

A study by Nawar et al. [2] shows that from 1995-2005, emergency department visits have increase from an estimated 96.5 million in 1995 to 115.3 million. This represents a 31 percent increase in visits. In addition, the number of EDs has decreased during that same period of time from 4,176 to 3,795. Certainly this is going to have an adverse affect on the waiting time of a patient in the ED. Unfortunately, a comparison cannot be made between waiting times during this period. Waiting times in the survey used were capped through 2004, but are no longer capped.

In this study, we seek to identify those factors that most impact modern day waiting times. Specifically, we attempt to predict a patient's waiting time based on factors known at the time of their arrival.

## METHODS GIVE A REFERENCE TO THE NHAMCS

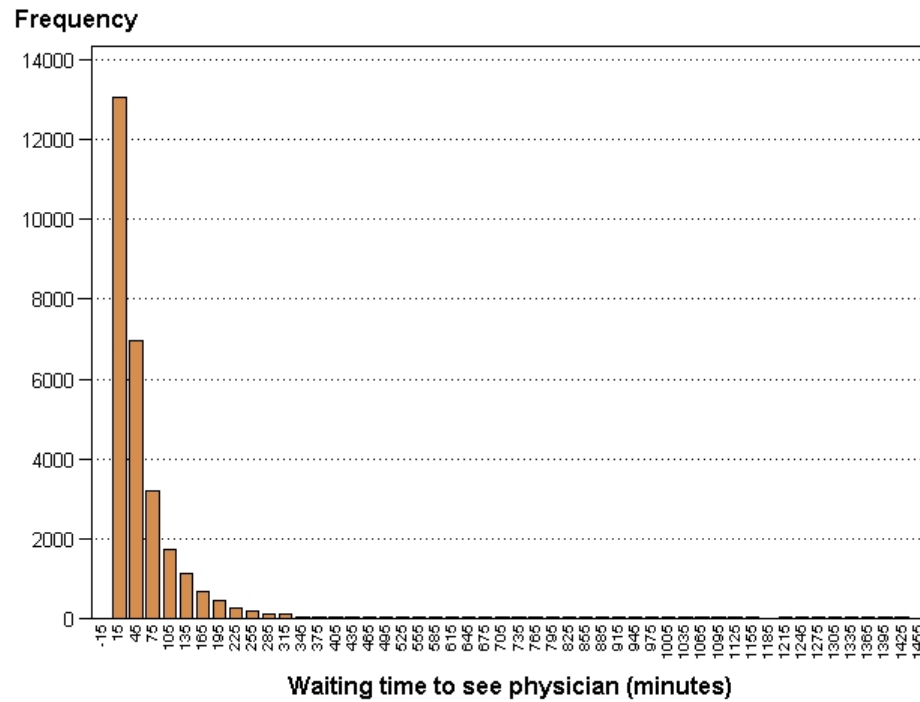
In this study, we use data collected from 360 EDs in the National Hospital Ambulatory Medical Care Survey (NHAMCS) [1] for the year 2006. There are 35,849 patients recorded in this sample. Of those, only 28,391 had known waiting times. We used SAS Enterprise Guide to process and model the data.

The original distribution of waiting times can be seen in Figure 1. From this figure, we can see that there are a large number of small waiting times and a few extreme waiting times. The longest waiting time was 1430 minutes or 23 hours and 50 minutes, while there were 770 cases with waiting time of zero minutes. Rather than model the actual waiting times, we chose instead to work with the log of the wait time. A histogram of the log of the waiting time can be seen in Figure 2, where one was added to the waiting time to avoid taking the log of zero. The shape of the data is now more normal and will allow more reliable linear regression analysis.

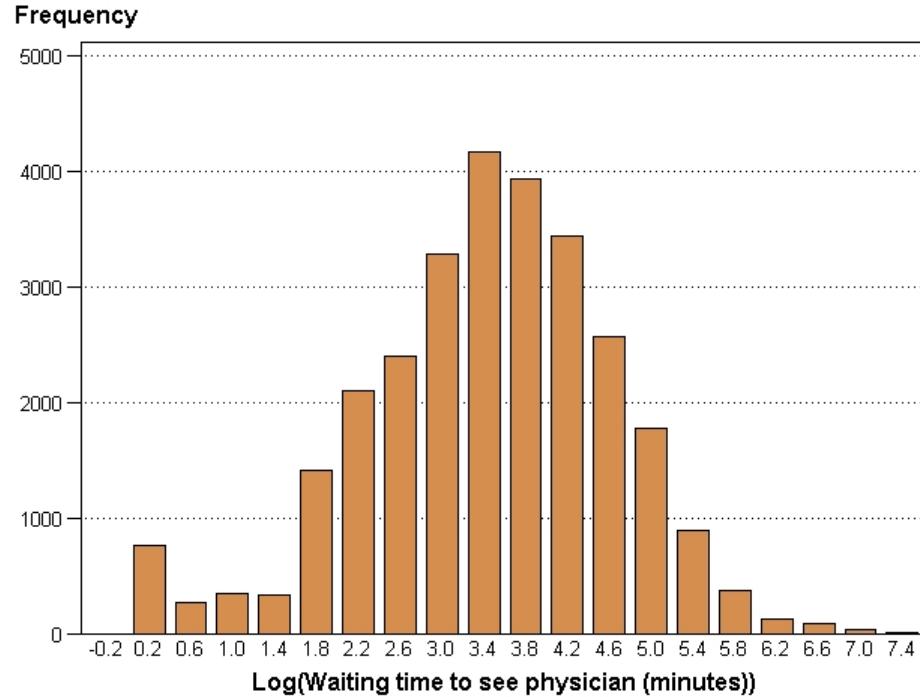
We originally computed a linear regression model on a random sample of 2,000 records, enabling the graphing of residuals and as a means to avoid inflated p-values merely based on a large sample size.

For each patient record in the survey, there was an associated hospital ID (HOSPCODE). We would like to obtain a measure of past waiting times for each hospital. We did not have enough information to order records by date, and instead sampled randomly from each hospital as an approximation to using an historical mean. We did this by sampling from each hospital using PROC SURVEYSELECT with STRATA set to HOSPCODE. We then used PROC SQL to combine the average log waiting from this sample while at the same time removing the sample from our data. The PROC SQL code can be seen in Figure 3. We computed the average log waiting time by hospital ID. We then compute the regression model on the 67% of the records for each hospital not used to obtain the average log waiting time. The regression model itself will be shown in the following section.

**Figure 1: Histogram of known waiting times.**



**Figure 2: Histogram of the log of known waiting times.**



**Figure 3: PROC SQL obtaining an average waiting time from a random sample and removing that sample from the data.**

```
PROC SQL;
CREATE TABLE SASUSER.TRAIN_WITH_AVG AS
SELECT LOG_WAITTIME, WAITTIME, IMMED, IMMEDIACY, AMBULANCE, BLACK, TIME, TIME2, HOSP_AVG
FROM SASUSER.QED2006 Q
JOIN (
    SELECT HOSPCODE, AVG(LOG_WAITTIME) AS HOSP_AVG
    FROM SASUSER.HOSP_TRAIN_SAMPLE
    GROUP BY HOSPCODE
) AS HOSP_AVG_T
ON Q.HOSPCODE=HOSP_AVG_T.HOSPCODE
WHERE NOT EXISTS (
    SELECT PATCODE
    FROM SASUSER.HOSP_TRAIN_SAMPLE AS TRAIN
    WHERE Q.PATCODE=TRAIN.PATCODE AND Q.HOSPCODE=TRAIN.HOSPCODE);
QUIT;
```

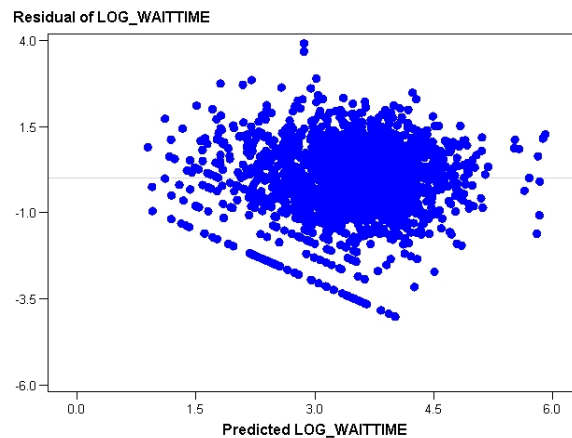
## RESULTS

Our computed regression model can be seen in Table 1 obtained through PROC GLM. All parameters can be seen to be significant. Additionally, the model achieved an  $R^2$  value of 0.307. This means that there is still a large percentage of variability that is not accounted for, but a linear correlation in the data exists.

**Table 1: Regression model for log waiting time.**

Parameter	Estimate		Standard Error	t Value	Pr >  t
Intercept	-0.585317570	B	0.05467727	-10.70	<.0001
AMBULANCE	-0.447478461		0.02084333	-21.47	<.0001
HOSP_AVG	0.873940524		0.01215321	71.91	<.0001
TIME2	0.011577446		0.00124005	9.34	<.0001
IMMEDIACY 1 - 14 min	0.579582632	B	0.03901214	14.86	<.0001
IMMEDIACY 15 - 60 min	0.967296553	B	0.03519738	27.48	<.0001
IMMEDIACY >1 hr - 2 hrs	1.139121110	B	0.03768034	30.23	<.0001
IMMEDIACY >2 hrs- 24 hrs	1.120894437	B	0.04034898	27.78	<.0001
IMMEDIACY Immediate	0.000000000	B	.	.	.

**Figure 4: Residuals of model on sample of 2,000 data points.**

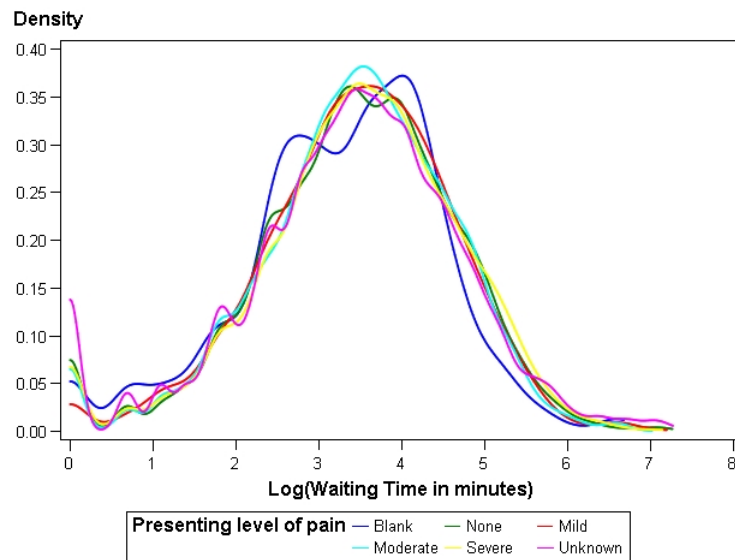


The residuals for this model could not be plotted directly in SAS. There were too many points to be kept in memory to display all residuals. Instead, we took a random sample of 2,000 points from this data set to compute a duplicate model. The residuals can be seen in Figure 4. For the most part, there is no obvious pattern in the residuals. We do see a couple of lines formed by the residuals on the lower left hand side of the plot. These come from the large number of patients with a waiting time of 0. The regression model for the sampled 2,000 points can be seen in Table 2. Again all p-values are significant and the coefficients for both the model based on 28,000 data points and the model based on 2,000 data points are very similar.

**Table 2: Regression model using a sample of 2,000 data points.**

Parameter	Estimate		Standard Error	t Value	Pr >  t
Intercept	-0.822908953	B	0.16029293	-5.13	<.0001
AMBULANCE	-0.453747277		0.06590805	-6.88	<.0001
HOSP_AVG	0.937947565		0.03509879	26.72	<.0001
TIME2	0.016791701		0.00373242	4.50	<.0001
IMMEDIACY 1 - 14 min	0.523127438	B	0.11768707	4.45	<.0001
IMMEDIACY 15 - 60 min	0.913631641	B	0.10631610	8.59	<.0001
IMMEDIACY >1 hr - 2 hrs	1.021281228	B	0.11508944	8.87	<.0001
IMMEDIACY >2 hrs- 24 hrs	1.019208467	B	0.12212009	8.35	<.0001
IMMEDIACY Immediate	0.000000000	B	.	.	.

**Figure 5: Kernel density estimators for log waiting time by pain levels.**



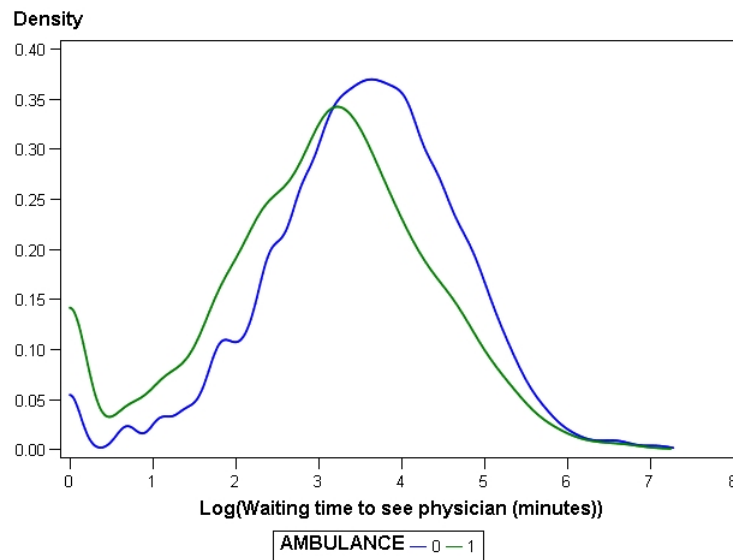
One surprising find in this model is that pain did not correlate with waiting time. This can be seen from a kernel density estimator for the log waiting time separated by various self reported pain levels seen in Figure 5. From this figure, we can see that there are no major differences in the distribution of waiting times when comparing different pain levels. The only noticeable differences are the blue line, representing a blank pain level, does not peak at the same point as the rest of the distribution and the pink line, representing an unknown pain level, is highest at a log waiting time of zero. The differences in the blank pain level can be attributed to the fact that there were very few patients with a blank pain level, only 2.5% or 716 patients, which is far fewer than any other category. This would lead to a less smooth KDE. A reason that an unknown pain level has a higher density at zero may be that patients who are unconscious when entering the emergency department are recorded as an unknown pain level. The overall distribution of an unknown pain level is very similar in every other way to the rest of the distributions of log waiting times. This means that having an unknown pain level still does not provide much information about log waiting time.

Figure 5 was produced using the PROC KDE function. This function uses kernel density estimators to approximate a the generalized probability density function for the given data. The code used to generate Figure 5 is given in Figure 6.

**Figure 6: SAS code used to produce Figure 5.**

```
PROC KDE DATA=TMPSORT_PAIN;  
UNIVAR LOG_WAITTIME / OUT=LOGWAITKDE_PAIN;  
BY PAIN;  
RUN;
```

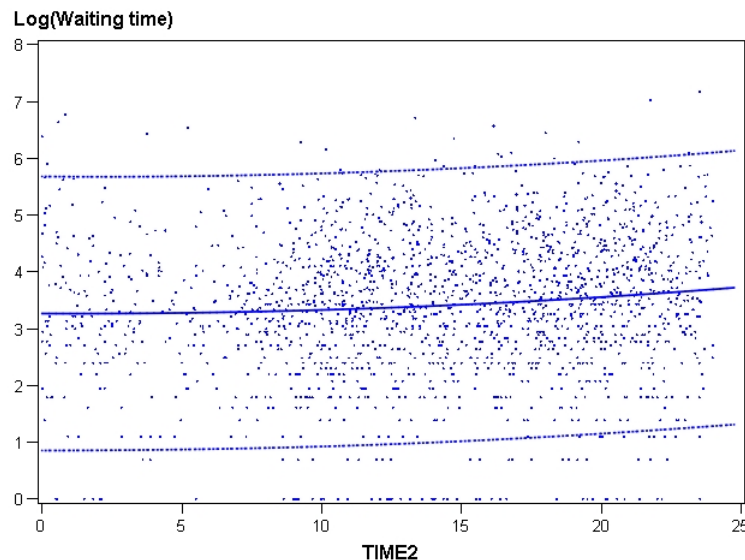
**Figure 7: Kernel Density Estimator of log wait time by ambulance arrival.**



There were a number of variables that were found to correlate with waiting time. Of these, the first is an indicator variable called AMBULANCE, which indicates that the patient was known to have arrived by ambulance to the ED. In Figure 7, we show the distributions of wait time for the ambulance variable. This graph was produced using the same steps as Figure 5 using PROC KDE. This graph shows us two very similar distributions, but those that came by ambulance are shifted to the left and have a much higher density near the value zero. This means both that patients who arrived by ambulance on average have a shorter waiting time and that a higher percentage of patients who arrive by ambulance have a waiting time of zero. Additionally, the coefficient for the ambulance variable suggests a negative correlation between waiting time and arriving by ambulance. One would expect that people who need to be taken to the ED by ambulance oftentimes have more serious needs and should be seen before those who arrived under their own power or with the help of friends or family.

The next variable we will discuss is TIME2. This variable was calculated by taking the hour of the arrival by military time and adding the minutes divided by 60 where the minutes of the current time become the decimal part of TIME2. The correlation of the time of arrival and the length of waiting time can be seen in a scatter plot of the two variables as is shown in Figure 8. This graph was created using Enterprise Guide's scatter plot functionality. On this scatter plot, we are using the "Interpolations" option to add a quadratic regression line with corresponding 95% confidence interval. From this graph, we can see that there is a slightly increasing waiting time as the day wears on. In the early morning from about 1 AM to 8 AM, there are very few people visiting the emergency department as few people are active during this time of day. With far fewer visits during the early morning hours, there are fewer people in the ED competing for the same staff. However, as the day goes on, there are an increasing number of visits to the ED, which increases the lines and waits of those patients to follow. Expect longer waits if you visit the ED later in the day.

**Figure 8: Scatter plot of time of arrival and log waiting time.**



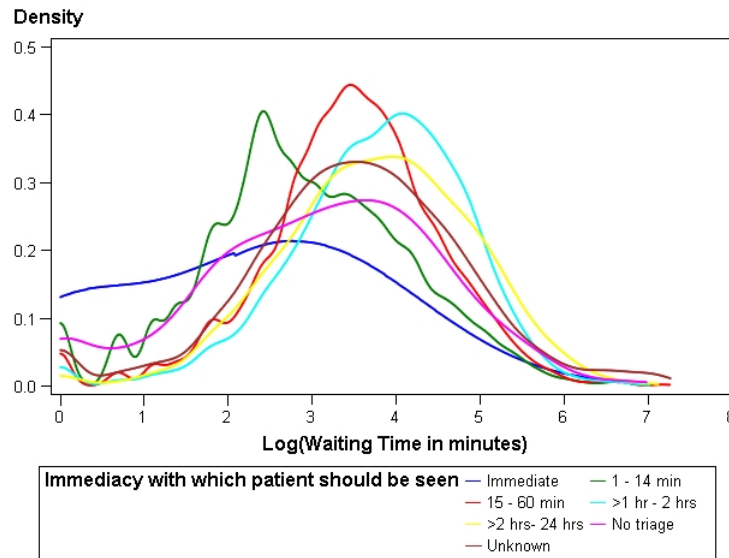
We now turn our attention to the variable, IMMEDIACY. This variable refers to the prioritization system used in many hospitals today. This form of prioritization is referred to as triage. In our data, there were 5 main groups that an incoming patient could be assigned to: “Immediate”, “1 – 14 min”, “15 – 60 min”, “>1 hr – 2 hrs”, and “>2 hrs – 24 hrs”. These categories refer to how long a patient may be able to wait to be seen. Some patients require immediate attention due to a life threatening condition while others may visit the ED to take advantage of medical attention that does not require health insurance or for service during odd hours. For example, there were 361 patients of the 35,849 in the original data sample who were not charged for the medical service provided. Additionally, there was a slight increase in patients during the hours following normal doctor office hours.

In Figure 9, we can see the distribution of waiting times according to the various priorities listed above. The figure shows that the distribution for the “Immediate” category has more density near the values 0 and 1 (0 minutes and 2 minutes) than any other category. These patients are generally seen more quickly than any other group. The next group in priority, “1 – 14 min” peaks around the value 2.5, or around 10 minutes. This peak is much sooner than the peaks of lower priority categories. The distributions generally show that higher priority results in a short waiting time, making this variable valuable for prediction.

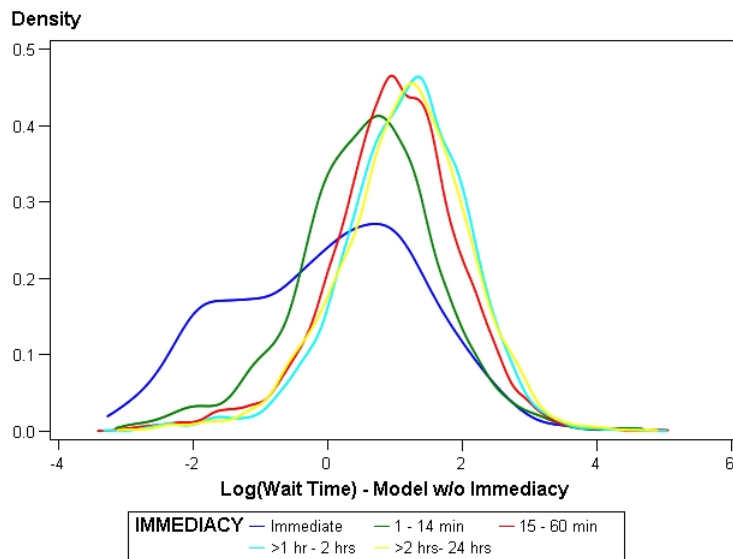
However, what is surprising is that there is large overlap among all of the groups. The “Immediate” group still includes patients who have long waits. For example, the longest known waiting time for patients put in the “Immediate” category in the data was 15 hrs. 25 min. There were 165 cases where patients who were put in the “Immediate” category were seen by a physician in an hour or more. On closer look, none of the patients who were put in the “Immediate” category and seen by a physician an hour or more later died. There were 62 cases in this data set of patients either dead on arrival (DOA) or dying in the ED. Of these 62 cases, 48 had known waiting times, and of these 48 patients who died, only 2 of them had waiting times over 16 minutes. The waiting times were 2 hrs. 35 min. and 1 hr. 13 min. It appears that very rarely is a patient with a severe threat of death made to wait for an extended period of time in the ED.

Before being used in the regression model, it was necessary to combine the categories “Unknown” and “No triage” with the most similar category in the data as an imputation technique. The group “15 – 60 min” was chosen, being the middle group with a distribution found to be most similar visually to each of these groups.

**Figure 9: Log waiting time according to priority.**



**Figure 10: Kernel Density Estimator for log wait time after subtracting out the part of the model that uses all other variables.**



**Error! Reference source not found.**In

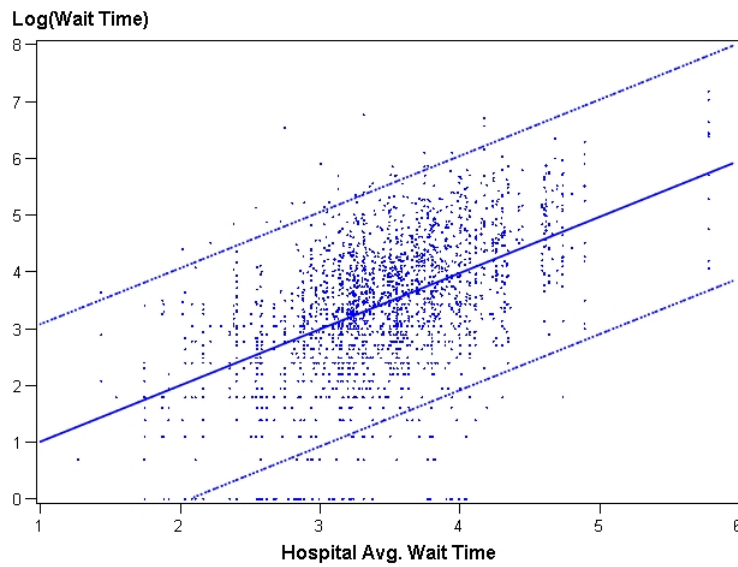
Figure 10, we see a very different kernel density estimator graph from that pictured in Figure 9. In Figure 10, we took the log wait time for each patient and subtracted out our regression's prediction except for the contribution in the model from triage. We did this using PROC SQL as shown in Figure 11. We then plotted the distribution of the resulting log wait time. The new distribution plot shows us that the model accounts for much of the variability shown in Figure 9. However, there is still a large difference in the log wait time for the "Immediate" group compared to the other groups. The averages for the other groups are all slightly different, with the exception of ">1 hr – 2 hrs" and ">2 hrs – 24 hrs" being even more similar than before. There appears to be very little difference in these two last groups. The model's coefficients even put a slightly lower log wait time on ">2 hrs – 24 hrs" group than on ">1 hr – 2 hrs". The lowest priority group has a coefficient of 1.121 versus 1.139. The difference is slight, but surprising since it rearranges the priorities in terms of actual wait time.



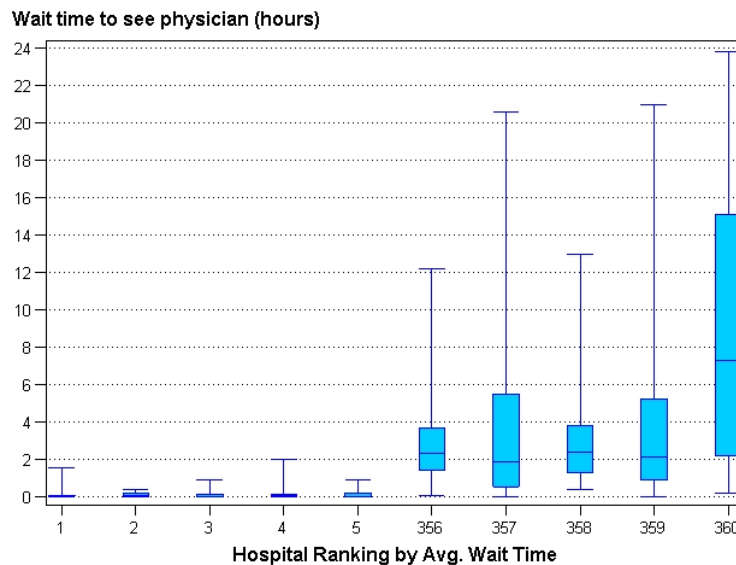
**Figure 11: PROC SQL code used to produce data behind Figure 10.**

```
PROC SQL;
CREATE TABLE SASUSER.TRAIN_PLUS_MODEL AS
SELECT LOG_WAITTIME-(-0.585317570 -0.447478461*AMBULANCE +0.873940524*HOSP_AVG
+0.011577446*TIME2) AS WAIT_FOR_IMMEDIACY
FROM SASUSER.TRAIN_WITH_AVG;
QUIT;
```

**Figure 12: Average sampled wait time per hospital versus wait time.**



**Figure 13: Box plot demonstrating variation between hospital wait times with whiskers extending to minimum and maximum values.**



The most useful of the variables for predicting waiting time is previous waiting times for a given ED. There were 360 different hospital EDs in the data. Each patient record had a hospital ID associated with it. The patients' log waiting times that are predicted were not used to calculate the sampled average for that hospital. One can see a scatter plot of log waiting times versus sampled average hospital log waiting time in Figure 12. Additionally, we have plotted a linear regression line only using hospital average waiting

time with its corresponding 95% confidence interval. It is apparent from this graph that a linear relationship exists between hospitals' previous performance and current performance, as would be expected.

Figure 12 also indicates that there is a great deal of variation between hospitals. This is more clearly shown in the box plot in Figure 13 drawn using Enterprise Guide. Here, we have the top 5 and bottom 5 hospitals based on their average waiting time. We can see that the top 5 hospitals in the survey had very little variance compared to that of the bottom 5. For example, the top hospital had 111 patients recorded in the data. The average wait time for the top hospital was 6.6 minutes with a maximum recorded waiting time of 93 minutes. Now compare this to the bottom ranked hospital, which had 87 patients recorded with known waiting times. Of these patients, the average waiting time was 9 hrs. 21 min., the minimum was 13 min. and the maximum was 23 hrs. 50 min. The shortest recorded waiting time at the bottom ranked hospital was nearly twice what the average waiting time was at the best hospital.

## DISCUSSION

In building a regression model in SAS, we learned that arriving by ambulance, time of arrival, previous waiting times in the ED and the immediacy given to a patient all are correlated with a patient's actual waiting time. The variable most strongly correlated is the previous waiting times in a given ED. Clearly not all EDs are able to have short waiting times. Those EDs with the worst waiting times need a much closer analysis to determine the root causes of such long waits (23+ hrs. in the worst cases). It may help to revisit the reasons why some 381 EDs have been closed from 1995 to 2005. Additional analysis may also provide hints as to which geographical regions are most need of additional emergency departments.

Another improvement that could be made would be to adjust when doctors are scheduled to be in EDs so that there are an increasing number of doctors throughout the day. This study does not provide clues on how to decrease the number of visits to EDs across the country, but we can see that waiting time increases gradually through each day. With improved scheduling, Figure 8 could potentially be flattened out to remove the correlation between time of day and waiting time.

This survey suggests that current prioritization methods are effective for those patients closest to death. Patients who did die before arriving at an ED, in an ED or in the hospital after visiting an ED rarely had to wait very long. In terms of treating patients most at risk of dying quickly, the current emergency department protocols are meeting needs. However, given that there are individuals waiting nearly 24 hours, there are major improvements that still need to be made in emergency care.

## REFERENCES

- [1] Centers for Disease Control and Prevention. "Ambulatory Health Care Data," 4/20/2009; <http://www.cdc.gov/nchs/about/major/ahcd/ahcd1.htm>.
- [2] E. W. Nawar, R. W. Niska, and J. Xu, "National Hospital Ambulatory Medical Care Survey: 2005 emergency department summary," *Adv Data*, no. 386, pp. 1-32, Jun 29, 2007.

## CONTACT INFORMATION

Contact the author at:

Brent Wenerstrom  
University of Louisville  
J.B. Speed Building – Room LL14  
Louisville, KY 40292  
bkwene01@louisville.edu