David Weller

Abstract

A pilot study to determine the feasibility of predicting how many examinations will be performed over the next 12 months.

Fargo Health Group  
Cardiovascular Examinations

Table of Contents

[Overview 2](#_Toc521083817)

[Data 2](#_Toc521083818)

[Data Cleansing 3](#_Toc521083819)

[Outliers 4](#_Toc521083820)

[Missing Data 5](#_Toc521083821)

[Forecasts 7](#_Toc521083822)

[Holt Winters Forecast 7](#_Toc521083823)

[ARIMA Forecast 8](#_Toc521083824)

[Choice of Models 8](#_Toc521083826)

[Appendix A: Cleaned and Imputed Data 9](#_Toc521083827)

Fargo Health Group  
Cardiovascular Examinations

# Overview

The Fargo Health Group (FHG) has decided to do a pilot-study to determine if a data solution can accurately predict the incoming volume of medical requests in order to improve the scheduling of physicians. FHG has provided a dataset on the historical monthly examinations volume of cardiovascular examinations from the Health Center located in Abbeville, LA.

In order to effectively schedule physicians, FHG needs to be able to predict how many examinations to expect at the different Health Centers. This pilot program attempts to predict a small subset of these examinations, specifically the cardiovascular exams at just one of their Health Centers.

The dataset that FHG provided will need to be examined, the data will need to be cleaned, and a forecasting method will need to be chosen.

# Data

FHG provided a dataset in the form of an Excel spreadsheet. This spreadsheet consisted of eight worksheets that contained the data.

The first sheet has three columns: Incoming Examinations, Year, and Month. It lists the number of cardiovascular examinations for each month from January 2006 through December 2013 that originated at the Abbeville, LA Health Center.

The next four worksheets contain examinations that were sent from one Health Center to another. Each of four worksheets contains information about examinations that were sent to another Health Center in May 2007. There is one worksheet for each of four Health Centers: Violet, New Orleans, Lafayette, and Baton Rouge. These worksheets include all examinations sent from all Health Centers in May 2007.

One worksheet contained information about all examinations done by FHG in December 2013. It just contains the Routing SYSID. The SYSID is a 17 character string that contains the location code (first four characters), Condition Code (next six characters), among other information.

There is a worksheet that contains the heart-related Condition Codes and another that contains all of the Condition Codes.

# Data Cleansing

The Abbeville worksheet was copied in order to transform the data without losing the original values. The Cleaned Data worksheet was sorted by Year and Month. Figure 1 shows a graph of the Incoming Examinations before any cleansing has been performed.

Figure 1

The four worksheets containing data for the examinations that were sent to other Health Centers had several dates that were not formatted properly. These were manually changed in order to get an accurate count of examinations in each month. Columns for Month and Year were then added to these worksheets. There were numerous duplicate Request IDs, which were removed.

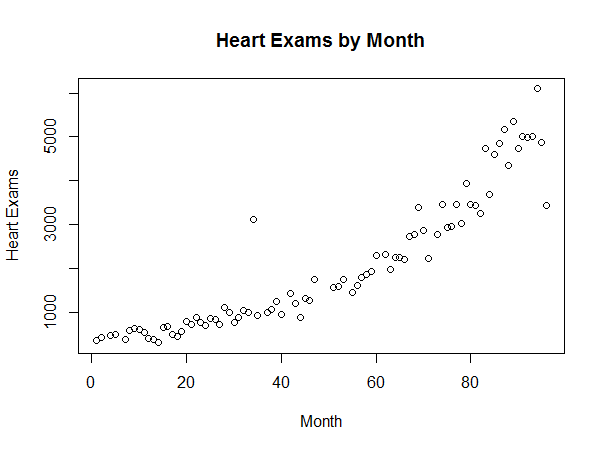
Pivot tables were created based each of the worksheets in order to filter the data. Each pivot table was filtered to show the cardiovascular examinations sent from Abbeville. They were grouped by Year and Month. The examinations were then added to the corresponding months in the Cleaned Data worksheet.

Added columns in the December 2013 Data worksheet for Location Code and Condition Code. These columns were filled in from the SYSID. A third column was added to represent whether or not the examination was sent from Abbeville and was a cardiovascular exam. This count was added to the Cleaned Data worksheet.

A quick examination of the data showed that there were 10 values that were obviously inaccurate. Two of them were over 99 million and eight were not numbers. These 10 values were reformatted in the Cleaned Data worksheet. The result was saved in a csv file and examined in R.

# Outliers

There were two values that did not appear to conform to the general trends of the data. October 2008 had 3110 exams listed, well above the number of exams around that time. December 2013 had 3442 exams listed, well below previous marks. Figure 2 shows the scatterplot of Heart Exams over time.

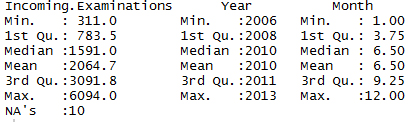
  
Figure2

The value for October 2008 does indeed appear to be a mistake, perhaps a simple typo. The value for December 2013, however, upon closer inspection appears to be part of a downward trend after a high of 6094 in October 2013.

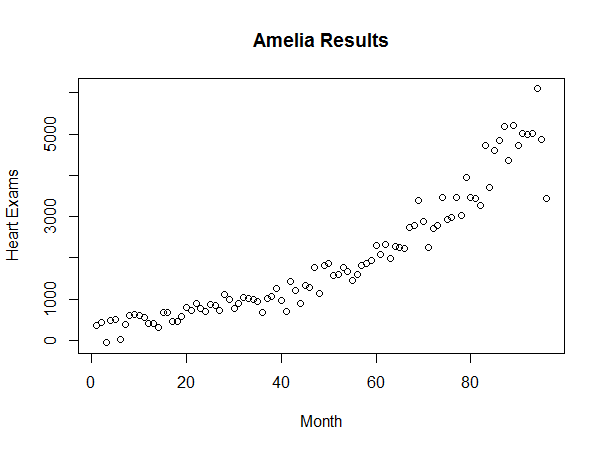
The value for October 2008 value was replaced using a 3-period weighted moving average value. The weights used were 0.5, 0.3, and 0.2. That value turned out to be 995.

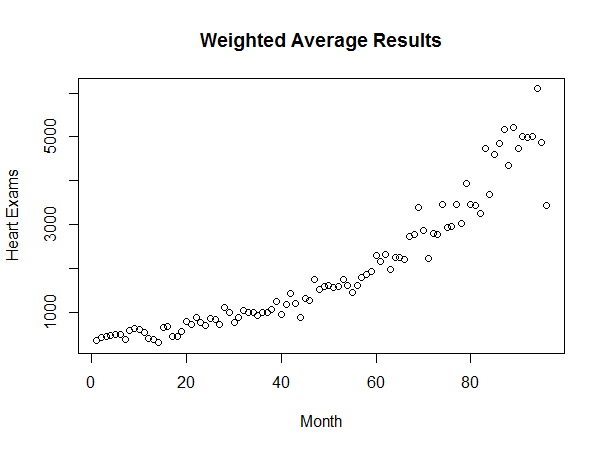
# Missing Data

The summary of the Cleaned Data is shown in Figure 3. There are 10 missing values that need to be imputed.

  
Figure 3

Two methods of imputing the missing values were tried: Amelia and Weighted Average. Figures 4 and 5 show the results of these two methods.

  
Figure 4

  
Figure 5

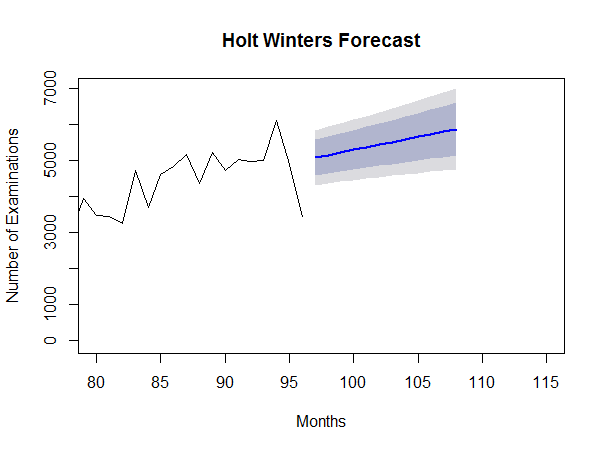
The Weighted Averages result was chosen because the Amelia results produced negative results early on. The Cleaned and Imputed Data can be found in Appendix A.

# Forecasts

Two forecasts were attempted on the dataset: Holt Winters and ARIMA. Both forecasts were analyzed and plotted.

## Holt Winters Forecast

The Holt Winters forecast was run without specifying the alpha value, allowing R to come up with the best fit. The predicted values for the next 12 months can be seen in Figures 6 and 7.

  
Figure 6

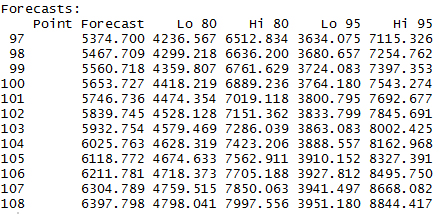
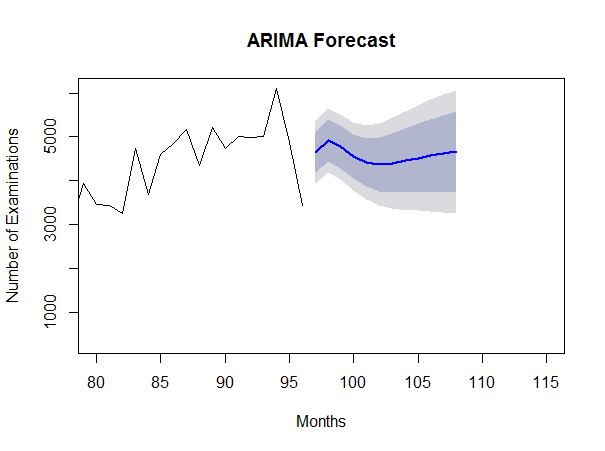
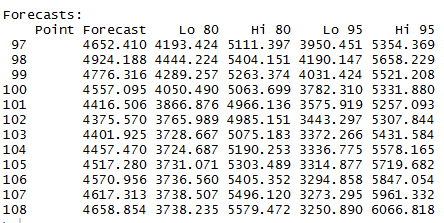


Figure 7

## ARIMA Forecast

The ARIMA forecast was run next. Once again, R was used to automatically fit the model. The predicted values for the next 12 months can be seen in Figures 8 and 9.

  
Figure 8



# Figure 9

# Choice of Models

The two models were compared on the basis of MAD. The Holt Winters forecast yielded a MAD of 0.13 while the Arima forecast yielded a MAD of 235.23. The Holt Winters forecast was chosen based on having a much better MAD.

# Appendix A: Cleaned and Imputed Data

|  |  |  |
| --- | --- | --- |
| Year | Month | Weighted Average |
| 2006 | 1 | 362 |
| 2006 | 2 | 436 |
| 2006 | 3 | 451.24 |
| 2006 | 4 | 490 |
| 2006 | 5 | 508 |
| 2006 | 6 | 495.3214286 |
| 2006 | 7 | 393 |
| 2006 | 8 | 596 |
| 2006 | 9 | 634 |
| 2006 | 10 | 613 |
| 2006 | 11 | 545 |
| 2006 | 12 | 411 |
| 2007 | 1 | 398 |
| 2007 | 2 | 311 |
| 2007 | 3 | 664 |
| 2007 | 4 | 680 |
| 2007 | 5 | 456 |
| 2007 | 6 | 467 |
| 2007 | 7 | 566 |
| 2007 | 8 | 806 |
| 2007 | 9 | 732 |
| 2007 | 10 | 886 |
| 2007 | 11 | 776 |
| 2007 | 12 | 698 |
| 2008 | 1 | 875 |
| 2008 | 2 | 840 |
| 2008 | 3 | 724 |
| 2008 | 4 | 1115 |
| 2008 | 5 | 997 |
| 2008 | 6 | 775 |
| 2008 | 7 | 886 |
| 2008 | 8 | 1041 |
| 2008 | 9 | 1011 |
| 2008 | 10 | 995 |
| 2008 | 11 | 939 |
| 2008 | 12 | 1011.166667 |
| 2009 | 1 | 1004 |
| 2009 | 2 | 1065 |
| 2009 | 3 | 1263 |
| 2009 | 4 | 962 |
| 2009 | 5 | 1174.466667 |
| 2009 | 6 | 1429 |
| 2009 | 7 | 1205 |
| 2009 | 8 | 890 |
| 2009 | 9 | 1320 |
| 2009 | 10 | 1276 |
| 2009 | 11 | 1757 |
| 2009 | 12 | 1525 |
| 2010 | 1 | 1584.142857 |
| 2010 | 2 | 1608.588235 |
| 2010 | 3 | 1578 |
| 2010 | 4 | 1604 |
| 2010 | 5 | 1758 |
| 2010 | 6 | 1627.655172 |
| 2010 | 7 | 1457 |
| 2010 | 8 | 1607 |
| 2010 | 9 | 1808 |
| 2010 | 10 | 1866 |
| 2010 | 11 | 1934 |
| 2010 | 12 | 2294 |
| 2011 | 1 | 2165.833333 |
| 2011 | 2 | 2334 |
| 2011 | 3 | 1973 |
| 2011 | 4 | 2262 |
| 2011 | 5 | 2259 |
| 2011 | 6 | 2217 |
| 2011 | 7 | 2739 |
| 2011 | 8 | 2772 |
| 2011 | 9 | 3383 |
| 2011 | 10 | 2869 |
| 2011 | 11 | 2239 |
| 2011 | 12 | 2796.866667 |
| 2012 | 1 | 2789 |
| 2012 | 2 | 3455 |
| 2012 | 3 | 2940 |
| 2012 | 4 | 2968 |
| 2012 | 5 | 3466 |
| 2012 | 6 | 3037 |
| 2012 | 7 | 3946 |
| 2012 | 8 | 3459 |
| 2012 | 9 | 3446 |
| 2012 | 10 | 3258 |
| 2012 | 11 | 4729 |
| 2012 | 12 | 3694 |
| 2013 | 1 | 4610 |
| 2013 | 2 | 4841 |
| 2013 | 3 | 5172 |
| 2013 | 4 | 4351 |
| 2013 | 5 | 5213 |
| 2013 | 6 | 4729 |
| 2013 | 7 | 5014 |
| 2013 | 8 | 4978 |
| 2013 | 9 | 5008 |
| 2013 | 10 | 6094 |
| 2013 | 11 | 4874 |
| 2013 | 12 | 3442 |