Medical Exam Model for Fargo Health

Data Prep and Forecasting

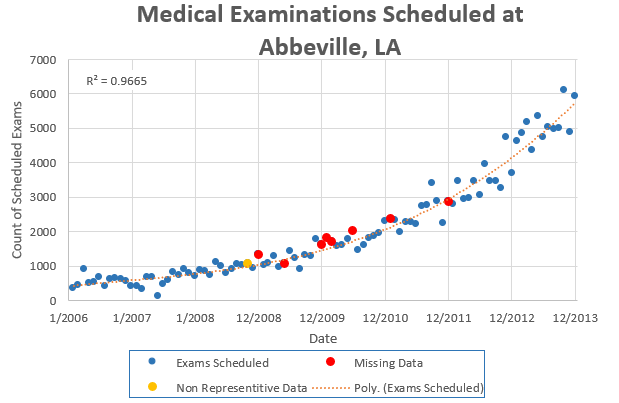
By Pete Kelley

1. In-scope Data, The study is limited to the following:
   1. Fargo Health Group’s Abbeville, LA Health Center
   2. Medical examinations related to heart conditions
   3. Dates between Jan 2006 and Dec 2013
2. Removed Non-Representative Data
   1. Removed Oct 2008 Incoming exams because of the one-time event of Hurricane Katrina and used Amelia to fill in the missing value.
3. Cleaning Data
   1. May-2007 Violet, LA Worksheet:
      1. Date format fixed by removing a comma between the month and year.
   2. May-2007 New Orleans,LA Worksheet:
      1. None
   3. May-2007 Lafayette, LA Worksheet:
      1. Date format fixed by removing a comma between the month and year.
   4. May-2007 Baton Rouge, LA Worksheet:
      1. Date format fixed by removing a comma between the month and year.
4. Aggregating Data from other Tabs of the Workbook
   1. Examination codes:
      1. Lack of uniformity between Health Centers required building a lookup table that list all examination descriptions across all of the HC that are heart-related
   2. December 2013 Data Worksheet:
      1. The data had to be filtered as follows:
         1. SYSID starts with “L839” and ends either in “TGU3” or “ROV8”
         2. SYSID included one of the condition codes listed on Worksheet Heart-related Condition Codes
         3. Include rows that meet both conditions
   3. May-2007 Violet, LA Worksheet / May-2007 New Orleans,LA Worksheet / May-2007 Lafayette, LA Worksheet / May-2007 Baton Rouge, LA Worksheet:
      1. Aggregate values
         1. Count rows that contained the following for May 2007:
            1. Original Hospital Location = “Abbeville”
            2. Examination was heart-related
            3. Date = May 2, 2007
         2. Count rows that contained the following for May 2013:
            1. Original Hospital Location = “Abbeville”
            2. Examination was heart-related
            3. Date = May 2013
         3. Count rows that contained the following for June 2013:
            1. Original Hospital Location = “Abbeville”
            2. Examination was heart-related
            3. Date = June 2013
         4. Count rows that contained the following for July 2013:
            1. Original Hospital Location = “Abbeville”
            2. Examination was heart-related
            3. Date = July 2013
5. Prep data for Imputing Missing Value:
   1. Save CSV that only had Incoming Examinations and running row number for importing into R.
6. Using Amelia for Imputing Missing Values
   1. Amelia gave me 5 models with missing values being placed
   2. Assumptions when picking a best fit model to use for imputation of data
      1. The data has seasonality that should be preserved
      2. The new values should fall close to a trend line of the data after factoring in the replacement values
         1. I calculated the average change between subsequent months for the raw data and each of the 5 models. I picked the one that minimized the sum of the square distances between the model and the raw data.
            1. For example: One set of imputed values between Dec and Jan

(Average Jan without imputing missing values)/ (Average Dec without imputing missing values) = 1.091

(Average Jan with imputing missing values)/ (Average Dec with imputing missing values) = 1.072

Square of distance = (1.091 – 1.072)^2 = 0.00035

* + - 1. I graphed the model that minimized the sum of the square distances determine if the imputed values fell close to the trend line.
      2. Imputed Values highlighted in gold and pink:
         1. Gold highlight indicates the month that had to be replaced because of hurricane Katrina.
         2. Pink indicates months that had no data in original data

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Year** | **Month** | **Exams\_imputed** |  | **Year** | **Month** | **Exams\_imputed** |  | **Year** | **Month** | **Exams\_imputed** |
| 2006 | 1 | 362 |  | 2007 | 1 | 398 |  | 2008 | 1 | 875 |
| 2006 | 2 | 436 |  | 2007 | 2 | 311 |  | 2008 | 2 | 840 |
| 2006 | 3 | 900 |  | 2007 | 3 | 664 |  | 2008 | 3 | 724 |
| 2006 | 4 | 490 |  | 2007 | 4 | 680 |  | 2008 | 4 | 1115 |
| 2006 | 5 | 508 |  | 2007 | 5 | 114 |  | 2008 | 5 | 997 |
| 2006 | 6 | 662 |  | 2007 | 6 | 467 |  | 2008 | 6 | 775 |
| 2006 | 7 | 393 |  | 2007 | 7 | 566 |  | 2008 | 7 | 886 |
| 2006 | 8 | 596 |  | 2007 | 8 | 806 |  | 2008 | 8 | 1041 |
| 2006 | 9 | 634 |  | 2007 | 9 | 732 |  | 2008 | 9 | 1011 |
| 2006 | 10 | 613 |  | 2007 | 10 | 886 |  | 2008 | 10 | 1064 |
| 2006 | 11 | 545 |  | 2007 | 11 | 776 |  | 2008 | 11 | 939 |
| 2006 | 12 | 411 |  | 2007 | 12 | 698 |  | 2008 | 12 | 1335 |
|  |  |  |  |  |  |  |  |  |  |  |
| **Year** | **Month** | **Exams\_imputed** |  | **Year** | **Month** | **Exams\_imputed** |  | **Year** | **Month** | **Exams\_imputed** |
| 2009 | 1 | 1004 |  | 2010 | 1 | 1812 |  | 2011 | 1 | 2357 |
| 2009 | 2 | 1065 |  | 2010 | 2 | 1702 |  | 2011 | 2 | 2334 |
| 2009 | 3 | 1263 |  | 2010 | 3 | 1578 |  | 2011 | 3 | 1973 |
| 2009 | 4 | 962 |  | 2010 | 4 | 1604 |  | 2011 | 4 | 2262 |
| 2009 | 5 | 1061 |  | 2010 | 5 | 1758 |  | 2011 | 5 | 2259 |
| 2009 | 6 | 1429 |  | 2010 | 6 | 2011 |  | 2011 | 6 | 2217 |
| 2009 | 7 | 1205 |  | 2010 | 7 | 1457 |  | 2011 | 7 | 2739 |
| 2009 | 8 | 890 |  | 2010 | 8 | 1607 |  | 2011 | 8 | 2772 |
| 2009 | 9 | 1320 |  | 2010 | 9 | 1808 |  | 2011 | 9 | 3383 |
| 2009 | 10 | 1276 |  | 2010 | 10 | 1866 |  | 2011 | 10 | 2869 |
| 2009 | 11 | 1757 |  | 2010 | 11 | 1934 |  | 2011 | 11 | 2239 |
| 2009 | 12 | 1615 |  | 2010 | 12 | 2294 |  | 2011 | 12 | 2868 |
|  |  |  |  |  |  |  |  |  |  |  |
| **Year** | **Month** | **Exams\_imputed** |  | **Year** | **Month** | **Exams\_imputed** |  |  |  |  |
| 2012 | 1 | 2789 |  | 2013 | 1 | 4610 |  |  |  |  |
| 2012 | 2 | 3455 |  | 2013 | 2 | 4841 |  |  |  |  |
| 2012 | 3 | 2940 |  | 2013 | 3 | 5172 |  |  |  |  |
| 2012 | 4 | 2968 |  | 2013 | 4 | 4351 |  |  |  |  |
| 2012 | 5 | 3466 |  | 2013 | 5 | 5329 |  |  |  |  |
| 2012 | 6 | 3037 |  | 2013 | 6 | 4736 |  |  |  |  |
| 2012 | 7 | 3946 |  | 2013 | 7 | 5015 |  |  |  |  |
| 2012 | 8 | 3459 |  | 2013 | 8 | 4978 |  |  |  |  |
| 2012 | 9 | 3446 |  | 2013 | 9 | 5008 |  |  |  |  |
| 2012 | 10 | 3258 |  | 2013 | 10 | 6094 |  |  |  |  |
| 2012 | 11 | 4729 |  | 2013 | 11 | 4874 |  |  |  |  |

1. Building Predictive Model with Holt-Winters
   1. Forecast using Holt-Winters Exponential Smoothing and let R calculate best values for alpha and beta
   2. exams\_imputed.mean <- HoltWinters(exams\_imputed, gamma=FALSE)

Smoothing parameters:

alpha: 0.2287065

beta : 0.1296049

gamma: FALSE

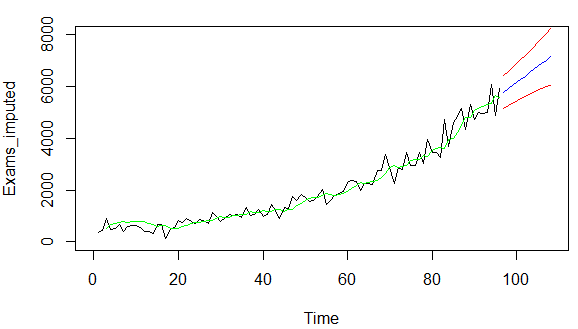
Coefficients:

[,1]

a 5670.3386

b 123.3381

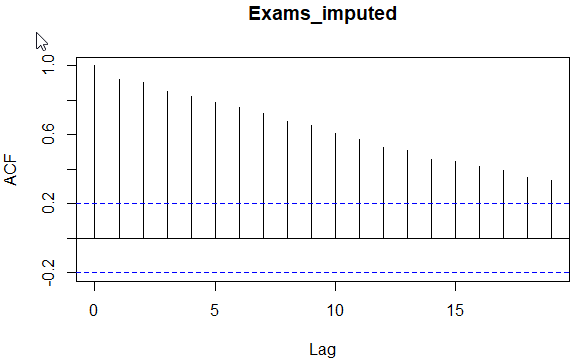
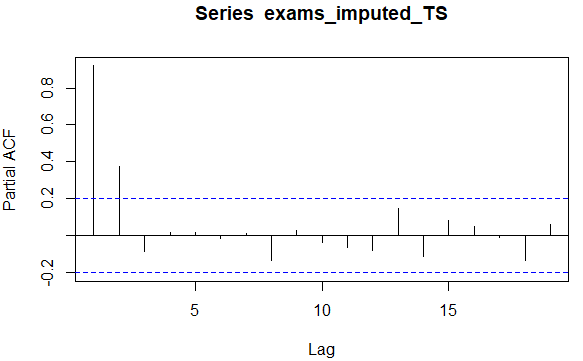
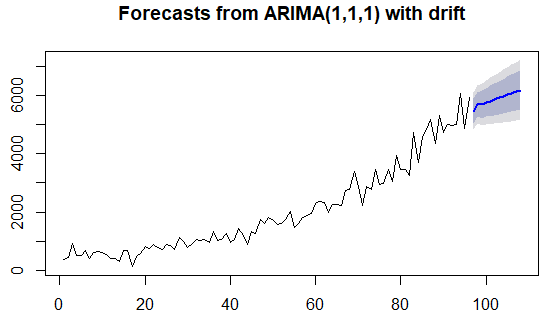
* 1. Seasonality:
     1. Having a beta value indicates there is seasonality in the data.
  2. Plot the model:
     1. plot.ts(exams\_imputed, xlim=c(0, 108), ylim=c(0,8000) )
     2. lines(exams\_imputed.mean$fitted[,1], col="green")
     3. lines(exams\_imputed.predict[,1], col="blue")
     4. lines(exams\_imputed.predict[,2], col="red")
     5. lines(exams\_imputed.predict[,3], col="red")



* 1. # Error measures
     1. accuracy(test,exams\_imputed[,1][3:96])

ME RMSE MAE MPE MAPE

Test set 17.70739 326.1431 249.068 -9.263005 21.83722

1. Building Predictive Model with ARIMA
   1. Assess the time series using (autocorrelation function)ACF and PACF (partial autocorrelation function)
      1. acf(exams\_imputed\_TS)
      2. pacf(exams\_imputed\_TS)
      3. Both graphs have large spikes to the left that die out, this indicates there are both autoregressive and have moving averages in the data.
      4. An ARIMA model (Combonation of Autoregression and moving averages) is a good option.
      5. myBestForecast <- auto.arima(x = exams\_imputed\_TS)
      6. plot the model with forceasts
      7. Error measures:

ME RMSE MAE MPE MAPE MASE ACF1

Training set 1.078 321.187 232.143 -11.632 21.224 0.766 -0.003

1. Compare Models
   1. Holt-Winters Model

|  |  |  |
| --- | --- | --- |
| Mean Absolute Deviation or Error, Avg(ABS(Error)) | | |
| MAD | 249 |  |
| Mean Square Error, Avg(Error ^2) | | |
| MSE | 106,365 |  |
| Mean Absolute Percent Error, Avg(ABS(Error)/Actual)) | | |
| MAPA | 21.84% |  |

* 1. ARIMA Model

|  |  |  |
| --- | --- | --- |
| Mean Absolute Deviation or Error, Avg(ABS(Error)) | | |
| MAD | 235 |  |
| Mean Square Error, Avg(Error ^2) | | |
| MSE | 104,243 |  |
| Mean Absolute Percent Error, Avg(ABS(Error)/Actual)) | | |
| MAPA | 21.45% |  |

* 1. The ARIMA model minimizes the MAD, MSE, and MAPA and is therefor the better model.

1. Predicted heart-related exams over the next 12 months.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Year | Month | Lo.95 | Lo.80 | Point.Forecast | Hi.80 | Hi.95 |
| 2014 | 1 | 4832 | 5054 | 5475 | 5895 | 6118 |
| 2014 | 2 | 5049 | 5275 | 5702 | 6130 | 6356 |
| 2014 | 3 | 4980 | 5229 | 5697 | 6166 | 6414 |
| 2014 | 4 | 5017 | 5278 | 5771 | 6264 | 6526 |
| 2014 | 5 | 5022 | 5298 | 5818 | 6339 | 6615 |
| 2014 | 6 | 5041 | 5329 | 5875 | 6420 | 6709 |
| 2014 | 7 | 5057 | 5359 | 5928 | 6497 | 6798 |
| 2014 | 8 | 5076 | 5390 | 5982 | 6574 | 6888 |
| 2014 | 9 | 5096 | 5422 | 6036 | 6650 | 6975 |
| 2014 | 10 | 5118 | 5454 | 6090 | 6725 | 7062 |
| 2014 | 11 | 5140 | 5488 | 6144 | 6800 | 7147 |

1. Model results vs imputed exams

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Year** | **Month** | **Exams\_imputed** | **Modeled Exams** |  |  |  |
| 2006 | 1 | 362 | 362 |  |  |  |
| 2006 | 2 | 436 | 421 |  |  |  |
| 2006 | 3 | 900 | 497 |  |  |  |
| 2006 | 4 | 490 | 629 |  |  |  |
| 2006 | 5 | 508 | 767 |  |  |  |
| 2006 | 6 | 662 | 698 |  |  |  |
| 2006 | 7 | 393 | 699 |  |  |  |
| 2006 | 8 | 596 | 703 |  |  |  |
| 2006 | 9 | 634 | 651 |  |  |  |
| 2006 | 10 | 613 | 701 |  |  |  |
| 2006 | 11 | 545 | 735 |  |  |  |
| 2006 | 12 | 411 | 731 |  |  |  |
| 2007 | 1 | 398 | 683 |  |  |  |
| 2007 | 2 | 311 | 611 |  |  |  |
| 2007 | 3 | 664 | 557 |  |  |  |
| 2007 | 4 | 680 | 565 |  |  |  |
| 2007 | 5 | 114 | 692 |  |  |  |
| 2007 | 6 | 467 | 656 |  |  |  |
| 2007 | 7 | 566 | 510 |  |  |  |
| 2007 | 8 | 806 | 578 |  |  |  |
| 2007 | 9 | 732 | 687 |  |  |  |
| 2007 | 10 | 886 | 808 |  |  |  |
| 2007 | 11 | 776 | 868 |  |  |  |
| 2007 | 12 | 698 | 930 |  |  |  |
| 2008 | 1 | 875 | 908 |  |  |  |
| 2008 | 2 | 840 | 903 |  |  |  |
| 2008 | 3 | 724 | 954 |  |  |  |
| 2008 | 4 | 1115 | 946 |  |  |  |
| 2008 | 5 | 997 | 974 |  |  |  |
| 2008 | 6 | 775 | 1098 |  |  |  |
| 2008 | 7 | 886 | 1078 |  |  |  |
| 2008 | 8 | 1041 | 1013 |  |  |  |
| 2008 | 9 | 1011 | 1047 |  |  |  |
| 2008 | 10 | 1064 | 1111 |  |  |  |
| 2008 | 11 | 939 | 1141 |  |  |  |
| 2008 | 12 | 1335 | 1150 |  |  |  |
| 2009 | 1 | 1004 | 1184 |  |  |  |
| 2009 | 2 | 1065 | 1275 |  |  |  |
| 2009 | 3 | 1263 | 1217 |  |  |  |
| 2009 | 4 | 962 | 1246 |  |  |  |
| 2009 | 5 | 1061 | 1273 |  |  |  |
| 2009 | 6 | 1429 | 1201 |  |  |  |
| 2009 | 7 | 1205 | 1267 |  |  |  |
| 2009 | 8 | 890 | 1383 |  |  |  |
| 2009 | 9 | 1320 | 1306 |  |  |  |
| 2009 | 10 | 1276 | 1239 |  |  |  |
| 2009 | 11 | 1757 | 1346 |  |  |  |
| 2009 | 12 | 1615 | 1469 |  |  |  |
| 2010 | 1 | 1812 | 1665 |  |  |  |
| 2010 | 2 | 1702 | 1747 |  |  |  |
| 2010 | 3 | 1578 | 1833 |  |  |  |
| 2010 | 4 | 1604 | 1815 |  |  |  |
| 2010 | 5 | 1758 | 1769 |  |  |  |
| 2010 | 6 | 2011 | 1783 |  |  |  |
| 2010 | 7 | 1457 | 1888 |  |  |  |
| 2010 | 8 | 1607 | 1924 |  |  |  |
| 2010 | 9 | 1808 | 1781 |  |  |  |
| 2010 | 10 | 1866 | 1799 |  |  |  |
| 2010 | 11 | 1934 | 1886 |  |  |  |
| 2010 | 12 | 2294 | 1960 |  |  |  |
| 2011 | 1 | 2357 | 2084 |  |  |  |
| 2011 | 2 | 2334 | 2277 |  |  |  |
| 2011 | 3 | 1973 | 2387 |  |  |  |
| 2011 | 4 | 2262 | 2366 |  |  |  |
| 2011 | 5 | 2259 | 2286 |  |  |  |
| 2011 | 6 | 2217 | 2345 |  |  |  |
| 2011 | 7 | 2739 | 2365 |  |  |  |
| 2011 | 8 | 2772 | 2455 |  |  |  |
| 2011 | 9 | 3383 | 2681 |  |  |  |
| 2011 | 10 | 2869 | 2911 |  |  |  |
| 2011 | 11 | 2239 | 3136 |  |  |  |
| 2011 | 12 | 2868 | 2955 |  |  |  |
| 2012 | 1 | 2789 | 2769 |  |  |  |
| 2012 | 2 | 3455 | 2878 |  |  |  |
| 2012 | 3 | 2940 | 3024 |  |  |  |
| 2012 | 4 | 2968 | 3228 |  |  |  |
| 2012 | 5 | 3466 | 3155 |  |  |  |
| 2012 | 6 | 3037 | 3220 |  |  |  |
| 2012 | 7 | 3946 | 3343 |  |  |  |
| 2012 | 8 | 3459 | 3420 |  |  |  |
| 2012 | 9 | 3446 | 3678 |  |  |  |
| 2012 | 10 | 3258 | 3634 |  |  |  |
| 2012 | 11 | 4729 | 3574 |  |  |  |
| 2012 | 12 | 3694 | 3748 |  |  |  |
| 2013 | 1 | 4610 | 4143 |  |  |  |
| 2013 | 2 | 4841 | 4147 |  |  |  |
| 2013 | 3 | 5172 | 4502 |  |  |  |
| 2013 | 4 | 4351 | 4810 |  |  |  |
| 2013 | 5 | 5329 | 4922 |  |  |  |
| 2013 | 6 | 4736 | 4874 |  |  |  |
| 2013 | 7 | 5015 | 5076 |  |  |  |
| 2013 | 8 | 4978 | 5022 |  |  |  |
| 2013 | 9 | 5008 | 5084 |  |  |  |
| 2013 | 10 | 6094 | 5106 |  |  |  |
| 2013 | 11 | 4874 | 5324 |  |  |  |
| 2013 | 12 | 5933 | 5576 |  |  |  |