4785-2020

Forecasting Healthcare Statistics at the Cleveland Clinic

Michael Bromley, Michael Lewis, Colleen Houlahan, Eric Hixson

Cleveland Clinic

ABSTRACT

A thorough understanding of key volumes is of utmost importance to effectively manage a hospital system. As the Cleveland Clinic's analytical culture matures, we see a shift from reactive decision making to proactive decision making. This shift in emphasis calls for the increased application of statistical forecast models inserted into stakeholder work flows and at the fingertips of leaders and key decision makers. However, model development and deployment are not enough. Active end-user engagement and buy in are critical in order for forecasting methods to be accepted and used to make decisions. This requires consistent interaction between our data scientists and end users beyond just the modeling and validation process. This presentation covers Cleveland Clinic's complete journey from model inception through end-user acceptance and adoption for weekly business management. We discuss the forecast model building; application of Monte Carlo simulations and programming; output visualization; and the challenges achieving end user adoption.

INTRODUCTION

Access, financial management, optimized care processes, and quality patient outcomes are fundamental to the enterprise mission. Senior leadership and service line owners need more accurate forecasting of key performance metrics to (1) anticipate future volumes to better monitor and optimize multiple service lines and locations and (2) inform preemptive intervention strategies and/or resource realignment efforts.

This initiative sought to address those needs by developing forecast metrics for key organizational metrics including Emergency Department (ED) visits, Hospital Observations, Hospital Admissions, Hospital Discharges, and Surgeries in all health system facilities and for all service lines; and Outpatient office Evaluation & Managment (E&M) Visits for all service lines.

Project objective statement:

Provide leadership with actionable insights to improve organizational performance

- *Leverage enhanced statistical methodologies to support business cycle*
- *Build statistically validated and accurate forecast models to predict key metrics*
- *Simplify and streamline information delivery through automation to support operational meetings*
- *Align corporate statistics and financial performance management reporting and reduce or eliminate manual custom report creation efforts*

We will emphasize the Outpatient E&M Visits statistic; but ED Visits, Observations, Admissions, Discharges, and Surgical Cases had similar structures for data preparation, variable usage, model development, and deployment.

ANALYTIC STRATEGY & APPROACH

Statistical forecasting uses time series approaches to model metric behavior and can account for explanatory feature effects. Monte Carlo (MC) simulation utilizes random variable generation to model risk or uncertainty and used forecast model output as the source to generate required levels of detail for reporting.

Statistical forecasting methods require time-stamped data in fixed time intervals; do not assume observation independence; and fit historical data to forecast estimates. Explanatory features must be available for historical and future time periods for model development and forecasting.[1,2]

Autoregressive integrated moving average – ARIMA

A time series regression method generally labeled *ARIMA(p,d,q)(P,D,Q)*$_m$ that can include a vector of independent feature variables and terms for autocorrelation *(p)*, moving average *(q)*, integration or differencing *(d)*, and seasonal effects *(P,D,Q)* over period *m*.

The general form of non-seasonal ARIMA is

$$z_t = \delta + \varphi_1 z_{t-1} + \cdots + \varphi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

ARIMA models that include exogenous variables (*X*) are often referred to as ARIMAX models, and provide an additional method with which to inform forecasts.

Exponential Smoothing - ESM

A method that uses smoothing constants to weight the level *(L)*, trend *(B)*, and seasonal *(SN)* effects of more recent time periods higher than that of more remote. Effects can be additive or multiplicative.

The general exponential smoothing form is

$$\text{Additive Holt-Winters: } y_t = L_{t-1} + B_{t-1} + SN_{t-L} + \varepsilon_t$$

$$\text{Multiplicative Holt-Winters: } y_t = (L_{t-1} + B_{t-1}) + SN_{t-L}(1 + \varepsilon_t)$$

Unexplained Components Decomposition - UCM

An alternative method that decomposes a time series into several component factors: trend *(TR)*, seasonal *(SN)*, cyclical *(CL)*, and irregular *(IR)* effects. Effects can be additive or multiplicative.

The general UCM form is

$$\text{Additive Decomposition: } y_t = TR_t + SN_t + CL_t + IR_t$$

$$\text{Multiplicative Decomposition: } y_t = TR_t * SN_t * CL_t * IR_t$$

Intermittent Demand - IDM

A time series method appropriate when observations are sporadic or where demand - the dependent variable - is frequently any constant number or zero (0). Intermittent demand models address the assumptions of other weight-based methods which when applied to intermittent data will bias forecasts toward zero (0).[3,4,5]

The general form is:

$$\text{If } X_t \neq 0 \text{ } then \begin{cases} Z_{t+1} = \alpha X_t + (1-\alpha) Z_t \\ V_{t+1} = \alpha q + (1-\alpha) V_t \\ Y_{t+1} = \dfrac{Z_{t+1}}{V_{t+1}} \end{cases}$$

$$\text{If } X_t = 0 \text{ } then \begin{cases} Z_{t+1} = Z_t \\ V_{t+1} = V_t \\ Y_{t+1} = Y_t \end{cases}$$

Where

Z_t = estimate of mean non-zero demand at time t

V_t = estimate of mean interval size between non-zero demands

X_t = observed demand

α = smoothing parameter between 0 and 1

q = current number of consecutive zero-demand periods

Y_t = estimate of mean demand

MONTE CARLO SIMULATION

Monte Carlo simulation is a mathematical technique that uses variables to model a desired numerical output based on pre-defined probability distributions and repeated sampling. A simulation models that utilizes random variables that change instantaneously would be classified as a *discrete stochastic model*[6]. This is the type of model we will utilize.

Monte Carlo methods rely on the core statistical principle 'The Law of Large Numbers' which states:

> "*As the number of identically distributed, randomly generated variables increases, their sample mean approaches their theoretical mean.*" [7]

The statistical forecast is treated as the theoretical mean with its calculated standard deviation. MC simulation allocates the forecast values to lower levels of detail not present in the forecast models that are consistent with the aggregate level (i.e. theoretical mean) and forecast uncertainty.

MC simulation is applied using the sample size generated by a normally distributed random variable by day of week (excluding holidays) for each level of the forecast hierarchy. The random variable uses the forecast mean and standard deviation calculated from the prior 365 days of historical data.

The Monte Carlo approximation of a random variable using samples can be mathematically defined by:

$$E(X) \approx \frac{1}{N} \sum_{n=1}^{N} x_{\mathrm{n}}$$

Where:

E(X) = expected value of our random variable.

N =total number of random values

X_n = value of the random variable in relation to *n*

DATA PREPARATION, EXPLORATION AND SETUP

All data for this analysis originates from operational and financial systems that is consolidated and maintained in the enterprise data warehouse. Data starting from January 1st, 2007 was used for the time series. E&M volumes are aggregated by visit date as the primary time variable and service line as primary hierarchy variable.

Two operationally relevant sub-statistic rollups were created using case statements to group visit types into: (1) New, Established, and Consultation Visits and (2) Observation, Emergency, Other Visits. Time series models require continuous and consecutive historical and forecast horizon time periods. Missing data or time periods prevent forecasts generating accurately. We did identify instances where no visits occurred for particular date-service line pairings (true nulls) and created imputation rules to substitute zero (0) values.

Feature engineering included day-of-week cycle and month-of-year seasonal effects; major and minor holidays; and major professional medical conferences. The number of time slots available for patients was forecast separately and added as a learned feature.

Finally, consideration was made to account for the time lag in operational processing of data prior to it being available for forecasting. A separate analysis determined 13 days was the optimal lag period from the processing date to assure complete historical data. This lag was then programmatically set as the maximum historical time period and lag-1 as the first period of the forecast horizon. Actual, incomplete data between the processing date and lag date is excluded from the analysis.

FORECAST MODELING, OUTPUT, AND ASSESSMENT

SAS Forecast Studio® 14.2 was used to generate all forecasts. Forecast Studio automates model selection, parameter estimation, and forecast generation by leveraging routines to assess and fit data using a native model repository to produce system generated models.

Candidate models are presented in a tournament fashion with the best performing model, based on the specified model diagnostic criteria, selected as the final model. Custom models can be developed and were added to the tournament for consideration.

Forecast settings included:

1. Hierarchy = Service line, Sub-statistic roll-up
2. Time interval = Day
3. Seasonal cycle length = 7
4. Missing values = 0
5. Independent variable significance = 0.05
6. Model methodologies: System generated ARIMA, exponential smoothing, unexplained components decomposition, intermittent demand, ensemble system generated models; and customized ARIMA models.
7. Holdout sample for model selection = 150 days
8. Model diagnostic criteria= Mean Absolute Error (MAE)
9. Reconciliation Method= Bottom Up

System generated models were inspected for each level of the hierarchy. Custom ARIMA models were developed and added for consideration. A holdout sample of 150 days at the end of the time series was withheld from model development. This sample was used to assess model accuracy, determine final model selection, and provide an estimate of post-deployment forecast performance. Final model selection was based on the minimum MAE of the holdout sample and used to forecast future time periods.

Forecasts were generated at the lowest level of the hierarchy and reconciled from the bottom up to aggregate forecasts. This approach has the advantage of increased sensitivity to local effects.

Rolling simulation was performed on out-of-sample time periods to assess stability and lead-time performance of the final forecast model. Uncertainty increases as the forecast horizon increases and is often visualized as an increasing funnel of forecast confidence intervals. Outpatient E&M model statistical results are summarized in **Appendix A**.

MONTE CARLO SIMULATION

Monte Carlo (MC) simulation was used to allocate forecasts from the bottom level of the model hierarchy (sub-statistic rollup) to further levels of drill-down detail necessary for reporting. Simulations were developed in Base SAS® 9.4 and implemented using SAS Enterprise Guide®.

We opted to build aggregated models and then allocate down via simulation for the following reasons:

1. Adding additional hierarchy levels exponentially increases the number of forecast models required to develop and manage.
2. Aggregating to higher levels will inherently remove some of the noise that exists in extremely low levels of detail. This makes fitting models to pattern more realistic. We want to avoid fitting to random noise.
3. Stable forecasts are produced for at higher levels while still capturing the distribution patters at lower levels of detail with simulation.
4. Forecast uncertainty can be also be represented in the lower levels of detail.

The E&M forecast model hierarchy has 39 unique service lines levels and 2 unique sub-statistic rollup levels. In total, there are 77 champion forecast models selected at the bottom level that project forward 84 days. In total, this is 77*84=6,468 separate forecast values that will be used in our sample algorithm. All 6,468 values will have 100 samples taken via a looping macro (646,800 samples). Each loop iteration will generate a random variable using a normal distribution with mean (μ) equal to the forecast value and standard deviation (σ) equal to the calculated standard deviation of the relevant model hierarchy level over the last year excludes holidays.

Key SAS® procedures used in the simulation code:

- *Data* step - standard SAS data manipulation, relatable to SQL.
- *Proc SQL* – procedure for executing SQL like code in SAS.
- *Proc SurveySelect* – Sampling procedure in SAS
  - *Method=URS* – this proc stands for 'Unrestricted random sampling'. This means sampling with replacement
  - *Option Outhits* – this specifies that instead of returning a row with a count of how many times it was selected, a row will be returned each time selection occurs.
- *Proc Summary* –Used to calculate summary statistics. In this case, specifically standard deviation.

Key SAS® functions utilized in the simulation:

- *Rand* – used to generate random numbers via a specified distribution and its parameters.

- *MDY* – creates a date formatted field with a month, day, and year numeric input.
- *Datepart* – used to convert SAS datetimes into dates or other numbers.
- *Weekday* – input a date and return the number of that day of the week (i.e. Sunday =1)

A number of data preparation and calculation steps are completed prior to simulation loop execution and include:

1. Create a table with the time series predictions and standard deviation.
2. Create a table called that contains the encounters to sample from the data source.
3. Create a control table of the sample size input for the simulation procedure. The sampling function used is PROC SURVEYSELECT.

Execute the simulation loop macro:

```
%macro Simulation();
        Options Compress=No;
        %do week=&MinWeek %to &MaxWeek;
                data work.Samplesize_1;
                        set brom_t.samplestrata_week_5;
                        do Sim=1 to &SimsInpt;
                                round0=PREDICT;
                                round1=round(PREDICT,1);
                                round2=abs(PREDICT-round1);
                                _NSIZE_=rand('Normal', PREDICT,STD);
                                RoundNsize=Round(_NSIZE_,1);
                                weeknum=&Week;
                                where WeekCount=&Week;
                                output;
                        end;
                        keep Strata _NSIZE_ weeknum;
                        ;run;
                PROC SQL;
                        CREATE TABLE WORK.SAMPLESIZE_1 AS
                                SELECT t1.Strata,
                        (case
                                when (round(SUM(t1._NSIZE_),1))<0 then 0
                                else (round(SUM(t1._NSIZE_),1))
                                end)
                        AS _NSIZE_
                        FROM WORK.SAMPLESIZE_1 t1
                                GROUP BY t1.Strata;QUIT;
                PROC SURVEYSELECT noprint data=brom_t.Bank_3
sampsize=work.samplesize_1
                        method=URS
                        out=work.Sampling outhits;
                        strata strata;
                data work.sampling_2;
                        set work.sampling;
                        Sims=100;
                        WeekNum=&Week;
                        ;run;
                data brom_t.SimResults_&Week;
                        set brom_t.SimResults_&Week work.Sampling_2;
```

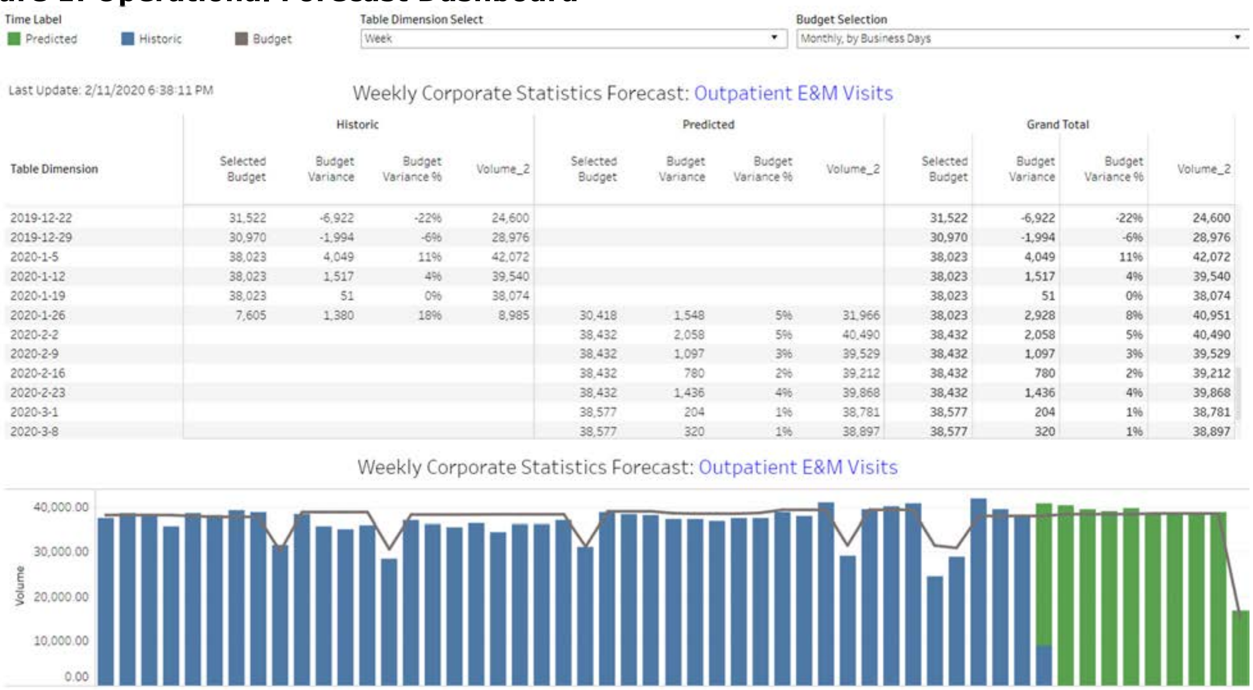
```
                    drop /*SelectionProb*/
                    ExpectedHits SamplingWeight;
                run;
        %end;
%mend;
```

The final output of the simulation contains the full range of samples for each day of the forecast for all original forecast hierarchy levels and the additional drill-down levels required for reporting.

VISUALIZATION AND REPORTING

The final output of the forecasting and simulation programs was loaded into the enterprise data warehouse and transformed to be integrated into enterprise dashboard reporting. Figure 1 provides daily and weekly forecasts displayed with organizational budget targets. Controls allow users to aggregate by facility or drill down into specific service lines and/or sub-statistics.

Figure 1: Operational Forecast Dashboard



	Historic				Predicted				Grand Total			
Table Dimension	Selected Budget	Budget Variance	Budget Variance %	Volume_2	Selected Budget	Budget Variance	Budget Variance %	Volume_2	Selected Budget	Budget Variance	Budget Variance %	Volume_2
2019-12-22	31,522	-6,922	-22%	24,600					31,522	-6,922	-22%	24,600
2019-12-29	30,970	-1,994	-6%	28,976					30,970	-1,994	-6%	28,976
2020-1-5	38,023	4,049	11%	42,072					38,023	4,049	11%	42,072
2020-1-12	38,023	1,517	4%	39,540					38,023	1,517	4%	39,540
2020-1-19	38,023	51	0%	38,074					38,023	51	0%	38,074
2020-1-26	7,605	1,380	18%	8,985	30,418	1,548	5%	31,966	38,023	2,928	8%	40,951
2020-2-2					38,432	2,058	5%	40,490	38,432	2,058	5%	40,490
2020-2-9					38,432	1,097	3%	39,529	38,432	1,097	3%	39,529
2020-2-16					38,432	780	2%	39,212	38,432	780	2%	39,212
2020-2-23					38,432	1,436	4%	39,868	38,432	1,436	4%	39,868
2020-3-1					38,577	204	1%	38,781	38,577	204	1%	38,781
2020-3-8					38,577	320	1%	38,897	38,577	320	1%	38,897

Forecasts were integrated into several additional dashboards based on stakeholder requirements to facilitate operational support and executive leadership meetings and goals.

Executive and operational stakeholder engagement has been a constant feature in the conception, development, and deployment of the forecasts. This ongoing dialog has been key to the project's success; but it is critical for forecast deployment and adoption.

The organization is increasingly shifting from reactive to proactive decision making. The change requires new predictive tools like the forecast, but also requires different approaches for evaluating and interpreting analytics. Looking retrospectively, an outcome can be compared to a target. Both are known and fixed values. Forecasts are estimates of a future outcome with a calculated level of uncertainty.

Model deployment was accompanied by comprehensive user documentation, multiple stakeholder education sessions, and ongoing coaching and mentoring. Statistical literacy and understanding variation are necessary to properly weigh the analytic in light of the decision to be made including whether and how to intervene proactively in the process.

CONCLUSION

Statistical forecasting is playing an increasing important role in assessing and managing key performance indicators at the Cleveland Clinic. Time series forecasting in combination with Monte Carlo simulation allow for developing stable forecasts at higher hierarchy levels and reliably allocating those estimates to lower levels of detail while still capturing important distribution patters. Collaborative development and deployment are necessary for effective stakeholder adoption.

REFERENCES

[1] Bowerman BL, O'Connell RT, Koehler AB. *Forecasting, Time Series, and Regression. An Applied Approach.* Books/Cole, 2005.

[2] SAS©Forecast Studio 14.2 User's Guide. SAS Institute.

[3] Croston, J. "Forecasting and stock control for intermittent demands", *Operational Research Quarterly*. 1972; 23(3); 289-303.

[4] Shenstone, L., and Hyndman, R.J. Stochastic models underlying Croston's method for intermittent demand forecasting. *Journal of Forecasting*. 2005; 24:389-402.

[5] West KD. Encompassing tests when no model is encompassing. *Econometrics.* 2001;105(1):287-308.

[6] Rubinstein, Reuven Y. and Kroese, Dirk P. Simulation and the Monte Carlo Method. 2017; 91-101.

[7] Routledge, Richard. Encyclopedia Britannica, https://www.britannica.com/science/law-of-large-numbers.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Michael Bromley
Cleveland Clinic
bromlem@ccf.org

APPENDIX A

RESULTS SUMMARY AND MODEL VALIDATION:

Details of model training and validation performances are summarized in Table 1 and 2.

Table field definitions:

1. ***Sub Statistic*** – Sub-statistic rollup used in the model hierarchy
   a. *New_Est_Consult* -New, Established and Consult visit types
   b. *Other* - ED, Observation, Other visits types
2. ***Model Type*** – Model used for hierarchy level
3. ***Adj RSQ*** – Adjusted R-Squared statistic; how well the model explains variation in the data
4. ***MAE*** – Mean Absolute Error; averaged daily error over the sample period (150 for holdout, all data for in-sample)

Table 1: Hold-out Sample Statistics

	Levels	Min Adj RSQ2	Ave Adj RSQ2	Max Adj RSQ2	Min MAE	Ave MAE	Max MAE
New_Est_Consult	**33**	**-3.26**	**0.70**	**1.00**	**0**	**15.5**	**138**
ARIMA	30	-3.26	0.79	1.00	2	20.1	138
ESM	1	-0.05	-0.05	-0.05	0	0.0	0
IDM	2	-0.45	-0.32	-0.19	0	0.5	1
Other	**31**	**-0.52**	**0.11**	**0.83**	**0**	**3.5**	**56**
ARIMA	16	-0.16	0.32	0.83	0	7.3	56
ESM	10	-0.30	-0.07	0.01	0	0.5	2
IDM	5	-0.52	-0.23	0.05	0	0.2	1

Table 2: In-Sample Statistics

	Levels	Min Adj RSQ2	Ave Adj RSQ2	Max Adj RSQ2	Min MAE	Ave MAE	Max MAE
New_Est_Consult	**38**	**-60.72**	**-0.83**	**1.00**	**0**	**15.9**	**179**
ARIMA	30	0.48	0.93	1.00	1	19.8	179
ESM	5	0.00	0.31	0.60	0	2.4	14
IDM	2	-60.72	-30.60	-0.47	2	4.50	7
Other	**37**	**-42.24**	**-2.84**	**0.99**	**0**	**3.3**	**47**
ARIMA	17	0.00	0.46	0.99	0	6.0	47
ESM	15	-42.24	-2.65	0.58	0	0.9	8
IDM	5	-37.29	-14.62	-5.23	1	1.4	2

Weekly Aggregated Results:

Weekly aggregations are an important operations and management perspective making it important to assess model accuracy from that vantage point.

Aggregations are calculated with the following logic:

1. All weekly forecasts are combined and to calculate the overall accuracy
2. Each individual forecast is stored as a snapshot to calculate an average accuracy for a specific time period

Table field definitions:

1. ***Week*** – Sunday marking the first day of the week
2. ***Total Daily Forecasts*** – Count of forecasts for specified week
3. ***Error %*** - Week Error divided by Weekly Actuals
4. ***MAPE*** – Absolute value of the error %

Table 3: Weekly Accuracy for ED Visits Type

Week	Total Daily Forecasts	Error%	MAPE
13-Oct-19	7	-1.16%	1.16%
20-Oct-19	12	-1.48%	1.48%
27-Oct-19	19	-0.42%	0.42%
3-Nov-19	26	-1.50%	1.50%
10-Nov-19	33	1.41%	1.41%
17-Nov-19	40	-2.41%	2.41%
Total	**137**	**-0.92%**	**1.40%**