Predicting COVID-19 Hospitalizations

By: Nick Seah, Michael Shim, and Kesin Dolwani

Introduction

- COVID-19 Pandemic
- Exposure of flawed healthcare system and infrastructure
- Unprepared hospitals
- Lack of ICU's
- Multi-State Data with the following information:
    - Date, death, death confirmed, death increase, death probable, hospitalized, hospitalized cumulative, hospitalized currently, **daily hospitalizations**, in ICU currently, negative, negative increase, etc.
- Data Interested: **Daily Hospitalizations (3/8/21- 3/14/21)**

Formulation

- Time-Series Forecasting
- Stationary Dataset
  - Mean, Variance, etc.
- ARMA Model
  - Autoregressive (AR)
  - Moving Average (MA)
- ARIMA Model
  - Autoregressive (AR)
  - Moving Average (MA)
  - Integrated with order-d differencing

ARMA Model

$$X_t = \left(\alpha_1 X_{t-1} + .. + \alpha_a X_{t-a}\right) + \left(\beta_1 E_{t-1} + .. + \beta_m E_{t-m}\right) + E_t + \lambda$$

- Autoregressive Model

  AR(n)

  $$X_t = \alpha X_{t-1} + E_t + \lambda$$
  $$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-1} + E_t + \lambda$$

- Moving Average Model

  MA(m)

  $$X_t = \beta E_{t-1} + E_t + \lambda$$
  $$X_t = \beta_1 E_{t-1} + \beta_2 E_{t-2} + E_t + \lambda$$

ARIMA Model

$$X_t = \left(\alpha_1 X_{t-1} + .. + \alpha_a X_{t-a}\right) + \left(\beta_1 E_{t-1} + .. + \beta_m E_{t-m}\right) + E_t + \lambda + \left(\theta_1 X_{t-1}^{(d)} + .. + \theta_m X_{t-m}^{(d)}\right)$$

- Autoregressive Model
- Moving Average Model
- Integrated with Order-d Differencing

$$X_t^{(d)} = X_t^{(d-1)} - X_{t-1}^{(d-1)}$$

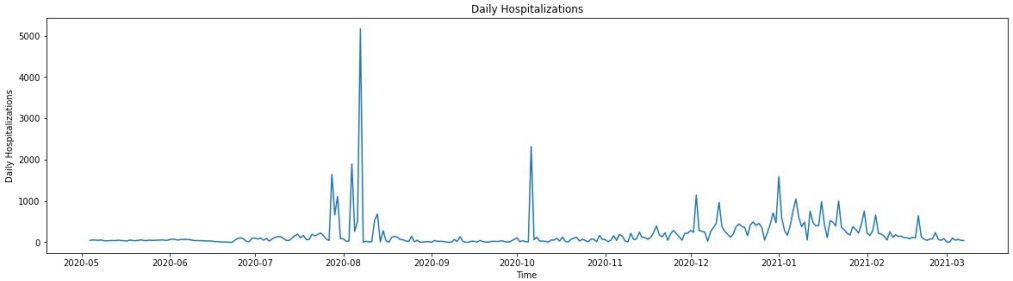
$$X_t^{(1)} = X_t - X_{t-1}$$

Order-1 Differencing: Speed of Change

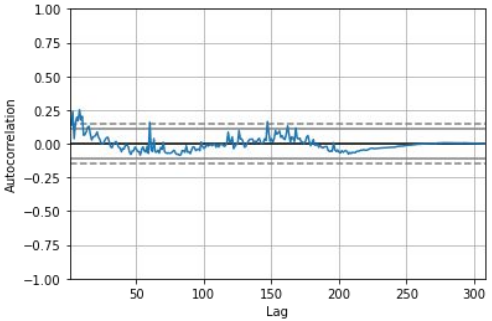
$$X_t^{(2)} = X_t^{(1)} - X_{t-1}^{(1)}$$

Order-2 Differencing: Degree of Acceleration

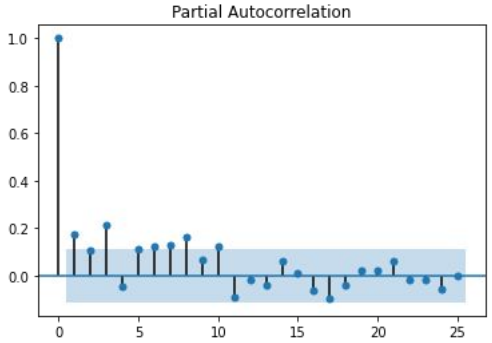
Visualizations



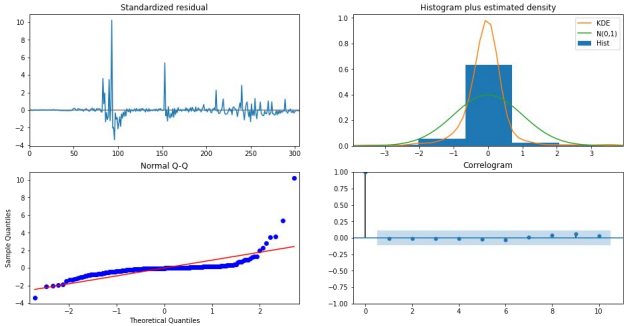
Daily Hospitalizations Training Data



ACF Plot (q = 0-10)



PACF Plot (p = 0-10)



ARIMA Plot Diagnostics

Methodology

- Training data from the COVID tracking project from 5/4/2020 to 3/7/2021
- Testing data from the ADHS hospitalizations public data
- Autoregressive Integrated Moving Average Model (ARIMA)
- Autoregressive Moving Average (ARMA)
- Heuristics on ACF plot to select a lag range of [0-10] for hyperparameter tuning
- Heuristics on PACF to select a moving average window of [0-10] for hyperparameter tuning
- Grid Search to select optimal ARIMA and ARMA parameters
- RMSE to evaluate forecasting accuracy

Results

Hospitalizations

		3/8/21	3/9/21	3/10/21	3/11/21	3/12/21	3/13/21	3/14/21	Total	RMSE
ARMA	Prediction	84	80	87	99	102	105	112	669.9	47.8
	Actual	68	74	66	36	57	45	37	383	
ARIMA	Prediction	45	64	54	49	56	46	44	358	11.8
	Actual	68	74	66	36	57	45	37	383	

Interpretation

- The ARMA Model performed significantly worse than the ARIMA Model with a RMSE of 47.8 compared to 11.8
- Arizona hospitals were not overwhelmed from 3/8/21 - 3/14/21
- ARIMA is better for time series forecasting of COVID-19 Hospitalizations
- Use ARIMA over ARMA for future epidemic predictions.
- While ARIMA performed relatively well, there is room for improvement as shown by plot diagnostics.

Questions?