ARIMA(BMOPD)

—Lumin @August 2022
4441 Probability and Statistics 1

INTRODUCTION

Why torture the social sciences with math? Because it's good for them!! Descriptive statistics are everywhere, though not usually labeled as such. A GPA is a descriptive statistic, the number describes an average of cumulative data. We are pretty familiar with this idea of central tendency, but one of the things we don't really hear about is how psychology is measuring people. But it's no different than the grades you got in school. There are rubrics established, and they take a large consensus to change. Instead of saying "good," "bad," or "ugly," applying calculus to any *Bob* is a better way to get have more people be able to access what is changing and hopefully plied with at least one layer of "why" (as seen in social data triangulation methods). In a more recent development: "A new public opinion poll released 'today' by the American Psychiatric Association (APA), 62% of Americans feel more anxious than they did at this time last year." October 21, 2020.[1] In my opinion, any jump in a population measured at %62 is in an immediate need of study. So how can we consistently measure *Bob,* is the question; and can we use that structure of measurement as a remedy tool?

Getting meanings of words down to consistent starting position does take more than one approach. "Triangulation" used to associate meaning with an equal image and intention, in this case, can be recorded in a series of three digit codes.

For every interaction with *Bob*, we can use the responses to *words*, *images* or the *future* to correlate *feelings, awareness* and *intention* then rate the collection for severity and projection. Assume that *Bob* suffers from panic attacks. Creating a workflow to be as complex as what *Bob* experiences is difficult, but the diagram (right) can gather a lot of information interactively, as well as add information to attempt to direct his reactions to an environment.

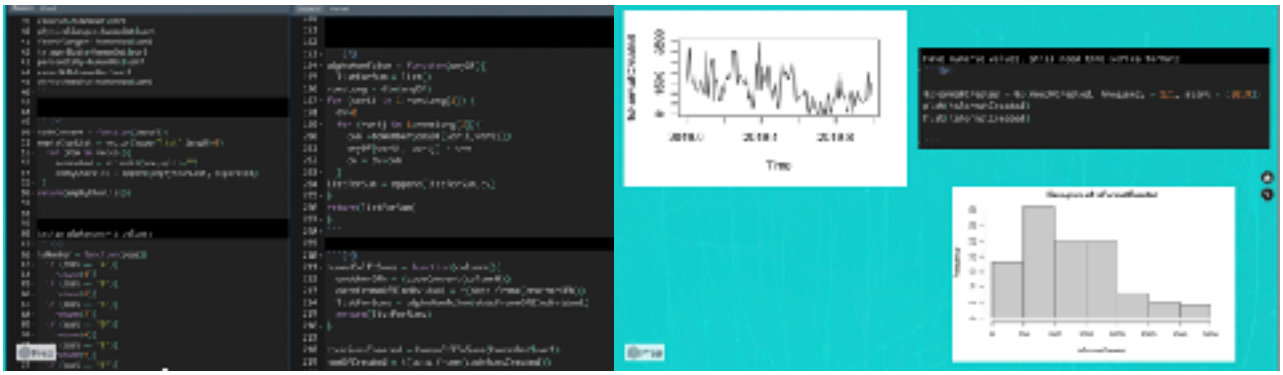
— Behavior Model of Panic Disorder (BMOPD)

By placing a history of progression through the behavior model, we can look at each event's score and determine the over-all influence the structure might have on *Bob*. These scores are seen through out disciplines such as psychology and history[2] and even though the translation from alphanumerical data into finite statistics is time consuming, the density of indicators is large. Each event is unique and once turned into a data point, quite telling. Ultimately, we can attempt to forecast general tendency to panic and hopefully add to the answer of resolve. Looking only at the category "created" we can assess *Bob's* severity of symptoms over a period of time—from the first recorded event to the last.

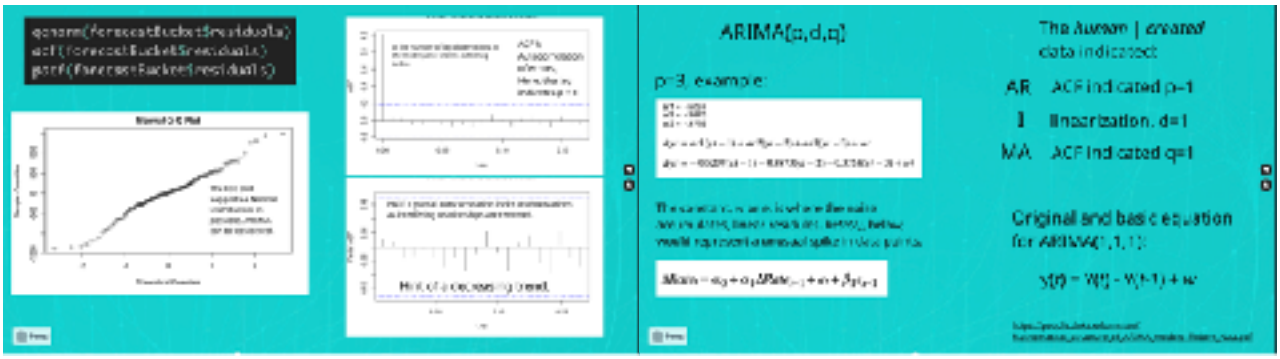
The ‘created’ score values that I have worked with might appear to have a positive outlook for this person, but, as I have already said, society would like to see a particular outcome, for the general health and wellbeing of the population. To gather empirical data and be able verify the probability that a person is migrating toward one outcome or another is how we can remove the symbolism and potential bias from this exploration.

METHODOLOGY

Using the R language in the R Studio API, I wrote a function that converts any length of historical events to values. If that data, in this case 121 assessments, can be seen as normally distributed set, we can use R to get an objective forecast the subjects progression by using the ARIMA(*p*,*d*,*q) time-series* function. The AutoRegressive Integrated Moving Average (ARIMA) assumes that “…an autoregressive specification is theoretically consistent with settings where dynamic forces bringing about change are expected to have some degree of persistence.”[3]



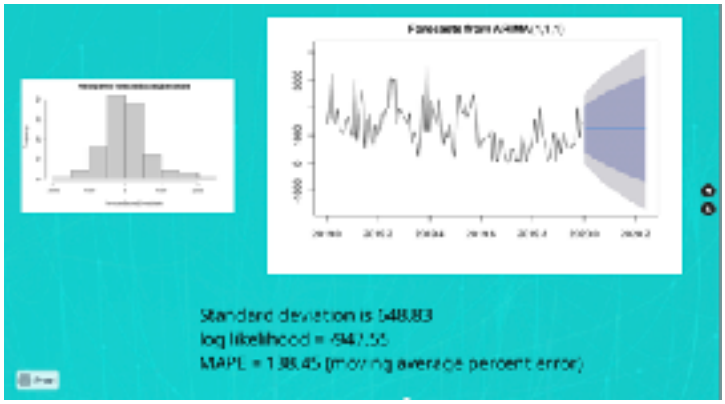
Above is an example the the conversion function (left) and initial residual assessment of the raw data. Given the difference between a theoretical Normal Distribution (right), there is no indication that a forecast of behavior wouldn’t be plausible—there are no radically different scores relative to the sets own standard deviation.



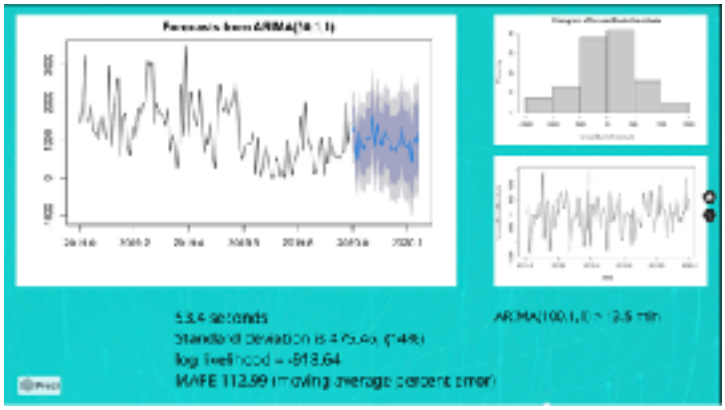
AUTOREGRESSION

ARIMA(BMOPD)

—Lumin @August 2022
4441 Probability and Statistics 1



Values for the *p*,*d*,*q* parameters are determined by the data's ability to be examined as a linear function including the latest recorded point and the previous event's score out to the number of ACF (AutoCorrelation Function) fitments it took to derive the proper relationship factor so the the degree that the past could be used to predict the future. The PACF (Partial Autocorrelation Function) plot has one significant spike (above, left), meaning that all the higher-order autocorrelations are effectively explained by the lag-1 autocorrelation, and since there is only one, the differencing (Integration, if on non-finite time curve) only needs to happen one time—as rule of thumb. This puts the estimate of the parameters of as: ARIMA(1,1,1) (above, right).



"If the lag-1 autocorrelation is zero or even negative, then the series does not need further differencing. You should resist the urge to difference it anyway just because you don't see any pattern in the autocorrelations! One of the most common errors in ARIMA modeling is to "overdifference" the series and end up adding extra AR or MA terms to undo the damage."[4] And if the values should deviate by *p* or *q*, they should not deviate at the same time.

CONCLUTION

The most promising forecast adheres to that *pq*-rule mentioned above and uses 50 previous data points to the last in the time-series. Considering that we are specifically looking for an answer to a over-all question of increase or decrease in the severity of panic attacks for *Bob*, spending the near-minute time to run the program's AR (*p*) parameter at 50 is appropriate to get a cursory balance of how progressive higher or scores might relate to one another or not. An "ordinal" variable implies order in the data. Both are categorical. And usually produce a ratio within an interval given the related continuity between categories. These details are important with testing the efforts of the BMOPD workflow—which, because of its rich density of causal information, can account for parallel intersections of reasons for any change in behavior. The probable many layers or "why." This is also why the behavior model can be referred to a structure and evaluated for accuracy at any phase of interaction. With the proper adaptive intelligence and machine learning, we just might be able to measure the intangible nature of our society in real time and steer it to a desired resolution.

Quoted references:

[1]https://www.psychiatry.org/newsroom/news-releases/anxiety-poll-2020
[2]https://www.researchgate.net/publication/328982105_Data_scopes_for_digital_history_research
[3]https://www.researchgate.net/publication/327896034_ARIMA_Time_Series_Models_for_Full_Truckload_Transportation_Prices
[4]https://people.duke.edu/~rnau/411arim2.htm

Contributing reading:

https://www.statisticshowto.com/akaikes-information-criterion/
https://stats.stackexchange.com/questions/179558/auto-regression-versus-linear-regression-of-xt-with-t-for-modelling-time-series
https://otexts.com/fpp2/non-seasonal-arima.html
https://people.duke.edu/~rnau/Mathematical_structure_of_ARIMA_models--Robert_Nau.pdf
https://analyticsindiamag.com/5-conditions-when-the-arima-model-should-be-avoided/
https://en.wikipedia.org/wiki/Akaike_information_criterion