Luke Momjian Data Challenge response, part 1

Nov 30 2019

This discussion of results should be read in conjunction with the Excel file Luke Momjian Data Challenge Visualizations.

**Insights (Trends over Time tab)**

The data provided in the CSV indicates that progress has been made among the cohort in 3 of the 4 response types tracked (all but sleep).

Anticipatory stress exhibits a declining trend (good) when average monthly values are plotted from August 2018 to July 2019 (see trendline). There are two peak values in this graph, in December 2018 and March 2019. I am hypothesizing that the December peak is due to Christmas. The March peak could be due to tax season and/or Easter. In 2019, tax day was April 15 and Easter was April 20. Healthcare providers should be aware of and expect these peaks so as to better combat the stress experienced by patients in these two months.

Mood exhibits a slight increasing trend (good), with a low value in Aug 2018.

Rumination stress exhibits a declining trend (good), with a high value in Aug 2018.

The prior mentioned improvements in anticipatory stress, mood, and rumination suggest that participants are generally making progress in their treatment.

Sleep demonstrates a slight declining trend (bad), with a high value in Aug 2018. This is the only variable among the 4 that has not shown signs of improvement.

The outliers present in most of these variables for August 2018 suggest to me that the patient population responding in August 2018 is different in some relevant characteristic(s). It is also possible that an adjustment in the patient response methodology was made around August to cause scores to seem more positive.

**Insights (Distribution of Averages tab)**

The graphs included in this tab group the participants by their averages scores for each of the 4 response types tracked. Participants with the same average, when rounded to one decimal place, and grouped into the same cohort. The resulting graphs can be used by healthcare providers as a reference to compare to an individual patient’s numerical responses.

For an individual patient, an average response value of less than 2.3 indicates that the patient is significantly below-par in terms of their sleep quality, where par is determined by the scores of all participants in these data.

A mood average of <2.2 indicates that the patient is below par in their mood.

An average value of over 3.4 (approx.) for anticipatory stress indicates that the user has a greater than average propensity for this sort of stress.

For rumination, a value of >3.0 suggests significantly above-par rumination stress.

These values are determined solely by visual analysis of the graphs. A more rigorous, accurate approach involving statistical analysis of variance could be used given more time.

**Assumptions**

Like I mentioned previously, I am thinking that some change was made in the way responses were recorded and/or the population responding in August 2018 was slightly different from the entire population as a whole in this period (Aug 2018 to Jul 2019). Considering this, I would consider the August 2018 values outliers or bad data and hesitate to include them in data analysis.

I am assuming that for the rumination and anticipatory stress responses, the value 0 corresponds to the response “Not at All”, and the value 4 corresponds to “Extremely”.

I assume that patients are generally capable of self-reporting their stress, mood, and sleep levels in a way that accurately represents their objective level or progress in these aspects.

**Additional information to collect**

It would be very useful to collect information on the treatment that patients are undergoing. Knowing when a patient has begun medication or therapy, as well as when there is a shift or cessation, would be extremely informative and allow for determining how effective or ineffective treatment options are for patients. Data on the exercises that the patient is assigned and completing should be collected. Treatment information could be compared to the responses to see if they have made any improvements.

The patient’s doctor or hospital affiliation would be useful in analysis.

Demographic information would help enrich the data, as age, sex, income level, race, location, or other characteristics might have an effect on participant scores and responses to treatment.