



What I did in the two graphs above was the following:

1. For each point , I generated N patients with 2 years worth of monthly seizure counts
   1. For each patient, I did the shapiro-wilk test to see if the data was reasonably gaussian
   2. ‘Reasonably gaussian’ in this case means that the p-value returned by the shapiro-wilk test was greater than 0.05
      1. If the p-value was less than 0.05, then that means the data got rejected as not being gaussian enough
   3. I counted up the number of patient over all N patients who passed the shapiro-wilk test for gaussianity
   4. I divided the number of patients who passed by N in order to get the probability that data generated with that point would be able to pass as gaussian
2. I did this for every point on the heatmap except for the region where the data was literally incalculable according to the negative binomial
3. After doing that, I overlaid scatter plots of model 1 and model 2 patients on top of the heatmap of probabilities

The heatmap shows that there’s a huge region where gaussianity absolutely should not be considered, and the location of the model 1 and model 2 patient populations show that while the sicker patients of model 2 can be approximated as gaussian, the population of model 1 cannot be safely assumed to be gaussian.

Furthermore, since the SNR paper is going to be looking at all possible regions of the heatmap and not just the location of the model 1 and model 2 patient populations, having a huge region where gaussianity cannot be assumed means that I’m just going to use negative binomial instead.