Signed, Blueno

zhasan1, jfox10, nkeirste, zsaljook

Hypothesis

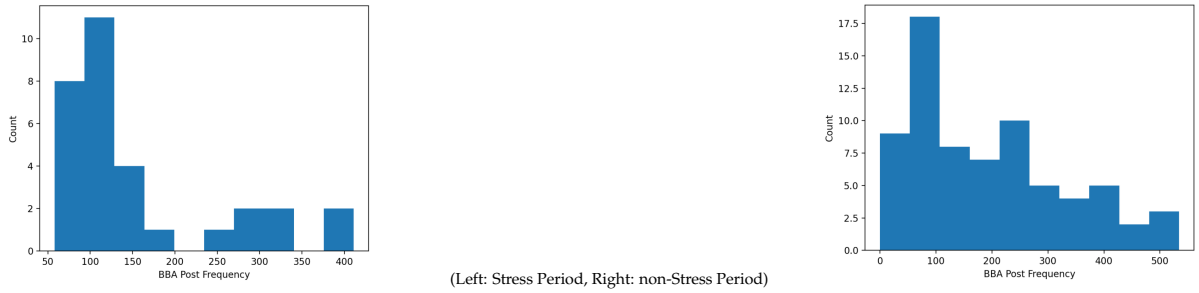
We wanted to see if the post behavior of anonymous online forums changed in times of stress versus times of non-stress. Specifically: **The average weekly frequency of posts on Blueno Bears Admirers (BBA) or the average weekly sentiment of posts on Dear Blueno (DB) are different in high and low stress periods.**

Data

We downloaded HTML files of the BBA and DB Facebook pages. We received these files from the moderators of both pages. We parsed out the relevant data from each HTML file, which was just the date and content of posts, and turned them into CSV files. We then aggregated the BBA data by number of posts in a week, and we blocked off weeks as either stressful or non-stressful. For DB, we used a machine learning algorithm to complete sentiment analysis on each post, and then took averages over a week, so we had data similar to the BBA data.

Findings

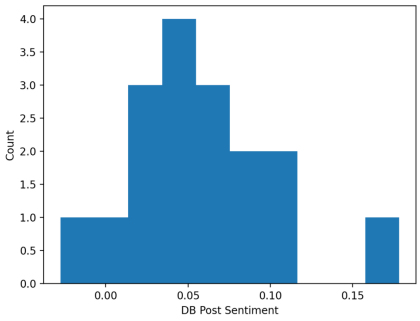
Claim #1: The average weekly frequency of BBA posts does not change in between high and low stress periods.
Support for Claim #1: From the BBA data, we calculated the number of posts made per week. We then divided the weeks into groups of stressful weeks and non-stressful weeks. We then used the Mann Whitney U test to find if the two sets of data were significantly different, and we calculated a **p-value** of **0.66** and a **U score** of **1039.5**. This indicates with high confidence that there is not much of a difference between the 2 sets of data, and therefore supports the claim above. We used the Mann Whitney U test because when we graphed both sets of data, we realized that the distribution was not normal (see below). This meant that we needed a form of non-parametric testing, and the Mann-Whitney U test was the right fit.

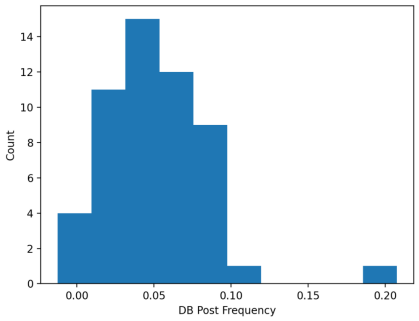


(Left: Stress Period, Right: non-Stress Period)

Claim #2: The average weekly sentiment of DB posts does not change in between high and low stress periods.

Support for Claim #2: Like the BBA data, the goal here was to get an average weekly score for DB posts, except now with sentiment scores instead of post frequency. We used VADER from nltk to find sentiment scores for each individual post, and then averaged those scores with posts in the same week, giving us an average sentiment score. We used the Mann-Whitney U test for the same reason as before (also shown below). We ended up with a **p-value** of **0.366** and a **U score** of **517**.





(Left: Stress Period, Right: non-Stress Period)