In neuroimaging, a common practice is to identify the spatial locations that is statistically significant to an effect. After image repressing which includes spatial registration and normalization, a voxel-based statistical analysis consist of model fitting and computation is performed. Therefore, we want to investigate the problem that whether the empirical null model fits more closely to the actual distribution than the theoretical null model at the same FDR level in neuroimaging. We chose FDR as the thresholding strategy because FDR tends to be more flexible with false positive and have lower thresholds, making it easier to detect the difference between empirical null and theoretical null distribution.

Generating correct inference in terms of any of the strategies for thresholding is highly dependent on the test statistics of null distribution. The paper we are trying to replicate shows example that demonstrate how the theoretical null distribution doesn't fit well with the observed the distribution. Consequently, the empirical null distribution is thought to be the solution to this problem. What's more, We believe this problem is worth studying since we may be able to apply it to much boarder subject such as targeted medicine or even gene therapy for cancer.

In the Jongho Bold dataset, from the research paper, the z-scores of whether the spatial location is significant were collected from the fMRI scan of the brain while receiving a stimulus. Certain image processing techniques like spatial registration and normalization were employed to quantify these scanned images. In order to find the specific spatial locations where an effect is statistically significant, the corresponding data after imaging processing were further calculated into z-scores, for which z-score close to 0 means that part of the brain was not correlated to that stimulus, while large z-score, either positive or negative, shows strong correlation between that part of the brain and the stimulus. As a result, analyzing these z-scores and finding statistically significant ones are critical to solve the problem.

Another dataset called 'Mootha', being used as practice for us, contains the gene expressions collected from diabetes patients and controlled healthy group for

comparison. In total of 10983 genes' expression levels were presented, and for each gene, 34 samples were collected, half from patients with Type 2 diabetes mellitus, and the other half from the group with normal glucose tolerance. As a result, we can conduct a standard two-sample t-test to determine the correctness of null hypothesis, which is for each gene, there is no difference between the mean of diabetes group and the mean of healthy group. Testing null hypothesis using theoretical method is that we have done so far in this quarter.

For the observed data in Jongho Bold dataset, we plot the histogram of z-scores and overlay the theoretical null distribution as well as the empirical null distribution on top of it, allowing us to check directly which distribution fits better. Similarly, for the Mootha dataset, we calculate the p-values and t-statistics of the observed data and overlay them with theoretical null, including uniform distribution for p-values and normal distribution for t-statistics, to check for the compatibility. As a result, we believe that with the two datasets, we should have sufficient data and knowledge to replicate the result of original paper.