Position Paper 1-

Statistical Significance Tests

Merrimack College

Data Exploration

Amol Gote

(Data science student)

May 26th, 2020

"Statistical significance is the probability that the observed difference between two groups is due to chance." (Sullivan and Feinn, 2012). Statistical significance tests (SST's) are computationally simple, provide a logical framework, and have been tried and tested, there are 3 types of SST's Chi-Square, t-test, and F-test, these tests comprise of three terms hypothesis testing, normal distribution, and p-value. Smaller the p-value (< 0.05) significant is the relationship between the variables in the dataset. P-values have become universal and depend heavily on the sample size, p-value can be significant if the sample size is large and but would fail the criteria of < 0.05 if the sample size is small. Finding the right sample size and composition for significance testing is a major challenge and there are various ways in which incorrect statistically significant relationships can be avoided.

The first way is to define proper extraction rules for the sample size, the sample should be representative for each of the categories of the data and appropriate size should be chosen from each of the categories. Random sampling can result in skewed significance results, so instead stratified sampling needs to be selected. For example, if the dataset under study is for students from class 1 to 12, then in that case, data set can be divided into following groups (or strata's) elementary, upper elementary, middle, and high school and chooses records randomly from the from each strata. The second step would be to measure variability and identifying outliers from the dataset, an outlier is data set record which falls outside of what otherwise is considered an acceptable range. Outliers can be identified by the simple scatterplots and then removed from the dataset, alternatively this can be done by calculating the mean and standard deviation of the variable under study, and then marking plus-minus 3 standard deviations away from mean (95%), this will also illustrate the variability associated with the dataset.

Key variables from dataset chosen for significance test should be normally distributed and not skewed, in statistical terms skewed variable is one which has a mean, not at the center of the distribution, it will be either right or left skewed. One of the ways to identify normal distribution is to plot a histogram a visual which shows variables response category frequency distribution, alternatively computational tests like Pearson's coefficient, quartile measures can be utilized. Along with frequency, availability (% of missing values) of variables under study should be analyzed.

Statistical significance test when applied to larger tests raises questions about its applicability, typically for larger dataset sample size is larger. The test results are based on p-value which is based on standard deviation and standard error. "*With a very large sample, the standard error becomes extremely small, so that even minuscule distances between the estimate and the null hypothesis become statistically significant". (Shmueli, 2012)*. On the contrary for the same data with a smaller sample size, the results would have been statistically not significant. In the case of a smaller sample size, it results in a high p-value which favors the null hypothesis and in case of a larger sample size, it leads to rejection of the null hypothesis due to a lower p-value due to very minor potential deviations. A larger dataset with larger sample size tends to introduce over-specification which is the inclusion of variables that are of negligible importance, which can have an impact on SST results. To avoid over-specification metric-aggregation, indexing and transformation can be incorporated. Another issue with these tests is that the results cannot be factored in for future inferences, which means today's results cannot be used for future rollouts which is a major issue for promotional campaigns. Statistical significance tests have its pros and cons and data analysts need to identify proper demarcation between their practical and statistical significance.

Bibliography

Sullivan, G. & Feinn R (2012 Sept). *Using Effect Size—or Why the P Value Is Not Enough.* Retrieved from Journal of Graduate Medical Education: https://www.jgme.org/doi/full/10.4300/JGME-D-12-00156.1

Shmueli, G. (2012, December 6). *Too Big to Fail: Large Samples and the P-Value Problem.* Retrieved from author's website: https://www.galitshmueli.com/system/files/Largesample-12-6-2012.pdf

Kenton, W. (2020, Jan 27). *Statistical Significance*. Retrieved from Investopedia: https://www.investopedia.com/terms/s/statistically_significant.asp

Koehrsen, W. (2018, Feb 2). *Statistical Significance Explained*. Retrieved from towards data science website: https://towardsdatascience.com/statistical-significance-hypothesis-testing-the-normal-curve-and-p-values-93274fa32687

Laurie. (2016, Nov 13). *Large amount of observations: Statistical Test not so Statistical*. Retrieved from medium: https://medium.com/data-design/large-amount-of-observations-statistical-test-not-so-statistical-3d8ed0e94be