Data Mining (RFT and FDR)

Name

Professor

Institutional Affiliation

Date

Course

**Denote the study**

The study's primary objective was to compare False Discovery Rate (FDR) and Random Field Theory (RFT) in examining one-dimensional biochemical data sets. The authors, Naouma & Pataky (2019), highlight that the RFT has been adopted in biomechanics in analyzing one dimension continuum data, but the FDR has not been used earlier. Therefore, the study's authors considered both RFT and FDR as substitute approaches for conducting different tests in evaluating one-dimension biomechanical data. The research included an examination of how the RFT and FDR could impact biomechanical interpretations. Besides that, it involved a reanalysis of multiple publicly accessible experimental datasets in helping to understand the features that lead to the divergence and convergence of the results of the FDR and RFT. Furthermore, the study also comprised several numerical simulations containing smooth, random Gaussian one-dimension data to offer complimentary clarifications for the experimental outcomes.

**How RFD was used in the case study.**

The random field theory involves arithmetic used in defining theoretical outcomes for smooth statistical maps. In the case study, the random field theory was used as a substitute testing process to re-establish the balance between false-negatives and false-positives as it is able to consider correlation. According to Naouma & Pataky (2019), the benefit of using the RFT is that it appropriately handles smoothness in analyzing one-dimension data. Such data can be computed using field smoothness Full Width at Half Maximum (FWHM) along with sample-size.

In the case study, random field theory was used in stages, and the first one was to approximate the smoothness level. Different Simulations were carried out to evaluate the effects of full-width-at-half-maximum in the multiple one-dimension SNR datasets. As White, Van der Ende & Nichols (2019) highlighted, the second stage involves using the smoothness values in the correct RFT equation to calculate the projected Euler characteristic at various thresholds. The purpose of the step is to calculate the threshold at which researchers would expect five percent of the corresponding statistical map originating under the null hypothesis to comprise at least one region above the threshold. As the smoothness parameter increased, the thresholds of RFT decreased.

**What were the results of the FDR in the study?**

False Recovery Rate simply uses a single parameter which is the sample size, in a 1-D signal. In the case study, two simulations were conducted. The first one involved a 1-D signal, and the second one involved smooth 1-D noise. Using three parameters; Pulse Centre (q), Pulse Breadth, and amplitude (amp), the researchers manually adjusted the 1-D noise. In the study, the results of the FDR thresholds were qualitatively sensitive to 1-D signal changes. Therefore, these results indicate that the false discovery rate is sensitive to the one-dimension signal changes and a less restrictive process.

When it comes to the sample size, the false discovery rate thresholds decreased with an increase in SNR regardless of the sample size within five to fifty, which signifies small to moderate sample sizes. The FDR has been shown to maintain better statistical power, hence maximizing findings of the true signal. Besides that, it also regulates the proportion of falsely rejected null hypothesis, which according to Komiyama et al. (2017), is a quantity that symbolizes scientific relevance than various testing correction techniques. The results of the FDR in the case study, therefore, indicate that for researchers that aim to maximize their findings, the FDR is beneficial.

References

Komiyama J, Ishihata M, Arimura H, Nishibayashi T, Minato S. 2017. Statistical emerging pattern mining with multiple testing corrections. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining— KDD '17. New York: ACM Press, 897–906 DOI 10.1145/3097983.3098137.

Naouma, H., & Pataky, T. C. (2019). A comparison of random-field-theory and false-discovery-rate inference results in the analysis of registered one-dimensional biomechanical datasets. *PeerJ*, *7*, e8189. <https://peerj.com/articles/8189.pdf>

White T, Van der Ende J, Nichols TE. 2019. Beyond Bonferroni revisited: concerns over inflated false-positive research findings in the fields of conservation genetics, biology, and medicine. Conservation Genetics 20:927–937 DOI 10.1007/s10592-019-01178-0.