Clinical Integration and Downstream Analysis

Summary

This service encompasses the entirety of desired downstream analyses following generation of results from genomic, epigenomic, transcriptomic, proteomic, or metagenomic services. Leverage our expertise in study design and statistical methodology to provide quality hypothesis development and testing, sample size calculations, data analysis and visualizations, and actionable interpretations.

*Study design, research question and hypothesis development*

This portion of the service includes generation or review of your study design, research question, and hypothesis. We will discuss your population of interest, outcome of interest, and the clinical significance of the work.

In addition to more traditional techniques (here including linear and logistic regression tests, t-tests, and ANOVA tests) we will evaluate the suitability of machine learning techniques for your research question and hypothesis. These techniques could include dimension reduction techniques (ex: Linear Discriminant Analysis or Principal Component Analysis) regression techniques (ex: Random Forest or Decision Trees,), clustering techniques (ex: k-means or k-nearest neighbors), or classification techniques (ex: Neural Networks or Support Vector Machines)

Deliverables:

Word Document containing:

* Research Question
* Hypothesis
* Population of Interest
* Sampling Frame & Strategy
* Agreed upon analytical methodologies and visualizations
* Example interpretations of results

Requirements:

* outcome of interest
* population of interest
* clinical context of results

*Power and sample size calculations*

This portion of the service includes a power and sample size calculation based on the outline generated from the *Study design, research question and hypothesis development* service portion.

Please note: If required sample size based on the *Power and sample size calculations* deliverable is not achieved, any results and associated confidence intervals or p-values generated from analyses with insufficient sample sizes will not be accurate to the desired level of confidence.

Excel sheet containing:

* Statistical Power calculation results
* Required sample size (n) for the outcome of interest in the population of interest with the acceptable and desired levels of confidence
* Complete study design
* 1-2 publications from your field which use the analysis you are interested in for reference
* Desired and acceptable levels of confidence

Requirements: Deliverables:

*Clinical Data Integration Analysis*

Building upon the *Study design, research question and hypothesis development* service deliverables The Core will carry out the agreed upon statistical tests and having met our requirements of the *Power and sample size calculations* deliverableswe can be confident that our analyses includes data from enough sampling units (ex: participants) to test your hypothesis and that the study is appropriately powered to assess that hypothesis within the agreed upon level of confidence.

Requirements: Deliverables:

* Cleaned data (Excel spreadsheet)
* Descriptive analysis and hypothesis-specific analysis with associated output (PDF)
* Agreed upon data visualizations (Power point presentation)
* Raw clinical, gene expression, proteomic, or microbiome data (Excel spreadsheet)
* 1-2 examples of desired analysis and associated data visualizations for reference

Optional service portion: *Methods and Results sections of manuscript*

This portion of the service includes all described deliverables above, description of methodology, and interpretation of the analysis for your Methods and Results sections of a manuscript. A condition of production of this deliverable is co-authorship on the manuscript produced.

If you have additional requirements or questions, please feel free to contact us at EICC@emory.edu

References

American Statistical Association. (2018). Ethical Guidelines for Statistical Practice. Retrieved June 17, 2019, from https://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx

American Statistical Association. (2019a). ASA ISSUES STATEMENT ON ROLE OF STATISTICS IN DATA SCIENCE. Retrieved from https://www.amstat.org/asa/files/pdfs/POL-ASAStatementonRoleofStatsinDataScience.pdf

McDonald, J. H. (2014). Introduction - Handbook of Biological Statistics. Retrieved June 17, 2019, from Handbook website: http://www.biostathandbook.com/index.html

Miliken, George A., Johnson, D. E. (2009). Analysis of messy data. Volume 1: Designed experiments.

Raschka, S. (2015). Python machine learning : unlock deeper insights into machine learning with this vital guide to cutting-edge predictive analytics.

Waskom, M. (2018). An introduction to seaborn — seaborn 0.9.0 documentation. Retrieved April 14, 2019, from https://seaborn.pydata.org/introduction.html

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA’s Statement on p-Values: Context, Process, and Purpose. The American Statistician, 70(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Weiss, N. A. (2016). Introductory Statistics, 9th Edition.