SAS for Genomics

Polygenic Risk Scoring – Bioinformatics for Genomic Health Risks

SAS Facts

SAS is ranked as a leader in advanced analytics and artificial intelligence.

SAS software is open, cloud-based, unified and powerful.

SAS has customers in 147 countries

SAS software is installed at more than 83,000 customer sites.

Genomics: increasing value for Life Sciences and Health Care

Genes in our genetic code are not the only valuable information across our genome. One percent of the code translates into the proteins that make us human. However, the other 99% has deeply valuable information.

Polygenic Risk Scoring (PRS) and related approaches are gaining prominence in identifying the right patients for the right treatments. This is one aspect of algorithmic approaches used to find unique patterns in the code – a form of bioinformatics. The value is that variation can be linked to diseases. The greater the association, the more confidence in matching patients to treatments. The holy grail is combining genomics with all of the other data types available from the patient – multiple 'Omics analytics.

Challenges

PRS and other genomic/genetic approaches are not a settled data science, but what is clear is that more data (Big Data) is necessary.

- **Big Data.** To find signals of the genetic code's influence on disease progression requires more data, the rarer the disease; sometimes 10's to 100's of thousands of patient genomes.
- **Multi-Software.** Bioinformatics typically requires a universe of many software(s) – some to perform quality control, some for sequence variation summation, and some for visualization. What if the entire process could be executed at scale in one software with governance?
- **Code-level-only access by a select few.** Much of modern, scaled bioinformatics is performed with "R" and Python coding languages. Insights and decisions at the citizen scientist level move through a tortured path to get there.
- **Accessible Advanced Analytics.** Access to advanced models, such as multivariate statistic models, is limited. And, plugging those models into new data as it becomes available is difficult for citizen scientists and decision makers.
- **Identifying the best analytic approach.** Comparing and contrasting advanced analytic models is inconvenient and sometimes not possible in the same GUI environment. How to measure the accuracy and precision so that the best can be deployed?
- **Control, Control, Control.** Keeping track of data flows and applications of statistics/analytics/reports is difficult when it is managed by different platforms. This is a liability when processing pre-clinical and clinical data.

The full genetic code is yet to be fully understood, but it is increasing in the depth of the insights that can be realized.

Our Approach

SAS helps organizations use large and diverse volumes of historical data to set up and train statistical and deep learning/machine learning models. We also blend third party, open source, and SAS-constructed statistical routines so that it is all controlled in one GUI interface and modeling platform; your choice.

We approach the problem by providing software and services to help you:

- **Access all relevant data**. Quickly access and prepare relevant research and patient data for modeling, simulation, and insight generation.
- **Predict the expected error explanation by 'Omics data.** Identify cohorts of patients that demonstrate genetic patterns associated with specific indications.
- **Combine all the data.** Get the most out of data by combining genetics and proteomics, for example. We know the whole of this data is greater than the sum of the parts. We can do this.
- **Take advantage of the open-source community AND the power of SAS.** SAS is open to all.
- **Governance.** We "track and trace" all the data and applications. This is not only valuable for auditing, but also so that the organization doesn't leave IP orphaned on individual machines or in the hands of individuals.

Business Impact

As genetic data, as well as quantitative proteomic, metabolomic, etc., has become cheaper to produce, there has been a corresponding increase in the rate of data accumulation. All parties in health care and life sciences recognize this challenge. The questions now revolve around realizing the "greater than the sum of the parts" challenge with enterprise analytics.

SAS can help by providing:

- The means of ingesting **all of the data** collected through research, in clinical settings and the rest from the real world, not just genomics.
- Data management and data quality tools to prepare data critical to generating insights at **modern speed**.
- Insightful and interactive **visualizations** of the outcomes from deep mathematical model approaches
- Predictive scenario analysis, unique 'omics risk scoring models, and all the **statistics** that have made SAS famous
- Automatic, large-scale model runs that enable time series and predictive modeling that **improve automatically** as new data is ingested.
- A solution that can use your institutional knowledge and models from wherever they were developed, allowing **automated learning** to ensure the best models are implemented.