TAEHOOON HA

(919) 308-3505 | tah4002@med.cornell.edu | taehoonh.me | taehoonh | taehoonha

Education

Weill Cornell Medicine — New York, NY
M.S., Biostatistics — 2018 – Present

- Advisor: Xi Kathy Zhou, PhD

Duke University — Durham, NC
M.S., Business Analytics — 2017 – 2018

- Capstone project: Duke University Hospital (Duke Health)

Sungkyunkwan University — Seoul, Korea
B.B.A. with an emphasis on Quantitative Methods — 2009 – 2017

- Dean's list with distinction
- Study-abroad: School of Arts and Sciences at the University of Pennsylvania (2014)
- Military service: Log analyst at Radar Operation and Computer Emergency Readiness Team (CERT), Republic of Korea Marine Corps (2011-2013)

Independent Math Coursework — Online

- Intended (Spring 2020):
  - Differential Equations (University of North Dakota)
  - Real Analysis (University of Illinois at Urbana-Champaign)
- In progress (Fall 2019): Multivariable Calculus (University of North Dakota)
- Completed:
  - Calculus I, Calculus II, and Principles of Statistics (Brigham Young University)
  - Linear Algebra (University of North Dakota)

Research Experience

Johns Hopkins Bloomberg School of Public Health — Baltimore, MD (Remote)
Voluntary Researcher (Advisor: Bongsoo Park, PhD) — 04/2019 – Present

Transcriptome and epigenome atlas for air pollution $PM_{2.5}$

- Construct data pipeline of both EdgeR and DESeq2 to identify differentially expressed genes associated with the exposure to ultra-fine dust, $PM_{2.5}$
- Check the distribution by Relative Log Expression (RLE) and Principal Component Analysis (PCA) plots
- Normalize each sample and removed unwanted variances and filtered top 5,000 genes by p-values from EdgeR and DESeq2
- Perform DE (Differential Expression) analysis using EdgeR and Gene Ontology (GO) Term analysis on six different brain sub-areas and liver cells.
- Conduct pathway analysis using Ingenuity Pathway Analysis (IPA) software
- Prepare for publication "*Brain transcriptome map of air pollution $PM_{2.5}$*"

Weill Cornell Medicine — New York, NY
Research Assistant (Advisor: Xi Kathy Zhou, PhD) — 04/2019 – Present

Methodology: Bayesian model averaging

- Assist the advisor with developing a new statistical method using Bayesian model averaging to identify differentially expressed genes associated with one or more phenotypes, as well as their interactions.
- Develop R package '*BMAseq*' using Bayesian model averaging to analyze observational gene-expression data
- Apply the Bayesian model averaging method to large scale public RNA-seq gene expression data
- Apply the Bayesian model averaging method ("BMA-seq") to observational metabolomics data to improve differentially expressed (DE) metabolites identification in high dimensional setting
- Prepare for publication "*Bayesian model averaging approach for RNA-seq counts data (BMA-seq) and its application"*

Application: Collaboration with Andrew J. Dannenberg, MD group

- Provide statistical consulting support to clinical and lab research projects using R
- Collaborate with Andrew J. Dannenberg, MD group to identify the link along the obesity→inflammation→Aromatase pathway in excess adipose tissue of mice and humans
- Update and customize R package 'BTKR' which includes multiple functions that implement some commonly used biostatistics analysis methods for a simple summary of data
- Prepare for publication "*Body composition is a determinant of breast adipose inflammation and Aromatase levels, including in normal-sized women*"

JB Lab & Clinic — Seoul, Korea (Remote)
Research Scientist — 08/2018 – Present

- Identified the association of sodium intake and hypertension, metabolic syndrome, and ARB treatment effect from a 10K+ hypertension patients data acquired from a K-MetS study

Weill Cornell Medicine — New York, NY
Leading Project Researcher

The Effect of Right Heart Catheterization (RHC) During the First 24 Hours on 30-day Mortality of Critically Ill Patients in Intensive Care Units — 05/2019 – 08/2019
[Taehoon Ha, Jingjing Qi]

- Applied data adaptation method to estimate the propensity score
- Generated causal model estimands using multiple methods (random forest, GLM, elastic net, LASSO, and gradient boosting with SuperLearner)
- Conducted 5-fold cross-validation to evaluate each estimand's performance and coefficient in the ensemble estimator

Rate of General Anesthesia Use for Cesarean Delivery Among Anesthesiologists with and without Fellowship Training in Obstetric Anesthesia — 01/2019 – 05/2019
[Taehoon Ha, Yu Cai, Siyang Pei, Xuewei Quan]

- Fitted the logistic regression model to identify whether the fellowship-trained anesthesiologists are more or less likely to provide general anesthesia for non-routine C-section as compared to non-fellowship trained anesthesiologists
- Identified the significant difference between general and neuraxial anesthesia and the odds of using general anesthesia is 0.471 times lower in attending fellowship training anesthesiologists compared the non-fellowship training anesthesiologists

Cost-Effective Optimization of Model-Based Prediction of Cardiovascular Disease (CVD) *[Taehoon Ha]* 03/2019 – 05/2019

- Aimed to increase the accuracy of CVD diagnosis using a model-based approach
- Determined the demographic factors and medical tests that help predict the likelihood of heart disease using proportional odds model and logistics regression
- Identified a combination of necessary medical tests that help predict the probability of heart disease in a cost-effective manner and reduced the exam cost by $110.17

Neuroendocrine Prostate Cancer (NEPC) *[Taehoon Ha, Diane Li, Lingchen Lou]* 11/2018 – 12/2018

- Identified that there is no difference in the proportions of 3-month progression status between patients in the NEPC and non-NEPC group treated with alisertib
- Determined that there is no association between 3-month progression-free survival for all clinical characteristics and gene abnormalities
- Identified an association between PSA level and prior systemic therapies between NEPC and non-NEPC groups treated with alisertib

Major Risk Factors of Low Birth Weight Babies *[Taehoon Ha]* 10/2018 – 12/2018

- Determined that low birth weight is associated with the mother's premature labor history, race, smoking status, weight of the last menstrual period, and history of hypertension
- Fitted multiple generalized linear models with the model selection based on deviance and p-values

Study design of Tommy John Surgery for MLB Pitchers by *Statcast* Measurements *[Taehoon Ha]* 09/2018 – 10/2018

- Identified the primary injury factors of Ulnar Collateral Ligament Reconstruction for Major League Baseball pitchers by Statcast measures
- Designed a study to identify major factors of Ulnar Collateral Ligament Reconstruction (a.k.a. Tommy John Surgery) for Major League Baseball Pitchers by Statcast measurements

Policrew Research Group 06/2015 – 06/2017

Co-founder and Head Research Scientist (Data)

Leadership

- Managed all research projects, earning a reputation for excellent understanding of trends and insight of youth
- Developed strong and long-term relationships with donors and Partner Research Organizations, beyond the scope of engagement

Problem-solving

- Conducted network analysis to publish annual youth employment and health trend report
- Defined and evaluated the key performance indicators of the Korean government's current youth unemployment rate and healthcare policies

Teaching

Design and Analysis of Biomedical Studies (Master's level) Spring 2020

Teaching Associate for Prof. Xi Kathy Zhou, Weill Cornell Medicine

- Lead lab sessions for 21 Master's candidate students to teach biostatistics methods with R
- Review and grade weekly homework and provide guidance on lab assignments
- Hold regular office hours regarding questions on course materials, assignments, and academic concerns

Categorical and Censored Data Analysis (Master's level) Fall 2019

Teaching Associate for Prof. Oleksandr Savenkov, Weill Cornell Medicine

- Lead lab sessions for 32 Master's candidate students to teach biostatistics methods with R
- Review and grade weekly homework and provide guidance on lab assignments
- Hold regular office hours regarding questions on course materials, assignments, and academic concerns

Publications

In Preparation

- Hanhan Wang, Lingsong Meng, **Taehoon Ha**, Xi Kathy Zhou. *A Bayesian model averaging approach for RNA-seq counts data (BMA-seq) and its application*

- B Park, S Kim, **Taehoon Ha**, JE Park, V Vinayachandran, KD Hansen, B Paul, S Rajagopalan, and S Biswal. *Brain transcriptome map of air pollution $PM_{2.5}$*

Awards

Academic Excellence (Over 4.0 cumulative GPA), *Weill Cornell Medicine* 07/2019

Exchange Student Scholarship at the University of Pennsylvania (Travel Funding), *Sungkyunkwan University* 03/2015

Academic Excellence Scholarship, *Sungkyunkwan University* 01/2015

College Scholarship (Academic Excellence), *Sungkyunkwan University* 03/2014

Pursuit of Excellence Scholarship, *Sungkyunkwan University* 01/2014

Pursuit of Excellence Scholarship, *Sungkyunkwan University* 01/2010

Pursuit of Excellence Scholarship, *Sungkyunkwan University* 07/2009

Invited Talks

How to organize, manage, share, collaborate and process bio/healthcare data, *Konkuk University Hospital* 12/2018

RNA-seq data analysis 101, *JB Lab & Clinic* 08/2018

How to collaborate with a statistician, data analyst, or data scientist, 06/2017
Sungkyunkwan University

Relevant Coursework

Weill Cornell Medicine

- Biostatistics I, Biostatistics II – Regression, Statistical Learning, Causal Inference, Foundations Biomedical Applications, Study Design, Categorical Censored Data Analysis, Statistical Programming, Data Management

Duke University

- Applied Probability (Math), Data Infrastructure (SQL), Decision Analytics and Modeling (R), Data Visualization (Tableau), Data Science for Business (Python), Pricing (R), Empirical Economic Analysis: Advanced Econometrics (R), Operations Analytics (R)

Sungkyunkwan University

- Computer Data Processing and Computer Word Processing (Java, SQL), Marketing Research Methodology (SAS), Consumer Behavior (SAS), Introduction to Social Science Research (SAS)

Independent Math Coursework

- Intended (Spring 2020):
  - Differential Equations (University of North Dakota)
  - Real Analysis (University of Illinois at Urbana-Champaign)
- In progress (Fall 2019): Multivariable Calculus (University of North Dakota)
- Completed:
  - Calculus I, Calculus II, and Principles of Statistics (Brigham Young University)
  - Linear Algebra (University of North Dakota)

Service

Prudential Foundation 2015
Project CD34: Stem Cell Donation Trend Visualization

- Integrated and constructed the stem cell donation trend databases in Korea
- Generated real-time dashboard to help the board's decision-making process

S-ONE, Sungkyunkwan University 2013 – 2014
Data Analysis & Visualization Education Session Chief

- Led lab sessions for 40 students to teach statistical methods and data visualization tools when analyzing social science data

Samsung Dream Class Program 2009 – 2010
High School Mathematics Tutor

- Taught high school and pre-college math to 20 low-income students

Technical Skills

- **Programming**: R, SAS, Python, Java

- **Database**: MySQL, SQL Server, Microsoft Access
- **Document**: LaTeX, Markdown
- **Cloud / Parallel computing**: Google Cloud Platform, Amazon Web Services, Microsoft Azure, Sun Grid Engine (SGE)
- **Visualization:** Tableau Software

Membership

- American Statistical Association (ASA)
- Korean Statistical Society (KSS)
- American Heart Association (AHA)
- American Association for the Advancement of Science (AAAS)