DAVID LIM

deelim@live.unc.edu · https://DavidKLim.github.io
Department of Biostatistics, University of North Carolina at Chapel Hill
135 Dauer Dr, Chapel Hill, NC, 27599

EDUCATION

Ph.D. Candidate, Biostatistics *2014 - Present*
Department of Biostatistics
University of North Carolina at Chapel Hill
Advisors: Dr. Naim U. Rashid, Dr. Joseph G. Ibrahim
Dissertation Committee: Dr. Michael I. Love, Dr. Di Wu, Dr. Wei Sun, Dr. Junier Oliva
Research areas: Missing data, Clustering, Bayesian deep learning, RNA-seq

B.S. Physics *2009 - 2013*
Department of Physics and Astronomy
University of California, Los Angeles

B.S. Applied Mathematics *2009 - 2013*
Department of Mathematics
University of California, Los Angeles

PROFESSIONAL EXPERIENCE

Graduate Research Assistant *2014-2018*
University of North Carolina at Chapel Hill
Supervisor: Dr. Bahjat F. Qaqish

Data Analyst Intern *2016-2018*
Syngenta AG, Durham, NC

HONORS AND AWARDS

Genomics and Cancer Training Grant recipient, 2017 - present

PROFESSIONAL MEMBERSHIPS

American Statistical Association (ASA), 2018 - present

SUBMITTED MANUSCRIPTS

1. **Lim, D.K.**, Rashid, N.U., Ibrahim, J.G. Model-based feature selection and clustering of RNA-seq data for unsupervised subtype discovery. *Annals of Applied Statistics (accepted).* (pre-print: https://biorxiv.org/cgi/content/short/2020.05.23.111799v1)
2. **Lim, D.K.**, Rashid, N.U., Ibrahim, J.G., Oliva, J. Handling Non-ignorably Missing Features in Electronic Health Records Data Using Importance-Weighted Autoencoders. *Journal of the American Statistical Association (under review).* Pre-print pending

WORKING MANUSCRIPTS

1. **Lim, D.K.**, Rashid, N.U., Ibrahim, J.G., Oliva, J. Deeply-Learned Generalized Linear Models with Missing Data.

ORAL PRESENTATIONS

1. Missing Data in Deep Learning, *Eastern North American Region (ENAR)*. Nashville, TN. March 2020.
2. FSCseq: Simultaneous Feature Selection and Clustering of RNA-Seq Data, *Joint Statistical Meeting (JSM)*. Denver, CO. August 2019.
3. Unsupervised Clustering and Variable Selection for RNA-seq Data, *Eastern North American Region (ENAR)*. Atlanta, GA. March 2018.

SOFTWARE

1. **FSCseq**: Simultaneous feature selection and clustering of RNA-seq gene expression count data. Can adjust for differences in sequencing depth, and for effects of confounders. Fitted model can be used for prediction. `https://github.com/DavidKLim/FSCseq`
2. **NIMIWAE (in preparation)**: Deeply-learned importance-weighted autoencoder that can properly handle up to Missing Not At Random (MNAR) data

SKILLS

Computing: R, SAS, Python, C++, SQL

Languages: English, Korean