

# 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

## 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 (under review)*. (bioRxiv site here)

## WORKING MANUSCRIPTS

---

1. **Lim, D.K.**, Rashid, N.U., Ibrahim, J.G., Oliva, J. Missing Data in Public Health Data using Variational Autoencoders.

## 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>

## SKILLS

---

Computing: R, SAS, Python, C++

Languages: English, Korean