

# SOM B. BOHORA

---

CONTACT 940 NE 13th Street, Suite 4938 405-215-6470  
INFORMATION Oklahoma City, OK 73117 som-bohora@ouhsc.edu

RESEARCH Data science, analytics, big data  
INTERESTS Statistical and machine learning  
Real-time decision systems  
Dynamic documentation and reporting  
Data visualization and graphics  
Management information Systems  
Bayesian statistical inference

EDUCATION **University of Oklahoma**, Oklahoma City, OK

**M.S., Biostatistics**, July 2014

- Thesis area: *Generalization of semi-parametric AUC regression model with discrete covariates*

**Louisiana State University**, Baton Rouge, LA

**M.Ap.Stat., Master of Applied Statistics**, August 2012

- Project area: *Non-linear quantile and mixed-effects models*

**Louisiana State University**, Baton Rouge, LA

**M.S., Forest Science**, May 2012

- Thesis area: *Multivariate analysis and time series in tree rings*

**Tribhuvan University**, Nepal

**B.S.**, February 2008

SOFTWARE DEVELOPED

- **sAUC**, R package to perform semi-parametric AUC regression(finished)
- **sAUCpy**, Python package to perform semi-parametric AUC regression (developing)
- **sAUC.jl**, Julia package to perform semi-parametric AUC regression (almost finished)

SCHOLARSHIP, AWARDS, & HONORS

- **FASD Travel Award** recipient for the 2016 Fetal Alcohol Spectrum Disorders Study Group at Research Society on Alcoholism (RSA) Scientific Meeting held in New Orleans, Louisiana June 25-29, 2016
- **Above and Beyond Award** recipient, Center on Child Abuse and Neglect on April 10, 2015 Department of Pediatrics, College of Medicine, University of Oklahoma Health Sciences Center
- **Above and Beyond Award** recipient, Center on Child Abuse and Neglect on December 5, 2014 Department of Pediatrics, College of Medicine, University of Oklahoma Health Sciences Center

- **Excellence in Pediatric Research Day Presentation Award** recipient on Pediatric Research Day held on May 16, 2014, Department of Pediatrics, College of Medicine, University of Oklahoma Health Sciences Center
- **RSA Travel Award** recipient for the 2014 Research Society on Alcoholism (RSA) Scientific Meeting held in Seattle, Washington June 21-25, 2014
- **RSA Student Merit Award** recipient for the 2014 Research Society on Alcoholism (RSA) Scientific Meeting to be held in Seattle, Washington June 21-25, 2014
- **Sooner Heritage Tuition Scholarship** recipient, University of Oklahoma Health Sciences Center, Oklahoma City, OK Fall 2013
- **Outstanding MS student award** recipient in recognition of high scholarship and outstanding achievement awarded by Gamma Sigma Delta chapter at Louisiana State University April 2012
- **Outstanding Graduate Student Merit Honor Roll Award** in recognition of high scholarship and outstanding achievement awarded by Gamma Sigma Delta chapter at Louisiana State University April 2012
- **Honored member of Phi Kappa Phi**, an academic honor society 2012
- **Honored member of Gamma Sigma Delta**, an academic honor society 2011

EMPLOYMENT	<b>Biostatistician</b>	10/2014 - present
EXPERIENCE	Department of Pediatrics, University of Oklahoma Health Sciences Center Oklahoma City, OK	
	<b>Certified Data Scientist</b>	12/2015 - present
	Data Science Central	
	<b>Graduate Research Assistant</b>	08/2012 - 07/2014
	University of Oklahoma Health Sciences Center Oklahoma City, OK	
	<b>Graduate Teaching Assistant</b>	01/2012 - 08/2012
	Department of Experimental Statistics Louisiana State University Baton Rouge, LA	
	<b>Graduate Research Assistant</b>	08/2008 - 12/2011
	School of Renewable Natural Resources Louisiana State University Baton Rouge, LA	

PROGRAMMING R (**more than 6 years of experience**)

- CERTIFICATION
- R Programming by Johns Hopkins University on Coursera . Certificate earned on February 1, 2015 with **Distinction**
  - Enrolled in online courses offered by "The Georgia R School"
  - Certificate of Participation in "Learn R In-Depth" presented by Dr. Paul Geissler from USGS
  - Enrolled in Data Science Specializations in [Coursera](#) that focuses on using R
  - Completed R Training Path for R Programmer and Data Analysts in [DataCamp](#)
    1. Introduction to R
    2. Intermediate R
  - Enrolled in [Statistical Learning](#) from Stanford University online OpenEdx program
  - Enrolled in [Big Data University](#) for big data analytics
  - Enrolled in MITx: 15.071x The Analytics Edge
- SAS*<sup>®</sup> (**about 6 years of experience**)
- Certified Base Programmer for *SAS*<sup>®</sup> 9
  - Applied Analytics Using *SAS*<sup>®</sup> Enterprise Miner

- PROGRAMMING & STATISTICAL SOFTWARE
- **R** 3.1.2 for Statistical computing (proficient, currently creating a package, sAUC)
  - Revolution R Enterprise for big data analytics (RevoR)
  - **Python** programming (currently writing a package sAUCpy)
  - **Julia** programming (currently writing a package sAUC.jl)
  - Git & GitHub, a local or remote version control software
  - C, C++, C#, Java, Perl, Scala
  - SQL & MySQL programming
  - *Mplus* for statistical modeling
  - L<sup>A</sup>T<sub>E</sub>X- A document preparation system
  - MATLAB
  - LyX - the document processor
  - *SAS*<sup>®</sup> 9.2 Programming (proficient)
  - *SAS*<sup>®</sup>/SQL; *SAS*<sup>®</sup> Macro Language; *SAS*<sup>®</sup>/IML, *SAS*<sup>®</sup>/STAT
  - SPSS (Statistical Packages for the Social Sciences)
  - JMP 8 for Statistical Analysis
  - STATISTICA, a statistics and analytics software package
  - *ArcGIS*<sup>®</sup> 9, *ArcMap*<sup>®</sup> version 9.3
  - CANOCO for Windows Version 4.5
  - PC-ORD version 5.31, software for multivariate statistical analysis
  - Citation Management: EndNote
  - MS-Office Package, Adobe Photoshop

- SOFTWARE FOR BAYESIAN STATISTICS
- WinBUGs, Stan (RStan, PyStan, Stan.jl)

IDE	<ul style="list-style-type: none"> <li>• RStudio, powerful and productive user interface for R</li> <li>• RCommander (knitr is embedded within it)</li> <li>• Visual Studio for <math>C\#</math>, VB.NET, <math>C^{++}</math></li> <li>• Visual Studio code</li> <li>• Eclipse IDE for Python, Java, C, <math>C\#</math>, <math>C^{++}</math>, R, and more</li> <li>• PyDev &amp; Pycharm IDEs for Python</li> <li>• CLion for <math>C^{++}</math></li> <li>• IntelliJ IDEA for Java</li> <li>• ReSharper, a productivity tool for Visual Studio</li> <li>• DataGrip, UI for managing databases (sqlite3, Sybase, DB2 etc.)</li> </ul>
BIG DATA	<ul style="list-style-type: none"> <li>• Apache Spark, Scala</li> <li>• sparklyr (R interface to Spark)</li> </ul>
DATABASES	<ul style="list-style-type: none"> <li>• SQL Server 2012</li> <li>• MySQL, SQLite3</li> <li>• MS Access</li> <li>• ETO, Qualtrics, Survey Monkey, D2L, Moodle</li> <li>• REDCap<sup>TM</sup> (Research Electronic Data Capture) &amp; REDCapR (a package to communicate with REDCap from R)</li> </ul>
WEB DEVELOPMENT	<ul style="list-style-type: none"> <li>• HTML5, CSS3, JavaScript, PHP, MySQL, Python</li> </ul>
PROFESSIONAL MEMBERSHIP	<ul style="list-style-type: none"> <li>• Member, Data Science Central</li> <li>• Institute of Mathematical Statistics (IMS) since 2011</li> <li>• Association of Nepalese Mathematicians in America (ANMA)</li> <li>• American Statistical Association (ASA) since 2010</li> <li>• Statistical Modelling Society (SMS) since 2010</li> <li>• Vice-president, OUHSC Biostatistics and Epidemiology Student Association (BSESA)</li> <li>• Statistical Computing Users' Group (SCUG), Oklahoma City, OK</li> </ul>
ORAL ENGLISH PROFICIENCY	<ul style="list-style-type: none"> <li>• Received a <b>Graduate Teaching Certificate</b> from Department of English, Louisiana State University (LSU) Summer 2010</li> </ul>
PUBLICATIONS & THESES	<ol style="list-style-type: none"> <li>1. <b>Bohora, S.</b> &amp; Cao, Q. (2014). Prediction of Tree Diameter Growth Using Quantile Regression and Mixed-Effects Models. Forest Ecology and Management, 319(0), 62-66.</li> <li>2. <b>Bohora, S.</b>; Zhao, Y.; &amp; Balachova, T. Generalization of the semi-parametric AUC regression model with discrete covariates. (Submitted)</li> <li>3. <b>Bohora, S.</b>; Balahcova, T; Shaboltas, A.; Nasledov, A.; Chaffin, M.; Batluk, Y.; Bonner, B.; Skitnevskaya, L.; Tsvetkova, L; Volkova, E. Analysis of HIV risk patterns among Russian women at-risk for AEP using Latent Class Analysis (in preparation)</li> </ol>

4. Balachova, T.; Sobell, L.; Agrawal, S.; Isurina, G.; Tsvetkova, L.; Volkova, E.; **Bohora, S.** 2015. Using a Single Binge Drinking Question to Identify Russian Women at Risk for an Alcohol-Exposed Pregnancy, Addictive Behaviors, 46(0), 53-57.
5. Balachova, T.; Shaboltas, A.; Nasledov, A; **Bohora, S.** Using Structural Equation Modeling to Analyze HIV Risk and Alcohol Abuse Behaviors in Russian Women (in preparation)

POSTER, Oral presentation

- ORAL TALK & MEETINGS
- **Bohora, S.**; Chaffin, M.; Balachova, T.; and Prevent FAS Group. **Using Multilevel Latent Transition Analysis to Model Transitions of Alcohol Use Profiles in Women at Risk for Alcohol-Exposed Pregnancies** presented at Fetal Alcohol Spectrum Disorders Study Group at Research Society on Alcoholism (RSA) Scientific Meeting held in New Orleans, Louisiana June 25-29, 2016
  - **Bohora, S.**; Zhao, D. **Extending Semiparametric AUC Model with Discrete Covariates into General Framework** presented in 'Statistical Computing Users' Group meeting, University of Oklahoma Health Sciences Center, Oklahoma City April 2015
  - **Bohora, S.** **Using knitr with logistic multilevel models and decision support trees** presented in 'Statistical Computing Users' Group meeting, University of Oklahoma Health Sciences Center, OK February 2015
  - **Bohora, S.** & Zhao, D. (presenter). **Generalization of the Semiparametric AUC Regression Model with Discrete Covariates** submitted to Joint Statistical Meetings 2015 organized in Seattle, WA August 2015
  - **Bohora, S.** & Zhao, D. (presenter). **General Semiparametric AUC Regression Model with Discrete Covariates** invited talk at the Department of Mathematical Sciences, University of Memphis, Memphis, TN April 2015
  - **Bohora, S.** **Generalization of the Semiparametric AUC Regression Model with Discrete Covariates** presented at the OUHSC College of Public Health, Bistatistics and Epidemiology Seminar Series, November 19, 2014
  - **Bohora, S.** **Dealing with simple 2x2 tables using SAS PROC FREQ**, presented in SAS/R Users' Group meeting, University of Oklahoma Health Sciences Center, Oklahoma City August 2013

Poster

- **Bohora, S.**, Balachova, T., Shaboltas, A., Nasledov, A., Chaffin, M., Batluk, Y., Bonner, B., Skitnevskaya, L., Tsvetkova, L., Volkova, E. "Analysis of HIV Risk Patterns among Russian Women At-risk for AEP Using Latent Class Analysis". A poster presented at the 38th

Annual RSA Scientific Meeting in San Antonio, TX. June 2015

- Balachova, T., Shaboltas, A., Nasledov, A., Chaffin, M., **Bohora, S.**, Batluk, Y., Bonner, B., Skitnevskaya, L., Tsvetkova, L., Volkova, E. "Using Structural Equation Modeling to Analyze HIV Risk and Alcohol Abuse Behaviors in Russian Women". A poster presented at the 38th Annual RSA Scientific Meeting in San Antonio, TX. June 2015
- **Bohora, S.**, Zhao, D., Balachova, T. **Extending Semiparametric AUC Model with Discrete Covariates into General Framework.** Poster accepted at The 4th Workshop on Biostatistics and Bioinformatics, Department of Mathematics and Statistics, Georgia State University, Atlanta, GA May 2015
- Bard, D. (presenter), Hecht, D., Beasley, W., **Bohora, S.**, & Peters, A. "Using Predictive Analytics for Systems-level Targeting of Child Welfare Prevention Services". A poster submitted to American Psychological Association Annual Convention in Toronto, Canada August 2015

#### Meetings/workshops attended

- One-day workshop on "Structural Equation Modeling" organized at LSU, Baton Rouge, LA (led by Dr. James B. Grace) April 2010
- "Multivariate Data Analysis using PC-ORD<sup>TM</sup>" organized at USGS Utah water science center, Salt Lake City, Utah February 2010

#### ABSTRACTS

- **Bohora, S.**, Balachova, T., Bonner, B., Chaffin, M. "Formatting health messages to prevent fetal alcohol exposure and brain damage in children" August 2013

#### JOURNAL PAPERS REVIEW

##### Ad hoc Statistical Reviewer

- Participated in review of submitted manuscript for peer-reviewed 'BioMed Central Pregnancy and Childbirth' July 2014
- Participated in review of submitted manuscript for peer-reviewed 'Journal of European Child and Adolescent Psychiatry' August 2013
- Participated in review of submitted manuscript for peer-reviewed 'Journal of Alcoholism: Clinical and Experimental Research' April 2013
- Participated in review of submitted manuscript for peer-reviewed 'Scandinavian Journal of Psychology' May 2013
- Participated in review of submitted manuscript for peer-reviewed 'Journal of The International Journal of Alcohol and Drug Research' May 2013

BOOKS READ	Statistics & biostatistics <ul style="list-style-type: none"> <li>• one</li> </ul> Computer & programming <ul style="list-style-type: none"> <li>• one</li> </ul> General <ul style="list-style-type: none"> <li>• one</li> </ul>
------------	---

RELEVANT COURSES	Courses	Taken in
	BSE 6553 Theory of Linear Models I	
	STAT Time Series Analysis	OSU
	STAT Bayesian Statistics	OSU
	STAT Multivariate Methods	
	BSE 5153 Clinical Trials	OUHSC
	BSE 6151 Applied Statistical Methods for Clinical Trials	OUHSC
	Data Management for Clinical Research	Coursera 2013
	Design and interpretation of Clinical Trials	Coursera 2013
	BSE 6643 Survival Data Analysis	OUHSC
	BSE 6563 Longitudinal Data Analysis	OUHSC
	BSE 5363 Epidemiology of Chronic Disease	OUHSC
	BSE 5001 Problems in Biostatistics & Epidemiology	OUHSC
	BSE 5453 US Healthcare Systems	OUHSC
	BSE 5153 Biostatistics Methods II	OUHSC
	BSE 5733 Mathematical Statistics I	OUHSC
	BSE 5111 Scientific Integrity in Research	OUHSC
	BSE 5113 Principles of Epidemiology	OUHSC
	BSE 5163 Biostatistics Methods I	OUHSC
	BSE 5013 Microcomputer to Data Analysis	OUHSC
	BSE 5703 Theory of Probability	OUHSC
	EXST 7031 Experimental Design	LSU
	EXST 7061 Statistical Theory II	LSU
	EXST 7060 Probability and Statistics	LSU
	MATH 3355 Probability	LSU
	EXST 7011 Nonparametric Statistics	LSU
	MATH 2085 Linear Algebra	LSU
	MATH 2057 Multivariable Calculus	LSU
	EXST 7037 Multivariate Statistics	LSU
	EXST 7142 Statistical Data Mining	LSU
	EXST 7036 Categorical Data Analysis	LSU
	EXST 7015 Statistical Techniques II	LSU
	EXST 7005 Statistical Techniques I	LSU
	SFM 507 Experimental Design & Statistical Analysis	Nepal
	BSH 403 Statistics	Nepal
	BSH 361 (F) Mathematics for Forestry	Nepal
	BSH 324 (F) Mathematics for Forestry	Nepal
	BSH 104 Mathematics for Forestry (1 Year)	Nepal

**References**    Available upon request