

Arun Kaushik

Curriculum Vitae

Department of Statistics
Banaras Hindu University
☎ +918853392016
✉ arundevkaushik@gmail.com



"An honest and sincere approach towards success and that of my benefactors brought about my ability to work with team spirit. To become a high successful person in life and to do better in more better."

Doctoral Thesis

Title ***Bayesian Inference for Some Life-Time Models Under Recent and Modern Censoring Scheme***
Supervisor Prof. Umesh Singh (Co-Ordinator DST-CIMS, BHU, Varanasi)
3 years Research Experience

Research Interests

- Bayesian Inference
- Data Analysis
- Applied Statistics
- Statistical Software
- Multivariate Study
- Mathematical and Statistical Computation

Academic Qualification

2010–2012 **Masters of Science**, CCS University, Meerut.
Specialized in Statistics
2006–2009 **Bachelor of Science**, CCS University, Meerut.
PCM
Massive Online Open Courses are attached in Separate sheet , .

Grant and Fellowship

Nov–2012 **UGS-BHU Reserach Fellowship.**
Oct–2013 **Junior Reserach Fellowship**, (NET), CSIR.
Oct–2015 **Senior Reserach Fellowship**, (NET), CSIR.

Awards

Oct–2014 **Department of Statistics, Kumaun University, Nainital, India.**, Best paper presentation award.
Dec–2014 **Department of Statistics, Jammu University, Jammu, India.**, Best poster presentation award.

Membership of Profession Bodies

- Life time membership of **International Indian Statistics Association**

- Student membership of **International Society of Bayesian Analysis**
- Life time membership of **Indian chapter in ISBA**
- Ordinary member of **Indian Statistical Institute Council, Kolkata, WB, India**
- Life time membership of **Indian Mathematics Society**
- Life time membership of **Indian Bayesian Society**
- Life time membership of **Mathematical Society Banaras Hindu University**
- Student membership of **International Mathematical Society**
- Student membership of **Bernoulli Society**

Computer Skills

Operating System	Microsoft Windows, Server, Unix, Linux, Macintosh
Statistical Softwares	SPSS, STATA, Statistica, Mathematica, Matlab, Maple, \LaTeX
Languages	R, S^+ , C, C^{++} , PYTHON, HTML, WinBUGS and ForTran
Office Tools	Adobe Illustrator, OpenOffice and MS-Office
Basic	Networking and Computer Hardware

Communication Skills

Aug 2013	Oral Presentation at the Annual CORTMAS Conference, DDU University Gorakhpur
Nov 2013	Oral Presentation at the RASA Conference, Pt. Shree Ravi Shankar University, Raipur
Feb 2014	Oral Presentation at the RTMMS Conference, ACMS-BHU, Varanasi
Oct 2014	Oral Presentation at the RASMSA -2014 Conference, Kumaun University, Nainital
Nov 2014	Poster Presentation at the Indian Association for the Study of population Conference, Kerala
Dec 2014	Poster Presentation National Meet of Research Scholars in Mathematical Sciences, University of Jammu, Jammu and Kashmir
Oct 2015	Oral Presentation at the NSSICW-2015 Conference, Banaras Hindu University, Varanasi
Nov 2015	Oral Presentation at the SRAESD/SPS2015 Conference, Lucknow University, Lucknow
Dec 2015	Oral Presentation at the IISA Conference 2015, Pune University, Pune

Publication

- 2015 with U. Singh and S.K. Singh, Bayesian inference for the parameters of Weibull distribution under progressive Type-I interval censored data with beta-binomial removals in Communication in Statistics: Simulation and Computation(2015), DOI:10.1080/03610918.2015.1076469

Languages

Hindi	Mother tongue	
English	Intermediate	<i>Conversationally fluent</i>

Personal Details

Nationality	Indian	Religion	Hindu
Date of Birth	March 6 th , 1990	Age	25 years
Gender	Male	Marital Status	Single

Declaration

The information given above is true to the best of my knowledge.

List of Massive Open Online Courses (MOOCs)

1. R Programming

Tutor: Roger D. Peng, PhD, Jeff Leek, PhD & Brian Caffo, PhD

University : Johns Hopkins University at Coursera

Certificate Verification Link:

<https://www.coursera.org/account/accomplishments/certificate/KEHCTNU5SX>

Course Description

This course contains how to program in R and how to use R for effective data analysis, how to install and configure software necessary for a statistical programming environment and describe generic programming language concepts as they are implemented in a high-level statistical language. The course covers practical issues in statistical computing which includes programming in R, reading data into R, accessing R packages, writing R functions, debugging, profiling R code, and organizing and commenting R code. Topics in statistical data analysis will provide working examples.

2. Getting and Cleaning Data

Tutor: Roger D. Peng, PhD, Jeff Leek, PhD & Brian Caffo, PhD

University : Johns Hopkins University at Coursera

Certificate Verification Link:

<https://www.coursera.org/account/accomplishments/certificate/NN5VD6RNHB>

Course Description

This course will cover the basic ways that data can be obtained, obtaining data from the web, from APIs, from databases and from colleagues in various formats. It will also cover the basics of data cleaning and how to make data “tidy”. Tidy data dramatically speed downstream data analysis tasks. The course will also cover the components of a complete data set including raw data, processing instructions, codebooks, and processed data. The course also covers the basics needed for collecting, cleaning, and sharing data.

3. Statistical Inference

Tutor: Roger D. Peng, PhD, Jeff Leek, PhD & Brian Caffo, PhD

University : Johns Hopkins University at Coursera

Certificate Verification Link:

<https://www.coursera.org/account/accomplishments/certificate/J7W3S8LL8L>

Course Description

Statistical inference is the process of drawing conclusions about populations or scientific truths from data. There are many modes of performing inference including statistical modeling, data oriented strategies and explicit use of designs and randomization in analyses. Furthermore, there are broad theories (frequentists, Bayesian, likelihood, design based, ...) and numerous complexities (missing data, observed and unobserved confounding, biases) for performing inference. A practitioner can often be left in a debilitating maze of techniques, philosophies and nuance. This course presents the fundamentals of inference in a practical approach for getting things done. After taking this course, students will understand the broad directions of statistical inference and use this information for making informed choices in analyzing data.

4.The Data Scientist's Toolbox

Tutor: Roger D. Peng, PhD, Jeff Leek, PhD & Brian Caffo, PhD

University : Johns Hopkins University at Coursera

Certificate Verification Link:

<https://www.coursera.org/account/accomplishments/certificate/7KLA3RZ964>

Course Description

This course you provide an introduction to the main tools and ideas in the data scientist's toolbox. The course gives an overview of the data, questions, and tools that data analysts and data scientists work with. There are two components to this course. The first is a conceptual introduction to the ideas behind turning data into actionable knowledge. The second is a practical introduction to the tools that will be used in the program like version control, markdown, git, GitHub, R, and RStudio.

5. Exploratory Data Analysis

Tutor: Roger D. Peng, PhD, Jeff Leek, PhD & Brian Caffo, PhD

University : Johns Hopkins University at Coursera

Certificate Verification Link:

<https://www.coursera.org/account/accomplishments/certificate/FFPXTK754T>

Course Description

This course covers the essential exploratory techniques for summarizing data. These techniques are typically applied before formal modeling commences and can help inform the development of more complex statistical models. Exploratory techniques are also important for eliminating or sharpening potential hypotheses about the world that can be addressed by the data. This course also cover in detail the plotting systems in R as well as

some of the basic principles of constructing data graphics, some of the common multivariate statistical techniques used to visualize high

h-dimensional data.

6. Case-Based Introduction to Biostatistics

Tutor: Roger D. Peng, PhD, Jeff Leek, PhD & Brian Caffo, PhD

University : Johns Hopkins University at Coursera

Certificate Verification Link:

<https://www.coursera.org/account/accomplishments/certificate/396SUKEZSN>

Course Description

The course objective is to enable each student to enhance his or her quantitative scientific reasoning about problems related to human health. Biostatistics is about quantitative approaches - ideas and skills - to address bioscience and health problems. This course reviews the scientific method and the role of experimentation and observation to generate data, or evidence, relevant to selecting among competing hypotheses about the natural world. Bayes theorem is used to quantify the concept of evidence. We used stratification and weighted averages to compare subgroups that are otherwise similar in an attempt to estimate the effects of smoking and smoking-caused diseases on medical expenditures. Finally, we study what factors influence child-survival in Nepal using data from the Nepal Nutritional Intervention Study Sarlahi or NNIPPS. We estimate and obtain confidence intervals for infant survival rates, relative rates and odds ratios within strata defined by gestational period, singleton vs twin births, and parental characteristics.

7. Reproducible Research

Tutor: Roger D. Peng, PhD, Jeff Leek, PhD & Brian Caffo, PhD

University : Johns Hopkins University at Coursera

Certificate Verification Link:

<https://www.coursera.org/account/accomplishments/certificate/E3KNC3XKXM>

Course Description

This course focuses on the concepts and tools behind reporting modern data analyses in a reproducible manner. Reproducible research is the idea that data analyses, and more generally, scientific claims, are published with their data and software code so that others may verify the findings and build upon them. The need for reproducibility is increasing dramatically as data analyses become more complex, involving larger datasets and more sophisticated computations. Reproducibility allows for people to focus on the actual content of a data analysis, rather than on superficial details reported in a written

summary. In addition, reproducibility makes an analysis more useful to others because the data and code that actually conducted the analysis are available.

8. Mathematical Biostatistics Boot Camp 1

Tutor: Brian Caffo, PhD

University : Johns Hopkins University at Coursera

Certificate Verification Link:

<https://www.coursera.org/account/accomplishments/certificate/6TTAXPVU88>

Course Description

Statistics is a thriving discipline that provides the fundamental language of all empirical research. Biostatistics is simply the field of statistics applied in the biomedical sciences. This course puts forward key mathematical and statistical topics to help students understand biostatistics at a deeper level. After completing this course, students will have a basic level of understanding of the goals, assumptions, benefits and negatives of probability modeling in the medical sciences. This understanding will be invaluable when approaching new statistical topics and will provide students with a framework and foundation for future self learning.

9. Experimentation for Improvement

Tutor: Kevin Dunn

University : McMaster University at Coursera

Certificate Verification Link:

<https://www.coursera.org/account/accomplishments/certificate/9YFAL3BRGP>

Course Description

It is clear: **better experiments save time and money, and lead to improvement.**

This course covers to use efficient factorial experiments, fractional factorials and response surface methods. If these terms sound intimidating, don't fear! We include a gentle introduction to basic statistical concepts, where required, throughout the course. End of this course you will be able to design your own experimental program, changing multiple variables, and interpret the experimental data using simple tools, based on sound statistical principles. And you will quickly see how to make those improvements. These tools and methods can be beneficial to solve the challenges you set for yourself above.

10. Regression Models

Tutor: Roger D. Peng, PhD, Jeff Leek, PhD & Brian Caffo, PhD

University : Johns Hopkins University at Coursera

Certificate Verification Link:

<https://www.coursera.org/account/accomplishments/certificate/S6LCHFBD76>

Course Description

Linear models, as their name implies, relates an outcome to a set of predictors of interest using linear assumptions. Regression models, a subset of linear models, are the most important statistical analysis tool in a data scientist's toolkit. This course covers regression analysis, least squares and inference using regression models. Special cases of the regression model, ANOVA and ANCOVA will be covered as well. Analysis of residuals and variability will be investigated. The course will cover modern thinking on model selection and novel uses of regression models including scatterplot smoothing.

11. Practical Machine Learning

Tutor: Roger D. Peng, PhD, Jeff Leek, PhD & Brian Caffo, PhD

University : Johns Hopkins University at Coursera

Certificate Verification Link:

<https://www.coursera.org/account/accomplishments/certificate/ZFX7J55ZHE>

Course Description

One of the most common tasks performed by data scientists and data analysts are prediction and machine learning. This course will cover the basic components of building and applying prediction functions with an emphasis on practical applications. The course will provide basic grounding in concepts such as training and tests sets, overfitting, and error rates. The course will also introduce a range of model based and algorithmic machine learning methods including regression, classification trees, Naive Bayes, and random forests. The course will cover the complete process of building prediction functions including data collection, feature creation, algorithms, and evaluation.

13. Developing Data Products

Tutor: Roger D. Peng, PhD, Jeff Leek, PhD & Brian Caffo, PhD

University : Johns Hopkins University at Coursera

Certificate Verification Link:

<https://www.coursera.org/account/accomplishments/certificate/2QXT3ZTG4R>

Course Description

A data product is the production output from a statistical analysis. Data products automate complex analysis tasks or use technology to expand the utility of a data informed model, algorithm or inference. This course covers the basics of creating data products using Shiny, R packages, and interactive graphics. The course will focus on the statistical fundamentals of creating a data product that can be used to tell a story about data to a mass audience.

14. Pattern Discovery in Data Mining

Tutor: Jiawei Han

University : University of Illinois at Urbana-Champaign at Coursera

Certificate Verification Link:

Course Description

Learn the **general concepts of data mining** along with basic methodologies and applications. Then dive into one subfield in data mining: **pattern discovery**. Learn in-depth concepts, methods, and applications of pattern discovery in data mining. We will also introduce methods for pattern-based classification and some interesting applications of pattern discovery. This course provides you the opportunity to learn skills and content to practice and engage in **scalable pattern discovery methods** on massive transactional data, discuss **pattern evaluation measures**, and study **methods for mining** diverse kinds of patterns, sequential patterns, and sub-graph patterns.

15. Text Retrieval and Search Engines

Tutor: ChengXiang Zhai

University : University of Illinois at Urbana-Champaign at Coursera

Certificate Verification Link:

<https://www.coursera.org/account/accomplishments/certificate/RRGPAVLH2J>

Course Description

Recent years have seen a dramatic growth of natural language text data, including web pages, news articles, scientific literature, emails, enterprise documents, and social media such as blog articles, forum posts, product reviews, and tweets. **Text data are unique** in that they are usually generated directly by humans rather than a computer system or sensors, and are thus especially valuable for discovering knowledge about people's

opinions and preferences, in addition to many other kinds of knowledge that we encode in text.

This course will cover search engine technologies, which play an important role in any data mining applications involving text data for two reasons. First, while the raw data may be large for any particular problem, it is often a relatively small subset of the data that are relevant, and a search engine is an **essential tool** for quickly discovering a small subset of **relevant text data** in a large text collection. Second, search engines are needed to help analysts **interpret any patterns discovered** in the data by allowing them to examine the relevant original text data to make sense of any discovered pattern. You will learn the basic **concepts, principles, and the major techniques in text retrieval**, which is the underlying science of search engines.

16. Data Analysis and Statistical Inference

Tutor: Dr. Mine Çetinkaya-Rundel

University : Duke University at Coursera

Certificate Verification Link:

<https://www.coursera.org/account/accomplishments/certificate/JL6H25RYZ2>

Course Description

The goals of this course are as follows:

- Recognize the importance of data collection, identify limitations in data collection methods, and determine how they affect the scope of inference.
- Use statistical software (R) to summarize data numerically and visually, and to perform data analysis.
- Have a conceptual understanding of the unified nature of statistical inference.
- Apply estimation and testing methods (confidence intervals and hypothesis tests) to analyze single variables and the relationship between two variables in order to understand natural phenomena and make data-based decisions.
- Model and investigate relationships between two or more variables within a regression framework.
- Interpret results correctly, effectively, and in context without relying on statistical jargon.
- Critique data-based claims and evaluate data-based decisions.
- Complete a research project that employs simple statistical inference and modeling techniques.

17. Programming for Everybody (Python)

Tutor: Charles Severance

University : University of Michigan at Coursera

Certificate Verification Link:

<https://www.coursera.org/account/accomplishments/certificate/B9UDWJQFVJ>

Course Description

This course is specifically designed to be a first programming course using the popular Python programming language. The pace of the course is designed to lead to mastery of each of the topics in the class. This course contain simple data analysis as the programming exercises through the course. Understanding how to process data is valuable for everyone regardless of your career. This course might kindle an interest in more advanced programming courses or courses in web design and development or just provide skills when you are faced with a bunch of data that you need to analyze. You can do the programming assignments for the class using a web browser or using your personal computer. All required software for the course is free.

18. Statistical Learning

Tutor: Trevor Hastie and Rob Tibshirani

University : Stanford University

Certificate Verification Link:

<https://verify.lagunita.stanford.edu/SOA/5afa49e30da54e4fa59018e4992ff985>

Course Description

This course is an introductory-level course in supervised learning, with a focus on regression and classification methods. The syllabus includes: linear and polynomial regression, logistic regression and linear discriminant analysis; cross-validation and the bootstrap, model selection and regularization methods (ridge and lasso); nonlinear models, splines and generalized additive models; tree-based methods, random forests and boosting; support-vector machines. Some unsupervised learning methods are discussed: principal components and clustering (k-means and hierarchical). Course focus on what we consider to be the important elements of modern data analysis. Computing is done in R.

19. Hadoop Fundamentals I

University : Big Data University with IBM

Course Description

The complexity of modern analytics needs is outstripping the available computing power of legacy systems. With its distributed processing, Hadoop can handle large volumes of structured and unstructured data more efficiently than the traditional enterprise data warehouse. This course is specifically designed to be a first introduction to Hadoop Fundamentals. The course covers the following topic:

- Hadoop (including common utilities, HDFS, MapReduce)
- Pig (programming / query language)
- Flume (data collection/aggregation)
- Hive (data summarization/querying)

20. Data Mining with Weka

Tutor: Prof. Ian H. Witten

University : The University of Waikato

Course Description

This course is specifically designed to learn about Data Mining and Big Data nowadays with weka software. Weka is a powerful, yet easy to use tool for machine learning and data mining. This course introduces you to practical data mining.

21. Machine Learning

Tutor: Prof. Andrew Ng

University : Stanford University

Course Description

This course is specifically designed to learn about Data Mining and Big Data nowadays with weka software. Weka is a powerful, yet easy to use tool for machine learning and data mining. This course introduces you to practical data mining.