

Areas of Interest	<ul style="list-style-type: none"> <li>Econometrics, Development Economics, Urban Studies, Health Economics, Labor Economics, AI, Machine Learning, NLP, Causal Inference, Historical Panel Data</li> </ul>		
Qualification Highlights	<ul style="list-style-type: none"> <li>Graduate in <b>Statistics with Development Economics, Econometrics, Real Analysis, Linear Models</b>, and <b>Advanced Time Series</b> courses. 5+ years of research experience specializing in <b>RCT design</b>, <b>data analysis</b>, <b>econometrics</b> and behavioral economics</li> <li>Proficient in <b>R, Python, Stata, SAS, C++</b>, <b>SQL, HTML, CSS, oTree</b>, and <b>MATLAB</b>.</li> <li>Experienced in <b>Data Mining, Visualization, Large Data Analysis Execution, Results Interpretation</b>, and <b>Report Writing</b>. Strong skills in <b>critical thinking, communication</b>, and <b>collaboration</b> with diverse backgrounds.</li> <li>Excellent <b>organizational skills</b> with the ability to <b>manage multiple projects</b>. Demonstrated <b>attention to detail</b> in all aspects of work. Capable of <b>working independently</b> and utilizing various <b>scripting languages</b> for data manipulation.</li> </ul>		
Education and Test Scores	<b>Degree</b>	<b>Institution</b>	<b>Year</b>
	B.Sc.(Hons) in Statistics	University of Calcutta	2015
	M.Sc. in Statistics	University of Calcutta	2017
	<b>GRE Score:</b> Quantitative: 170, Verbal: 158		
	<b>IIT JAM Rank:</b> All India Rank 86		
	<b>ISI (M.Stat) Rank:</b> All India Top 25		
Certificate Courses	<b>Course Name and Institution</b>		<b>Year</b>
	Managing Successful Field Research (MSFR), World Bank		2024
	Statistics and Mathematics MicroBachelors (Proctored Exam), LSE		2024
	DEDP MicroMasters, MIT		2024
	Macroeconometric Forecasting/Macroeconomics for Climate Change, IMF		2024

## Research and Publications

- R-package (DHSr):** R-package to facilitate developing regression summary statistics datasets from repeated nesetd regressions on large datasets (e.g., DHS) and a novel panel/spatial-panel econometric clustering tool (under CRAN review). [\[Github Documentation\]](#)
- The Estimates of Inter-Community Educational Participation in Indian Districts and Stochasticity of Wealth (Working Paper):** I address my previous limitations by developing district-level mixed-effects regression models to analyze educational participation in India as a compound decision problem. The study examines variations in gender, caste, and religion, explores inter-district spillover, and uses Markov Chain Analysis to assess education's impact on wealth transitions and pre-market discrimination within similarly skilled groups. [\[Draft Paper\]](#)

### Other Research:

- Multidimensional Poverty Analysis:** assessing poverty patterns among different demographic groups of India: (Published at Journal of Progressive Research in Social Sciences Vol 13) [\[JPRSS\]](#)
- Two Case Studies: Statistically Observable Data for LGBTQI+ Individuals in the Indian Subcontinent**
  - Non-parametric analysis on transgender individuals matched with their most similar male/female counterparts from the same households to assess the effect of being transgender on education and health deprivation
  - Comparing scores for GAD-7, PHQ-9, and GHQ-12 mental health indicators between LGBTQI+ students and a control group of heterosexual students of the same age from 28 different Indian cities, during the COVID-19 pandemic
- MSc Dissertations:**
  - Time Series and Regression Analysis of Measles Cases (1928-1972):** Applied ARIMA and GARCH models for prediction and used VAR and SEM models to capture instantaneous causality.
  - Randomized Response Technique for Multiple Categories:** Developed a generalized RRT for populations with multiple categories to reduce bias in survey responses.

## Academic Work Experience and Collaborations

Institution	Role	Project/Responsibilities
Cambridge/Kingston University PI: Prof. Jalal Siddiki	Research Assistant & Co-author	<ol style="list-style-type: none"> <li>We measure <b>home ownership segregation</b> represented by an <b>entropy index</b>.</li> <li>Research objective is to assess the home ownership segregation, its determinants, association with <b>generational mobility and convergence</b> using <b>historical and cross-sectional data</b> and data from the <b>European Central Bank</b>.</li> <li>I am performing the <b>empirical section - writing and coding (Stata)</b>.</li> </ol>

Indian Institute of Management, Ahmedabad PI: Prof. Jeevant Rampal	Associate/Programmer	<b>Project 1:</b> Helped in conducting a <b>lab-in-field</b> experiment to evaluate a <b>novel matching mechanism</b> for assigning Indian bureaucrats, and testing empirical versus theoretical findings on equilibrium and strategy-proof properties. Tools used: R, Python, HTML, oTree; Survey Population: UG students/job aspirants. <b>Project 2:</b> Assisted to create an ideal dataset for NGO Swadhaar FinAccess for social-network/peer-effect analysis among individuals near the poverty line in India.
University of Calcutta Supervisor: Prof. Sugata Sen Roy	Graduate Research Apprentice	<b>1.</b> Computed <b>Multidimensional Poverty Index</b> using <b>IPUMS DHS data</b> . <b>2.</b> Applied analytical tools to assess government policies, introducing neural networks and unsupervised methods. <b>3.</b> Adapted a <b>Cluster Analysis-based Missing Data imputation method</b> . <b>4.</b> Conducted literature reviews and academic writing for research publications. <b>5.</b> Performed literature reviews of public finance literature to assist senior fellows' research.

## Industry Work Experience

Organization	Designation and Duration	Job Deliverables
GSK/Cytel/Merck	Senior Data Analyst (May 2024 - Present)	<b>1.</b> Lead <b>Health Economist/Data Analyst</b> for a <b>Clinical Trial</b> to research efficacy of a new <b>respiratory disease medication</b> , optimizing <b>Carbon Footprint</b> reduction. <b>2.</b> Developing <b>R packages/Dashboard</b> for data analysis.
IQVIA Biotech	Statistical Programmer II (July 2022 – April 2024)	<b>1.</b> Developed and documented programs for <b>Phase 3 clinical trials</b> , adhering to <b>FDA guidelines</b> . <b>2.</b> Interacted with global <b>stakeholders</b> across the <b>EU, USA, and Japan</b> . <b>3.</b> Enhanced process <b>quality and efficiency</b> using <b>AI and ML</b> . <b>4.</b> Trained <b>group members</b> and <b>field-enumerators</b> on new processes and <b>Electronic Data Capture</b> tools. <b>5.</b> Managed multiple <b>projects simultaneously</b> . Awards: <b>Excellence in Data Innovation 2023</b> .
Clinipace Clinical Research	Senior Statistical Programmer (Sep 2021 – June 2022)	<b>1.</b> Lead <b>Statistician</b> for multiple <b>randomized controlled trials</b> . <b>2.</b> Designed and modified <b>Randomized Controlled Trial</b> . <b>3.</b> Created complex <b>statistical tables and graphs</b> . Utilized <b>MATLAB</b> for various analyses.
Novartis Healthcare Pvt. Ltd.	Statistical Programmer (Apr 2019 – June 2021)	<b>1.</b> Lead <b>SAS Statistical Programmer</b> for <b>clinical trials</b> . <b>2.</b> Designed and modified <b>clinical trials</b> . <b>3.</b> Collaborated globally on <b>data quality</b> . <b>4.</b> SME for <b>AI-based clinical programming</b> . <b>5.</b> Utilized scripting for <b>automation</b> . Awards: <b>Best SME 2020, IMPACT Performer 2019</b> .
Novartis Healthcare Pvt. Ltd.	Associate Statistical Programmer (Aug 2017 – Apr 2019)	<b>1.</b> Cleaned and analyzed <b>clinical trial datasets</b> using <b>SAS and R</b> . <b>2.</b> Ensured robustness in <b>P21 checks</b> and developed outputs. <b>3.</b> Organized <b>departmental activities</b> . <b>4.</b> Implemented knowledge in <b>clinical data standards</b> . <b>5.</b> Edited and finalized <b>clinical reports</b> .

- **Conference Volunteering:** Participated in the “Ninth International Triennial Calcutta Symposium on Probability and Statistics”, organized by the Department of Statistics, University of Calcutta.
- **Pre-College Academic Achievement:** Recipient of a State-level Scholarship for Excellence in Secondary Exam (Awarded by the Government of India to the top 1% of students).
- **PhD Thesis Assistance:**
  - Assisted Dr. Kunal Datta in his PhD thesis “Mixed-halide perovskite semiconductors for multijunction photovoltaics” – submitted to TU Eindhoven in The Netherlands.
  - Supported Dr. Saurav Ghosh (University of Calcutta) in his paper “Prevalence of Work-related Musculoskeletal Disorders in Traffic Police Personnel in Maintaining Transportation: A Cross-sectional Study” (submitted to the Journal of Transport and Health).

## Additional Courses

- Advanced Linear Algebra, UT Austin
- Stanford University CS229 (Machine Learning), CS230 (Deep Learning), CS231 (Deep Learning for Computer Vision)
- Microeconomic Theory I-IV (graduation level Microeconomics course 14.121, 14.122, 14.123, 14.124) - MIT
- Spatial Data Analysis course by Prof. Michael J. Pyrcz (University of Texas)

## Personal Information

**Language Proficiency:** English, Bengali, Hindi

**Base SAS: 9.4 Other Software Knowledge:** R, Python, MATLAB, ArcGIS, QGIS, STATA, MINITAB

I hereby declare that the above information is correct to the best of my knowledge.

Date: 19-Oct-2024

Place: Kolkata

**Arnab Samanta**