

Areas of Interest	<ul style="list-style-type: none"> Econometrics, Development Economics, Urban Studies, Health Economics, Labor Economics, AI, Machine Learning, NLP, Causal Inference, Historical Panel Data 		
Qualification Highlights	<ul style="list-style-type: none"> Graduate in Statistics with Development Economics, Econometrics, Real Analysis, Linear Models, and Advanced Time Series courses. 5+ years of research experience specializing in RCT design, data analysis, econometrics and behavioral economics Proficient in R, Python, Stata, SAS, C++, SQL, HTML, CSS, oTree, and MATLAB. Experienced in Data Mining, Visualization, Large Data Analysis Execution, Results Interpretation, and Report Writing. Strong skills in critical thinking, communication, and collaboration with diverse backgrounds. Excellent organizational skills with the ability to manage multiple projects. Demonstrated attention to detail in all aspects of work. Capable of working independently and utilizing various scripting languages for data manipulation. 		
Education and Test Scores	Degree	Institution	Year
	B.Sc.(Hons) in Statistics	University of Calcutta	2015
	M.Sc. in Statistics	University of Calcutta	2017
	GRE Score: Quantitative: 170, Verbal: 158		
	IIT JAM Rank: All India Rank 86		
Certificate Courses	Course Name and Institution		Year
	Managing Successful Field Research (MSFR), World Bank		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 [[CRAN Publication](#)] [[Github Documentation/Workflow](#)]
- Community Effect on Educational Participation in Contemporary India: A District Level Analysis(Working Paper):** I address the limitations in my previous work by developing district-level mixed-effects regression models to analyze educational participation in India. The study examines variations in effect of gender, caste, and religion across Indian districts, accounting for inter-district spillover, and uses Markov Chain Analysis to assess community effect on education, and role of wealth. Most importantly, it stresses the need to take regional variations into account. [[Link: 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	Position	Responsibilities
Cambridge/ Kingston University PI: Prof. Jalal Siddiki	Research Assistant & Co-author	1. We measure home ownership segregation within specific travel-to-work areas , where the level of diversity in home ownership is represented by an entropy index . 2. Research objective is to assess the influence of home ownership segregation on inter-generational upward mobility using historical and cross-sectional data from the European Central Bank . 3. I am supervising the empirical section - writing and coding (Stata) .
Indian Institute of Management, Ahmedabad PI: Prof. Jeevant Rampal and Prof. Shantanu Khanna	Programmer/Associate	Project 1: Designing a lab-in-field experiment to evaluate a novel matching mechanism for assigning Indian bureaucrats, comparing it to the current method, and testing empirical versus theoretical findings on equilibrium and strategy-proof properties. Tools used: R, Python, HTML, oTree; Survey Population: UG students/job aspirants. Project 2: 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	1. Computed Multidimensional Poverty Index using IPUMS DHS data . 2. Applied analytical tools to assess government policies, introducing neural networks and unsupervised methods. 3. Adapted a Cluster Analysis-based Missing Data imputation method . 4. Conducted literature reviews and academic writing for research publications. 5. Performed literature reviews of public finance literature to assist senior fellows' research.
---	------------------------------	---

Industry Work Experience

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

- **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