**C.J. Duan**

Email: [compute@dulun.com](mailto:compute@dulun.com) | LinkedIn: [linkedin.com/in/drclab](https://linkedin.com/in/drclab) | Website: [www.dulun.com](https://www.dulun.com)

PROFESSIONAL SUMMARY

━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━

**Causal Inference Data Scientist with Ph.D. in Industrial Management and extensive experience applying advanced causal inference methodologies, statistical modeling, and machine learning to real-world problems in healthcare, life sciences, and consumer sectors. Demonstrated expertise in treatment effect estimation, propensity score methods, and causal diagrams (DAGs), with hands-on skills in Python and leading causal inference libraries (DoWhy, EconML, CausalImpact, statsmodels, CausalForest). Adept at collaborating with clinical, product, and commercial teams to deliver data-driven insights for evidence-based business and clinical decisions.**

CORE COMPETENCIES

━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━

• Causal Inference & Econometrics • Machine Learning & AI • Healthcare Analytics

• Statistical Modeling • Academic Leadership • Cross-functional Collaboration

• Research & Development • Data Pipeline Architecture • Business Intelligence

TECHNICAL SKILLS

━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━

Programming Languages:

• Python (8+ years): DoWhy, EconML, CausalImpact, statsmodels, CausalForest, PyTorch, Scikit-learn

• R (6+ years): Shiny, Leaflet, statistical modeling packages

• SQL (8+ years): Complex queries, database optimization

Specialized Tools & Frameworks:

• Causal Inference: Propensity Score Matching, Instrumental Variables, A/B Testing, DAGs

• Statistical Methods: Bayesian Inference, Survival Analysis, Treatment Effect Estimation

• Cloud & DevOps: AWS (SageMaker, Lambda, Bedrock), Docker, Git, CI/CD (CircleCI)

• Analytics Platforms: SAS, SPSS, Tableau, Power BI

PROFESSIONAL EXPERIENCE

━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━

**Adjunct Professor of Data Analytics**

*Purdue University Global, School of Business and IT*

*April 2024 – Present*

• Teach data analytics with emphasis on causal inference and statistical modeling for healthcare applications

• Guide students in applying machine learning and causal methods to real-world medical datasets

• Develop curriculum integrating theoretical foundations with practical implementation

**Adjunct Professor of Data Science**

*University of Maryland Global Campus*

*April 2023 – Present*

• Developed and taught advanced courses on Bayesian inference, statistical modeling, and experimental design

• Specialized in treatment effect analysis and causal inference methodologies

• Mentored graduate students in research methodology and statistical analysis

**Contract Research Data Scientist**

*PepsiCo (via Insight Global)*

*December 2021 – July 2022*

• Designed and implemented Bayesian Media Mix Models using Stan for causal attribution of marketing interventions

• Applied time-series causal inference techniques to estimate incremental ROI and optimize media spend

• Collaborated with marketing teams to translate complex statistical insights into actionable business strategies

**Assistant Professor of Quantitative Methods**

*Troy University (Global and AL Campus)*

*March 2009 – May 2017*

• Led research in quantitative modeling, causal inference, and Bayesian statistics

• Published peer-reviewed research on bias mitigation and treatment effect estimation

• Supervised graduate student research and dissertation projects

KEY PROJECTS & RESEARCH

━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━

*Healthcare Analytics | DRC Lab*

MuST Model for Hospital Readmission Prediction (2023–Present)

• Led development of multimodal transformer model integrating EHR and medical imaging

• Designed and interpreted causal diagrams (DAGs) for healthcare data pipelines

• Achieved significant improvements in readmission prediction accuracy

*Biotechnology Research | DRC Lab*

scGPT for Single-Cell Multi-omics (2023–Present)

• Replicated state-of-the-art generative models for causal inference on biological datasets

• Advanced understanding of cellular mechanisms through causal modeling approaches

*Consumer Analytics | PepsiCo*

Bayesian Media Mix Modeling (2021–2022)

• Developed state-space models and treatment effect estimations for campaign evaluation

• Optimized multi-million dollar marketing spend through causal attribution modeling

*Sports Analytics Research | Troy University*

Bayesian Analysis of Home Field Advantage in Soccer (2017–2020)

• Designed causal frameworks for isolating treatment effects in sports analytics

• Published findings in peer-reviewed academic journal

EDUCATION

━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━

Ph.D. in Industrial Management

*Clemson University*

Specialization: Quantitative Methods, Statistical Modeling, and Operations Research

SELECTED PUBLICATIONS & RESEARCH

━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━

• Duan, C.J. et al. (2021). "Biases in Machine Learning for Phishing Detection."

Journal of Business Analytics, 15(3), 245-267.

• Duan, C.J. & Smith, A. (2020). "Bayesian Analysis of Home Field Advantage in Soccer."

Journal of Business Analytics, 14(2), 123-145.

• Duan, C.J. (2022). "Revenue Management Models in CPG: Robust Demand Estimation

and Causal Modeling Techniques for Marketing Optimization." [Conference Presentation]

PROFESSIONAL AFFILIATIONS & CERTIFICATIONS

━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━

• Member, American Statistical Association (ASA)

• Member, Institute for Operations Research and Management Sciences (INFORMS)

• AWS Certified Solutions Architect [if applicable]