Iris Ivy Gauran

Statistician · PhD in Statistics · Open to my next role

irisivyg@gmail.com · statistical.systems · github.com/gauranii · Full CV

Summary

I build the tests that decide which discoveries are real. PhD in Statistics (UMBC), former Associate Professor (University of the Philippines), postdoctoral fellow at KAUST. Methods published in *Biometrics* and *The Annals of Applied Statistics*; 15 peer-reviewed papers. In 2026, eight public replications of clinical trials, field experiments and A/B tests, each rebuilt from raw data and matched value by value in Python, R and Lean.

Projects: Replications, 2026

Each one is rebuilt from the raw data, checked against the original's own numbers, then extended. Write-ups at statistical.systems/projects.

- **AncovaThreeWays** (Python, R, Lean). The primary-result table of a public Alzheimer's trial submission (CDISC pilot), rebuilt and proved value by value; 70 of 70 values match the 2006 SAS output. Refitted with MMRM, the way a trial would handle dropouts today.
- **ColonAdjuvant** (Python, R, Lean). The trial behind the 1990 approval of levamisole plus 5-FU for colon cancer; 74 of 74 quantities agree across languages, with the verdict proved in exact arithmetic and the mismatches with the paper left in view.
- **PlaceboVsObserved** (Python, R, Lean). The 1954 Salk polio trial's randomized and observed designs side by side; 1,064 of 1,068 printed values match, and the 9-point gap between the designs is taken apart.
- **HeadlineExpectedLoss** (Python, Lean). Thousands of Upworthy headline A/B tests replayed under the Bayesian expected-loss stopping rule: which version wins, when to stop, how much traffic it took. 511 checks, 0 mismatches.
- **NeighborsMailer** (Python). A large voter-turnout field experiment, matched to the authors' Stata log to the last printed digit (130 of 130), then turned into the yes-or-no a campaign needs.
- **BcgLatitude** (Python, R, Lean). The BCG vaccine meta-analysis, 164 of 164 published values matched, its latitude pattern turned into a vaccination decision.
- **GapYears** and **BloodVsMethylation** (R). The healthspan–lifespan gap rebuilt from live WHO and UN data (4,070 country-years); two measures of biological age that disagree in the same people.

Experience

Postdoctoral Research Fellow, Biostatistics Group, KAUST, Saudi Arabia 2021–2026

- Developed high-dimensional hypothesis tests, false discovery rate control and Bayesian methods, applied to imaging genetics, protein domains and neuroimaging.
- Built a cross-validation-based test of predictive performance for high-dimensional data (submitted to *JMLR*) and correlation-adjusted simultaneous testing for grouped data (submitted to *CSDA*).
- Taught probability and statistics to the Ministry of Interior's AI training program as Head Teaching Assistant, KAUST Academy (2023).
- Presented at JSM, ENAR, ICSDS and CMStatistics, and at seminars at UCSF and Seoul National University.

Lead Statistician, Philippine Genome Center and Lung Center of the Philippines 2019–2021

- SARS-CoV-2 biosurveillance through whole genome sequencing (Philippine Genome Center, 2021).
- Proteomics-based discovery of non-small cell lung cancer biomarkers and drug targets (Clinical Proteomics for Cancer Initiative, 2019–2021).

Project Leader, Philippine Statistical Research and Training Institute and Philippine Statistics Authority 2019

- Sustainable Development Goals research for a national disability registration framework.

Associate Professor, School of Statistics, University of the Philippines Diliman 2018–2021

- Taught parametric inference, experimental design, multiple hypothesis testing, Bayesian analytics and computational statistics.
- Centennial Professorial Chair, 2019 and 2020. Outstanding Teacher, 2018 and 2019.

Instructor and Graduate Teaching Assistant, University of Maryland, Baltimore County 2013–2018

Instructor and Assistant Professor, University of the Philippines Manila and Diliman 2009–2013

Selected Publications

- **Gauran, I.**, et al. (2022). Bayesian local false discovery rate for sparse count data with application to the discovery of hotspots in protein domains. *The Annals of Applied Statistics*, 16(3).
- **Gauran, I.**, et al. (2017). Empirical null estimation using zero-inflated discrete mixture distributions and its application to protein domain data. *Biometrics*, 74(2).
- **Gauran, I.**, et al. (2022). Ridge penalization in high-dimensional testing with applications to imaging genetics. *Frontiers in Neuroscience*, 16.
- Toonsi, S., **Gauran, I.**, et al. (2024). Causal relationships between diseases mined from the literature improve the use of polygenic risk scores. *Bioinformatics*, 40(11).
- Rafor, B., **Gauran, I.**, et al. (2025). Analysis of experiments with high frequency time series responses and the implications for power and sample size. *Scientific Reports*, 15(1).
- Albao, D., ..., **Gauran, I.**, et al. (2019). Methylation changes in the peripheral blood of Filipinos with type 2 diabetes. *Human Molecular Genetics*, 28(24).

15 peer-reviewed papers in all; the full list is in the CV.

Skills

Methods: multiple testing and false discovery rates, high-dimensional inference, Bayesian methods, cross-validation, experimental design, clinical trial analysis (ANCOVA, MMRM), survival analysis, meta-analysis, A/B testing.
Tools: R, Python, Lean (machine-checked proofs), Nix (reproducible environments), C++ with Rcpp, LaTeX and Typst.

Education and Honors

- **PhD, Statistics**, University of Maryland, Baltimore County, 2018. **MS** (2012) and **BS, cum laude** (2009), Statistics, University of the Philippines Diliman.
- Advised or co-advised five master's theses (2020–2023) at the University of the Philippines and KAUST; one won 1st Place, Best Graduate Student Paper, National Faculty-Student Conference on the Statistical Sciences.
- Jan Tinbergen Award for Outstanding Young Statistician, International Statistical Institute (2013). Outstanding Teacher, five times (2010–2019). Harvard Derek Bok Center Higher Education Teaching Certificate (2021).