AARON MAURER

1267 Page St, Apt F ◇ San Francisco, CA 94117
610 · 717 · 8858 ◇ ajmaurer89@gmail.com
www.github.com/ajmaurer

SUMMARY

I am a experienced statistician and data scientist with a zeal for answering hard questions with big data. After spending three years analyzing huge Medicare data sets at Acumen LLC, I went back to get my masters in statistics. Now I'm looking for a company that could use my help gaining insights and predictions from its data.

Specialties: Nonparametric, Bayesian, and Classical Statistics; Large Scale Data Mining; Machine Learning; Regression Analysis; Mathematical Optimization; Econometric Modeling

TECHNICAL SKILLS

Programing Languages	Python, Stata, SAS, R, SQL, Visual Basic, Some MATLAB
Programing Tools	Apache Spark, AWS EC2, MapReduce, LaTeX, Git, SVN

EDUCATION

Masters of Science in Statistics *Expected August 2015*
University of Chicago - Chicago, Illinois
Thesis: *Using Probabilistic Knockoffs of Binary Variables to Control the False Discovery Rate*
Selected Coursework: Machine Learning and Large-Scale Data Analysis, Bayesian Analysis, Nonparametric Inference, Convex Optimization, Mathematical Statistics, Generalized Linear Regression

Bachelor of Arts in Mathematics & History, Cum Laude *June 2011*
Carleton College - Northfield, Minnesota
Distinction in the Math Major, Sigma Xi, Varsity Football & Track

PROFESSIONAL EXPERIENCE

Acumen, LLC Burlingame, CA
Policy Associate *August 2011 - August 2014*

- Provided analytical and programing expertise to federal agencies studying healthcare topics.
- Cleaned huge Medicare data sets, complied statistics, and built statistical models in Stata and SAS to provide prediction or inference for these studies.
- Worked with federal clients to develop analysis and explicate results; produced data visualization and documentation to assist client understanding.
- Project work included:
  - ***Market and Enrollment Projections of the Affordable Care Act***
    - Developed microsimulation model to estimate ACA's effect on each individual.
    - Employed series of probit regressions on person level variables to predict future enrollment.
    - Forecast federal budget implications, guiding $25 billion in federal spending.
  - ***Flu Vaccine Comparative Effectiveness***
    - Studying vaccine effectiveness presented difficulty due to distinct populations receiving the vaccine, outcome mismeasurment, and the variable latent infection risk.

- Carefully designed comparable test and control cohorts from Medicare enrollees.
- Implemented and tested a number of regression models to account for remaining issues, including proportional hazard and measurement error models.

◇ ***Financial Projections of Policy Reform***

- Projected budget impact of Bipartisan Policy Center Medicare reform program.
- Developed a log-normal mixture model to simulate beneficiary level expenditures.
- Total net cost forecasts included in final report on the Domenici-Rivlin plan

◇ ***Active Surveillance of Flu Vaccine Safety***

- Tracked approximately 15 million yearly vaccinations in Medicare population for cases of Guillain-Barre syndrome.
- Employed sequential probability test to signal when hazard was above historical levels as quickly as possible.

OTHER SELECTED PROJECTS

- Used City of Chicago crime database to estimate weather's effect on citywide number of robberies.
  - ◇ Implemented Kernel and Poisson regressions on hourly crime counts across three years in R.
  - ◇ Models estimated change in robbery rate with temperature by hour and day of week.
- Employed Bayesian Latent Dirichlet Allocation model to categorize Wikipedia articles by topics.
  - ◇ Utilized python Onlineldavb package to download and parse 12,800 random articles on which to estimate topic posterior distributions.
  - ◇ Using topic weighting as a low dimensional representation of each article, fit k-medians clustering to group similar articles.
- Built L2 regularized logistic regression model to predict sentiment across 1.5 million tweets.
  - ◇ Parsed tweets into 4,000 word/emoticon feature vector in parallel using pyspark.
  - ◇ Fit model with stochastic gradient descent, choosing optimal parameters via cross validation.
- Created models to predict overall review score for RateBeer.com reviews based on description.
  - ◇ Built feature vector of tf-idf statistics of hashed words over 3 million review corpus.
  - ◇ Fit both random forest and gradient boosted trees using Spark MLLib on AWS EC2.

PUBLICATIONS

- W. Baird, A. Beveridge, A. Bonato, P. Codenotti, J. MacCauley, A. Maurer, S. Valeva, "On the minimal order of k-cop-win graphs", *Contributions to Discrete Mathematics*, Vol. 9, No. 1 (2014), pp. 1-15
- Hector Izurieta et al., "Comparative effectiveness of high-dose versus standard-dose influenza vaccines in US residents aged 65 years and older from 2012 to 2013 using Medicare data: a retrospective cohort analysis", *The Lancet*, Vol 15, No. 3 (2015), pp. 293-300

MISCELLANEOUS/PERSONAL

Erdos Number of 3; has or will soon visit all but 4 of continuous 48 states; expert on Byzantine military history; former member of International Brotherhood of Teamsters; avid board game player