Marta Karas

CONTACT INFORMATION

929 N Wolfe St Apt 607D
Baltimore, MD 21205
USA

Voice: +1 317-665-4551
E-mail: marta.karass@gmail.com
Web: https://martakarass.github.io/

EDUCATION

Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Ph.D., Biostatistics, Aug 2017–present

- Advisor: Ciprian Craniceanu

Wroclaw University of Science and Technology, Wroclaw, Poland

M.S., Mathematics (Mathematical Statistics), Jul 2015

- Dissertation Topic: "Theoretical and practical issues in change point detection."
- Advisor: Malgorzata Bogdan
- Final Grade: 5.5 (Excellent)

Wroclaw University of Science and Technology, Wroclaw, Poland

B.S., Mathematics, Jun 2013

- Final Grade: 5.0 (Very good)

HONORS AND AWARDS

Wroclaw University of Science and Technology: Graduation with Academic Distinction, 2015.

ACADEMIC EXPERIENCE

Indiana University Bloomington, Department of Epidemiology and Biostatistics, School of Public Health, Bloomington, IN, USA

Research Assistant **Jan 2017 - present**

Indiana University – Purdue University Indianapolis, Department of Biostatistics, Richard M. Fairbanks School of Public Health, Indianapolis, IN, USA

Research Assistant **Jan 2016 - Jul 2016**

- Developing extension of the existing graph-constrained regularization methods. Determining what imaging markers obtained from brain imaging data are associated with alcohol use disorder risk factors.
- Preparing materials in R and conducting a series of lectures on statistical methods for brain imaging data for participants from Indiana University Center for Neuroimaging.

PUBLICATIONS

Brzyski, D., **Karas, M.**, Ances, B., Dzemidzic, M., Goni, J., Randolph, T.W., Harezlak, J. Connectivity-Informed Adaptive Regularization for Generalized Outcomes (May 2018) (*preprint*)

Karas, M., Bai, J., Straczkiewicz, M., Harezlak, J., Glynn, N W., Harris, T., Zipunnikov, V., Crainiceanu, C., Urbanek, J.K. Accelerometry data in health research: challenges and opportunities. Review and examples (Submitted to *Statistics in Biosciences* Dec 2017.) (*preprint*)

Karas, M., Brzyski, D., Dzemidzic, M., Goni, J., Kareken, D.A., Randolph, T., Harezlak, J. Brain connectivity-informed regularization methods for regression. Stat Biosci, Nov 2017. (*link*)

CONFERENCE POSTERS

Harezlak, J., **Karas, M.**, Dzemidzic, M., Goni, J., Oberlin, B. G., Karaken, D. A. Association of gray matter imaging markers with alcoholism incorporating structural connectivity information: a regularized statistical approach. Neuroscience 2016, San Diego, CA, USA, Nov 2016.

Karas, M. Penalized regression inference regarding variable selection in high dimensions: presentation of selected methods implemented in R. European R Users Conference, Poznan, Poland, Oct 2016.

Karas, M., Dzemidzic, M. Goni, J., Karaken, D. A., Harezlak, J. Association of structural brain imaging markers with alcoholism incorporating structural connectivity information: a regularized statistical approach. IUPUI Research Day, Indianapolis, IN, USA, Apr 2016.

Oral Presentations

Karas, M., Harezlak, J., Straczkiewicz, M., Fadel, W., Crainiceanu, C., Urbanek, J.K. ADaptive Empirical Pattern Transformation (ADEPT) with application to walking stride segmentation. JSM 2018, Vancouver, Canada, Aug 2018.

Karas, M.. Wearable accelerometers, accelerometry data and automatic steps segmentation in R: strideter and convo R packages. Why R? 2018 Conference, Wroclaw, Poland, Jul 2018.

Karas, M.. Bayesian analysis with R and Stan: introduction and application to a business case problem. Wroclaw Data Science Meetup, Wroclaw, Poland, Nov 2016.

Karas, M.. Convex Clustering and Biclustering with application in R. Cracow R Users Meetup, Krakow, Poland, Sep 2016.

Software

`mdpeer` R package: Graph-Constrained Regression with Enhanced Regularization Parameters Selection. (*link*)

Industry Experience

Opera Software (Software), Wroclaw, Poland

Analyst **Aug 2016 - Dec 2016**

- Developing time series forecasting model for Opera browser's core metrics.
- Performing Bayesian analysis to describe and infer about browser's users performance.

Opera Software (Software), Wroclaw, Poland

Analyst **Jul 2015 - Dec 2015**

- Establishing methodology for A/B-test results analysis, including selection of statistical methods and implementation of a complete tool in R.
- Performing user base analysis for software product improvements.

Datarino (Big Data Services & Data Management Solutions), Wroclaw, Poland

Data Scientist **Jul 2014 - Mar 2015**

- Analyzing user activity and monetization KPIs of a Polish social networking service.
- Retrieving knowledge from business partners' big-data size data sets.

KRUK S.A. (Debt collection), Wroclaw, Poland

Intern **Jun 2014**

- Comparing feature selection methods, addressing serious correlation problem.
- Implementing and application of a text mining tool for utilizing unstructured data in R.

QuantUp (Data analysis, modeling and training), Wroclaw, Poland

Intern **Jul 2013 - Apr 2014**

- Performing research and case-study analysis of building and validating scoring models.
- Writing articles about reproducible research tools and large-size data analysis tools in R.

Computer Skills

- Languages / Tools: R, Python, Stan, SQL, Bash, Apache Spark.
- Operating Systems: Linux, OS X, Windows.

Competitions and Hackatons

- 2nd place in Data Analysis Marathon: "Determining tabloidization index". Krakow, Poland, Nov 2015.
- Participation in "HackZurich // The biggest European hackathon". Zurich, Switzerland, Oct 2015.
- The Winner of the Schneider Electric Business Analyst Competition 2015. Wroclaw, Poland, Jun 2015.
- 2nd place in the Independent National Data Analysis Competition "NOMAD" V edition 2015. Wroclaw, Poland, Jun 2015.
- The Winner of the Wroclaw Local Final of European BEST Engineering Competition 2015. Wroclaw, Poland, Apr 2015.
- 2nd place in the Independent National Data Analysis Competition "NOMAD" IV edition 2014. Wroclaw, Poland, Jun 2014.
- The Winner of the KRUK Analytics Challenge 2014. Wroclaw, Poland, May 2014.

Independent Coursework

- University of Washington: *Summer Institute in Statistics for Big Data (SISBID)*, Jul 11 - 27, 2016, Seattle, WA, USA:
  - Reproducible Research for Biomedical Big Data
  - Supervised Methods for Statistical Machine Learning
  - Unsupervised Methods for Statistical Machine Learning
- Johns Hopkins University & University of Colorado Boulder: *Principles of fMRI 1*, Coursera.
- Johns Hopkins University: *Genomic Data Science* Specialization, Coursera:
  - Introduction to Genomic Technologies
  - Genomic Data Science with Galaxy
  - Python for Genomic Data Science
- Stanford University: *Mining Massive Datasets*, Coursera.
- MIT: *Introduction to Computer Science and Programming Using Python*, edX.
- Johns Hopkins University: *Data Science* Specialization, 2014-2015, Coursera:
  - The Data Scientists Toolbox
  - R Programming
  - Getting and Cleaning Data
  - Exploratory Data Analysis
  - Reproducible Research
  - Statistical Inference
  - Regression Models