REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
05-02-2019	Final Report	30-Sep-2017 - 29-Sep-2018

4. TITLE AND SUBTITLE
Final Report: Interactive data analysis with statistical guarantees

5a. CONTRACT NUMBER
W911NF-17-1-0416

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Stanford University
3160 Porter Drive
Suite 100
Stanford, CA 94304 -8445

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
70940-MA.3

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Jonathan Taylor
UU	UU	UU	UU		19b. TELEPHONE NUMBER 650-725-8977

RPPR Final Report
as of 05-Feb-2019

Agency Code:

Proposal Number: 70940MA **Agreement Number: W911NF-17-1-0416**
INVESTIGATOR(S):

Name: Ph.D Jonathan E. Taylor
Email: jonathan.taylor@stanford.edu
Phone Number: 6507258977
Principal: Y

Organization: **Stanford University**
Address: 3160 Porter Drive, Stanford, CA 943048445
Country: USA
DUNS Number: 009214214 EIN: 941156365
Report Date: 29-Dec-2018 Date Received: 05-Feb-2019
Final Report for Period Beginning 30-Sep-2017 and Ending 29-Sep-2018
Title: Interactive data analysis with statistical guarantees
Begin Performance Period: 30-Sep-2017 **End Performance Period:** 29-Sep-2018
Report Term: 0-Other
Submitted By: Ph.D Jonathan Taylor Email: jonathan.taylor@stanford.edu
Phone: (650) 725-8977
Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 0 **STEM Participants:** 1

Major Goals: This project considered the problem of inference (with frequentist guarantees) when statistical models are not specified before data collection. We termed this type of data analysis "inferactive data analysis".

Two of the main goals of this project involved working out a framework for inference, the second being development of software usable in this context. In terms of the framework, questions include: under what kind of exploratory regimes is it possible to do inference; what are the appropriate corrections to the likelihood? For computations we considered specific randomization mechanisms that allowed tractable inference algorithms.

Accomplishments: We worked out a framework for inference after randomized exploration, specifically in the context of looking at the solutions to randomized convex programs. Given a sequence of such queries of the data, we derive an explicit representation of the selection probability that can be used for inference. This is described in the paper titled "Inferactive Data Analysis" in the publications section of this project.

When this randomization is Gaussian, inference is particularly tractable. Even Bayesian analysis where a prior is elicited only after looking at the data is feasible. Using this representation of the selection probability, we considered approximate Bayesian inference
in the paper titled "Pliable Methods for Post-Selection Inference Under Convex Constraints".

In terms of software, all methods have been implemented in python code, available at:

http://github.com/jonathan-taylor/selective-inference.git

We hope to complete an officially released version of this software within the next calendar year. Some methods have also been implemented in the `R` package `selectiveInference`.

Training Opportunities: Graduate student Nan Bi participated in the research funded by this report, her work focusing on describing the "inferactive data analysis" framework and implementation in python. She also participated in some separate work on selection effects in model specification for classical instrumental variable models under the breakdown of the instrumental variable assumptions.

RPPR Final Report
as of 05-Feb-2019

Results Dissemination: The PI spoke on work funded by this grant at several conferences as an invited speaker:

- A selective survey of selective inference. International Congress of Mathematics, Brazil, 2018. 2018.

- Approximate selective inference with maximum likelihood. Workshop on Higher Order Asymptotics and Post-Selection Inference, St-Louis, MO. 2018.

- Inference after selection through a black box. Jerusalem Joint Statistical Event 2018, EMR-BIS and 70th birthday of Yuval Benjamini. 2018.

- Selective Inference in Linear Regression. Symposium on Data Science and Statistics, Reston, VA. 2018.

Honors and Awards: The PI was invited to give a talk at the 2018 ICM (International Congress of Mathematicians) in Rio de Janeiro. This is considered a prestigious invited talk, with roughly 12 talks in probability and statistics every 4 years.

Protocol Activity Status:

Technology Transfer: As mentioned in the "Accomplished" section, software in R and python have been developed to address selective inference problems.

PARTICIPANTS:

Participant Type: PD/PI
Participant: Jonathan Edward Taylor
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Graduate Student (research assistant)
Participant: Nan Bi
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Nothing to report in the uploaded pdf (see accomplishments)