



General Services Administration (GSA)  
Federal Acquisition Service (FAS)  
Assisted Acquisition Services (AAS) Region 9



ANALYTICS SAMPLE TASK ORDER A

Modernizing and Optimizing OPA Statistical Processes

**1.0 INTRODUCTION**

Two goals of the Office of People Analytics (OPA) Data Science team are: 1) to transform current statistical processes (e.g. sampling, weighting, estimation) from their SAS and mainframe environment to free open source software (FOSS) and the Department of Defense (DoD) ADVANA analytics environment, and 2) to develop training and how-to materials for the transition of current users to the new system and for future users. In collaboration with the Data Science team and survey operation analysts, the Data Science team would like to translate and update code used in survey sampling, weighting, and estimation to facilitate a transition from SAS-based procedures to R-based procedures. This effort will provide, at a minimum, a back-up system to current workflows that can become the primary workflow.

**2.0 BACKGROUND**

OPA conducts many DoD-wide scientific surveys that require substantial development time to properly identify and sample the survey population. After the survey has fielded, there is additional work done in weighting and estimation. These steps in the survey process are currently being conducted through a combination of Defense Manpower Data Center (DMDC) mainframe database access and through statistical programming software such as SAS and Microsoft Access. OPA is currently being directed to transition its work and major IT infrastructure to the ADVANA environment. The ADVANA environment will provide storage and analytical solutions different from our current paradigm. While there is ongoing discussion of acquiring SAS on the ADVANA platform, our focus is to develop a fully functional parallel process for these survey activities. We anticipate that in the near future, we will have fully updated and transitioned these statistical procedures from SAS to R.

While some OPA statisticians and operations analysts may have familiarity with the R language, many do not. There is a learning curve associated with not only learning the R language, but of utilizing the R language to do tasks previously done in SAS. There should be a structured and consistent manner in which to train and transition SAS users to R in order to perform work duties. There currently exists no traditional or organized training. Most individuals have learned R through trial and error and utilizing free courses available online. These online courses vary broadly from general R usage down



General Services Administration (GSA)  
Federal Acquisition Service (FAS)  
Assisted Acquisition Services (AAS) Region 9



to specific packages and functions. Navigating these as a brand new user is a daunting task. That may work for some, but it is not sufficient nor efficient to transition OPA's workforce.

Attracting new talent is always a consideration. OPA has been successful in utilizing modern tools and best practices where possible. From a statistical programming perspective, there are many facets of the job that would seem antiquated and unattractive. SAS is not obsolete, but would be hard to find as a top requirement for many positions in private industry, broadly speaking. Academia may be more likely than in the past to train students using tools other than SAS. Some metrics and resources online (TIOBE, PYPL, RedMonk, Indeed, Stack Overflow, IEEE Spectrum, Rexer, and others) demonstrably show a wider adoption of R and R-like tools over SAS in the last 10 years. We would argue that the lift to train or convert new hires in R (or other comparable language) is lighter than to train them in SAS. SAS is also proprietary and costly. Given our move to ADVANA, the costs are unnecessary in light of capable alternatives that already exist.

OPA needs to be proactive in dealing with the changing environment. We would like to develop a solution to transition away from SAS before being forced to. The Data Science team has already started to work on the estimation portion of SAS procedures with great success. To continue this work and start work in other areas of the statistical team's portfolio, more rigorous scripting, testing, and validation is required. More specifically, the goals of this task order are to:

- (1) Work at the direction of government staff in transforming, modernizing, optimizing SAS procedures for sampling, weighting, and estimation fully into the R programming language.
- (2) Work at the direction of government staff to assist in development of an R package encompassing the transformed procedures above. Complete with help documentation, vignettes, and helper functions. This package should be easily digestible and usable by a new user via the documentation and training materials developed. This package will also need to keep the following in mind:
  - a. Code must be well-commented and modular to facilitate updates to the package going forward by staff who may see the code for the first time.
  - b. Follow any guidelines and/or restrictions that are hard boundaries in the ADVANA environment in terms of usable packages, functions, and utilities.



General Services Administration (GSA)  
Federal Acquisition Service (FAS)  
Assisted Acquisition Services (AAS) Region 9



- (3) Work with government staff to rigorously test and validate developed package functions against existing processes to 99.9% accuracy.
- (4) Based on the work and package developed above, develop training materials aimed at the current workforce. This training would start from the basics of R to the use of the developed R package to perform the statistical functions that users previously performed in SAS. This is not intended to be a comprehensive R tutorial. Rather, this is aimed, minimally, at getting someone completely unfamiliar with the R interface acquainted with it and enabling them to follow steps and procedures outlined in a reference guide to perform duties associated with their position. The training could be at a self-taught pace or done in a group setting.

### 3.0 SCOPE

The Contractor shall provide all personnel, equipment, materials, and other items necessary to perform: training materials, user-guides, reference materials, and summary reports. The Contractor shall utilize Government Furnished Equipment (GFE), DMDC IT systems, and ADVANA to perform all other stages of the project, including but not limited to: scripting, testing, and package development. All code, all steps taken, and all results (successful and unsuccessful) should be transparent and include such measures as (1) carefully commented scripts and (2) documentation in a markdown (or similar) file. Progress will be tracked and reported regularly to the OPA Data Science via the project management software JIRA (provided by DMDC) with biweekly summaries provided in writing and discussed in regular scrum sessions and meetings.

### 4.0 REQUIREMENTS

- 4.1 Attend a kickoff meeting with OPA personnel in person or via teleconference. The purpose of the kickoff meeting is to clarify OPA's objectives and to discuss how the Contractor will address those objectives. The Contractor shall provide OPA with a written summary of the kick-off meeting discussion along with a plan of research.
- 4.2 Work at direction of and with OPA staff for further refinement and development of:
  - 4.2.1 Estimation procedures (RSAM). RSAM currently exists in an alpha stage of completion. Further development to incorporate all SAM2 functionality and estimation needs of the survey divisions. Will need to work within the final package framework and ADVANA environment;



General Services Administration (GSA)  
Federal Acquisition Service (FAS)  
Assisted Acquisition Services (AAS) Region 9



- 4.2.2 Sampling procedures. To include refined sampling tool (currently done using SAS and Microsoft Access) and development of fully R-based sampling frame development (currently done on Mainframe). R-based tool for sampling exists but will need to be modified, if necessary, to work within the final package framework and ADVANA environment;
- 4.2.3 Weighting procedures. Weighting programs are currently partially converted at roughly 50/50 SAS and R. This requirement will be aimed at completing the conversion of this process to R, ensuring compatibility within the final package framework, and to ensure compatibility in the ADVANA environment.

This work can be developed and tested on GFEs and other DMDC resources. But final testing should be successfully completed in the ADVANA environment.

- 4.3 Work at direction of and with OPA staff in comprehensive testing of developed R procedures vs currently used SAS procedures. This work can be developed and tested on GFEs and other DMDC resources. Final testing should be successfully completed in both the DMDC environment and the ADVANA environment. These testing and validation results should be presented to OPA in either an R markdown notebook or Python Jupyter Notebook in a narrative fashion. Accuracy, or comparable metric, should strive to be within 99.9% of current procedures.
- 4.4 Work at direction of and with OPA staff in the development and final polish of an R package compilation of the various functions and scripts created through 4.2. The package should be completed with commented, modular code and also have fully functioning helper functions, vignettes, and help documentation.
- 4.5 Develop training materials as outlined in the background section above for use in transitioning current SAS users to R with the package developed. These materials should be usable in a self-taught or group setting. User guides and reference materials for statisticians and analysts to use as they familiarize themselves with the R environment. A draft of these materials should be provided to OPA 2 months prior to the end of the period of performance. The final version should be provided at the end of the period of performance.



General Services Administration (GSA)  
Federal Acquisition Service (FAS)  
Assisted Acquisition Services (AAS) Region 9



- 4.6 Work should follow an agile methodology with at least weekly meetings to align team efforts and project tracking to allow transparency into the status of relevant issues associated with each tasking and a description of the remaining work to be done on the task.
- 4.7 All code should be delivered to OPA and written clearly with appropriate documentation. All code should be commented following the OPA Data Science style guide and standards and should provide the level of detail a team of independent analysts would require to replicate the work that was done and what work may be done in future extensions of this work. Where applicable, version control should be used. Where applicable, package control should be used. Auxiliary functions should be created when there is an emphasis on reproducibility.
- 4.8 Provide a brief report summarizing the development effort but with particular emphasis on describing the results of testing and validating the R procedures versus current SAS procedures. The report should discuss any major roadblocks and considerations towards system/network limitations and potential solutions (both local and ADVANA environment), and a discussion of the implications and way forward. A draft of the report should be provided to OPA 2 months prior to the end of the period of performance. The final version should be provided at the end of the period of performance. No outline is required.
- 4.9 Provide regular status reports in a format to be agreed upon with OPA, which includes the status of relevant issues associated with each tasking and a description of the remaining work to be done on the task. Written deliverables are expected to be in development as early as 1 week following kick-off.
- 4.10 RESERVED

## 5.0 DELIVERABLES

Deliverable	Ref.	Delivery Date (on or about)
Kick-off meeting	4.1	Within 1 week of award
Written summary of Kick-off meeting	4.1	1 week following kick-off
Estimation functions	4.2.1	Deliverable schedule set during kick-off meeting

---

SAMPLE TASK ORDER A PWS - RFP 47QFPA23R0001 Attachment 5

---



General Services Administration (GSA)  
Federal Acquisition Service (FAS)  
Assisted Acquisition Services (AAS) Region 9



Deliverable	Ref.	Delivery Date (on or about)
Sampling Functions	4.2.2	Deliverable schedule set during kick-off meeting
Weighting Functions	4.2.3	Deliverable schedule set during kick-off meeting
Testing and Validation	4.3	Deliverable schedule set during kick-off meeting
R package	4.4	Deliverable schedule set during kick-off meeting. Final delivery End of PoP
Training Materials	4.5	Deliverable schedule set during kick-off meeting. Draft for review 2 months before end of PoP. Final delivery End of PoP
Delivery of code/supporting documentation	4.7	Ongoing during PoP
Draft summary report	4.8	2 months before end of PoP
Final summary report	4.8	End of PoP
Status Reports	4.9	Ongoing during PoP

**6.0 PLACE OF PERFORMANCE.**

The place of performance shall primarily be at the Government facility, but may be performed offsite if prior approval of the COR is obtained.

**7.0 PERIOD OF PERFORMANCE.**

The period of performance shall be twelve (12) months from award.

**8.0 GOVERNMENT FURNISHED PROPERTY (GFP)**

As specified in the contract.

**9.0 SECURITY CLEARANCE.**

No classified data will need to be accessed for performance of this task order.

**10.0 TRAVEL**

N/A

**11.0 ODCS**

N/A



General Services Administration (GSA)  
Federal Acquisition Service (FAS)  
Assisted Acquisition Services (AAS) Region 9

