

Developing Seamless Tools to Support Metrics for Adaptive Survey Designs

Stephen H. Cohen¹, Dean Resnick¹, Zachary H. Seeskin¹, Imad Lakhali¹

¹NORC at the University of Chicago, 4350 East-West Highway, Suite 800,
Bethesda, Maryland 20514

Abstract

NORC has used adaptive survey design (ASD) principles in its survey operations for many years based upon quantitative metrics such as response rates by key characteristics and minimum sample targets. To facilitate use of these metrics and for introducing additional metrics such as R-indicators, NORC has developed and is enhancing tools to make applications of ASD metrics more efficient at lower costs. SharePoint is used as the interface given its ease of application to survey directors and field staff with SAS as the tabulation engine underlying the interface. The file set-ups allow access to both the collection and survey databases to provide seamless real-time statistics via the tool. Outputs can be Excel spreadsheets with detailed statistics or graphs that can be used to easily discern trends. R package *shiny* can be used to allow project teams to dynamically visualize adaptive design-related measures.

Key Words: Adaptive Design, R-indicators, graphics, real-time estimates

1. Introduction

NORC's adaptive survey design (ASD) strategy is to maximize the use of quantifiable metrics to adjust collection protocol targeting in response to results of collection efforts in the field. Often software implementing these metrics is developed separately for each survey resulting in increased costs and inefficiencies.

NORC has developed and is expanding an ASD IT tool that will make metrics that assess data collection more readily available across the entire organization as well as easily integrated into survey operations in a single web-based tool. These metrics can be used by non-statisticians after the models have been derived and validated. The system will permit updating model parameters with the latest information. The ASD team is providing support on the application and interpretation of the metrics.

The SharePoint interface allows for easy integration of additional metrics that require access to real-time collection status of sample units.

2. Background

It is well-known that gaining responses to social science survey questionnaires has become very challenging and more difficult over time. At the same time, there has also been increased pressure to deliver survey results faster and at lower cost while maintaining quality. Many past efforts at survey quality have focused on increasing response rates. However, it well documented that response rates alone are poor indicators of non-response survey bias (Groves, 2006; Groves and Peytcheva, 2006). Also, high response rates alone do not ensure that survey results are truly representative of the population. A survey with a high nonresponse rate could still be acceptable if the responding sample is representative of the population depending upon the user's tolerance for higher sampling errors.

Groves and Heeringa (2006) proposed that current, real-time paradata should be used to change treatment of cases to improve performance and control costs, rather than following predetermined collection protocols. Schouten et al. (2009) extended efforts at improving response by developing R-indicators that highlight where efforts should be devoted to minimize survey bias based upon response propensities exhibited by sample members. R-indicators may complement response rates and facilitate the assessment of survey quality over time as well as the comparison of the effectiveness of different adaptive design strategies.

Traditional sample designs assign treatment protocols to different groups at the start of collection based upon frame data or past experience with the respondent. They are left in their assigned protocols for the entire collection cycle, or mid-course corrections are made on an ad hoc basis and may not be well documented or supported. Adaptive survey designs assign sample members to treatment protocols as is classically done, but carefully monitor survey response and other metrics to change the treatment of the sample members based upon analysis of survey paradata and statistical model outputs during data collection. Thus, adaptive design uses information available during data collection to help achieve survey objectives. In adaptive designs, individually tailored, time-varying treatment approaches are applied (Wagner, 2008). The goals of adaptive design may include:

- 1) Improving the representativeness of key subgroups,
- 2) Effecting meaningful improvement in response rates,
- 3) Reducing the design effects due to weighting;
- 4) Reducing collection costs,
- 5) Improving timeliness of survey estimates, and
- 6) Reducing efforts on unproductive cases.

Adaptive design recognizes the limitation of response rates and as measure of survey success and attends to different sources of survey error, including nonresponse bias, measurement error, processing error, and sampling error. Adaptive design has improved upon traditional designs by adding data-driven design choices at the start of data collection or at mid-course sample releases. Decisions are informed through continuous monitoring of the sample representativeness using questionnaire data from the sample in the field. One applies different intervention strategies across subpopulations to reduce bias. At the same time, costs are controlled by analyzing technique effectiveness using paradata in addition to propensity models that predict the probability of a sample unit to respond. Paradata are valuable for the information they provide about the raw data collection, such as skill of the field agents, time of successful completions, and number of contacts until a successful completion.

3. NORC Corporate Goals

NORC field staff have practiced executing adaptive survey designs for many years. Over the years NORC has used paradata, weekly collection reports, past experience, incentives to adjust collect protocols. Commonly used metrics include:

- Response rates, overall and by subgroups of interest.
- Stability of estimates as response increases.
- Item nonresponse rates.
- Non-sampling error.
- Comparison of survey estimates to historical estimates for critical variables.
- Measures of field effort.
- Representative indicators.

4. ASD IT Tool

The ASD IT tool is designed as a web-based instrument accessed through the NORC Intranet via SharePoint. The tool requires access to the data monitoring systems to access real-time response data to produce quantifiable metrics. Once set up to access the correct sample units for a survey, the system seamlessly accesses the current status of every unit.

The system also requires links to survey data elements as needed to run metrics. No survey specific data is downloaded to the ASD IT server in order to meet client restrictions on access to confidential data, e.g., Confidential Information Protection and Statistical Efficiency Act of 2002, Health Insurance Portability and Accountability Act of 1996 (HIPAA). The tool accesses data needed from the protected survey folders. The computational engine for the tool is SAS.

Initially R-indicators measures were included in the beta version of the tool. We are currently adding an integrated graphics module and a metric that predicts the response over a sample unit based upon the last round for longitudinal surveys or results from similar surveys that can be applied to the current effort.

4.1 Representativity Indicators for Adaptive Survey Designs

Representativity indicators or R-indicators as described by the Cathie Marsh Institute for Social Research's RISQ Project are supported by the ASD IT tool. R-indicators are measures of response rate variance. The three indicators answer the questions:

- How does the collected sample reflect the underlying population?
- How much does each key variable (covariate) contribute to the lack of sample representativeness?
- Which variable (covariate) category is (are) causing lack of representativeness?

The first task in applying R-indicators is identification of objectives that identify the statistics of primary importance. This provides a basis for measuring the representativeness of the key statistics in the adaptive/responsive survey design using the ASD tool. Using these key statistics, the tool supports estimation of response propensities using logistic regression. We describe the formulae to calculate R-indicators and the interpretation of their values.

4.2 Overall R-Indicator

The sample R-indicator is a measure of how representative the collected sample is compared to the underlying population:

$$R(\hat{\rho}) = 1 - 2 \left(\sqrt{\frac{1}{N-1} \sum_{i=1}^N \frac{s_i}{\pi_i} (\hat{\rho}_i - \hat{\rho})^2} \right),$$

where s_i indicates whether case i is in the sample, π_i reflects selection probability, $\hat{\rho}_i$ is the estimated response propensity, and $\hat{\rho}$ is the average response propensity over the entire sample. The sample R-indicator has a lower bound of 0 and an upper bound of 1, when the responding sample is fully representative of the population. It should be noted that this indicator is independent of specific survey outcome variables.

Two other metrics help answer the second and third questions above.

4.3 Variable-Level R-Indicator

The unconditional variable-level partial R-indicator is used to ascertain the relative strength of each covariate (e.g., beneficiary's age or location) to predict response propensity. This unconditional R-indicator measures the contribution of a specific variable to the overall R-indicator. This value is obtained by computing the variation in the propensities across the categories or domains, $k = 1, \dots, K$ defined by the variable x in question using the following equation:

$$R_u(x, \hat{\rho}) = \sqrt{\sum_{k=1}^K \frac{N_k}{N} (\hat{\rho}_{x,k} - \hat{\rho}_x)^2}.$$

In this formula, k indexes the categories of variable x , $\hat{\rho}_{x,k}$ is the estimated response propensity of category k , and $\hat{\rho}_x$ is the average response propensity across the categories. This indicator is bound between 0.0 and 0.5. When comparing two variables, the one associated with a higher R-indicator means that it causes more variation in response propensity than the other variable does.

The unconditional variable-level R-indicator measures the representativeness of the responding sample associated with each variable. For example, the sample may be balanced by geographic region but not by age categories. However, it does not indicate which age group(s) is (are) over- or underperforming.

4.4 Category-Level R-Indicator

A category-level partial R-indicator complements the variable-level measure and is defined as

$$R_u(x, k, \rho) = \sqrt{\frac{N_k}{N} (\bar{\rho}_{x,k} - \bar{\rho}_x)^2}.$$

This indicator, which is bound between -0.5 and $+0.5$, shows which beneficiary's age categories (for example) are over- or under-represented in the responding sample. Therefore, the second indicator can provide useful guidance about where to focus data collection resources to achieve a more representative final sample.

The values of the partial R-indicators inform data collection managers where they need to concentrate data collection resources to improve the representativeness of the responding sample. When run after a successful intervention, partial R-indicators can be expected to show a more representative responding sample compared to the pre-intervention sample.

4.5 Overview—Interpreting the R-indicators

In summary, the overall R-indicator provides an indication of representativeness of the collected sample. As noted above, the measure ranges from 0 to 1 with a value of 1 for the collected sample indicating perfect sample representativeness of the underlying population. However, other indicators are needed to guide where to concentrate collection resources.

The variable- and category-level R-indicators are needed to identify where to focus collection resources. These statistics are designed to measure how representative the collected data are for key survey statistics. Comparing the variable-level R-indicators across covariates identifies which variables are not covered well by the collected data. Covariate R-indicators run from 0 to 0.5 with 0 being ideal. The categorical R-indicator narrows the problem area down to specific values. The value of the categorical R-indicator measure goes from -0.5 to $+0.5$ with a value of 0 being ideal.

To identify where intervention is needed, one should review the R-indicators as follows:

- ➔ Review the overall R-indicator with a goal of the final collected sample having a value of greater than the survey goal (such as 0.75). However, as cited above, this indicator will not identify how to focus resources when the values are below survey targets.
- ➔ Review the partial variable (covariate) R-indicators to identify sample characteristics that fall short of representativeness goals, e.g. covariate R-indicator values greater than a predetermined goal. Note that at this step, one should prioritize which covariates to target.
- ➔ Review the partial categorical variables of covariates to identify which sample members need attention (e.g. covariate categorical values below and above a pre-set value). It should be noted that positive values indicate the portion of the sample has better representativeness than those categories with a lower value. Note that focusing solely on sample members with high negative categorical values might result in an undesirable effect, i.e., sample members originally showing positive characteristic values could become problematic subsequent to the intervention.

Typical interventions include:

- 1) Offering a pre- or post-paid monetary incentive after a certain number of contact attempts,
- 2) Assigning the more difficult cases to more experienced interviewers,

- 3) Shifting mail solicitations for interviews from standard mail to priority mail or FedEx,
- 4) Changing strategies for obtaining appointments for sample members from one survey mode to another (e.g., mail to CAPI) for subgroups where a particular mode seems effective,
- 5) Proposing different interview schedules,
- 6) Subsampling nonrespondents to concentrate intensive effort on fewer case, and
- 7) Offering an abbreviated interview to capture critical items during the late stages of data collection.

These interventions may or may not be applicable to or useful for every survey. Interventions will need to be developed and vetted within the constraints of the budget and with an eye toward maintaining rapport with respondents, especially for longitudinal surveys.

5. Employing the ASD IT Tool

As noted earlier, the ASD IT tool is a SharePoint application that uses SAS as the computation engine. In order to set up the adaptive design propensity models along with the survey's collection status in real-time, the tool must be input with the entire selected sample, including sampling weights. Once the models are developed, the model form should remain fixed throughout the current field cycle with only the parameters updated with new collection of sample member's status. The response propensities will be used to calculate the R-indicators and partial unconditional R-indicators.

If the model is changed, comparisons of R-indicators across time periods reflecting different model versions will be problematic.

SharePoint screens are used to complete the steps necessary to compute R-indicators. A manual provides step-by-step instructions on using the tool. The sequence of steps for computing R-indicators includes:

- Access the sample input file and real-time collection status codes.
- Create an analysis file of variables from the sample input file that will be used to develop the propensity models for monitoring the representativeness of the survey collection.
- Utilize both main effects and interaction terms in constructing the response propensity model.
- Review the model fit after each run to inform the decision on whether to accept the model or change the main effects/interaction terms used.
- Run the R-indicators in real-time.
- Develop a report including past R-indicators for use in assessing collection progress.
- Export as Excel files or a file to a graphics package.

6. ASD IT Tool Screen Shots

We will present a selected few screen shots to illustrate how the tool works. The first screen shot identifies the survey data needed to construct the propensity models that are used to generate R-indicators. First as discussed above, one must decide the key statistics for which the final sample must be representative. The user selects from the variables available from the data source. Then, the user then must select the variable that provides real-time collection status for the entire sample. Many tracking systems include more status codes than needed to compute R-indicators so collapsing of codes might be necessary.

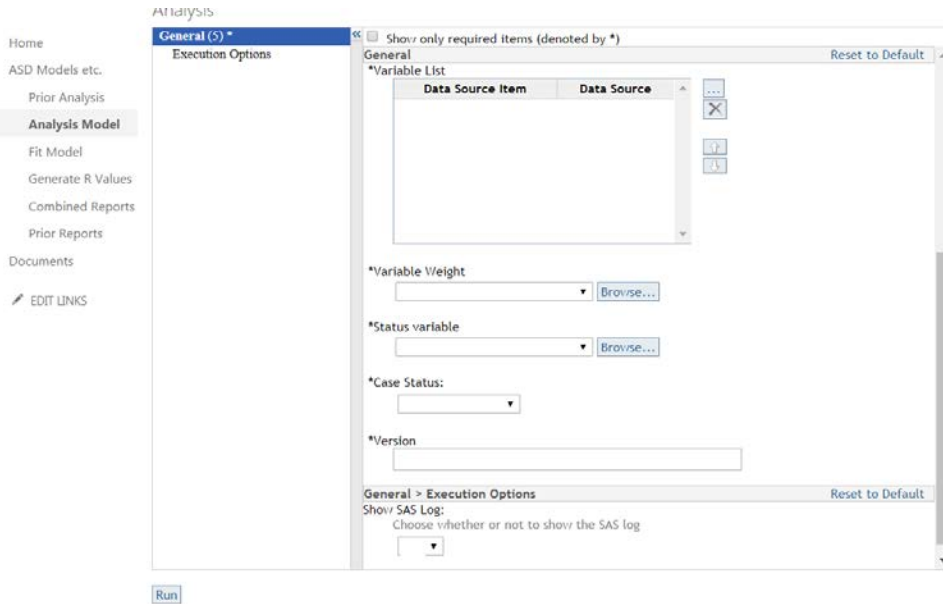


Figure 1: Screen shot of ASD IT Tool's data input facility

Once the analysis file is ready, the ASD IT tool returns statistics on the analysis file.

- Home
- Notebook
- Documents
- Recent
- JoyTest
- SAS TESTS
- Analysis Model**
- Fit Model
- Generate R Values
- Prior Reports
- Combined Reports
- Site Contents

Analysis File

The CONTENTS Procedure

Data Set Name	AFANALYSIS_V1	Observations	12686
Member Type	DATA	Variables	6
Engine	META	Indexes	0
Created	12/29/2016 15:18:59	Observation Length	48
Last Modified	12/29/2016 15:18:59	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

Engine/Host Dependent Information

Data Set Page Size	65536
Number of Data Set Pages	10
First Data Page	1
Max Obs per Page	1361
Obs in First Data Page	1326
Number of Data Set Repairs	0
ExtendObsCounter	YES
Filename	D:\Adaptive Design\Analysis file\analysis_v1.sas7bdat
Release Created	9.0401M3
Host Created	X64_SRV12

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
1	SAMPLE_RACE_78SCRN	Num	8	BEST1.	BEST1.
2	SAMPLE_SEX_1979	Num	8	BEST1.	BEST1.
4	SAMPWEIGHT_1979	Num	8	BEST7.	BEST7.
5	Status	Num	8	BEST1.	BEST1.
3	URBAN_RURAL_1979	Num	8	BEST2.	BEST2.
6	_status	Num	8		

Figure 2: Screen shot of ASD IT Tool’s statistical output on analysis file

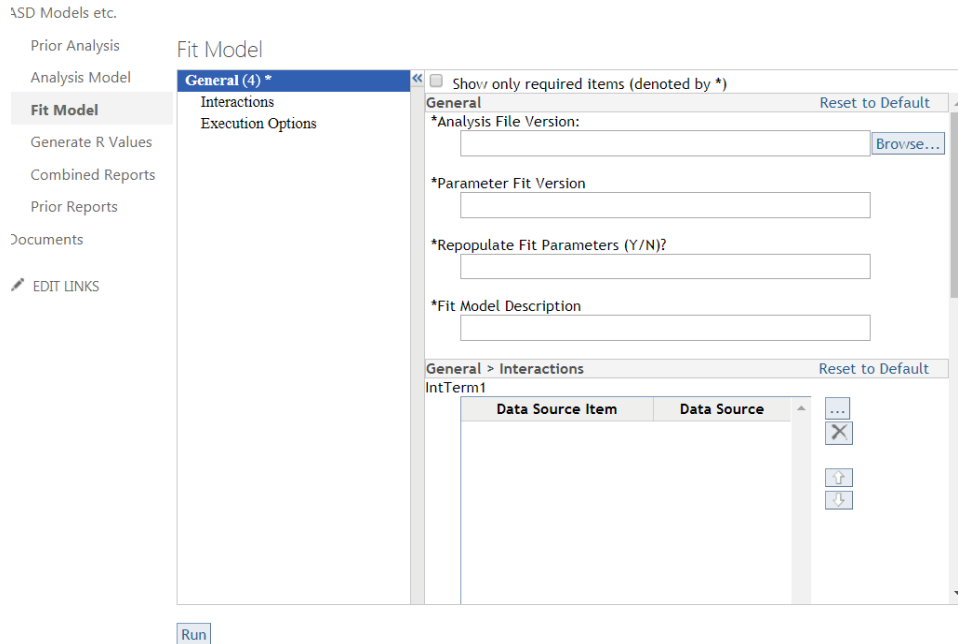


Figure 3: Screen shot of ASD IT Tool’s modeling input

The above screen allows a statistician to develop the logistic regression model for calculating response propensities. Models may not be stable until sufficient sample has been collected such as the response rate is around 50%. Usually just including main effects results in a satisfactory fit but the software allows for adding interaction terms as necessary. Submitting the model with the run command returns SAS logistic model diagnostics.

After developing the model the following screen generates R-indicators.

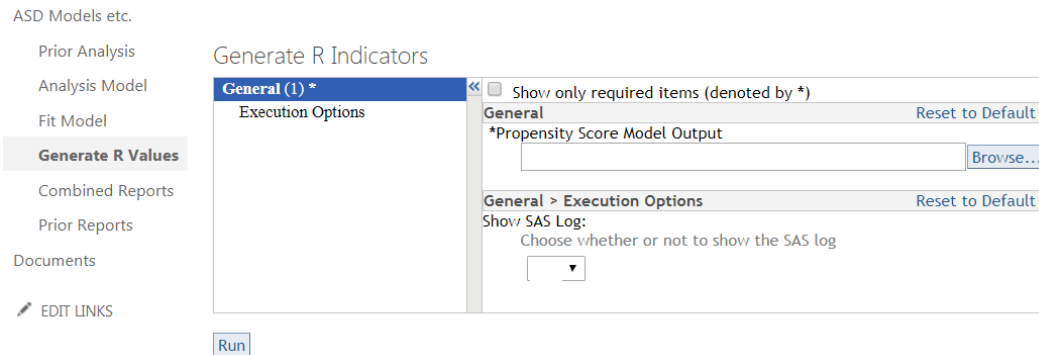


Figure 4: Screen shot of ASD IT Tool’s facility to generate R-indicators

The model returns an Excel spreadsheet. The user is given an option of how many past cycles of results to include in the spreadsheet.

7. Visualizing R-Indicators

Survey managers working on a survey day-to-day can easily spot trends from the detail numbers in the Excel spreadsheets. However many users might find it is easier to spot trends visually. Survey organizations might prefer to send graphs rather than Excel spreadsheets to clients.

We are in the process of building a SAS graphics capability into the tool. Currently, to produce graphics, the Excel output file can be visualized in an independent R Shiny dashboard tool available at NORC.

Graphic packages need flexibility to be customized for client presentations but also should be easy-to-use for novices. R-indicator graphs should allow the user to change the scale on the vertical and which time periods that should be displayed to minimize weekly changes that might not be significant.

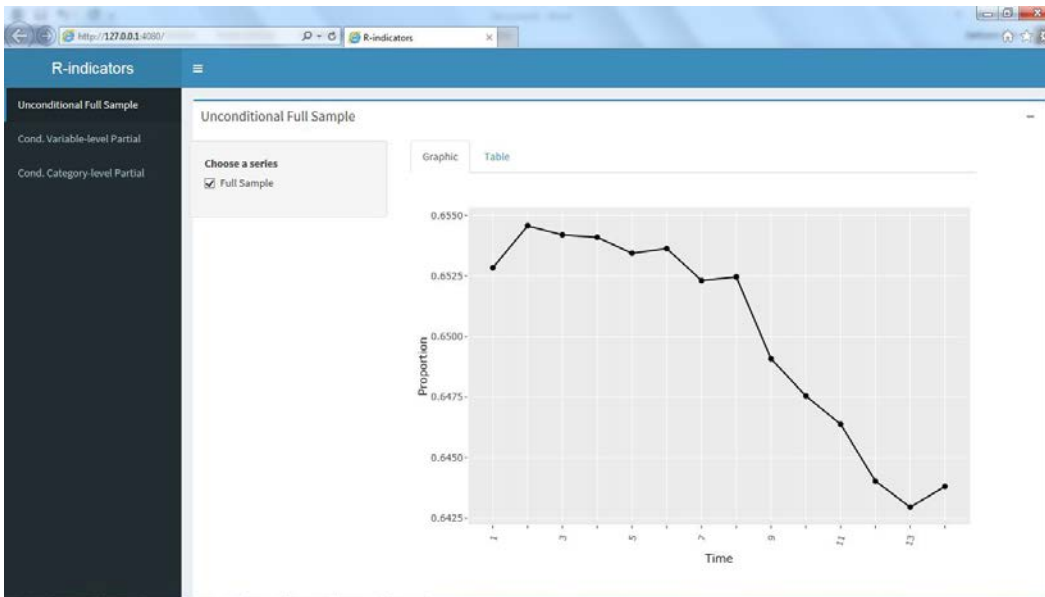


Figure 5: Screen shot of dynamic visualization of overall R-indicator

Graphs for the covariate R-indicators can display all covariate measures and can be restricted to the ones of interest for clean displays.

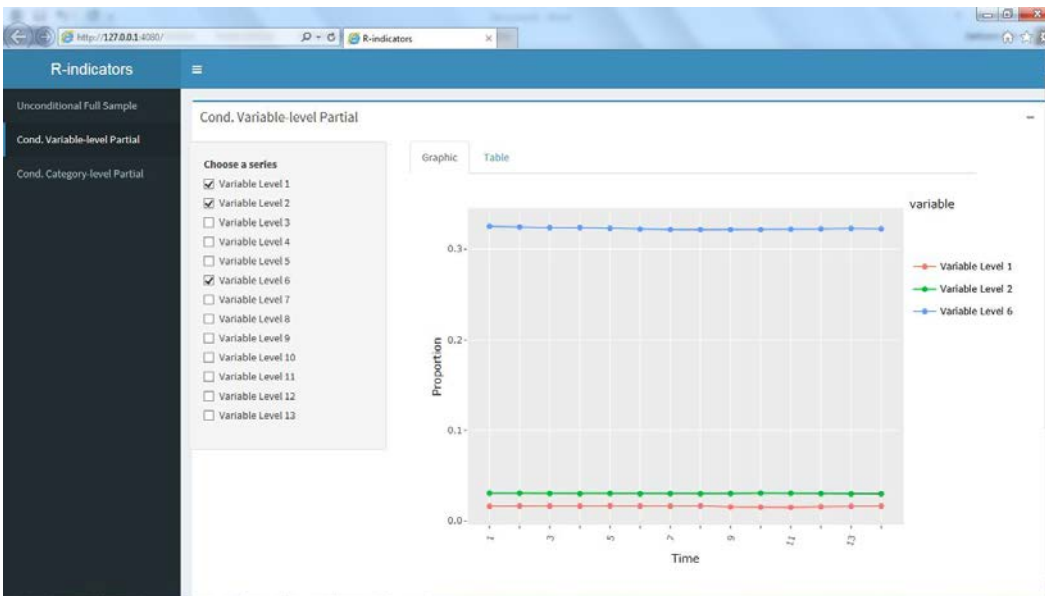


Figure 6: Screen shot of dynamic visualization of variable-level R-indicators

Then one can display the categories of the covariate measures that are of interest.

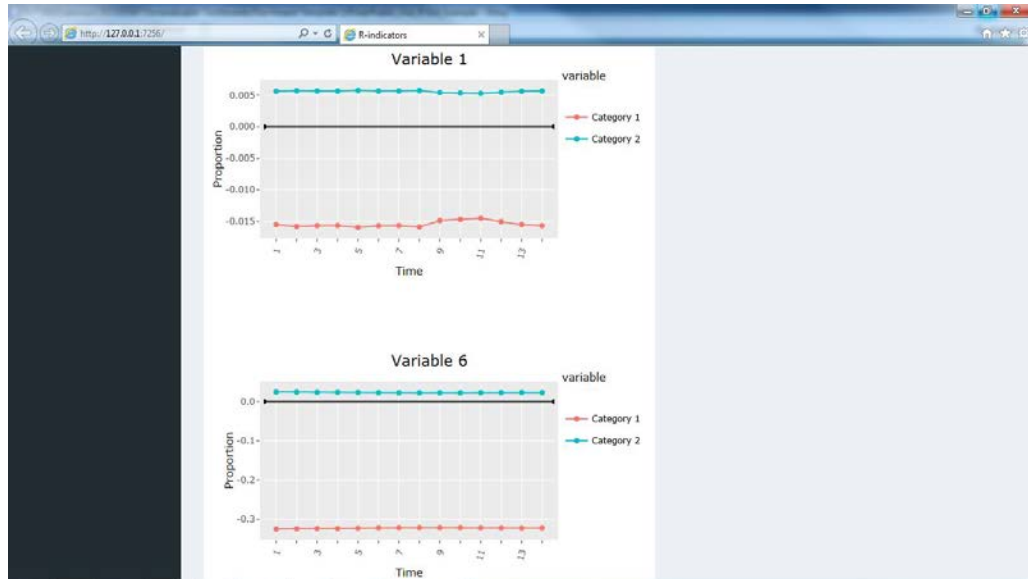


Figure 7: Screen shot of dynamic visualization of category-level R-indicators

7. Conclusion

NORC has deployed an ASD IT tool that can seamlessly produce and visualize R-indicators in real-time. The tool has been developed to be easily used by non-statisticians. Additional measures over time can be readily added based upon staff input.

References

- Groves, R. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys, *Public Opinion Quarterly*.
- Groves, R. & Heeringa, S. (2006). *Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs*. J. R. Statistical Society A 169, Part 3, 439-457.
- Groves, R., & Peytcheva, E. (2006). *The Impact of Nonresponse rates on Nonresponse Bias: A Meta-Analysis*. 17th International Workshop on Household Nonresponse.
- Schouten, B., Cobben, F., & Bethlehem, J. (2009). *Indicators for the Representativeness of Survey Response*, Survey Methodology 35, 101-113.