William R. Whitledge

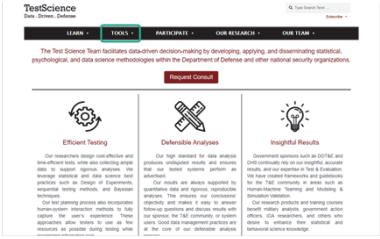
Institute for Defense Analyses

Test Science Apps

In operational testing and evaluation of Departments of Defense (DOD) and Homeland Security (DHS) acquisition systems, analysts repeatedly encounter certain types of data, metrics, and research questions. For example, researchers often estimate a system's reliability as a function of usage or the probability that it will detect or destroy a target depending on range or other variables. And researchers often use surveys to assess system usability, user satisfaction, training adequacy, and other human factors related to the system's effectiveness or suitability.

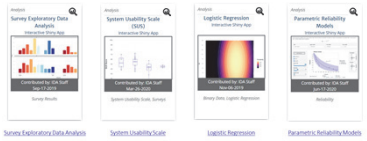
This poster describes four web-based tools I developed to automate analyses that IDA routinely does during test and evaluation work. These tools are available for use at IDA's Test Science Tools webpage:

https://testscience.org/interactive-tools/



Source: https://testscience.org/

Examples



These apps are a subset of roughly 30 free interactive apps and downloadable spreadsheet tools available through the Test Science Tools public webpage covering topics in test planning, design, and analysis. They ingest simple numeric values and text-based tables of survey responses and series of numbers. No coding or special software required!

Analysis Apps for the Operational Tester

Use Test Science web apps to improve the efficiency, aesthetics, and reproducibility of your analysis.

Apps presented here are coded in R using the Shiny package.

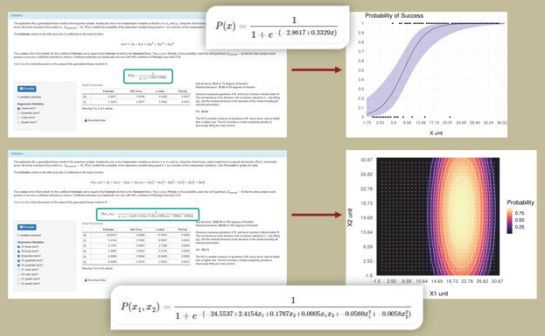
Analyze Likert scale survey responses[1]

Quickly review and plot groups of Likert scale survey responses as column graphs, histograms, and box plots to assess user satisfaction, training adequacy, and other human factors.

[1] A Likert scale survey response is a multiple choice numeric response (e.g., 1 through 7) indicating level of agreement or confidence with a survey question or statement. The response 1 often indicates "strong disagreement," and 7 often indicates "strong agreement."

Estimate the probability of an event occurring

Fit a logistic regression model to one or two independent continuous variables and plot the probability of mission success, threat detection, target destruction, or other success (1) or failure (0).

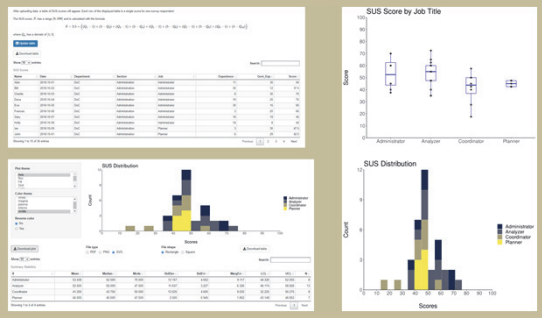


All graphs on this poster show imaginary randomly-generated data.

R logo downloaded on February 8, 2022 from https://www.pngitem.com/middle/ibwTimx_r-studio-icon-png-transparent-png/

Assess usability using the system usability scale (SUS)[2]

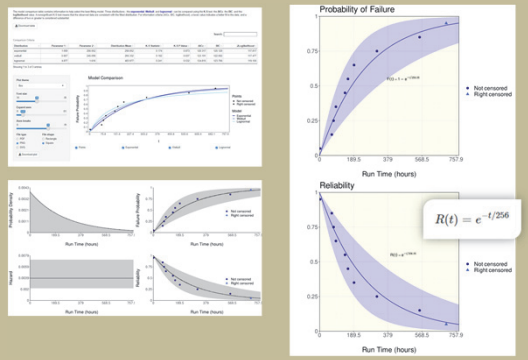
Review survey responses, calculate SUS scores, sort tables, and plot scores by independent variables to assess system usability.



[2] The SUS is a scale with range [0, 100]. Ten 5-point Likert scale survey questions are used to calculate a single SUS score from the questions. The SUS score is a metric of usability that is comparable across different systems. See https://uxpajournal.org/determining-what-individual-sus-scores-mean-adding-an-adjective-rating-scale/.

Model reliability using simple parametric distributions[3]

Review and sort data, fit and compare reliability models (Exponential, Weibull, and Lognormal), and plot system reliability.



[3] This app is designed to use time or interval length data to fit and plot reliability distributions. It has limited ability to plot reliability distributions based only on a parameter, such as mean time to failure.

The right app for the right research question

- Human Factors (Often measured with Likert scale survey responses) → Survey Exploratory Data Analysis
- Usability (Measured with a specific Likert-like survey in which responses form a single-number metric or usability score) → System Usability Scale
- Probability of Success (Estimated from binary event data: something either happens or does not) → Logistic Regression
- Reliability (Often estimated using service tickets and other sources of event times, durations, distances traveled, etc.) → Parametric Reliability Models

Why use these apps?

- Better reproducibility of results
- Faster analysis of new similar data
- Standard and more beautiful aesthetics in figures
- Easier data uploads and table and figure downloads
- Smaller workloads in future analyses
- Free and web-accessible

Where do these apps live?

Apps are available for public Internet use, and the source code is currently available for IDA-internal use.

Survey Exploratory Data Analysis	•App link: https://test-science.shinyapps.io/survey_data_analysis/ •Code repository: https://code.ida.org/projects/TSSHINY/repos/survey_data_analysis/browse
System Usability Scale	•Link: https://test-science.shinyapps.io/system_usability/ •Code: https://code.ida.org/projects/TSSHINY/repos/system_usability/browse
Logistic Regression	•Link: https://test-science.shinyapps.io/logistic_regression/ •Code: https://code.ida.org/projects/TSSHINY/repos/logistic_regression/browse
Parametric Reliability Models	•Link: https://test-science.shinyapps.io/ParametricReliabilityModels/ •Code: https://code.ida.org/projects/TSSHINY/repos/parametric_reliability_models/browse

Acknowledgments

Thank you to Kelly Avery, Brian Conway, John Haman, Bram Lillard, and Kelly Tran for reviewing this poster and presentation.

Special thanks to all the people who reviewed these applications: Lee Allison, Kelly Avery, Jonathan Bell, Rose Clark, Brian Conway, Caitlan Fealing, Thomas Johnson, Curtis Lansdell, Peter Mancini, Keyla Pagan-Rivera, Conor Schlick, Jason Schlup, Jason Sheldon, Kelly Tran, Brian Vickers.