



INSTITUTE FOR DEFENSE ANALYSES

## **DATAWorks 2022: Analysis Apps for the Operational Tester**

Vincent Lillard, Project Leader

William Whitledge

April 2022

Public release approved. Distribution is  
unlimited.

IDA Document NS D-32959

Log: H 2022-000030

INSTITUTE FOR DEFENSE ANALYSES  
730 East Glebe Road  
Alexandria, Virginia 22305



The Institute for Defense Analyses is a nonprofit corporation that operates three Federally Funded Research and Development Centers. Its mission is to answer the most challenging U.S. security and science policy questions with objective analysis, leveraging extraordinary scientific, technical, and analytic expertise.

#### About This Publication

This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-19-D-0001, Task BD-229990, "Test Science App," for the Office of the Director, Operational Test and Evaluation. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

#### Acknowledgments

The IDA Technical Review Committee was chaired by Mr. Robert R. Soule and consisted of: Dr. Kelly M. Avery, Dr. Brian T. Conway, Dr. Kelly Tran, Dr. John T. Haman, and Dr. Vincent A. Lillard from the Operational Evaluation Division.

#### For more information:

Dr. Vincent A. Lillard, Project Leader  
vlillard@ida.org • (703) 845-2230

Robert R. Soule, Director, Operational Evaluation Division  
rsoule@ida.org • (703) 845-2482

#### Copyright Notice

© 2022 Institute for Defense Analyses  
730 East Glebe Road, Alexandria, Virginia 22305 • (703) 845-2000

This material may be reproduced by or for the U.S. Government pursuant to the copyright license under the clause at DFARS 252.227-7013 [Feb. 2014].

Rigorous Analysis | Trusted Expertise | Service to the Nation

INSTITUTE FOR DEFENSE ANALYSES

IDA Document NS D-32959

**DATAWorks 2022: Analysis  
Apps for the Operational Tester**

Vincent Lillard, Project Leader

William Whitley



## Executive Summary

---

In the world of Department of Defense (DOD) and Homeland Security (DHS) acquisition and testing, researchers and data 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 (e.g., time or distance until a failure, alert, or detection), and they often estimate the probability that a system will detect, destroy, or survive a threat 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. Although common in testing of new systems, these types of analyses are generally not trivial, quick, or easy, especially when it comes to visualizing the data intuitively.

Researchers and analysts need tools that enable them to produce *and reproduce* quality and timely analyses of the data they acquire during testing. This document includes slides and a poster describing four web-based apps designed to enable researchers to answer these types of questions or evaluate these types of metrics. The slides and poster are for presentation at the April 2022 Defense and Aerospace Test and Analysis Workshop (DATAWorks). The poster presentation will include live demonstration of the web-based apps using randomly generated imaginary data.

The apps are designed to assist analysts and researchers with simple repeatable analysis tasks, such as building summary tables and plots for reports or briefings. The first app calculates summary statistics and produces plots of groups of Likert-scale<sup>1</sup> survey question responses. The second calculates the system usability scale (SUS) scores<sup>2</sup>

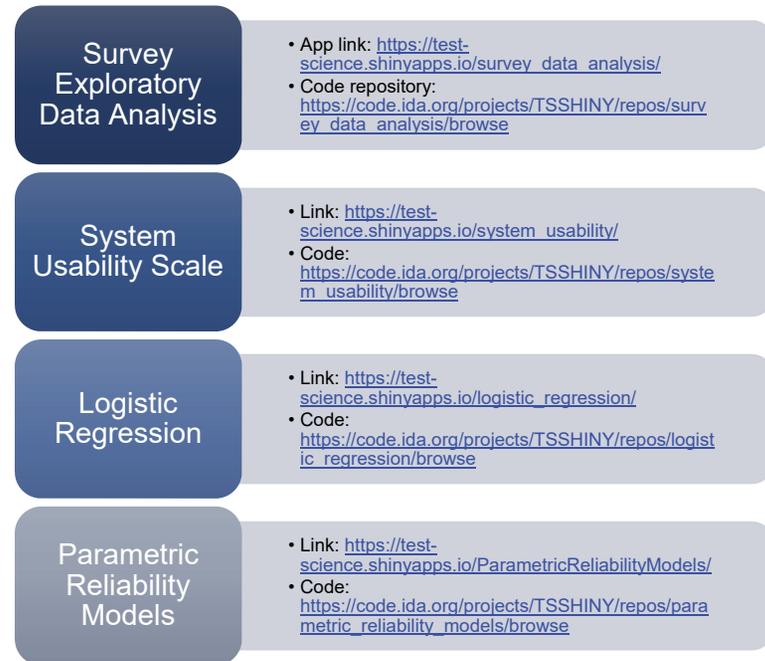
---

<sup>1</sup> 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.”

<sup>2</sup> 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/>.

for SUS survey responses and lets the app user plot scores versus an independent variable. The third app fits a logistic regression model to binary data with one or two independent continuous variables as predictors. The fourth app models reliability of a system or component by fitting parametric statistical distributions to interval-to-failure data (e.g., time to failure, miles to breakdown, etc.).<sup>3</sup>

These four apps are a subset of about 30 interactive web-based apps and downloadable spreadsheet tools maintained by the IDA Test Science group and are available for public use on the Test Science Interactive Tools webpage <https://testscience.org/interactive-tools/>.<sup>4</sup> Figure 1 shows how to access the apps directly.<sup>5</sup> It is worth noting that apps and tools hosted on the Test Science Tools webpage are free, easy to use, and available to anyone with a modern web browser and Internet connection.



**Figure 1. Four data analysis apps with locations**

Using software tools like these apps can increase reproducibility and accuracy of results, timeliness of analysis and reporting, and attractiveness and standardization of aesthetics in figures. Reproducibility of

---

<sup>3</sup> The reliability 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 or  $\alpha$  and  $\beta$ .

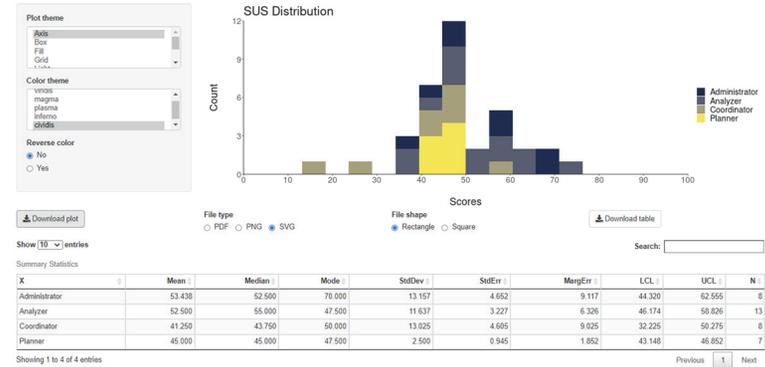
<sup>4</sup> IDA Test Science applications cover topics such as reliability, Bayesian analysis, t-tests, calculating statistical power in designed

experiments, and generating binomial operating characteristic curves. They are designed to be intuitive with sufficient internal documentation to enable an analyst with basic or intermediate statistical knowledge to use them.

<sup>5</sup> Apps are available for public use via the Internet. Source code repositories are only available for IDA-internal use at this time.

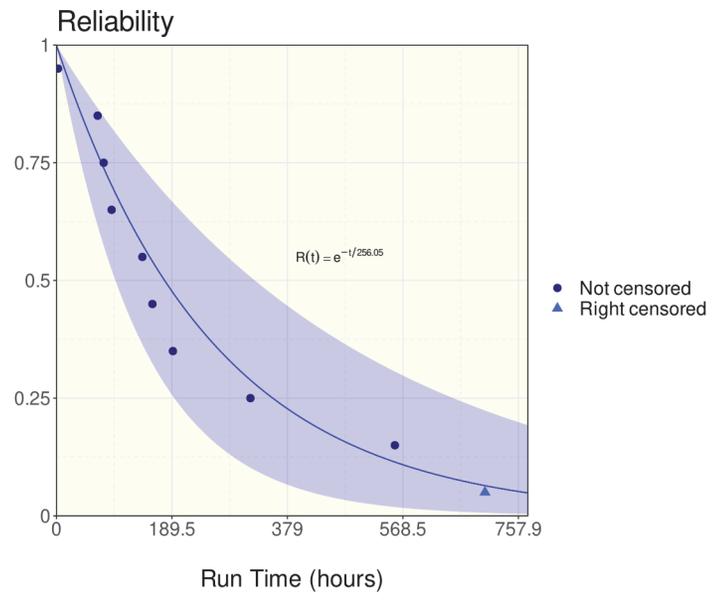
results is important, because separate analysts may need to review or modify an analysis of the same data set years or miles apart. The ability to reproduce the same result (e.g., a graph or table) increases credibility and confidence in the conclusions and recommendations drawn by the separate analysts from the common data. Timeliness and accuracy of results need no explanation, and figures and tables with standard and attractive aesthetics enable communication of results with less explanation needed.

Each app is designed to read a data table in a simple commonly-used format and produce downloadable images and spreadsheets, such as box plots, histograms, scatter plots, and tables of summary statistics. For example, the System Usability Scale app produces an output shown in Figure 2, which includes a histogram of usability scores and a table of summary statistics for data categories. The table is sortable, searchable, and filterable, and it can be downloaded as a new spreadsheet file. The user can also download the histogram in multiple image formats, such as .png, .svg, and .pdf.



**Figure 2. Sample output from System Usability Scale**

Figure 3 shows an example from the Parametric Reliability Models app. This app lets the user analyze time-to-failure data and produce estimates of reliability over time like the chart shown here. This type of chart is also downloadable in the three common formats (.png, .svg, and .pdf), and the app contains features that enable the user to change chart elements such as color, labels, legends, equations, and grid lines. (Both Figure 2 and Figure 3 show plots of imaginary randomly generated data.)



**Figure 3. Sample output from Parametric Reliability Model**

# Contents

---

1. Briefing – Analysis Apps for the Operational Tester, Introduction .....	1
2. Analyze Likert scale survey responses.....	6
3. Assess usability using the system usability scale (SUS) .....	8
4. Estimate the probability of an event occurring .....	10
5. Model reliability using simple parametric distributions.....	12
6. Backup.....	19
Appendix A Acronyms .....	A-1
References.....	R-1
Analysis Apps for the Operational Tester, Poster.....	P-1





# Analysis Apps for the Operational Tester

William R. Whitley

Defense and Aerospace Test and Analysis Workshop (DATAWorks)  
April 26-28, 2022  
Presented on April 27, 2022

**Institute for Defense Analyses**

730 East Glebe Road • Alexandria, Virginia 22305

Since 2012, I have worked at IDA on test and evaluation projects for both the departments of Defense (DOD) and Homeland Security (DHS), including as an operational tester for DHS. In that time, I have repeatedly encountered certain types of data, metrics, and research questions that are analyzed during test and evaluation on acquisition programs.

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

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

# Recurring research questions and data in operational testing

- Human factors  
Often measured with Likert scale survey responses
- Usability of a system or interface  
Measured with a specific Likert-like survey in which responses form a single-number metric or usability score
- Probability of success, target detection, etc.  
Estimated from binary event data: something either happens or does not
- Reliability of a system or component  
Often estimated using service tickets and other sources of event times, durations, distances traveled, etc.

# One way to answer recurring research questions or analyze recurring data consistently and quickly is using tailored software tools.

```
# Source the file 'config.R'.
source("config.R")

# 2. User Interface -----

# Source the file 'UI.R'.
source("UI.R")

# 3. Server Function -----

# Define server logic to provide the application
server <- function(input, output, session){ }

# 4. Execute Application -----

# Run the whole R Shiny application
shinyApp(ui = ui, server = server)
```

```
# This is a function to name a plot file for downloading
download_name <- function(ID, Fshape, Ftype){
  if(Fshape == shape_values[1]){paste(ID, '_plot_rect.', switch(Ftype, pdf = dow
  } else if(Fshape == shape_values[2]){paste(ID, '_plot_sq.', switch(Ftype, pdf =
  }

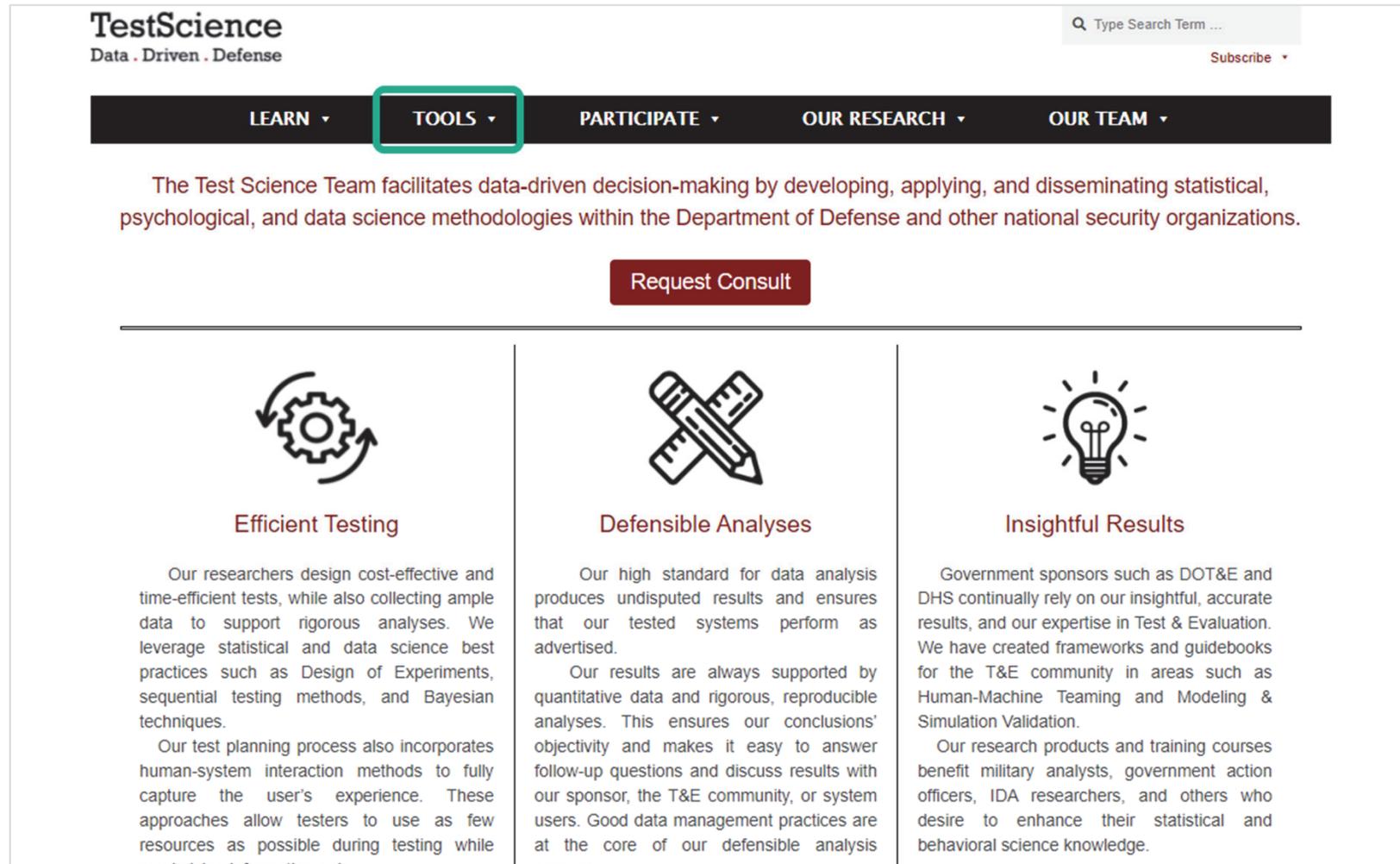
# This is a function to write a ggplot object to an image file
download_write <- function(File, FUN, Fshape, Ftype){
  if(Fshape == shape_values[1]){
    switch(Ftype, pdf = pdf(File, width = 15, height = 8), png = png(File, width
    FUN()
    dev.off()
  }
  else if(Fshape == shape_values[2]){
    switch(Ftype, pdf = pdf(File, width = 9, height = 7), png = png(File, width
    FUN()
    dev.off()
  }
}
```

Sample app run code (left) and a download function (above)



The R “Shiny” package enables development of web-based applications suitable for data analysis and visualization.

# You can find useful analysis apps and tools on the Test Science [Tools](#) webpage.



The screenshot shows the Test Science website header with the logo "TestScience Data . Driven . Defense" and a search bar. A navigation bar contains "LEARN", "TOOLS" (highlighted with a red box), "PARTICIPATE", "OUR RESEARCH", and "OUR TEAM". Below the navigation bar is a paragraph about the team's mission and a "Request Consult" button. The main content area features three columns: "Efficient Testing" with a gear icon, "Defensible Analyses" with a pencil and ruler icon, and "Insightful Results" with a lightbulb icon. Each column contains a brief description of the service.

**TestScience**  
Data . Driven . Defense

Q Type Search Term ...  
Subscribe ▾

LEARN ▾ **TOOLS ▾** PARTICIPATE ▾ OUR RESEARCH ▾ OUR TEAM ▾

The Test Science Team facilitates data-driven decision-making by developing, applying, and disseminating statistical, psychological, and data science methodologies within the Department of Defense and other national security organizations.

[Request Consult](#)

---

  
**Efficient Testing**

Our researchers design cost-effective and time-efficient tests, while also collecting ample data to support rigorous analyses. We leverage statistical and data science best practices such as Design of Experiments, sequential testing methods, and Bayesian techniques.

Our test planning process also incorporates human-system interaction methods to fully capture the user's experience. These approaches allow testers to use as few resources as possible during testing while maximizing information gain.

  
**Defensible Analyses**

Our high standard for data analysis produces undisputed results and ensures that our tested systems perform as advertised.

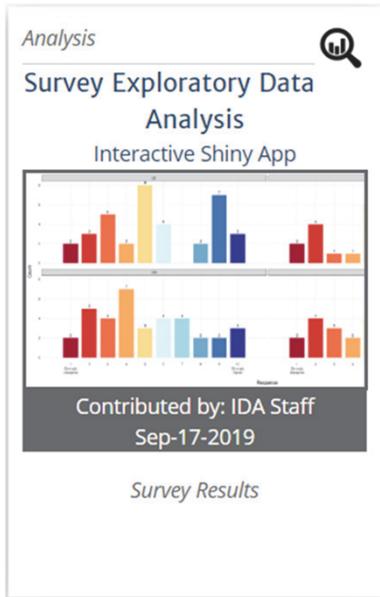
Our results are always supported by quantitative data and rigorous, reproducible analyses. This ensures our conclusions' objectivity and makes it easy to answer follow-up questions and discuss results with our sponsor, the T&E community, or system users. Good data management practices are at the core of our defensible analysis process.

  
**Insightful Results**

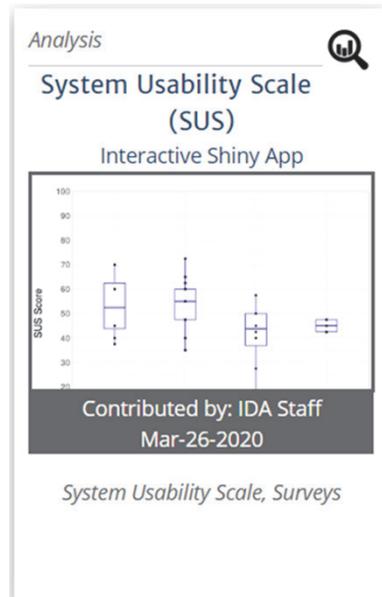
Government sponsors such as DOT&E and DHS continually rely on our insightful, accurate results, and our expertise in Test & Evaluation. We have created frameworks and guidebooks for the T&E community in areas such as Human-Machine Teaming and Modeling & Simulation Validation.

Our research products and training courses benefit military analysts, government action officers, IDA researchers, and others who desire to enhance their statistical and behavioral science knowledge.

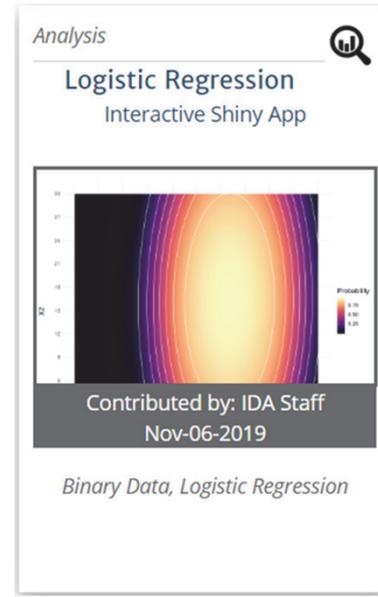
# Examples



[Survey Exploratory Data Analysis](#)



[System Usability Scale](#)



[Logistic Regression](#)



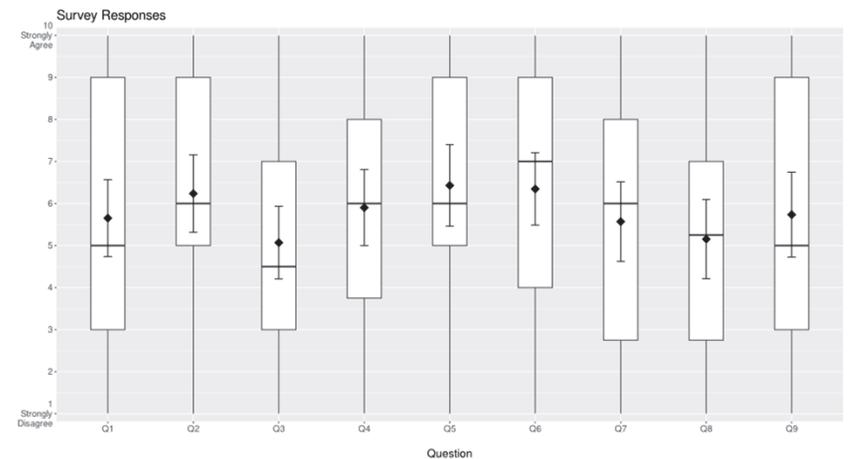
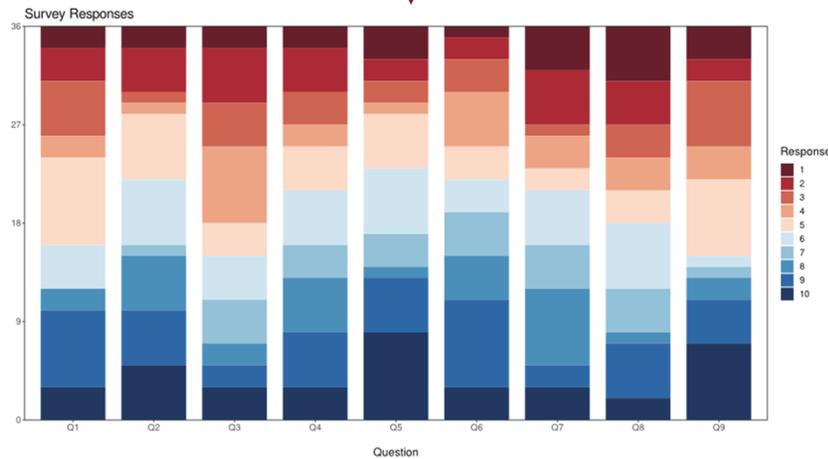
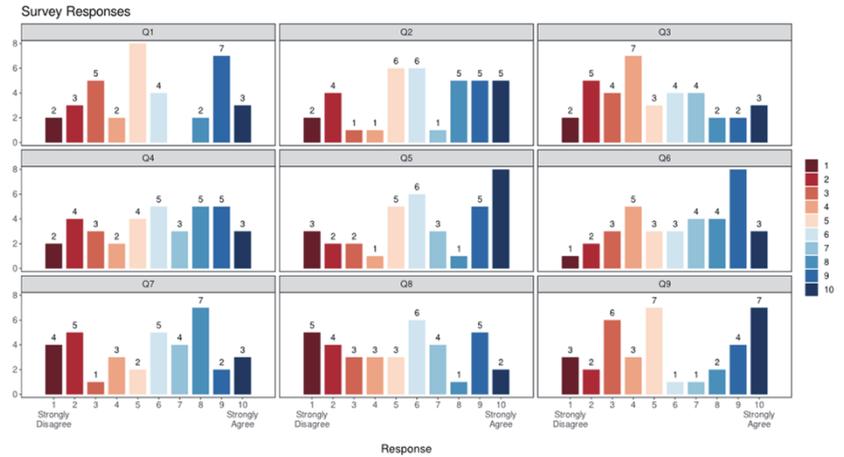
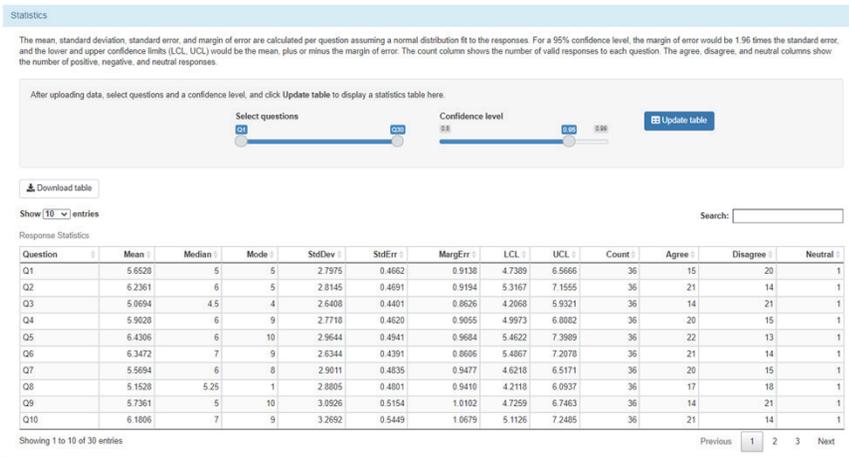
[Parametric Reliability Models](#)

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!

# Analyze Likert scale survey responses

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.

# Use the app to do early exploratory analysis of Likert-response survey data.<sup>1</sup>



Graphs show imaginary randomly generated data.

<sup>1</sup> 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.”

# Assess usability using the system usability scale (SUS)

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

# Use the app to analyze and plot system usability<sup>1</sup> data.

After uploading data, a table of SUS scores will appear. Each row of the displayed table is a single score for one survey respondent.

The SUS score,  $S$ , has a range [0, 100] and is calculated with the formula

$$S = 2.5 \times ((Q_1 - 1) + (5 - Q_2) + (Q_3 - 1) + (5 - Q_4) + (Q_5 - 1) + (5 - Q_6) + (Q_7 - 1) + (5 - Q_8) + (Q_9 - 1) + (5 - Q_{10}))$$

where  $Q_n$  has a domain of [1, 5].

Update table

Download table

Show 10 entries

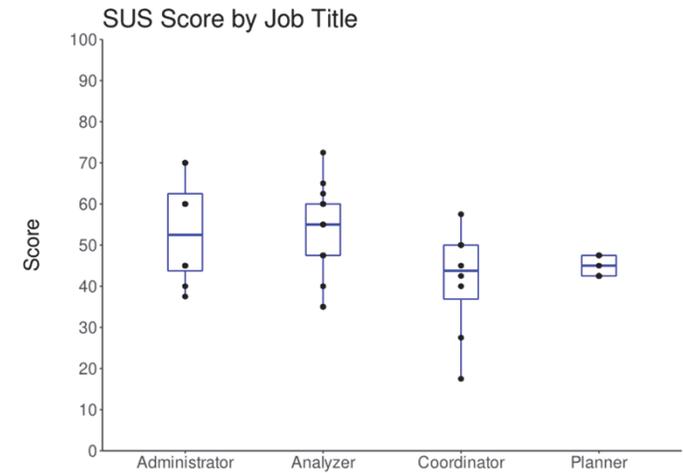
Search:

SUS Scores

Name	Date	Department	Section	Job	Experience	Govt_Exp	Score
Alex	2018-10-01	DoC	Administration	Administrator	11	30	40
Bill	2018-10-02	DoC	Administration	Administrator	30	12	37.5
Charlie	2018-10-03	DoC	Administration	Administrator	8	30	70
Dave	2018-10-04	DoC	Administration	Administrator	19	20	70
Eva	2018-10-05	DoC	Administration	Administrator	30	15	60
Frances	2018-10-06	DoC	Administration	Administrator	3	20	60
Gary	2018-10-07	DoC	Administration	Administrator	18	19	45
Holly	2018-10-08	DoC	Administration	Administrator	18	8	45
Ian	2018-10-09	DoC	Administration	Planner	3	30	47.5
John	2018-10-01	DoC	Administration	Planner	6	29	42.5

Showing 1 to 10 of 36 entries

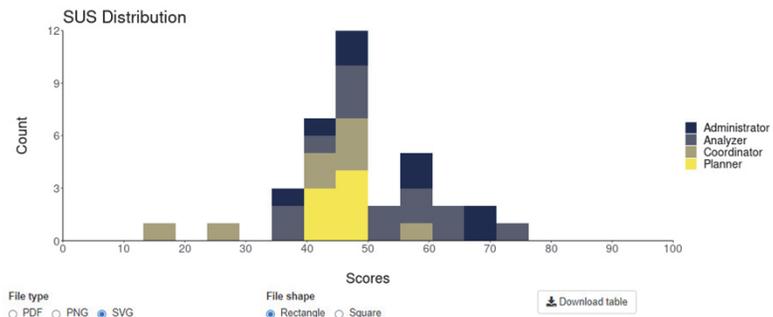
Previous 1 2 3 4 Next



Plot theme: Axis, Box, Fill, Grid

Color theme: viridis, magma, plasma, inferno, cividis

Reverse color:  No,  Yes



File type:  PDF,  PNG,  SVG

File shape:  Rectangle,  Square

Download table

Search:

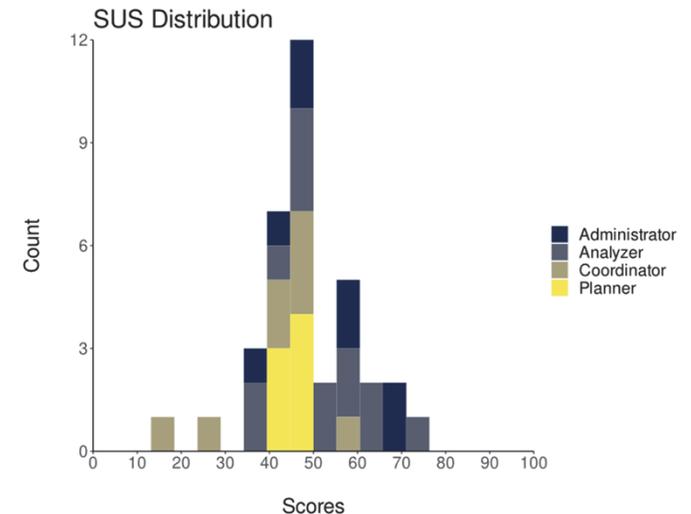
Show 10 entries

Summary Statistics

X	Mean	Median	Mode	StdDev	StdErr	MargErr	LCL	UCL	N
Administrator	53.438	52.500	70.000	13.157	4.652	9.117	44.320	62.555	8
Analyzer	52.500	55.000	47.500	11.637	3.227	6.326	46.174	58.826	13
Coordinator	41.250	43.750	50.000	13.025	4.605	9.025	32.225	50.275	8
Planner	45.000	45.000	47.500	2.500	0.945	1.852	43.148	46.852	7

Showing 1 to 4 of 4 entries

Previous 1 Next



## SUS – System Usability Scale

Graphs show imaginary randomly generated data.

<sup>1</sup> 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/>.

# 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).

# Use the app to build a logistic regression model of probability for one or two independent variables.

Statistics

The application fits a generalized linear model to the response variable, treating the one or two independent variables as factors  $x$  or  $x_1$  and  $x_2$ . Using the check boxes terms will not be included in the model (i.e.,  $\beta_{unselected} = 0$ ).  $P(x)$  models the probability of the dependent variable being equal to 1 as a function of the independent  $x$ .

The Estimate column in the table gives the  $\beta$  coefficients in the linear function

$$u(x) \approx \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4$$

The  $z$  value is the Wald statistic for the coefficient Estimate and is equal to the Estimate divided by the Standard Error. The  $p$ -value,  $\text{Pr}(|z|)$ , is the probability under the null hypothesis ( $\beta_{unselected} = 0$ ) that the data sample would produce a non-zero coefficient estimate by chance. Coefficient estimates are statistically non-zero with 95% confidence if  $\text{Pr}(|z|)$  is less than 0.05.

[Click here for a brief discussion on the output of the generalized linear model in R.](#)

$$P(x) = \frac{1}{1 + e^{-(-2.8617 + 0.3329x)}}$$

**Fit model**

2 variables detected.

Regression Variables

- Linear term?
- Quadratic term?
- Cubic term?
- Quartic term?

	Estimate	Std. Error	z value	Pr(> z )
$\beta_0$	-2.8617	0.9108	-3.1420	0.0017
$\beta_1$	0.3329	0.0877	3.7942	0.0001

Showing 1 to 2 of 2 entries

[Download table](#)

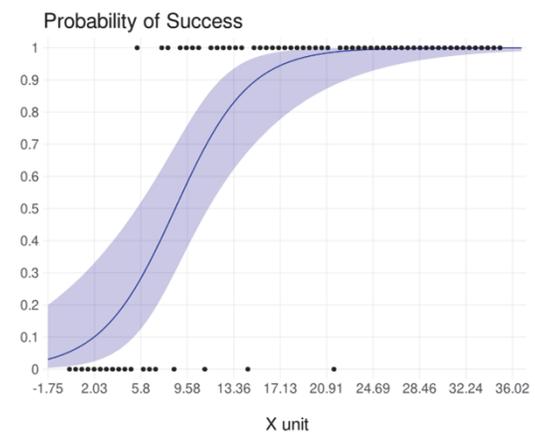
Null deviance: 80.4 on 70 degrees of freedom  
Residual deviance: 36.45 on 69 degrees of freedom

Deviance measures goodness of fit, and lower numbers indicate better fit. The null deviance is the deviance with no factors selected (i.e., only fitting  $\beta_0$ ), and the residual deviance is the deviance of the model including all selected parameters.

AIC: 40.45

The AIC is another measure of goodness of fit, and a lower value is better than a higher one. The AIC includes a model-complexity penalty to discourage fitting too many factors.

$$P(x) = \frac{1}{1 + e^{-(-2.8617 + 0.3329x)}}$$



Statistics

The application fits a generalized linear model to the response variable, treating the one or two independent variables as factors  $x$  or  $x_1$  and  $x_2$ . Using the check boxes, select model terms to specify the function  $P(x)$ . Unselected terms will not be included in the model (i.e.,  $\beta_{unselected} = 0$ ).  $P(x)$  models the probability of the dependent variable being equal to 1 as a function of the independent variable(s). Click **Fit model** to update the table.

The Estimate column in the table gives the  $\beta$  coefficients in the linear function

$$u(x_1, x_2) \approx \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2 + \beta_6 x_1^2 + \beta_7 x_2^2 + \beta_8 x_1^2$$

The  $z$  value is the Wald statistic for the coefficient Estimate and is equal to the Estimate divided by the Standard Error. The  $p$ -value,  $\text{Pr}(|z|)$ , is the probability under the null hypothesis ( $\beta_{unselected} = 0$ ) that the data sample would produce a non-zero coefficient estimate by chance. Coefficient estimates are statistically non-zero with 95% confidence if  $\text{Pr}(|z|)$  is less than 0.05.

[Click here for a brief discussion on the output of the generalized linear model in R.](#)

$$P(x_1, x_2) = \frac{1}{1 + e^{-(-24.5537 + 2.4154x_1 + 0.1767x_2 + 0.0005x_1x_2 - 0.0569x_1^2 - 0.0058x_2^2)}}$$

**Fit model**

3 variables detected.

Regression Variables

- X1 linear term?
- X2 linear term?
- Interaction term?
- X1 quadratic term?
- X2 quadratic term?
- X1 cubic term?
- X2 cubic term?
- X1 quartic term?
- X2 quartic term?

	Estimate	Std. Error	z value	Pr(> z )
$\beta_0$	-24.5537	2.0599	-11.9197	0.0000
$\beta_1$	2.4154	0.1903	12.6937	0.0000
$\beta_2$	0.1767	0.0647	2.7296	0.0063
$\beta_3$	0.0005	0.0022	0.2139	0.8306
$\beta_4$	-0.0569	0.0044	-12.8455	0.0000
$\beta_5$	-0.0058	0.0015	-3.9323	0.0001

Showing 1 to 6 of 6 entries

[Download table](#)

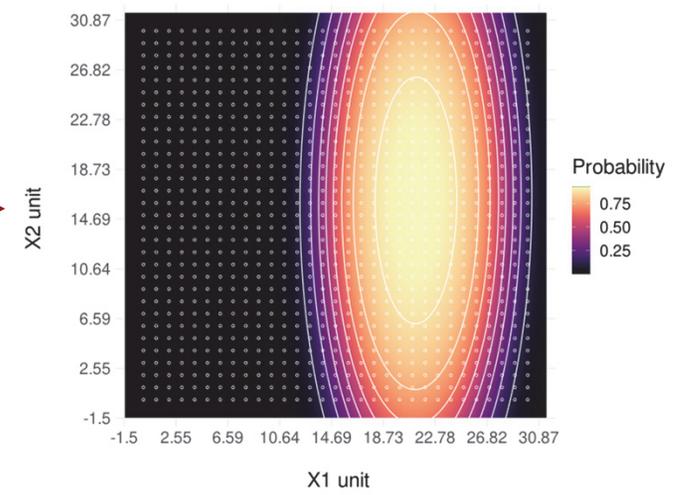
Null deviance: 1242.18 on 960 degrees of freedom  
Residual deviance: 564.66 on 955 degrees of freedom

Deviance measures goodness of fit, and lower numbers indicate better fit. The null deviance is the deviance with no factors selected (i.e., only fitting  $\beta_0$ ), and the residual deviance is the deviance of the model including all selected parameters.

AIC: 583.11

The AIC is another measure of goodness of fit, and a lower value is better than a higher one. The AIC includes a model-complexity penalty to discourage fitting too many factors.

$$P(x_1, x_2) = \frac{1}{1 + e^{-(-24.5537 + 2.4154x_1 + 0.1767x_2 + 0.0005x_1x_2 - 0.0569x_1^2 - 0.0058x_2^2)}}$$



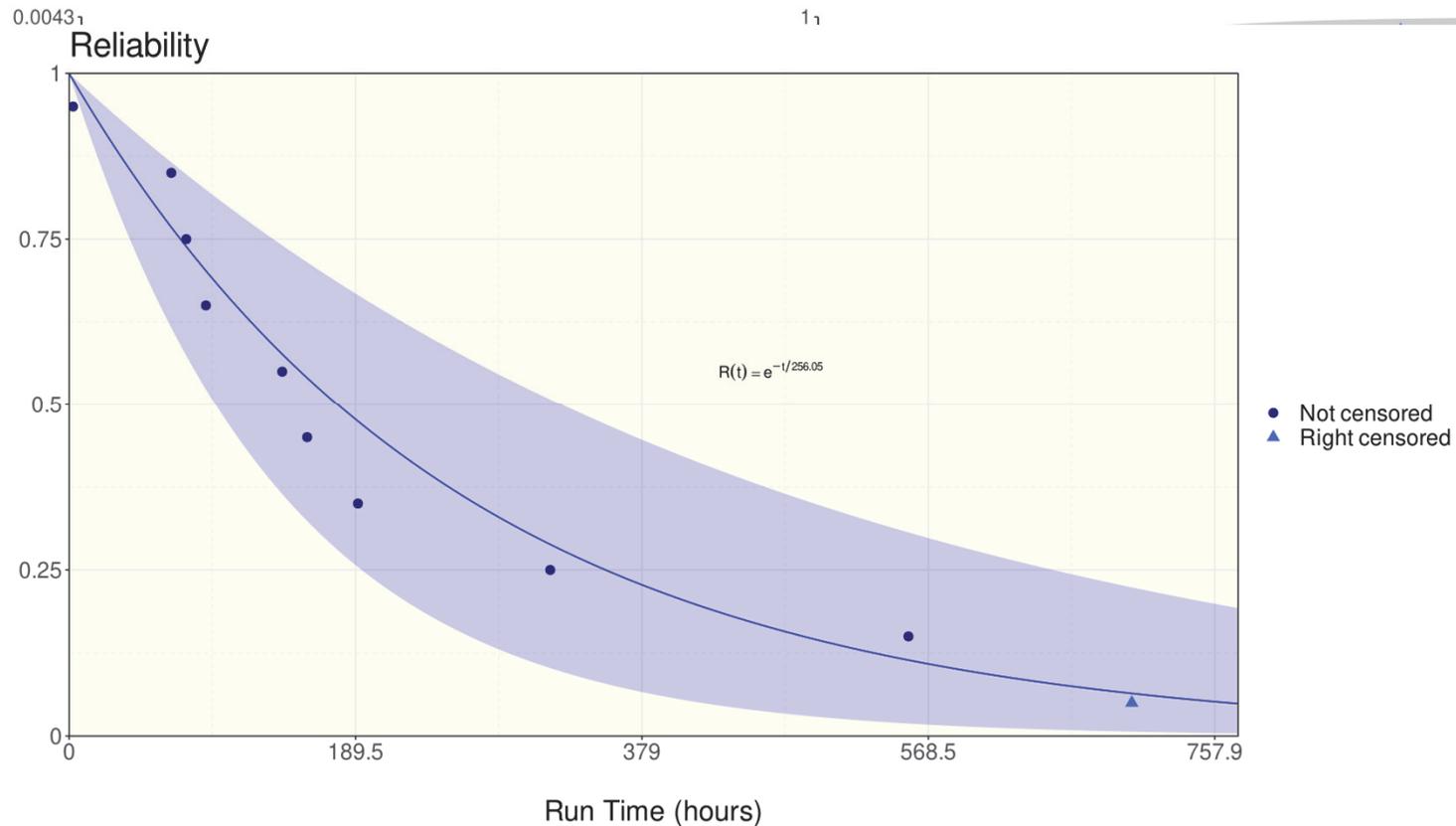
Graphs show imaginary randomly generated data.

# Model reliability using simple parametric distributions

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

# Use the app to compare, analyze, and plot parametric reliability models.<sup>1</sup>

The model comparison table contains information to help select the best-fitting model. Three distributions - the **exponential**, **Weibull**, and **lognormal** - can be compared using the K-S test, the AICc, the BIC, and the loglikelihood value. A nonsignificant K-S test means that the observed data are consistent with the fitted distribution. For information criteria (AICc, BIC, loglikelihood), a lower value indicates a better fit to the data, and a



Create final plot of reliability  
Compare models

Graphs show imaginary randomly generated data.

<sup>1</sup> 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.

**Use these apps to answer specific research questions, save time, improve aesthetics, and increase reproducibility of your results.**

Human Factors  Survey Exploratory Data Analysis  
(Often measured with Likert scale survey responses)

Usability  System Usability Scale  
(Measured with a specific Likert-like survey in which responses form a single-number metric or usability score)

Probability of Success  Logistic Regression  
(Estimated from binary event data: something either happens or does not)

Reliability  Parametric Reliability Models  
(Often estimated using service tickets and other sources of event times, durations, distances traveled, etc.)

## Why use the software tools presented here?

- 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

**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/](https://test-science.shinyapps.io/survey_data_analysis/)
- Code repository\*: [https://code.ida.org/projects/TSSHINY/repos/survey\\_data\\_analysis/browse](https://code.ida.org/projects/TSSHINY/repos/survey_data_analysis/browse)

## System Usability Scale

- Link: [https://test-science.shinyapps.io/system\\_usability/](https://test-science.shinyapps.io/system_usability/)
- Code: [https://code.ida.org/projects/TSSHINY/repos/system\\_usability/browse](https://code.ida.org/projects/TSSHINY/repos/system_usability/browse)

## Logistic Regression

- Link: [https://test-science.shinyapps.io/logistic\\_regression/](https://test-science.shinyapps.io/logistic_regression/)
- Code: [https://code.ida.org/projects/TSSHINY/repos/logistic\\_regression/browse](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](https://code.ida.org/projects/TSSHINY/repos/parametric_reliability_models/browse)

\* Source code repositories are only available for IDA-internal use at this time.

Use the Test Science software tools to improve the efficiency, aesthetics, and reproducibility of your analysis. Just bring your own web browser and simple text data files.

Most importantly, after analyzing the data you have, say *what you know* and *what you think* as directly and succinctly as possible.

Or else, no one will listen or care about your figures.

## Special thanks to:

Technical reviewers for this presentation:

Kelly Avery, Brian Conway, John Haman, Bram Lillard, Kelly Tran

Prior app reviewers:

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



# Backup

# [RELIABILITY ALTERNATE SLIDE] Use the app to compare, analyze, and plot parametric reliability models.

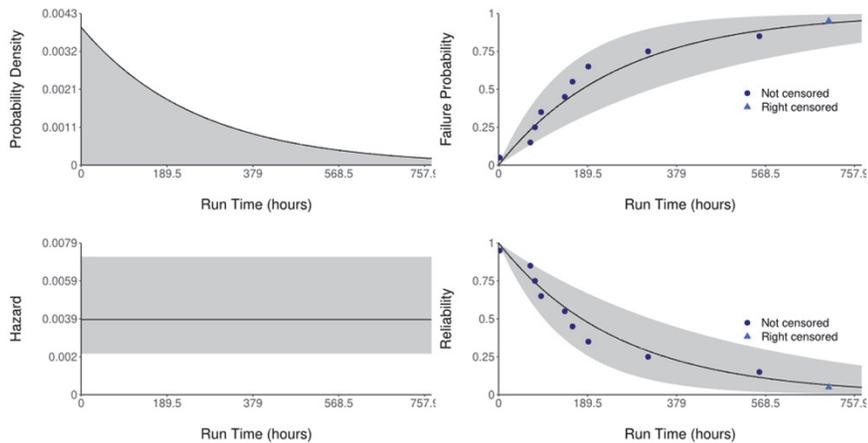
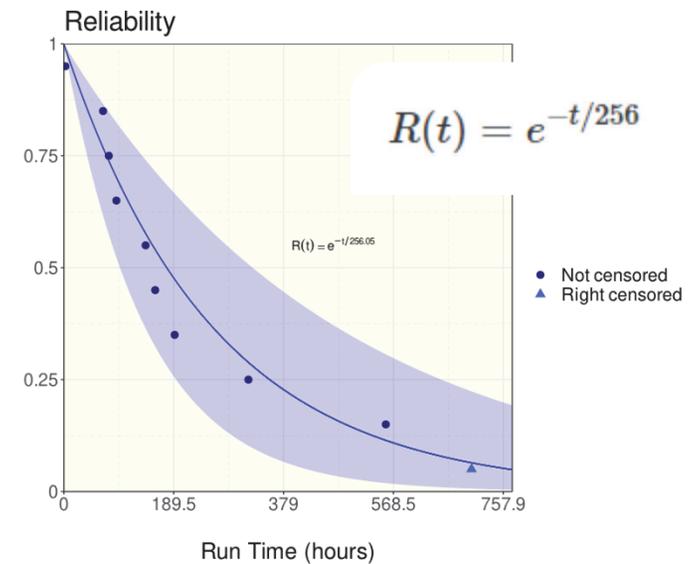
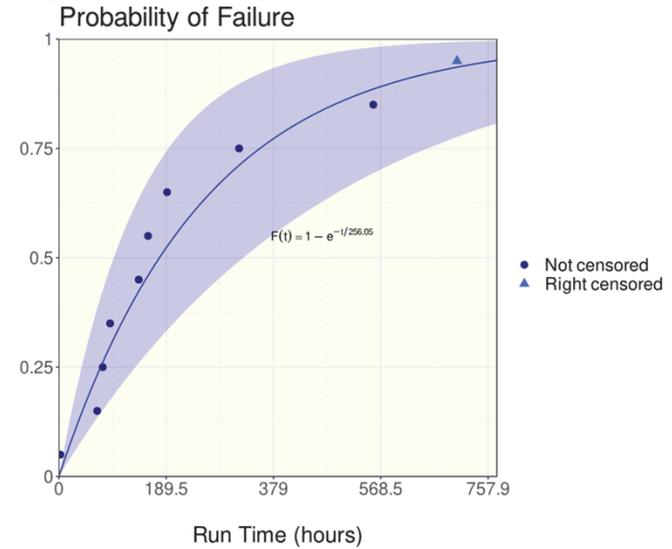
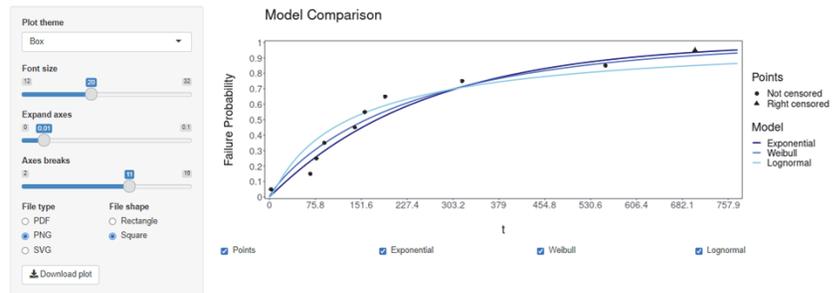
The model comparison table contains information to help select the best-fitting model. Three distributions - the exponential, Weibull, and lognormal - can be compared using the K-S test, the AICc, the BIC, and the loglikelihood value. A nonsignificant K-S test means that the observed data are consistent with the fitted distribution. For information criteria (AICc, BIC, loglikelihood), a lower value indicates a better fit to the data, and a difference of two or greater is considered substantial.

Download table

Search:

Distribution	Parameter 1	Parameter 2	Distribution Mean	K-S Statistic	K-S P-Value	AICc	BIC	-2*loglikelihood
exponential	1.000	256.052	256.052	0.174	0.873	120.317	120.120	117.817
weibull	0.857	245.999	265.332	0.182	0.837	123.191	122.682	117.477
lognormal	4.877	1.616	483.977	0.241	0.532	124.815	123.706	119.100

Showing 1 to 3 of 3 entries



Graphs show imaginary randomly generated data.

## **Appendix A**

### **Abbreviations and Acronyms**

---

DATAWorks	Defense and Aerospace Test and Analysis Workshop
DHS	Department of Homeland Security
DOD	Department of Defense
IDA	Institute for Defense Analyses
SUS	System Usability Scale



## References

---

Whitledge, William R., *Survey Exploratory Data Analysis*, IDA Web Application: [https://test-science.shinyapps.io/survey\\_data\\_analysis/](https://test-science.shinyapps.io/survey_data_analysis/).

Whitledge, William R., *Survey Exploratory Data Analysis Web Application*, Memorandum to Dr. Greg Zacharias and Dr. Ray O'Toole, DOT&E, May 9, 2019.

Whitledge, William R., *System Usability Scale*, IDA Web Application: [https://test-science.shinyapps.io/system\\_usability/](https://test-science.shinyapps.io/system_usability/).

Whitledge, William R., *System Usability Web Application*, Memorandum to Dr. Greg Zacharias and Dr. Ray O'Toole, DOT&E, February 13, 2020.

Bangor et al., *Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale*, Journal of Usability Studies, Vol. 4, Issue 3, May 2009, pp. 114-123, <https://uxpajournal.org/determining-what-individual-sus-scores-mean-adding-an-adjective-rating-scale/>.

Whitledge, William R., *Logistic Regression*, IDA Web Application: [https://test-science.shinyapps.io/logistic\\_regression/](https://test-science.shinyapps.io/logistic_regression/).

Whitledge, William R., *Logistic Regression Web Application*, Memorandum to Dr. Greg Zacharias and Dr. Ray O'Toole, DOT&E, October 7, 2019.

*Logistic Regression*, Wikipedia, [https://en.wikipedia.org/wiki/Logistic\\_regression](https://en.wikipedia.org/wiki/Logistic_regression).

Whitledge, William R., *Parametric Reliability Models*, IDA Web Application: <https://test-science.shinyapps.io/ParametricReliabilityModels/>.

Whitledge, William R., *Parametric Reliability Models Web Application*, Memorandum to Dr. Laura Freeman, DOT&E, June 19, 2018.

Whitledge, William R., Pinelis, Yevgeniya K., *Tutorial: Parametric Reliability Models*, IDA Non-Standard Document NS D-9171, September 2018, <https://www.ida.org/research-and-publications/publications/all/t/tu/tutorial-parametric-reliability-models>.

# 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 in test and evaluation work. These tools are available for use at IDA's Test Science 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!

[www.pngitem.com](https://www.pngitem.com/middle/shw/Twm_r-studio-icon-png-transparent-png/)



# Analysis Apps for the Operational Tester

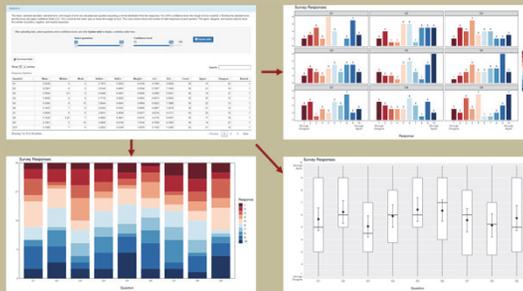
Use Test Science web apps to improve the efficiency, aesthetics, and reproducibility of your analysis.

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



## Analyze Likert scale survey responses<sup>1</sup>

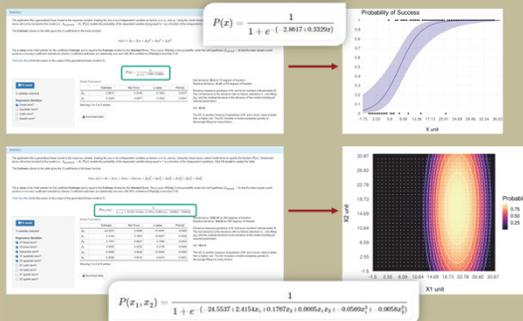
Quickly review and plot groups of Likert scale survey responses as column graphs, histograms, and box plots to assess user satisfaction, training adequacy, and other human factors.



<sup>1</sup> 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

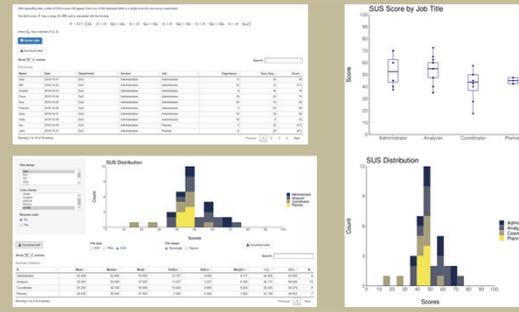
Fit a logistic regression model to one or two independent continuous variables and plot the probability of mission success, threat detection, target destruction, or other success (1) or failure (0).



All graphs on this poster show imaginary randomly-generated data.  
R logo downloaded on February 8, 2022 from [https://www.pngitem.com/middle/shw/Twm\\_r-studio-icon-png-transparent-png/](https://www.pngitem.com/middle/shw/Twm_r-studio-icon-png-transparent-png/).

## Assess usability using the system usability scale (SUS)<sup>2</sup>

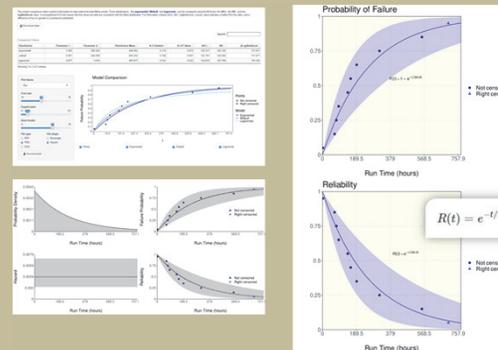
Review survey responses, calculate SUS scores, sort tables, and plot scores by independent variables to assess system usability.



<sup>2</sup> 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://www.researchgate.net/publication/221214141\\_The\\_SUS\\_questionnaire\\_a\\_simple\\_5-point\\_Likert-scale\\_usability\\_questionnaire](https://www.researchgate.net/publication/221214141_The_SUS_questionnaire_a_simple_5-point_Likert-scale_usability_questionnaire).

## Model reliability using simple parametric distributions<sup>3</sup>

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



<sup>3</sup> 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



## 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.

## 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.



## REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION

<b>1. REPORT DATE</b> 04-2022	<b>2. REPORT TYPE</b> Final	<b>3. DATES COVERED</b>	
		<b>START DATE</b>	<b>END DATE</b> Apr 2022
<b>4. TITLE AND SUBTITLE</b> DATAWorks 2022: Analysis Apps for the Operational Tester			
<b>5a. CONTRACT NUMBER</b> HQ0034-19-D-0001	<b>5b. GRANT NUMBER</b>	<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>5d. PROJECT NUMBER</b> BD-09-22990	<b>5e. TASK NUMBER</b> 229990	<b>5f. WORK UNIT NUMBER</b>	
<b>6. AUTHOR(S)</b> Whitledge, William, R.			
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> Institute for Defense Analyses 730 East Glebe Road Alexandria, Virginia 22305		<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b> NS D-32959 H 2022-000030	
<b>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> Director, Operational Test and Evaluation 1700 Defense Pentagon Room 1D548 Washington, DC 20301-1700		<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>	<b>11. SPONSOR/MONITOR'S REPORT NUMBER</b>
<b>12. DISTRIBUTION/AVAILABILITY STATEMENT</b> Public release approved. Distribution is unlimited.			
<b>13. SUPPLEMENTARY NOTES</b>			
<b>14. ABSTRACT</b> In the acquisition and testing world, data analysts repeatedly encounter certain categories of data, such as time or distance until an event (e.g., failure, alert, detection), binary outcomes (e.g., success/failure, hit/miss), and survey responses. Analysts need tools that enable them to produce quality and timely analyses of the data they acquire during testing. This poster presents four web-based apps that can analyze these types of data. The apps are designed to assist analysts and researchers with simple repeatable analysis tasks, such as building summary tables and plots for reports or briefings. Using software tools like these apps can increase reproducibility of results, timeliness of analysis and reporting, attractiveness and standardization of aesthetics in figures, and accuracy of results. The first app models reliability of a system or component by fitting parametric statistical distributions to time-to-failure data. The second app fits a logistic regression model to binary data with one or two independent continuous variables as predictors. The third calculates summary statistics and produces plots of groups of Likert-scale survey question responses. The fourth calculates the system usability scale (SUS) scores for SUS survey responses and enables the app user to plot scores versus an independent variable. These apps are available for public use on the Test Science Interactive Tools webpage <a href="https://testscience.org/interactive-tools/">https://testscience.org/interactive-tools/</a> .			
<b>15. SUBJECT TERMS</b> Survey; System Usability Scale (SUS); Reliability; Test Science; Logistic Regression; Binary Response; Likert Response; Interactive Web Application; Reproducibility; Shiny			
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>
<b>a. REPORT</b> Unclassified	<b>b. ABSTRACT</b> Unclassified	<b>c. THIS PAGE</b> Unclassified	SAR
			<b>18. NUMBER OF PAGES</b> 39
<b>19a. NAME OF RESPONSIBLE PERSON</b> Vincent Lillard			<b>19b. PHONE NUMBER</b> 703-845-2230