Automated Software Weibull Analysis & Reliability Modeling (SWARM) Tool

Rhonda Brown, Bob Stoddard

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

DM18-0289

Agenda

- Automation Tools Overview
- SEI Weibull modeling tool for prediction of DR closure dates
- Weibull Tool Operation
  - User interface
  - Input file
  - results
- What's happening in Minitab
- Download Instructions

Automation Tools Overview

Provide awareness of Visual Basic for Applications (VBA) tools in use in SEMA and CTSD that have been used for customers and projects.

Demonstrate the tools that automate processes and to provide more effective ways to perform data analysis tasks and visualization.

This presentation's focus is the Weibull Analysis and Reporting Tool.

Tools are available for stated purposes or can be adapted for wider use.

Why Use VBA Tools

Automate repetitive tasks

Reduce human error

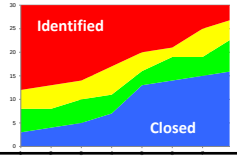
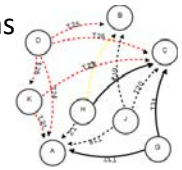
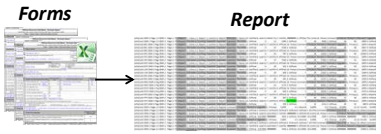
Perform tasks quickly

Compensate for limitations of other tools or to augment other tools

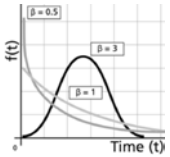
Perform specialized functions

- Reformat/normalize/transform data
- Extract data from forms & documents
- Generate technical reports
- Automate reporting processes
- Data Visualization
- Process a large number of files

VBA Tools in Use at SEI

Tool	Description
Weibull Analysis 	Improves efficiency in status reporting by replacing manual data processing and analysis with automation.
Cumulative Flow Diagram (CFD) Generator 	Automates the collection of information about the time of events and transactions from a change management workflow database so that activities can be understood through a CFD visualization. In addition, finds durations that each DR remained in each state for prediction of defect closure dates using Weibull analysis algorithm to acquire predictions of date ranges for closure of defect reports.
Database Visualizations 	A data repository that includes a feature for exploring the data represented in tables using the D3 open source graph visualization library. Visualizations include node graphs, hierarchy charts, stacked bar charts, and sunburst diagrams and others. This feature builds the visualization in a web browser based on database parameters sent to a JSON file that is read by a D3 html file. Graph visualizations use a force-directed layout algorithm to draw a graph unfolded and detangled.
Dependency Structure Matrix (DSM) Tool 	Streams in data from an Excel spreadsheet, builds a matrix, and provides a selection of options for matrix manipulation. Makes use of linked lists to build and reorder a data structure (graph) representing the given matrix. The DSM tool produces visualizations using the D3 open source graph visualization library.
Data Scraper 	A VBA program that captures all data from Software Resource Data Report (SRDR) forms into an Excel file in preparation for entering the data into a database.

VBA Tools in Use at SEI

Tool	Description
Weibull Analysis & Reporting 	Automates the prediction of software defect closure rates and reliability growth with Weibull Modeling. Effort normalized data (time points & defect counts) is streamed to a statistical tool (Minitab). A Distribution ID Plot is run, confirming the Weibull goodness of fit. Then the distribution analysis is conducted which gives additional desired output tables & parameters. This information is used for calculating the prediction in terms of dates to close a specified percentages of modeled defects. Comparisons can be made among releases, determining factors such as the highest increasing rate of closure across time.

Weibull Analysis & Reporting

Automates the prediction of software defect closure rates and reliability growth with Weibull Modeling. Effort normalized data (time points & defect counts) is streamed to a statistical tool (Minitab)

A Distribution ID Plot confirms Weibull is a good fit.

Then the distribution analysis is conducted which gives additional desired output tables & parameters. This information is used for calculating the prediction in terms of dates to close a specified percentages of modeled defects.

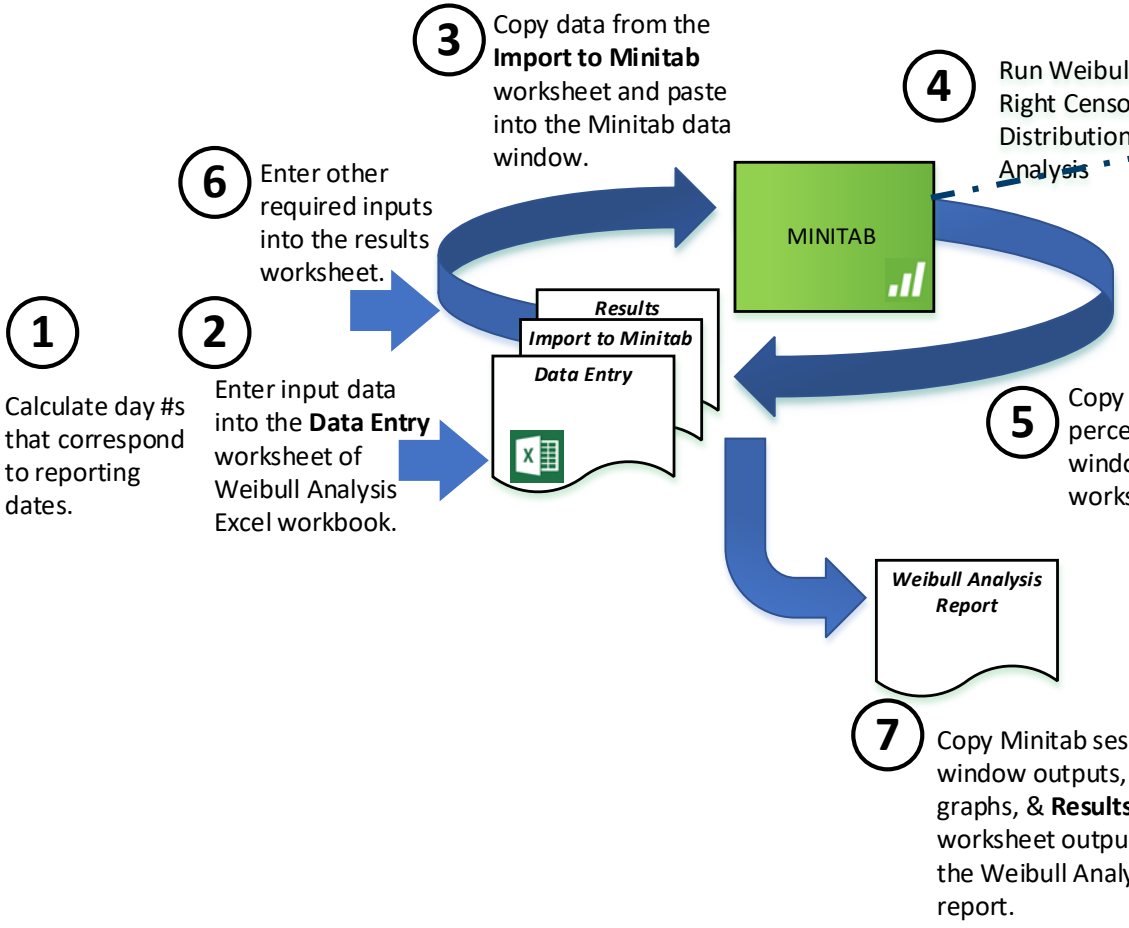
Comparisons can be made among releases, determining factors such as the highest increasing rate of closure across time.

Approach to SEI Weibull Method for Predicting Defect Report (DR) Closure Dates

- Weibull modeling is the default choice by hardware reliability engineers in fitting a statistical distribution to life data without having to guess which distribution best fits the data.
- The fitting process involves calculating a “shape” parameter from the input data which determines the best fitting distribution. Predictions are made about the shape of the Weibull output curve.
- Number of DR closures from past weeks are used to model when the remaining DRs will be closed.
- The model produces a 95% estimate of the percentage of DRs to be closed at any future date along with a 95% confidence interval.
- The confidence interval is then used to calculate how many work days it will take to close different cumulative percentages of the DRs.
  - The model excludes weekends and user-specified holidays.
  - Equal effort is applied each week in the past for DR closure.
  - Future weeks will incur a constant effort similar to the past toward DR closure.

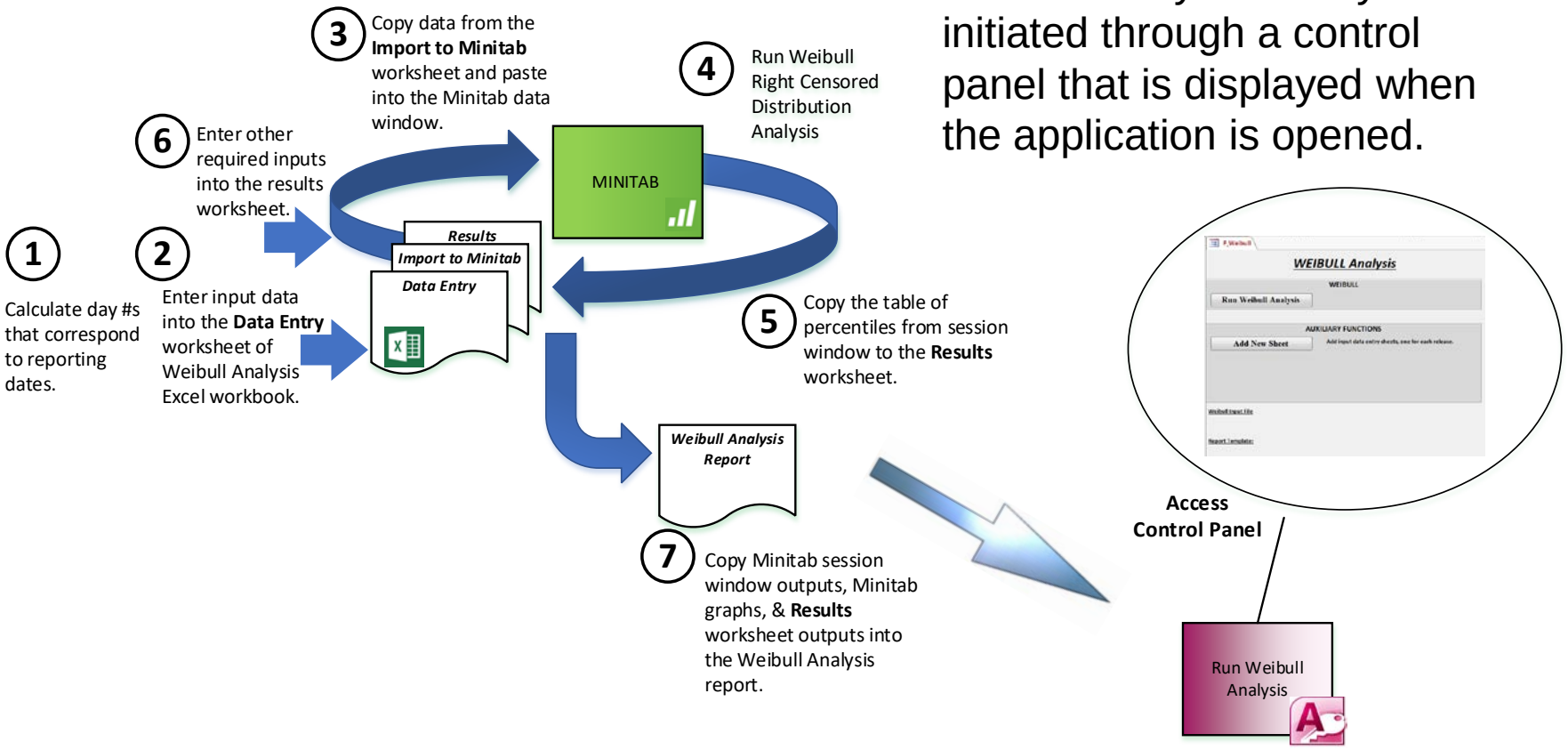
  If these assumptions are incorrect, we have the ability to re-run the Weibull using different assumptions.

Weibull Analysis & Reporting Process

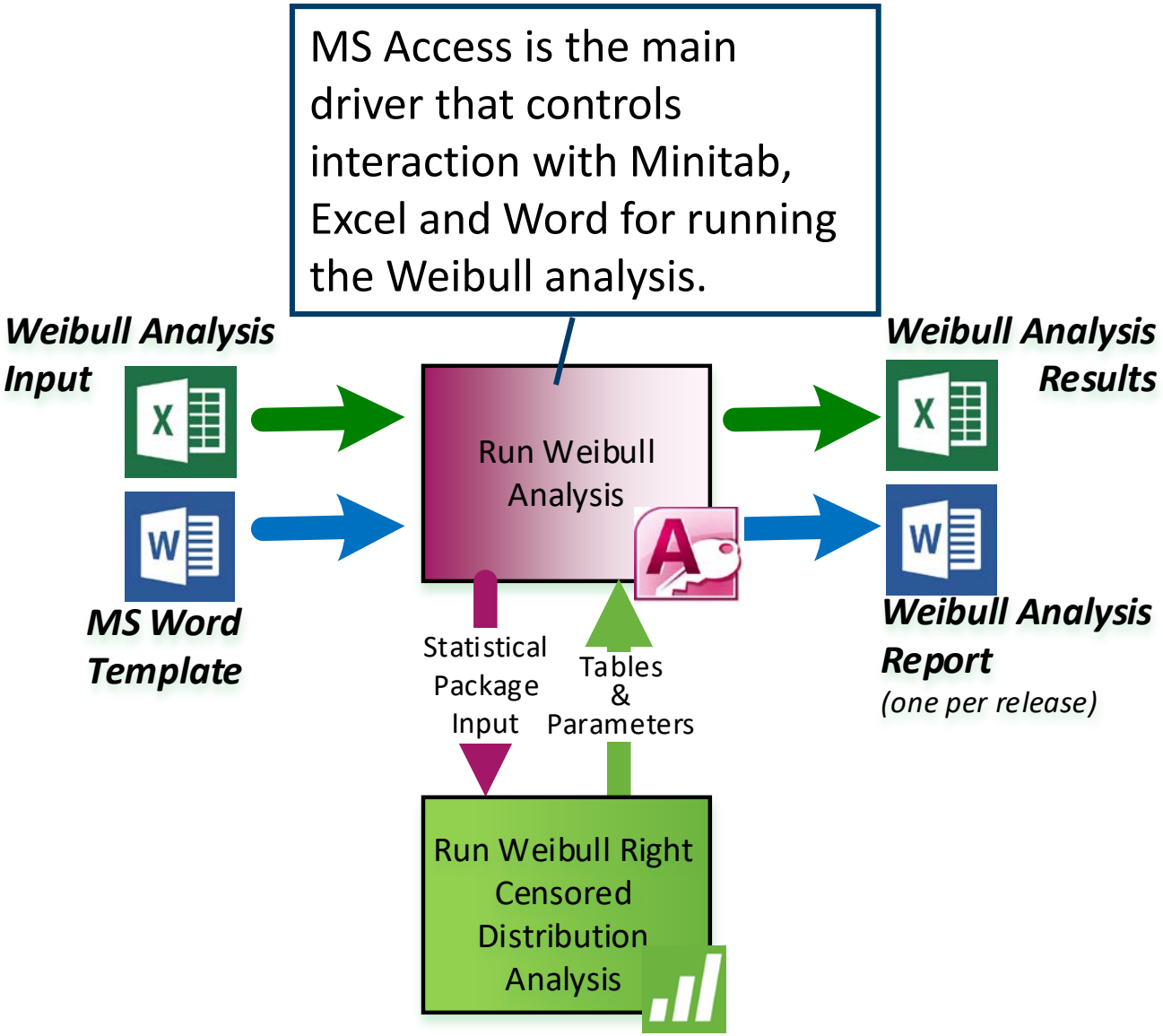


Weibull Automation Tool

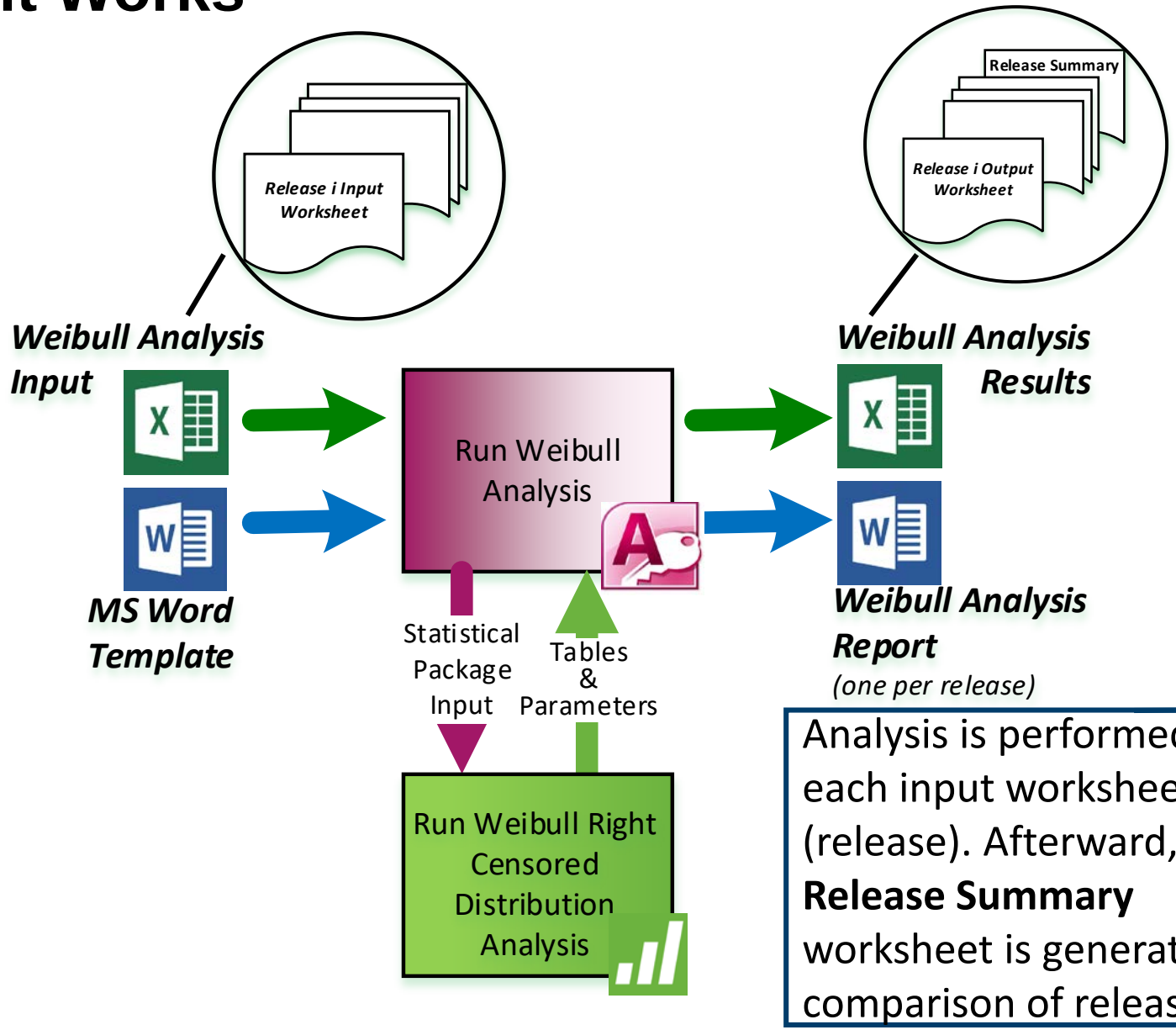
The Weibull Automation tool is an MS Access application that executes all steps for Weibull analysis. Analysis is initiated through a control panel that is displayed when the application is opened.



How it Works

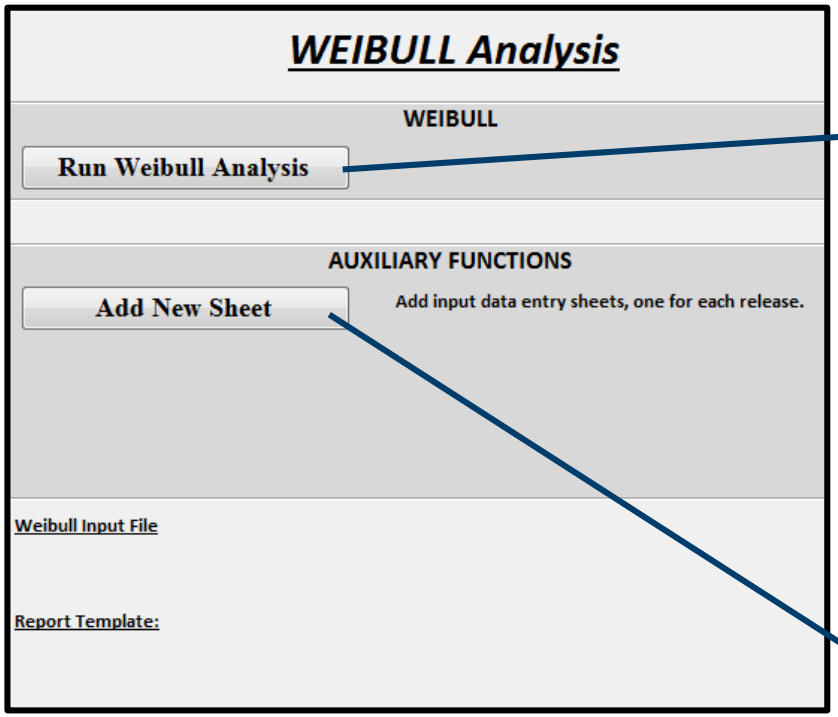


How it Works



Analysis is performed for each input worksheet (release). Afterward, a **Release Summary** worksheet is generated for comparison of releases.

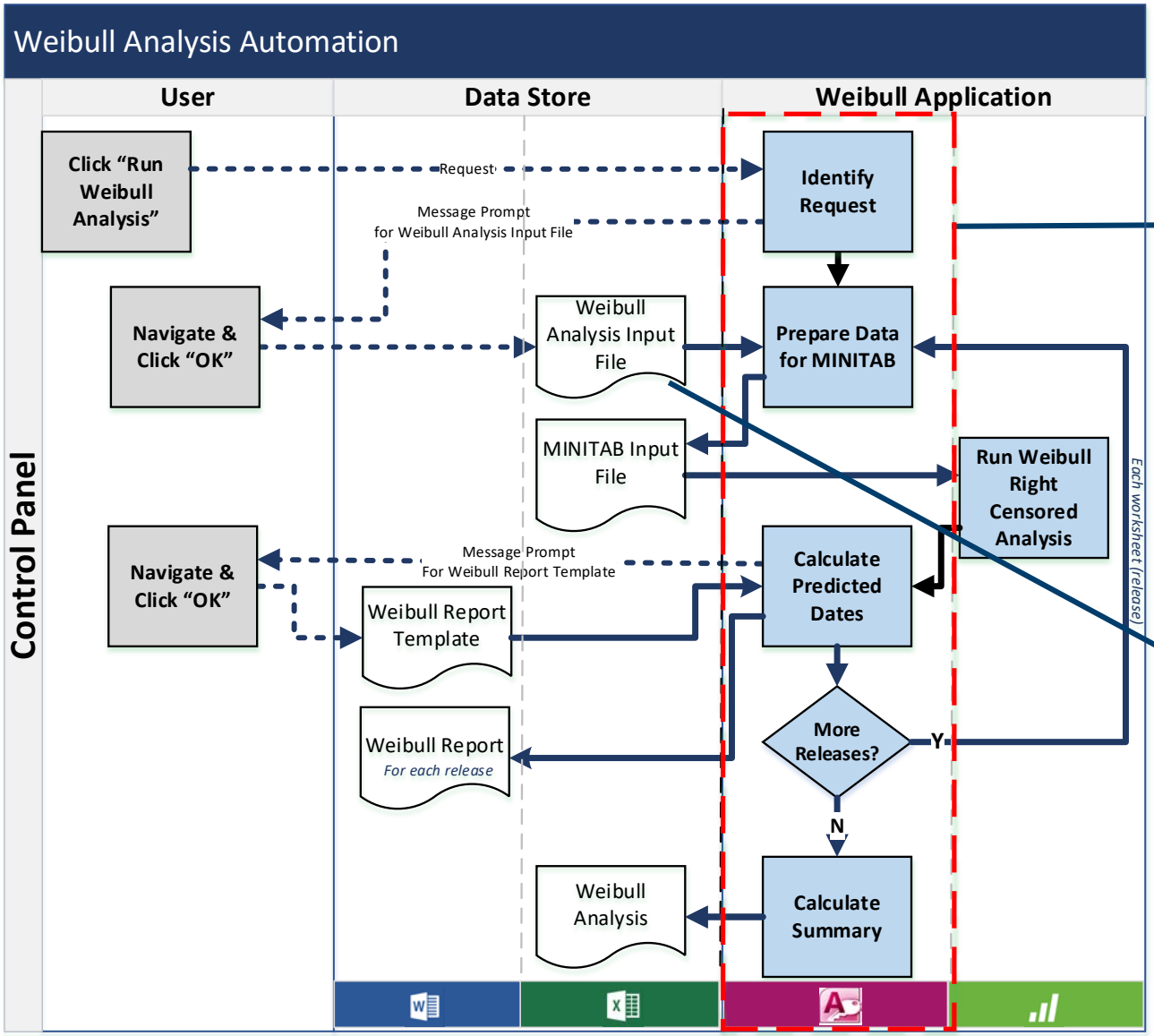
User Interface Control Panel



Run Weibull Analysis initiates the Weibull modeling, prompting the user for the Weibull analysis input file and the Weibull report template.

The tool operates on multiple populations or releases so that comparisons of the results may be made. The **Add New Sheet** function provides ease of adding additional worksheets to the input file template, one for each release to be analyzed. This prevents the user from having to manually add sheets, define named ranges, and cut & paste.

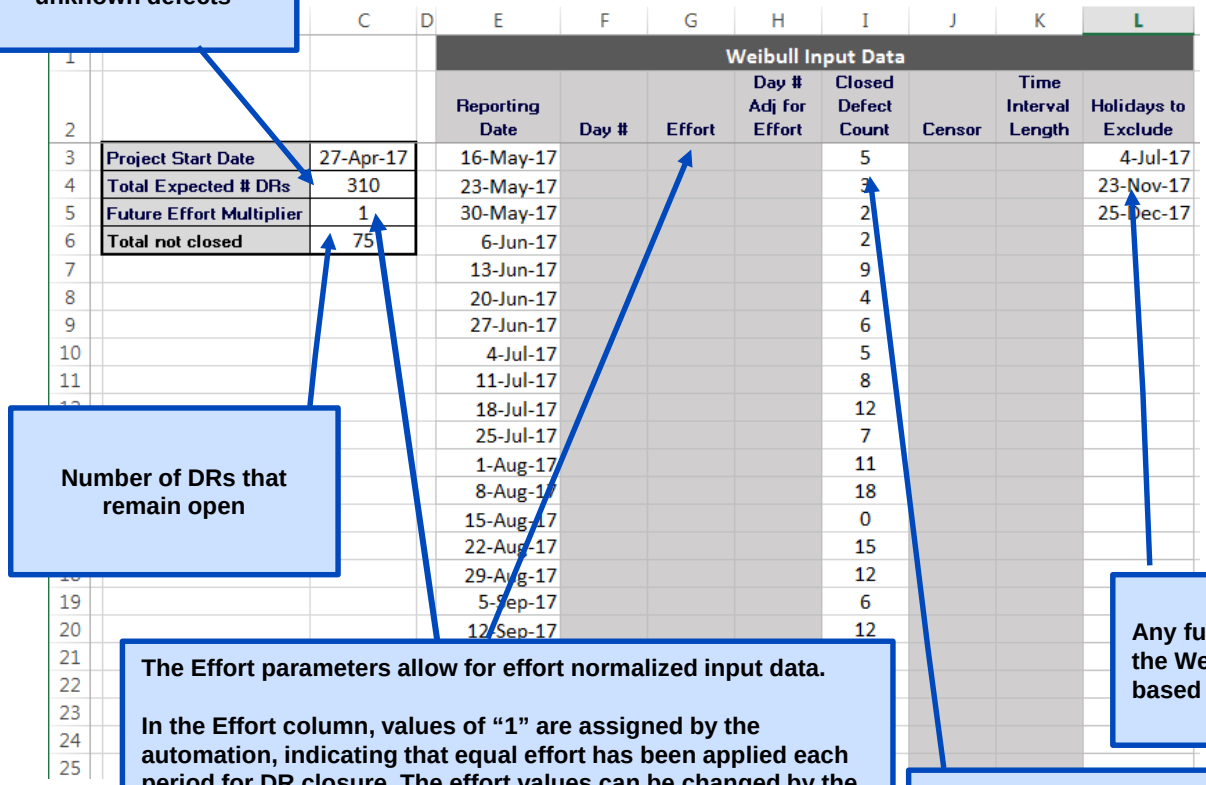
Weibull Analysis Automation Steps



MS Access controls process execution and makes calls to other applications as needed.

The user is prompted to select the Weibull input file which contains one worksheet per release that is being analyzed.

Weibull Analysis Input



	C	D	E	F	G	H	I	J	K	L
1			Weibull Input Data							
2			Reporting Date	Day #	Effort	Day # Adj for Effort	Closed Defect Count	Censor	Time Interval Length	Holidays to Exclude
3	Project Start Date 27-Apr-17		16-May-17				5			4-Jul-17
4	Total Expected # DRs 310		23-May-17				3			23-Nov-17
5	Future Effort Multiplier 1		30-May-17				2			25-Dec-17
6	Total not closed 75		6-Jun-17				2			
7			13-Jun-17				9			
8			20-Jun-17				4			
9			27-Jun-17				6			
10			4-Jul-17				5			
11			11-Jul-17				8			
12			18-Jul-17				12			
13			25-Jul-17				7			
14			1-Aug-17				11			
15			8-Aug-17				18			
16			15-Aug-17				0			
17			22-Aug-17				15			
18			29-Aug-17				12			
19			5-Sep-17				6			
20			12-Sep-17				12			
21										
22										
23										
24										
25										

Total expected number of defects including unknown defects

Number of DRs that remain open

The Effort parameters allow for effort normalized input data.

In the Effort column, values of "1" are assigned by the automation, indicating that equal effort has been applied each period for DR closure. The effort values can be changed by the user in future runs.

The future effort multiplier is usually the same relative effort as in historical time periods. To pull in dates, the future effort multiplier applied to defect closure can be increased.

Counts of the DR closures in the reporting period/ Report 0 count in an interval as needed.

Any future holidays listed enables the Weibull model to predict dates based on future work days only.

The user completes the non-shaded areas of the Weibull Analysis Input template, reporting DR closures in batch form at different time intervals.

The shaded columns are computed by the automation in preparation for Minitab.

Historical vs. Future Effort Parameters

Knowledge of effort in a given week will affect the prediction of DR closure. The automation treats a week that had 2 times the normal effort as 2 weeks of time. Therefore, that number of closed DRs are treated as if they were closed at the end of a 2 week period, therefore increasing the number of days in time. The day numbers represent standard time as opposed to calendar time.

A historical effort that is greater than one indicates that it took more effort to get the work done in the past, therefore pushing predicted dates out. Conversely if the future effort multiplier is greater than one, it will pull predicted dates in.

Weibull Analysis Input Example

	B	C
3	Project Start Date	27-Apr-17
4	Total Expected # DRs	310
5	Future Effort Multiplier	1
6	Total not closed	75

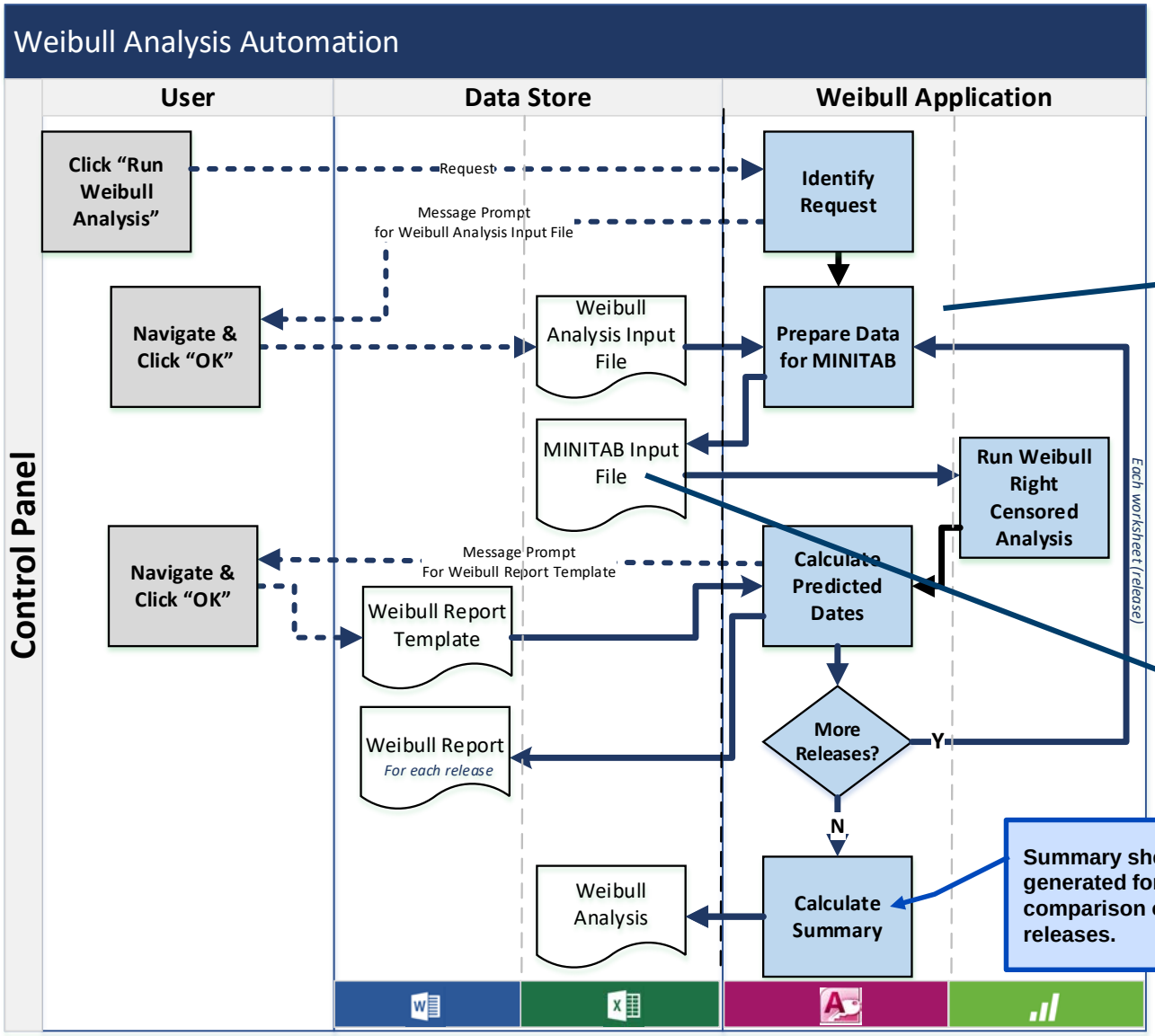
	E	F	G	H	I	J	K	L
1	Weibull Input Data							
2	Reporting Date	Day #	Effort	Day # Adj for Effort	Closed Defect Count	Censor	Time Interval Length	Holidays to Exclude
3	16-May-17				5			4-Jul-17
4	23-May-17				3			23-Nov-17
5	30-May-17				2			25-Dec-17
6	6-Jun-17				2			
7	13-Jun-17				9			
8	20-Jun-17				4			
9	27-Jun-17				6			
10	4-Jul-17				5			
11	11-Jul-17				8			
12	18-Jul-17				12			
13	25-Jul-17				7			
14	1-Aug-17				11			
15	8-Aug-17				18			
16	15-Aug-17				0			
17	22-Aug-17				15			
18	29-Aug-17				12			
19	5-Sep-17				6			
20	12-Sep-17				12			
21	19-Sep-17				19			
22	26-Sep-17				7			
23	3-Oct-17				5			
24	10-Oct-17				4			
25								

Assumptions:

1. Assume 75 DRs remain to be closed with no more unknown DRs to be found and opened

2. As of this week, 75% of the original projected 310 opened DRs have been identified (235); we assume no more unknown DRs are yet to be found and opened.

Preparation for Minitab



From the input data, Access calculates day numbers (adjusted for effort) and censor values.

A data file is generated for import into Minitab to conduct the Weibull Modeling.

Summary sheet is generated for comparison of releases.

Preparation for Minitab – Calculated Fields

	B	C
3	Project Start Date	27-Apr-17
4	Total Expected # DRs	310
5	Future Effort Multiplier	1
6	Total not closed	75

	E	F	G	H	I	J	K	L
1	Weibull Input Data							
2	Reporting Date	Day #	Effort	Day # Adj for Effort	Closed Defect Count	Censor	Time Interval Length	Holidays to Exclude
3	16-May-17	5	1	5	5	0	0	04-Jul-17
4	23-May-17	10	1	10	3	0	5	23-Nov-17
5	30-May-17	15	1	15	2	0	5	25-Dec-17
6	06-Jun-17	20	1	20	2	0	5	
7	13-Jun-17	25	1	25	9	0	5	
8	20-Jun-17	30	1	30	4	0	5	
9	27-Jun-17	35	1	35	6	0	5	
10	04-Jul-17	40	1	40	5	0	5	
11	11-Jul-17	45	1	45	8	0	5	
12	18-Jul-17	50	1	50	12	0	5	
13	25-Jul-17	55	1	55	7	0	5	
14	01-Aug-17	60	1	60	11	0	5	
15	08-Aug-17	65	1	65	18	0	5	
16	15-Aug-17	70	1	70	0	0	5	
17	22-Aug-17	75	1	75	15	0	5	
18	29-Aug-17	80	1	80	12	0	5	
19	05-Sep-17	85	1	85	6	0	5	
20	12-Sep-17	90	1	90	12	0	5	
21	19-Sep-17	95	1	95	19	0	5	
22	26-Sep-17	100	1	100	7	0	5	
23	03-Oct-17	105	1	105	5	0	5	
24	10-Oct-17	110	1	110	4	0	5	
25	10-Oct-17			110	75	1		

From the input data, Access calculates additional data columns that will be used by Minitab and the automation for predicting dates.

An additional row of data is also added, representing the total number of unclosed defects to date.

Input into Minitab

	A	B	C
1	time	count	censor
2	1	5	0
3	10	3	0
4	15	2	0
5	20	2	0
6	25	9	0
7	30	4	0
8	35	6	0
9	40	5	0
10	45	8	0
11	50	12	0
12	55	7	0
13	60	11	0
14	65	18	0
15	70	0	0
16	75	15	0
17	80	12	0
18	85	6	0
19	90	12	0
20	95	19	0
21	100	7	0
22	105	5	0
23	110	4	0
24	110	25	1
25			

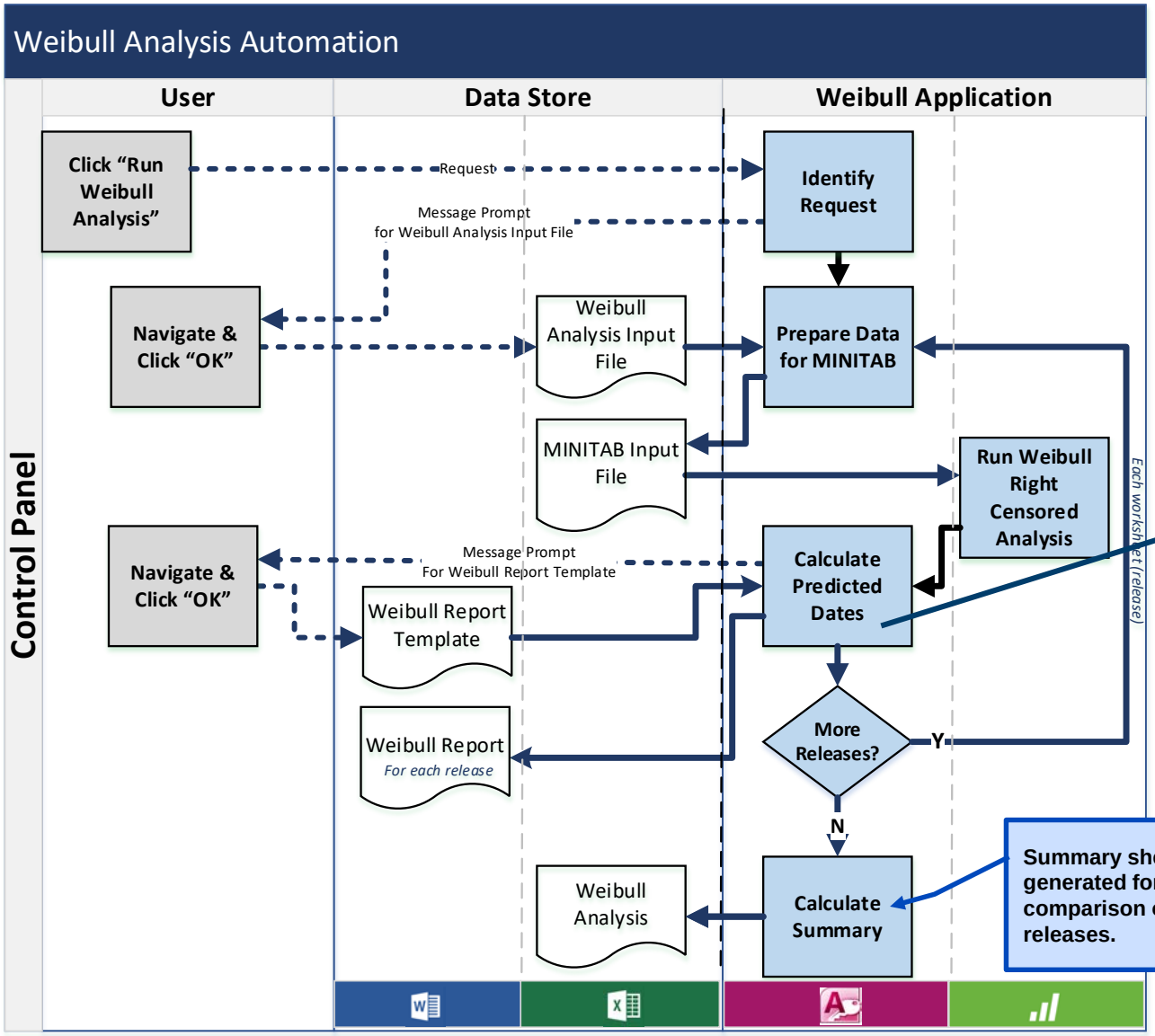
Access produces a statistical input package file containing the required fields to support the Minitab Weibull analysis.

From the Closed Defect Count and Censor columns of the Weibull input file.

Minitab requires time points that were calculated in the Day # column of the Weibull input file. Day numbers represent cumulative time, normally recorded in work days.

The final row of data representing the total number of unclosed defects to date is carried into Minitab. Cumulative day 110 is identified as the last of defects closed.

Minitab Outputs to Calculate Predicted Dates



Access captures the table of percentiles, shape and scale parameters, and graphs for the report (MS Word). Access uses data from Minitab to calculate predictions for remaining effort days based on the expected future effort multiplier. Then the number of workdays needed to closed cumulative percentages of DRs is calculated.

Summary sheet is generated for comparison of releases.

Predicted Defect Report (DR) Closure Dates

Weibull Input Data								95% CI of Number of Days to Close DRs					95% CI & Expected Future Effort Multiplier			Weibull Date Prediction					Parameter Estimates	
Reporting Date	Day #	Effort	Day # Adj for Effort	Closed Defect Count	Censor	Time Interval Length	Holidays to Exclude	Percent of DRs Closed	Number of DRs to be Fixed	Lower CI Days	Nominal Days	Upper CI Days	Lower CI Date	Expected Date	Upper CI Date	Percent of DRs Closed	Number of DRs to be Fixed	Lower CI Date	Expected Date	Upper CI Date	Shape Parameter	Scale Parameter
16-May-17	1	1	1	5	0	0	04-Jul-17	1	3	5.75346	8.08245	11.3542	5.75346	8.08245	11.3542	1	3	05/23/2017	05/26/2017	05/31/2017	1.81407	102.054
23-May-17	10	1	10	3	0	9	23-Nov-17	2	6	8.88868	11.8767	15.8632	8.88868	11.8767	15.8632	2	6	05/26/2017	05/31/2017	06/06/2017		
30-May-17	15	1	15	2	0	5	25-Dec-17	3	9	11.4757	14.8932	19.3286	11.4757	14.8932	19.3286	3	9	05/31/2017	06/05/2017	06/12/2017		
06-Jun-17	20	1	20	2	0	5		4	12	13.7661	17.5022	22.2524	13.7661	17.5022	22.2524	4	12	06/02/2017	06/08/2017	06/15/2017		
13-Jun-17	25	1	25	9	0	5		5	16	15.8624	19.8499	24.8398	15.8624	19.8499	24.8398	5	16	06/06/2017	06/12/2017	06/19/2017		
20-Jun-17	30	1	30	4	0	5		6	19	17.8188	22.0121	27.1923	17.8188	22.0121	27.1923	6	19	06/08/2017	06/15/2017	06/22/2017		
27-Jun-17	35	1	35	6	0	5		7	22	19.6683	24.0343	29.3696	19.6683	24.0343	29.3696	7	22	06/12/2017	06/19/2017	06/26/2017		
04-Jul-17	40	1	40	5	0	5		8	25	21.4328	25.9464	31.4105	21.4328	25.9464	31.4105	8	25	06/14/2017	06/20/2017	06/28/2017		
11-Jul-17	45	1	45	8	0	5		9	28	23.1278	27.7692	33.342	23.1278	27.7692	33.342	9	28	06/16/2017	06/22/2017	06/30/2017		
18-Jul-17	50	1	50	12	0	5		10	31	24.7649	29.518	35.1835	24.7649	29.518	35.1835	10	31	06/19/2017	06/26/2017	07/05/2017		
25-Jul-17	55	1	55	7	0	5		20	62	39.2395	44.6418	50.7879	39.2395	44.6418	50.7879	20	62	07/11/2017	07/18/2017	07/26/2017		
01-Aug-17	60	1	60	11	0	5		30	93	52.0735	57.8128	64.1847	52.0735	57.8128	64.1847	30	93	07/28/2017	08/04/2017	08/15/2017		
08-Aug-17	65	1	65	18	0	5		40	124	64.3691	70.4723	77.1542	64.3691	70.4723	77.1542	40	124	08/15/2017	08/23/2017	09/01/2017		
15-Aug-17	70	1	70	0	0	5		50	155	76.6872	83.3849	90.6675	76.6872	83.3849	90.6675	50	155	08/31/2017	09/11/2017	09/20/2017		
22-Aug-17	75	1	75	15	0	5		60	186	89.5442	97.2529	105.625	89.5442	97.2529	105.625	60	186	09/19/2017	09/29/2017	10/11/2017		
29-Aug-17	80	1	80	12	0	5		70	217	103.692	113.05	123.253	103.692	113.05	123.253	70	217	10/09/2017	10/23/2017	11/06/2017		
05-Sep-17	85	1	85	6	0	5		80	248	120.622	132.666	145.912	120.622	132.666	145.912	80	248	11/01/2017	11/17/2017	12/07/2017		
12-Sep-17	90	1	90	12	0	5		90	279	144.653	161.622	180.583	144.653	161.622	180.583	90	279	12/06/2017	01/01/2018	01/26/2018		
19-Sep-17	95	1	95	19	0	5		91	282	147.931	165.658	185.51	147.931	165.658	185.51	91	282	12/11/2017	01/05/2018	02/02/2018		
26-Sep-17	100	1	100	7	0	5		92	285	151.503	170.077	190.928	151.503	170.077	190.928	92	285	12/15/2017	01/12/2018	02/09/2018		
03-Oct-17	105	1	105	5	0	5		93	288	155.444	174.976	196.963	155.444	174.976	196.963	93	288	12/21/2017	01/18/2018	02/19/2018		
10-Oct-17	110	1	110	4	0	5		94	291	159.861	180.497	203.796	159.861	180.497	203.796	94	291	12/28/2017	01/26/2018	02/28/2018		
10-Oct-17			110	75	1			95	295	164.918	186.854	211.707	164.918	186.854	211.707	95	295	01/04/2018	02/05/2018	03/12/2018		
								96	298	170.885	194.402	221.156	170.885	194.402	221.156	96	298	01/12/2018	02/15/2018	03/26/2018		
								97	301	178.256	203.796	232.995	178.256	203.796	232.995	97	301	01/24/2018	02/28/2018	04/10/2018		
								98	304	188.111	216.466	249.095	188.111	216.466	249.095	98	304	02/07/2018	03/19/2018	05/03/2018		
								99	307	203.765	236.833	275.269	203.765	236.833	275.269	99	307	02/28/2018	04/16/2018	06/08/2018		

After running Weibull Right Censored Distribution Analysis in Minitab, the table of percentiles and shape and scale parameters from the Minitab session and calculation results from the automation are copied into the Excel worksheet.

Predicted Defect Report (DR) Closure Dates

95% CI of Number of Days to Close DRs				
Percent of DRs Closed	Number of DRs to be Fixed	Lower CI Days	Nominal Days	Upper CI Days
1	3	5.75346	8.08245	11.3542
2	6	8.88868	11.8767	15.8692
3	9	11.4757	14.8932	19.3286
4	12	13.7661	17.5022	22.2524
5	16	15.8624	19.8499	24.8398
6	19	17.8188	22.0121	27.1923
7	22	19.6683	24.0343	29.3696
8	25	21.4328	25.9464	31.4105
9	28	23.1278	27.7692	33.342
10	31	24.7649	29.518	35.1835
20	62	39.2395	44.6418	50.7879
30	93	52.0735	57.8128	64.1847
40	124	64.3691	70.4723	77.1542
50	155	76.6872	83.3849	90.6675
60	186	89.5442	97.2529	105.625
70	217	103.692	113.05	123.253
80	248	120.622	132.666	145.912
90	279	144.653	161.622	180.583
91	282	147.931	165.658	185.51
92	285	151.503	170.077	190.928
93	288	155.444	174.976	196.963
94	291	159.861	180.497	203.796
95	295	164.918	186.854	211.707
96	298	170.885	194.402	221.156
97	301	178.256	203.796	232.995
98	304	188.111	216.466	249.095
99	307	203.765	236.833	275.269

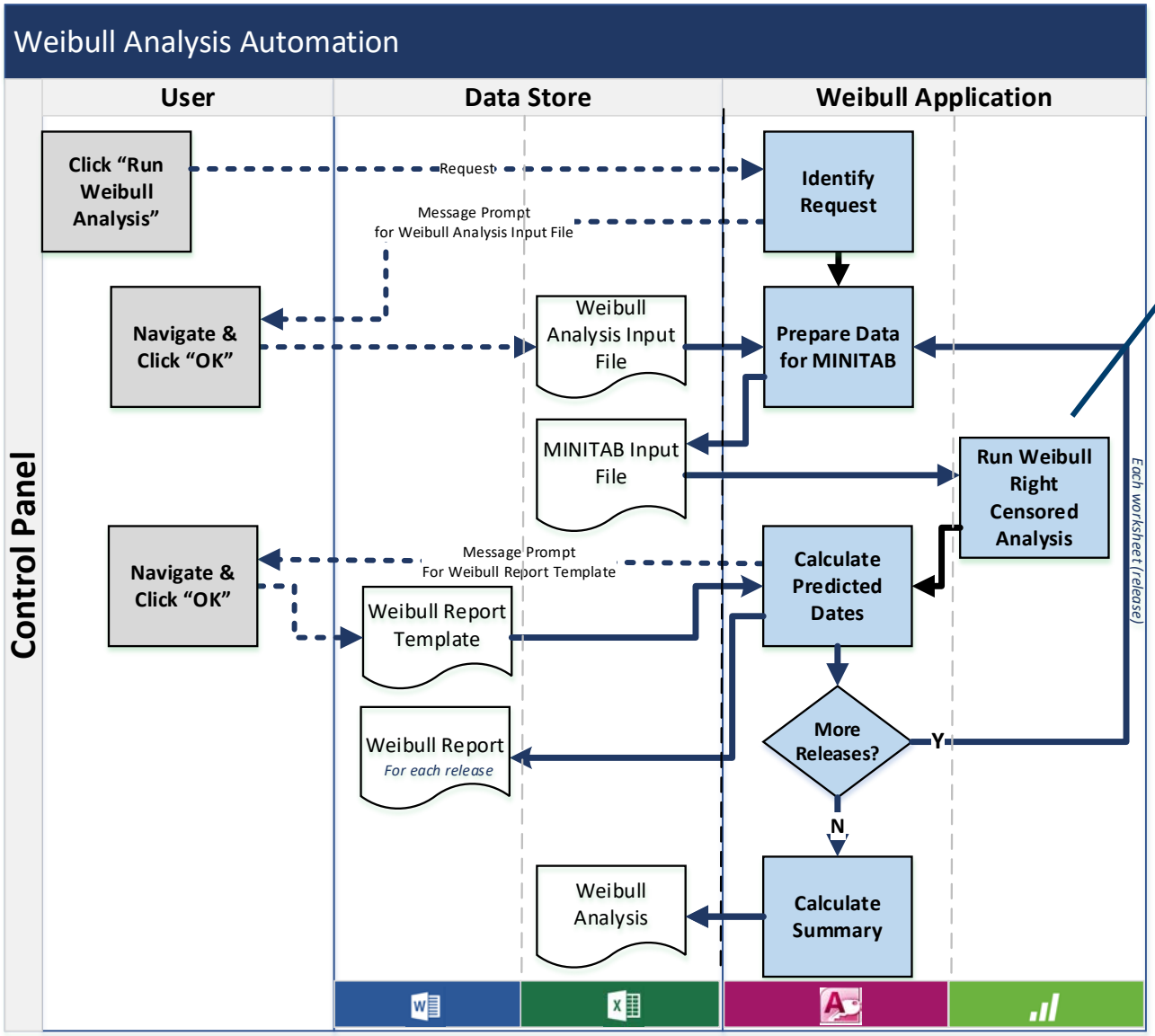
Weibull Date Prediction		
Lower CI Date	Expected Date	Upper CI Date
05/23/2017	05/26/2017	05/31/2017
05/26/2017	05/31/2017	06/06/2017
05/31/2017	06/05/2017	06/12/2017
06/02/2017	06/08/2017	06/15/2017
06/06/2017	06/12/2017	06/19/2017
06/08/2017	06/15/2017	06/22/2017
06/12/2017	06/19/2017	06/26/2017
06/14/2017	06/20/2017	06/28/2017
06/16/2017	06/22/2017	06/30/2017
06/19/2017	06/26/2017	07/05/2017
07/11/2017	07/18/2017	07/26/2017
07/28/2017	08/04/2017	08/15/2017
08/15/2017	08/23/2017	09/01/2017
08/31/2017	09/11/2017	09/20/2017
09/19/2017	09/29/2017	10/11/2017
10/09/2017	10/23/2017	11/06/2017
11/01/2017	11/17/2017	12/07/2017
12/06/2017	01/01/2018	01/26/2018
12/11/2017	01/05/2018	02/02/2018
12/15/2017	01/12/2018	02/09/2018
12/21/2017	01/18/2018	02/19/2018
12/28/2017	01/26/2018	02/28/2018
01/04/2018	02/05/2018	03/12/2018
01/12/2018	02/15/2018	03/26/2018
01/24/2018	02/28/2018	04/10/2018
02/07/2018	03/19/2018	05/03/2018
02/28/2018	04/16/2018	06/08/2018

Output tables from Weibull modeling in Minitab are used to obtain the DR closure rate and translate into calendar dates.

Assumptions:

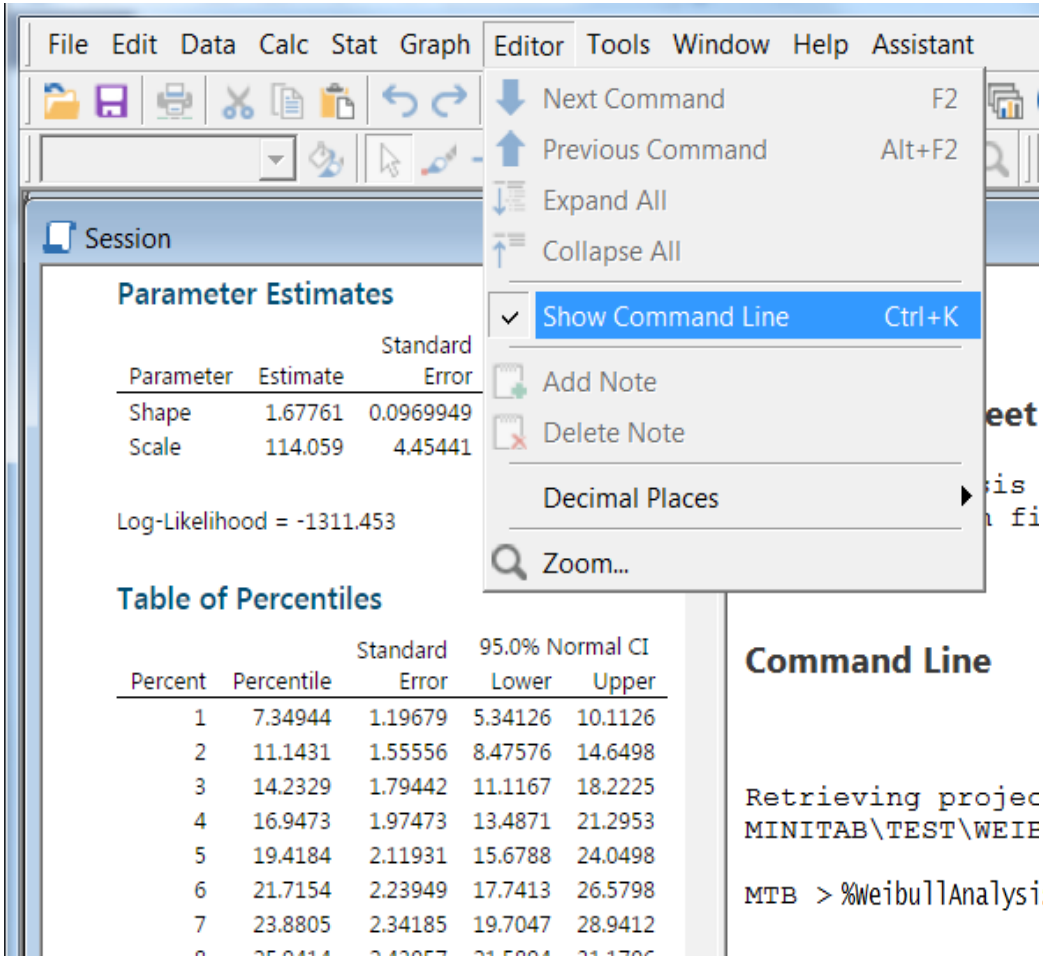
- Only work days are modeled (excludes weekends and specified holidays).
- Assumes constant man/months of effort from first reporting date to end of DR closure work.
- Average complexity of the DRs over each week is constant.
- With constant effort expect all DRs to close between 2/29/2018 to 6/8/2018.

What's Happening in Minitab?



Minitab runs macros to calculate a 95% estimate of the percentage of DRs to be closed at any future date.

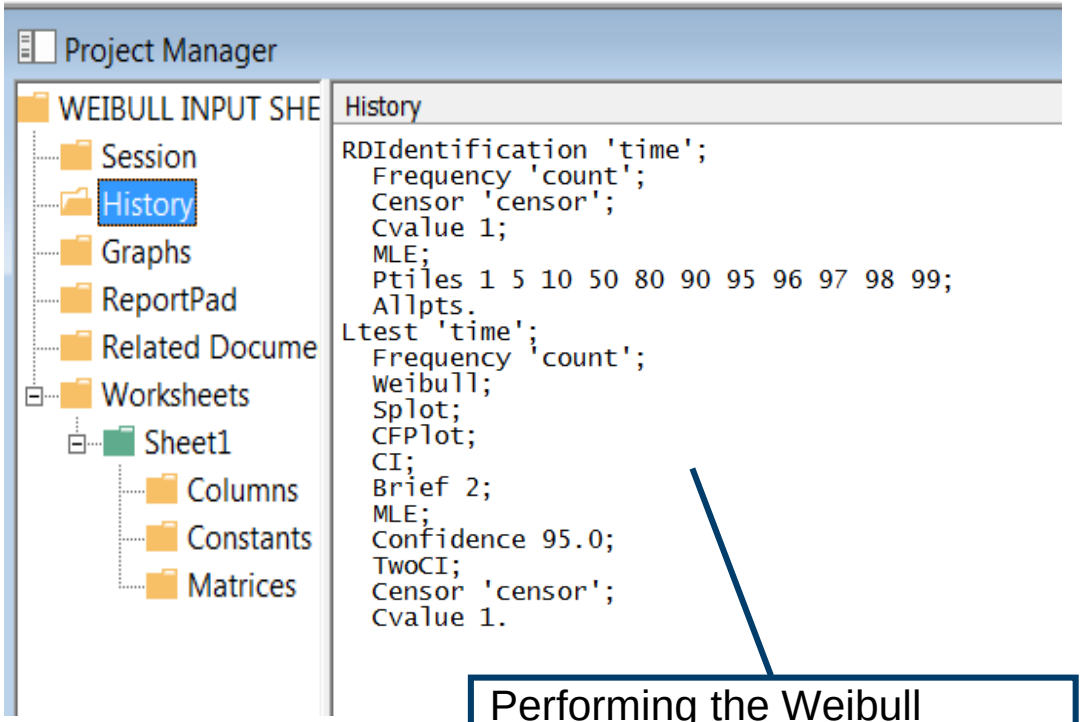
Minitab Macros



A Minitab macro can be manually invoked in the Session window by opening the Command Line Editor and typing % followed by the macro file name.

In the Weibull automation, Access instructs the Weibull macro to automatically run & process the data that was imported into Minitab. When complete, Access reads the Table of Percentiles & Parameter Estimates (shown in the session window) and uses the information to continue processing. The Minitab session is then saved, one .MPJ file per release.

Minitab Macro for Weibull Analysis



Performing the Weibull analysis manually produces these Minitab macro commands. These are the Minitab commands that are executed in the macro **WeibullAnalysis.mac**.

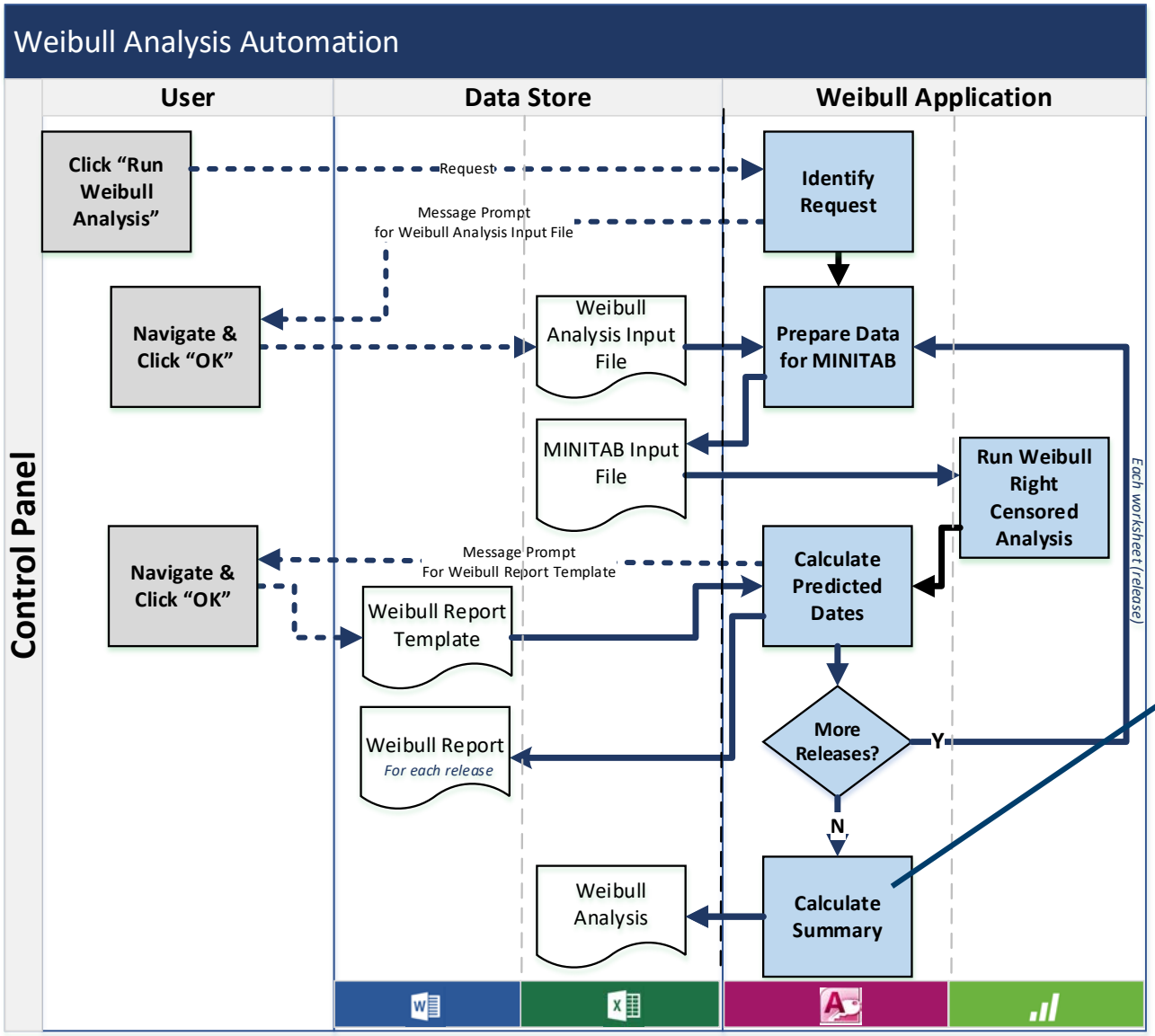
The History pane of the Project Manager window shows a recording, or log, of all actions done in a Minitab session.

Commands in this history pane were used to compile the Weibull analysis macro used by Access. The Weibull analysis macro commands perform the following:

Distribution ID Plot - to first confirm Weibull is a good if not best fit.

Parametric Distribution Analysis - which yields the desired output tables and parameter estimates.

Summary Sheet for Comparison of Releases



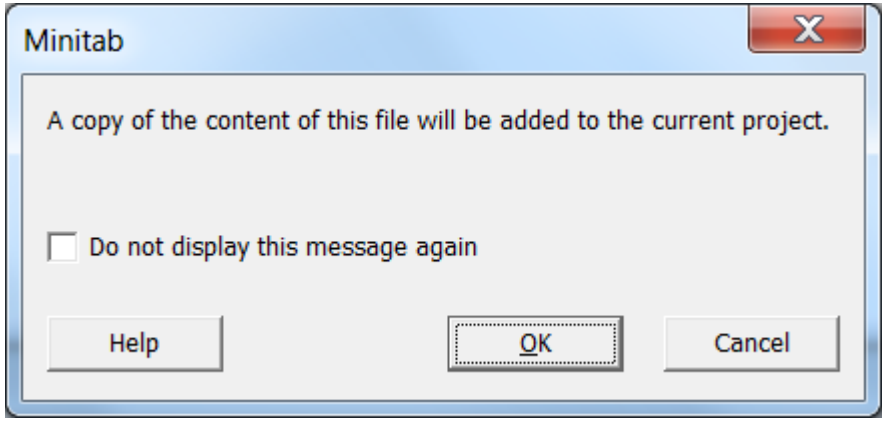
After Weibull analysis is completed for each release of the input file workbook, a summary sheet is generated listing line-by-line results from each release.

Download Instructions

1. Install Minitab from https://www.cmu.edu/computing/software/all/minitab/download.html. Installation will default to C:\Program Files (x86)\Minitab
2. Download WEIBULL AUTOMATION - MINITAB Download Files.zip
3. Unzip to your desktop or any area on your local drive.
   - Weibull Analysis Input Template.xlsx
   - WEIBULL Analysis MINITAB.mdb
   - Weibull Report Template.docx
   - WeibullAnalysis.mac
4. Use your administrator login to copy WeibullAnalysis.mac to C:\Program Files (x86)\Minitab\Minitab 18\English\Macros.
5. Open WEIBULL Analysis MINITAB.mdb

Additional Notes on Running the Weibull Analysis

1. Minitab will require acknowledgement that data is being imported into Minitab. Operation will be halted until the user clicks ok.



2. At times during the run of the Weibull automation tool, processing may halt & wait for the user to give permission to overwrite files from a previous run.