SENG 438 - Software Testing, Reliability, and Quality

Lab. Report #5 – Mutation Testing and Web app testing

Group #	1
Student Names:	Huda Abbas
	Nuha Shaikh
	Lubaba Sheikh
	Rajpreet Gill

1 Introduction

This lab enabled the team to concentrate on fully grasping the analysis techniques of integration testing using reliability assessment tools. The team learned about the two methods that are used to assess failure data including reliability growth testing and reliability assessment using reliability demonstration charts (RDC). The team further understood the importance of reliability growth model and the importance it serves in ensuring the software product is set for release. The testing tool that was used was C-SFRAT which was used to assess reliability growth testing in this lab.

Along with the fundamental reliability assessment knowledge the lab provided, it also gave us an essence into the strength and influence teamwork offers. Through strong collaboration and cooperation as well as by employing effective teamwork capabilities, we all ensured to maximize our expertise by taking part in active discussions throughout the lab and providing each other guidance and feedback frequently. Overall, this lab was a success, and the following report captures all of the components required for this lab assignment. The lab also provides reasoning and conclusions on how the requirements set in the lab document were met.

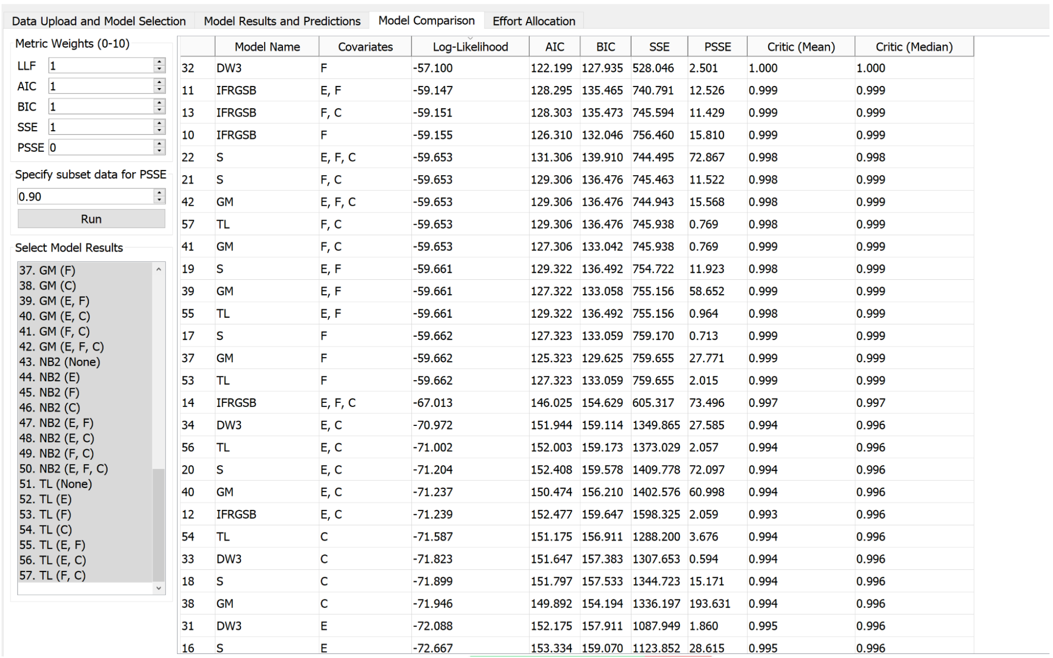
2 Assessment Using Reliability Growth Testing

Result of model comparison (selecting top two models) using C-SFRAT as our testing tool

The top two models that provide the best fit for the project data using model ranking are as follows:

- DW3 (F)
- IFRGSB (E, F)

We performed model ranking using highest log-likelihood values which represent the greatest goodness-of-fit



	Model Name	Covariates	Log-Likelihood	AIC	BIC	SSE	PSSE	Critic (Mean)	Critic (Median)
32	DW3	F	-57.100	122.199	127.935	528.046	2.501	1.000	1.000
11	IFRGSB	E, F	-59.147	128.295	135.465	740.791	12.526	0.999	0.999
13	IFRGSB	F, C	-59.151	128.303	135.473	745.594	11.429	0.999	0.999
10	IFRGSB	F	-59.155	126.310	132.046	756.460	15.810	0.999	0.999
22	S	E, F, C	-59.653	131.306	139.910	744.495	72.867	0.998	0.998
21	S	F, C	-59.653	129.306	136.476	745.463	11.522	0.998	0.999
42	GM	E, F, C	-59.653	129.306	136.476	744.943	15.568	0.998	0.999
57	TL	F, C	-59.653	129.306	136.476	745.938	0.769	0.998	0.999
41	GM	F, C	-59.653	127.306	133.042	745.938	0.769	0.999	0.999
19	S	E, F	-59.661	129.322	136.492	754.722	11.923	0.998	0.999
39	GM	E, F	-59.661	127.322	133.058	755.156	58.652	0.999	0.999
55	TL	E, F	-59.661	129.322	136.492	755.156	0.964	0.998	0.999
17	S	F	-59.662	127.323	133.059	759.170	0.713	0.999	0.999
37	GM	F	-59.662	125.323	129.625	759.655	27.771	0.999	0.999
53	TL	F	-59.662	127.323	133.059	759.655	2.015	0.999	0.999
14	IFRGSB	E, F, C	-67.013	146.025	154.629	605.317	73.496	0.997	0.997
34	DW3	E, C	-70.972	151.944	159.114	1349.865	27.585	0.994	0.996
56	TL	E, C	-71.002	152.003	159.173	1373.029	2.057	0.994	0.996
20	S	E, C	-71.204	152.408	159.578	1409.778	72.097	0.994	0.996
40	GM	E, C	-71.237	150.474	156.210	1402.576	60.998	0.994	0.996
12	IFRGSB	E, C	-71.239	152.477	159.647	1598.325	2.059	0.993	0.996
54	TL	C	-71.587	151.175	156.911	1288.200	3.676	0.994	0.996
33	DW3	C	-71.823	151.647	157.383	1307.653	0.594	0.994	0.996
18	S	C	-71.899	151.797	157.533	1344.723	15.171	0.994	0.996
38	GM	C	-71.946	149.892	154.194	1336.197	193.631	0.994	0.996
31	DW3	E	-72.088	152.175	157.911	1087.949	1.860	0.995	0.996
16	S	E	-72.667	153.334	159.070	1123.852	28.615	0.995	0.996

Result of range analysis (an explanation of which part of data is good for proceeding with the analysis)

Using a Laplace trend test we determined the suitable range of useful data as those with a score greater than 1.96 as this indicates we are at least 95% confident there is a significant trend upward.

Formula used for laplace test:

$$\text{Laplace score} = [((\sum_{i}^{n} t_i) / n) - T / 2] \Big/ T\sqrt{1/12n}$$

Table 1.1 Laplace Trend Test

Time (T)	Failure Count (FC)	Cumulative Failure Count (n)	Laplace Trend Test
1	2	2	-0.724744871
2	11	13	-0.249615231
3	2	15	-0.047213595
4	4	19	0.128956082
5	3	22	0.312543709
6	1	23	0.551885919
7	1	24	0.813113276
8	2	26	1.044932274
9	4	30	1.183772234
10	0	30	1.517105567
11	4	34	1.644132208
12	1	35	1.935801407
13	3	38	2.113761099
14	0	38	2.482182151
15	1	39	2.799572979
16	1	40	3.126138721
17	2	42	3.375595901
18	1	43	3.712608914
19	8	51	3.482954571
20	9	60	3.276393202
21	6	66	3.286799284
22	7	73	3.263032211
23	4	77	3.387030076

24	3	80	3.556350833
25	0	80	3.868850833
26	4	84	3.989589192
27	1	85	4.259191536
28	0	85	4.588603301
29	2	87	4.814304662
30	2	89	5.041122083
31	3	92	5.210725568

The range of data that is good is from time 13 to 31 and we can proceed with analysis for this part of the data.

Plots for failure rate and reliability of the SUT for the test data provided.

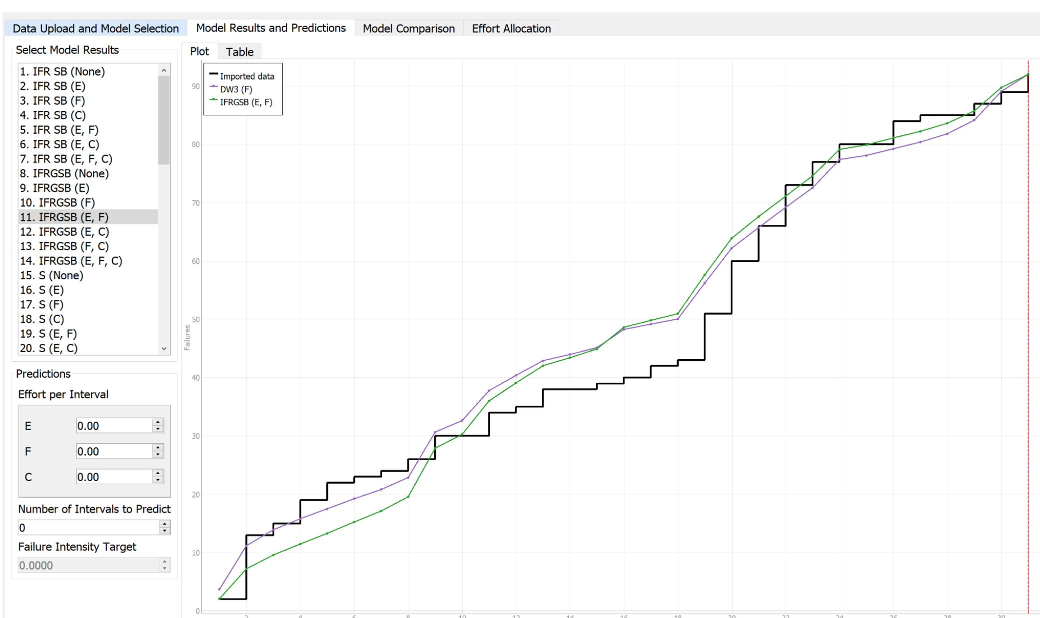


Figure 1.1 MVF (Mean Value Function) Graph

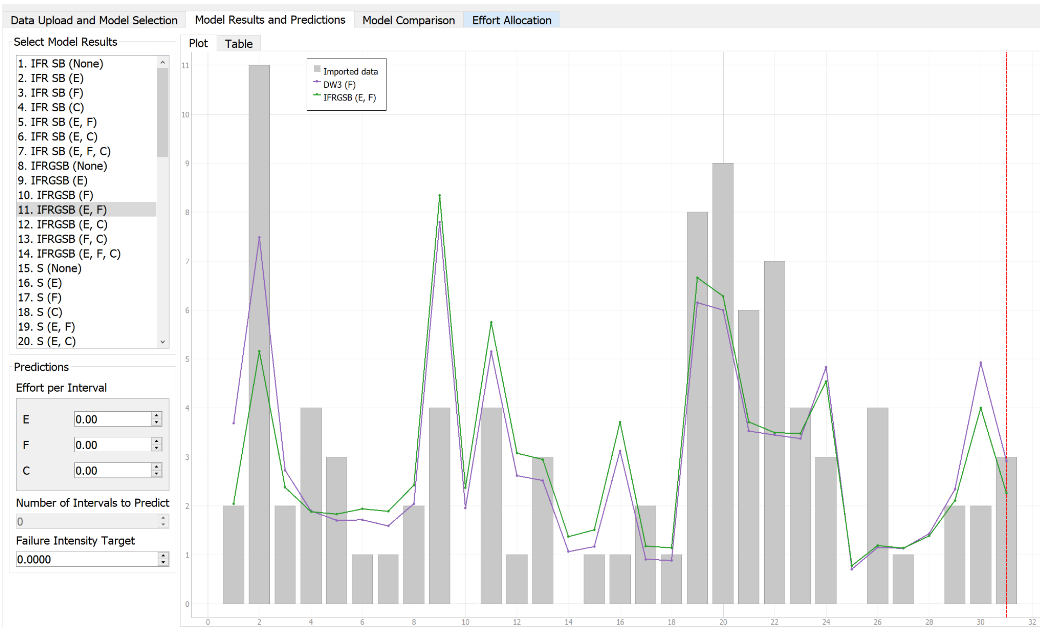


Figure 1.2 Failure Intensity Graph

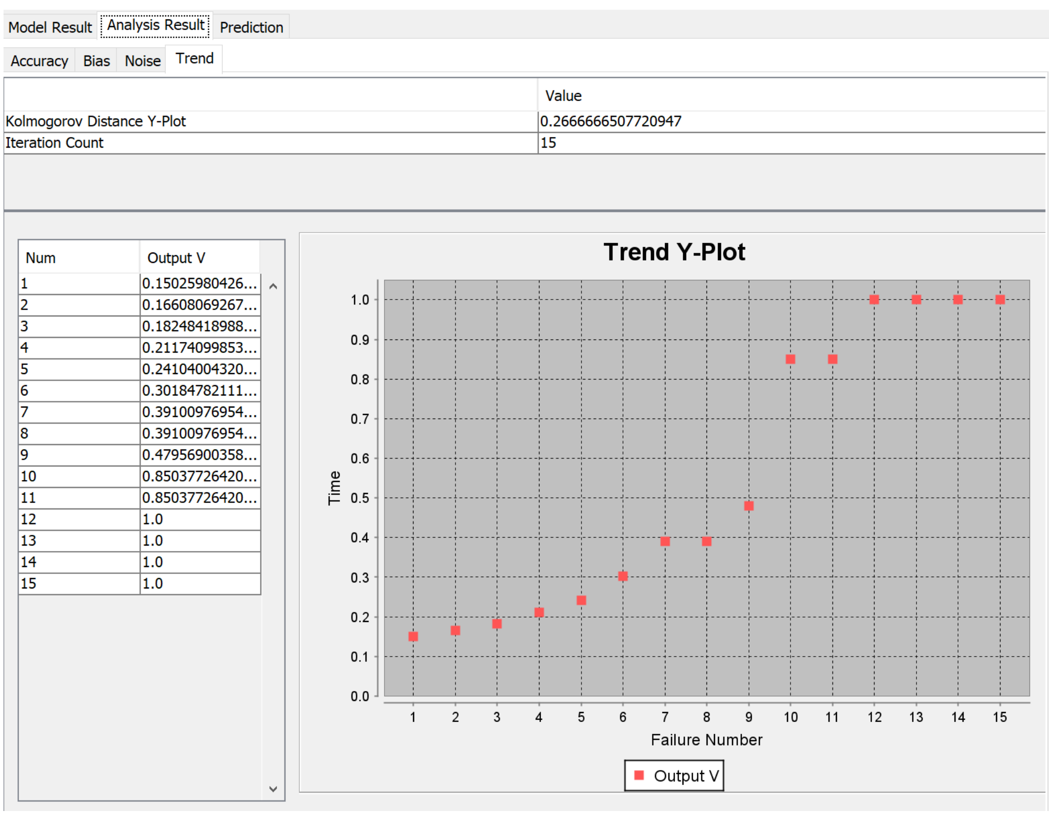
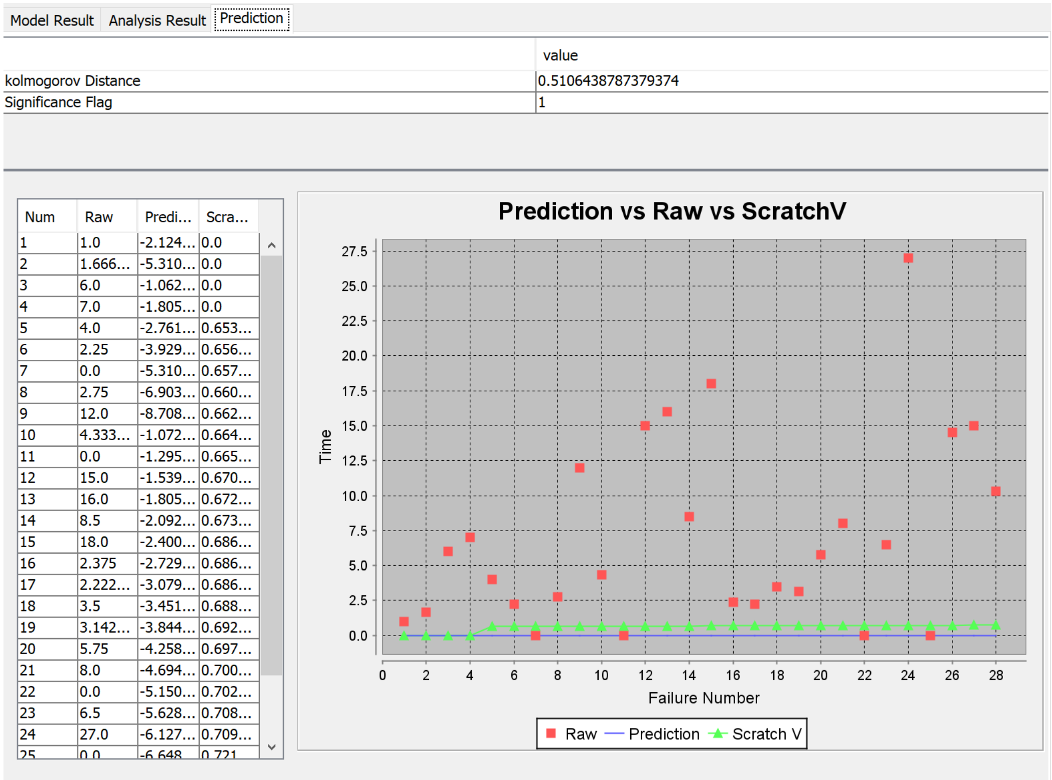


Figure 1.3 Geometric Trend Analysis Graph using time-between-failure data with the SRTAT tool



	value
kolmogorov Distance	0.5106438787379374
Significance Flag	1

Num	Raw	Predi...	Scra...
1	1.0	-2.124...	0.0
2	1.666...	-5.310...	0.0
3	6.0	-1.062...	0.0
4	7.0	-1.805...	0.0
5	4.0	-2.761...	0.653...
6	2.25	-3.929...	0.656...
7	0.0	-5.310...	0.657...
8	2.75	-6.903...	0.660...
9	12.0	-8.708...	0.662...
10	4.333...	-1.072...	0.664...
11	0.0	-1.295...	0.665...
12	15.0	-1.539...	0.670...
13	16.0	-1.805...	0.672...
14	8.5	-2.092...	0.673...
15	18.0	-2.400...	0.686...
16	2.375	-2.729...	0.686...
17	2.222...	-3.079...	0.686...
18	3.5	-3.451...	0.688...
19	3.142...	-3.844...	0.692...
20	5.75	-4.258...	0.697...
21	8.0	-4.694...	0.700...
22	0.0	-5.150...	0.702...
23	6.5	-5.628...	0.708...
24	27.0	-6.127...	0.709...
25	0.0	-6.648	0.721

Figure 1.4 Littlewood and Varral's Bayesian Reliability Prediction Graph

A discussion on decision making given a target failure rate

Given an acceptable range of failure rate for the test data one can use this target failure and the model parameters to make predictions of the remaining failures in the system and the time to release. The selected reliability models above fit the failure data against a model. Through reliability acceptance analysis a decision can be made on whether the system is acceptable to be released. A target failure rate ensures reliability and by identifying the lowest failure intensity for the models in Figure 1.2 a desired failure rate prediction using C-SFRAT was set to 0.5 to be predicted in 50 intervals. This target failure rate can be said to be MTTF = 0.5/50 = 0.01.

A discussion on the advantages and disadvantages of reliability growth analysis

The reliability growth model allows us to understand how a system's reliability changes over time and this is recorded throughout the testing process. An advantage to the reliability growth model is that it allows us to measure improvements, achieve effectiveness and ensure efficient testing and debugging for our program and the system under test is ultimately achieved. It further allows us to understand when the release date of the project should be and allows us to estimate the number of services that will be required for the release of this product. Hence, the reliability growth model facilitates the measurement and predictions for improvements of reliability programs through the testing process. Some decisions related to reliability growth testing are the needs to track bugs in pre-release and product release. A disadvantage to this system is that when doing this type of analysis, it is almost an exhaustive approach since we are unsure when testing is enough, test phase time and or resources are usually exhausted. Good measurement techniques to ensure the amount of accurate testing is needed and should be determined by the team to ensure an accurate amount of testing is performed without exhausting resources.

3 Assessment Using Reliability Demonstration Chart

3 plots for MTTFmin, twice and half of it for your test data using RDC-11 as our testing tool

For all figures below the following risk profile was used:
Discrimination Ratio (y) = 2.000
Developer Risk (a) = 0.1
Customer Risk (b) = 0.1

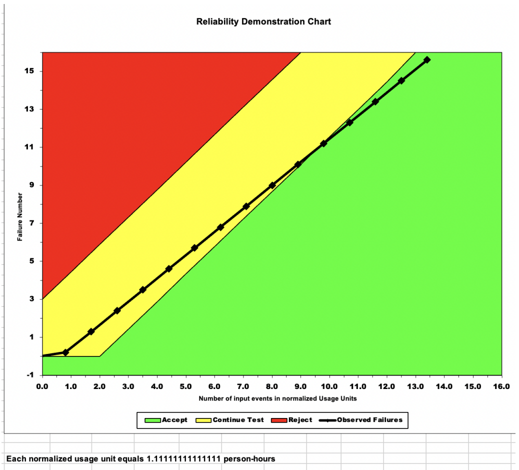


Figure 2.1 MTTFmin as 1.25 with a FIO of 8 failures

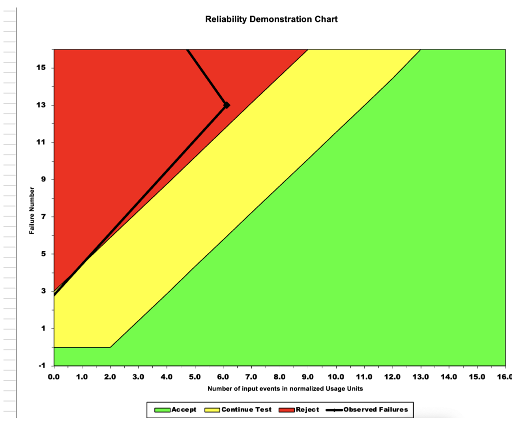


Figure 2.2 MTTFtwice as 2.5 with a FIO of 4 failures

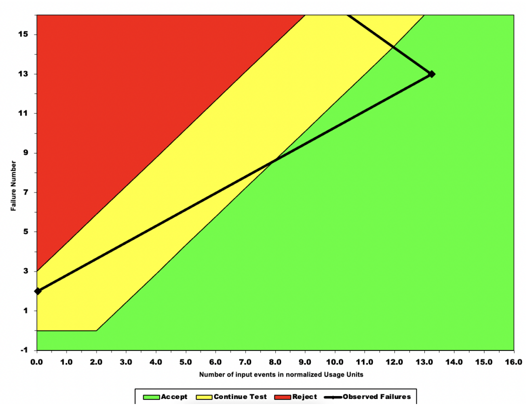


Figure 2.3 MTTFhalf as 1 with a FIO of 10 failures

3 plots for MTTFmin, twice and half of it for your test data using SRTAT as our RDC testing tool

Time Between Failure, calculated from T/FC was used as the input.

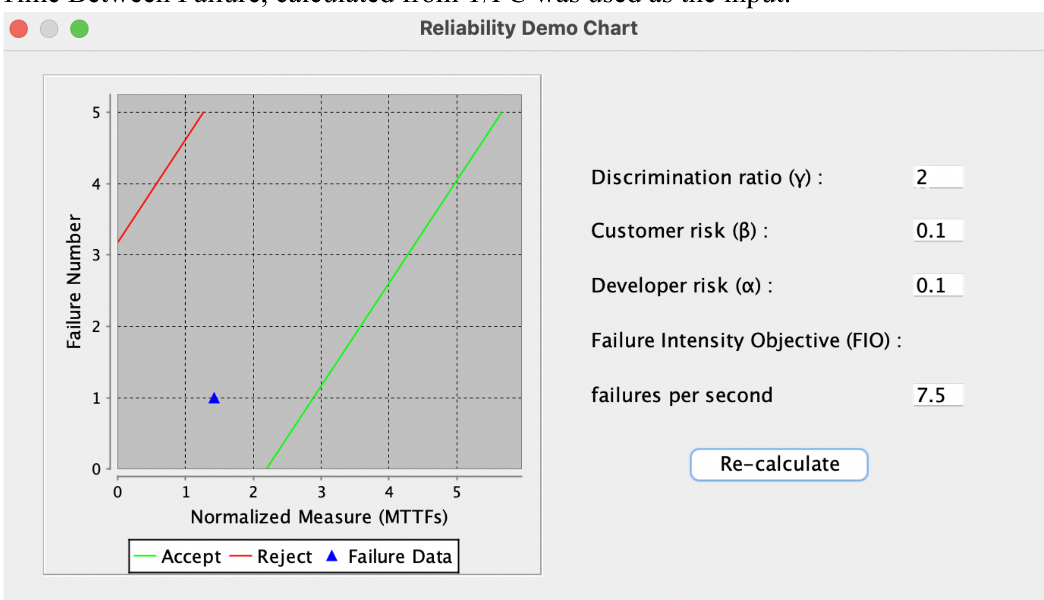


Figure 2.4 MTTFtwice with a FIO of 7.5 failures per second

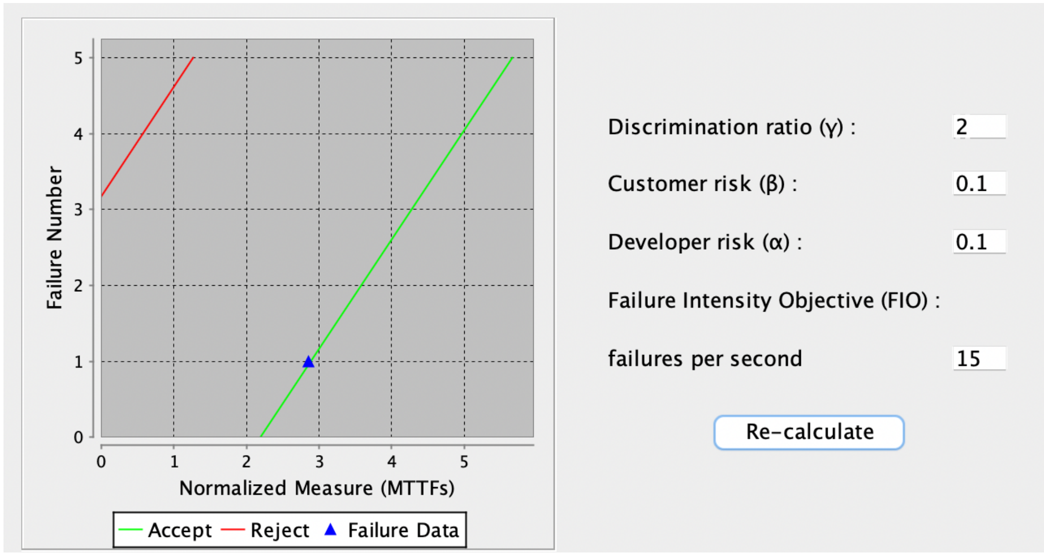


Figure 2.5 MTTFmin with a FIO of 15 failures per second

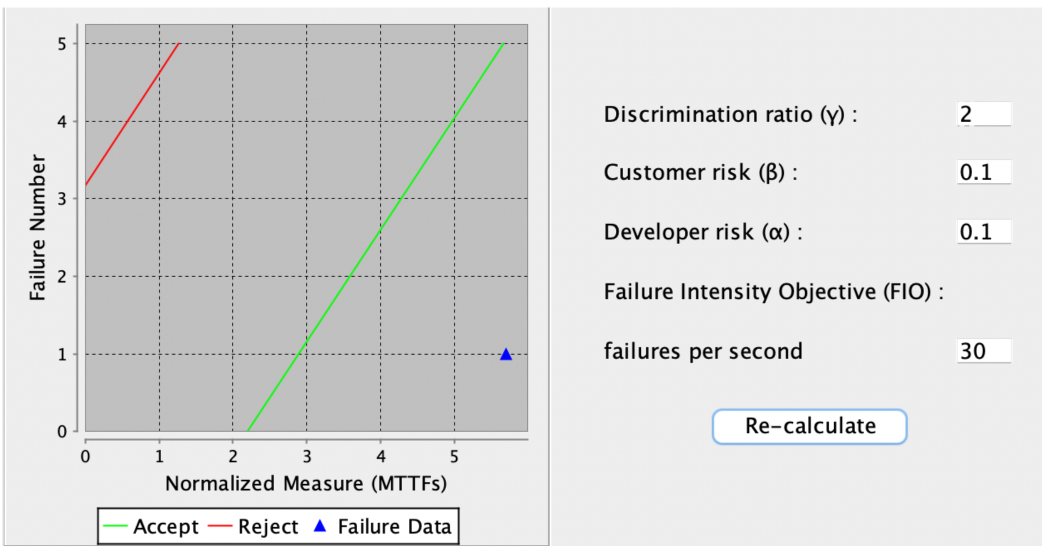


Figure 2.6 MTTFhalf with a FIO of 30 failures per second

Explain your evaluation and justification of how you decide the MTTFmin

The MTTFmin was determined by adjusting the maximum number of failures to result in a graph with the greatest number of failure data in the accept region. When the observed failures only entered the accept and continue test regions the MTTFmin was selected.

A discussion on the advantages and disadvantages of RDC

Reliability Demonstration Chart allows you to measure system failure against your reliability goals. Advantages of RDC is that it clearly indicates whether a reliability objective is satisfied or not through a certain boolean value. Additionally, it's good when failure data is limited to a few failures and time is known. Furthermore RDC propagates “what if?” scenarios for the system which means you are able to experiment with different values of confidence levels and MTTF. RDC is versatile and a time & cost efficient way of analyzing the reliability of a system. The RDC also incorporates customer and developer risk when determining the reject, continue and accept regions, hence providing a good overview when decision making. Disadvantages are that it requires more data and additional calculations to get the model parameters. Additionally, it cannot be used to calculate the exact quantitative value for the reliability (or availability) of the system under study, it can only indicate that the SUT is acceptable or not.

Comparison of Results

The results from Part 1 predict the reliability of a system which were determined by using the covariates and different models to determine acceptable failure rates. In Part 2 we used RDC to find the MTTF. The minimum MTTF was found to be 1.25 whereas using the reliability growth analysis the resulting MTTF was found to be 0.1. These numbers are not extremely similar, due to the error observed using the excel sheet and the interval and target selected on the C-SFRAT tool. However, relatively these results can be comparative as they are both small.

Discussion on Similarity and Differences of the Two Techniques

Reliability Growth Testing differs from RDC testing as it is used to estimate or predict the improvement of system reliability as a function of the amount of system testing that is carried out. Whereas RDC is only used to determine if an SUT is acceptable or not by seeing if the target failure rate is met or not. The main difference for RDC is that its good when one wants to find out what is the trend for reliability of the system and Reliability Growth Testing is good when one wants to track bugs in pre-release and determine a suitable product release date. A similarity between the two techniques is that they both require time between failure data. To summarize, RDC is based on inter failure times only and target failure rate or MTTF where reliability growth analysis is based on inter failure times and/or failure count and target failure rate or MTTF.

How the team work/effort was divided and managed

Huda: Plotted our Reliability Demonstration Chart using the excel file in part 2, comparisons between techniques and advantages/disadvantages of RDC

Nuha: Comments/feedback on lab, difficulties encountered, reliability growth model advantages/disadvantages, RDC advantages/disadvantages

Lubaba: Created graphs from C-SFRAT in part 1, difficulties encountered, RDC advantages and disadvantages

Rajpreet: Introduction, reliability growth model advantages/disadvantages, edits to other written portions throughout the lab (difficulties encountered, comments on lab)

Difficulties encountered, challenges overcome, and lessons learned

Some of the difficulties that we encountered during the lab was modifying the data provided in the excel sheet in order to be imported to STRAT. We also struggled with understanding the attributes under the model comparison tab in CSFRAT. We were also unsure of how to change the metric weights in model comparison and how to filter the data based on which model was best. For part 2, the RDC chart required lots of scaling and manipulation as it only displays failure numbers to 15. This caused the failure data to not be properly plotted and resulted in inaccurate results.

Another large difficulty our team encountered is that C-SFRAT is outdated and does not function on all operating systems. This severely hinders a group's progress when a group is using MAC OS which the software does not support. Additionally, it also takes away from the learning capabilities of the members unable to utilize the software due to their OS since they are unable to test functionalities, manipulate the data and plot various things. Also struggled getting the CASRE application to open for all operating systems, due to the Windows 3 requirement.

We overcame these compatibility and tool difficulties by attempting all applications on all operating systems until we found a tool that was functioning. We learned a lot about system requirements and how to properly format our environment and files to fit the specificity required by an application.

Comments/feedback on the lab itself

Overall the lab allowed us to understand the importance of analyzing failure data and its significance in terms of software reliability through reliability growth testing as well as reliability assessment using

Reliability Demonstration Chart (RDC) during integration testing. We were able to gain an understanding of what reliability growth testing is and why it is useful. Additionally we learned how to measure the failure rate, MTTF and reliability of the SUT through analyzing the test data and became familiar with the features and usage of a reliability growth testing tool. The lab was further able to equip us with valuable knowledge regarding the importance of understanding the reliability growth model and the role it plays in ensuring the successful release of a software system.

The lab instructions and goals required more clarity as there were lots of questions that had to be asked to ensure what we did was what was actually required. Additionally we felt that there was a huge knowledge gap between what was taught in lectures and the applications of the lab. Additionally, the setup and OS issues with software was time consuming and some of the software was outdated and would not open. Furthermore, with both parts the lab fet a little too long and covered too much material.