Investigations Using the Weibull Analysis for Predictability of Failure in the Service Industry

Michael J. Anderson
Laguna Niguel, CA 92677

Abstract
There is significant literature on the Weibull Distribution's common use as a time-to-failure model in manufacturing. This paper will investigate whether a Weibull analysis can be applied when considering inter-arrival time for customers in the service sector. Inter-arrival time will be defined as the time (in days) between a customer's services at an establishment.

KeyWords: Inter-arrival Time, Cumulative Distribution Function (CDF)

1. Introduction

The data for this project was taken from the alternative health (maintenance) service industry. In order for independency, the datasets were limited to only those customers/clients that had an identical service. Even though the Weibull distribution is typically utilized with failure times, this analysis was carried out investigating inter-arrival time, (the number, in days, between each customer's appointments), and many conclusions can be made utilizing the results of this analysis.

Based on the nature of the customer, both complete data as well as multi-censored was utilized. Due to the idiosyncrasies of the particular clients in the dataset, it was determined that the client's inter-arrival times would be censored from 1 – 6 days between appointments, as well as > 73 days.

Clients were categorized by those having the same service as well as those who came in the same number of times. The client appointment dataset was broken down and analyzed according to the number of appointments (visits) for each client. For example, all clients that have frequented the establishment 3 times were grouped and analyzed together, similarly, 4 visits, 5 visits through 31 visits, were all grouped together. The following notation will be used:

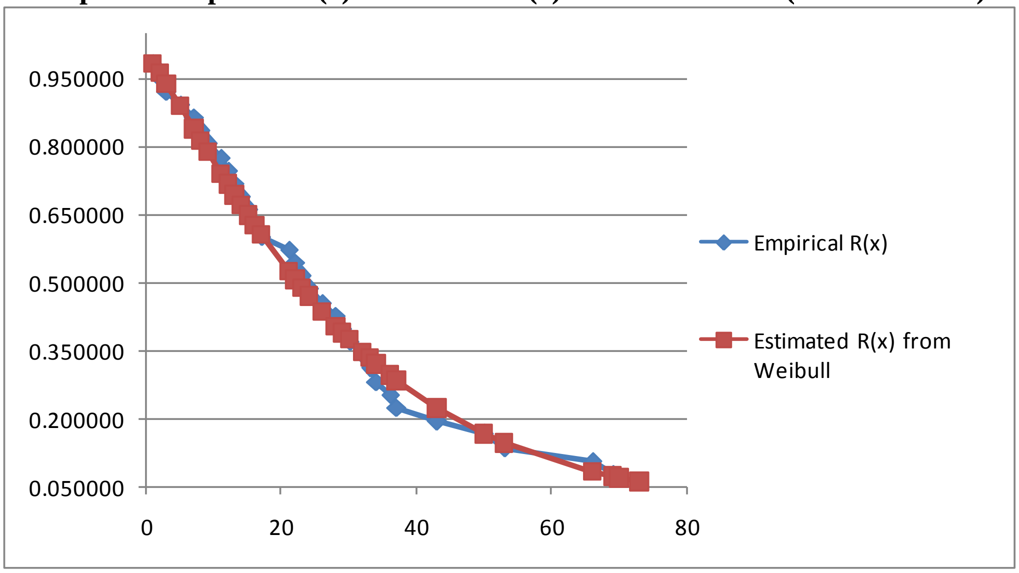
Visits:	Used interchangeably with "appointments"
X – Visits:	The quantity (X) of visits that a client frequented the service establishment
CDF – X:	The Cumulative Distribution Function (CDF) for clients having X-Visits.
X Value:	The X number of days between visits.

Finally, the appointments were analyzed in more refined groups where the β parameters were similar, and these are color coded in the tables below. After observing the results of the analysis, the multi-censored data proved more accurate when compared to complete data, and therefore, the discussion will be focused on the censored data results only.

2. Method

The first step was to see if the Weibull calculations could be used with the dataset. Graph 1 shows how the empirical Reliability Function - R(x), compares very favorably with the Weibull Estimated R(x).

Graph 1 – Empirical R(x) vs. Weibull R(x) – Censored Data (21-Visit Client)



There are several methods that can be used for calculating the parameters of the Weibull CDF. The Maximum Likelihood Method (MLE) and graphical methods were studied here. Due to the fact that very satisfactory results were achieved, as well as the speed in calculating, the graphical method will be the focus of this report.

The first step is to plot the data to see if it follows a straight line. In order to do this the CDF, given by:

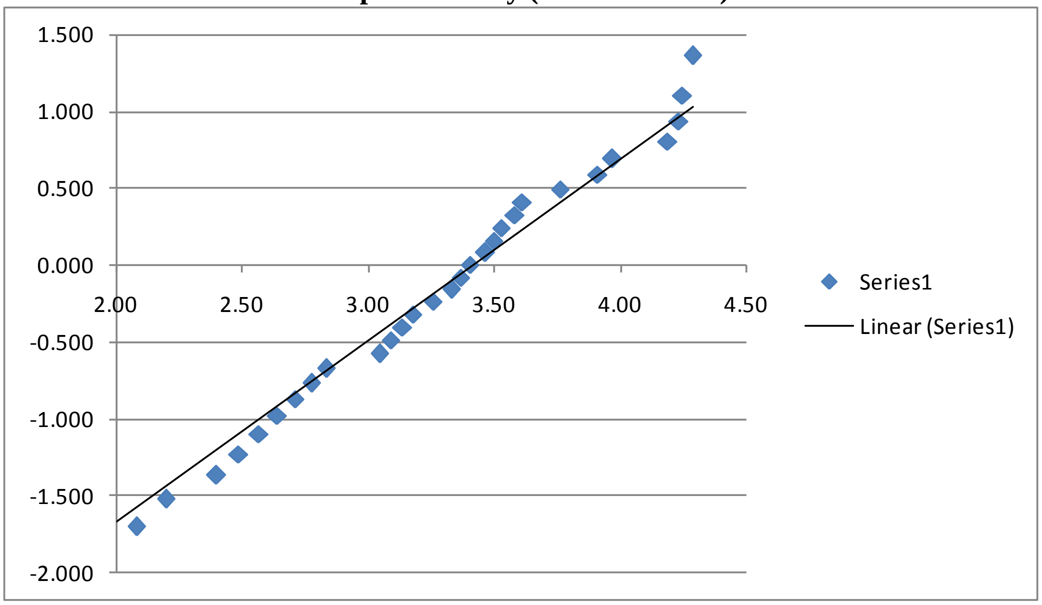
$$f(x) = 1 - e^{[-(\frac{x-\delta}{\theta})^{\beta}]}$$

was put in the form of a line as:

$$\ln(-\ln(R(x))) = \beta\ln(x) - \beta ln\theta$$

From this, we plot x as ln(x) versus y as ln[-ln(Emperical(R(x)))]. Graph 2 shows one such plot, i.e. one category of client (21-Visits), which depicts a good linear fit.

Graph 2 – x vs. y (21-Visit Client)



From this, we solve for the slope β (the Weibull shape parameter), and y-intercept. These values are then used to solve for θ (the Weibull scale parameter) by:

$$\theta = e^{-[\frac{b}{\beta}]}$$

Where:
b = Intercept

Finally, the Weibull CDF is calculated, and a partial summary of the results are shown in Table 1.

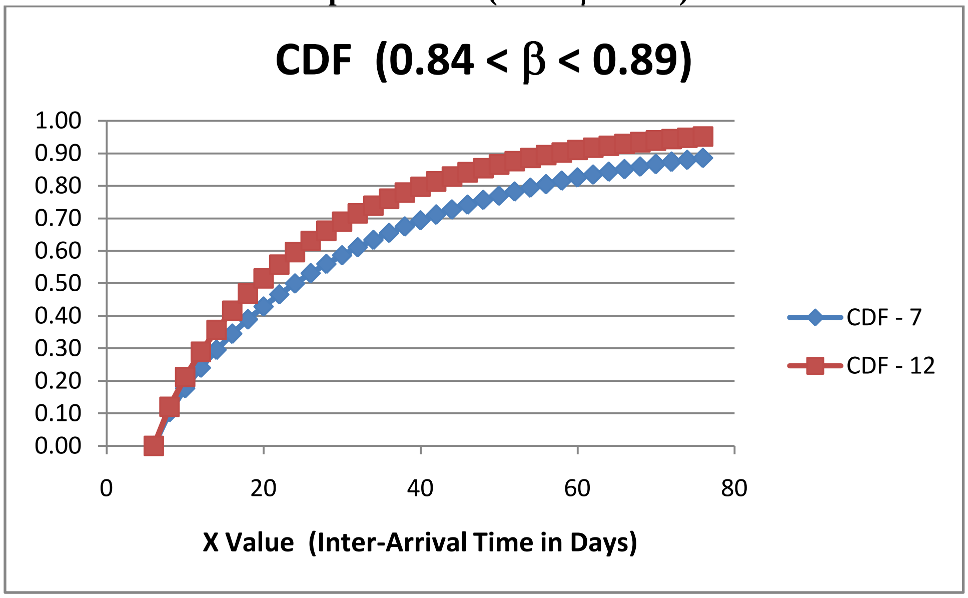
Table 1 – CDF Output – Censored Data

Similar β Color Code																	
	Test Client	3 Visit	4 Visit	5 Visit	7 Visit	8 Visit	9 Visit	10 Visit	11 Visit	12 Visit	13 Visit	14 Visit	15 Visit	18 Visit	19 Visit	21 Visit	31 Visit
β:	1.28	1.14	1.03	1.06	0.84	1.16	0.98	1.45	1.12	0.89	1.16	1.28	1.00	1.14	1.15	1.18	1.10
θ:	49.87	30.38	32.53	39.59	27.88	31.96	27.89	41.35	30.66	20.11	18.18	34.50	30.87	30.43	25.56	30.43	22.31
δ:	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
μ:		224.4	162.5	165.8	122.5	93.4	112.1	105.1	70.6	107.6	38.0	63.1	64.3	68.4	64.7	70.2	52.3
X Value	**CDF Test**	CDF 3	CDF 4	CDF 5	CDF 7	CDF 8	CDF 9	CDF 10	CDF 11	CDF 12	CDF 13	CDF 14	CDF 15	CDF 18	CDF 19	CDF 21	CDF 31
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.016	0.044	0.055	0.042	0.103	0.040	0.072	0.012	0.046	0.121	0.074	0.026	0.063	0.044	0.052	0.040	0.068
10	0.039	0.094	0.109	0.085	0.177	0.086	0.138	0.033	0.097	0.212	0.158	0.061	0.122	0.095	0.112	0.087	0.140
12	0.064	0.146	0.161	0.128	0.240	0.135	0.198	0.059	0.149	0.289	0.241	0.100	0.177	0.146	0.172	0.137	0.210
14	0.091	0.196	0.210	0.169	0.295	0.183	0.254	0.088	0.199	0.356	0.320	0.142	0.229	0.197	0.231	0.187	0.277
16	0.120	0.246	0.257	0.209	0.344	0.230	0.306	0.120	0.248	0.416	0.393	0.185	0.277	0.246	0.288	0.236	0.339
18	0.149	0.293	0.301	0.247	0.389	0.275	0.354	0.153	0.295	0.468	0.461	0.227	0.323	0.293	0.342	0.284	0.397
20	0.178	0.339	0.343	0.284	0.429	0.320	0.398	0.188	0.340	0.516	0.522	0.270	0.365	0.339	0.394	0.330	0.451
22	0.208	0.382	0.382	0.319	0.466	0.362	0.440	0.223	0.383	0.558	0.578	0.311	0.405	0.382	0.442	0.374	0.500
24	0.237	0.423	0.419	0.353	0.499	0.402	0.478	0.259	0.424	0.596	0.628	0.352	0.442	0.423	0.487	0.416	0.546
26	0.267	0.463	0.454	0.385	0.531	0.441	0.514	0.294	0.462	0.630	0.673	0.391	0.477	0.462	0.530	0.456	0.588
30	0.324	0.534	0.519	0.445	0.586	0.512	0.578	0.365	0.532	0.690	0.749	0.466	0.541	0.534	0.605	0.530	0.662
38	0.432	0.654	0.626	0.550	0.675	0.633	0.682	0.498	0.650	0.779	0.855	0.597	0.645	0.653	0.726	0.654	0.774
42	0.482	0.703	0.670	0.595	0.711	0.683	0.723	0.559	0.698	0.813	0.890	0.652	0.688	0.702	0.773	0.705	0.816
46	0.529	0.745	0.710	0.636	0.742	0.726	0.759	0.614	0.740	0.842	0.918	0.702	0.726	0.744	0.813	0.749	0.850
50	0.573	0.782	0.745	0.673	0.770	0.765	0.791	0.665	0.776	0.865	0.939	0.745	0.759	0.781	0.846	0.787	0.879
54	0.614	0.814	0.775	0.706	0.794	0.798	0.818	0.711	0.808	0.886	0.954	0.783	0.789	0.813	0.873	0.819	0.902
58	0.652	0.842	0.802	0.736	0.815	0.827	0.842	0.752	0.836	0.902	0.966	0.816	0.814	0.841	0.896	0.848	0.921
62	0.687	0.866	0.826	0.764	0.834	0.852	0.862	0.788	0.859	0.917	0.975	0.845	0.837	0.865	0.915	0.872	0.936
66	0.718	0.886	0.847	0.788	0.851	0.874	0.880	0.820	0.880	0.929	0.982	0.869	0.856	0.885	0.931	0.892	0.949
70	0.748	0.903	0.866	0.810	0.866	0.893	0.896	0.848	0.898	0.939	0.987	0.890	0.874	0.903	0.944	0.910	0.959
74	0.774	0.918	0.882	0.830	0.880	0.909	0.909	0.872	0.913	0.948	0.990	0.908	0.889	0.917	0.954	0.924	0.967
76	0.786	0.925	0.890	0.839	0.886	0.916	0.915	0.883	0.919	0.952	0.992	0.916	0.896	0.924	0.959	0.931	0.970

3. Discussion

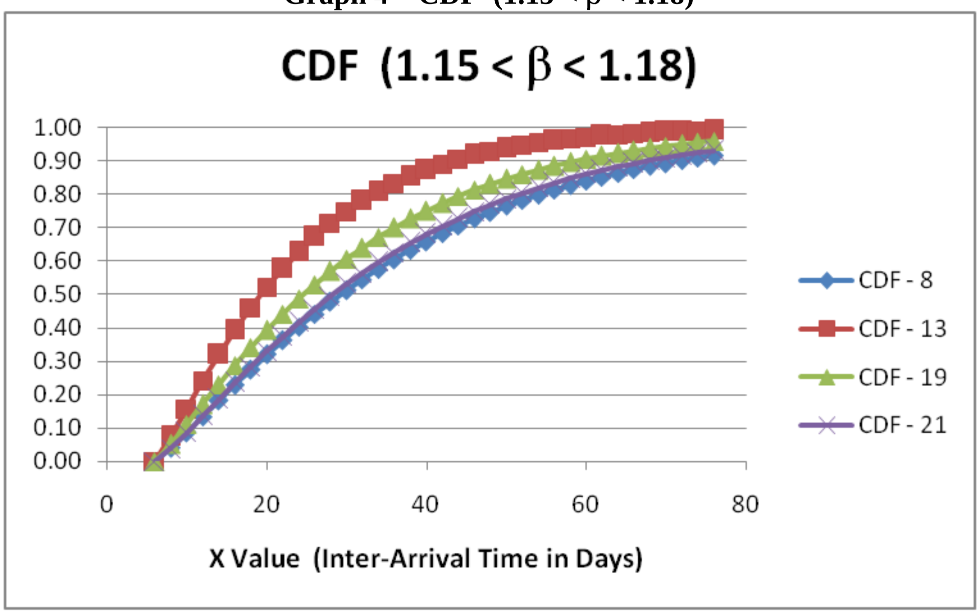
Within the service industry, if a customer is pleased with the quality of service, as well as with the establishment as a whole, we would expect that client to continue to frequent the establishment. Further, we would expect this to be even more prevalent in those service industries that are not necessities, such as the restaurant industry, or in the alternative health maintenance industry as studied here. The analysis contained herein proved this to be true as can be seen in the CDF table and graph 3. The graph makes it clear to see that those clients who frequented the establishment on more occasions, did so with a greater degree of frequency as well. Focusing on the inter-arrival time of 20, we can see that almost 10% more of the 12-Visit clients came in than did the 7-Visit clients. Similarly, there is a 20% difference between the 5 and 24 visit customers as can be seen in Table 1.

Graph 3 – CDF (0.84 < β < 0.89)



However, an anomaly occurs in Chart 4 when comparing the 13 versus the 21-Visit customers. Here we see that the 13-Visit client has a far greater degree of frequency, approximately 20% more than its counterpart.

Graph 4 – CDF (1.15 < β < 1.18)



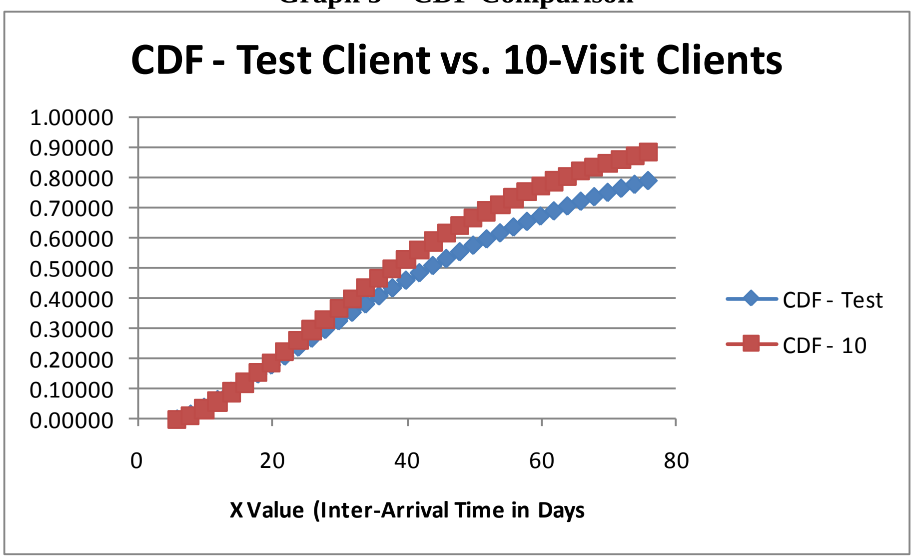
If we look at the mean inter-arrival time for all customers in each group and see that the 13-Visit group is abnormally low, and without further investigation, might be considered an outlier – Table 2.

Table 2 – Inter-Arrival Mean Values

Similar β Color Code:	[purple]	[blue]	[blue]	[yellow]	[red]	[blue]	[green]	[purple]	[yellow]	[red]	[green]	[blue]	[purple]	[red]	[red]	[purple]
Visits/Customer:	3	4	5	7	8	9	10	11	12	13	14	15	18	19	21	31
Inter-Arrival μ:	224.4	162.5	165.8	122.5	93.4	112.1	105.1	70.6	107.6	38.0	63.1	64.3	68.4	64.7	70.2	52.3

It was desirous to see if there could be predictive power using the analysis that has been concluded. Therefore a "test client" was created assigning random inter-arrival times of 11 days between visits, 17 days, 27, 63, 64 and 73 days. The model was then run and the results can be seen in Table 1 under the heading of Test Client. A careful observation of the results from the Test Client with those of the 10-Visit clients will show how the two distributions are almost identical. A graphical summary of the comparison can also be seen in Graph 5. Note: In comparing the Test Client with any other client group, we have made the simplifying assumption that within each group, the individuals all have the same inter-arrival time distributions and that all inter-arrival times are independent.

Graph 5 – CDF Comparison



It is believed that within the non-essential service sector that it is advantageous to minimize the inter-arrival times in order to maximize profits. Further, relating to the specific industry that was analyzed for this paper, it is believed that 16 to 22 days between appointments is the optimum for the type of health benefits this establishment provides. Graph 5 shows that our CDF distribution can be relied upon to make inferences about our test client, especially within this optimum inter-arrival range. Additionally, several other inter-arrival times can be introduced, and subsequently running the model again. By doing this, it could be determined what the next arrival time should be to get that client to a higher level, say 15 visits, 18, 21 visits or beyond. Once this is known, marketing efforts can be utilized to insure said arrival time.

References

Dodson, Bryan, (2006). The Weibull Analysis Handbook, Milwaukee, WI: ASQ Press.

Gertsbakh, I.B., (1989). Statistical Reliability Theory. New York, NY: Marcel Dekker, Inc.

Cohen, A. Clifford, (1965). “Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and on Censored Samples.” Technometrics, Vol. 17, No. 4, pp 579-588.

_____, (1975). “Multi-Censored Sampling in the Three Parameter Weibull Distribution.” Technometrics, Vol. 17, No. 3, pp 347-351.

Nelson, Wayne, (1982). Applied Life Data Analysis, New York, NY: John Wiley & Sons.

Pham, Hoang (Ed.), (2003). Handbook of Reliability Engineering, London, UK: Springer-Verlag.