

# Using Microsoft EXCEL for Weibull Analysis

by William W. Dorner

Many people use Microsoft Excel on a daily basis. Yet few people realize the extent of Excel's analytical capabilities. Fewer still put these capabilities to work for process improvement, product improvement and profit. Most Excel users are aware of the common formulas and charts. But with some creativity, users can produce tools like control charts, Pareto charts and box-and-whisker plots (see "Using Excel for Data Analysis," *Quality Digest*, October 1997). And with a little guidance, users can employ more advanced statistical methods with Excel. This article presents a how-to approach for one such advanced technique-Weibull analysis.

**Download the Excel worksheet used in this article**

**Netscape users may need to download this Zipped version instead.**  
[Weibull.zip](#)

You haven't turned the page yet? Those of you who remain probably fall under one of two categories: those familiar with reliability data analysis, and Excel enthusiasts who are curious to learn one more way to exploit this versatile software. I predict readers in both groups will be glad they stuck around.

For the uninitiated, Weibull analysis is a method for modeling data sets containing values greater than zero, such as failure data. Weibull analysis can make predictions about a product's life, compare the reliability of competing product designs, statistically establish warranty policies or proactively manage spare parts inventories, to name just a few common industrial applications. In academia, Weibull analysis has modeled such diverse phenomena as the length of labor strikes, AIDS mortality and earthquake probabilities.

## Learning by example

Let's ignore the formulas for now and start by looking at an example of Weibull analysis in action. Imagine that you work for a toy company that wants to compare the reliability of two proposed designs for a jack-in-the-box spring housing. The desired reliability at 400,000 cycles is 0.90. In other words, the toy company would like 90 percent of the spring housings to survive at least 400,000 cycles. This reliability goal is expressed mathematically as  $R(400,000) \geq 0.90$ . Ten units were assembled with each of the two housing designs (Design A and Design B). These 20 units were tested until their spring housings failed. Figure 1 shows the number of cycles before failure for each item tested.

The data in Figure 1 don't clearly indicate whether either design meets the desired reliability goal. Both designs had at least one failure before 400,000 cycles, yet clearly the average number of cycles before failure exceeds 400,000 for both designs. A comparison of sample averages using a Student's t test reveals no statistical difference between the average cycles for Design A and the average cycles for Design B ( $p\text{-value} = 0.965$ ). But as a simple measure of central tendency, the sample average gives no information about the spread or shape of the distribution of failure times. Could the two designs' averages be the same, but their reliability be quite different? How can you be more scientific about comparing the reliability of the two proposed designs?

**Figure 1: Failure Data**

Design A		Design B	
Sample	Cycles	Sample	Cycles
1	726,044	11	529,082
2	615,432	12	729,957
3	508,077	13	650,570
4	807,863	14	445,834
5	755,223	15	343,280
6	848,953	16	959,903
7	384,558	17	730,049
8	666,686	18	730,640
9	515,201	19	973,224
10	483,331	20	258,006

## Preparing to analyze

Modeling the data using Weibull analysis requires some preparation. For now, focus on the data from Design A.

1. Open Excel and into cell A1, type the label: Design A Cycles. Enter the failure data for Design A into cells A2:A11. Highlight cells A1:A11 and click on the Sort Ascending button to order the failure cycles from lowest to highest.
2. In cell B1, type the label: Rank. In cells B2:B11, type the integers 1-10 (see Figure 2).

**Figure 2: Preparing Design A for Weibull Analysis**

	A	B	C	D	E	F
1	Design A Cycles	Rank	Median Ranks	1/(1-Median Rank)	ln(ln(1/(1-Median Rank)))	ln(Design A Cycles)
2	384,558	1	0.067307692	1.072164948	-2.663843085	12.8598499
3	483,331	2	0.163461538	1.195402299	-1.72326315	13.088457
4	508,077	3	0.259615385	1.350649351	-1.202023115	13.13838829
5	515,201	4	0.355769231	1.552238806	-0.821666515	13.15231239
6	615,432	5	0.451923077	1.824561404	-0.508595394	13.33007974
7	666,686	6	0.548076923	2.212765957	-0.230365445	13.41007445
8	726,044	7	0.644230769	2.810810811	0.032924962	13.4953659
9	755,223	8	0.740384615	3.851851852	0.299032932	13.53476835
10	807,863	9	0.836538462	6.117647059	0.593977217	13.60214777
11	848,953	10	0.932692308	14.85714286	0.992688929	13.6517591

3. In Column C, put an estimate of the proportion of the population that will fail by the number of cycles listed in Column A. This can be accomplished using several different methods, the most common of which is median ranks. In cell C1, type the label: Median Ranks. In cell C2, enter the formula:  $=(B2-0.3)/(10+0.4)$ . Next, copy cell C2 down through cell C11. Note that in the formula for median ranks, the 10 in the denominator is the total number of Design A units tested.

4. Type into cell D1 the label:  $1/(1-\text{Median Rank})$ . Then, in D2, enter the formula:  $=1/(1-C2)$ . Copy cell D2 down through cell D11.

5. Into cell E1, enter the label:  $\ln(\ln(1/(1-\text{Median Rank})))$ . In cell E2, type the formula:  $=\text{LN}(\text{LN}(D2))$ . Copy cell E2 down through cell E11.

6. Finally, you'll need to transform the Cycles data. In cell F1, type the label:  $\ln(\text{Design A Cycles})$ . In cell F2, type the formula:  $=\text{LN}(A2)$ . Copy cell F2 down through cell F11.

7. Again, compare your spreadsheet with Figure 2. After you confirm that everything is correct, save your workbook.

### Fitting a line to the data

At this point, you're ready to perform the Weibull analysis. The beauty of this method is that you can expect to see a straight line when you plot the data in Column E vs. Column F. By performing a simple linear regression, you can obtain parameter estimates that will enable you to make inferences about Design A's reliability.<sup>3</sup>

First, be sure that the Analysis ToolPak Add-In is loaded into Excel. From the menu bar, select Tools . Add-Ins. Click on the checkbox for Analysis ToolPak, and then click OK.

To perform the simple linear regression:

1. While on the page you just created, from the menu bar, select Tools and Data Analysis. Scroll down and highlight "Regression" and click OK. A data-entry window will pop up.
2. Under "Input Y Range," type:  $\$E\$1:\$E\$11$ .
3. For "Input X Range," type:  $\$F\$1:\$F\$11$ .
4. Click to add a checkmark in the box for "Labels."
5. For "Output Options," select "New Worksheet Ply."
6. Click to add a checkmark in the box for "Line Fit Plots."
7. Click OK. Excel will perform the regression and place the output on a new worksheet.

### Reformatting the output

Before interpreting the output, you'll need to do some tidying up. The columns do not automatically adjust to their optimal widths. To do this, within the worksheet that you just created, click on column heading A and drag to column heading I. Now double-click on the

### Estimating Weibull Parameters

Why can we expect the graph of the  $\ln(\text{Cycles})$  vs. the transformed median ranks to plot as a straight line?

With some effort, the Weibull cumulative distribution function can be transformed so that it appears in the familiar form of a straight line:  $Y=mX+b$ . Here's how:

$$F(x) = 1 - e^{-\left(\frac{x}{\alpha}\right)^\beta}$$

$$1 - F(x) = e^{-\left(\frac{x}{\alpha}\right)^\beta}$$

$$\ln(1 - F(x)) = -\left(\frac{x}{\alpha}\right)^\beta$$

$$\ln\left(\frac{1}{1 - F(x)}\right) = \left(\frac{x}{\alpha}\right)^\beta$$

$$\ln\left[\ln\left(\frac{1}{1 - F(x)}\right)\right] = \beta \ln\left(\frac{x}{\alpha}\right)$$

$$\ln\left[\ln\left(\frac{1}{1 - F(x)}\right)\right] = \beta \ln x - \beta \ln \alpha$$

Comparing this equation with the simple equation for a line, we see that the left side of the equation corresponds to Y,  $\ln x$  corresponds to X,  $\beta$  corresponds to  $m$ , and  $-\beta \ln \alpha$  corresponds to  $b$ . Thus, when we perform the linear regression, the estimate for the Weibull  $\beta$  parameter comes directly from the slope of the line. The estimate for the  $\alpha$  parameter must be calculated as follows:

$$\alpha = e^{-\left(\frac{b}{\beta}\right)}$$

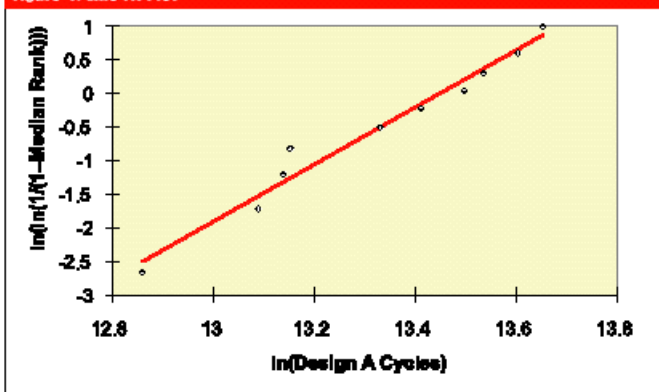
boundary to the right of any column heading. Your table should look similar to Figure 3.

**Figure 3: Results of Linear Regression for Design A**

	A	B	C	D	E	F	G
1	SUMMARY OUTPUT						
2							
3	Regression Statistics						
4	Multiple R	0.98538223					
5	R Square	0.97097815					
6	Adjusted R Square	0.96735041					
7	Standard Error	0.20147761					
8	Observations	10					
9							
10	ANOVA						
11		df	SS	MS	F	Significance F	
12	Regression	1	10.86495309	10.86495309	267.6543319	1.96274E-07	
13	Residual	8	0.324745817	0.040593227			
14	Total	9	11.18969891				
15							
16		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
17	Intercept	-57.1930531	3.464488033	-16.5083708	1.82953E-07	-65.18218204	-49.2039242
18	ln(Design A Cycles)	4.2524822	0.259929377	16.36014462	1.96274E-07	3.653083597	4.851880811
19	Beta (or Shape Parameter) = 4.25						
20	Alpha (or Characteristic Life) = 693,380						
21							

Now scroll to the right and click once on the graph. Stretch the graph by clicking and dragging on the handle in the lower right corner. From the menu bar, with the graph still selected, click View . Chart Window. Reformat the graph according to your preferences. It's best to use a solid line and no point markers for the Predicted line and delete the legend. Also, move the horizontal axis by clicking on the vertical axis with the right mouse button, selecting "Format Axis," clicking on the "Scale" tab and changing the "Value (X) Axis Crosses At" to -3 (see Figure 4).

**Figure 4: Line Fit Plot**



In cell A19, type the label: Beta (or Shape Parameter)=. In cell B19, type the formula: =B18. In cell A20, type the label: Alpha (or Characteristic Life)=. In cell B20, type the formula: =EXP(-B17/B18). Your results should closely resemble Figure 3. For Design A,  $b=4.25$  and  $a=693,380.4$

An identical analysis using the Design B data yields a  $\beta = 2.53$  and an  $\alpha = 723,105$ .

### Interpreting the results

The Weibull shape parameter, called  $\beta$ , indicates whether the failure rate is increasing, constant or decreasing. A  $\beta < 1.0$  indicates that the product has a decreasing failure rate. This scenario is typical of "infant mortality" and indicates that the product is failing during its "bum-in" period. A  $\beta = 1.0$  indicates a constant failure rate. Frequently, components that have survived burn-in will subsequently exhibit a constant failure rate. A  $\beta > 1.0$  indicates an increasing failure rate. This is typical of products that are wearing out. Such is the case with the spring housings-both designs A and B have  $\beta$  values much higher than 1.0. The housings fail due to fatigue, i.e., they wear out.

The Weibull characteristic life, called  $\alpha$ , is a measure of the scale, or spread, in the distribution of data. It so happens that  $\alpha$  equals the number of cycles at which 63.2 percent of the product has failed. In other words, for a Weibull distribution  $R(\alpha) = 0.368$ , regardless of the value of  $\beta$ . For example, with Design A housings, about 37 percent of the housings should survive at least 693,380 cycles.

While this is interesting, it still doesn't reveal whether either jack-in-the-box design meets the reliability goal of  $R(400,000) \geq 0.90$ . For this, you need to know the formula for reliability assuming a Weibull distribution:

$$R(t) = e^{-\left(\frac{x}{\alpha}\right)^\beta}$$

where  $x$  is the time (or number of cycles) until failure.

The formula looks intimidating, but by simply plugging in the known values for  $\alpha$ ,  $\beta$  and  $x$ , you can obtain the desired reliability estimate.

$$R(400,000)_{\text{Design A}} = e^{-\left(\frac{400,000}{693,380}\right)^{4.25}} = 0.908$$

$$R(400,000)_{\text{Design B}} = e^{-\left(\frac{400,000}{723,105}\right)^{2.53}} = 0.800$$

Computing the above formulas can be confusing and laborious using a calculator. Besides, you can't visualize or compare the reliability of each design for multiple cycle values. Excel provides a better way.

### Creating a reliability calculator worksheet

1. From your Design A regression output worksheet, highlight and copy cells A19:B20. Activate a new worksheet and locate the cursor in cell A1. Select Edit . Paste Special, click on Values, and click OK. This will paste your  $\beta$  and  $\alpha$  labels and values into cells A1:B2 of the new worksheet. Resize the columns as needed.
2. In cell D1, type the label: Cycles.
3. In cells D2:D11, type the values 100,000-1 million in increments of 100,000.
4. In cell E1, type the label: Survival Probability.
5. In cell E2, type the formula: =WEIBULL(D2,\$B\$1,\$B\$2,TRUE).
6. Copy cell E2 down through cell E11.
7. In cell F1, type the label: Reliability.
8. In cell F2, type the formula: =1-E2.
9. Copy cell F2 down through cell F11.
10. Reformat cells as desired. Compare your worksheet with the top portion of Figure 5.

Figure 5: Weibull Reliability Calculator—Design A					
A	B	C	D	E	F
1	Beta (or Shape Parameter) =	4.25	Cycles	Survival Probability	Reliability
2	Alpha (or Characteristic Life) =	693,380	100,000	0.0003	0.9997
3			200,000	0.0050	0.9950
4			300,000	0.0280	0.9720
5			400,000	0.0919	0.9081
6			500,000	0.2204	0.7796
7			600,000	0.4176	0.5824
8			700,000	0.6470	0.3530
9			800,000	0.8407	0.1593
10			900,000	0.9518	0.0482
11			1,000,000	0.9913	0.0087
12					
13		Reliability	Cycles		
14		0.01	992,975		
15		0.1	843,624		
16		0.5	636,122		
17		0.9	408,458		
18		0.99	235,056		

You've now created a Weibull reliability calculator. You supply the  $\alpha$ ,  $\beta$  and cycles of interest, and Excel calculates the reliabilities for you. By merely changing the inputs in cells B1, B2 and D2:D11, you can get reliability estimates for any Weibull distribution of interest.

Likewise, sometimes you'll need to compute the number of cycles (or time to failure) corresponding to a certain reliability level. For example, 99 percent of Design A housings will have failed by how many cycles?

Unfortunately, Excel doesn't have an inverse Weibull function. To perform this calculation (called solving for "critical values"), follow these steps:

1. On your Weibull reliability calculator worksheet, type in the label and values as shown in cells C13:C18 in Figure 5.
2. In cell D13, type the label: Cycles.
3. In cell D14, enter the formula:  $=\$B\$2*(-LN(C14))^{(1/\$B\$1)}$ .
4. Copy cell D14 down through D18.

We find that for Design A of the jack-in-the-box,  $R(992,975)=0.01$ , or 99 percent of the housings will have failed by 992,975 cycles.

### Creating a survival graph

Perhaps the best way to compare the reliability of Design A with that of Design B is by using a survival graph. This line graph depicts the survival probabilities of each housing type at various numbers of cycles. Using the formulas discussed above, enter the data into a new worksheet (see Figure 6). Use the Chart Wizard to construct an X-Y scatterplot. Select line styles of your choice and delete the point markers. The resulting survival graph looks like Figure 7.

Figure 7 allows a comprehensive comparison of the two designs' survival rates. Note that at 400,000 cycles, about 90 percent of Design A housings have survived, whereas only about 80 percent of Design B housings have survived. Therefore, for the stated reliability goal of  $R(400,000) \geq 0.90$ , Design A is clearly superior. However, about 10 percent of Design B housings will survive to 1 million cycles, vs. fewer than 1 percent of Design A. This graph clearly shows the importance of defining the reliability goal in order to choose the more desirable design.

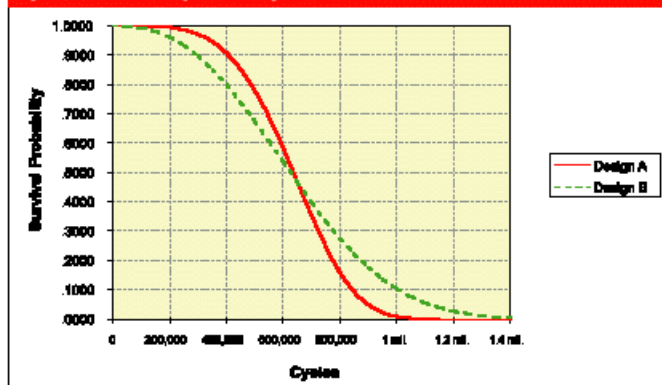
### A warranty example

Having settled upon Design A as the superior alternative, suppose your company plans to offer a warranty on the jack-in-the-box. Of course, you would want to allocate suitable funds to honor the warranty, so as not to be blindsided by unexpected warranty costs. You've decided to set the warranty period so that no more than 1 percent of the units sold would fail before the warranty period expires. How can you determine what length of warranty to offer?

**Figure 6: Survival Data for Designs A & B**

	A	B	C
1	Cycles	Design A	Design B
2	0	1.0000	1.0000
3	50,000	1.0000	0.9988
4	100,000	0.9997	0.9933
5	150,000	0.9985	0.9815
6	200,000	0.9950	0.9620
7	250,000	0.9870	0.9342
8	300,000	0.9720	0.8976
9	350,000	0.9468	0.8526
10	400,000	0.9081	0.7997
11	450,000	0.8529	0.7399
12	500,000	0.7796	0.6749
13	550,000	0.6884	0.6063
14	600,000	0.5824	0.5360
15	650,000	0.4678	0.4660
16	700,000	0.3530	0.3981
17	750,000	0.2475	0.3339
18	800,000	0.1593	0.2749
19	850,000	0.0928	0.2219
20	900,000	0.0482	0.1756
21	950,000	0.0220	0.1361
22	1,000,000	0.0087	0.1032
23	1,050,000	0.0029	0.0766
24	1,100,000	0.0008	0.0556
25	1,150,000	0.0002	0.0394
26	1,200,000	0.0000	0.0273
27	1,250,000	0.0000	0.0184
28	1,300,000	0.0000	0.0121
29	1,350,000	0.0000	0.0078
30	1,400,000	0.0000	0.0049

**Figure 7: Survival Graph for Designs A and B**



The established Weibull model shows 99 percent of the housings should survive at least 235,056 cycles (see Figure 5). Market research shows that a heavily used jack-in-the-box is cycled 100 times per day. We find that 235,056 cycles equates to about 6.4 years of use.

$$\frac{235,056 \text{ cycles}}{100 \frac{\text{cycles}}{\text{day}} \times 365 \frac{\text{days}}{\text{year}}} = 6.4 \text{ years}$$

Armed with this information, and knowing that the competition only offers a two-year warranty on its jack-in-the-boxes, your company might choose to be conservative and offer a five-year warranty. This would ensure domination of the competition from a marketing standpoint, yet still



allow for warranty costs to stay at or below the desired levels.

The above example is somewhat simplistic. Interested readers can find more sophisticated illustrations of warranty strategy using Weibull analysis in academic articles, such as Jayprakash Patankar and Amitava Mitra's "Effects of Warranty Execution on Warranty Reserve Costs" (*Management Science*, 1995).

## A brief statistics overview

Weibull analysis involves fitting a data set to the following cumulative distribution function (cdf):

$$F(x) = 1 - e^{-\left(\frac{x}{\alpha}\right)^\beta}, \text{ for } x > 0$$

Confusion has arisen in the past due to the lack of standardized nomenclature for the Weibull cdf. Its creator, Waloddi Weibull, himself published multiple versions of this formula using different nomenclatures. Arthur Hallinan Jr. provides an excellent history of the various forms of the Weibull distribution in "A Review of the Weibull Distribution" (*Journal of Quality Technology*, 1993).

The format above is the most commonly accepted one. Unfortunately, in Excel, the "Help" screen for the "=WEIBULL" function gives the formula with the  $\alpha$  and  $\beta$  parameters reversed (i.e., the characteristic life is labeled  $\beta$  and the shape parameter is labeled  $\alpha$ ).

## Conclusion

The Weibull distribution's strength is its versatility. Depending on the parameters' values, the Weibull distribution can approximate an exponential, a normal or a skewed distribution.

The Weibull distribution's virtually limitless versatility is matched by Excel's countless capabilities. An astute data analyst who understands the theory behind a given analysis can often get results from Excel that others might assume require specialized statistical software. With Excel, Weibull analysis lies well within reach for most engineers with a statistics background.

## For more information

The Excel file used in this article and an explanation of estimating Weibull parameters are available from our Web site at [www.qualitydigest.com/jan99/html/weibull.html](http://www.qualitydigest.com/jan99/html/weibull.html).

## Notes

1. For simplicity, this article deals with complete failure data, i.e., all samples were tested until they failed. In practice, reliability data analysis frequently involves censored data, or samples for which, for one reason or another, failure times are unknown. Often, tests are suspended before all samples fail. Or perhaps items may fail due to a cause other than the one being studied.

The issues involved in analyzing and interpreting censored data are complex. Improper analysis of censored data can yield misleading results, which Margaret Mackisack and Ronald Stillman point out in "A Cautionary Tale About Weibull Analysis" (*IEEE Transactions on Reliability*, 1996). For further technical details about analyzing censored life data, readers also can consult Wayne Nelson's book *Applied Life Data Analysis* (John Wiley & Sons, 1982) or William Meeker and Luis Escobar's book *Statistical Methods for Reliability Data* (John Wiley & Sons, 1998).

2. For a full explanation of why you can expect a straight line, see this article at our Web site: [www.qualitydigest.com/jan99/html/weibull.html](http://www.qualitydigest.com/jan99/html/weibull.html).

3. Many methods exist for estimating Weibull distribution parameters from a set of data. This article uses the method called probability plotting. Readers interested in other methods, such as maximum likelihood estimation or hazard plotting, should consult Nelson's book, Meeker and Escobar's book or Bryan Dodson's book *Weibull Analysis with Software* (ASQ Quality Press, 1994).

4. Some software packages may give slightly different parameter estimates than the ones in this article. That is because these applications regress the transformed median ranks ( $Y$ ) on the transformed lifetimes ( $X$ ) rather than vice versa. "Simulation studies show that  $Y$  on  $X$  regression produces almost double the bias in the estimation of the shape parameter as the  $X$  on  $Y$  regression," according to Dodson. Moreover, the universal convention for displaying a Weibull probability plot is to depict " $\ln(\text{lifetime})$ " on the horizontal axis. The regression method presented in this article automatically generates the plot in this standard format.

5. The type of Weibull distribution discussed in this article is called the two-parameter Weibull

distribution. This simple form is adequate for a majority of Weibull analysis scenarios. However, if the transformed failure data plot has a curved rather than a straight line appearance, or if  $\beta$  is found to be greater than 6.0, then a third parameter may be needed to adequately model the data. The third parameter, included in the aptly named three-parameter Weibull distribution, effectively shifts the entire distribution to the right. This location parameter is most commonly called  $\gamma$  (the Greek letter gamma). In practice,  $\gamma$  can be interpreted as the earliest possible time at which failure may occur. Of course,  $\gamma$  may never be larger than the value of the earliest failure from the data set. Readers who encounter a curved regression plot or a  $\beta$  value greater than 6.0 should consult Hallinan's article or John McCool's article "Inference on the Weibull Location Parameter" (*Journal of Quality Technology*, 1998) for guidance on fitting a three-parameter Weibull model.

### About the author

*William W. Dorner is a senior quality/process engineer at Best Access Systems in Indianapolis. He is a Certified Quality Engineer and a member of ASQ and ASA.*

For more information, you may e-mail Dorner at [wdorner@qualitydigest.com](mailto:wdorner@qualitydigest.com)

[\[QD Online\]](#) [\[Choose CMM\]](#) [\[CMM Guide\]](#) [\[CE Mark\]](#) [\[weibull\]](#)



Copyright 1999 QCI International. All rights reserved. Quality Digest can be reached by phone at (530) 893-4095. E-mail: [Click Here](#)