Problem Of The Month
August 2001---Weibull Problem #1 For Steel Yield Strength

Download this problem of the month as a PDF file (300KB) revised 10/3/01.

Waloddi Weibull's original 1951 paper, was published by ASME . This paper told about a distribution which became know as the Weibull distribution—a widely used tool in reliability circles, The ASME paper presented seven problems from widely different fields to show applicability and versatile for solving problems. This problem of the month discusses his first example for yield strength of steel.

Weibull's seven problems were solved the hard way by slide ruel, hand-crank calculators, graph papers, and the application of plenty of brainpower. The wide variety of problems were:

1. Yield strength of a Bofors steel (Bofors was a European producer of steel). **←This problem is described below**.
2. Size distribution of fly ash. ←See http://www.barringer1.com/sep01prb.htm
3. Fiber strength of Indian cotton.
4. Length of cytoidea (Worm length for ancient sedimentary deposits).
5. Fatigue life of a St-37 steel
6. Statures for adult males, born in the British Isles
7. Breadth of beans of Phaseolux Vulgaris.

Weibull used these examples to demonstrate the new distribution "may sometimes render good service".

Details of the Weibull distribution are described in the ASME Journal of Applied Mechanics, September 1951, pages 293-297 along with an important discussion published June 1952 on pages 233-234.

In the original papers, Waloddi Weibull has some interesting quotations:

> "…this distribution functions has no theoretical basis."
>
> "…the only practicable way of progressing is to choose a simple function, test it empirically, and stick to it as long as none better has been found"
>
> "The author has never been of the opinion that this function is always valid."
>
> "On the contrary, he very much doubts the sense of speaking of the "correct distribution functions, just as there is no meaning in asking for the correct strength values of an SAE steel, depending as it does, not only on the material itself, but also upon the manufacturer and many other factors. In most cases, it is hoped that these factors will influence only the parameters. However, accidentally they may even affect the function itself."

Weibull's first example in the paper is very difficult because it describes a three-parameter situation. Weibull's first example involves the yield strength of steel from Bofors which practical people know always has a minimum value which is far distance from the zero value. Thus a statistical distribution, which would identify this "failure free" strength value, would be very helpful as the normal distribution covers the range of negative infinity to positive infinity and if you have every tried to describe negative strengths to a crowd of doubter—then you know why the Weibull distribution is helpful for preserving your reputation among the doubters.

The three-parameter distribution is described by Dr. Abernethy in The New Weibull Handbook. In modern parlance:

the shift of the origin is known as t_0 which Dr. Weibull identified it as x_u,

the shape factor slope of the distribution is referred to a β whereas Dr. Weibull labeled it as m.

the characteristic value as η compared to Weibull designation as x_o.

Why the nomenclature changed during the past 50 years is lost in antiquity.

Table 1 from Weibull's paper is shown below as Table 1. Table 1 contains raw data highlighted in yellow for yield strength and the cumulative observed values . The expected values column is a result of his back calculation and the normal distribution column is likewise a back calculation using the normal distribution which 50 years ago was the standard of excellence.

Weibull's Table 1: Yield Strength Of A Bofors Steel (x - yield strength in 1.275 kg/mm^2

Plot Point	Yield Strength	Expected values	Observed Values	Normal distribution
n_i	x	n	n	n
1	32	10	10	8
2	33	36	33	28
3	34	84	81	71
4	35	150	161	141
5	36	224	224	225
6	37	291	289	301
7	38	340	336	351
8	39	369	369	376
9	40	383	383	386
10	42	389	389	388

Today, we would see the 389 data points in "age-to-failure format" along with the "number of occurrences". (Isn't it strange to see a data set with 389 data points?—well not when you consider the use of the mean ranks method of plotting which will employ a term N+1 which will be 390—a value which is easy to set on slide rules which were the "high tech" tool of preference in 1951!) The 389 data are reduced to only 10 points for plotting.

The data in a frequency table would appear as (yield_strength * occurrences):
32*10
33*23
34*48
35*80
36*63
37*65
38*47
39*33
40*14
42*6
You can copy this frequency data from this web page and paste it directly into WinSMITH Weibull. The data has coarse intervals so the WinSMITH Weibull feature of "Inspection Data" and "Mean Ranks" must be employed to get better analysis.

If you're using the **demonstration version** of WinSMITH Weibull, you'll get a correct analysis but *the input data will not be exact,* as the no-cost demo version will randomize input data. For authentic results use the full strength software.

Contrast the [Copy], [Paste] effort with modern software to the large effort required in 1951 to work out the problem the hard way (as illustrated below)! Also remember in 1951 it required a very high degree of engineering skill to solve these problems as compared with the ease today in solving the problems with software! (Lest you forget, to put the software into high quality, accurate output, the details must be explained in terms the programmers can use to get the correct output—so even today, you must have a high degree of engineering skills in a few knowledgeable people or else you get screwed-up and erroneous output from the software.)

Table 2: Data For X-axis Plot		
	x_u = 30.25	Use For X-axis on regression plot
x	$(x-x_u)$	$\log_{10}(x-x_u)$
32	1.7500	**0.24304**
33	2.7500	**0.43933**
34	3.7500	**0.57403**
35	4.7500	**0.67669**
36	5.7500	**0.75967**
37	6.7500	**0.82930**
38	7.7500	**0.88930**
39	8.7500	**0.94201**
40	9.7500	**0.98900**
42	11.7500	**1.07004**

Weibull's paper tells the x_o correction is t_0 = 38.57. You will quickly conclude that subtracting the x_o correction from yield strength gives you a big problem with imaginary numbers when logarithms are used. Such is not the case. Use the clue in Table 1 that you must apply a 1.275 factor to the x_o value (I found this clue as I tried to get a $x_u = t_0$ correction that did not produce negative values and wondered why Weibull had such strange units in his Table 1). The correct correction is obtained from 38.57/1.275 = 30.25. Therefore subtract 30.25 from the yield strength numbers. Table 2 shows how the calculations are made for finding "X" values for his regression plot which are highlighted in the right hand column of Table 2.

Also note the middle column of Table 2 for $(x-x_o)$. Do you really think Mother Nature arranges for this ordered event to occur in such a regularly occurring manner? The answer is it's unlikely. Weibull's problem #1 is ***probably simulated*** to illustrate the technique (Weibull's paper says "…was chosen for purposes of demonstration only"), and the results of yield strengths "look" reasonable to most people (if my opinion is correct, the data is simulated---this **does not lessen** the utility of the distribution or the validity of the technique).

The next step in solving Weibull yield strength problem is to find plotting position for values, which represent the percentage of occurrences represented by the $\log_{10}\log_{10}$ scale on the vertical axis of the regression plot. The details for the methodology are not described at any place in Weibull's paper! Likewise, the numerical figures on the Y-axis and just inside the plot are not described. However, it is clear the divisions of the regression graph are uniform.

We can assume Weibull *probably* would have used "mean ranks" plotting position in 1951. In later years he switched to Benard's median rank plotting position for his regression techniques. The mean ranks method was a common tool for pre-computer calculations because it is simple to use (although 50 years of experience has shown the median rank estimate by Benard is a better plotting position which Dr. Weibull found to be an agreeable replacement for plotting position). Mean ranks methodology employs a n/(N+1) term where N is the total number of data points and n is the cumulative plot point and for this case the denominator is 390—an easy number of manual calculations. The plotting position details for the Y-axis are shown in Table 3

Table 3: Data For Y-axis Plot

Plot Point	Observed Values	cum $p = n_i/(N+1)$	1/(1-p)	$\log_{10}(\log_{10}(1/(1-p)))$
n	n_i	N = 389		Use For Y-axis
1	10	0.0256	1.0263	**-1.94765**
2	33	0.0846	1.0924	**-1.41571**
3	81	0.2077	1.2621	**-0.99522**
4	161	0.4128	1.7031	**-0.63596**
5	224	0.5744	2.3494	**-0.43068**
6	289	0.7410	3.8614	**-0.23155**
7	336	0.8615	7.2222	**-0.06617**
8	369	0.9462	18.5714	**0.10341**
9	383	0.9821	55.7143	**0.24204**
10	389	0.9974	390.0000	**0.41348**

Table 2 for X-values and **Table 3 for Y-values** give pairs of data for plotting. The paired data in format for copying to WinSMITH Visual is given in (X*Y).

The dataset can be copied from this web page and pasted directly into the software for regression. If you're using the full strength version of WinSMITH Visual, you'll get authentic results. If you're using the demonstration version of WinSMITH Visual, it will slightly randomize the data

The data used for WinSMITH Visual is (X-data * Y-data):
0.24304*-1.94765
0.43933*-1.41571
0.57403*-0.99522
0.67669*-0.63596
0.75967*-0.43068

0.82930*-0.23155
0.88930*-0.06617
0.94201*0.10341
0.98900*0.24204
1.07004*0.41348

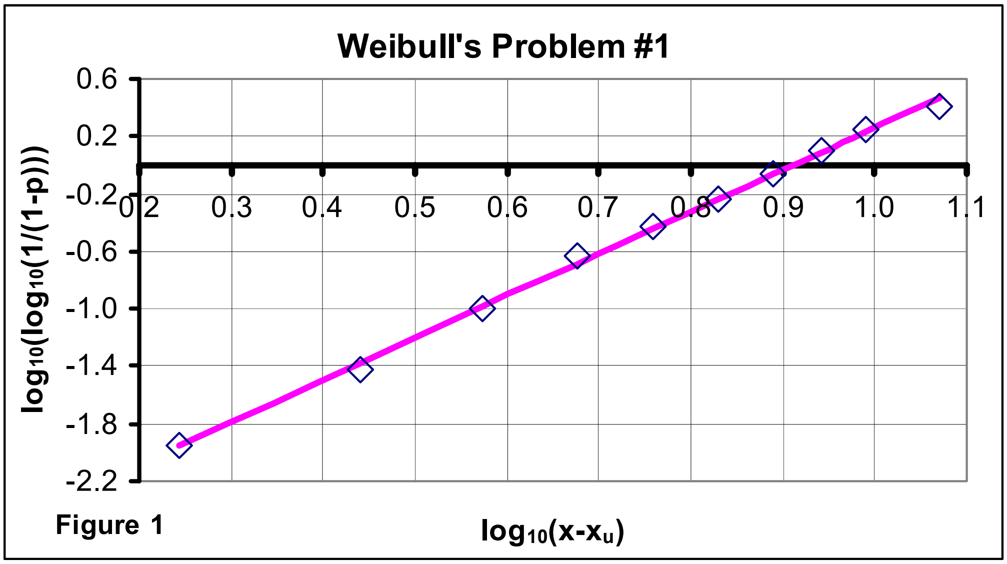

Figure 1 shows a regression with dimensions on the X- and Y-axis which looks similar to Weibull's Figure 1.-

Regressing the Y-axis onto the X-axis as you learned at the university gives a regressed equation Y = 2.91799*X - 2.65934 with r^2 = 0.99864 for the coefficient of determination. The equation is dimensionless.

Regressing the X-axis onto the Y-axis to solve for regression coefficients and then converting it back into a Y = mX+b equation form gives Y = 2.92195*X + 2.662275 and this is today considered best practice for handling errors in the data (although in this particular case, the difference between the two views for regression is very small.

The characteristic value must be calculated from the regression line at p = 63.2% (literally [1-1/e]= 0.63212056…). This gives a Y-value = log(log(1/(1-0.632))) = -0.3622157. Inserting this Y-altitude (-0.3622157) into the regression equation learned at the university shows
–0.3622157 = 2.91799*X-2.65934. Solving for X = 0.78722793 = log($x - x_u$) which is the same as writing the characteristic value as 10^0.78722793 = **6.126781** in the transformed régime. In the scale as recorded or 6.126781 = ($x - x_u$). where x_u = 30.25. Thus x = 6.126781 + 30.25 = **36.3767** in the scale as recorded. This regression of Y onto X is **not too close** to Wallodi Weibull's numbers (remember m = β = 2.918 is invariant in the transform).

Swapping the axis in WinSMITH Visual is easily done by icon click. Thus X is easily regressed on into Y, as is current best practice. The reason for this best practice is because the largest potential for errors occurring in the X-data gives. X= 0.91113 + 0.34224*Y with r^2=0.99864. The equation is dimensionless.

How do you find Weibull's slope and characteristic value from the regression equation X = 0.91113 + 0.34224*Y? Weibull's slope m = β = 1/0.34224 = 2.92195. The characteristic value x_o **in the t_0 domain** occurs at 63.2%. X = 0.91113+0.34224*(-0.3622157) = 0.78716552 which is the same as writing the characteristic values as 10^0.78716552 = **6.12583811 in the transformed régime**.

Converting this value to the **scale as recorded**, the value is 6.12583811+30.25= **36.3758** in the scale as recorded. Thus the regression of X onto Y shows no significant difference between regression methodologies for this case and the results are not too close to Weibull's numbers either and the slope m = β = 2.92195 for regressing X onto Y which today is considered best practice.

So what's the score card after this analysis:

Weibull's slope = m = 2.934, this analysis the slope β = 2.92195

Weibull's characteristic value x_o is given as:

In the x_u domain:

as 8.32 for Weibull's papers and η_{t0}=6.1258 for this analysis using his 32.5 offset in the t_0

In the scale as recorded domain

x_u =38.57 for Weibull and η = 36.3758 for this analysis

Today's Excel/WinSMITH Visual calculated results are not too close to the data calculated from Weibull's paper. Of course one big source of differences lies in the plotting position details for which Weibull is mute in his paper and we assumed he used mean ranks!

Weibull didn't have the advantage of Excel spreadsheets and the Add-In feature of solver to obtain Table 4 {In Excel, under Tools then Options and on the Calculation tab, set Maximum Iterations to 10,000 and Maximum Change to 0.000001}.

Table 4: Use Excel Solver To Find $x_o = t_0$

$x_o = t_0 =$ 30.4006638 **<--Solver changes this number to find $x_u = t_0$**

$\log(x-x_u)$	$\log_{10}(\log_{10}(1/(1-p)))$
0.20393977	-1.94765
0.41486246	-1.41571
0.55622242	-0.99522
0.66269516	-0.63596
0.74813655	-0.43068
0.81950025	-0.23155
0.88077566	-0.06617
0.93446493	0.10341
0.9822412	0.24204
1.06443314	0.41348

Excel solver uses this array

Regress X on to Y		Convert to X_u domain		Scale as recored	
slope	y-intercept	$m = \beta$	$\eta_{\tau o}$	$m = \beta$	$x_u = \eta$
0.35537	0.90313	2.81399	5.94856972	2.81399	36.34923

r^2	
0.99876466	**<--Solver maximizes R^2 by changing $x_o = t_0$**

Table 4 uses Excel solver to find the optimum value of $x_o = t_0$ and the correct answer is $x_o = t_0 = 30.4006638$ (see the yellow block in the upper left hand corner of Table 4). This produces $m = \beta = 2.81399$ with a $x_u = \eta = 36.34923$ in the scale as recorded or 5.94856972 in the $x_o = t_0$ domain.

Assume that the calculated results from Excel's solver are the "True" values (because all the results and all the calculations are visible for audit inspections). Watch the variation in the parameters for various inputs to
$x_o = t_0$ for various conditions:

1) Weibull $x_o = t_0 = 30.25$
2) Excels $x_o = t_0 = 30.4006638$ (selected as the datum)
3) WinSMITH Weibull's $x_o = t_0 = 30.3688$ which is obtained quickly by means of a robust method that produces a good answer but not necessarily the perfect answer.

The comparison results are summarized in Table 5 using Excel solver as the datum and you can see the yellow blocks from use of the spreadsheet calculations shown in Figure 4.

Table 5: Results For x_0 = t_0's Using Table 4			
	Solver	WSW	Weibull's x_o
$x_o = t_0 =$	30.400664	30.368800	30.250000
R^2=	0.998765	0.998759	0.998643
xo = t_0 domain for $x_u = \eta$	5.9486	5.9863	6.1258
Recorded Scale for $x_u = \eta$	36.3492	36.3551	36.3758
$m = \beta$	2.8140	2.8370	2.9340
Errors			
$x_o = t_0 =$	Datum	0.1048%	0.4956%
R^2=	Datum	0.0006%	0.0122%
xo = t_0 domain for $x_u = \eta$	Datum	-0.6336%	-2.9800%
Recorded Scale for $x_u = \eta$	Datum	-0.0160%	-0.0732%
$m = \beta$	Datum	-0.8177%	-4.2648%

Finding the true value of $x_o = t_0$ requires huge amounts of data. The New Weibull Handbook recommends not less than 20 data points of well-behaved data! Remember in this case we have 389 data, which are in 10 groups of data points. The data must be viewed as "course intervals" as you would get with attribute data as compared to variables data where the yield strengths would have been reported with several digits beyond the decimal point.

If the data is not well behaved, i.e., scatter, then the $x_o = t_0$ requirement may be for 100's of data points and/or for you must increase the measurement divisions between the data so as to avoid coarseness in the data. This issue requires some use of good engineering judgment.

Look at the yield strength data from beside of Table 1. Put it into a Weibull probability plot as shown in Figure 2. You will be impressed with a simple plot of the 10 data points and the steepness of the slope **m = β = 20.69**. The New Weibull Handbook tells you to beware of more information hiding in steep betas. In this case, a click of the dimensions distribution analysis icon shows the data set needs a $x_o = t_0$ correction.

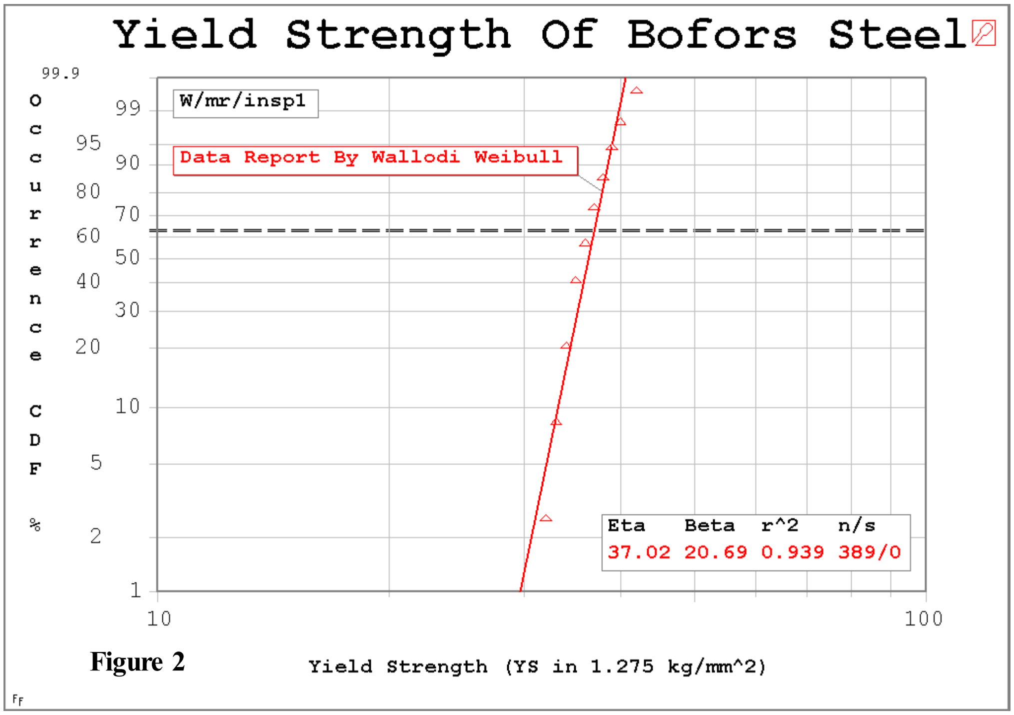

Figure 2

When you apply a $x_o = t_0$ correction to the data, you need two satisfying conditions:

1. More than 20 data points for the regression
2. A physical relationship justifying the change

We're weak on the first requirement for 20 data points for the regression because of course data, we only have 389 data going into the 10 stacked data points shown in Figure 2.

We're strong on the second requirement as it is well known that steel has a failure free stress level for a given grade of steel along with it's surface treatment and heat treatment when used in the proper environment under stated conditions.

When the X-axis is transformed into the scale as recorded following a $x_o = t_0$ correction, the probability trend line is curved. From the curve, you get a good idea of the x_o or t_0 value as the lower end of the concave curve approaches $x_o = t_0$ value asymptotically. This is shown in Figure 3 by taking the data in Figure 2 and asking WinSMITH Weibull to find the $x_o = t_0$ and plotting the data in the scale as recorded.

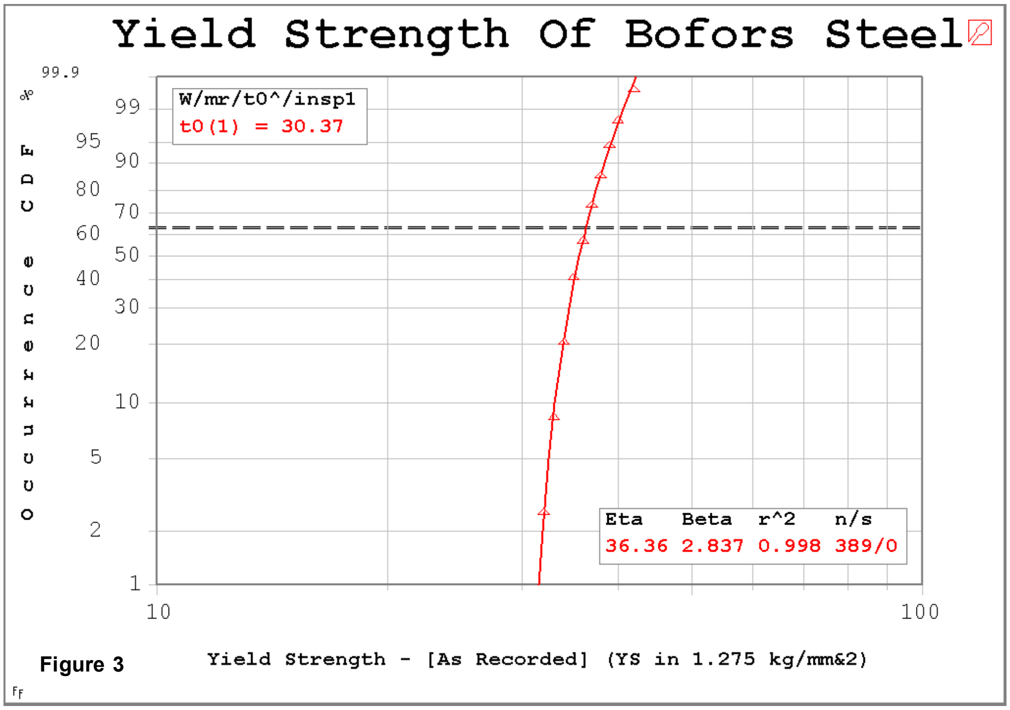

Figure 3

Figure 4 shows the same data as Figure 2 and Figure 3, **but** it is displayed in the $x_o = t_0$ regime. If you consult the data used for this plot (assuming you were to apply the $x_o = t_0 = 30.25$) you could get the same results shown in the first two columns of Table 2.

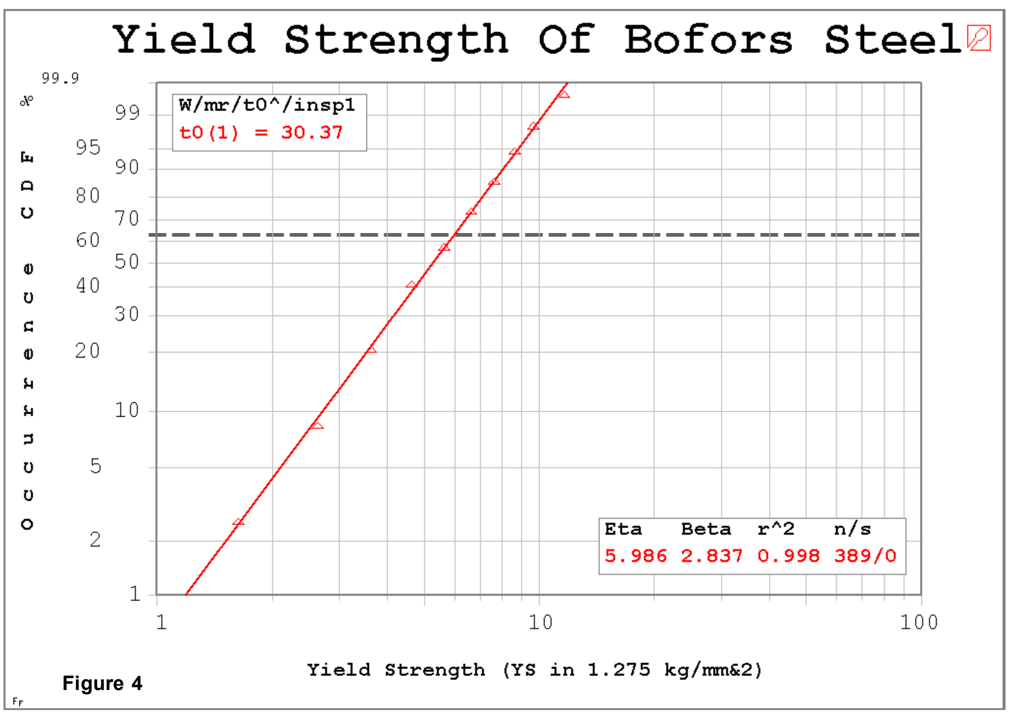

Figure 4

Finally we have calculated Weibull's original results with fair agreement given the fact that no details were provided in for his plotting position. We have carried the analysis further by calculating a better value for $x_o = t_0$ correction and to show the sensitivity of how it alters the results in the Weibull parameters.

We must also acknowledge that 50 years of applications and experience (hindsight) with the technology provide greater information into the Weibull analysis technology than Dr. Weibull had at his disposal.

Thanks to **Wes Fulton** of Fulton Findings for showing me the errors in my spreadsheet where I tried to make Weibull calculations the hard way. Also thanks to **Carl Tarum** for reminding me that Excel Solver needs more iterations and stiffer demands for better accuracy by resetting the default conditions.

Do not underestimate the complexities that each engineer had to struggle with before the days of software and Weibull handbooks. Now with smart engineers, like Dr. Bob Abernethy who writes The New Weibull Handbook and Wes Fulton who writes WinSMITH Weibull software, they have organized much complexity into a well written text and easy to use software, so that the entire effort appears easy and you too can look like an expert. We thank them for their efforts. If you plan to write your own Weibull spreadsheets, you too can also screw up the results (as I did a couple of times) and think—who's going to clean up your mess?—the morale of this embarrassment is to use validated software and validated methodology.

Comments:

Refer to the caveats on the **Problem Of The Month Page** about the limitations of the following solution. Maybe you have a better idea on how to solve the problem. Maybe you find where I've screwed-up the solution and you can point out my errors as you check my calculations. E-mail your comments, criticism, and corrections to: Paul Barringer by **clicking here**.

Last revised 10/3/2001

Simple Tools Strong Results **Return to Barringer & Associates, Inc. homepage**