ORiON, Vol. 3, No. 1, pp. 23-41 *ISSN 0259-191X*

TWO SIMPLE TOOLS FOR INDUSTRIAL O.R.

KEITH SANDROCK

UNIVERSITY OF THE WITWATERSRAND
Nº 1 JAN SMUTS AVENUE
JOHANNESBURG 2001

ABSTRACT

At the 1985 Annual Congress of the South African Production & Inventory Control Society it was pointed out that the productivity growth rate for South Africa is completely out of kilter with that for the western industrialised nations. The latter all display positive rates (some as high as that of Japan) whereas the rate for South Africa is - NEGATIVE.

Partly as a result of this situation, more and more attention is being given to quality control and reliability engineering by our industrialists in their attempts to improve productivity.
This is going hand in hand with the introduction of better techniques and better use of the latest technology.

We should also give attention to analytical tools that may be used in a simple inexpensive way to improve our methods of analysing industrial data, and in this way to improve our performance at little or no additional cost.

To this end two tools are discussed. They are by means new. But it does seem as though they could be more widely applied in the industrial milieu.

INTRODUCTION

The reasons for fitting a distribution to data have been clearly stated by Hahn and Shapiro [6]. These are:

1) The desire for objectivity
2) The need for automating the data analysis
3) Interest in the values of the distribution parameters.

To this list we may also add:

4) Making probability statements.

To precisely these ends, various empirical distributions are being used by industrial analysts daily. Foremost of them is the normal distribution on which most forms of quality control charts are based, and which is also the de facto standard for hypothesis testing.
The Poisson distribution is used extensively for acceptance sampling procedures because it is somewhat easier to handle than the Binomial. And the number one distribution for reliability testing and analysis is our old friend the negative exponential distribution.

There are of course many variations on the above themes. For instance the lognormal distribution is sometimes resorted to if the range of the data is several powers of 10 or if they display a long right hand tail such as for example metal fatigue and electrical insulation life.

There are also the instances for which the use of the normal distribution is suspect. One such instance is that in which data are obtained in terms of absolute units. These data may be distributed according to the folded normal distribution in which case the use of normal probability tables for hypothesis testing etc. should be exercised with caution.

With regard to the use of the negative exponential distribution for the analysis of life data, the main advantage is that it is a single parameter distribution (which makes it easy to use). The Weibull distribution on the other hand has a great variety of shapes which makes it much more flexible. It empirically fits many kinds of data. And it is not as difficult to use as its rather formidable expression for the probability density function would suggest.

In this paper two tools are discussed in the hope that the analyst will find them useful additions to his tool kit (if he is not already using them). They are the Folded Normal Distribution and the Weibull Distribution. Both are easy to use and require no specialized knowledge beyond that already possessed by the average analyst working with industrial data.

A] THE FOLDED NORMAL DISTRIBUTION.

It may be opportune to reconsider our widespread use of the normal distribution for drawing inferences about data which exhibit a characteristic Gaussian (humped) frequency curve.
Very often our assumption of normality is quite valid. Over the years it has been demonstrated time and time again that many physical measurements are closely approximated by the normal distribution.
Christopher Chatfield (Bath University) says: "Indeed non-normality is so rare that it is a useful clue when it does occur" - ref [9] page 93.

The problem with using the normal approximation occurs when we examine data which have been recorded in absolute units. When for example the deviation from a specified norm is recorded without algebraic sign.
The measurement of out-of-round or out-of-true (skewness) are typical examples. In cases such as these, where all defects are positive, if we regard the data as having a normal distribution then we are ignoring the fact that strictly speaking the distribution requires them to exhibit both positive and negative deviations from the norm.

In fact what occurs is that there is a geometric folding of the distribution as shown in Figure 1.

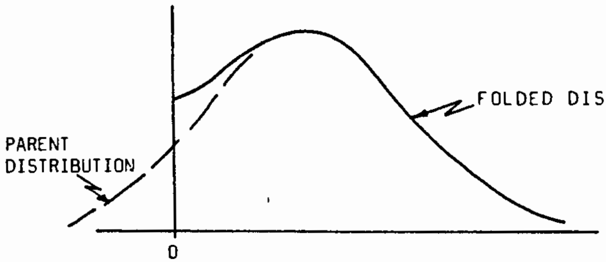


FIGURE 1

Three questions naturally arise. They are:

1) How serious is this geometric folding?
2) To what extent will it nullify inferences based on the use of normal tables?
3) What alternative methods for analysis should be used?

The answers to these questions are:

Firstly, if the mean of our data divided by the standard deviation is greater than about 2,0 say then there is no problem associated with using the normal approximation as long as the data satisfy requirements which we would apply in any case to confirm the hypothesis that they are normally distributed.*

Secondly, if the mean divided by the standard deviation is less than 2,0 then we should exercise caution even if the data satisfy our usual normality test criteria.

Thirdly, there is a very simple procedure for analysing data from a folded normal distribution. This method is presented below. It does not require the analyst to make use of the folded distribution per se but instead it allows him to map the parameters obtained from the field data on to a parent normal distribution which is then used as the reference distribution for hypothesis testing and for drawing inferences.

The best way of illustrating the method is to consider an example.

EXAMPLE:

Consider an inspection process in which measurements are recorded in absolute units.
Assume that the data clearly exhibit a Gaussian shape and that we are satisfied (on closer examination which need not be dealt with here) that they are approximately normally distributed.
The mean is $\tilde{X}$ = 30,3 units and the standard deviation is S = 21,43 units.
In the usual course of events we would draw inferences etc using our well worn tables of the normal distribution. However, having been alerted to the fact that some folding may have occured, we form the ratio $\tilde{X}/S$ and find that it is equal to 1,4.
This value indicates to us that we need to consider an alternative approach to the use of the tables.
This approach is as follows:

* This is not necessarily so if we are interested in the small percentiles of the distribution, say the 1% point. Throughout this paper it is assumed that we are dealing with percentiles in the region of 5%.

STEP 1: With $\bar{X}/S = 1,4$ enter Table 1 below.

It will be found that we have $S/S_N = 0,753$ where S_N is the standard deviation of the parent normal distribution which we seek.
From this ratio we have:

$$S_N = 28.46.$$

Similarly, from row two of the table the value of the mean of the parent distribution is found to be $\bar{X}_N = 23,44$.

This simple step results in our being in possesion of a parent normal distribution which will serve as a reference distribution. The mean is 23,44 units and the standard deviation is 28,46 units.

STEP 2: Use the parent distribution and the normal tables for hypothesis testing or for drawing inferences in the usual way.

COMMENT: As expected, the parent normal distribution has a lower mean value and a higher standard deviation than the original data.

Assuming that we wish to determine a 95% upper limit for the inspection process how would we go about this?

ANSWER: $\text{Prob}(x_i \leqslant X) = 0,95$ and for the one-tailed case $z = 1,64$.

Hence:

$$X = (1,64)(28,46) + 23,44 = 70,12.$$

In the long run 95% of our measurements will be less than or equal to 70,12. Only 5% are expected to exceed this value.

Similarly, if we wish to establish 2σ limits in order to monitor quality or in order to construct a control chart, then we have:

$$\bar{X} + 2\sigma = 2(28,46) + 23,44 = 80,36.$$

COMPARISONS.

At this point some comparisons may be useful.

Firstly, if we use the folded normal distribution per se to determine our 95% upper limit we obtain a value of 70,7 (in a much less straightforward way than the above method).
Obviously 70,7 and 70,12 are close enough for most practical quality control situations.

Secondly, if we had not found the parent distribution and used the sample values of $\bar{X}$ and S as the parameters for our reference distribution the 95% upper limit would be $(1,64)(21,43) + 30,3 = 65,44$.

This is a more stringent quality limit than 70,12 and by a fairly significant margin.

CONCLUSION:

In a number of industrial applications we may find it necessary to regard our data as being approximated by a folded normal distribution and not a normal distribution.
When this is the case we should relate the folded distribution to a parent normal distribution before testing hypotheses, or designing control charts, or drawing inferences etc.

The method for doing so is described above and is sufficiently accurate for most practical quality control situations. There are of course much more sophisticated techniques which make use of the folded distribution - but they provide only marginally greater precision. They are really not worth the effort.

TABLE 1

$\bar{X}/S$	=	1,320	1,340	1,360	1,380	1,400	1,450
S/S_N	=	0,620	0,673	0,708	0,733	0,753	0,793
$\bar{X}_N/S_N$	=	0,265	0,517	0,653	0,748	0,824	0,975
$\bar{X}/S$	=	1,500	1,550	1,600	1,650	1,700	1,750
S/S_N	=	0,824	0,848	0,869	0,886	0,901	0,934
$\bar{X}_N/S_N$	=	1,097	1,203	1,298	1,386	1,468	1,545
$\bar{X}/S$	=	1,800	1,850	1,900	1,950	2,000	2,500
S/S_N	=	0,924	0,934	0,942	0,949	0,956	0,989
$\bar{X}_N/S_N$	=	1,619	1,690	1,758	1,825	1,889	2,469.

THIS TABLE IS TAKEN FROM REF [2] WHICH ALSO EXPLAINS THE THEORY UNDERLYING THE APPROXIMATION.

The use of TABLE 1 makes life easy since it greatly simplifies the application of the folded normal distribution.

B] THE WEIBULL DISTRIBUTION.

In 1939 Wallodi Weibull introduced his distribution as follows:

"Consider a variable X which is random and has cdf F(x).
We may write any F(x) in the form:

$$F(x) = 1 - \exp(-\varphi(x)) \text{ where the function } \varphi(x)$$

must be:
1) Positive.
2) Non-decreasing.
3) Zero at a value x_u where x_u is not necessarily zero.

The simplest function satisfying these criteria is:

$(x - x_u)^m/x_0$ and we may therefore write:

$$F(x) = 1 - \exp[-(x - x_u)^m / x_0]$$

The only merit of this cdf is that it is the simplest expression of the appropriate form.
And since we cannot hope to expect a theoretical basis for distribution functions of random variables such as the strength of steel, or particle sizes etc. we should choose a simple function, test it empirically, and stick to it as long as no better has been found."

There is no doubt that the Weibull distribution is an attractive one to deal with. It has come to play a very important part in reliability theory and in quality engineering.
In an early paper (1964) J N Berrettoni [7] demonstrates the wide spread of applications of the distribution for describing empirical data.
The applications he describes are:

1 Corrosion resistance of magnesium alloy plates.
2 Return goods classified by number of weeks after shipment. That is the number of weeks it took the customer to return a defective product.
3 Number of down times per shift.
4 Leakage failure of dry batteries.
5 Life expectancy of ethical drugs.
6 Reliability of electric motors.
7 Reliability of solid tantalum capacitors.

Another early application is due to J H K Kao (Technometrics 1959) [8] on the life-testing of electron tubes.

More recently, B P Gu et al. (Materials Science & Engineering 1986) used it to describe the AL,Li particle distribution in binary AL-Li alloys.

These are only a few examples. Many more are to found in the literature.

Generally we deviate from the original way of writing the distribution and use the following Greek notation instead:

$$F(x) = 1 - \exp[-(x/\alpha)^{\beta}] \text{ for } x > 0$$

Where: α = characteristic life-time of the function. This is the life-time at which 63,2% of the items have failed. The matter of characteristic life-time is referred to again later.

β = the shape parameter of the distribution.

There are two aspects which we will immediately recognise. The first is that this is a shifted exponential distribution when $\beta = 1$ and when $\tau \neq 0$. And we have our well known negative exponential distribution so often used for life-testing when $\beta = 1$ and $\tau = 0$.
Not so obvious is the fact that when $\beta = 3,5$ the distribution approximates the normal distribution.

In the above formulation of F(x) we have assumed that $x_u = 0$
If this is unrealistic then the following modification must be introduced:

$$F(x) = 1 - \exp[-((x - \tau)/\alpha)^{\beta}] \quad \tau > 0 \text{ and } x > 0.$$

What this implies is that our data start at a point $\tau > 0$ and not at zero.

In many life-time situations this is usually the case. There is a certain minimum time below which no items have been observed to fail.

The first step in using the Weibull distribution is of course to estimate the parameters. This is relatively simple. There is no need to resort to the use of Weibull probability paper.

STEP 1. Tabulate the data and determine the cumulative frequencies (cum f). From the cum f find the cumulative probabilities F(x).

STEP 2. Calculate $\ln(x_i)$ and $\ln[\ln\{1/(1 - F(x_i))\}]$ for all i.

STEP 3. Optional. Plot $\ln(x_i)$ on the horizontal axis and the values of $\ln[\ln\{1/(1 - \hat{F}(x_i))\}]$ on the vertical axis using ordinary linear graph paper.

Your points should form a straight line. If not then revise your value of τ as per the procedure described on page 12.

A common occurence is that some points fall on a straight line whereas others fall on another straight line. In this case there may be more than one set of parameters. We will return to this situation later.

STEP 4. Perform a linear regression analysis.

Obtain the parameters in the model:

$\ln[\ln\{1/(1 - \hat{F}(x))\}] = a + b\ln(x).$

We then have:

$$1 - F(x) = \exp - [x/\exp(a)^{-1/b}]^b$$

$$= \exp - [x/\alpha]^\beta$$

Where: $\alpha = [\exp(a)]^{-1/b}$

$\beta = b$

Note: In the foregoing we have assumed that $\tau = 0$.

EXAMPLE:

Eighty specimens of Kevlar belting which is used on a popular belt driven motor cycle were tested to the limit of serviceability.

The results are tabulated below.

The lifetime is in units of 500km.

LIFETIME	=	20	30	40	50	60	70	80	90	100	110+
N° OF FAILED SPECIMENS	=	6	8	6	14	11	10	7	5	4	9
CUM f	=	6	14	20	34	45	55	62	67	71	80

Using these data the following table is obtained as described in STEP 2.

$\hat{F}(x)$	$\ln(x_i)$	$\ln[\ln\{1/(1-\hat{F}(x_i))\}]$
0,075	3,00	-2,55
0,175	3,40	-1,65
0,250	3,69	-1,25
0,425	3,91	-0,59
0,563	4,09	-0,19
0,690	4,25	0,16
0,780	4,38	0,42
0,840	4,50	0,61
0,890	4,61	0,79
1,000	4,70	N/A

A linear regression analysis is then performed as required by STEP 4. The results of this analysis for the above data are:

$$a = -8{,}9$$

$$b = 2{,}12 \quad \text{also } r = 0{,}998$$

The estimates of the parameters of the approximating Weibull distribution are:

$$\alpha = 66{,}98 \quad \text{and} \quad \beta = 2{,}12.$$

Note: In the above table it will be observed that when $\hat{F}(x) = 0{,}632$ then the value in the last column is $\simeq 0$. We expect this since the characteristic life of the distribution is the lifetime at which 63,2% of the items have failed.
This is also the value of the parameter α.

Having performed STEPS 1 to 4 and obtained values for α and β we are in a position to answer a number of relevant questions concerning the experiment.

Q 1. The warranty period on these belts is 35 000km. How many are expected to fail before the warranty expires?

ANS: $1 - \{F(x) = \exp[-(70/66{,}98)^{2{,}12}]\} = 67\%$.

Q 2: How do we determine percentiles? What is the 10% percentile?

ANS: Let x_p be the p^{th} percentile.

Then: $x_p = \alpha[-\ln(1-p)]^{1/\beta}$

Hence $x_{,1} = 66,98[-\ln(0,90)]^{1/2,12} = 23,2.$

Q 3: What is the most likely life of these belts, i.e. the mode?

ANS: Let x_m be the mode.

then $x_m = \alpha[1 - 1/\beta]^{1/\beta}$

$= 66,98[1 - 1/2,12]^{1/2,12} = 49,57 \approx 50.$

Q 4: What is the hazard function of this distribution? *

ANS: The hazard function is given by $h(x) = (\beta/\alpha)(x/\alpha)^{\beta-1}$ for $x > 0$

This is a very useful function. From the above expression it is apparent that it can describe failure rates which are increasing, decreasing or constant. The clue is the value of β.

The correspondence is:

a) Increasing rate with age	$\beta > 1$
b) Decreasing rate with age	$\beta < 1$
c) Constant	$\beta = 1$**

In the Kevlar belt example we have $\beta > 1$ and hence the belts have an increasing failure rate with age. They tend to wear out.

A table of the Kevlar belt data together with the Weibull estimates is given below. The approximation is a good one, as confirmed by the χ^2 test.

KEVLAR BELT EXAMPLE - RESULTS OF WEIBULL FIT

LIFE = x =	20	30	40	50	60	70	80	90	100	110
OBSERVED $\hat{F}(x)$ =	0,075	0,175	0,250	0,425	0,563	0,690	0,770	0,840	0,890	1,00
WEIBULL $F(x)$ =	0,07	0,17	0,29	0,42	0,55	0,67	0,77	0,85	0,90	0,94

* The hazard function is a measure of the proneness to failure as a function of the age of the item. It is defined as follows:

Hazard Function $h(y) = f(y)/[1 - F(y)]$

** Remember that when this parameter is unity we have the negative exponential distribution - it is the unique distribution with a constant hazard rate.

A NOTE ON THE ESTIMATION OF τ.

It is frequently found in practice that the plot resulting from STEP 3 is NOT a straight line. In these situations we need to obtain a value of τ such that when $\ln(x - \tau)$ is used in place of $\ln(x)$ a straight line is in fact obtained.

This requires a lot of trial and error, particularly when graphical procedures are resorted to exclusively.
And even when a straight line does result we may ask ourselves: How straight is straight? Is the choice of τ a good one? The optimum perhaps?

These questions can be answered to some degree by making use of the fact that when $\ln[\ln\{1/(1 - \hat{F}(x))\}] \approx -7$ then $F(x) = 0$ for all practical purposes.

Let $a + b\ln(x - \tau)_0 = -7$

Then $\ln(x - \tau)_0 = (-7 - a)/b$

$(x - \tau)_0 = \exp[(-7 - a)/b]$

If $(x - \tau)_0 \simeq 0$ our choice of τ is a good one.

Naturally one cannot adopt this little procedure blindly. The analyst will need to bring his knowledge of the system that is generating the data to bear on the problem of finding τ.

In the Kevlar belt example we have:

$(x)_0 = \exp[(-7 + 8{,}9)/2{,}12] = 2{,}45 \neq 0.$

Our knowledge of the system leads us to disregard the small value of $(x)_0$ obtained above and to leave $\tau = 0$.

COMPLEX WEIBULL DISTRIBUTIONS.

A number of researchers have commented on the sensitivity of the Weibull plot and its ability to show up heterogeneous and/or mixed distributions. Weibull himself found them when he was investigating the fatigue life of steel samples. Berretonni found dichotomies in the leakage failure of batteries. Investigation brought to light the fact that early leakage occured at the top of the battery whereas later failures were due to bottom leaks. Berretonni also found that ethical drug failure has two phases. In the first phase the failure per se is increasing at a decreasing rate which terminates at a high point and then remains constant.
In the case of tests on capacitors it has been found that the failure rate decreases initially and then enters a second phase with a further decrease in failure rate. Interestingly enough this implies that the product actually tends to improve with age. This is the well known burn-in phenomenon.

EXAMPLE OF A CASE IN WHICH WE NEED TO DETERMINE A COMPLEX DISTRIBUTION.

Consider the following data:

LIFE = x	$\hat{F}(x)$	ln(x)	$\ln[\ln\{1/(1-\hat{F}(x))\}]$
10	0,022	2,300	-3,81
15	0,130	2,710	-1,97
20	0,522	3,000	-0,30
25	0,739	3,220	0,295
30	0,826	3,400	0,662
35	0,913	3,560	0,893
40	0,939	3,690	1,029
45	0,961	3,810	1,177
50	0,983	3,910	1,405
55	0,991	4,010	1,550
60	1,000	4,090	N/A

Casual examination of these data may not reveal the fact that they describe a complex distribution. In many cases such as this one, if we plot the original frequency distribution using x and f(x) then we see the situation depicted in Figure 2 below whereas the real situation may be similar to that shown in Figure 3.

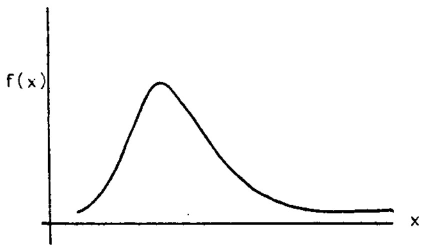


FIGURE 2

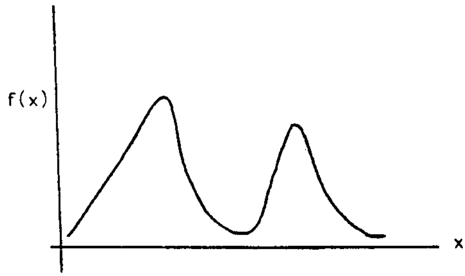
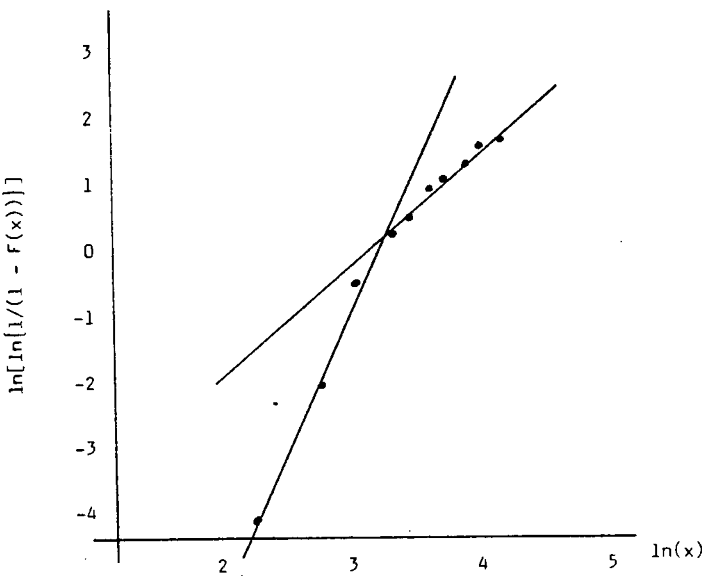


FIGURE 3

When the last two columns of the above table are plotted we have the following result:



Clearly there are two straight lines that can be drawn through the data. From these separate regression models two sets of Weibull parameters can be obtained.

A simple distribution is just not good enough for modelling this complex failure situation.
Fortunately the Weibull distribution lends itself admirably to complexity. It will be shown below that it is fairly simple to obtain mixed models and composite models of two (or more) Weibull distributions.

It is possibly for this reason that many researchers have opted for the Weibull instead of choosing lognormal, Pearson type III, or one of the Johnson distributions.
After all, there is little theoretical justification for choosing it above the others.

There are two commonly used complex models.

THE MIXED WEIBULL DISTRIBUTION.

The cdf is given by:

$$F(x) = \Sigma p_i F_i(x) \quad ; \qquad x > 0 \; ; \; \Sigma p_i = 1$$

STEP M_1:

Plot the $\ln(x)$ and $\ln[\ln\{1/(1 - \hat{F}(x)\}]$ data and obtain regression estimates etc.

STEP M_2: Assuming that we have only two Weibull distributions (as is the case here) formulate the model as follows:

Initial failure is given by:

$F_1(x) = 1 - \exp[-\{(x - \tau_1)/\alpha_1\}^{\beta_1}]$ and wear out failure by:

$$F_2(x) = 1 - \exp[-\{(x - \tau_2)/\alpha_2\}^{\beta_2}]$$

The mixed cdf is:

$$F(x) = p_1 F_1(x) + p_2 F_2(x)$$

$$= 1 - p_1 \exp-[(x - \tau_1)/\alpha_1]^{\beta_1} - p_2 \exp-[(x - \tau_2)/\alpha_2]^{\beta_2}$$

Where p_1 and p_2 are the mixing parameters and $p_1 + p_2 = 1$.

The explanation given to $F(x)$ is that the life of the component is a mixed chance of surviving failure in the initial stages and then surviving wear out later on.

Usually one finds that the initial failure rate is a flatter line than the wear out line if there is no initial burn-in.

The mixing parameters p_1 and p_2 may be obtained graphically or may be computed. In the computational procedure the value of $\ln(x)$ for which the line with the greatest b-value reaches a value of $\ln[\ln\{(1/(1-F(x))\}]$ equal to +2 is used to estimate p_1. The rationale behind the choice of the value +2 is that at this point $F(x) = 1$ for all practical purposes.
We have:

$$\ln(x) = (2 - a_2)/b_2 \quad \text{if } b_2 > b_1.$$

The mixing proportion p_1 is then obtained from:

$$p_1(\text{est}) = 1 - 1/[\exp\{\exp(a_1 + b_1(2 - a_2)/b_2)\}]$$

An analogous expression pertains when $b_1 > b_2$.

EXAMPLE:

Consider the data given on page 13.

The first four points fall on the line with constants $a_1 = -14{,}37$; $b_1 = 4{,}6081$ and incidently we have r_1 = correlation coeff = 0,995.
On calculating $(x)_0$ we find $(x)_0 = \exp[-(a_1 + 7)/b_1] = 5$.

For this situation this is a significant departure from zero.
Hence it is necessary to assign a value other than zero to τ.
Try $\tau = 7$.
Then: $a_1 = -6{,}494$; $b_1 = 2{,}339$; $(r = 0{,}991)$ and $(x - \tau)_0 = 0{,}8$.

This is satisfactory. We find therefore that $\alpha_1 = 16{,}06$; $\beta_1 = 2{,}34$.

In a similar way it is found that:

$$\alpha_2 = 16{,}61 \; ; \; \beta_2 = 1{,}4 \; \text{ and } \; \tau_2 = 5$$

The mixed model is to be composed of:

$$F_1(x) = 1 - \exp-[(x - 7)/16{,}06]^{2{,}339}$$

and

$$F_2(x) = 1 - \exp-[(x - 5)/16{,}61]^{1{,}4}$$

The next step is to find p_1 and p_2, the mixing parameters.

From $p_1 = 1 - 1/[\exp\{\exp(a_2 + b_2(2 - a_1)/b_1)\}]$

we have $p_1 \doteq 0,95$ and hence $p_2 = 0,05$.

The mixed Weibull model is:

$$F(x) = 1 - p_1\exp-[(x - 7)/16,06]^{2,339} - p_2\exp-[(x - 5)/16.61]^{1,4}$$

THE COMPOSITE WEIBULL DISTRIBUTION.

The cdf is:

$$F(x) = F_j(x) \quad ; \quad \delta_j \leqslant x \leqslant \delta_{j+1} \quad j = 0,1,2 \ldots . n.$$

Where: $F_j(x) = 1 - \exp-[(x - \tau_j)/\alpha_j]^{\beta_j}$

The δ_j's are points at which partition occurs.

Note that $\delta_0 = \tau_1$ and that $\delta_{n+1} = \infty$. Hence if we have a distribution with two components there is only one partition parameter.

The value of δ is obtained from:

$$1 - \exp-[(\delta - \tau_1)/\alpha_1]^{\beta_1} = 1 - \exp-[(\delta - \tau_2)/\alpha_2]^{\beta_2}$$

or

$$[(\delta - \tau_1)/\alpha_1]^{\beta_1} = [(\delta - \tau_2)/\alpha_2]^{\beta_2}$$

When graphical procedures are used δ is taken as the intersection of the two Weibull plots.

EXAMPLE:

Referring to the data on page 13 we have the following composite Weibull model:

$$F(x) = 1 - \exp-[(x - 7)/16,06]^{2,339} \quad \text{for } 7 \leqslant x \leqslant 25$$

$$= 1 - \exp-[(x - 5)/16,61]^{1,4} \quad \text{for } 25 \leqslant x < \infty$$

$$= 0 \text{ otherwise.}$$

RESULTS:

		COMPOSITE WEIBULL		MIXED WEIBULL
x	$\hat{F}(x)$	$F_1(x)$	$F_2(x)$	$F(x)$
10	0,022	0,020		0,027
15	0,130	0,180		0,188
20	0,522	0,457		0,463
25	0,739	0,730	0,727	0,729
30	0.826		0,830	0,898
35	0,913		0,900	0,973
40	0,939		0,940	0,992
45	0,961		0,967	0,997
50	0,983		0,982	0,999
55	0,991		0,991	
60	1,000		0,995	

NOTE: The better fit of the composite model indicates that failure is due to two different causes and is not a mixed chance. See also page 16 (top).

A general observation regarding the mixed model (which is borne out above) is that it favours the LHS of the distribution. For this reason it is often employed by life-data analysts who are especially interested in early failure rates.

The mean of the mixed distribution can be shown to be equal to the weighted average of the means of the subpopulations in the proportions p_1 and p_2.*

CONCLUSION

The value of the Weibull distribution and its derivative models has been illustrated using two examples.
Graphical means (which are often the norm for Weibull analysis) have not been resorted to.
The use of Weibull paper is not essential.

It is hoped that the easy-to-use property of the Weibull distribution has been illustrated. You really need nothing more than a cheap pocket calculator. Preferably one which can perform linear regression analysis automatically.

* Obtaining the first moment of the complex distribution is described in ref [8].

REFERENCES:

[1] ELANDT REGINA C, The Folded Normal Distribution: Two Methods For Estimating Parameters from Moments. Technometrics 3 (4) 1961.

[2] LEONE F C ET AL, The Folded Normal Distribution. Technometrics 5 (4) 1961.

[3] NELSON W, Applied Life Data Analysis, Wiley, 1982.

[4] WEIBULL W, An Analysis of the Size of Cyrtoidea, Nature 164 1949.

[5] WEIBULL W, A Statistical Distribution of Wide Applicability. Jnl of Applied Mechanics Sept 1951.

[6] HAHN G & SHAPIRO S, Statistical Models in Engineering, Wiley 1967.

[7] BERRETTONI J N, Practical Applications of the Weibull Distribution. Industrial Quality Control August 1964.

[8] KAO J H K, A Graphical Estimation of Mixed Weibull Parameters in Life Testing of Electron Tubes. Technometrics 4 (1) 1959.

[9] CHATFIELD C Statistics for Technology, Chapman & Hall (1981).