LifetimeReport

Student numbers: 741731,744485

August 30, 2018

1 Exercise

(a)

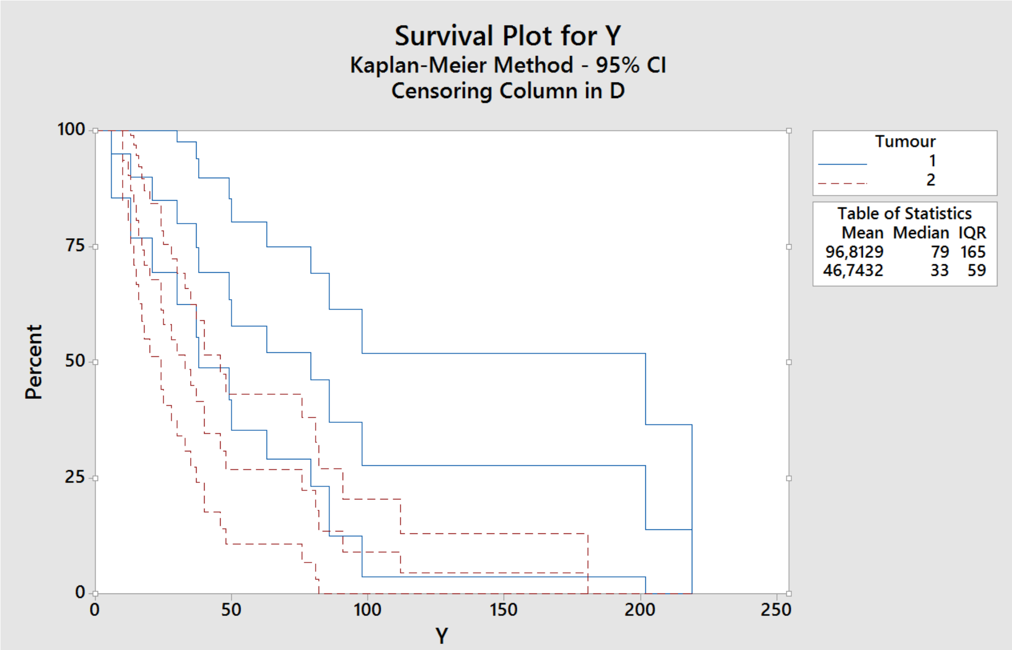

Figure 1: Kaplan-Meier plot of the data sets for Tumour 1 and Tumour 2.

Figure 1 shows the Kaplan-Meier estimator of both Tumour 1 and Tumour 2 with the corresponding 95% confidence intervals. Minitab calculates the median for Tumour 1 and 2 to be 79 and 33 respectively. They are found by checking which value in the KM-estimators that corresponds to 50% on the Y-axis.

The lower and the upper-quartiles are found by finding the corresponding values to 25% and 75% on the X-axis respectively. In Minitab we can check that

$$t_{0.25,1} = 202, \quad t_{0.75,1} = 37$$

$$t_{0.25,2} = 76, \quad t_{0.75,2} = 17$$

This can also be verified by observing that the difference in the values equals the lengths of the IQR's.

(b)

We set the null hypothesis to be that the data for Tumour 1 and Tumour 2 are equally distributed:

$$H_0 : R_1(t) = R_2(t) \text{ for all t>0},$$

and we would like to test it against the hypothesis

$$H_1 : R_1(t) \neq R_2(t) \text{ for all t} > 0.$$

One can check the null hypothesis in two different ways: One can use Minitab to perform the logrank test to find that the p-value of the null hypothesis is 0.006 and that the test-statistic is 7.49659. This results in rejecting the null hypothesis at 5% significance level because $0.006 < 0.05$. Otherwise we may compute the logrank test manually. To check this, we can compute a χ^2-test-statistic for the two tumour samples. If the obtained value for the test-statistic is not sufficiently probable, the null hypothesis is rejected. We use the test statistic as in Slide 6:

$$V = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \tag{1}$$

where O_1, O_2 are the observed deaths of people with Tumour 1 and 2 respectively, and E_1, E_2 are the expected number of deaths in Tumour 1 and 2.

Under the null hypothesis, V is distributed as χ^2_1 and $P(V \geq 3.84) = 0.05$, which means that when we perform the test with a significance level of 5%, the null hypothesis is rejected if $V > 3.84$. The values for E_1 and E_2 are the sum of all $E_{1,i}$ and $E_{2,j}$ in the data set, and it is given that they are equal to 22.48 and 19.52 respectively. $E_{1,i}$ and $E_{2,j}$ can be calculated by the following formula:

$$E_{1,i} = \text{Fail}_i \cdot \frac{\text{Risk1}_i}{\text{Risk}_i}, \qquad E_{2,i} = \text{Fail}_i \cdot \frac{\text{Risk2}_i}{\text{Risk}_i},$$

where Fail_i are the number of patiens that died at t_i, Risk1_i and Risk2_i are the number of patients with Tumour 1 and 2 at risk at time t_i, and Risk_i is the total number at risk.

Table 1: Displaying the calculations for lifetimes up to $t = 15$.

t_i	Risk1	Risk2	Risk	Fail1	Fail2	Fail	$E_{1,i}$	$E_{2,i}$
6	20	31	51	1	0	1	0.392	0.608
10	19	31	50	0	2	2	0.760	1.240
12	19	29	48	0	1	1	0.396	0.604
13	19	28	47	1	1	2	0.809	1.191
14	18	27	45	0	1	1	0.400	0.600
15	18	26	44	0	1	1	0.409	0.591

From the exercise sheet we have that $O_1 = 14$, $O_2 = 28$, $E_1 = 22.48$ and $E_2 = 19.52$. Using Equation 1, we calculate that $V = 6.883$. From Minitab, we know that this corresponds to the p-value 0.0087. This means that we reject the null hypothesis at 5% significance level. Tumour 1 and Tumour 2 does therefore not have the same distribution, and the mortality is different between people with Tumour 1 and 2. The reason that we get another answer than Minitab may be due to different test statistics used.

(c)

Let $\hat{Z}_1(t)$ be the Nelson-Aalen estimator for the cumulative hazard function $Z_1(t)$ of T_1. Then we know from Slide 6 page 5 that we can calculate $\hat{Z}_1(t)$ according to the following formula:

$$\hat{Z}_1(t) = \sum_{T_{(i)} \leq t} \frac{d_i}{n_i},$$

where $T_{(i)}$ are the ordered life times, n_i are the number at risk and d_i are the number of patients to die at lifetime number i. When $T_{(i)}$ is a censoring time, no patients will die, such that only times of deaths will be added in the interval. This gives the calculation of the Nelson-Aalen estimator as shown in Table 2.

Table 2: Calculation of the Nelson-Aalen estimator manually for $t \leq 38$

$T_{(i)}$	d_i	n_i	$\frac{d_i}{n_i}$	$\hat{Z}_1(T_{(i)} \leq t < T_{(i+1)})$
6	1	20	0.050	0.050
13	1	19	0.053	0.103
21	1	18	0.056	0.158
30	1	17	0.059	0.217
31	0	16	0.000	0.217
37	1	15	0.067	0.284
38	1	14	0.071	0.355

Table 3: The Nelson-Aalen estimator calculated by the Minitab macro.

T_i	Cum Haz Nelson	Survival Nelson
6	0.05000	0.951229
13	0.10263	0.902459
21	0.15819	0.853690
30	0.21701	0.804921
37	0.28368	0.753010
38	0.35511	0.701099
49	0.43844	0.645042
50	0.52935	0.588989
63	0.62935	0.532939
79	0.74046	0.476895
86	0.94046	0.390448
98	1.19046	0.304082
202	1.69046	0.184435
219	2.69046	0.067850

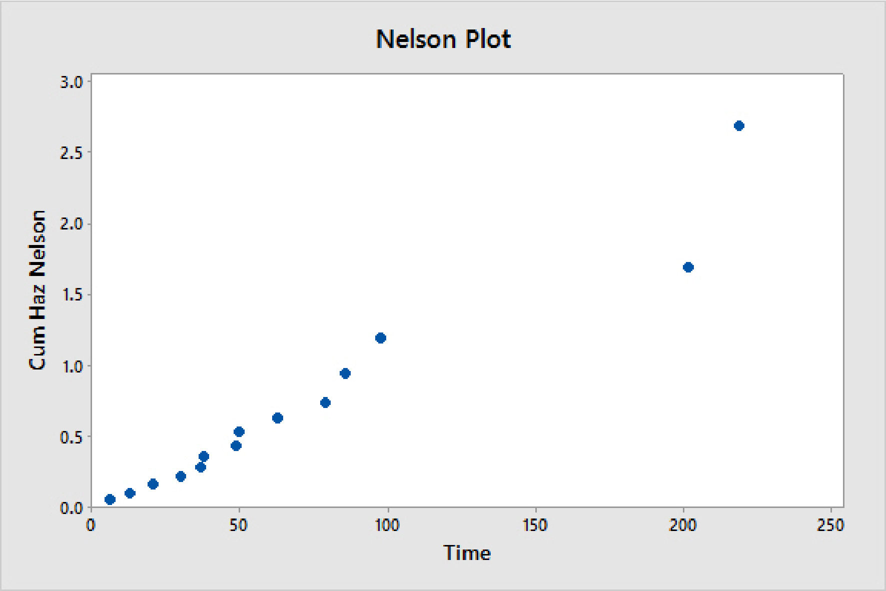

Figure 2: Scatter plot for the Nelson-Aalen estimator. Since the Nelson-Aalen plot is actually a step-function, the value between two points will actually equal the value at the most recent scatter plot point.

The same calculation as we did in Table 2 made by the Minitab macro produces Table 3. The scatter plot of the Nelson-Aalen plot from the same macro is shown in Figure 2. $Z(t)$ seems here to be slightly convex up until t=100, which indicates that $z(t)$ is increasing on this interval. The second last point may show some deviation, but the data still have some tendency of being convex. We therefore say under some uncertainty that the graph displays a failure rate that is slightly increasing.

(d)

First the TTT-statistic (Total Time on Test) for $\mathcal{T}(t_i)$ is calculated manually. The result is shown in Table 4.

Table 4: Total Time on Test (TTT) at time t_i. $\mathcal{T}(t_i)$ is calculated as $\mathcal{T}(t_{i-1}) + (t_i - t_{i-1}) \cdot n_i$, where n_i are the number at risk immediately before t_i.

t_i	d_i	n_i	$\mathcal{T}(t_i)$
6	1	20	120
13	1	19	253
21	1	18	397
30	1	17	550
31	0	16	566
37	1	15	656
38	1	14	670
40	0	13	696

We further perform the Barlow-Proschan's test and make a TTT-plot for Tumour 1:

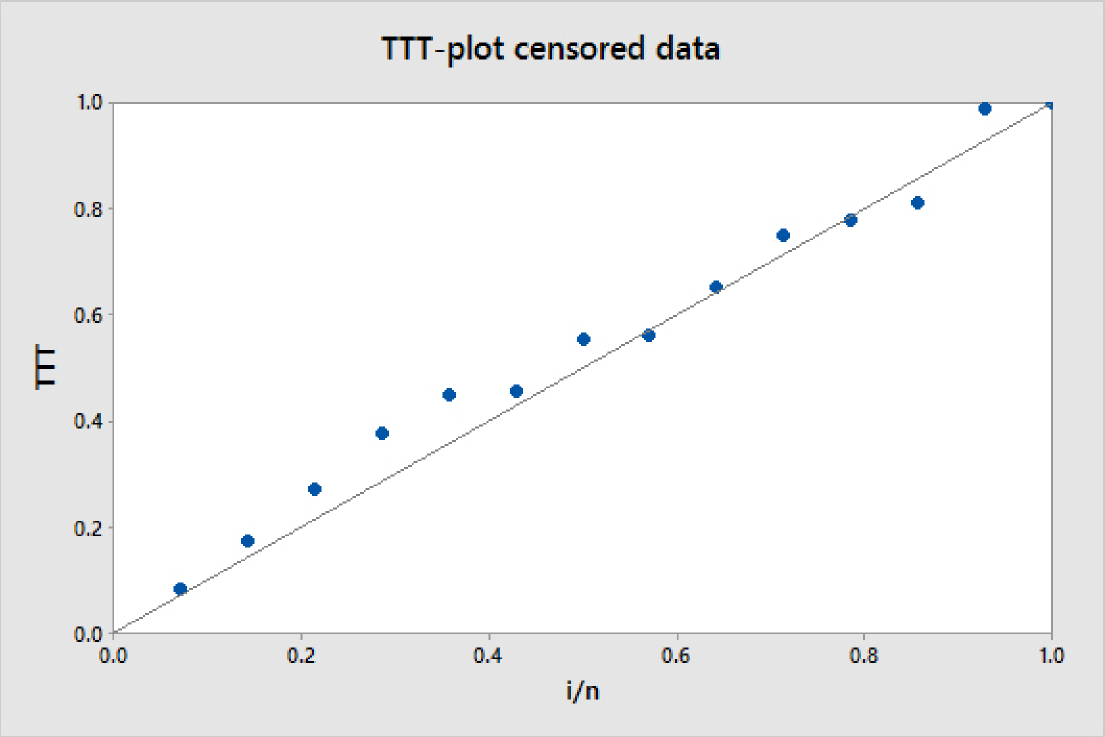

Figure 3: Scatterplot of the normalized TTT values at their failure times. The values are plotted against i/n.

To perform the Barlow-Proschan test we calculate the Barlow-Proschan test-statistic W:

$$W = \sum_{i=1}^{k-1} \frac{Y_i}{Y_k},$$

where Y_i/Y_k are the values of the points in the plot in Figure 3. This is the normalized values for $\mathcal{T}(t_i)$ at the times of death. We set the null hypothesis to be

$$H_0 : T \sim \text{expon}(\lambda).$$

Figure 3 shows the first half of the data set to be above the reference line. In the second half there are only one point that lies clearly below, so it is plausible to test the null hypothesis against the hypothesis

$$H_1 : \text{T is IFR}.$$

We test the null hypothesis at 5% significance level, so we have to check if the test-statistic Z is greater than $z_{0.05} = 1.65$. We calculate

$$Z = \frac{W - \frac{k-1}{2}}{\sqrt{\frac{k-1}{12}}},$$

where $k = 14$ is the number of non-censored failures. From the Minitab macro on the Barlow-Proschan's test, we get that $W = 6.906$, which further gives $Z = 0.390$. This results in not rejecting the null hypothesis at 5% significance level, as $z_{0.05} = 1.65 > 0.390$.

(e)

The Barlow-Proschan test performed in (d) gives a strong indication that the data set are distributed according to an exponential distribution, and it is natural to use this distribution while performing a parametric analysis of the data for Tumour 1. The test does however not guarantee that the data are exponentially distributed, but since the corresponding p-value to $Z = 0.390$ is rather large, it is plausible to assume that the data actually are exponentially distributed at 5% significance level.

We continue by calculate $E[T_1]$:

$$E[T_1] = \int_0^\infty t f_1(t;\theta)\, dt$$
$$= \int_0^\infty \frac{t}{\theta} e^{-t/\theta}\, dt = \left[-te^{-t/\theta}\right]_0^\infty + \int_0^\infty e^{-t/\theta}\, dt = 0 + \left[-\theta e^{-t/\theta}\right]_0^\infty$$
$$= \theta$$

Thus, the expected value of T_1 is θ. Next, we calculate the survivor function $R_1(t;\theta)$ for T_1. The survivor-function is equal to the probability that the works at time t. Therefore we have that

$$R_1(t;\theta) = 1 - \int_0^t f_1(u;\theta)\, du = \int_t^\infty \frac{1}{\theta} e^{-u/\theta}\, du$$
$$= \left[-e^{-u/\theta}\right]_t^\infty = e^{-t/\theta}$$

Next, we calculate the standard deviation of T_1. Observe that

$$E[T_1^2] = \int_0^\infty \frac{t^2}{\theta} e^{-t/\theta}\, dt = \left[-t^2 e^{-t/\theta}\right]_0^\infty + \int_0^\infty 2te^{-t/\theta}\, dt$$
$$= 0 + \left[-2te^{-t/\theta}\right]_0^\infty + \int_0^\infty 2\theta e^{-t/\theta}\, dt$$
$$= 2\theta^2$$

We know that the variance of T_1 can be expressed as

$$Var(T_1) = E[T_1^2] - E[T_1]^2 = \theta^2,$$

and it is well known that the standard deviation is the positive square-root of the variance. Thus,

$$SD(T_1) = \theta$$

Further, the likelihood-function is expressed as follows:

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-t_i/\theta} = \frac{1}{\theta^n} e^{-\left(\Sigma_{i=1}^{n} t_i\right)/\theta}$$

We can from this derive the maximum likelihood estimator of θ by taking the logarithm of L, differentiating with respect to θ, and then solving for θ:

$$\ell(\theta) = \log L(\theta) = -n \log \theta - \frac{\Sigma_{i=1}^{n} t_i}{\theta}$$

$$\frac{d\ell(\theta)}{d\theta} = -\frac{n}{\theta} + \frac{\Sigma_{i=1}^{n}}{\theta^2}$$

Setting the right hand expression equal to zero, we get the following estimator for θ:

$$\hat{\theta} = \frac{\Sigma_{i=1}^{n} t_i}{n}$$

(f)

We first observe in Figure 4 that every point is inside the confidence interval. The first five points shows some deviation from the reference line, but the rest of the data set has a strong correlation with the exponential distribution. Purely based on the probability plot it would be wrong to reject the exponential distribution for Tumour 1, but it may still exists other distributions that has a probability plot that correlates just as good as the exponential.

Using what we did in (e), we start by summing over the test set for Tumour 1. This gives

$$\Sigma_{i=1}^{20} \theta_i = 1462$$

So our ML-estimator for θ is

$$\hat{\theta} = \frac{1462}{14} \approx 104.429,$$

which corresponds to the value of the mean given by Minitab.

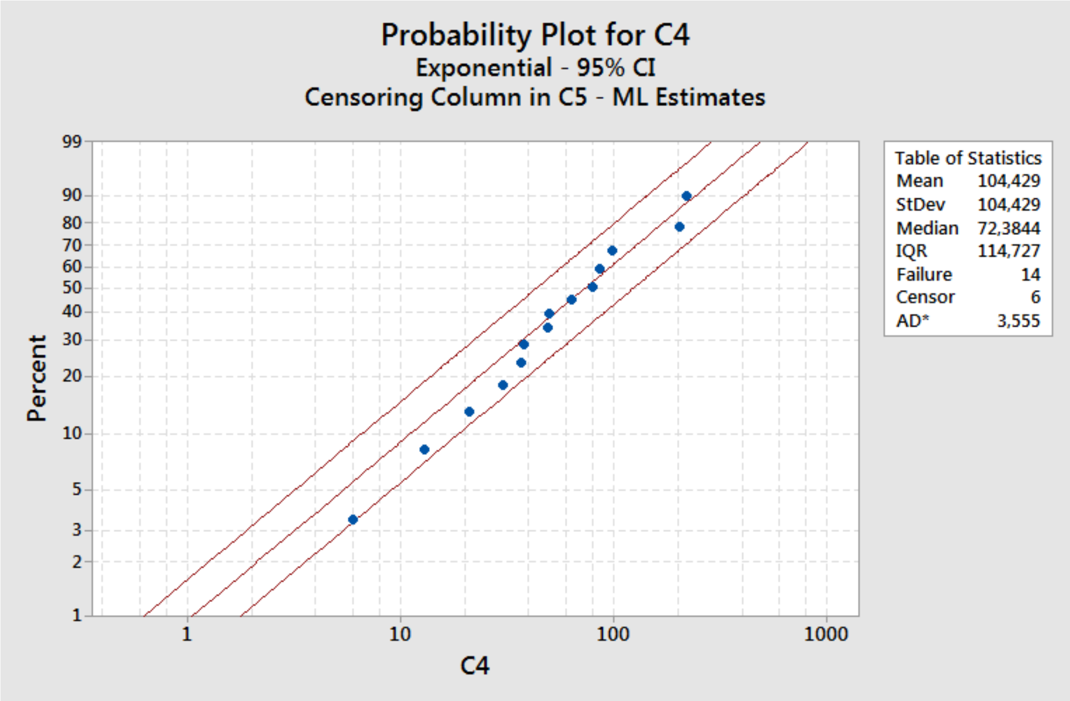

Figure 4: Weibull probability plot for Tumour 1.

(g)

We will restrict ourselves to four different probability plots for the data for Tumour 2.

In (d) the null hypothesis that the data for Tumour 1 might be exponentially distributed was not rejected. We therefore first try the probability plot for the exponential distribution on Tumour 2, as shown in Figure 5. The data clearly does not fit within the 95% confidence interval, so this is not an appropriate distribution.

The probability plot for the Weibull distribution in Figure 6 gives more values within the confidence interval, but all the values should be within when the data set consists of so few values. The plot also does not seem random inside the confidence intervals; It is a clear pattern in the data set, and this distribution is also rejected.

In Figure 7 we consider the lognormal distribution. The probability plot fits the points quite nicely. The points in the outer ends are the points that differ the most, but it is not enough to reject a lognormal distribution for Tumour 2.

The probability plot for the 3-parameter Weibull model in Figure 8 seems also like a god fit to the data. The mid values differ little from the expected value, while the first two points show some deviation from the reference line. Therefore, we are also unable to reject that Tumour 2 has a 3-parameter Weibull distribution. To further conclude in whether the distribution for Tumour 2 is lognormal or 3-parameter Weibull requires that we perform other tests that are more accurate. A bigger data set will also benefit the test. If we had to choose one, it would obviously be the 3-paramter Weibull as the exercise wants us to consider it particularly. It is difficult to make an independent decision from the probability plots alone.

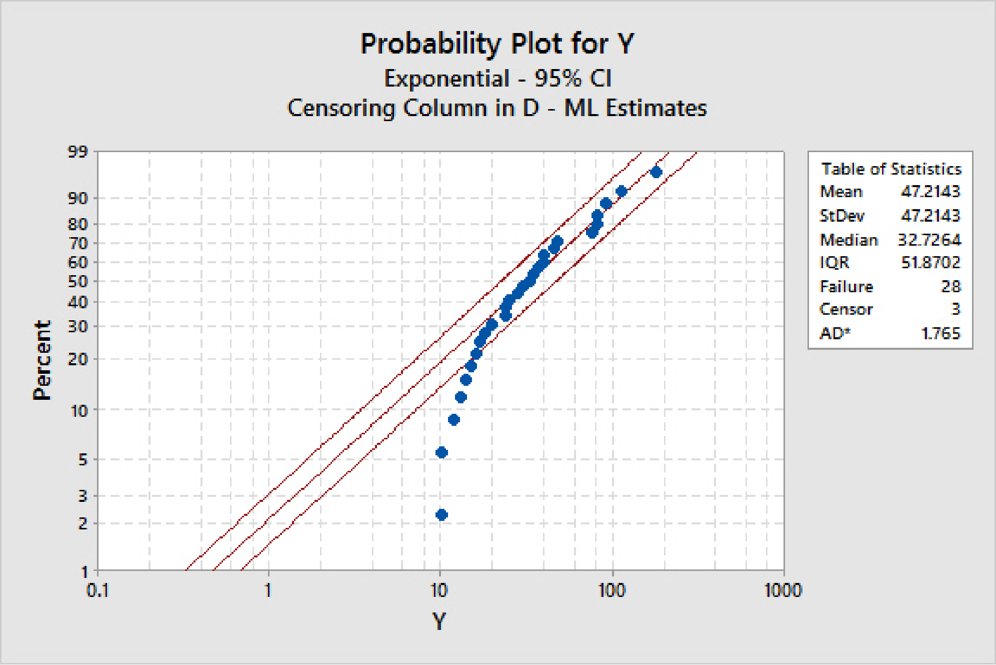

Table of Statistics	
Mean	47.2143
StDev	47.2143
Median	32.7264
IQR	51.8702
Failure	28
Censor	3
AD*	1.765

Figure 5: Exponential probability plot for Tumour 2.

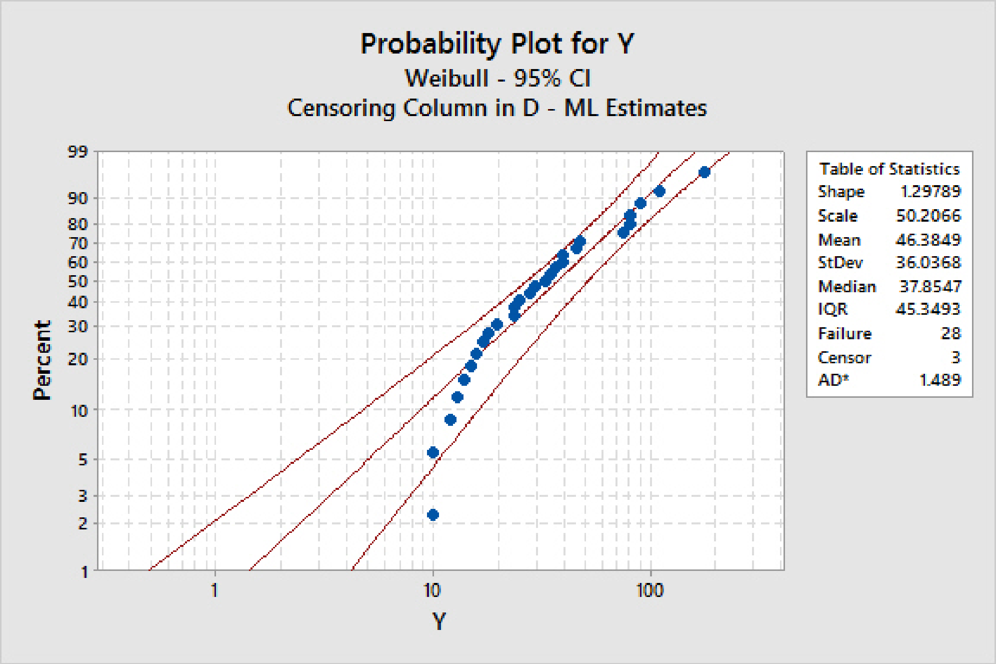

Table of Statistics	
Shape	1.29789
Scale	50.2066
Mean	46.3849
StDev	36.0368
Median	37.8547
IQR	45.3493
Failure	28
Censor	3
AD*	1.489

Figure 6: Weibull probability plot for Tumour 2.

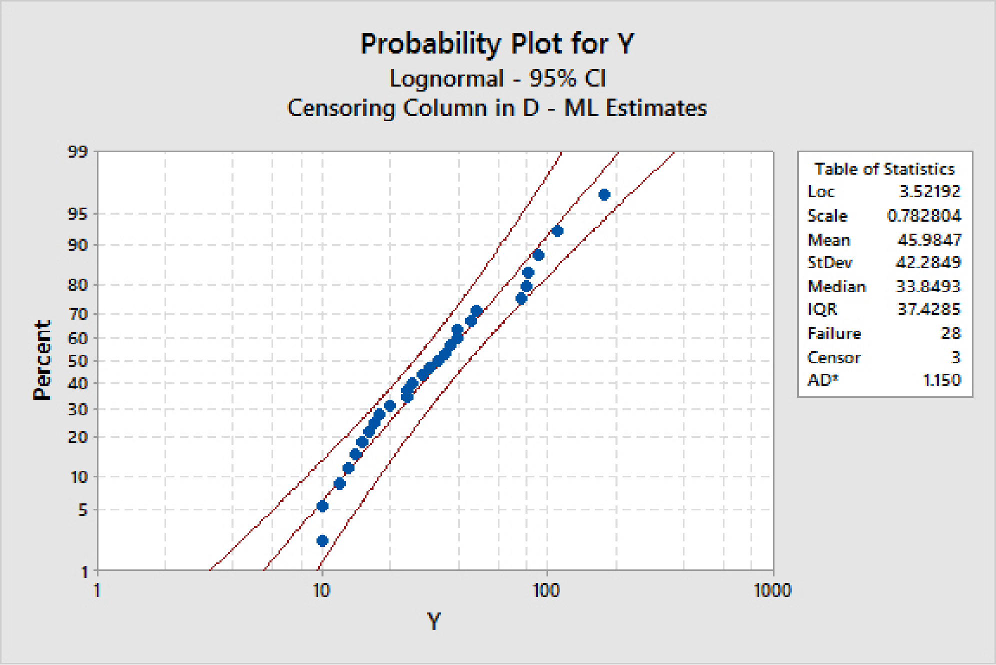

Table of Statistics	
Loc	3.52192
Scale	0.782804
Mean	45.9847
StDev	42.2849
Median	33.8493
IQR	37.4285
Failure	28
Censor	3
AD*	1.150

Figure 7: Lognormal probability plot for Tumour 2.

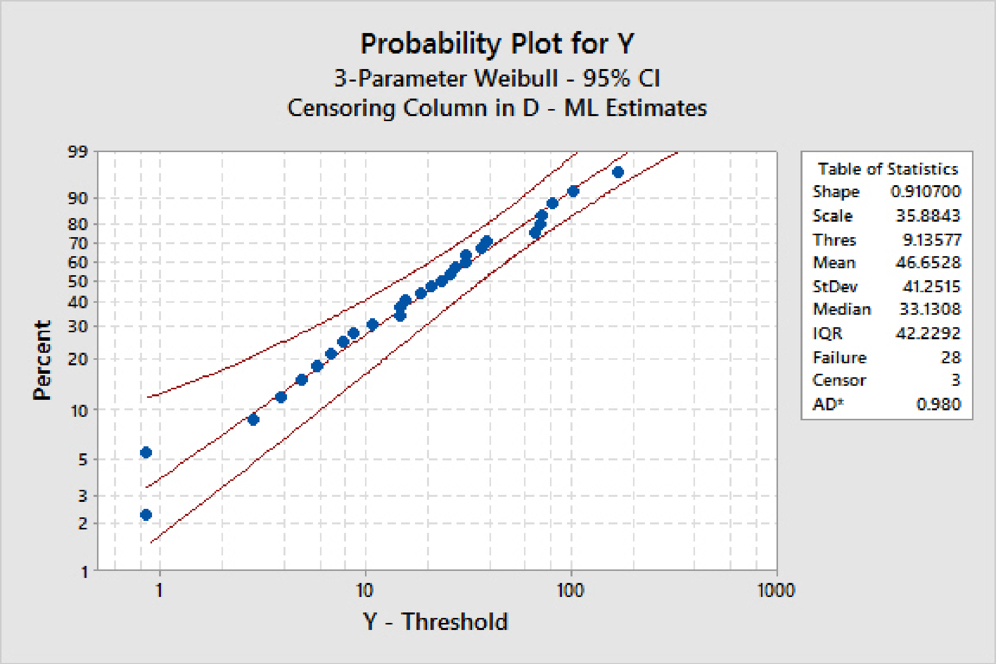

Table of Statistics	
Shape	0.910700
Scale	35.8843
Thres	9.13577
Mean	46.6528
StDev	41.2515
Median	33.1308
IQR	42.2292
Failure	28
Censor	3
AD*	0.980

Figure 8: 3-parameter Weibull probability plot for Tumour 2.