Reliability Analysis

Module 5A: Reliability Data Analysis & Model Selection

Prof. Katrina M. Groth
Mechanical Engineering Department
Center for Risk and Reliability
University of Maryland
kgroth@umd.edu

Objectives for Module 5

- To date, we've talked about how to define reliability and several related quantities, how to use parametric probability distributions, and how to calculate descriptive statistics.
- **Now we discuss:**
  - Types of reliability data and data sources
  - Non-parametric reliability modeling procedures
  - Model selection: how to use observed data to select an appropriate probability model for a set of data
  - Parameter estimation (w/uncertainty) from complete and censored data.

Key Assumptions in Module 5

- We have empirical data & we have verified the data quality.
- Data come from: **iid (Identical & independently distributed) exchangeable observations**
  - Practical interpretation in reliability: the elements of sample are obtained independently and under the same conditions.
- We're uncertain about
  - If a component might fail, when a specific component will fail, etc.
- …so we are trying to select distributions & estimate (unknown) parameters to predict these things.
- We are ignoring: Causes of failures.

Data → models → probabilities & descriptive stats

Some Excel, Matlab & R scripts available

- https://crr.umd.edu/computational-scripts. It can also be found under Research > Selected Publications > Computational Scripts.

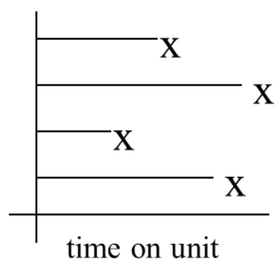
- Also see RARE2011 software
  - If you run into an error opening it in Excel – Be sure you have allowed add-ins.
    - Excel -> Options -> "Add-ins" then select "analysis toolpak VBA" and you'll see the option "go" -- press "go" it should enable.

Types of failure data

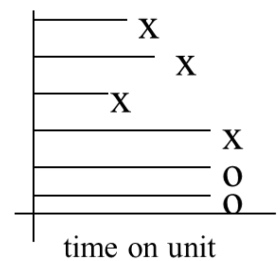
- Reliability data sources:
  - Operational or field data
    - Production and/or customer returns
    - Surveillance, maintenance, and field service
  - Generic databases
  - Testing
    - Reliability tests (prototype, production)
    - Environmental tests
    - Reliability growth tests
- Reliability data characteristics:
  - Grouped vs. ungrouped
  - Large samples vs. small samples
  - Complete vs. censored
  - With or without replacement

Complete vs. censored data

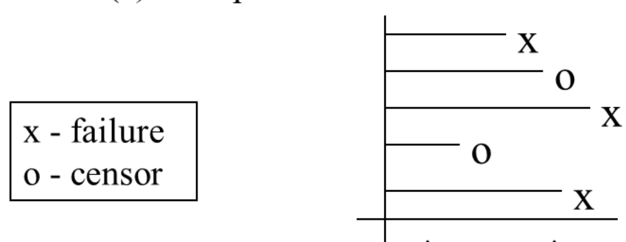
- Data are **complete** when t_i (the exact time/cycle of the specified failure mode) is available for all items $i = 1, \dots, n$
  - E.g., if a test is run until all items fail, and all failure times are recorded
- **Censored data** mean there is any item *i* with an unknown exact failure time, or failure via a different failure mode



(a) Complete Data



(b) Singly Censored



(c) Multiply Censored

Types of censored data

- Consider n items in a test
- Data are **right censored** if a failure time is not known, but is known to be greater than a given value
  - Common in reliability: only r ($r \leq n$) items fail in test duration t_{test}
  - The $n - r$ non-failed items thus have failure times $t \geq t_{test}$
- Data are **interval censored** if a failure time is not known, but is known to fall within an interval
  - This type of data may come from, e.g., periodic inspection. If an item fails in between inspections, only the interval will be known
- Data are **left censored** if a failure time is only known to be less than a given value

Type I and Type II censoring

- Data is **Type I** (time right-singly-censored) if the test is terminated at a **non-random** time t_{test}
  - Place n items on a **time-terminated test**, which will be terminated after a predetermined time t_{test} has elapsed.
  - Items that fail with $t \le t_{test}$ will have known, specific failure times. Items that do not fail are thus right censored.
- Data is **Type II** censored if the test is terminated at a **non-random** number of observed failures
  - Place n items on a **failure-terminated test**, which will be terminated after a predetermined number of failures r is observed.
  - Only the r smallest times to failure ($t_1 \le \ldots \le t_r$) out of n sample times to failure are known

Generic failure data sources

- Skim **Appendix B** in the textbook for failure data related to various mechanical and electrical components.
- See comprehensive list in **Chapter 5.2**, including:
  - NSWC
  - OREDA
  - Ignition
  - RIAC
  - IEEE Std. 500-1984
- Influencing factors are ofen used to adjust generic data
  - Environment, design and manufacturing, operational factors
- After class: spend a few minutes looking at one relevant to your industry.

Generic approach for identifying a distribution from data

- Multiple methods can help you identify candidate distributions:
  1. Construct a histogram of the data (e.g., failure or repair times)
  2. Compute descriptive statistics of the data
  3. Analyze empirical statistics of failure rate (non-parametric)
  4. Use prior knowledge of failure process, or properties of the theoretical distribution
  5. Construct a probability plot (& implement linear regression/least squares)

Selecting probability distributions for your failure data

- There are ways in which we can establish a distribution directly from data…or check fit against a known distribution form:

  - Non-parametric

(No distribution is assumed)

  - Parametric

(A distribution is assumed)

Probability Plotting

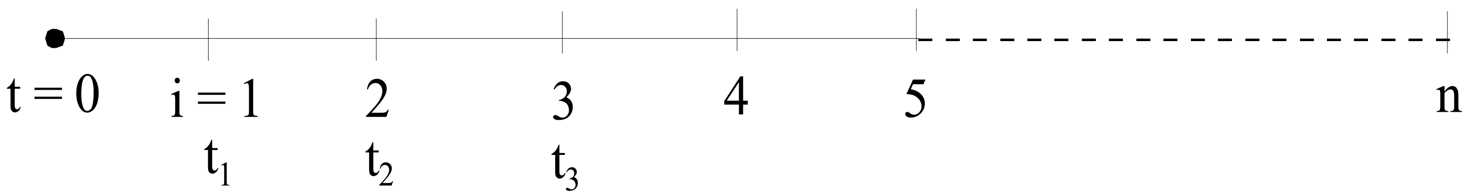
Goodness of fit tests

Nonparametric procedures for reliability functions

- We can estimate reliability parameters directly from data using **non-parametric (empirical, distribution-free)** approaches
- Nonparametric approaches attempt to directly estimate the reliability characteristics (e.g., $f(t), R(t), h(t)$) from a sample.
  - Useful for exploratory/preliminary analysis
  - Used in probability plotting
  - Key assumption: i.i.d. data
- We use various corrections/estimators depending on how much data we have & if the data are complete or censored
  - e.g., Blom, Kimball, Nelson-Aalen estimators, mean plotting position, Kaplan-Meier etc.

Non-parametric estimation

- Ordered data for *n* times to failure: $t_1 \leq t_2 \leq \; \dots \; \leq t_n$



- Recall that

$$h(t) = \frac{\text{probability of failure in } (t + \Delta t) \text{ given surival to t}}{\text{time interval}}$$

for an interval $\Delta t = t_{i+1} - t_i$

That is:

$$h(t) = \frac{\text{\# failures between t and t+}\Delta\text{t}}{\text{\# of units surviving past t}} * \frac{1}{\Delta t}$$

Non-parametric estimation (cont.)

- $\mathrm{h(t)} = \dfrac{\text{\# failures between t and t+}\Delta\text{t}}{\text{\# of units surviving past t}} * \dfrac{1}{\Delta t}$

$$\boxed{\hat{\mathrm{h}}(\mathrm{t}_i) = \frac{1}{(\mathrm{n}-\mathrm{i})(\mathrm{t}_{\mathrm{i}+1}-\mathrm{t}_{\mathrm{i}})}} \qquad i = 1,2,\ldots,n-1$$

For small samples (~n<25), to make an unbiased estimate of h(t), corrections are necessary

$$\hat{h}(t_i) = \frac{1}{(n-i+0.625)(t_{i+1}-t_i)}, \qquad i = 1,\ldots,n-1$$

Note: 0.625 and 0.25 come from statistical analyses by Kimball to minimizing bias for the Weibull distribution with small samples.

Non-parametric estimation (cont.)

Estimators for R(t) and f(t) are equally straightforward:

$$\mathrm{R(t_i)} = \frac{\mathrm{n - i}}{\mathrm{n}}$$

$$\mathrm{f(t) = h(t) * R(t)}$$

- And the unbiased estimators of R(t) and f(t) are the **Blom (Kimball) estimators or plotting positions**

$$\hat{R}(t_i) = \frac{n - i + 0.625}{n + 0.25}, \qquad i = 1, \dots, n$$

$$\hat{f}(t_i) = \frac{1}{(n + 0.25)(t_{i+1} - t_i)}, \qquad i = 1, \dots, n - 1$$

- These are designed to de-bias small ($n \lesssim 25$) samples and estimate $h(t), R(t)$, and $f(t)$.
- These are recommended for use with small samples but can be used for larger samples.

Non-parametric reliability estimation (cont.)

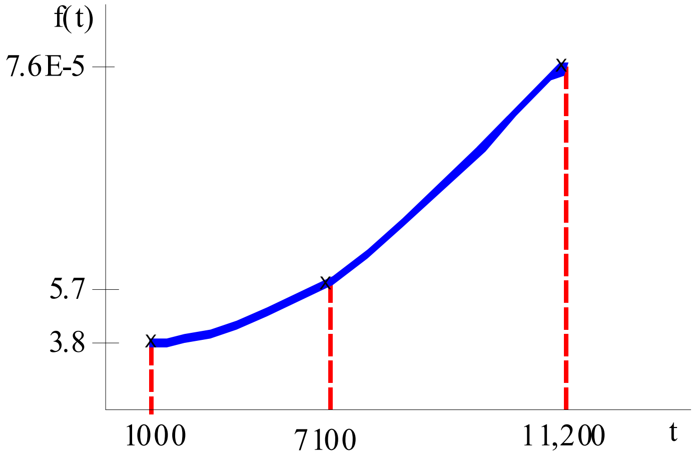
- **Example:** We observe 4 component failure times (in hrs): (1000, 7100, 11200, 14300). Use the Kimball plotting positions to obtain a non-parametric estimate of $\hat{f}(t)$, hazard rate $\hat{h}(t)$, and Reliability $\hat{R}(t)$.

i	$\mathbf{t_i}$ (hrs)	$\hat{\mathbf{f}}(\mathbf{t_i})$	$\hat{\boldsymbol{R}}(\boldsymbol{t_i})$	$\hat{\boldsymbol{h}}(\boldsymbol{t})$
1	1,000			
2	7,100			
3	11,200			
4	14,300			

Non-parametric estimation (cont.)

- **Example:** we observe 4 component failure times (hrs.): (1000, 7100, 11200, 14300)

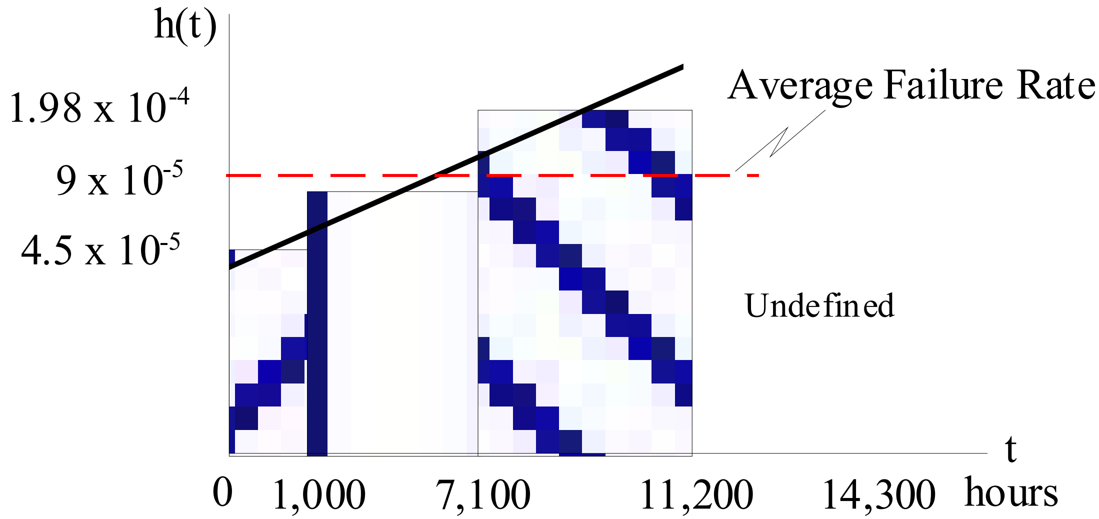
$\mathbf{i}$	$\mathbf{t_i}$ **(hrs)**	$\hat{\mathbf{f}}(\mathbf{t_i}) = \dfrac{\mathbf{1}}{(\mathbf{n}+\mathbf{0.25})(\mathbf{t_{i+1}}-\mathbf{t_i})}$
1	1,000	$\dfrac{1}{(4.25)(6100)} = 3.88 \times 10^{-5}\text{hr}^{-1}$
2	7,100	$\dfrac{1}{(4.25)(4100)} = 5.74 \times 10^{-5}\text{hr}^{-1}$
3	11,200	$\dfrac{1}{(4.25)(3100)} = 7.59 \times 10^{-5}\text{hr}^{-1}$
4	14,300	$\dfrac{1}{(4.25)(?)} = \text{undefined}$

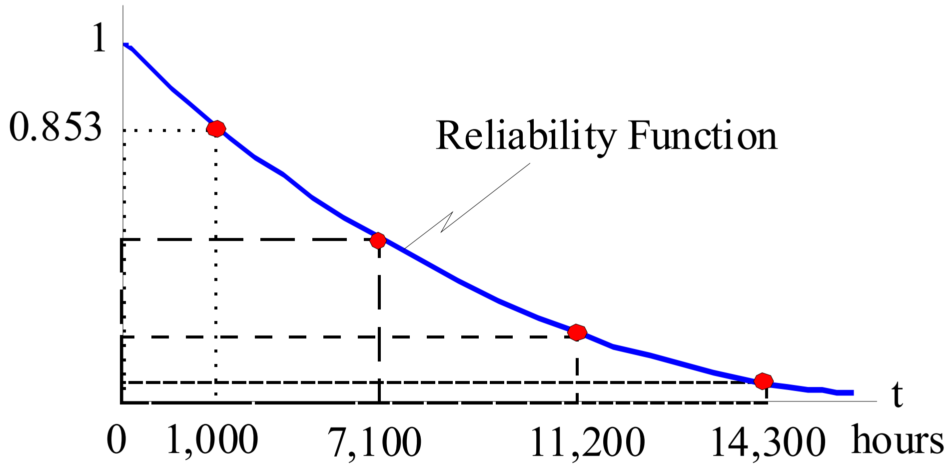


Non-parametric reliability estimation (cont.)

i	t_i $(hour)$	$\hat{R}(t_i) = \dfrac{(n-i+0.625)}{(n+0.25)}$	$\hat{h}(t) = \dfrac{1}{(n-i+0.625)(t_{i+1}-t_i)}$
1	1,000	$\dfrac{(4-1+0.625)}{4+0.25)} = 0.853$	$\dfrac{1}{(4-1+0.625)(7100-1000)} = 4.52 \times 10^{-5}$
2	7,100	$\dfrac{(4-2+0.625)}{(4+0.25)} = 0.617$	$\dfrac{1}{(4-2+0.625)(11200-7100)} = 9.29 \times 10^{-5}$
3	11,200	$\dfrac{(4-3+0.625)}{(4+0.25)} = 0.382$	$\dfrac{1}{(4-3+0.625)(14300-11200)} = 1.99 \times 10^{-4}$
4	14,300	$\dfrac{(4-4+0.625)}{(4+0.25)} = 0.147$	$\dfrac{1}{(4-4+0.625)(?-14300)} = undefined$

Non-parametric reliability estimation (cont.)





So the hazard rate is slightly increasing and the reliability is decreasing exponentially

Nonparametric procedure for large or grouped samples

- The **Nelson-Aalen Nonparametric Estimators** are designed for use with large or grouped samples
- Key assumption: Times to failure are grouped into equal increments Δt (required)

$$\hat{h}(t_i) = \frac{N_f(t_i)}{N_s(t_i)\Delta t} \qquad \hat{R}(t_i) = \frac{N_s(t_i)}{N} \qquad \hat{f}(t_i) = \frac{N_f(t_i)}{N\Delta t}$$

- $N_f(t_i) =$ # of failures observed in interval $(t_i, t_i + \Delta t)$
- $N_s(t_i) =$ # of surviving components in the interval starting at t_i
- t_i is usually the *lower endpoint* of interval Δt, but this can differ between practitioners, so be careful

Non-parametric reliability estimation example

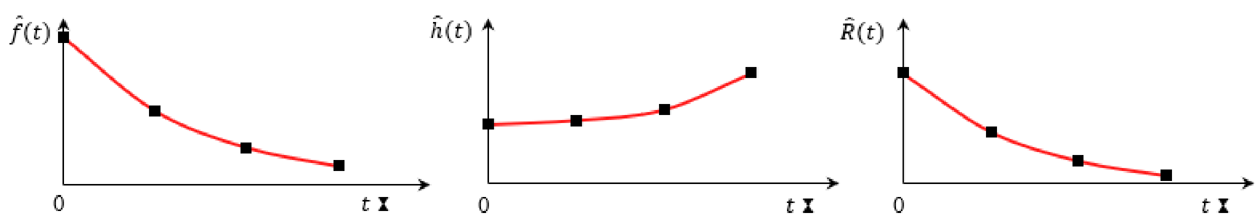
- **Example:** Given $n = 150$ observed failure times (hrs), estimate $\hat{f}(t), \hat{h}(t)$, and $\hat{R}(t)$. Use t_i as the lower endpoint of the interval in this example.

i	Interval	$N_f(t_i)$
1	$0 < t < 1000$	80
2	$1000 < t < 2000$	40
3	$2000 < t < 3000$	20
4	$3000 < t < 4000$	10

Non-parametric reliability estimation example

- **Solution:** Given $n = 150$ observed failure times, estimate $\hat{f}(t), \hat{h}(t)$, and $\hat{R}(t)$. Use t_i as the lower endpoint of the interval in this example.

i	Interval	$N_f(t_i)$	$N_S(t_i)$	$\hat{f}(t_i) = \frac{N_f(t_i)}{N\Delta t}$	$\hat{h}(t_i) = \frac{N_f(t_i)}{N_S(t_i)\Delta t}$	$\hat{R}(t_i) = \frac{N_S(t_i)}{N}$
1	$0 < t < 1000$	80	150	$\frac{80}{150 \times 1000} = 5.33 \times 10^{-4}$	$\frac{80}{150 \times 1000} = 5.33 \times 10^{-4}$	$\frac{150}{150} = 1.00$
2	$1000 < t < 2000$	40	70	$\frac{40}{150 \times 1000} = 2.67 \times 10^{-4}$	$\frac{40}{70 \times 1000} = 5.71 \times 10^{-4}$	$\frac{70}{150} = 0.47$
3	$2000 < t < 3000$	20	30	$\frac{20}{150 \times 1000} = 1.33 \times 10^{-4}$	$\frac{20}{30 \times 1000} = 6.67 \times 10^{-4}$	$\frac{30}{150} = 0.20$
4	$3000 < t < 4000$	10	10	$\frac{10}{150 \times 1000} = 6.67 \times 10^{-5}$	$\frac{10}{10 \times 1000} = 1.00 \times 10^{-3}$	$\frac{10}{150} = 0.07$



Parameter estimation

- We can also use data to estimate parameters of the underlying probability distribution

Parameter Estimation

- Probability Plotting(Linear regression w/ least squares estimates)
- Maximum Likelihood Estimation
- Bayesian Estimation

Probability plotting

- Observed data may be plotted on coordinates (previously: probability papers) such that the resulting life cdf falls on a straight line.
  - To visually assess fit
  - And then to use linear regression (least squares) to estimate parameters of the distribution.
- Probability papers for many types of distributions exist (several have been uploaded on ELMS; more commonly, we use software).

Probability plotting

- General procedure:
  - List your data in Excel or Matlab
    - *make sure that list is ordered; if it's not, sort it!
  - Find $\hat{R}(t_i)$ or $\hat{F}(t_i)$ for each time or interval (Use the non-parametric estimators)
  - Make the known expression for R(t) or F(t) linear by taking ln of both sides (as many times as needed)
  - Plot values
  - Find trendline (least squares fit) & match to parameters

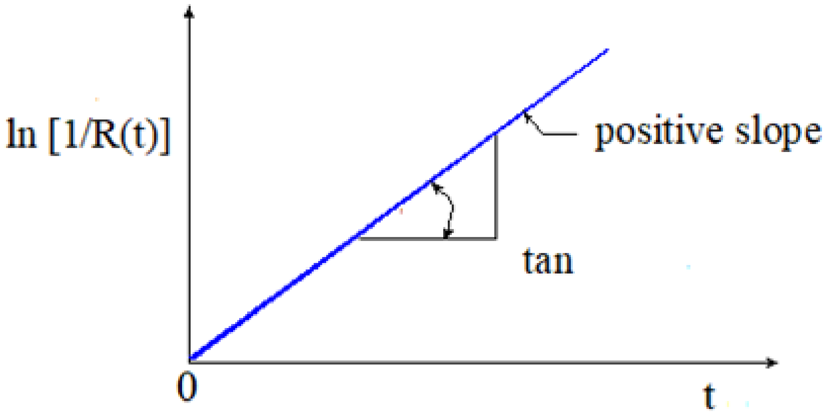
Exponential probability plotting

1. Order the *n* failure times $t_1 \leq t_2 \leq \ \dots \leq t_i \leq \ \dots \ \leq t_n$
2. Calculate the reliability estimator (using Kimball Plotting Position or Nelson-Aalen approach depending on sample size)
3. Recall: for exp. dist: $R(t) = e^{-\lambda t} \ or \ \frac{1}{R(t)} = e^{\lambda t}$

   By taking logarithms of both sides we get a linear equation: $ln\left[\frac{1}{R(t)}\right] = \lambda t$
4. Plot ln $\left(\frac{1}{R(t)}\right)$ vs. t. A straight line is a good fit and suggests that the exponential distribution is an adequate model.
5. Get the trendline:

   Set y-intercept = 0

   And the Slope = λ



Example

- The following 20 failure times (in days) were recorded for an electrical component: 51.1, 41.6, 12.9, 13.8, 22.8, 14.8, 18.5, 14.3, 27.1, 29.7, 32, 39.5, 41.3, 4.2, 3.3, 61.7, 92.2, 106.6, 148.8, 198.1 days.
- Use probability plotting to determine whether the data come from an Exponential distribution.
- Find the MTTF from this distribution.

Example – Solution

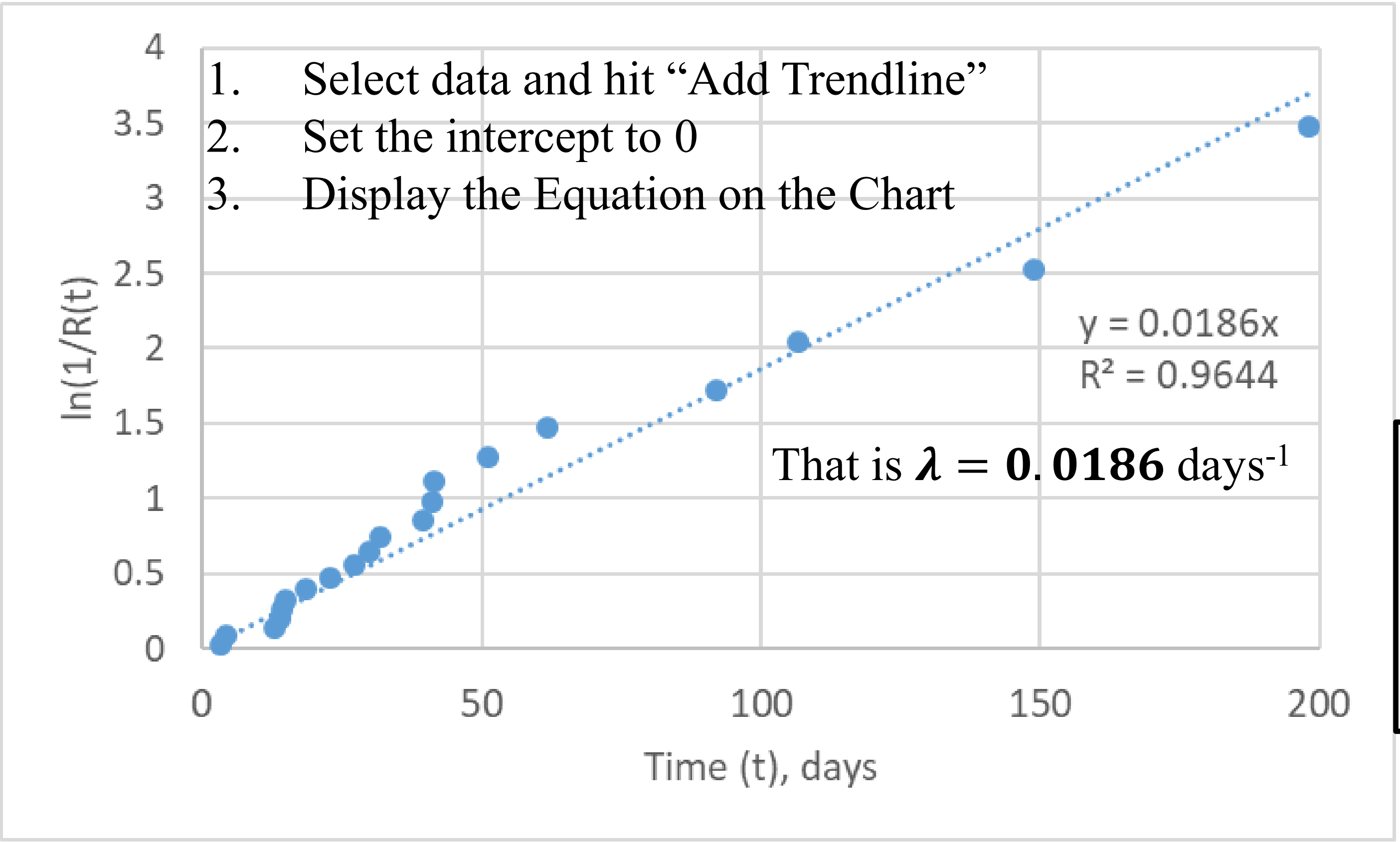
i	t_i	$R(t_i)$	$1/R(t_i)$	$\ln(1/R(t_i))$
1	3.3	0.969	1.032	0.031
2	4.2	0.920	1.087	0.084
3	12.9	0.870	1.149	0.139
4	13.8	0.821	1.218	0.197
5	14.3	0.772	1.296	0.259
6	14.8	0.722	1.385	0.325
7	18.5	0.673	1.486	0.396
8	22.8	0.623	1.604	0.472
9	27.1	0.574	1.742	0.555
10	29.7	0.525	1.906	0.645
11	32	0.475	2.104	0.744
12	39.5	0.426	2.348	0.853
13	41.3	0.377	2.656	0.977
14	41.6	0.327	3.057	1.117
15	51.1	0.278	3.600	1.281
16	61.7	0.228	4.378	1.477
17	92.2	0.179	5.586	1.720
18	106.6	0.130	7.714	2.043
19	148.8	0.080	12.462	2.523
20	198.1	0.031	32.400	3.478

With a relatively small, ungrouped sample – use the estimators to create $\hat{R}(t_i)$:

$$\hat{\mathrm{R}}(t_i) = \frac{\mathrm{n} - \mathrm{i} + 0.625}{\mathrm{n} + 0.25}$$

$$\frac{1}{\hat{\mathrm{R}}(t_i)} = \frac{\mathrm{n} + 0.25}{\mathrm{n} - \mathrm{i} + 0.625}$$

Example- Solution



MTTF= $\mathbf{53.76}$ days (compare to the mean of the data sample of 48.15)

Note: if you do this in RARE, it doesn't set intercept to 0. RARE gives $\boldsymbol{\lambda} = \mathbf{0.0179}$ days^{-1} but has a small intercept (0.101) which RARE neglects.

Exponential plotting (cont.)

- We may observe two or more straight lines. This happens when, e.g., there may be initially a particular failure mode and another failure mode become dominant.

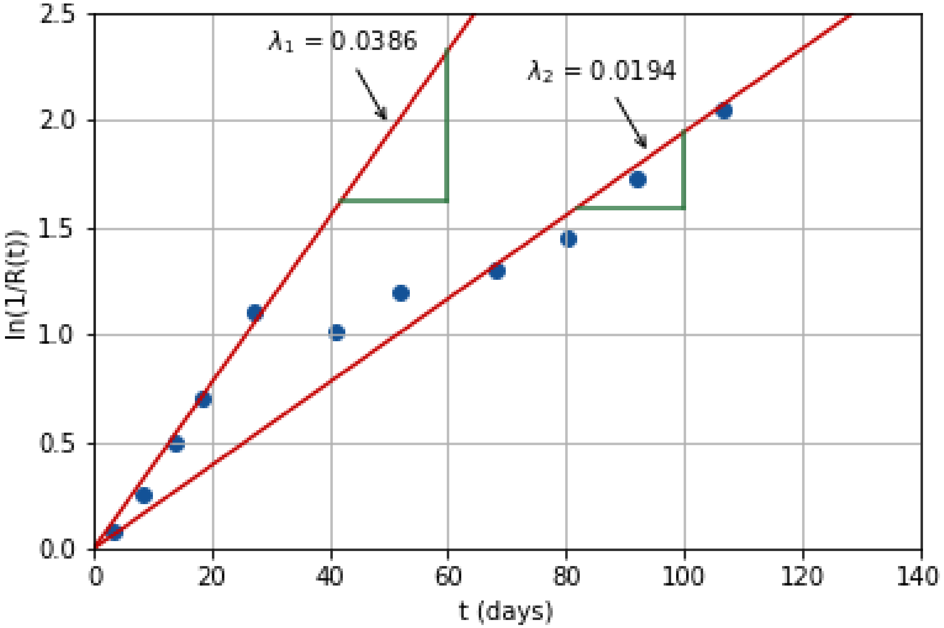
$$\lambda_E = \lambda_1 + \lambda_2$$

Therefore,

$$\mathrm{R} = \mathrm{R}_1 \cdot \mathrm{R}_2$$

$$= \mathrm{e}^{-\lambda_1 \mathrm{t}} \mathrm{e}^{-\lambda_2 \mathrm{t}}$$

$$= \mathrm{e}^{-\lambda_\mathrm{E} \mathrm{t}}$$



Weibull probability plotting (cont.)

- The goal is to calculate the shape parameter β and the scale parameter α.
- Weibull $R(t) = e^{-\left(\frac{t}{\alpha}\right)^{\beta}}$

$$\ln\left(\ln\left[\frac{1}{R(t)}\right]\right) = \beta \cdot \ln(t) - \beta \cdot \ln(\alpha)$$

****note double ln here**

- Plot $\text{Ln}\left(\text{Ln}\left[\frac{1}{R(t)}\right]\right)$ vs. $\ln(t)$
  - If data falls on a straight line, Weibull is a good fit
  - Add linear trendline.
  - Slope = β
  - Y-int = $-\beta \times \ln(\alpha)$
  - Solve for β, α

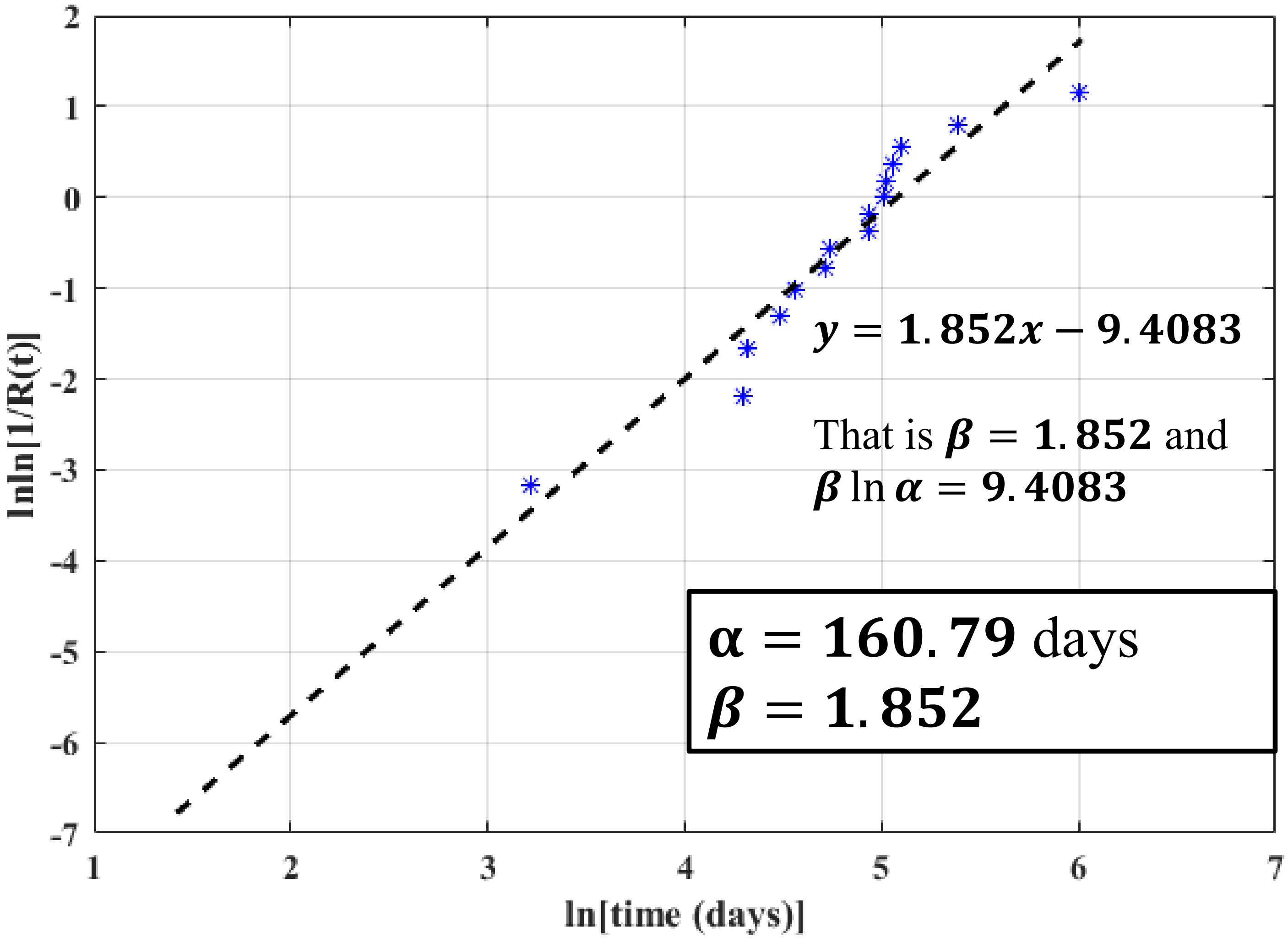
Example

- The following failure times were obtained from testing 15 units until each had failed: 25.1, 73.9, 75.5, 88.5, 95.5, 112.2, 113.6, 138.5, 139.8, 150.3, 151.9, 156.8, 164.5, 218, 403.1 days. Determine whether the data represent the Weibull distribution. If the data are a reasonable fit, find the shape and scale parameters.

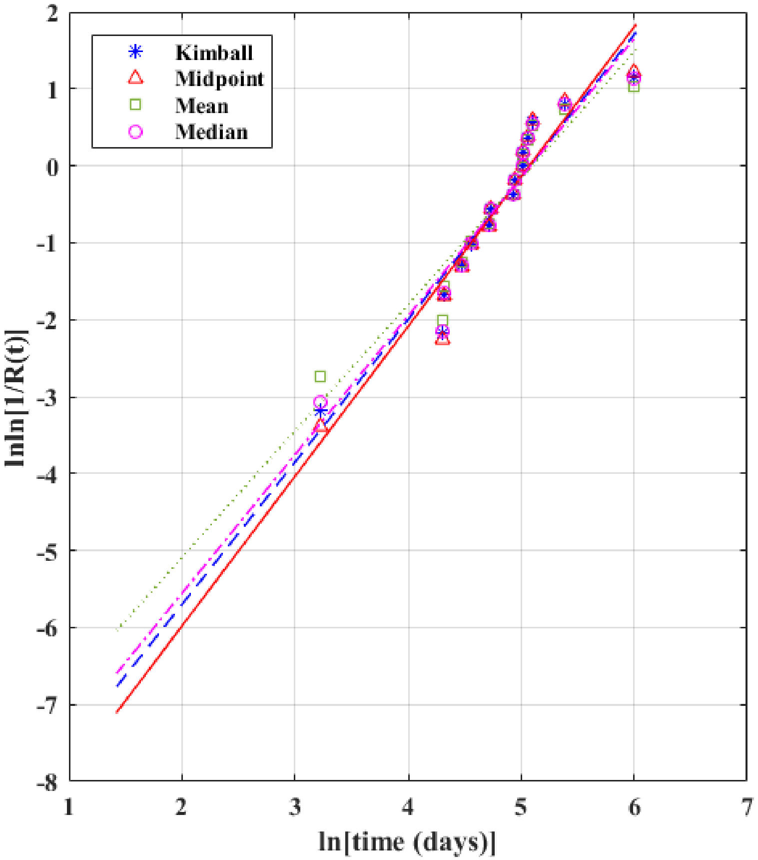
Example – Solution

i	ti	Ln[ti]	R(ti)	1/R(ti)	LnLn[1/R(ti)]
1	25.1	3.2229	0.9590	1.0427	-3.1737
2	73.9	4.3027	0.8934	1.1193	-2.1833
3	75.5	4.3241	0.8279	1.2079	-1.6665
4	88.5	4.4830	0.7623	1.3118	-1.3041
5	95.5	4.5591	0.6967	1.4353	-1.0179
6	112.2	4.7203	0.6311	1.5844	-0.7761
7	113.6	4.7327	0.5656	1.7681	-0.5623
8	138.5	4.9309	0.5000	2.0000	-0.3665
9	139.8	4.9402	0.4344	2.3019	-0.1818
10	150.3	5.0126	0.3689	2.7111	-0.0026
11	151.9	5.0232	0.3033	3.2973	0.1766
12	156.8	5.0550	0.2377	4.2069	0.3624
13	164.5	5.1029	0.1721	5.8095	0.5650
14	218	5.3845	0.1066	9.3846	0.8061
15	403.1	5.9992	0.0410	24.4000	1.1615

Example – Solution



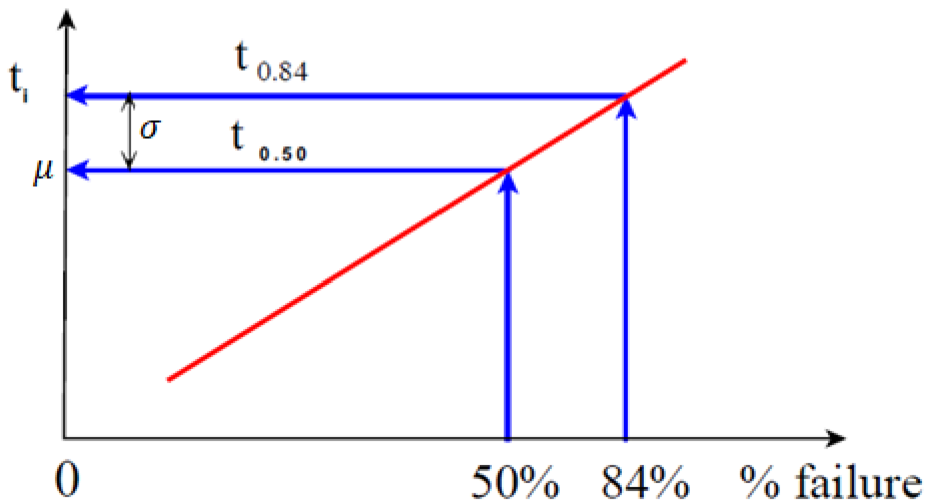
Alternative plotting positions exist – see textbook for details

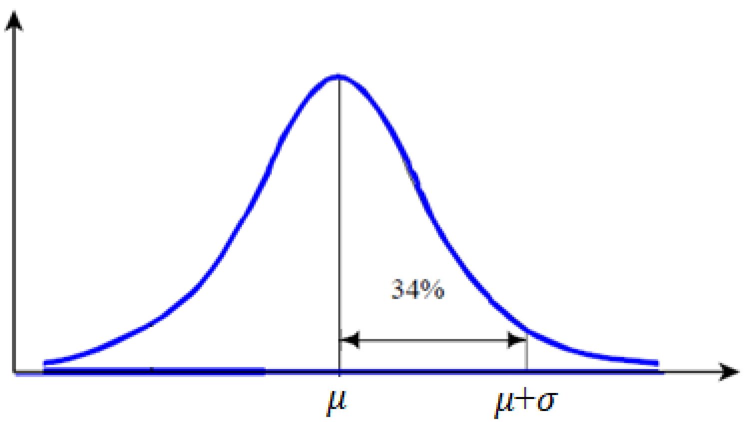


Midpoint:	$R(t) = \dfrac{n - i + 0.5}{n}$
Mean:	$R(t) = \dfrac{n - i + 1}{n + 1}$
Median:	$R(t) = \dfrac{n - i + 0.7}{n + 0.4}$

Normal distribution probability plotting

- We plot $\Phi^{-1}(F(t))$ against t for the normal distribution
  - Can also plot $\Phi^{-1}(F(t))$ against $\ln(t)$ for the lognormal distribution
- $F(t)$ is constructed using appropriate non-parametric method
  - E.g., $F(t) = \frac{i-0.375}{n+0.25}$ if using the Kimball estimators.
- $\Phi^{-1}(\cdot)$ is the inverse of the standard normal distribution





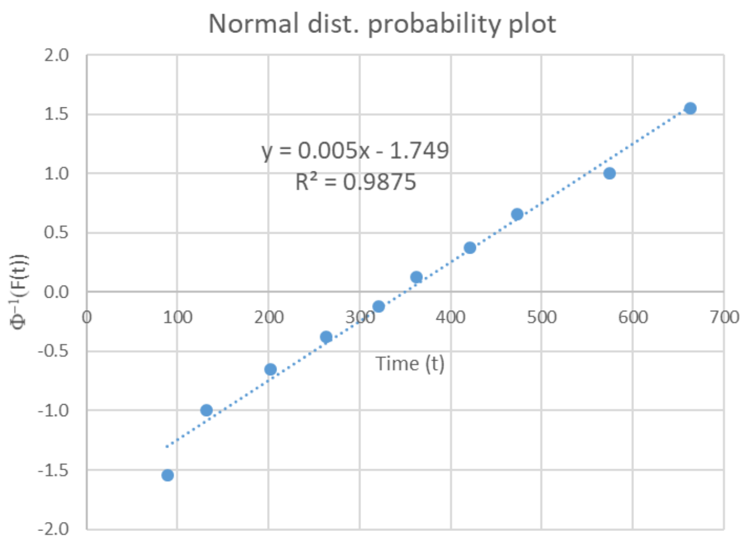
Normal distribution

Normal plot of the example

- **Example:** Normal dist. $F = \Phi(\frac{t-\mu}{\sigma})$, Linearizes as: $\Phi^{-1}(F) = \frac{t}{\sigma} - \frac{\mu}{\sigma}$

=Norm.S.inv(F)

i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
t_i (days)	89	132	202	263	321	362	421	473	575	663
$F(t)=\frac{i-0.375}{n+0.25}$	0.061	0.159	0.256	0.354	0.451	0.549	0.646	0.744	0.841	0.939
$\Phi^{-1}\left(\frac{i-0.375}{n+0.25}\right)$	-1.55	-1.00	-0.66	-0.38	-0.12	0.12	0.38	0.66	1.00	1.55

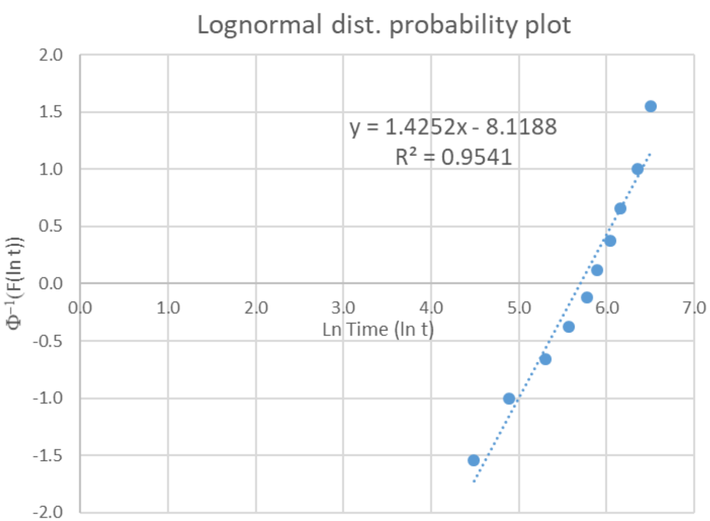


$\boldsymbol{t_{0.50} = \mu}$	$\boldsymbol{\mu = 349.8}$
$\boldsymbol{t_{0.84} = \mu + \sigma}$	$\boldsymbol{\sigma = 200}$

Lognormal plot of the example

- **Example: Lognormal distribution** $F = \Phi(\frac{\ln t - \mu}{\sigma})$, linearize as $\Phi^{-1}(F) = \frac{\ln t}{\sigma} - \frac{\mu}{\sigma}$

i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$F(t)=\frac{i-0.375}{n+0.25}$	0.061	0.159	0.256	0.354	0.451	0.549	0.646	0.744	0.841	0.939
$\Phi^{-1}\left(\frac{i-0.375}{n+0.25}\right)$	-1.55	-1.00	-0.66	-0.38	-0.12	0.12	0.37	0.66	1.00	1.55
t_i (days)	89	132	202	263	321	362	421	473	575	663
lnt_i	4.5	4.9	5.3	5.6	5.8	5.9	6.0	6.2	6.4	6.5



Lognormal parameters	
$\ln(t_{0.5})$	5.70
$\ln(t_{0.84})$	6.39
μ	5.70
σ	0.70

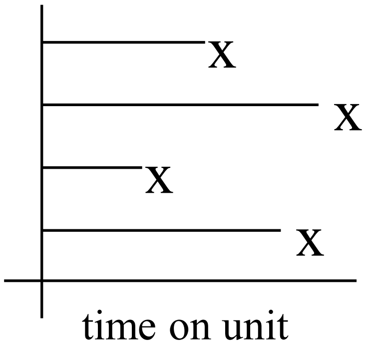
Conclusion: In this case, the Normal distribution fits better than the Lognormal distribution (i.e., higher R^2 value for the normal plot.)

Parameter estimation

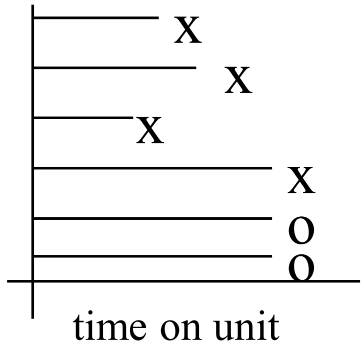
- Recall: In Module 4, you learned how to dervice the MLE estimators for many distributions.
- Now we'll present **point estimates** & **interval estimates** for the parameters of common reliability distributions

Reminder! Reliability data are often censored

- What happens now?

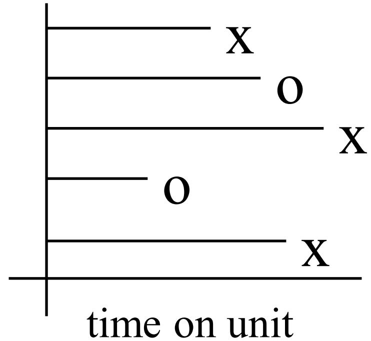


(a) Complete Data



(b) Singly Censored

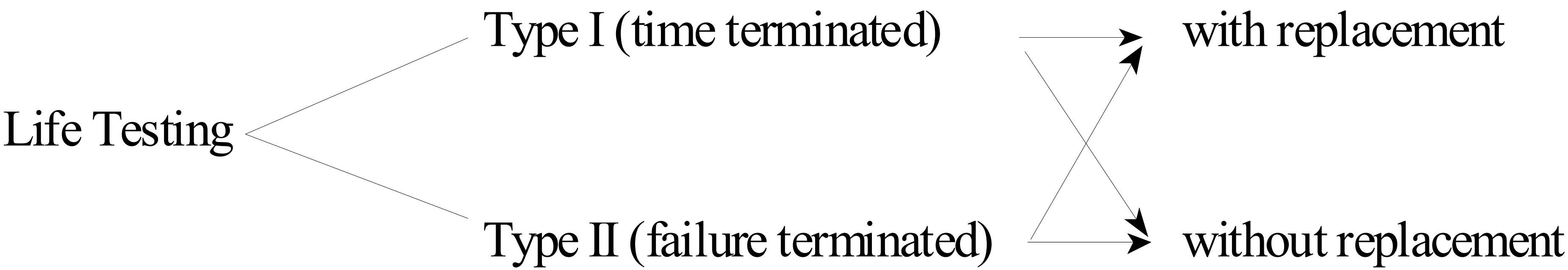
x - failure
o - censor



(c) Multiply Censored

Types of Life Testing Data

- Life testing is done to get failure data for reliability estimation methods:



- **The result is that reliability data are almost always censored.**

Recall: MLE

- **Likelihood Function for known failure times:**

$$L(\theta|x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} f(x_i|\theta)$$

$$\Lambda(\theta|E) = \ln\{L(\theta|E)\} = \sum_{i=1}^{n} \ln[f(x_i\,|\theta)]$$

- **Maximum Likelihood (ML)** estimate of θ is the value of $\hat{\theta}$ such that

$$L(\hat{\theta}|x_1, x_2, \dots, x_n) \geq L(\theta|x_1, x_2, \dots, x_n)$$

for every value of θ. Statistic $\hat{\theta}$ is a r.v. called the ML estimator (MLE) of θ.

Solve for θ

$$\left.\frac{\partial L}{\partial \theta}\right|_{\theta-\hat{\theta}} = \left.\frac{\partial lnL}{\partial \theta}\right|_{\theta-\hat{\theta}} = 0$$

Higher L →better fit.

Likelihood functions for different types of reliability data

Type of Observation	Likelihood Function	Example Description
Exact lifetimes	$L_i(\theta\|t_i) = f(t_i\|\theta)$	Failure time is known.
Left censored	$L_i(\theta\|t_i) = F(t_i\|\theta)$	Component failed before time t_i.
Right censored	$L_i(\theta\|t_i) = 1 - F(t_i\|\theta) = R(t_i\|\theta)$	Component survived to time t_i.
Interval censored	$L_i(\theta\|t_i) = F\left(t_i^{RI}\|\theta\right) - F\left(t_i^{LI}\|\theta\right)$	Component failed between t_i^{LI} and t_i^{RI}.
Left truncated	$L_i(\theta\|t_i) = \dfrac{f(t_i\|\theta)}{R(t_L\|\theta)}$	Component failed at time t_i where observations are truncated before t_L.
Right truncated	$L_i(\theta\|t_i) = \dfrac{f(t_i\|\theta)}{F(t_U\|\theta)}$	Component failed at time t_i where observations are truncated after t_U.
Interval truncated	$L_i(\theta\|t_i) = \dfrac{f(t_i\|\theta)}{F(t_U\|\theta) - F(t_L\|\theta)}$	Component failed at time t_i where observations are truncated before t_L and after t_U.

Example: Likelihood function creation

- Assume we have a sample of size *D*
  - Some are known, exact failure times ($\delta_i = 1$)
  - Some are right-censored times ($\delta_i = 0$)
- The likelihood function is constructed:

$$L(\theta|D) = c\prod_i \left\{ [f(t_i|\theta)]^{\delta_i} \times [1 - F(t_i|\theta)]^{1-\delta_i} \right\}$$

- Where:
  - c = combinatorial constant
  - $f(t_i|\theta)$ = likelihood function for exact data points
  - $1 - F(t_i|\theta)$ = likelihood function for right-censored data points

MLE parameters of various distributions

- For *complete data:* point estimates of the MLE estimates for the parameters of many relevant distributions are known.
  - Now we'll discuss confidence intervals on those parameters, too.
- *For censored data:* you must maximize the likelihood function for this data; then use the Fisher information matrix (or established, derived functional relationships) to come up with values needed to estimate confidence intervals.

MLE parameters of exponential dist for complete data

- **Exponential Distribution: $\boldsymbol{n}$ failures at times $\boldsymbol{t_i}$**

$$\mathrm{L} = \prod_{i=1}^{n} \lambda e^{-\lambda t_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} t_i}$$

Solve for n failures

$$\Lambda = \ln(L) = n \ln \lambda - \lambda \sum_{i=1}^{n} t_i$$

$$\left.\frac{\partial \Lambda}{\partial \lambda}\right|_{\lambda=\hat{\lambda}} = \frac{n}{\hat{\lambda}} - \sum_{i=1}^{n} t_i = 0 \rightarrow$$

$$\boldsymbol{\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i}}$$

Exponential distribution MLEs for censored data

- **Type I with replacement:**
  - n components are placed under test.
  - t_{end} time at which the test is terminated.
  - TTT accumulated component test hours (**total time on test**)
  - r failures have been observed (up to t_o)

$$TTT = nt_{end}$$

$$\hat{\lambda} = \frac{r}{TTT}$$

$$\widehat{MTTF} = \frac{TTT}{r}$$

And the number of units actually used in the test (n') is: $\mathrm{n}' = \mathrm{n} + \mathrm{r}$

Exponential distribution MLEs for censored data

- **Type I without replacement:**

$$TTT = \sum_{i=1}^{r} t_i + (n-r)t_{end}$$

Accumulated time on test of r failed components.

Accumulated time on test of the non-failing components.

$$\hat{\lambda} = \frac{r}{TTT}$$

Exponential distribution MLEs for censored data

- **Type II with replacement:**
  - n components placed on test.
  - t_r the time after which test is terminated when the r^{th} failure has occurred. So r^{th} failure time is specified by t_r and is a random variable.

$$TTT = nt_r$$

$$\hat{\lambda} = \frac{r}{TTT}$$

  - Total units put on test (n’) is:

$$n' = n + r - 1$$

Exponential distribution MLEs for censored data

- **Type II without replacement:**

$$\mathrm{TTT} = \sum_{i=1}^{r} t_i + (n - r)t_r$$

and

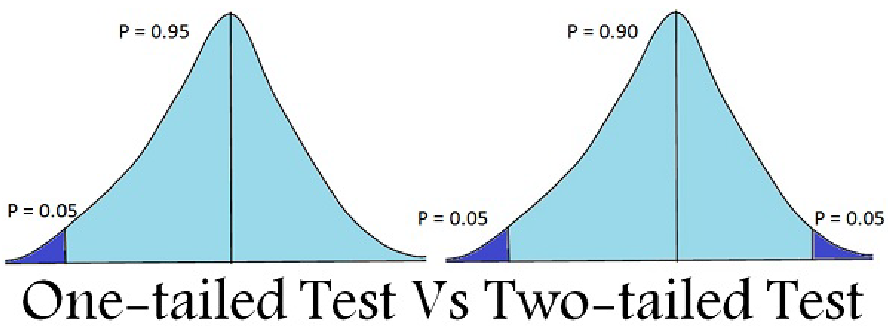
$$n' = n$$

Confidence intervals express uncertainty due to sample size

- **Example**: If 100 units are tested, consider two situations for exponential parameter estimation:
  - **Case 1: For r** $= \mathbf{1}\ \boldsymbol{failure}, \boldsymbol{t_0} = \mathbf{10}.$ hrs
    - $T = 10 \times 100 = 1000$
    - $\hat{\lambda} = \frac{r}{T} = \frac{1}{1000} = 10^{-3}\text{hr}^{-1}$
  - **Case 2: For r** $= \mathbf{10}$ **failures**, $\boldsymbol{t_0} = \mathbf{100}\ \boldsymbol{hrs}$
  - $T = 100 \times 100 = 10{,}000$
  - $\hat{\lambda} = \frac{\mathrm{r}}{\mathrm{T}} = \frac{10}{10{,}000} = 10^{-3}\text{hr}^{-1}$
  - Both gives you the same $\boldsymbol{\hat{\lambda}}$ estimate, but one has more data. The MLE parameter is the same, but the confidence interval is different for these two datasets.

Reminders: confidence interval

- The $1-\alpha$ confidence interval for a parameter θ is the interval such that:
  - $\Pr\left(\hat{\theta}_{lower} \leq \theta \leq \hat{\theta}_{upper}\right) = 1-\alpha$
  - e.g, for a 90% confidence interval, $\alpha = 0.1$. $\Pr\{R_L \leq R(t_0) \leq R_U\} = 1 - \alpha$, and thus in $100(1 - \alpha)\%$ of repetitions of that test, the population parameter falls between R_L and R_U.
  - Used to quantify uncertainty due to sampling error (i.e., limited number of samples),
    - Not uncertainty due to incorrect model selection or assumptions!



Exponential dist.: Confidence intervals

	Type I (Time Terminated Test for complete data)			
	One-Sided Confidence Limits		**Two-Sided Confidence Limits**	
Parameter	**Lower Limit**	**Upper Limit**	**Lower Limit**	**Upper Limit**
λ	0	$\frac{\chi^2_{(1-\gamma)}[2r+2]}{2TTT}$	$\frac{\chi^2_{\left(\frac{\gamma}{2}\right)}[2r]}{2TTT}$	$\frac{\chi^2_{\left(1-\frac{\gamma}{2}\right)}[2r+2]}{2TTT}$
MTTF	$\frac{2TTT}{\chi^2_{(1-\gamma)}[2r+2]}$	∞	$\frac{2TTT}{\chi^2_{\left(1-\frac{\gamma}{2}\right)}[2r+2]}$	$\frac{2TTT}{\chi^2_{\left(\frac{\gamma}{2}\right)}[2r]}$
$R(t)$	$e^{-\left[\frac{\chi^2_{(1-\gamma)}[2r+2]}{2Ttt}t_E\right.}$	1	$e^{-\left[\frac{\chi^2_{\left(1-\frac{\gamma}{2}\right)}[2r+2]}{2TTT}t_e\right.}$	$e^{-\left[\frac{\chi^2_{\left(\frac{\gamma}{2}\right)}[2r]}{2TTT}t_{\text{end}}\right]}$
	Type II (Failure Terminated Test for complete data)			
	One-Sided Confidence Limits		**Two-Sided Confidence Limits**	
Parameter	**Lower Limit**	**Upper Limit**	**Lower Limit**	**Upper Limit**
λ	0	$\frac{\chi^2_{(1-\gamma)}[2r]}{2TTT}$	$\frac{\chi^2_{\left(\frac{\gamma}{2}\right)}[2r]}{2TTT}$	$\frac{\chi^2_{\left(1-\frac{\gamma}{2}\right)}[2r]}{2TTT}$
MTTF	$\frac{2TTT}{\chi^2_{(1-\gamma)}[2r]}$	∞	$\frac{2TTT}{\chi^2_{\left(1-\frac{\gamma}{2}\right)}[2r]}$	$\frac{2TTT}{\chi^2_{\left(\frac{\gamma}{2}\right)}[2r]}$
$R(t)$	$e^{-\left[\frac{\chi^2_{(1-\gamma)}[2r]}{2TTT}t_{\text{end}}\right.}$	1	$e^{-\left[\frac{\chi^2_{\left(1-\frac{\alpha}{2}\right)}[2r]}{2TTT}t_{\text{end}}\right.}$	$e^{-\left[\frac{\chi^2_{\left(\frac{\gamma}{2}\right)}[2r]}{2TTT}t_{\text{end}}\right]}$

- Where $\chi^2_\gamma[\text{x}]$ is a chi-square distribution value, which has two parameters
  - Degree of freedom (x)
  - Some confidence level (γ)

- Note: uncensored data can be treated as a special case of a Type II (failure terminated) test.

Exponential confidence intervals

- These are type 1 data, so the confidence interval is calculated as:

$$\frac{\chi^2_{\left(\frac{\gamma}{2}\right)}(2\mathrm{r})}{2\mathrm{TTT}} \leq \hat{\lambda} \leq \frac{\chi^2_{\left(1-\frac{\gamma}{2}\right)}(2\mathrm{r}+2)}{2\mathrm{TTT}}, \qquad \text{For a 90\% confidence interval, } 1-\alpha = 0.9, \quad \alpha = 0.1$$

- **Case 1: For r = 1, TTT = 1000, the 90% confidence interval is:**
  - $\frac{\chi^2_{\left(\frac{\gamma}{2}\right)}(2)}{2(1000)} \leq \hat{\lambda} \leq \frac{\chi^2_{\left(1-\frac{\gamma}{2}\right)}(4)}{2(1000)} \Rightarrow \frac{0.1026}{2000} \leq \hat{\lambda} \leq \frac{9.49}{2000}$
  - $\mathbf{5.13 \times 10^{-5} \leq \hat{\lambda} \leq 4.75 \times 10^{-3}}$

- **Case 2: For r = 10, TTT = 10000, the 90% confidence interval is:**

- $\frac{\chi^2_{\left(\frac{\gamma}{2}\right)}(20)}{2(10{,}000)} \leq \hat{\lambda} \leq \frac{\chi^2_{\left(1-\frac{\gamma}{2}\right)}(22)}{2(10{,}000)} \Rightarrow \frac{10.85}{20{,}000} \leq \hat{\lambda} \leq \frac{33.92}{20{,}000}$
  - $\mathbf{5.43 \times 10^{-4} \leq \hat{\lambda} \leq 1.70 \times 10^{-3}}$

Example: Exponential confidence intervals

- 25 units are placed on test for 500 hours. Eight failures occur at times 75, 115, 192, 258, 312, 389, 410, 496 hours. Failed units are replaced.

- **Find:**
  A. The MLE of λ
  B. The two-sided 90% confidence limits on λ.

Example: Exponential confidence intervals

- **Solution:**
  A. This is time-terminated test, i.e., (**Type I**), with n=25 and $t_{end} = 500\ hrs,$ therefore,
- TTT = 25 * 500 hrs = 12,500 component-hrs.

$$\hat{\lambda} = \frac{number\ of\ failures}{total\ component\ time} = \frac{8}{12{,}500} = 6.4\ \times 10^{-4} hr^{-1}$$

Example: Exponential two-sided confidence interval

- **Solution B:**
- For a type I test, exponential distribution the two-sided confidence interval expression is:

- $$\frac{\chi^2_{\gamma/2}(2\mathrm{r})}{2\mathrm{TTT}} \le \hat{\lambda} \le \frac{\chi^2_{1-\gamma/2}(2\mathrm{r}+2)}{2\mathrm{TTT}}$$

Here, $\gamma = 0.1, \frac{\gamma}{2} = 0.05, TTT = 12{,}500, r = 8$

$$\frac{\chi^2_{0.05}(16)}{2 * 12500} \le \hat{\lambda} \le \frac{\chi^2_{0.95}(18)}{2 * 12500}$$

$$\frac{7.96}{25{,}000} \le \hat{\lambda} \le \frac{28.87}{25{,}000}$$

$$\boxed{3.18 \times 10^{-4} \le \hat{\lambda} \le 1.15 \times 10^{-3} hr^{-1}}$$

Exponential distribution: Right censored data summary

Type I data is time terminated and type II data is failure terminated. r is number of failures and n is the number of units being observed , t_i is the time to failure of a failed unit. t_r is the time after which test is terminated (for type I data this is the specified test time; for type II data, this is the time, when the r^{th} failure occurs)

Case	MLE	Total time, TTT	Confidence interval
Type II w/ replacement	$\hat{\lambda} = r/TTT$	$TTT = nt_r$ Where	$\frac{\chi^2_{\frac{\gamma}{2}}(2r)}{2TTT} \leq \lambda \leq \frac{\chi^2_{1-\frac{\gamma}{2}}(2r)}{2TTT}$ (Exact)
Type II w/o replacement	$\hat{\lambda} = r/TTT$	$TTT = \sum_{i=1}^{r} t_i + (n-r)t_r$	
Type I, w/ replacement	$\hat{\lambda} = r/TTT$	$TTT = n\, t_r$	$\frac{\chi^2_{\frac{\gamma}{2}}(2r)}{2TTT} \leq \lambda \leq \frac{\chi^2_{1-\frac{\gamma}{2}}(2r+2)}{2TTT}$ (Approximate)
Type I, w/o replacement	$\hat{\lambda} = r/TTT$	$TTT = \sum_{i=1}^{r} t_i + (n-r)t_r$	

Note: uncensored data can be treated as a special case of a Type II (failure terminated) test.

Example: Exponential two-sided confidence interval

- **Example:**
  - A plant had 50 instrument failures in a year among a total of 5613 such instruments.

A. Find 95% confidence limits on λ

B. Find 95% confidence limits on R (8760 hrs.),

C. Find the point estimate and 95% percentile estimate of the time at which R = 0.8 for each instrument.

Example: Exponential two-sided confidence interval

- **Solution A) Type I test**
  - $TTT = 5613\, units \times 8760\, hours = 4.9 \times 10^{7}$ component-hours, therefore,

$$\hat{\lambda} = \frac{50\, failures}{4.9 \times 10^{7}} = 1.0 \times 10^{-6} hr^{-1}$$

$\gamma = 1 - 0.95 = 0.05$, therefore,

$$\frac{\chi^2_{\frac{\gamma}{2}}(2 \times 50)}{2\mathrm{TTT}} \leq \lambda \leq \frac{\chi^2_{1-\frac{\gamma}{2}}(2 \times 50 + 2)}{2\mathrm{TTT}}$$

$$\frac{\chi^2_{0.025}(100)}{2\mathrm{TTT}} \leq \lambda \leq \frac{\chi^2_{0.975}(102)}{2\mathrm{TTT}}$$

$$\frac{74.55}{2TTT} \leq \hat{\lambda} \leq \frac{131.54}{2TTT}$$

With T = 4.9E7 hrs.

$$\boldsymbol{7.6E - 7 \leq \lambda \leq 1.3E - 6hr^{-1}}$$

Example: Exponential two-sided confidence interval

Solution B)

$$\hat{R}(8760) = e^{-\hat{\lambda}t} = e^{-1.0E-6(8760)} = 0.991$$

$$\exp\left(-\frac{\chi^2_{0.975}(102)}{2TTT} * 8760\right) \leq R \leq \exp\left(-\frac{\chi^2_{0.025}(100)}{2TTT} 8760\right)$$

$$\exp\left(-\frac{131.54 * 8760}{2(365)(24)(5613)}\right) \leq \hat{\mathrm{R}} \leq \exp\left(-\frac{74.53 * 8760}{2(365)(24)(5613)}\right)$$

$$\mathbf{0.9884 \leq \hat{R} \leq 0.9934}$$

Example: Exponential two-sided confidence interval

Solution C)

- To find reliable life at R = 0.8

$$R(t) = e^{-\lambda t} \rightarrow lnR = -\lambda t \rightarrow t = \frac{lnR}{-\lambda}$$

$$\hat{t}_{0.8} = \frac{-\ln(0.8)}{\hat{\lambda}} = 233{,}144\ hours\ \cong 25.5\ years$$

$$\frac{-\ln(0.8)}{1.3E-6} \leq t_{0.8} \leq \frac{-\ln(0.8)}{7.5E-7}$$

$$171{,}649 \leq t_{0.8} \leq 297{,}525\ hours$$
$$19.59 \leq t_{0.8} \leq 33.96\ years$$

MLE Parameters of various distributions

- For some distributions and data types (e.g., exponential with complete or right censored data), we have specific known relationship forms for certain data types for certain distributions.
  - E.g., exponential with complete or right censored data
  - E.g., others we will cover shortly.
- Other data types and parameters require deriving confidence intervals from the Fisher Information Matrix;

MLE: parameter uncertainty estimation

- Where parameter values have been estimated using the MLE process, the uncertainty of each parameter is quantified using the observed **Fisher Information Matrix ($\boldsymbol{J}(\boldsymbol{\theta})$)** as follows:
- For example, in the case of a distribution with *n* parameters, I is given by:
- $I(\widehat{\boldsymbol{\theta}}) = \begin{bmatrix} -\frac{\partial^2 \Lambda(\boldsymbol{\theta}|\,D\,)}{\partial\theta_1^2} & -\frac{\partial^2 \Lambda(\boldsymbol{\theta}|D)}{\partial\theta_1\partial\theta_2} & \cdots & -\frac{\partial^2 \Lambda(\boldsymbol{\theta}|\,D\,)}{\partial\theta_1\partial\theta_p} \\ -\frac{\partial^2 \Lambda(\boldsymbol{\theta}|\,D\,)}{\partial\theta_2\partial\theta_1} & -\frac{\partial^2 \Lambda(\boldsymbol{\theta}|\,D\,)}{\partial\theta_2^2} & \cdots & -\frac{\partial^2 \Lambda(\boldsymbol{\theta}|\,D\,)}{\partial\theta_2\partial\theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial^2 \Lambda(\boldsymbol{\theta}|\,D\,)}{\partial\theta_p\partial\theta_1} & -\frac{\partial^2 \Lambda(\boldsymbol{\theta}|\,D\,)}{\partial\theta_p\partial\theta_2} & \cdots & -\frac{\partial^2 \Lambda(\boldsymbol{\theta}|\,D\,)}{\partial\theta_p^2} \end{bmatrix}_{\theta_i=\widehat{\theta_i}}$
- $where\ \Lambda = \ln(\boldsymbol{\theta}|x_i)$

MLE: Parameter uncertainty estimation

- The inverse of the Fisher Information Matrix gives the covariance matrix which has the estimated variance of each parameter as follows:

$$Var(\hat{\theta}) = [I(\hat{\theta})]^{-1} = \begin{bmatrix} var(\theta_1) & cov(\theta_1, \theta_2) & \cdots & cov(\theta_1, \theta_n) \\ cov(\theta_2, \theta_1) & var(\theta_2) & \cdots & cov(\theta_2, \theta_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(\theta_n, \theta_1) & cov(\theta_n, \theta_2) & \cdots & var(\theta_n) \end{bmatrix}$$

- Using these values, the desired confidence intervals of each parameters can be found.

t-Distribution table

- Student's *t*-distribution (the *t*-distribution): a distribution that arises when estimating the mean of a normally distributed variable when sample size is small and (population) standard deviation is unknown.
- Note: t distribution is symmetrical, e.g., $t_{1-\frac{\gamma}{2}}(df)$=- $t_{\frac{\gamma}{2}}(df)$

- Lookup table in Appendix A.
- Excel: Use t.inv ($\frac{\gamma}{2}, df$) or t.inv.2t(γ, df)

MLE parameters of various distributions

- **Normal distribution: Complete data**
  - We know that for a sample of size n

$$\hat{\mu} = \frac{\sum_{i=1}^{n} t_i}{n} \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(t_i - \hat{\mu})^2}{n-1}$$

confidence interval for mean (when $\sigma$is unknown as estimated as s)

$$\hat{\mu} - \frac{s}{\sqrt{n}} * t_{\frac{\gamma}{2}}(n-1) \leq \mu \leq \hat{\mu} + \frac{s}{\sqrt{n}} * t_{\gamma/2}(n-1)$$

Where:

$(1-\gamma)$ = confidence level,

$t \rightarrow$ one-tailed t-distribution. (Or use t_γ with two-tailed distribution.)

$df = n - 1 =$ degrees of freedom

MLE parameters of various distributions

- **Normal distribution: right censored data**
- For a sample of size n where m components fail (n > m)

$$\hat{\mu} = \frac{\sum_{i=1}^{m} t_i}{n} \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{m} (t_i - \hat{\mu})^2}{n-1}$$

  - The confidence interval for mean (when σ is unknown and estimated as s)

$$\hat{\mu} - \frac{s}{\sqrt{n}} * t_{\frac{\gamma}{2}}(m-1) \le \mu \le \hat{\mu} + \frac{s}{\sqrt{n}} * t_{\frac{\gamma}{2}}(m-1)$$

where

$(1-\gamma)$ = confidence level,

$t \rightarrow$ one-tailed t-distribution (Or use t_γ with two-tailed t distribution)

$with\ m-1 = degrees\ of\ freedom$

MLE parameters of various distributions

- **Normal distribution (cont.): complete data and right censored**
  - Confidence limits for variance with complete data is:

$$\frac{(n-1)s^2}{\chi^2_{1-\frac{\gamma}{2}}[n-1]} \leq \sigma^2 \leq \frac{(n-1)\hat{s}^2}{\chi^2_{\frac{\gamma}{2}}[n-1]}$$

  - With right censored data, it is:

$$\frac{(n-1)s^2}{\chi^2_{1-\frac{\gamma}{2}}[m-1]} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{\frac{\gamma}{2}}[m-1]}$$

when m failures occur in n observations

MLE parameters of various distributions

- **Lognormal distribution: Complete data**

Note: numerical methods are required for dealing with *incomplete data.* So only complete data are presented for lognormal dist. See textbook.

$$\hat{\mu} = \frac{\sum_{i=1}^{n} lnt_i}{n}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(lnt_i - \hat{\mu}_t)^2}{n-1}$$

MLE parameters of various distributions

- **Lognormal distribution: complete data**
  - Two-sided confidence limits on $\hat{\mu}$ **t → one-tailed t-distribution**

$$\hat{\mu} - \frac{\hat{\sigma}}{\sqrt{n}} * t_{\frac{\gamma}{2}}(n-1) < \hat{\mu} < \hat{\mu} + \frac{\hat{\sigma}}{\sqrt{n}} * t_{\frac{\gamma}{2}}(n-1)$$

  - One sided confidence limits on $\hat{\mu}_t$

$$0 \leq \hat{\mu} < \hat{\mu} + \frac{\hat{\sigma}}{\sqrt{n}} * t_{\gamma}\ (n-1)$$

$$\hat{\mu} - \frac{\hat{\sigma}}{\sqrt{n}} * t_{\gamma}(n-1) < \hat{\mu} < \infty$$

  - And corresponding on σ_t^2:

$$\frac{(n-1)\hat{\sigma}^2}{\chi^2_{1-\frac{\gamma}{2}}(n-1)} \leq \sigma^2 \leq \frac{(n-1)\hat{\sigma}^2}{\chi^2_{\frac{\gamma}{2}}(n-1)}$$

$$0\ \leq \sigma^2 \leq \frac{(n-1)\hat{\sigma}^2}{\chi^2_{\gamma}(n-1)}$$

$$\frac{(n-1)\hat{\sigma}^2}{\chi^2_{1-\gamma}(n-1)} \leq \sigma^2 \leq \infty$$

Example: Lognormal MLE & confidence intervals

- **Example:** Consider the follow time-to-failure values t. Assuming the data are from a lognormal distribution, find:
- A) point estimates of the parameters
- B) the 90% confidence interval on the parameters
- C) The MTTF

$\boldsymbol{Lnt_i}$	4.3	4.7	5.3	5.7	5.9	6.0	6.2
$\boldsymbol{t_i}$	75	115	192	312	389	410	496

Solution: Lognormal MLE & confidence intervals

$\boldsymbol{Lnt_i}$	4.3	4.7	5.3	5.7	5.9	6.0	6.2
$\boldsymbol{t_i}$	75	115	192	312	389	410	496

- **Solution A)** $\hat{\mu}_t = \Sigma \frac{lnt_i}{n} = 5.46 \quad \hat{\sigma}_t = 0.71$
- **Solution B) For $\boldsymbol{\mu_t}$:**

$$\hat{\mu} - \frac{\hat{\sigma}}{\sqrt{n}} * t_{\frac{\gamma}{2}}(n-1) < \mu < \hat{\mu} + \frac{\hat{\sigma}}{\sqrt{n}} * t_{\frac{\gamma}{2}}(n-1)$$

n=7

$t_{\left(\frac{\gamma}{2}\right)}(7-1) = t_{0.05}(6) = 1.943$

$$5.46 - \frac{0.71}{\sqrt{7}} * 1.943 \leq \mu \leq 5.46 + \frac{0.71}{\sqrt{7}} * 1.943$$

$$\mathbf{4.93 \leq \boldsymbol{\mu} \leq 5.98}$$

Solution: Lognormal MLE & confidence intervals

- **Solution B: for $\boldsymbol{\sigma^2}$**

$$\frac{(n-1)\hat{\sigma}^2}{\chi^2_{1-\frac{\gamma}{2}}(n-1)} \le \sigma^2 \le \frac{(n-1)\hat{\sigma}^2}{\chi^2_{\frac{\gamma}{2}}(n-1)}$$

$$\frac{(6)0.71^2}{\chi^2_{0.95}(6)} \le \sigma^2 \le \frac{(6)0.71^2}{\chi^2_{0.05}(6)}$$

$$\frac{(6)0.71^2}{12.59} \le \sigma^2 \le \frac{(6)0.71^2}{1.64}$$

$$\mathbf{0.240 \le \boldsymbol{\sigma}^2 \le 1.844}$$

- **Solution C:**

$$\mathbf{MTTF} = \widehat{\boldsymbol{E}(\boldsymbol{t})} = \exp\left(\hat{\mu} + \frac{\hat{\sigma}^2}{2}\right) = \exp\left(5.46 + \frac{0.71^2}{2}\right) = \mathbf{302.49}$$

MLE parameter confidence intervals - Binomial

- **Binomial distribution**
  - Recall the binomial distribution:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

where

$$C_x^n = \binom{n}{x} = \frac{n!}{(n-x)!\,x!}$$

$p \rightarrow$ probability of failure

$q \rightarrow$ 1-p probability of success

$n \rightarrow$ number of trials

$x \rightarrow$ number of failures out of n trials

MLE parameter confidence intervals - Binomial

MLE for complete data: $\hat{p} = \frac{x}{n}$

Confidence limits on p can be found from the **Clopper-Pearson procedure:**

$$p_{Lower} = \left\{1 + \frac{(n - x + 1)}{x} F_{1-\frac{\gamma}{2}}(2n - 2x + 2; 2x)\right\}^{-1}$$

$$p_{Upper} = \left\{1 + \frac{n - x}{(x + 1) F_{1-\frac{\gamma}{2}}(2x + 2; 2n - 2x)}\right\}^{-1}$$

where $F_{1-\gamma/2}(f_1; f_2)$ is the F distribution with f_1 and f_2 degrees of freedom to the right and left, respectively, for $(1 - \gamma/2)$ confidence level.

See Tables in Appendix. Or in Excel: F.INV($1 - \gamma/2, f_1, f_2$)

MLE parameter confidence intervals (cont.)

- **Example:** An emergency pump is in standby mode. There have been 563 start tests for the pump, and 3 failures have been observed.
  A. Estimate the probability of failure on demand, p_{fod} and
  B. Find the 90% confidence interval for the probability of failure on demand.

MLE parameter confidence intervals (cont.)

- **Solution:**
  A. Number of trials, n = 563, and 3 failures, therefore:

$$\hat{\boldsymbol{p}}_{\boldsymbol{fod}} = \frac{x}{n} = \frac{3}{563} = \mathbf{0.0053}$$

MLE parameter confidence intervals (cont.)

- **Solution:**
  B. For confidence level 90%, $(1-\gamma) = .9$, thus $\frac{\gamma}{2} = 0.05, 1 - \frac{\gamma}{2} = 0.95$
- Given x=3 failures in n=563 tests:

$$p_L = \left\{1 + \frac{(563 - 3 + 1)}{3} * F_{(0.95)}\ (1122; 6)\right\}^{-1}$$

$$= \frac{1}{1 + \frac{561}{3} * 3.67} = 0.00145$$

$$p_U = \left\{1 + \frac{563-3}{(3+1)F_{0.95}(8;1120)}\right\}^{-1} = \frac{1}{1+\frac{560}{4*1.94}} = 0.0137$$

$$\boxed{\begin{array}{c}\hat{\boldsymbol{p}} = \mathbf{0.0053} \\ \mathbf{0.00145} \le \boldsymbol{p} \le \mathbf{0.0137,}\end{array}}$$

MLE parameter confidence intervals (cont.)

- **Weibull distribution**
  - When there is only complete failure and/or right censored data the point estimates can be solved using the following expressions.
  - Note that **numerical methods** are needed to solve $\hat{\beta}$ then substitute to find $\hat{\alpha}$. *To try this out: Use the SOLVER function in in Excel.*

$$\hat{\beta} = \left[\frac{\Sigma (t_i)^{\hat{\beta}} ln(t_i) + (n-r)(t_r)^{\hat{\beta}} ln(t_r)}{\Sigma (t_i)^{\hat{\beta}} + (n-r)(t_r)^{\hat{\beta}}} - \frac{1}{r}\sum ln(t_i)\right]^{-1}$$

Where t_i are complete data and t_r are right censored; r is the number of complete data points.

$$\hat{\alpha} = \left[\frac{\Sigma (t_i)^{\hat{\beta}}}{n} + (n-r)(t_r)^{\hat{\beta}}\right]^{\frac{1}{\hat{\beta}}}$$

MLE parameter confidence intervals (cont.)

- **Weibull distribution (cont.)**
  - Are derived from Fisher information Matrix and require numerical methods to solve.

$$\hat{\beta}\exp\left(-Z_{1-(\gamma/2)}\frac{\sqrt{\mathrm{var}(\hat{\beta})}}{\hat{\beta}}\right)\le\beta\le\hat{\beta}\exp\left(Z_{1-(\gamma/2)}\frac{\sqrt{\mathrm{var}(\hat{\beta})}}{\hat{\beta}}\right),$$

$$\hat{\alpha}\exp\left(-Z_{1-(\gamma/2)}\frac{\sqrt{\mathrm{var}(\hat{\alpha})}}{\hat{\alpha}}\right)\le\alpha\le\hat{\alpha}\exp\left(Z_{1-(\gamma/2)}\frac{\sqrt{\mathrm{var}(\hat{\alpha})}}{\hat{\alpha}}\right),$$

$$I(\alpha,\beta)=\begin{bmatrix}\frac{\beta^2}{\alpha^2} & \frac{\Gamma'(2)}{-\alpha}\\ \frac{\Gamma'(2)}{-\alpha} & \frac{1+\Gamma''(2)}{\beta^2}\end{bmatrix}=\begin{bmatrix}\frac{\beta^2}{\alpha^2} & \frac{1-\gamma}{\alpha}\\ \frac{1-\gamma}{\alpha} & \frac{\frac{\pi^2}{6}+(1-\gamma^2)}{\beta^2}\end{bmatrix}\cong\begin{bmatrix}\frac{\beta^2}{\alpha^2} & \frac{0.422784}{-\alpha}\\ \frac{0.422784}{-\alpha} & \frac{1.823680}{\beta^2}\end{bmatrix}$$

MLE parameter confidence intervals (cont.)

- **Example:** 5 components are put on a test with the following failure times: 535, 613, 976, 1031, 1875 hours
  - $\hat{\beta}$ is found by numerically solving:

$$\boxed{\hat{\beta} = \left[\frac{\Sigma\left(t_i^F\right)^{\hat{\beta}} ln\left(t_i^F\right)}{\Sigma\left(t_i^F\right)^{\hat{\beta}}} - 6.8118\right]^{-1} = 2.275}$$

  - $\hat{\alpha}$ is found by solving:

$$\boxed{\hat{\alpha} = \left[\frac{\Sigma\left(t_i^F\right)^{\hat{\beta}}}{n_F}\right]^{\frac{1}{\hat{\beta}}} = 1140}$$

$$\boxed{cov\left(\hat{\alpha},\hat{\beta}\right) = \begin{bmatrix} 278386 & 293 \\ 293 & 3.1463 \end{bmatrix}}$$

MLE Parameters of various distributions

Gamma Distribution

$$L(\alpha,\beta|t_i \dots) = \frac{1}{\beta^{\alpha n}\Gamma(\alpha)^n}\prod_{i=1}^{n} t_i^{\alpha-1}\exp\left(-\frac{t_i}{\beta}\right)$$

$$\ln L = \alpha n \ln\left(\frac{1}{\beta}\right) - n\ln[\Gamma(\alpha)] + (\alpha-1)\sum_{i=1}^{n}\ln(t_i) - \frac{1}{\beta}\sum_{i=1}^{n} t_i$$

$$\left.\frac{\partial \ln L}{\partial \alpha}\right|_{\alpha=\hat{\alpha}} = n\ln\left(\frac{1}{\beta}\right) - n\Psi(\alpha) + \sum_{i=1}^{n} \ln(t_i) = 0$$

$$\left.\frac{\partial \ln L}{\partial \beta}\right|_{\beta=\hat{\beta}} = \alpha\beta n - \sum_{i=1}^{n} t_i = 0$$

where $\Psi(\alpha) = \frac{d}{dx}\ln[\Gamma(\alpha)]$ digamma function

- Solve by using numerical methods on both equations simultaneously

MLE Parameters of various distributions

Continuous Uniform Distribution

$$\boldsymbol{\hat{a} = \min(t_i, \dots, t_n)}$$
$$\boldsymbol{\hat{b} = \max(t_i, \dots, t_n)}$$

Beta Distribution

$$L(\alpha, \beta; \mathrm{t_i} \dots) = \frac{\Gamma(\alpha+\beta)n}{\Gamma(\alpha)\Gamma(\beta)} \prod_{i=1}^{n} t_i^{\alpha-1}(1-t_i)^{\beta-1}$$

$$lnL = n\{\ln[\Gamma(\alpha+\beta)] - \ln[\Gamma(\beta)]\} + (\alpha-1)\sum_{i=1}^{n} lnt_i + (\beta-1)\sum_{i=1}^{n} \ln(1-t_i)$$

$$\boldsymbol{\left.\frac{\partial lnL}{\partial \alpha}\right|_{\alpha=\hat{\alpha}} = \Psi(\alpha) - \Psi(\alpha+\beta) - \frac{1}{n}\sum_{i=1}^{n} \ln t_i = 0}$$

$$\boldsymbol{\left.\frac{\partial lnL}{\partial \beta}\right|_{\beta=\hat{\beta}} = \Psi(\beta) - \Psi(\alpha+\beta) - \frac{1}{n}\sum_{i=1}^{n} \ln(1-t_i) = 0}$$

where $\Psi(a) = \frac{d}{d\alpha}\ln[\Gamma(\alpha)]$ is the digamma function

- Solve by using numerical methods on both equations simultaneously

MLE Parameters of various distributions

Truncated Normal Distribution

- First find point estimates for $z_a = \frac{a_L - \mu}{\sigma}$ and $z_b = \frac{b_U - \mu}{\sigma}$

$$H_1(z_1, z_b) = \frac{\hat{\mu} - a_L}{b_U - a_L} \qquad \hat{\mu} = \frac{1}{n}\sum_0^n x_i$$

$$H_2(z_a, z_b) = \frac{\sigma^2}{(b_U - a_L)^2} \qquad \hat{\sigma}^2 = \frac{1}{n-1}\sum_0^n (x_i - \mu)^2$$

- Solving for H_1 and H_2 simultaneously gives:
- $\boldsymbol{\hat{\sigma} = \frac{b_U - a_L}{\hat{z}_b - \hat{z}_a}}$ $\qquad$ $\boldsymbol{\hat{\mu} = a_L - \hat{\sigma}}$

MLE Parameters of various distributions

Multivariate Normal Distribution

$$\widehat{\vec{\boldsymbol{\mu}}} = \frac{1}{n}\sum_{t=1}^{n} \vec{x}_t$$

$$\hat{\Sigma}_{ij} = \frac{1}{n-1}\sum_{t=1}^{n}\left(x_{i,t} - \hat{\mu}_i\right)\left(x_{j,t} - \hat{\mu}_j\right)$$

$$\vec{x}_t = \begin{bmatrix} x_{1,t} \\ x_{2,t} \\ \vdots \\ x_{d,t} \end{bmatrix}, t = 1,2,\dots,n$$

Bivariate Normal Distribution

$$\hat{\mu}_{x_1} = \frac{1}{n}\sum_{i=1}^{n} x_{i,1};\ \hat{\sigma}_{x_1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{i,1} - \hat{\mu}_{x_1}\right)^2$$

$$\hat{\mu}_{x_2} = \frac{1}{n}\sum_{i=1}^{n} x_{i,2};\ \hat{\sigma}_{x_2}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{i,2} - \hat{\mu}_{x_2}\right)^2$$

$$\hat{\rho} = \frac{1}{n\hat{\sigma}_{x_1}\hat{\sigma}_{x_2}}\sum_{i=1}^{n}\left(x_{i,1} - \hat{\mu}_{x_1}\right)\left(x_{i,2} - \hat{\mu}_{x_2}\right)$$

Nonparametric methods for censored data

- Nonparametric (Empirical) reliability estimates with censored data - Apply to right censored data.
  - Relates to the nonparametric methods we learned earlier: Kimball, Kaplan Meier Plotting methods we discussed earlier.
  - There are ways to do this by relating the sample to the binomial distribution – see textbook.
- Start with sorted (ordered) data, and note the units that actually failed vs. censored data.
  - Right censored data are shown by a + sign).
  - 150, 340+, 560, 800, 1130+, 1720, 2470+, 4210+, 5230, 6890

Rank Adjustment Method

- The *rank adjustment method* is the most accurate method for plotting censored failure data.
  - For *n* units tested, ordered from $t_{i=1} \leq t_{i=2} \leq \cdots \leq t_{i=n}$where *m* units have survived including and beyond the ith unit..
- We calculated a rank adjustment (or order) for each data point:

$i_{t_i} = j_{t_{i-1}} + \frac{(n+1)-j_{t_{i-1}}}{1+m}$ (Call i_{t_i} the *initial rank*)

- Note: $\mathrm{i}_{\mathrm{t}_1} = 1$;
  - The *adjusted rank,* j_{t_i}, for non-censored units can be calculated as:

$$\mathrm{j}_{\mathrm{t_i}} = \mathrm{j}_{\mathrm{t_{i-1}}} + \frac{(\mathrm{n}+1) - \mathrm{j}_{\mathrm{t_{i-1}}}}{\mathrm{n} - \mathrm{i}_{\mathrm{t_i}} + 2}$$

where (*i*) is the initial rank (order) and (*j*) is the adjusted rank. We then use the non-parametric estimators to plot.

Rank adjustment method for censored data plots (cont.)

- **Example:** Given the following failure times for n = 10 components, 150, 340+, 560, 800, 1130+, 1720, 2470+, 4210+, 5230, 6890

A) Use the rank adjustment method to create a plot of the Weibull distribution for the failure times.

B) Assuming the Weibull is a suitable fit, use the plot to estimate the parameters of the Weibull distribution.

Rank adjustment method for censored data plots (cont.)

- **Solution A)**

Weibull plotting positions

i	t_i (hrs)	$j_{t_i} = j_{t_{i-1}} + \frac{(n+1) - j_{t_{i-1}}}{2 + n - i_{t_i}}$	$F(t_i) = \left(\frac{j_{t_i} - 0.375}{n + 0.25}\right)$	R(t) =1-F(t)	Ln(t)	$\ln(\ln(\frac{1}{R(t)}$
1	150	1	$\left(\frac{1 - 0.375}{10 + 0.25}\right) = .061$	0.939	5.011	-2.77
2	340+	-	-			
3	560	$1 + \frac{(10+1) - 1}{2 + 10 - 3} = 2.111$	$\left(\frac{2.111 - 0.375}{10 + 0.25}\right) = 0.169$	0.831	6.328	-1.68
4	800	$2.111 + \frac{(10+1) - 2.111}{2 + 10 - 4} = 3.222$	.2778	0.722	6.685	-1.12
5	1130+	-	-			
6	1720	$3.222 + \frac{(10+1) - 3.222}{2 + 10 - 6} = 4.519$	0.4042	0.596	7.450	-0.66
7	2470+	-	-			
8	4210+	-	-			
9	5230	6.679	0.615	0.385	8.562	-0.05
10	6890	8.84	0.8258	0.174	8.838	0.56

Rank adjustment method for censored data plots (cont.)

Solution B)

Least Squares regression equation:
y=0.815x-6.777
R^2=0.98

Weibull parameters:
α=4084.8
$\beta = 0.815$
(Reminder, slope=β)

(You can also compare to the MLE estimate of parameters: α=3926.86; $\beta = 0.97$)

Kaplan-Meier method (aka product-limit method)

- Applicable for right censored data. And grouped data (see textbook)
- Each term in the expression below is **Conditional Probability of Surviving** past time t. The product is the **Unconditional Surviving Probability (aka the survival function):**

$$\hat{R}(t) = \Pi_{t_j \le t}\left(\frac{n_j - 1}{n_j}\right)$$

where j = reverse rank.

- For $0 \le t \le t_i \quad R(t) = 1$

A measure of uncertainty of the estimated reliability is:

$$Var\left[\hat{R}(t)\right] = \Sigma_{t_j \le t}\frac{1}{n_j(n_j - 1)}$$

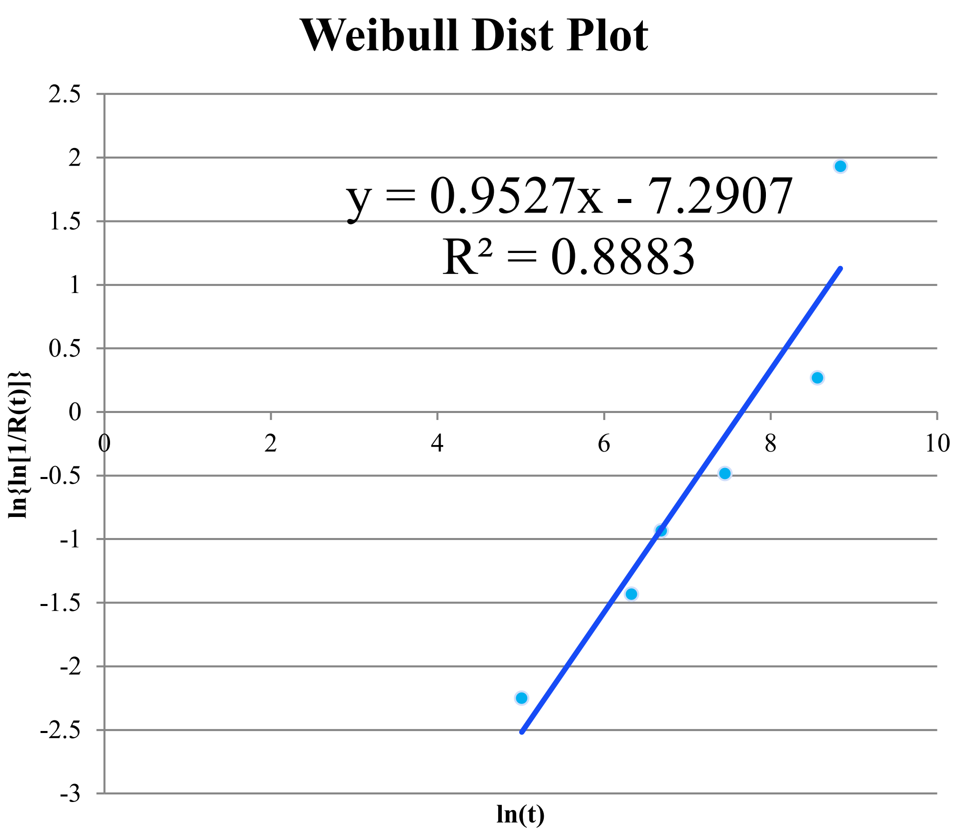
Kaplan-Meier method (cont.)

- **Example using same data as before:**

Weibull plotting positions

i	t_i	j	$\frac{n_j - 1}{n_j}$	$\hat{R}(t_i)=\Pi_{t_j \le t}\left(\frac{n_j-1}{n_j}\right)$	ln(t)	Ln(ln(1/R(t))
1	150	10	0.900	0.900 x 1.000 = 0.900	5.01	-2.25
2	340+	9	-	-		
3	560	8	0.875	0.875 x 0.900 = 0.788	6.33	-1.43
4	800	7	0.857	0.857 x 0.788 = 0.675	6.68	-0.93
5	1130+	6	-	-		
6	1720	5	0.800	0.800 x 0.675 = 0.540	7.45	-0.48
7	2470+	4	-	-		
8	4210+	3	-	-		
9	5230	2	0.500	0.500 x 0.540 = 0.270	8.56	0.27
10	6890	1	0.000	0.000 x 0.270 = 0.000	8.84	1.93

Kaplan-Meier plot



- $\beta = 0.953, \alpha = 2105$

Kaplan-Meier method (cont.)

- **Note:** If two or more failure occurs at time t_j then

$$\hat{R}(t) = \Pi_{t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right)$$

where d = number of failure in the jth time ranking