Chapter 12

Comparing Failure-Time Distributions

W. Q. Meeker, L. A. Escobar, and F. G. Pascual
Iowa State University, Louisiana State University, and Washington State University.

Based on Meeker, Escobar, and Pascual (2021): *Statistical Methods for Reliability Data, Second Edition*, John Wiley & Sons Inc.

May 24, 2021
10h 58min

Chapter 12
Comparing Failure-Time Distributions

Topics discussed in this chapter are:

- Background and motivation for comparing failure-time distributions.
- Nonparametric methods for comparing failure-time distributions.
- Parametric comparison of two distribution quantiles without making the assumption that the distribution σ values are equal.
- Parametric comparison of two distribution quantiles assuming that the distribution σ values are equal.
- Generalizations of the procedures for comparing three or more processes or populations.

Chapter 12

Segment 1

Motivation for Comparing Failure-Time Distributions and Graphical Nonparametric Comparison

Motivation for Comparing Failure-Time Distributions

There are many reasons for comparing failure-time distributions. Examples include:

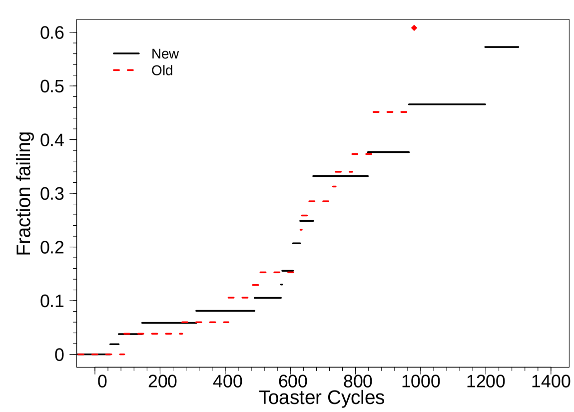
- Comparing two materials.
- Comparing product-design options.
- Compare products manufactured at different points in time or coming from different manufacturing locations.
- Comparing different vendors.
- Comparing different groups to see if data can be pooled or not.

It is important to distinguish between practically and statistically significant differences.

Snubber Life Test Data

- A snubber is a component in a pop-up toaster.
- Multiple censoring due to another failure mode.
- Purpose of the test was to compare the failure-time distribution for a New design with that of an existing Old design.
- Data first presented in Nelson (1982).

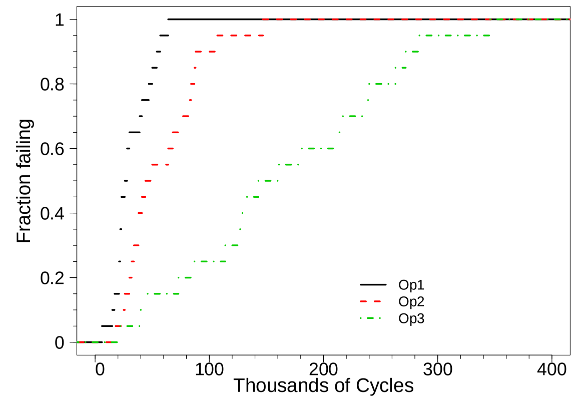
Kaplan-Meier Estimates for the Failure-Time Distributions of the Two Snubber Designs



Comparison of Operators Testing Part-A

- Part-A is a component of a cutting tool.
- Life tests are used to assess life length of Part-A and assure that the life length of current production is satisfactory.
- The life tests are conducted by operators who subject Part-A to constant simulated use to accelerate the test.
- There is concern that operator-to-operator variability is complicating the interpretation of the life-test data.
- Are there important differences among the operators?

Kaplan-Meier Estimates for the Failure-Time Distributions of the Different Part-A Operators



Chapter 12

Segment 2

Nonparametric Tests to Compare Distributions

Motivation for Nonparametric Tests to Compare Failure-Time Distributions

- Visual comparison of nonparametric estimates of different failure-time distributions is important.
- When there appears to be practically important differences, it is generally useful to see if the differences are statistically significant.
- Nonparametric tests do not require an assumed parametric distribution.
- Many different kinds of nonparametric tests have been developed.
- The available tests differ in the kinds of differences one wants to detect.
- In some applications, it may be important to detect changes
  - Near the center of a distribution.
  - In the lower tail of a distribution.

Differences Between the Observed and Expected Number of Failures

- Want to test that $F_1(t) = F_2(t) = \cdots = F_m(t)$.
- Complete or right-censored data are available for each of the m groups.
- $t_{(1)} < \cdots < t_{(r)}$ are the r unique failure in the data pooled across all m groups.
- d_{ij} and n_{ij} denote the number of failures at time $t_{(i)}$ and the number of units at risk just before time $t_{(i)}$, respectively, from group j.
- $d_i = \sum_j^m d_{ij}$ and $n_i = \sum_j^m n_{ij}$ denote the number of failures at time $t_{(i)}$ and the number of units at risk just before time $t_{(i)}$, respectively, for the pooled data.
- The difference between the observed and expected number of failures at time $t_{(i)}$ for group j when $F_1(t) = \cdots = F_m(t)$ is

$$\left(d_{ij} - \frac{n_{ij} d_i}{n_i}\right),\ i = 1, \ldots, r,\ j = 1, \ldots, m.$$

Weighted Logrank Test Statistic

- The weighted sum of the differences between the observed and expected number of failures for group j is

$$U_j = \sum_{i=1}^{r^*} w(i)\left(d_{ij} - \frac{n_{ij} d_i}{n_i}\right),\ j = 1, \ldots, m.$$

  where r^* is the index of the largest unique failure time where the size of the risk set is positive for at least two groups.
- The weighted logrank test statistic is $X^2 = \boldsymbol{U}' \widehat{V}^{-1} \boldsymbol{U}$.
- Because $\sum_{j=1}^m U_j = 0$, only $m-1$ components need to be used in the test statistic so $\boldsymbol{U} = (U_1, \ldots, U_{m-1})'$.
- $\widehat{V} = \widehat{\mathrm{Var}}(\boldsymbol{U})$ is an $(m-1) \times (m-1)$ estimated variance-covariance matrix.
- $w(i)$ is a specified weight function that can be chosen to focus the test on specified parts of a distribution.
- When samples are large and the m distributions are the same, $X^2 \dot\sim \chi^2_{(m-1)}$.

Variance-Covariance Matrix Estimate

The jk element of $\widehat{V} = \widehat{\text{Var}}(\boldsymbol{U})$ is

$$\widehat{V}_{jk} = \sum_{i=1}^{r^*} [w(i)]^2 \frac{n_{ij}}{n_i}\left(\delta_{jk} - \frac{n_{ik}}{n_i}\right)\frac{(n_i - d_i)d_i}{n_i - 1}, \; j,k = 1,\ldots,(m-1)$$

where

$$\delta_{jk} = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{otherwise.} \end{cases}$$

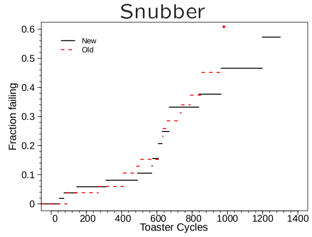
Choosing the Weight Function

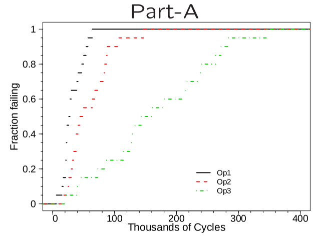
- Many different methods for choosing the weight function $w(i)$, $i = 1,\ldots,r^*$ have been suggested.
- The weight choices reflect the part of the distribution where it is important to detect differences.
- Two common choices are
  - ▶ **Equal weights** $w(i) = 1, \ldots\ i = 1, r$. This is sometimes called the **logrank** test.
  - ▶ **Survival weights**

$$w(i) = \widehat{S}(i) = \prod_{j=1}^{i} \frac{n_j + 1 - d_j}{n_j + 1}, \; i = 1, \ldots r^*.$$

where $\widehat{S}(i)$ is approximately equal to the Kaplan-Meier estimate of the survival function. This choice provides more sensitivity to detect differences in the lower tail of the distribution. The resulting test is sometimes called the **generalized Wilcoxon** test.

Weighted Logrank Test Examples





- For the Snubber example, there is no evidence of any difference:
  - ▶ The equal-weight $X^2 = 0.152$ and p-value $= 0.70$.
  - ▶ The survival-weight $X^2 = 0.034$ and p-value $= 0.85$.
- For the Part-A example, there is strong evidence of differences:
  - ▶ The equal-weight $X^2 = 46.1$ and p-value < 0.001.
  - ▶ The survival-weight $X^2 = 35.7$ and p-value < 0.001.

Weighted Logrank Test Comments

The weighted logrank tests:

- Are nonparametric and thus are valid for any continuous distribution.
- Assume that observations within each group are independent and identically distributed.
- Assume that censoring is non-informative.
- Are distribution-free (i.e., the distribution of X^2 does not depend on form of the underlying continuous distribution) when there is no censoring.
- Are asymptotically efficient (relative to a particular parametric model) when there is no censoring.
- Perform poorly when distributions cross. Alternative tests should be used in such cases.

Chapter 12

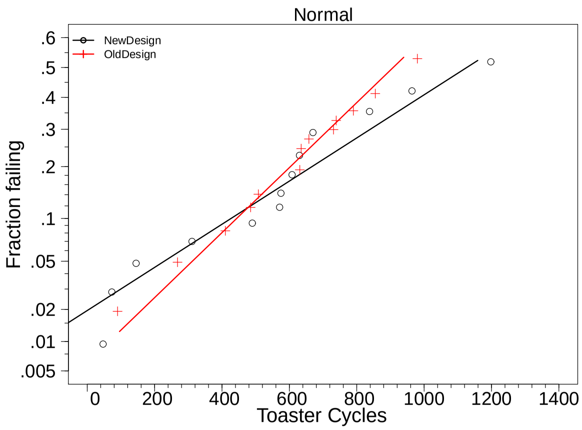
Segment 3

Parametric Comparison of Two Groups Using Separate Analyses

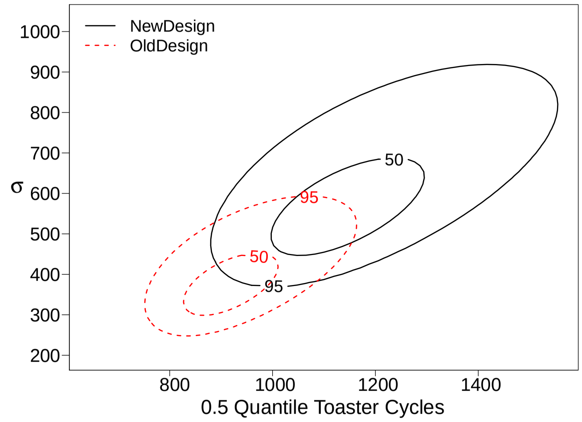
Strategy for Comparing the Failure-Time Distributions for Two or More Groups

- Comparison of nonparametric estimates of the different groups.
- Fit **separate distributions** to each group (usually log-location-scale or location-scale distributions). Compare different parametric distributions to find one that provides an adequate description of the data.
- Fit separate distributions to each group, under the constraint that the σ **parameter is the same** in all groups.
- Compare (using a likelihood-ratio test) the constant σ and separate distribution models to see if there is evidence of differences in the σ parameters.
- If there is no evidence of different σ values, check to see if there is evidence of differences in the μ parameters.
- If there is sufficient evidence of differences between (among) the groups, use the appropriate model to make the desired comparison(s).

Separate Normal Distribution ML Estimates for the Old and New Snubber Designs



Joint Confidence Regions for the Parameters of the Old and New Snubber Designs



Comparison of Snubber Designs—Separate Analyses (SepDists: different σ's and different μ's)

- In general comparison complicated. What should we compare? Typical choice: specified quantile or $F(t)$ at a specified t.
- Compare the $t_{0.5}$ (also μ for the normal distribution).

$\widehat{\mu}_{\text{new}} - \widehat{\mu}_{\text{old}} = 1126 - 908 = 218$

$$\text{se}_{\widehat{\mu}_{\text{new}} - \widehat{\mu}_{\text{old}}} = \sqrt{\text{se}^2_{\widehat{\mu}_{\text{new}}} + \text{se}^2_{\widehat{\mu}_{\text{old}}}} = \sqrt{(76.2)^2 + (123)^2} = 144.7$$

- An approximate 95% confidence interval for $\Delta = \mu_{\text{new}} - \mu_{\text{old}}$ is

$$\begin{aligned}[\underset{\sim}{\Delta}, \quad \widetilde{\Delta}] &= \widehat{\mu}_{\text{new}} - \widehat{\mu}_{\text{old}} \mp z_{(1-\alpha/2)} \text{se}_{\widehat{\mu}_{\text{new}} - \widehat{\mu}_{\text{old}}} \\ &= 218 \mp 1.96 \times 144.7 = [-66, \quad 501].\end{aligned}$$

Interval contains 0 and thus the difference between the means could be zero.

Chapter 12

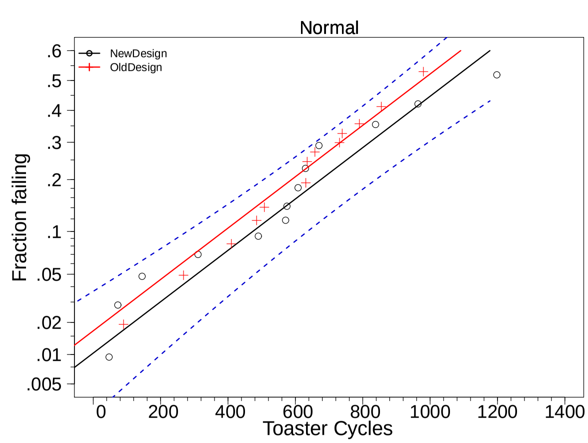
Segment 4

Parametric Comparison of Groups Using a Equal Spread Parameter

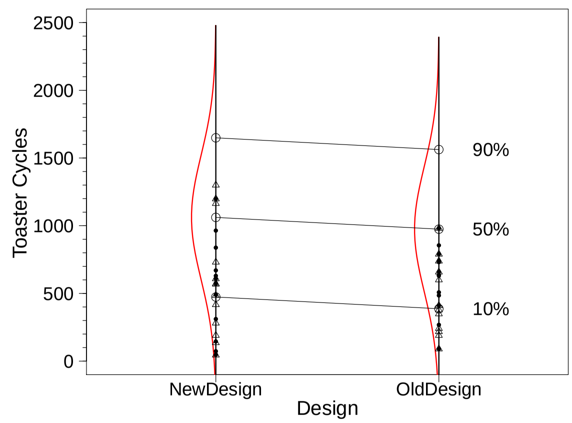
Likelihood Ratio Tests to Assess Evidence That the σ and μ Parameters Differ Between (Among) Groups

- Compare the log-likelihoods from the separate distributions for each group (SepDists) and σ constant (EqualSig) models.
- Log-likelihood for the SepDists model $\mathcal{L}_{\text{SepDists}}$ is the sum of the log-likelihoods for all of the groups.
- If $2(\mathcal{L}_{\text{SepDists}} - \mathcal{L}_{\text{EqualSig}}) > \chi^2_{(1-\alpha,\nu)}$, then there is statistical evidence of differences in the σ values in across the groups where degrees of freedom ν is the difference in the number of parameters in the SepDists and EqualSig models.
- If there is no evidence of different values of σ, then one might want to test to see if there are differences between (among) the different values of μ by fitting a single distribution to the data (Pooled model).
- If $2(\mathcal{L}_{\text{EqualSig}} - \mathcal{L}_{\text{Pooled}}) > \chi^2_{(1-\alpha,\nu)}$, then there is statistical evidence of differences in the μ values across the groups. The degrees of freedom ν is the difference in the number of parameters in the EqualSig and Pooled models.

Equal-σ Normal Probability Plot and ML Estimates from the Old and New Snubber Designs



Model Plot Showing the Equal-σ Normal Distribution ML Estimates from the Old and New Snubber Designs



Likelihood Ratio Tests for the Snubber Example

- For the SepDists model, there are 4 parameters and $\mathcal{L}_{\text{old}} = -138.6$ for the old design and $\mathcal{L}_{\text{new}} = -146.8$ for the new design, with a combined log-likelihood $\mathcal{L}_{\text{SepDists}} = \mathcal{L}_{\text{old}} + \mathcal{L}_{\text{new}} = -285.4$.
- For the EqualSig model, there are 3 parameters and $\mathcal{L}_{\text{EqualSig}} = -286.7$. Thus $\mathcal{L}_{\text{SepDists}} - \mathcal{L}_{\text{EqualSig}} = 2[-285.4-(-286.7)] = 2.6 < \chi^2_{(0.95,1)} = 3.8415$, indicating that differences between σ_{new} and σ_{old} are not statistically significant.
- For the Pooled model, there are 2 parameters and $\mathcal{L}_{\text{Pooled}} = -286.9$. Thus $\mathcal{L}_{\text{EqualSig}} - \mathcal{L}_{\text{Pooled}} = 2[-286.7-(-286.9)] = 0.573 < \chi^2_{(0.95,1)} = 3.84$, indicating that differences between μ_{new} and μ_{old} are not statistically significant.

Snubber Model-Fitting Summary

Model	−2LogLike	AIC	# Param
SepDists	570.72	578.72	4
EqualSig	573.32	579.32	3
Pooled	573.89	577.89	2

Snubber Likelihood Ratio Tests

Comparison	LR Statistic	dof	p-value
EqualSig vs SepDists	2.59991	1	0.11
Pooled vs EqualSig	0.57294	1	0.45

Comparison of Snubber Designs (EqualSig: different μ's and common σ)

- Simple regression model using dummy variables. $\mu(w) = \beta_0 + \beta_1 w$ where $w = 0$ for old design and $w = 1$ for the new design.
- Substituting $w = 0, 1$ into the model gives

$$\mu(0) = \beta_0, \quad \text{for the old design}$$
$$\mu(1) = \beta_0 + \beta_1, \quad \text{for the new design}$$

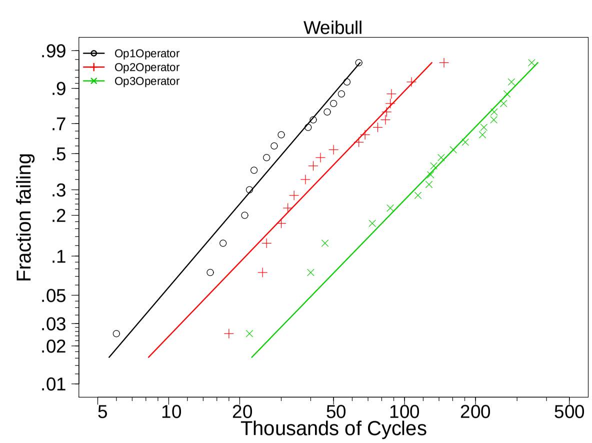
- The model assumes that σ is the same for both designs.
- Note that $\Delta = t_p(1) - t_p(0) = \mu(1) - \mu(0) = \beta_1$, so Δ does not depend on the choice of which quantile to compare.
- $[\underset{\sim}{\beta}_1, \quad \widetilde{\beta}_1] = \widehat{\beta}_1 \mp z_{1-\alpha/2} \text{se}_{\widehat{\beta}_1} = 86.7 \mp 1.96 \times 114 = [-137, \quad 311]$

Chapter 12

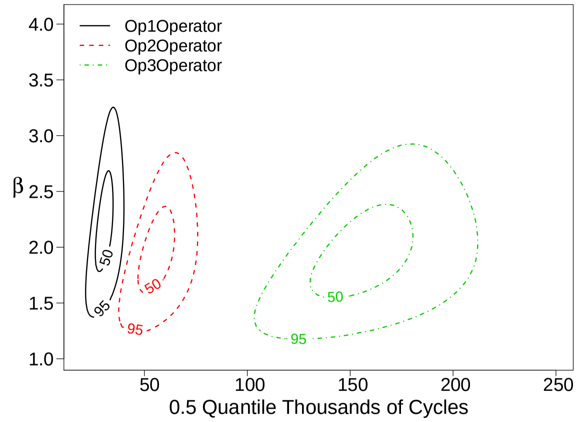
Segment 5

Parametric Comparison of the Part-A Operators

Separate Weibull Distribution ML Estimates for the Different Part-A Operators



Joint Likelihood Confidence Regions for the Parameters of the Different Part-A Operators



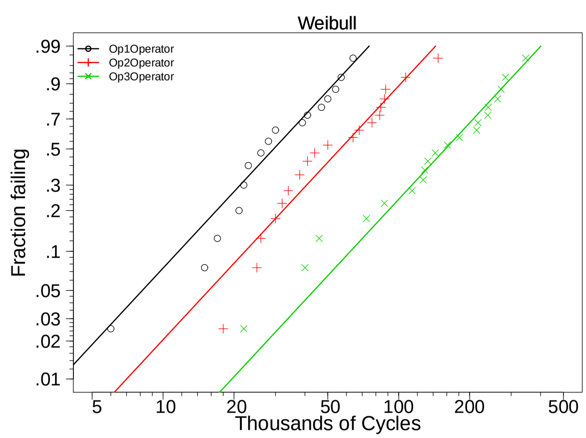
Comparison of Part-A Operators (EqualSig: different μ's and common σ)

- Simple regression model using dummy variables. $\mu = \beta_0 + \beta_1 w_2 + \beta_2 w_3$ where $w_i = 1$ for Operator i and $w_i = 0$ otherwise, $i = 2, 3$.
- Substituting w_2 and w_3 into the model gives

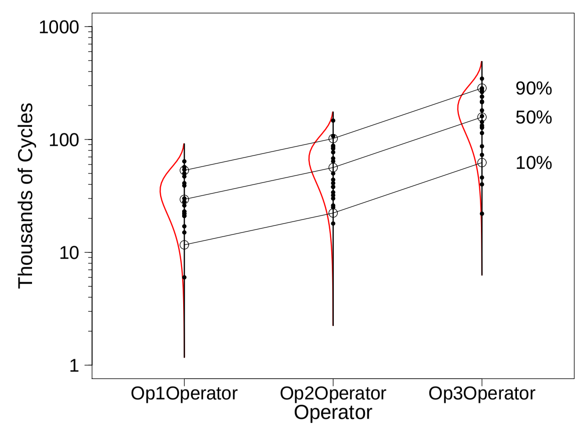
$$\mu_{\text{Op1}} = \beta_0$$
$$\mu_{\text{Op2}} = \beta_0 + \beta_1$$
$$\mu_{\text{Op3}} = \beta_0 + \beta_2$$

- The model assumes that σ is the same for both designs.
- This EqualSig model or the SepDists can be used to construct confidence intervals to compare quantiles (differences for location-scale and ratios for log-location-scale distributions) or failure probabilities for pairs of operators in the usual way.
- When constructing more than one confidence interval, methods of simultaneous inference should be used.

Weibull Probability Plot and Equal-β ML Estimates for the Part-A Operators



Model Plot Showing the Equal-β Weibull Distribution ML Estimates for the Part-A Operators



Part-A Model-Fitting Summary

Model	−2LogLike	AIC	# Param
SepDists	590.31	602.31	6
EqualSig	590.65	598.65	4
Pooled	651.96	655.96	2

Part-A Likelihood Ratio Tests

Comparison	LR Statistic	dof	p-value
EqualSig vs SepDists	0.34	2	0.84
Pooled vs EqualSig	61.31	2	< 0.001

References

Meeker, W. Q., L. A. Escobar, and F. G. Pascual (2021). *Statistical Methods for Reliability Data* (Second Edition). Wiley. [1]

Nelson, W. B. (1982). *Applied Life Data Analysis*. Wiley. []