# Introduction to Censored Data Analysis

Michael Crotty, PhD

Senior Statistical Writer, JMP

SAS Institute

# Outline

## Some Questions to Answer

- What is censoring?
- When do censored data appear?
- How do you specify censoring in JMP?
- What happens if you ignore the fact that your data are censored? [DEMO]
- What is an example of left-censored data in a regression setting? [DEMO]

# Many Types of Censoring

- Time Censoring (aka Type I)
  - Study ends at a specified time before all failures have occurred.
- Failure Censoring (aka Type II)
  - Study ends after specified number of failures have occurred.
- Interval Censoring
- Random Right Censoring
  - Can arise with multiple forms of failure, but only interested in one type.
- Systematic Multiple Censoring
  - Some failure times exceed some running times; arises from staggered entry.
- Left Censoring

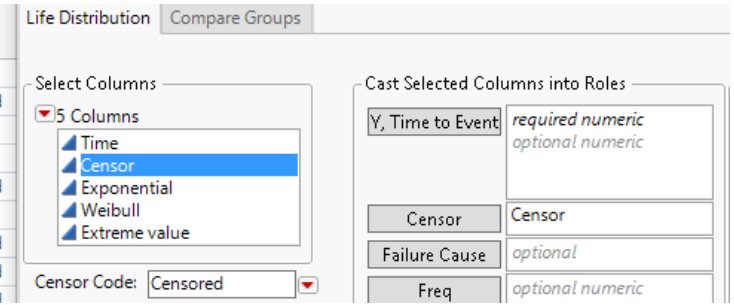
# Types of Censoring in JMP

- Right Censoring
- Left Censoring
- Interval Censoring
- Other
  - Censor Indicator in Failure Cause Column
  - Nevada format in Reliability Forecast platform
  - Destructive Degradation: censoring is for non-time-based response
  - Type II (failure censoring) in Reliability Growth and Recurrence

# Right Censoring

- Probably the most common form of censoring
- Event of interest does not have enough time to occur
- Common examples:
  - Reliability tests (light bulbs, etc.)
  - Survival models (patient survives past end of study)
- Two ways to specify right-censored observations in JMP:

| | Time | Censor |
|---|---|---|
| 1 | 450 | Failed |
| 2 | 460 | Censored |
| 3 | 1150 | Failed |
| 4 | 1150 | Failed |
| 5 | 1560 | Censored |
| 6 | 1600 | Failed |
| 7 | 1660 | Censored |
| 8 | 1850 | Censored |
| 9 | 1850 | Censored |




| | Start Time | End Time | Count | Censoring Type |
|---|---|---|---|---|
| 1 | • | 50 | 50 | Left |
| 2 | • | 100 | 6 | Left |
| 3 | 25 | • | 30 | Right |
| 4 | 75 | • | 10 | Right |
| 5 | 80 | 150 | 4 | Interval |
| 6 | 100 | 250 | 7 | Interval |

# Left Censoring

- Event of interest occurs before observation starts
- Common examples:
  - Failure at time of first inspection
  - Limit of detection

- Use two columns to specify left-censored observations in JMP:

| | Start Time | End Time | Count | Censoring Type |
|---|---|---|---|---|
| 1 | • | 50 | 50 | Left |
| 2 | • | 100 | 6 | Left |
| 3 | 25 | • | 30 | Right |
| 4 | 75 | • | 10 | Right |
| 5 | 80 | 150 | 4 | Interval |
| 6 | 100 | 250 | 7 | Interval |

# Interval Censoring

- Event of interest happens in between observation times
- Common examples:
  - Period inspection times (instead of continuous observation)

- Use two columns to specify interval-censored observations in JMP:

| | Start Time | End Time | Count | Censoring Type |
|---|---|---|---|---|
| 1 | • | 50 | 50 | Left |
| 2 | • | 100 | 6 | Left |
| 3 | 25 | • | 30 | Right |
| 4 | 75 | • | 10 | Right |
| 5 | 80 | 150 | 4 | Interval |
| 6 | 100 | 250 | 7 | Interval |

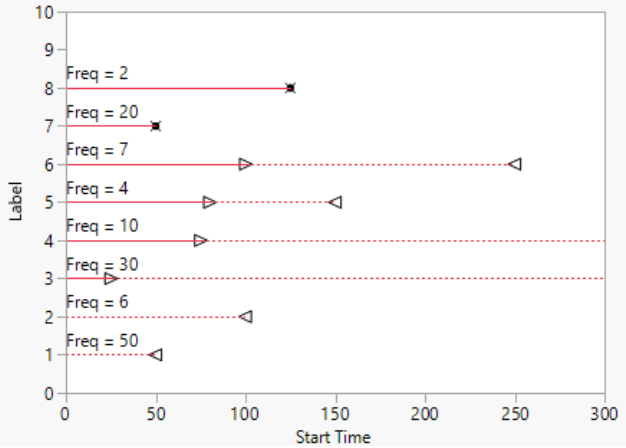
# Mixed Censoring in JMP

- Use two columns
- Specify both columns (in time order) in the response role
- For uncensored failures, use the same value in both columns
- Event Plot helps you visualize the censored observations

**Data**

| Label | Start Time | End Time | Count | Censoring Type |
|---|---|---|---|---|
| 8 | 125 | 125 | 2 | Exact |
| 7 | 50 | 50 | 20 | Exact |
| 6 | 100 | 250 | 7 | Interval |
| 5 | 80 | 150 | 4 | Interval |
| 4 | 75 | . | 10 | Right |
| 3 | 25 | . | 30 | Right |
| 2 | . | 100 | 6 | Left |
| 1 | . | 50 | 50 | Left |

**Event Plot**



# Background

## CDFs and PDFs

- Need some background to understand how censoring affects maximum likelihood (ML) estimation.
- The *cumulative distribution function* (CDF) is the probability of a random variable (generally time in censoring situations) being less than or equal to a particular value.

$$F(t) = \Pr(T \leq t)$$

- For continuous distributions, the *probability density function* (PDF) is the derivative of the CDF. Similar to a smoothed histogram of the responses. Relationship to CDF:

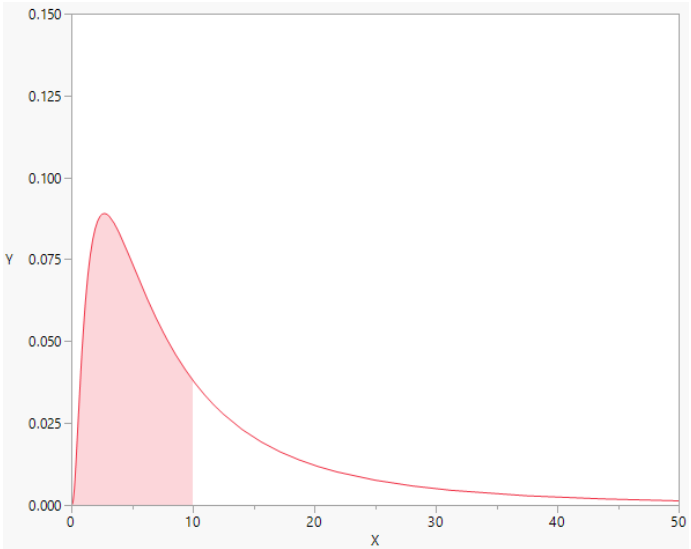
$$F(t) = \int_{-\infty}^{t} f(x)\, dx$$

  - Properties: PDFs are always nonnegative and the area under the PDF sums to 1.

# Background

## Area Under the PDF

- Another way to think of the CDF is as the area under the PDF:
  - Shaded region = F(10) = Pr( $X \leq 10$ )
  - This is a lognormal PDF, so it is positive only for $X > 0$.
  - So, F(0) = Pr( $X \leq 0$ ) = 0.
  - Then, F(10) = F(10) – F(0).



# Likelihoods

## Review of Non-Censored Case

- The *likelihood* is the product of the PDFs given the observed data, or the joint probability of the data.

$$L(\beta) = \prod_{i=1}^{N} f_i(x_i) = \prod_{i=1}^{N} L_i(\beta; x_i)$$

- We want to find the parameter values that maximize the likelihood.
  - These parameter values are the *maximum likelihood estimates* (MLEs).
  - They are the most likely since they maximize the joint probability of the data.
  - MLEs generally denoted $\hat{\beta} = (\hat{\mu}, \hat{\sigma})$, where $\mu$ and $\sigma$ are location and scale.

# Likelihoods

## Censored Data

- Interval-censored case:

$$L_i(\beta; t_i) = \int_{t_l}^{t_u} f(x)\, dx = F(t_u) - F(t_l)$$

- Left- and right-censored cases are special cases of interval-censored case.
- Left-censored case: $t_l = 0 \Rightarrow F(t_l) = F(0) = \Pr(T \leq 0) = 0$

$$L_i(\beta; t_i) = \int_{-\infty}^{t_u} f(x)\, dx = F(t_u) - F(0) = F(t_u)$$

- Right-censored case: $t_u = \infty \Rightarrow F(t_u) = F(\infty) = \Pr(T \leq \infty) = 1$

$$L_i(\beta; t_i) = \int_{t_l}^{\infty} f(x)\, dx = F(\infty) - F(t_l) = 1 - F(t_l)$$

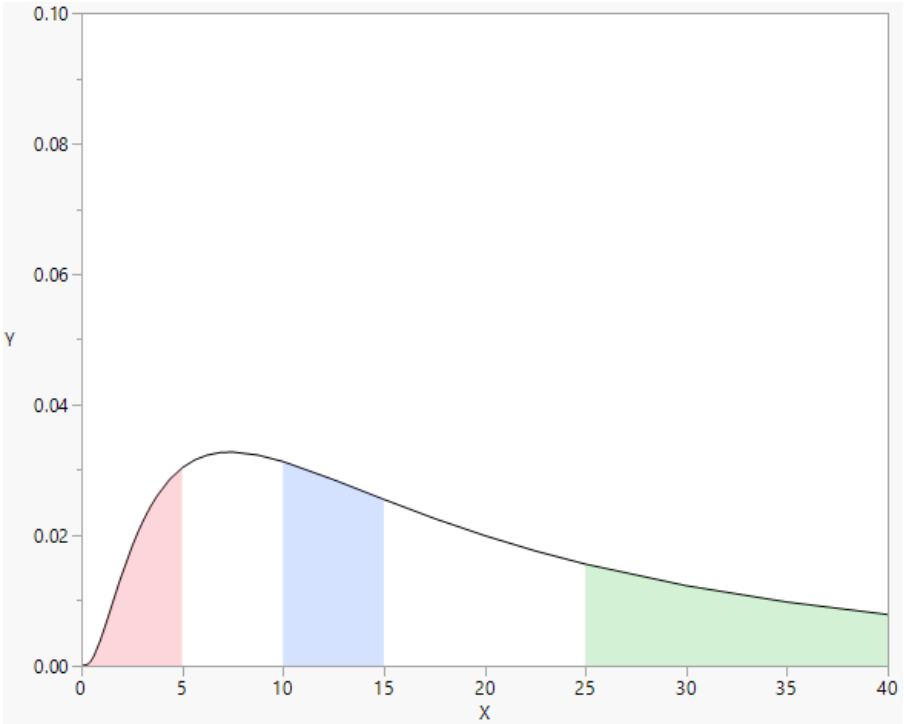
# Censored Data

## Area Under the PDF

- Left-censored:
  - $F(5)$
- Interval-censored:
  - $F(15) - F(10)$
- Right-censored:
  - $1 - F(25)$

- Likelihood:

$F(5) \times (F(15) - F(10)) \times (1 - F(25))$



# Software Demo

- Demo 1: Pitfalls of ignoring censored observations
- Demo 2: Limit of detection (left-censored observations)
- Demo 3: Fun example of interval censoring

# Conclusion

## What We Covered

- Better to use censoring information in your analysis.
- How to specify censoring in JMP.
  - Two-column response approach for right, left, and interval censoring.
  - One-column response + Censor column approach for right censoring.
- Censoring is often for time responses, but can be for other responses.
  - Limit of detection example.

# References


- Meeker, W. Q., and Escobar, L. A. (1998) *Statistical Methods for Reliability Data*. New York: John Wiley & Sons.
- SAS Institute Inc. (2017). *JMP® 14 Reliability and Survival Methods*. Cary, NC: SAS Institute Inc.

# Thank you!
# Questions?

sas.com



THE POWER TO KNOW®