**A Smart Girl’s Guide to Survival Content**

Absence or presence of an **event** is denoted with an **indicator variable**.

Δ = I(event happened)

Dealing with rates is a logistic regression type problem, but it ignores rates.

**Prospective** follows a group over time, and **retrospective** looks back at a group over time.

We define time variables with an **origin**, **scale**, and **event** of interest.

Question: When dealing with right censored data, should entry be standardized? Is it entry into a study or entry into the condition we’re looking for?

**Staggered entry** has individuals enter the study at different times.

**Censoring** is where our information cuts off.

**Right, left, interval, and double** are examples of censoring.

Censoring can be **random** or **non-random**.

Failure time (or time to event) is nonnegative and either discrete or continuous.

Censoring time is nonnegative.

**Failure time** has four related functions.

PDF

Discrete: p(x) = Pr(X = x) for all possible x

Continuous: f(x) =

Survival (probability of surviving past time x)

Pr(X > x) aka 1 – CDF in continuous case

Its negative derivative is the PDF

It’s a step function in the discrete cases

It’s also the product of the conditional survival probabilities

S = e^-H

Hazard (chance that one who survived to time x has an event in the next instant)

Conditioned on being upper bounded by that next instant

Good for looking at mechanism

Continuous: h = f/S

And f = -S’

Increasing failure-rate for nondecreasing h

Increasing failure-rate on average for nondecreasing H/x

Decreasing failure-rate for nonincreasing h

H’ = h

Mean Residual Life (E(time to event) given that event hasn’t yet happened by time x)

Cumulative Hazard (accumulated event risk up to time x) (H)

Integral from 0 to x of h(t) OR H(Xj) = sum from k = 1 to j of h(xk)

Die or disease recur is an event? Unsure of how to interpret a survival curve

Is hazard in terms of survival alone only true in discrete case?

Measures of Central Tendency for Failure Time (X)

Mean Residual Life (only considers those living at time x)

Mean time to event given event free until just before time x

The mrl(x) = E(X-x|X > x)

Expected remaining lifetime given they already lived to x

=

Mean and Median Survival

Pth quantile

Sinv(1 – p)

MRL def: E(X – x|X > x)

E(X|X>x) = x + mrl(x)

We use that adjustment

**A Smart Girl’s Guide to General Statistics Content**

* Non parametric estimators are fully unstructured estimators
  + Ex. estimate a CDF Fhat(y) as a long run frequency. However, it’s a step function. It’s usually stepwise, but if you have a conditional CDF Fhat(y|x) you can try to smooth it
* Semi parametric is the choice to replace x with a smooth x (it has some parameterization)
* Parametric: put a distribution on it
* [Weibull distributions1](https://stats.libretexts.org/Courses/Saint_Mary's_College_Notre_Dame/MATH_345__-_Probability_(Kuter)/4%3A_Continuous_Random_Variables/4.6%3A_Weibull_Distributions#:~:text=When%20%CE%B1%3D1%2C%20the%20Weibull,comparing%20the%20cdf's%20of%20each) is to model lifetimes that are not “memoryless”.
  + The parameter 𝛼 is referred to as the ***shape parameter***, and 𝛽 is the ***scale parameter***. When 𝛼=1, the Weibull distribution is an exponential distribution with 𝜆=1/𝛽, so the exponential distribution is a special case of both the Weibull distributions and the gamma distributions.
  + A random variable 𝑋 has a ***Weibull distribution*** with parameters 𝛼,𝛽>0 write 𝑋∼Weibull(𝛼,𝛽), if 𝑋 has pdf for nonnegative x, α, β

**A Smart Girl’s Guide to R**

* Cache = TRUE as a chunk option saves stuff that already ran

**A Smart Girl’s Guide to GitHub**

* Use GitKraken as a GUI to stage, commit, push, and pull
* Staging picks things to upload
* Committing is a mass upload to the history
* Push sends your commits to the cloud
* Pull grabs the most recent folder from the cloud

**A Smart Girl’s Guide to Model Types**

* Kaplan-Meier Estimator (1950s) – great nonparametric
  + Classical but not great with covariates because of sample size issues
  + Summarizes into life tables organized by each moment
  + Piecewise based on a time domain, drops incrementally
  + They call it the product limit (PL) estimator in their paper
  + So common that it isn’t always cited
  + Assumptions
    - Independence of death event and loss event (not always warranted)
    - We have an initial event observed
  + Notation
    - P(t): Proportion of the population whose lifetimes exceed time t without loss/censoring \*\*\*this is what we’re estimating\*\*\*
    - PL estimate: after listing N observed lifetimes from least to greatest,
    - (N = size of random sample, r = values where time to death tr’ ≤ t
  + Notes
    - We’re creating an MLE for the distribution (for our observations)

**A Smart Girl’s Guide to Having Receipts**

All instances of these sources are hyperlinked earlier in the document.

1. <https://stats.libretexts.org/Courses/Saint_Mary's_College_Notre_Dame/MATH_345__-_Probability_(Kuter)/4%3A_Continuous_Random_Variables/4.6%3A_Weibull_Distributions#:~:text=When%20%CE%B1%3D1%2C%20the%20Weibull,comparing%20the%20cdf's%20of%20each>.

**A Smart Girl’s Guide to Having a Plan**

1. Learn about non-parametric
   1. Kaplan Meier – (read sections 1 and 2, formula 2a (in light censoring mode and vary sample size))
2. Learn about semi-parametric
3. Learn about parametric

.

.