# Initial Posts

|  |
| --- |
| **What is Survival Analysis?**  Given a specific start event and end event, survival analysis aims to gain insight on the probability that a specific time interval might occur after a specified time. Although a lot of survival analysis deals with scenarios where the end date is death, destruction, or degradation, this is not a requirement for survival analysis.  Example survival analysis: Length of time before a patient is re-admitted after they have been discharged from a hospital.   * Start event: Patient discharged from hospital. * End event: Patient comes back to the hospital and is re-admitted.   For our analysis, we would be interested in the probability that a patient is re-admitted after <x> days (t). The end date has nothing to do with the patient dying but in my head I still try to associate the end date with survival which, if you think about it, is always time. For each observation, how much time survived between the start & end events? In the scenario above, I in my head I think about it as, "how much time survives between hospital discharge and the next hospital readmission? |
| **What is a Survival Curve**  A survival curve is a graphical representation of a survival function. There are various different survival functions which can represent survival:   * Kaplan-Meier: * Exponential: * Weibull:   *Image obtained from Wikipedia*  *Reference:*  Contributors to Wikimedia projects. (2022, February 14). Weibull distribution - Wikipedia. Retrieved from https://en.wikipedia.org/w/index.php?title=Weibull\_distribution&oldid=1071854085 |
| **What is a Hazard Function?**  For the clocks that were still ticking at the beginning of time period (t), Hazard measures the instantaneous rate of change in the survival for the current time period (t) and some other time in the future (t+delta). Below is an example of how it is calculated. |
| **Why you SHOULD Include Censored Data**  One very common problem in Survival Analysis is censored data. Censored meaning that that the duration is missing either a start time or end time. Some might be tempted to either use study start date for missing start time observations or study end date for missing end time observations which would produce underestimated survival probabilities. Even worse, someone may exclude censored data altogether which could potentially underestimate the survival even more. Below is a graph showing how this is the case.    As you can see, there is right-censored data in our set. If the study end date was used for the right-censored data, you can see how the predicted survival is underestimated. If we were to exclude them altogether it would make the problem even worse. |
| **Survival Diagram**  I found the below diagram in the Python package 'lifetimes' documentation and thought it was very cool and shows the relationship between CDF, Survival, Hazard, and Cumulative Hazard. It also shows the formula for each of them.    Reference:  Introduction to survival analysis — lifelines 0.26.4 documentation. (2022, January 19). Retrieved from <https://lifelines.readthedocs.io/en/latest/Survival%20Analysis%20intro.html> |

# Replies

|  |
| --- |
| Wow great article and very in-depth. Thank you for sharing Saima!  Hi Jeff, I think the main difference between a standard exponential model and Weibull model is that the Weibull model assumes that the Hazard is proportional to time (*t*). For example, as time increases, the Hazard will either increase or decrease. The standard exponential model however assumes that the Hazard is constant over time (*t*). |
| Very good point Chandrasekhar. Visually looking at your data may give you some insight on what distribution it might be but statistical testing is used to confirm. Also, picking a Kaplan-Meier vs. Exponential vs. Weibull vs. Cox PH, etc. etc. depends on what it is that you're wanting to measure. For example, a Kaplan-Meier can estimate survival probability but Cox PH cannot. |
| Hi Nicolas, I have also found this weeks topic to be very interesting as well. I work in healthcare which as you mentioned is very relatable. Do you work in healthcare as well Nicolas? |
| Hi Amelia & Robert. I haven't found much about how to handle censored data that you're including in your analysis. There of course may be a good reason to remove data that is censored (for example, if a patient quits a study) but for instances where the censored data still holds valuable information and should be included, I'm not sure what the best method would be to make an assumption about when the end date occurred. Professor Shankar Parajulee, what would be the recommendation here? |
| No problem James! Yes this diagram definitely helped me solidify how all of the survival concepts were related to each other. Glad you found it helpful as well. |