

Data description *Analysis of Survival Data, fall 2025*

Task: Hand in a short description of the data that you intend to use for home assignments 1 and 2 (as a pdf file).

Provide the following information:

- Where you found the data
- Sample size (total no. of observations)
- Degree of censoring
- Which variables you intend to use, what each variable measures, including its unit
- Type of censoring and/or truncation present

The purpose of handing this in is for me to check that you have found data that will function for both assignments. The information will not be graded in any way.

The data you find must contain the following variables for a number of objects (individuals, countries, companies, etc.):

- Time to event. Tip: age can sometimes be used as a time variable, and repeated measurements could be used to find out how long it takes from some start point until a certain event happens.
- Event indicator.
- At least two grouping variables, i.e. variables that can be used to divide the observations into a number of different groups. For one of the grouping variables there must be at least three groups.
- Other potential covariates, i.e. variables that could be associated with the time to event. There should be at least two covariates, in addition to the grouping variables.

If you find data where the objects are defined (countries, municipalities, companies, etc.), you might be able to combine different datasets to include all the requested variables.

Don't know where to look for data? Some examples are provided below.

- Kaggle: <https://www.kaggle.com/datasets>.
- Reliability data for hard drives: <https://www.backblaze.com/blog/hard-drive-data-feb2015>.
- The Failure trace archive: <http://fta.scem.uws.edu.au>.
- Gapminder: www.gapminder.org has data by country, e.g. age at first baby can be matched with other country specific data?

The sites below also have lots of data in general (not specifically checked for survival data).

- gov: <https://catalog.data.gov/dataset>.
- StatLib at Carnegie Mellon: <http://lib.stat.cmu.edu/DASL>.
- World Value Survey: <http://www.worldvaluessurvey.org/wvs.jsp>.
- S. Bureau of Labor Statistics: www.bls.gov, the American Time Use Survey has many interesting aspects
- S. Census Bureau: www.census.gov, see e.g. the American Community Survey
- University of Edinburgh: <http://www.ed.ac.uk/information-services/library-museum-gallery/finding-resources/library-databases/databases-a-z>.
- UCI's Machine Learning Repository: <https://archive.ics.uci.edu/ml>.