Open Data Sets

1. Titanic

- Link: https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/Titanic.html
- Description: This data set provides information on the fate of passengers on the fatal maiden voyage of the ocean liner 'Titanic', summarized according to economic status (class), sex, age and survival.

2. Data sets in package "Survival"

- Link: https://cran.r-project.org/web/packages/survival/survival.pdf
  https://www.rdocumentation.org/packages/survival/versions/2.44-1.1/topics/ovarian
  https://www.rdocumentation.org/packages/survival/versions/2.43-3/topics/bladder
  https://www.rdocumentation.org/packages/survival/versions/2.44-1.1/topics/cgd
  https://www.rdocumentation.org/packages/survival/versions/2.44-1.1/topics/colon
  https://www.rdocumentation.org/packages/survival/versions/2.44-1.1/topics/flchain
  https://www.rdocumentation.org/packages/survival/versions/2.44-1.1/topics/mgus
  https://www.rdocumentation.org/packages/survival/versions/2.44-1.1/topics/pbc
  https://www.rdocumentation.org/packages/survival/versions/2.44-1.1/topics/lung
  https://www.rdocumentation.org/packages/survival/versions/2.44-1.1/topics/heart
  https://www.rdocumentation.org/packages/survival/versions/2.44-1.1/topics/veteran

- Description: In the links.

3. Wcgs

- Link: https://www.rdocumentation.org/packages/epitools/versions/0.5-9/topics/wcgs
- Description: The Western Collaborative Group Study (WCGS), a prospective cohort study, recruited middle-aged men (ages 39 to 59) who were employees of 10 California companies and collected data on 3154 individuals during the years 1960-1961. These subjects were primarily selected to study the relationship between behavior pattern and the risk of coronary hearth disease (CHD). A number of other risk factors were also measured to provide the best possible assessment of the CHD risk associated with behavior type. Additional variables collected include age, height, weight, systolic blood pressure, diastolic blood pressure, cholesterol, smoking, and corneal arcus.

4. Framingham heart study

- Link:
  https://www.rdocumentation.org/packages/LocalControl/versions/1.1.2/topics/framingham
- Description: Data collected over a 24 year study suitable for competing risks survival analysis of hypertension and death as a function of smoking.

5. Addicts

- Link: http://web1.sph.emory.edu/dkleinb/surv3.htm#data
  http://web1.sph.emory.edu/dkleinb/allDatasets/surv2datasets/addicts.dat
- Description: (KK2012 P526) In a 1991 Australian study by Caplehorn et al., two methadone treatment clinics for heroin addicts were compared to assess patient time remaining under methadone treatment. A patient's survival time was determined as the time (in days) until the person dropped out of the clinic or was censored. The two clinics differed according to its live-in policies for patients. The variables are defined as follows:
  ID – Patient ID
  SURVT – The time (in days) until the patient dropped out
  of the clinic or was censored
  STATUS – Indicates whether the patient dropped out of the
  clinic (coded 1) or was censored (coded 0)
  CLINIC – Indicates which methadone treatment clinic the
  patient attended (coded 1 or 2)
  PRISON – Indicates whether the patient had a prison
  record (coded 1) or not (coded 0)
  DOSE – A continuous variable for the patient's maximum
  methadone dose (mg/day)
  https://rdrr.io/github/lbraglia/suanselete3/man/addicts.html

6. Anderson

- Link: http://web1.sph.emory.edu/dkleinb/surv3.htm#data
  http://statweb.stanford.edu/~olshen/hrp262spring01/spring01Assignments/anderson.txt
- Description: Survival times in weeks (in remission) of 42 leukemia patients in clinical trial to compare treatment with placebo. Data from Freireich et al., "The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia," Blook 21, 699-716, 1963.
  Column 1 = survival time (weeks)
  Column 2 = status (0=censored, 1=relapse)
  Column 3 = sex (1=male, 0=female)

Column 4 = log WBC
Column 5 = Rx (1=placebo, 0=treatment)

7. rhDNase

- Link: https://cran.r-project.org/web/packages/tpr/tpr.pdf
  http://www.math.uwaterloo.ca/~rjcook/cook-lawless-recurrent/example2/rhDNase.dat
- Description: (Cook&Lawless2007 P376-P377) Fuchs et al. (1994) report on a double-blind randomized multicenter clinical trial designed to assess the effect of rhDNase, a recombinant deoxyribonuclease I enzyme, versus placebo on the occurrence of respiratory exacerbations among patients with cystic fibrosis. The rhDNase operates by digesting the extracellular DNA released by leukocytes that accumulate in the lung as a result of bacterial infection (Therneau and Hamilton, 1997), and so it was expected that aerosol administration of rhDNase would reduce the incidence of exacerbations. Data on the occurrence and resolution of all exacerbations were recorded over approximately 169 days of followup for 645 patients in this trial; the data are discussed in some detail in Therneau and Grambsch (2000). Part of the data is given in Table D.2 for the first 20 patients. We include a patient identifier, the treatment assignment (T) (1 = rhDNase, 0 = placebo), two baseline measurements of forced expiratory volume (FEV1 and FEV2) reflecting lung capacity, and the date of randomization. In addition, the number of days from randomization to the beginning (B) of the exacerbations is recorded, as well as the day on which treatment for each exacerbation ended (E) and patients became at risk of a new exacerbation. Therefore, for patient number 589302, the first exacerbation began 8 days after randomization and antibiotic therapy for this exacerbation ended 22 days after randomization. The patient then remained at risk until the second exacerbation, which began 63 days after randomization, and became at risk again after therapy ended on day 88; the patient did not have another exacerbation over the remainder of followup which ended on day 169.

8. ACTG175

- Link:
  https://www.rdocumentation.org/packages/speff2trial/versions/1.0.4/topics/ACTG175
- Description: ACTG 175 was a randomized clinical trial to compare monotherapy with zidovudine or didanosine with combination therapy with zidovudine and didanosine or zidovudine and zalcitabine in adults infected with the human immunodeficiency virus type I whose CD4 T cell counts were between 200 and 500 per cubic millimeter.

9. Abortion

- Link: https://www.rdocumentation.org/packages/etm/versions/1.0.4/topics/abortion
- Description: Outcomes of pregnancies exposed to coumarin derivatives. The aim is to

investigate whether exposition to coumarin derivatives increases the probability of spontaneous abortions. Apart from spontaneous abortion, pregnancy may end in induced abortion or live birth, leading to a competing risks situation. Moreover, data are left-truncated as women usually enter the study several weeks after conception.

10.sir.adm

- Link: https://www.rdocumentation.org/packages/mvna/versions/2.0.1/topics/sir.adm
- Description: Pneumonia status on admission for intensive care unit (ICU) patients, a random sample from the SIR-3 study.

11.sir.cont

- Link: https://www.rdocumentation.org/packages/mvna/versions/2.0.1/topics/sir.cont
- Description: Time-dependent ventilation status for intensive care unit (ICU) patients, a random sample from the SIR-3 study.

12.icu.pneu

- Link: https://www.rdocumentation.org/packages/kmi/versions/0.5.4/topics/icu.pneu
- Description: This data set is a random sample drawn from the SIR-3 study that aimed at analysing the effect of nosocomial infections on the length of ICU stay. Patients were included in the study if they had stayed at least 1 day in the unit. The sample includes information to assess the effect of nosocomial pneumonia on the length of stay. The endpoint is either discharge alive from the ICU or dead in the unit. These data are censoring complete as the censoring time is known for all patients.
- PS：Package 'kmi' was removed from the CRAN repository. Formerly available versions can be obtained from https://cran.r-project.org/src/contrib/Archive/kmi/.