Biostatistics

Task2: Survival data

1. Consider a survival function with constant hazard $h(t) = 0.07$ when $0 \leq t \leq 5$, and $h(t) = 0.4$ for $t > 5$ (this is known as a piecewise constant hazard).

Plot this hazard function and the corresponding survival function for $0 < t < 10$.

What is the **median** survival time?

2. Suppose that we assume that the time-to-event is a *Rayleigh* distribution with density function:

$$f(y) = (\lambda_0 + \lambda_1 y) \exp\left[-\lambda_0 y - \frac{1}{2}\lambda_1 y^2\right]$$

where $y > 0$

Calculate the survival, hazard, and cumulative hazard functions.

3. The file `Henning.txt` contains data from a study of criminal recidivism by Henning and Frueh (1996), who followed 194 inmates released from a medium-security prison for a maximum of three years from the day of their release; during the period of the study, 106 of the released prisoners were rearrested.

The data set contains the following variables:

- `months`: The time of re-arrest in months (but measured to the nearest day).
- `censor`: A dummy variable coded 1 for censored observations and 0 for uncensored observations. Note that this is the **opposite** of our usual convention.
- `personal`: A dummy variable coded 1 for prisoners with a record of crime against persons and 0 otherwise.
- `property`: A dummy variable coded 1 for prisoners with a record of crime against property and otherwise.
- `cage`: Centered age in years at the time of release, that is age-average age.

- Compute and plot the Kaplan-Meier estimate of the survival function for all of the data.
- Compute and plot separate survival curves for those with and without a record of crime *against persons*; test for differences between the two survival functions.
- Compute and plot separate survival curves for those with and without a record of crime *against property*; test for differences between the two survival functions.

- Fit a Cox regression of time to re-arrest on the covariates `personal`, `property`, and `cage`.

  (i) Determine by a Wald test whether each estimated coefficient is statistically significant.

  (ii) Interpret each of the estimated Cox-regression coefficient.

4. Given a hazard function $h(t) = c$, where $c > 0$, derive the survival and the density function. Calculate the **median** failure time for $c = 5$.

5. Consider and explain briefly the case of **recurrent events**, and show an example by using R.

See, for example as references, these websites:

`https://www.sychiou.com/files/course.pdf`

`https://cran.r-project.org/web/packages/mets/vignettes/recurrent-events.html`

`https://www.jstatsoft.org/article/download/v105i05/4427`

6. Develop a brief example of Survival Analysis from a **Bayesian** point of view using the `rstanarm` and `dynsurv` packages.

See, for example as references, these websites:

`https://rpubs.com/kaz_yos/rstanarm_survival1`

`https://arxiv.org/pdf/2002.09633`

`https://zenodo.org/record/3866164/files/survmodelsrstanarm_eren.pdf`

`https://cran.r-project.org/web/packages/dynsurv/index.html`