

Modelos de tiempo de fallo acelerado para tiempos de sobrevivencia de pacientes con cáncer de laringe

La base de datos `larynx` del paquete `KMsurv` contiene información de pacientes que fueron diagnosticados con cáncer de laringe. La información es la siguiente:

- `stage` Etapa de la enfermedad de 1 a 4
- `time` tiempo (en meses) desde el primer tratamiento hasta la muerte o el final del estudio
- `age` edad (en años) en que se diagnosticó que el paciente presentaba cáncer de laringe
- `diagyr` año del diagnóstico de cáncer de laringe
- `delta` Indicador (1: si el paciente murió; 0: otro caso)

- a) Como usted sabe, los modelos de tiempo de fallo acelerado (AFT) consisten en modelar los tiempos de sobrevivencia T en escala logarítmica a través de una combinación lineal de las covariables \mathbf{x} . Es decir, sea $Y = \log(T)$ entonces un modelo AFT se escribe de la siguiente forma:

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i$$

Donde ϵ corresponde a los errores. Comente por qué utilizar un modelo AFT trae consigo una ventaja al momento de plantear modelos de regresión para los tiempos de sobrevivencia.

- b) Dependiendo de la distribución asumida para ϵ podemos plantear distintos modelos. Por ejemplo, asumir que ϵ_i distribuye independiente e idénticamente $N(0, \sigma^2)$ implica que Y_i distribuya Normal, por lo cual T distribuye lognormal. Plantee el modelo AFT asumiendo errores normales considerando todas las observaciones completamente observadas (ninguna censurada). Compruebe que este modelo coincide con un modelo lineal utilizando Y_i como variable respuesta. ¿Qué implica esto?
- c) Plantee el modelo AFT considerando la censura. La expresión para cada tiempo en escala logarítmica es la siguiente:

$$\log(T_i) = \beta_1 + \beta_2 \mathbb{1}_{stage=2} + \beta_3 \mathbb{1}_{stage=3} + \beta_4 \mathbb{1}_{stage=4} + \beta_5 age + \beta_6 diagyr + \epsilon_i$$

Obtenga los coeficientes en R y plantee las relaciones de los tiempos de sobrevivencia de las etapas de la enfermedad 2, 3 y 4 respecto a la etapa 1. Comente.

- d) Plantee un modelo utilizando como distribución de $\log(Y)$ de valores extremos, ¿qué distribuciones podría asumir para T ? Obtenga los coeficientes y compare con los coeficientes del modelo anterior.