

MTH6157 Survival Models

Questions – Week 4 – Covariates and Proportional Hazard Models

1. The Health & Safety Executive wish to investigate serious accidents and deaths at large construction projects in London. For all construction workers present at building sites on 1st November 2022 they collect the following data:

Data item	Data range collected
Length of experience in industry	Number of Months
Electricians' qualification	Yes / No
Colour of hard hat	White / Yellow / Other
Hours worked on site in last week	Number of hours
Witnessed accident in the last year	None / One / More than 1

- (a) Define the term *covariate*
- (b) List some weaknesses of the H&SE approach to data collection here
- (c) For each of the data items above suggest how they could be converted into a covariate matrix
- (d) What are the advantages and disadvantages of using a scoring system rather than using actual recorded numerical values?
2. A new treatment for pneumonia is being tested in two hospitals, one in Rome and one in Madrid. The team coordinating the trial use a Cox Proportional Hazard model to analyse the results.

The model fitted is $\lambda_i(t) = \lambda_0(t) \exp(\beta \cdot z^T)$ where

$\lambda_i(t)$ is the hazard at time t since administering the new treatment

$\lambda_0(t)$ is the baseline hazard

z is a vector of covariates where z_1 is 1 for Italian patients and 0 for those in Spain; z_2 is the period from diagnosis to treatment in weeks; and z_3 is 1 if the patient is age 55 or over and 0 otherwise.

β is a vector of parameters where $\beta_1 = 0.03$; $\beta_2 = 0.4$; $\beta_3 = -0.3$

- (i) To whom does the baseline hazard apply?

- (ii) For a Spaniard aged 56 who was given the new treatment 2 weeks after diagnosis,
 - a. Write down the hazard function in terms of $\lambda_0(t)$
 - b. Express the survival function in terms of $\lambda_0(t)$
- (iii) For 65 year-old Italian given the new treatment 1 week after diagnosis, the probability of survival for 3 months is 0.95. Calculate the probability that the Spaniard in (ii) above will survive 3 months.

3. Two Proportional Hazard models are found to model the survival function for patients taking a certain cancer drug. One uses age, time since diagnosis and body mass index data whereas the second model adds data on which hospital administered the drug, blood pressure and smoker status.
- (i) What is a likelihood ratio statistic and how might it be used in this case?
 - (ii) What is the distribution of the statistic in this test?