MÉTODOS DE REGRESIÓN EN BIOESTADÍSTICA

Fernando A Barrios, PhD
Universidad Nacional Autónoma de México
Instituto de Neurobiología

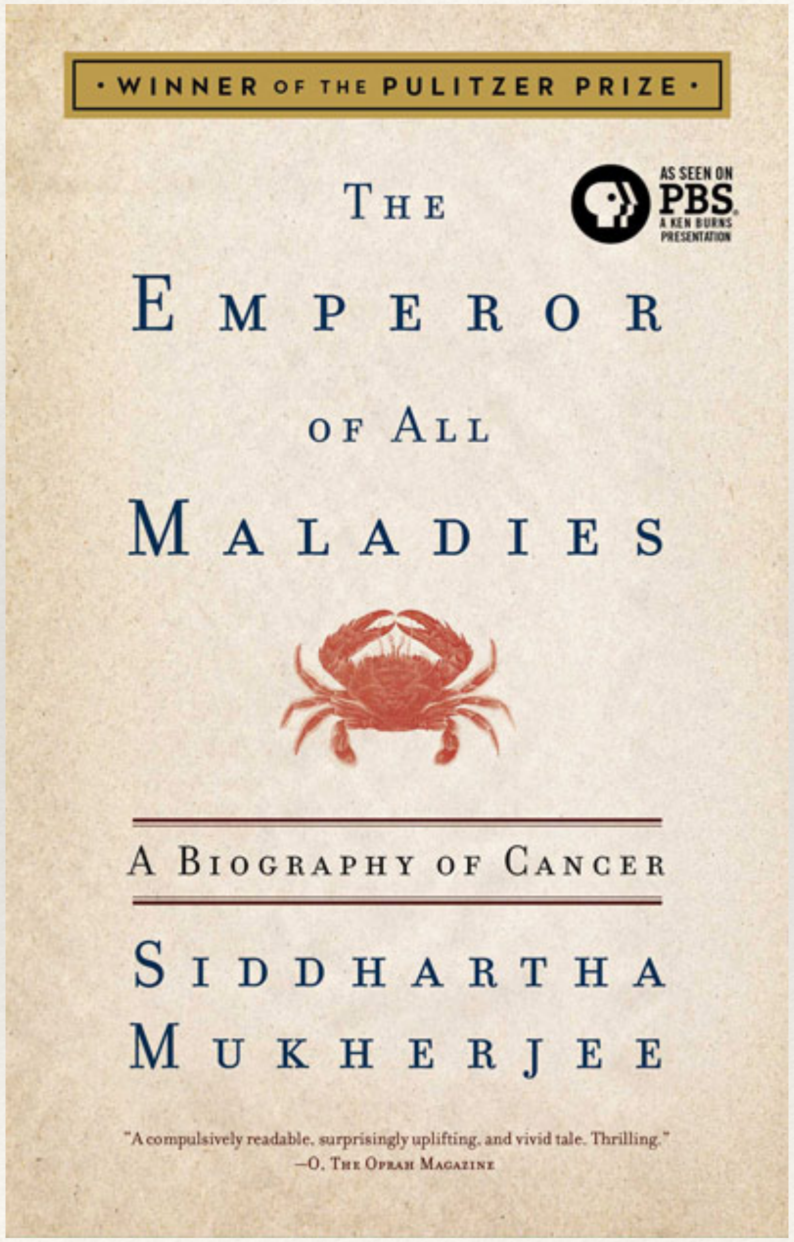
Modelos básicos de análisis de supervivencia

Survival analysis

Right censoring example “acute lymphoblastic leukemia (ALL)”, 6-mercaptopurine (6-MP)

42 patients achieved remission from induction therapy and then were randomized in equal numbers to 6-MP or placebo.

Siddhartha Mukherjee



http://siddharthamukherjee.com/the-emperor-of-all-maladies/

Survival Analysis

Freireich, ALL.

Some 6-MP patients will not have the exact time to relapse, 12 remained in remission at the time of the study, the exact time of relapse was unobserved. One patient was in observation only for 6 weeks, this survival time is said to be right-censored.

Definition: A survival time is said to be *right-censored* at time t if it only known to be grater that t.

Kaplan-Meier Estimator

The description of the probability of the remaining remission during each of the first ten weeks of ALL is called the survival function

Definition: The **survival function at time *t***, denoted $S(t)$, is the probability of being event free at t; equivalently, the probability that the survival time is grater the t.

Kaplan-Meier Estimator

Table 3.13 Follow-up table for placebo patients in the leukemia study

Week of follow-up	No. followed	No. relapsed	No. censored	Conditional prob. of remission	Survival function
1	21	2	0	$19/21 = 0.91$	0.91
2	19	2	0	$17/19 = 0.90$	$0.90 \times 0.91 = 0.81$
3	17	1	0	$16/17 = 0.94$	$0.94 \times 0.81 = 0.76$
4	16	2	0	$14/16 = 0.88$	$0.88 \times 0.76 = 0.67$
5	14	2	0	$12/14 = 0.86$	$0.86 \times 0.67 = 0.57$
6	12	0	0	$12/12 = 1.00$	$1.00 \times 0.57 = 0.57$
7	12	0	0	$12/12 = 1.00$	$1.00 \times 0.57 = 0.57$
8	12	4	0	$8/12 = 0.67$	$0.67 \times 0.57 = 0.38$
9	8	0	0	$8/8 = 1.00$	$1.00 \times 0.38 = 0.38$
10	8	0	0	$8/8 = 1.00$	$1.00 \times 0.38 = 0.38$

Kaplan-Meier Estimator

Right - censored

Table 3.14 Follow-up table for 6-MP patients in the leukemia study

Week of follow-up	No. followed	No. relapsed	No. censored	Condition. prob. of remission	Survival function
1	21	0	0	$21/21 = 1.00$	1.00
2	21	0	0	$21/21 = 1.00$	$1.00 \times 1.00 = 1.00$
3	21	0	0	$21/21 = 1.00$	$1.00 \times 1.00 = 1.00$
4	21	0	0	$21/21 = 1.00$	$1.00 \times 1.00 = 1.00$
5	21	0	0	$21/21 = 1.00$	$1.00 \times 1.00 = 1.00$
6	21	3	1	$18/21 = 0.86$	$0.86 \times 1.00 = 0.86$
7	17	1	0	$16/17 = 0.94$	$0.94 \times 0.86 = 0.81$
8	16	0	0	$16/16 = 1.00$	$1.00 \times 0.81 = 0.81$
9	16	0	0	$16/16 = 1.00$	$1.00 \times 0.81 = 0.81$
10	16	0	1	$16/16 = 1.00$	$1.00 \times 0.81 = 0.81$

Interpretation of Kaplan-Meier Curves

To include the extrapolation for the revival experience of censored observations the las table requires to be modified (case of competing risk data, where cumulative incidence functions).

$$S(t) = P(T \leq t)$$

$$\hat{S}(t) = \frac{\text{number individuals with survival times} \geq t}{n}$$

$$\text{variance estimate} = \frac{\hat{S}(t)(1 - \hat{S}(t))}{n}$$

Survivor function

The Kaplan-Meier estimarte of the survival function os obtained as

$$\hat{S}(t) = \prod_{j:t_{(t)} \leq t} (1 - \frac{d_j}{r_j})$$

Where r_j is the number of individuals at risk just before $t_{(j)}$ (including those censored at $t_{(j)}$) d_j is the number of individuals who experience the event of interest at time $t_{(j)}$

Survivor function

The estimated variance of the Kaplan-Meier estimarte of the survival function os obtained from

$$Var(\hat{S}(t)) = (\hat{S}(t))^2 \sum_{j:t_{(t)} \leq t} \frac{d_j}{r_j(r_j - d_j)}$$

Where r_j is the number of individuals at risk just before $t_{(j)}$ (including those censored at $t_{(j)}$) d_j is the number of individuals who experience the event of interest at time $t_{(j)}$

A formal test of the equality of the survival curves for two groups can be made using the ***log-rank test***.

Hazard function

The hazard function, *h(t),* defined defined as the probability that an individual experiences the event in a small time interval, given that the individual has survived up to the up to the beginning of the time interval. The hazard function and survivor function are related by the formula

$$S(t) = \exp(-H(t))$$

Where *H(t)* is known as the *integrated hazard* or *cumulative hazard*, defined as follows:

$$H(t) = \int_0^t h(u)du$$

Hazard function

The hazard function can be estimated as the proportion of individuals experiencing the event of interest in an interval per unit time, given that they have survived to the beginning of the interval

$$\hat{h}(t) = \frac{d_j}{n_j(t_{(j+1)} - t_j)}$$

Usually integrated hazard is used and can be estimated as follows:

$$\hat{H}(t) = \sum_j \frac{d_j}{n_j}$$

Cox's Regression

The most widely used regression technique for modeling of a possibly censored survival time is the *Cox's proportional hazards model*, or *Cox's regression*. Models the hazard function since it does not involve the cumulative history of events. Since *h(t)* is restricted positive, a more suitable model is

$$\log(h(t)) = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$$

This is a restrictive model since only accepts a constant function, so it is potable to propose (as Cox) an open general initial time dependence

$$\log(h(t)) = \log(h_0(t)) + \beta_1 x_1 + \ldots + \beta_n x_n$$

Where $h_0(t)$ is known as the baseline hazard function.

Interpretation of Kaplan-Meier Curves

Example in HSAUR3 Glioma Radioimmunotherapy

The Colon data comes from one of the first successful trials of adjuvant chemotherapy for colon cancer. Levamisole is a low-toxicity compound previously used to treat worm infestations in animals; 5-FU is a moderately toxic chemotherapy agent.
There are two records per person, one for recurrence and one for death estimarte of the survival function as obtained.