Intro and EDA

Xueqing Huang(xh2470)

Introduction

We are going to conduct a survival analysis by using lung cancer dataset in 'survival' package in R. The data describes survival of patients with advanced lung cancer from the North Central Cancer Treatment Group, as well as measures of the patients performance assessed either by the physician and by the patients themselves. This project aims to find if the nutritional factors such as caloric intake, will bring significant differences to the survival rate among patients with advanced lung cancer. The association between both the physician's assessments of performance status (PS) as well as the patient's assessment of their own performance status and the survival rate are also evaluated.

Exploratory Data Analysis(EDA)

The dataset contains a total of 228 patients and 10 variables. A brief description of variables in the dataset is shown below.

- inst: Institution code
- time: Survival time in days
- status: Censoring status(1=censored, 2=dead)
- age: Age in years
- sex: Male=1, Female=2
- ph.ecog: ECOG performance score (0=good 5=dead)
- ph.karno: Karnofsky performance score (from bad=0 to good=100) rated by physician
- pat.karno: Karnofsky performance score as rated by patient
- meal.cal: Calories consumed at meals
- wt.loss: Weight loss in last six months

Survival endpoint is the death of patients. The type of censoring is right censoring, which means patients left the study before their death. Among 228 patients, 63 of them are right censored and the number of events is 165. We group the patients by their survival status and provide the descriptive statistics of other variables. From the table, we can see that average survival time for censored and dead patients is 363 days and 283 days, respectively. Generally speaking, patients who alive were younger than patients who died, and female had higher chance of survival than female.

	Alive	Death	Total
	(N=63)	(N=165)	(N=228)
Survival Time (days)			
Mean (SD)	363 (221)	283 (203)	305 (211)
Median [Min, Max]	284 [92.0, 1020]	226 [5.00, 883]	256 [5.00, 1020]
Age			
Mean (SD)	60.3 (9.74)	63.3 (8.69)	62.4 (9.07)
Median [Min, Max]	62.0 [39.0, 77.0]	64.0 [40.0, 82.0]	63.0 [39.0, 82.0]
Sex			
Male	26 (41.3%)	112 (67.9%)	138 (60.5%)
Female	37 (58.7%)	53 (32.1%)	90 (39.5%)
ECOG Performance Score			
Asymptomatic	26 (41.3%)	37 (22.4%)	63 (27.6%)
Symptomatic but completely ambulatory	31 (49.2%)	82 (49.7%)	113 (49.6%)
In bed <50% of the day	6 (9.5%)	44 (26.7%)	50 (21.9%)
In bed > 50% of the day but not bedbound	0 (0%)	1 (0.6%)	1 (0.4%)
Bedbound	0 (0%)	0 (0%)	0 (0%)
Missing	0 (0%)	1 (0.6%)	1 (0.4%)
Karnofsky Performance Score(by physician)			
50	1 (1.6%)	5 (3.0%)	6 (2.6%)
60	3 (4.8%)	16 (9.7%)	19 (8.3%)
70	3 (4.8%)	29 (17.6%)	32 (14.0%)
80	20 (31.7%)	47 (28.5%)	67 (29.4%)
90	25 (39.7%)	49 (29.7%)	74 (32.5%)
100	11 (17.5%)	18 (10.9%)	29 (12.7%)
Missing	0 (0%)	1 (0.6%)	1 (0.4%)
Karnofsky Performance Score(by patients)			
30	1 (1.6%)	1 (0.6%)	2 (0.9%)
40	1 (1.6%)	1 (0.6%)	2 (0.9%)
50	0 (0%)	4 (2.4%)	4 (1.8%)
60	3 (4.8%)	27 (16.4%)	30 (13.2%)
70	10 (15.9%)	31 (18.8%)	41 (18.0%)
80	12 (19.0%)	39 (23.6%)	51 (22.4%)
90	22 (34.9%)	38 (23.0%)	60 (26.3%)
100	14 (22.2%)	21 (12.7%)	35 (15.4%)
Missing	0 (0%)	3 (1.8%)	3 (1.3%)
Calories Consumed (kcals)			
Mean (SD)	913 (453)	934 (384)	929 (402)
Median [Min, Max]	975 [96.0, 2450]	1030 [169, 2600]	975 [96.0, 2600]
Missing	16 (25.4%)	31 (18.8%)	47 (20.6%)
Weight Loss (pounds)			
Mean (SD)	9.11 (12.9)	10.1 (13.2)	9.83 (13.1)
Median [Min, Max]	4.00 [-10.0, 49.0]	8.00 [-24.0, 68.0]	7.00 [-24.0, 68.0]
Missing	1 (1.6%)	13 (7.9%)	14 (6.1%)

Reference

Loprinzi CL. Laurie JA. Wieand HS. Krook JE. Novotny PJ. Kugler JW. Bartel J. Law M. Bateman M. Klatt NE. et al. Prospective evaluation of prognostic variables from patient-completed questionnaires. North Central Cancer Treatment Group. *Journal of Clinical Oncology.* 12(3):601-7, 1994.

Things need to be discussed.

Karnofsky Score - catogorical variables, can be classify into 3 levels to make table shorter.

0-40: unable to care for self

50-70: unable to work

80-100: able to carry on normal activity and to work.