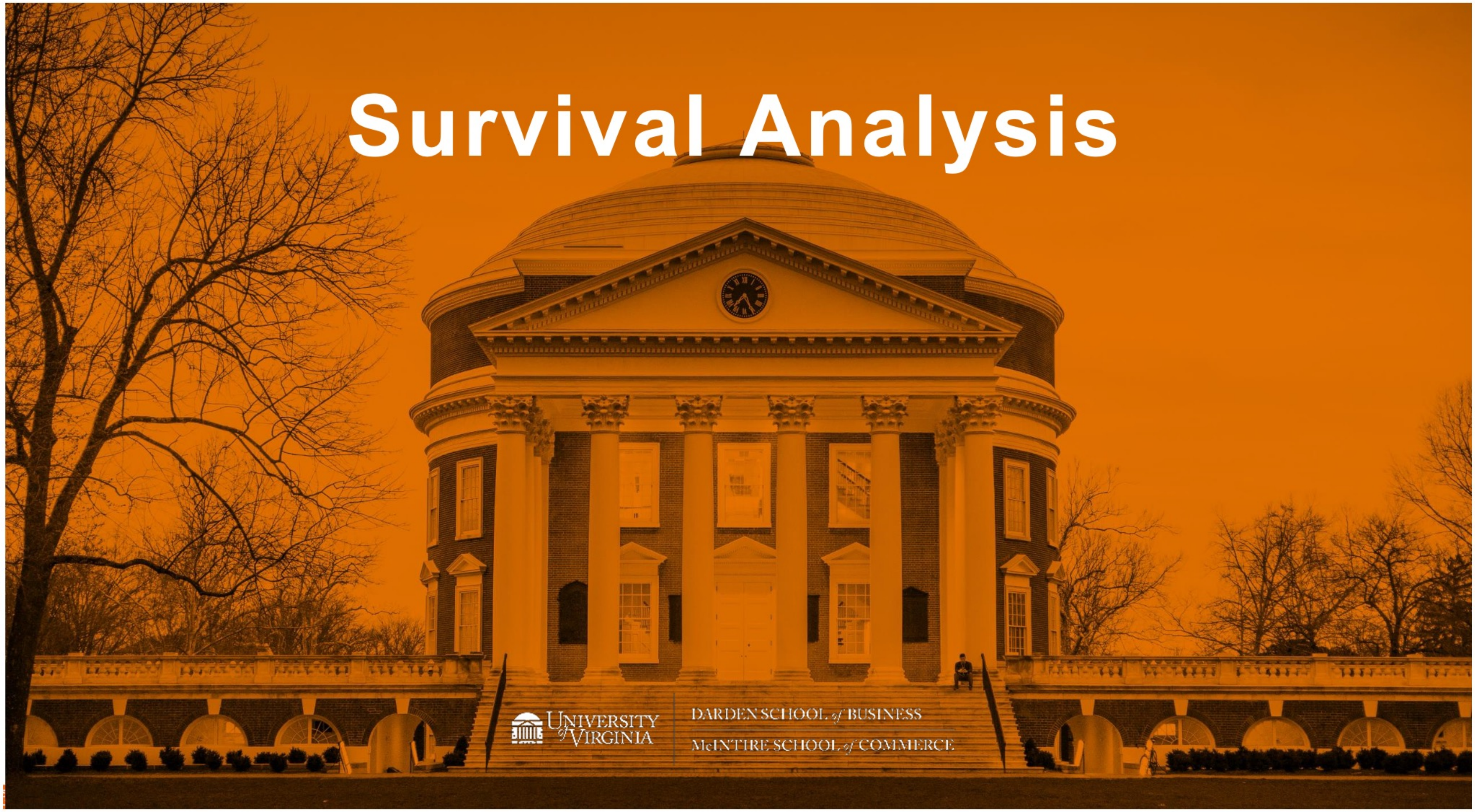
Survival Analysis



Survival Analysis Terminology

- **Event: The event/ experience of interest**
  - Examples:
    - Cardiovascular death after some treatment intervention
    - Machine part failure
    - Cancelation of a subscription
    - Departure of an employee

- **Survival Time: The duration of time UNTIL the event of interest occurs**
  - Examples:
    - Time until cardiovascular death after some treatment intervention
    - Time until a machine part fails
    - Time until a customer cancels subscription
    - Time until an employee leaves the company

Survival Analysis Terminology

- **Censoring**
  - Given that in many cases we are not able to observe the subjects (patients, customers, …) during their entire lifespan, the data in survival analysis is often "censored". That is, for some subjects we may not know the exact time of treatment, for some of them we may not know the exact time of event, and for some subjects we may know neither. In survival analysis, we learn how to deal with situations like these.

- *Fixed type I censoring* occurs when a study is designed to end after C years of follow-up. In this case, everyone who does not have an event observed during the course of the study is censored at C years.
- In *random type I censoring*, the study is designed to end after C years, but censored subjects do not all have the same censoring time. This is the main type of right-censoring we will be concerned with.
- In *type II censoring*, a study ends when there is a pre-specified number of events.

Survival Analysis Terminology

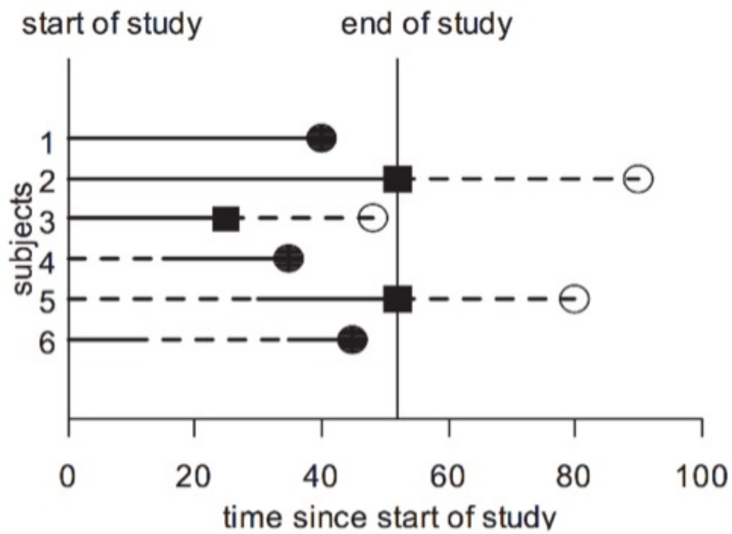


Figure 1. Data from an imagined study illustrating various kinds of subject histories: Subject 1, uncensored; 2, fixed-right censoring; 3, random-right censoring; 4 and 5, late entry; 6, multiple intervals of observation.

Quiz

- **If none of the data points were censored, could we use linear regression instead of survival analysis?**

  - Time to event is restricted to be positive and has a skewed distribution.
  - The probability of surviving past a certain point in time may be of more interest than the expected time of event.
  - The hazard function, used for regression in survival analysis, can lend more insight into the failure mechanism than linear regression.

Survival Data

- 1- The survival time of each subject, or the time at which the observation for the subject is censored.
- 2- Whether the subject's survival time is censored.
- 3- In most interesting analyses, the values of one or more explanatory variables (covariates) thought to influence survival time. The values of (some) covariates may vary with time.

	id	age	gender	hr	sysbp	diasbp	bmi	cvd	afb	sho	...	lenfol	fstat
0	1	83	0	89	152	78	25.54051	1	1	0	...	2178	0
1	2	49	0	84	120	60	24.02398	1	0	0	...	2172	0
2	3	70	1	83	147	88	22.14290	0	0	0	...	2190	0

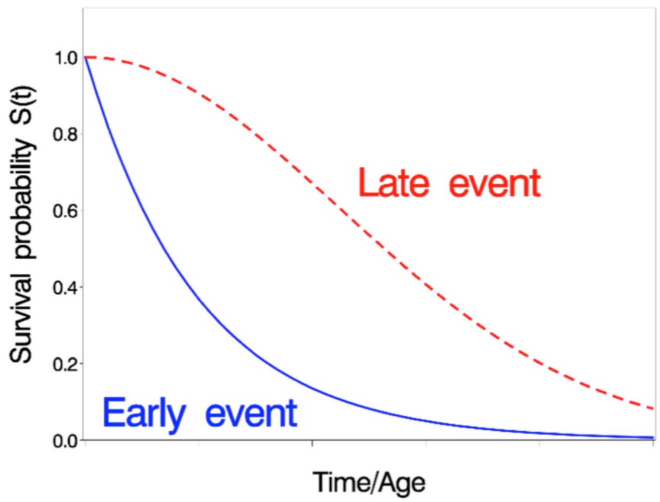
Survival Function

$$S(t) = P(\textbf{Outcome} > t)$$

"The probability of surviving time point t"

Properties of S(t):

- $S(0) = 1$: There is absolute certainty to 'survive' $t = 0$
- $S(+\infty) = 0$: There is absolute certainty to 'fail' eventually
- $S(t)$ is a decreasing function



Kaplan-Meier Survival Curve

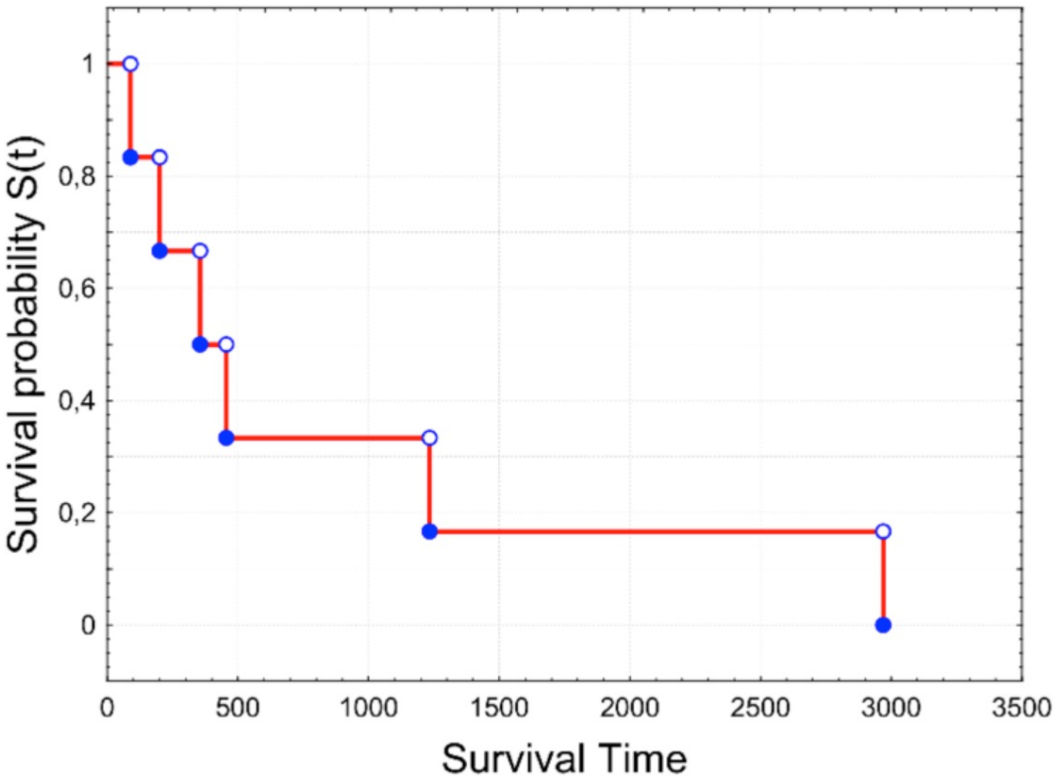
$$S(t) = P(\text{Outcome} > t) \quad \longrightarrow \quad \widehat{S}(t) = \frac{\#\text{ subjects surviving } t}{N}$$

$$S_t = \frac{\begin{array}{c}\text{Number of subjects}\\ \text{living at the start}\end{array} - \begin{array}{c}\text{Number of subjects}\\ \text{died}\end{array}}{\text{Number of subjects living at the start}}$$

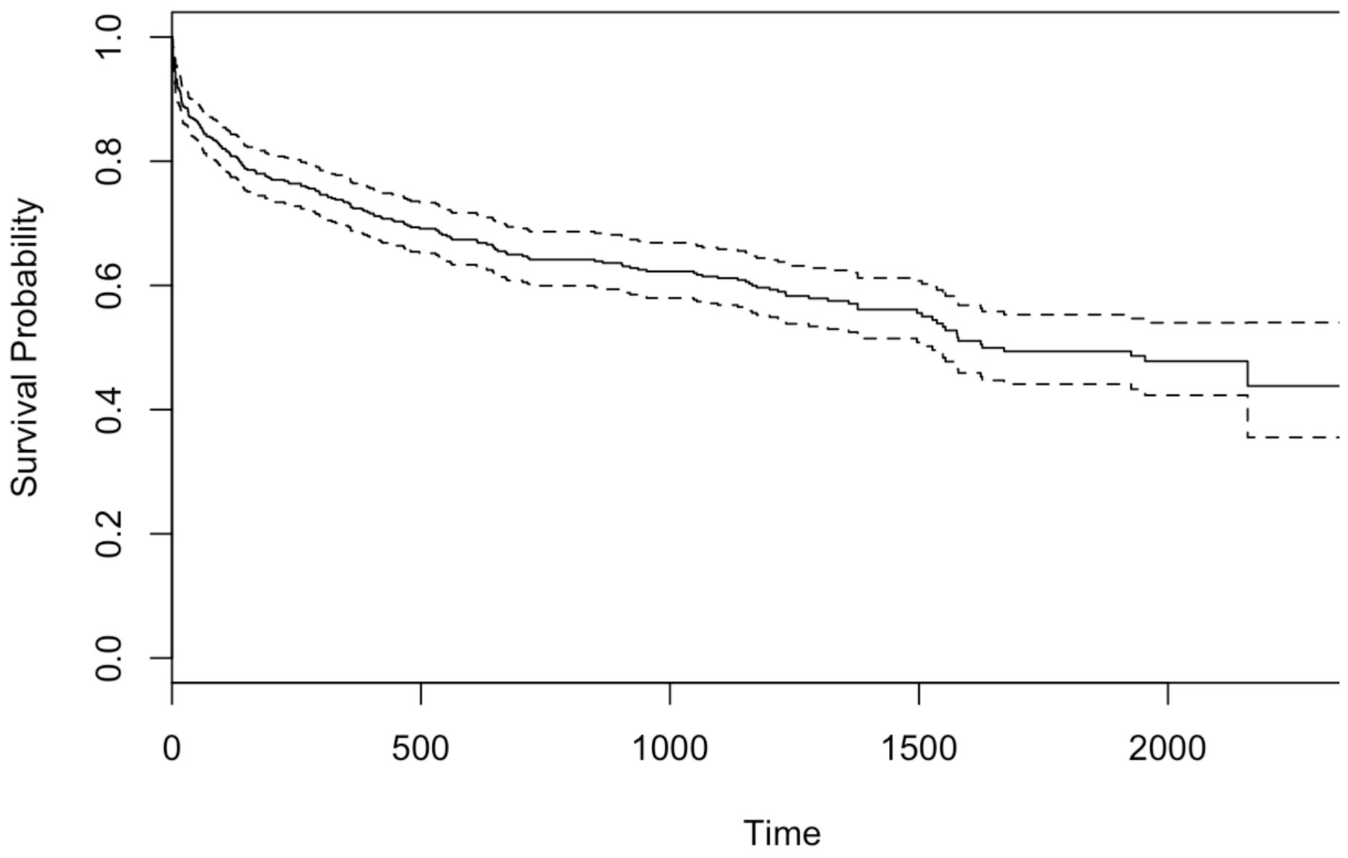
Illustrative Example

Time (t)	# Surving t	$\widehat{S}(t)$
0	6	$6/6 = 1.00$
30	6	$6/6 = 1.00$
89	5	$5/6 = 0.83$
100	5	$5/6 = 0.83$
201	4	$4/6 = 0.67$
356	3	$3/6 = 0.50$
400	3	$3/6 = 0.50$
556	2	$2/6 = 0.33$
1234	1	$1/6 = 0.17$
2970	0	$0/6 = 0.00$

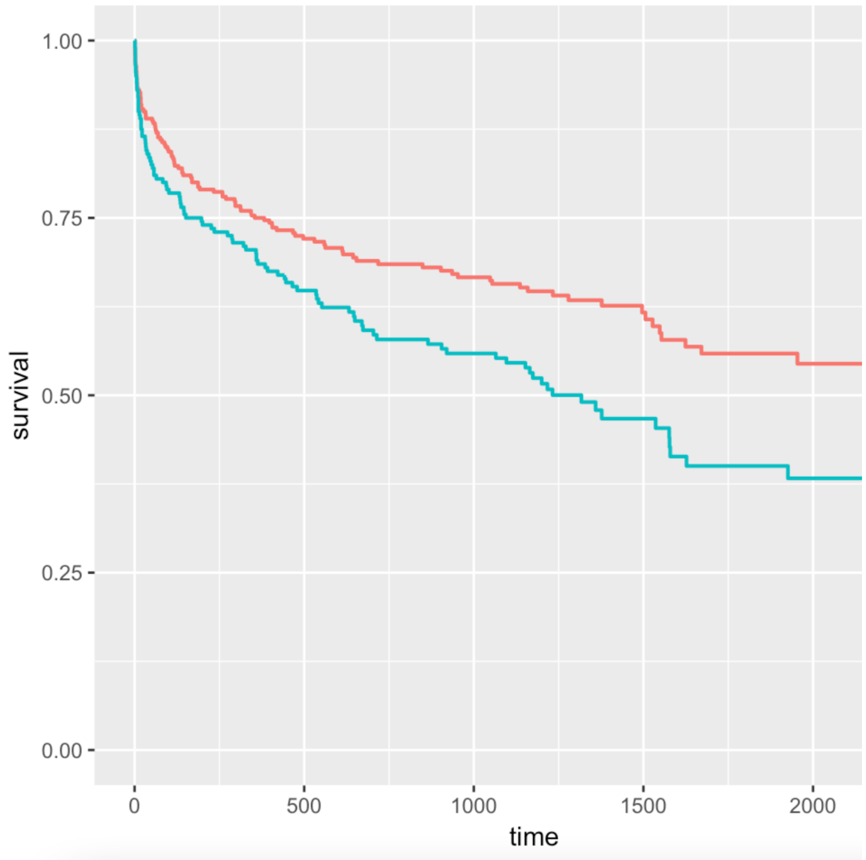
Illustrative Example



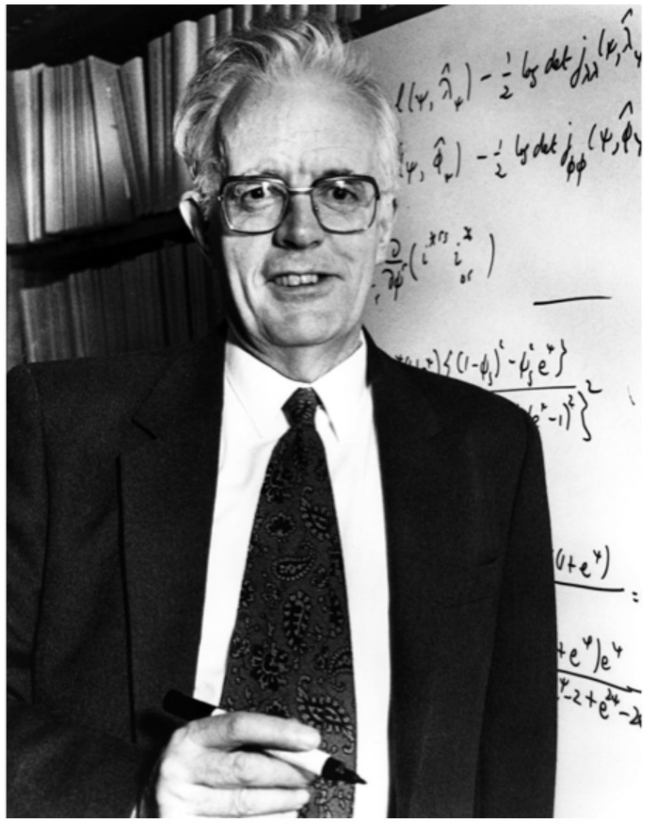
Illustrative Example



Illustrative Example



David Cox



Cox Proportional Hazard Model

$$log(h_i(t)) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$$

$$h_i(t) = exp(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})$$

Cox Proportional Hazard Model

$$h_i(t) = exp(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})$$

$$h_i(t) = h_0(t)exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})$$

$$\eta_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$$

$$\eta_{i'} = \beta_1 x_{i'1} + \beta_2 x_{i'2} + \cdots + \beta_k x_{i'k}$$

$$\frac{h_i(t)}{h_{i'}(t)} = \frac{h_0(t)e^{\eta_i}}{h_0(t)e^{\eta_{i'}}} = \frac{e^{\eta_i}}{e^{\eta_{i'}}}$$

The Output of Cox Model

```
coxph(formula = Surv(lenfol, fstat) ~ gender + cvd, method = "breslow")

  n= 500, number of events= 215

         coef exp(coef) se(coef)     z Pr(>|z|)
gender 0.3571    1.4291   0.1386 2.577  0.00997 **
cvd    0.2328    1.2621   0.1692 1.376  0.16892
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The Output of Cox Model

```
Concordance= 0.553  (se = 0.02 )
Likelihood ratio test= 9.56  on 2 df,   p=0.008
Wald test            = 9.51  on 2 df,   p=0.009
Score (logrank) test = 9.62  on 2 df,   p=0.008
```