Pontificia Universidad
JAVERIANA
Bogotá

Introduction to cox regression

Notes of class

Iván Andrés Trujillo Abella

FACULTAD DE INGENIERÍA

0.1 Exponential Distribution

Exponential distribution dont have memory equall to the geometric distribution. with one parameter λ the time between until a poisson process occur. Assume that $y \sim G(p)$ then

$$P(Y \leq z)$$

The exponential distribution have the following Probability Distribution Function:

$$f(x) = \lambda e^{-\lambda x}$$

Therefore we said, $X \sim exp(\lambda)$ that x follow a exponential distribution with one only parameter in this case lambda.

https://www.youtube.com/watch?v=yldSqu3WArw

Notes exponential distribution it is related with **gamma distribution**.

0.1.1 Survivor and hazard function

we have a variable y that means time of survivor $f(y)$ it is its PDF. the cdf will be $F(y) = \int_0^y F(t)dt$ and survivor function is defined by

$$S(y) = P(Y > y) = 1 - F(y)$$

that is the probability of survive beyond T.

in this order of ideas the *Hazard function* is defined as

$$h(y) = \frac{f(y)}{S(Y)}$$

then this will be undesertand as a risk. but

$$f(y) = \lim_{\Delta \to 0} \frac{F(y + \Delta y) - F(y)}{(\Delta y)}$$

$$\frac{f(y)}{S(y)} = \frac{\lim_{\Delta \to 0} \frac{F(y+\Delta y) - F(y)}{(\Delta y)}}{1 - F(y)}$$

if we translate the above expression in probability terms we have *the probality of failure in a small time period of change* due the survial period.

$$h(y) = -\frac{d}{dy} ln[1 - F(y)] \tag{1}$$

$$\int_0^y h(y) = -ln(S(Y)) \tag{2}$$

$$S(y) = e^{\int_0^y h(t)dt} \tag{3}$$

0.1.2 Hazard in exponential

we need get the **CDF** this is

$$f(x) = \lambda e^{-\lambda x}$$
$$\int_0^x f(t)dt = 1 - e^{\lambda x}$$
$$S(x) = e^{-\lambda x}$$
$$h(x) = \lambda$$

thus the hazard it is a contant when failure time it is exponential.

0.2 Normal distribution

0.3 Weibull distribution

1 Cox hazard

the main parameter is *hazard rate* the expected number of event by period time.

hazard ratio could be compated with odds ratio.

observed to expected this is very important this concept due it is have a notion about ramdon.

$_0(t) riskcanchangeovertimeaccordingtocovariates.theproportionalhazardconditionestablishthatcovariatesaremultiplic$

the measure of effect is the hazard rate, not is probability (the expected number of events per unit of time). Sometimes we are interested, in compare rates among groups.

the model has the following charace

$$\lambda(t|z) = \lambda_0(t) e^{\sum_{i=1}^{n} \beta_i x_i}$$

in this case there are not β_0 due if $z = 0$ then $\lambda_0(t)$ it is the baseline hazard.

Cox it is semiparametric model due Z, that it is a vector of covariates.

it is called proportional due hazard ratio is constant.

https://www4.stat.ncsu.edu/ dzhang2/st745/chap6.pdf This is a good book to understand this distribution probability.

We go to uses a hazard ratio to model bankrupcity..

```python
import pandas
```