# Survival Analysis

def: Models remaining time before an event happens.

1. GOALS
2. DG: Data generative Process => assumption , Y distribution (resp var)
3. Estimation: spread and interval estimation
4. Reporting: look at results and have some conclusions

## Logit

Likelihood function ???

1. Prepare input data as dataframe
2. Fitting a logistic regression model. Glm

30 May 2023

Effect plots

Forest Plot

Method

1. DGP
2. ESTIMATION / FIT
3. HYP testing, CI
4. Interpretation
5. Model Based predictions

### Keywords

Survival Analysis

Time to event

Duration data

Key concept: Censored duration

NB:

* it’s a mistake to considered it as a logistic regression.
* Mistake to drop censored duration / NA

Our project must be on this kind of data

Use case : Survival data / censored data

when will a rat touch a bright object ?

\* table on moodle

* Probabilistic description

S(t) = P(t >s)

T: We observe the minimum between time to event and U

U: time to interruption

* Observed => min(T,U, delta)
* Y => Delta(0 : if U > T or 1 if U <= T)
* Kaplan-Meier estimator
* In R. library(Survival)
* Plot : survival plot
* See Also Nelson\_Aalen

in R add option = “fh”

Plan in R

1. express your data in 2 cols : time & status / event indicator
2. Call survfit(time, censoring indicator/status)
3. Plot the curve / get median

Surv analysis

A - Time , status Indicator (SI)

B - start\_date, stop\_date, Event, timestamp

SI = 0 censored, 1 event

How to present

1. Data presentation
2. KM plot and median time students got an internship.

Factors affecting tti

1. Simple approach

* Stratified analysis

1. Qualitatively

Strong impact of age and children

Weak effect of age and sex

Q2 : what are significant factors ? (aka yes/no question way) =>

Does age has an effect on tti ?

* There is an effect is it significant ?
* To answer hyp testing. : H0 Sa(t) = Sb(t)
* Solution : logrank test
* Fun = survdiff( )
* Plan

1. Data obtain(duration, event)
2. curves in R survfit() => plot/figures
3. Compare group => survdiff()

R package ASAUR

Cumulated prob for 6 months

* Summary(fit\_object, time = 6)

DCHLB TRIAL

Drug test : test of plasma susbstitute.

Survival of patient

Randomized control trial with 2 arms

– SALIN

* Plasma

Case Study: PHARMACOSMOKING

Treatments

1. Randomized controlled trial

* Combo : combination treatment patch +psycho th.
* Psycho therapy

Endpoint: time to release.

Probabilistic description of survival data

Hazard function

Proportional hazard function

Cox proportional Hazard: Coxph

REGRESSION ANALYSIS

Linear regression

* 1. Risk of smth occurs
* 2. Cxph h(t|x) = h0(t)exp(xi\*beta)
* 3. Library survival

Coxph(srv(t,O)~covariates)

* Exp(beta) if x = [0,1 ] => hazard ratio

RECAP

2 groups :

1. H.R
2. One cont covariates
3. Multiple covariates

CASE STUDY / DSTI SURVEY : cox model => risk finding an internship (age & having children)

### RECAP

1. Cox regression compare 2 groups
2. Continuous variables
3. Mult regression covariates
4. Covariates with more than 2 levels. (example: race { white, Asian, Black, other })

### COX Regression recap

1. Model specification
2. Application to single binary covariate
3. Single continuous covariate
4. Multiple covariate
5. Categorical levels with more than 2 levels.

## Making predictions

RECAP tt event

1. sample KM
2. 2 sample log rank
3. Regression coxph

## Models Diagnosis

Martingale Residuals !