

# EDA

Runze Cui

2023-11-10

## Read File

```
dat <- read_csv("./data/heart_failure.csv") %>%
  arrange(TIME) %>% janitor::clean_names() %>%
  mutate(gender = factor(gender),
         smoking = factor(smoking),
         diabetes = factor(diabetes),
         bp = factor(bp),
         event = factor(event),
         anaemia = factor(anaemia)) %>%
  rename(platelets = pletelets)
# write_xlsx(dat, "./data/dat.xlsx")
# read_excel("./data/dat.xlsx")
dat_overview = head(dat)
```

## EDA:

- 299 patients of heart failure comprising of 105 women and 194 men.
- All the patients were more than 40 years old.

**Discrete Variable:** event (indicator), gender, smoking, diabetes, bp and anaemia.

**Continuous Variable:** time, age, ejection\_fraction (EF), sodium, creatinine, pletelets and cpk.

Overall	
(N=299)	
<b>Event indicator (censored = 0, event = 1)</b>	
0	203 (67.9%)
1	96 (32.1%)
<b>Gender (female = 0, male = 1)</b>	
0	105 (35.1%)
1	194 (64.9%)
<b>Smoking status (non-smoker = 0, smoker = 1)</b>	
0	203 (67.9%)
1	96 (32.1%)
<b>Diabetes (no diabetes = 0, diabetes = 1)</b>	
0	174 (58.2%)
1	125 (41.8%)
<b>Blood Pressure (normal = 0, hypertension = 1)</b>	
0	194 (64.9%)
1	105 (35.1%)
<b>Anaemia (no anaemia = 0, anaemia = 1)</b>	
0	170 (56.9%)
1	129 (43.1%)

Overall	
(N=299)	
<b>Survival time (days)</b>	
Mean (SD)	130 (77.6)
Median [Min, Max]	115 [4.00, 285]
<b>Age</b>	
Mean (SD)	60.8 (11.9)
Median [Min, Max]	60.0 [40.0, 95.0]
<b>Ejection Fraction (EF)</b>	
Mean (SD)	38.1 (11.8)
Median [Min, Max]	38.0 [14.0, 80.0]
<b>Serum Sodium</b>	
Mean (SD)	137 (4.41)
Median [Min, Max]	137 [113, 148]
<b>Serum creatinine</b>	
Mean (SD)	1.39 (1.03)
Median [Min, Max]	1.10 [0.500, 9.40]
<b>Plateletes</b>	
Mean (SD)	263000 (97800)
Median [Min, Max]	262000 [25100, 850000]
<b>CPK</b>	
Mean (SD)	582 (970)
Median [Min, Max]	250 [23.0, 7860]