

P8108 Survival Analysis Presentation (Title TBD)

Yiming Zhao (yz3955), Wenshan Qu (wq2160)
Tucker Morgan (tlm2152), Junzhe Shao (js5959),
and Benjamin Goebel (bpg2118)

2022-11-20

Presentation Outline

- ▶ Background / Research question of interest
- ▶ Exploratory Analysis
- ▶ Methods
 - ▶ Cox Proportional Hazard with Elastic Net
 - ▶ Survival Random Forest
 - ▶ Conformalized Analysis
- ▶ Results
- ▶ Discussion

Background

- ▶ We analyzed the Rotterdam data set from the `survival` package in R. This data includes 2982 breast cancer patients from the Rotterdam tumor bank.
- ▶ Notable measurements include age, menopausal status, tumor size, number of effected lymph nodes, and other baseline variables.
- ▶ Interventions were hormone treatment and chemotherapy.
- ▶ Subjects were followed until death or loss to follow up with occurrences of remission and death being recorded.

| year | age | meno | size | grade | nodes | pgr | er | hormon | chemo | rtime | recur | dtime | death |
|------|-----|------|-------|-------|-------|-----|-----|--------|-------|-------|-------|-------|-------|
| 1992 | 74 | 1 | <=20 | 3 | 0 | 35 | 291 | 0 | 0 | 1799 | 0 | 1799 | 0 |
| 1984 | 79 | 1 | 20-50 | 3 | 0 | 36 | 611 | 0 | 0 | 2828 | 0 | 2828 | 0 |
| 1983 | 44 | 0 | <=20 | 2 | 0 | 138 | 0 | 0 | 0 | 6012 | 0 | 6012 | 0 |
| 1985 | 70 | 1 | 20-50 | 3 | 0 | 0 | 12 | 0 | 0 | 2624 | 0 | 2624 | 0 |
| 1983 | 75 | 1 | <=20 | 3 | 0 | 260 | 409 | 0 | 0 | 4915 | 0 | 4915 | 0 |

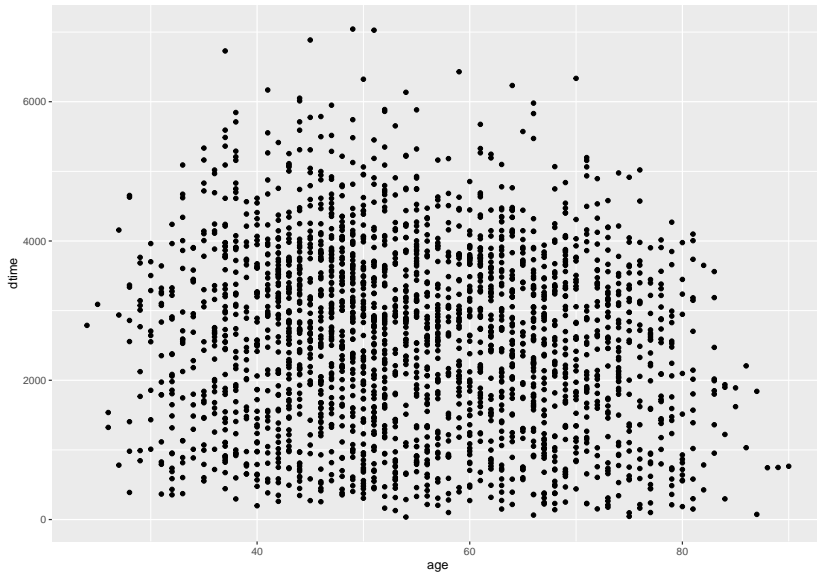
Exploratory Analysis

- Example: You can add arrows ">-"

| Overall (N=2982) | |
|------------------|-----------------|
| age | |
| Mean (SD) | 55.058 (12.953) |
| Range | 24.000 - 90.000 |
| meno | |
| Mean (SD) | 0.560 (0.496) |
| Range | 0.000 - 1.000 |
| size | |
| <=20 | 1387 (46.5%) |
| 20-50 | 1291 (43.3%) |
| >50 | 304 (10.2%) |
| grade | |
| Mean (SD) | 2.734 (0.442) |
| Range | 2.000 - 3.000 |
| nodes | |
| Mean (SD) | 2.712 (4.384) |

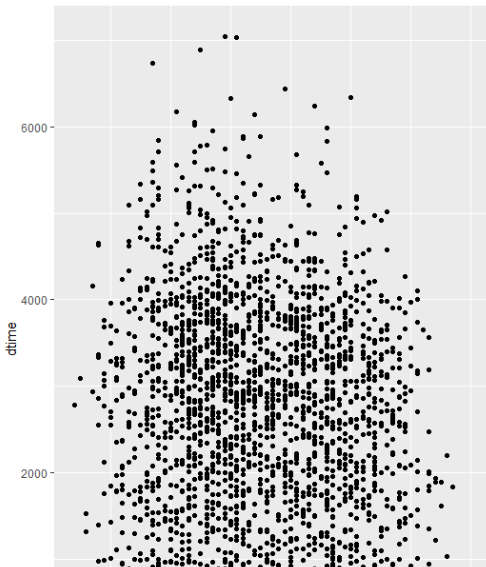
Methods

- ▶ We can use code chunks to illustrate plots



Methods

- Or you can also insert images like this (but the image dimensions have to be correct):



Log-rank Test - Hormon

- ▶ Hypothesis

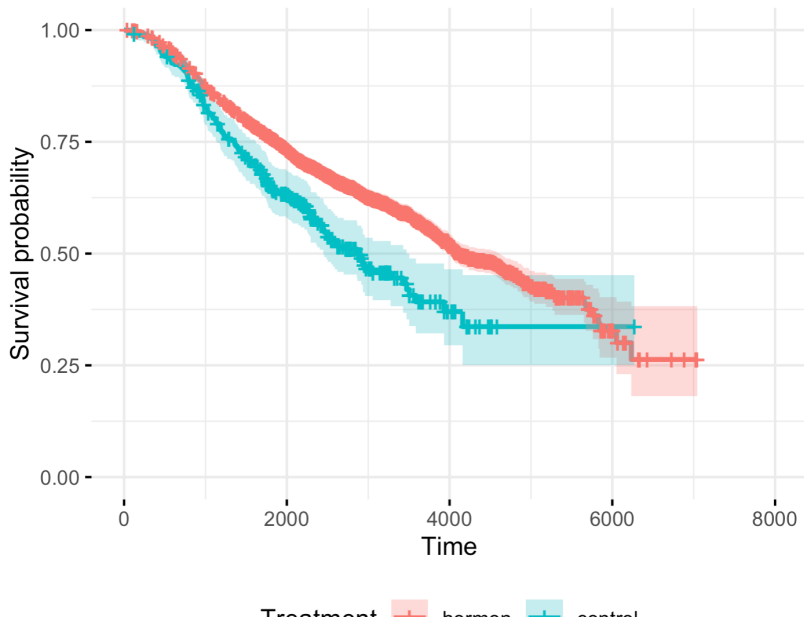
- ▶ $H_0 : S_{hormon}(t) = S_{control}(t); H_a : S_{hormon}(t) \neq S_{control}(t).$

- ▶ Log-rank test

- ▶ The test statistic is 23.7, and corresponding p-value is $1.13^{-6} < 0.05$, reject H_0 and conclude: we are 95% confident that the survival probability function of hormon group and control group are significantly different;
 - ▶ The test statistic is positive, thus conclude that hormon treatment is effective to the survival of patients with breast cancer.

Log-rank Test - Hormon (Continued)

Survival Curve of Hormon and Control group



Log-rank Test - Chemotherapy

- ▶ Hypothesis

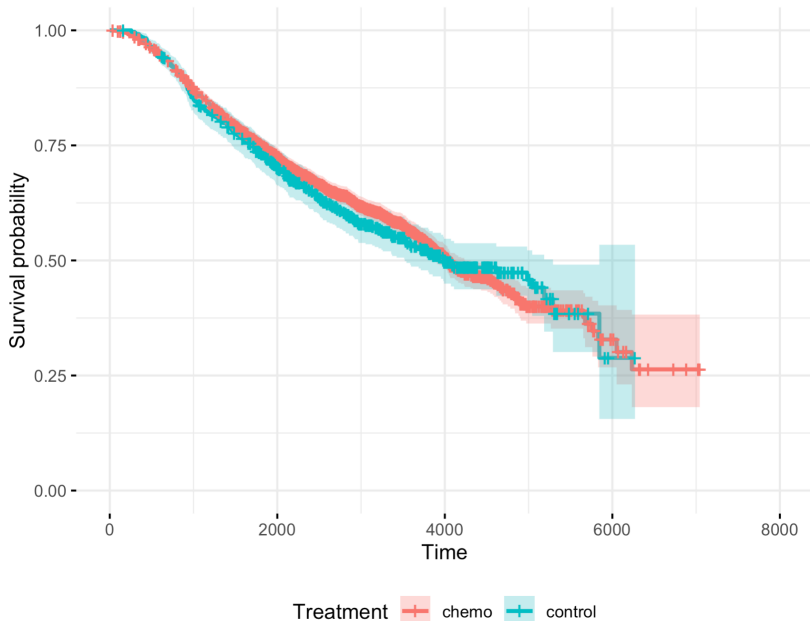
- ▶ $H_0 : S_{chemo}(t) = S_{control}(t); H_a : S_{chemo}(t) \neq S_{control}(t).$

- ▶ Log-rank test

- ▶ The test statistic is 0.495, and corresponding p-value is $0.48 > 0.05$, fail to reject H_0 and conclude that there are no difference between the survival probability function of chemotherapy group and control group;
 - ▶ Chemotherapy is not effective to breast cancer.

Log-rank Test - Chemotherapy (Continued)

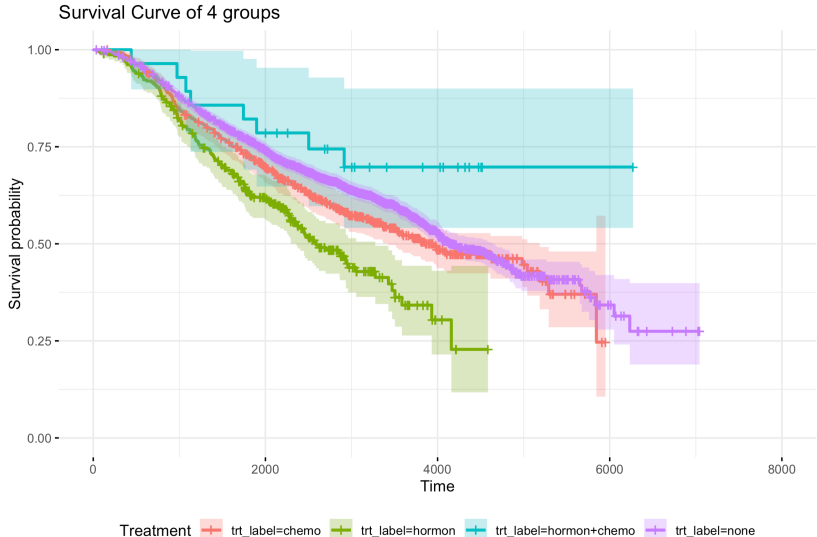
Survival Curve of Chemo and Control group



Log-rank Test - Hormon + Chemotherapy

- ▶ Regroup
 - ▶ 4 groups: hormon+chemo, hormon only, chemo only and none.
- ▶ Hypothesis
 - ▶ $H_0 : S_{hormon+chemo}(t) = S_{hormon}(t) = S_{chemo}(t) = S_{none}(t);$
 - ▶ $H_a : \text{at least two survival functions are not equal.}$
- ▶ Log-rank test
 - ▶ The test statistic is 40.4, and corresponding p-value is $9^{-9} < 0.05$, reject H_0 .

Log-rank Test - Hormon + Chemotherapy (Continued)



- Chemotherapy or Hormon alone does not have treatment effect on breast cancer?

- Hormon + Chemotherapy has a treatment effect on breast cancer

Log-rank Test - Hormon + Chemotherapy (Discussion)

| Var1 | Freq |
|--------------|------|
| chemo | 552 |
| hormon | 311 |
| hormon+chemo | 28 |
| none | 2091 |

- ▶ Discussion

- ▶ The sample size of hormon+chemo group is 28;
- ▶ Within hormon treatment group, only 8% people receive chemotherapy at the same time.

- ▶ Conclusion

- ▶ Reserve the results of the two previous separate log-ranks tests.

Prediction

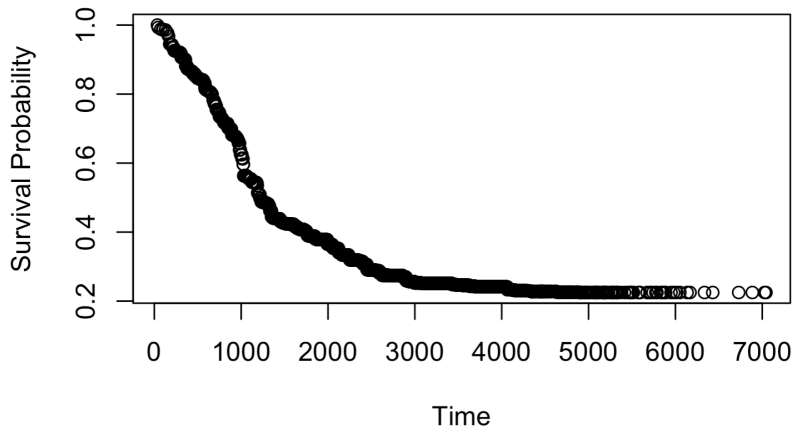
- ▶ Random Survival Forest
- ▶ Cox/Cox with Elastic Net
- ▶ Comparison with Brier Score

Random Survival Forest

- ▶ For a single survival tree:
 - ▶ Assign subjects to groups based on certain splitting rules regarding their covariates;
 - ▶ The subjects in each group will share a similar survival behavior;
 - ▶ Estimate survival probability of a given data point based on its “neighbors”.
- ▶ Assemble trees: Random Survival Forest (RSF)
 - ▶ use ranger package to train RSF;
 - ▶ non-parametric approach, no interpretable parameters;
 - ▶ prediction result for a single subject:
 - ▶ 7th subject in test data set, with pid = 58.

Random Survival Forest - Example

Survival Prediction for Patient 7



- Estimated median survival time: 1217 days.