STATISTICAL DEPARTMENT OF UC DAVIS

Final Summary Report

Bohao Zou
December 19, 2020

1 DEFINE THE ESSENTIAL PARTS OF SURVIVAL DATA

When we get a survival dataset from a institution the first thing we need to do is to define the objections, what we are interested in this data. The next step we need to define is what is the Time origin, Time scale , Event of interest and the Mechanism of censoring and/or truncation of this dataset.

2 NONPARAMETRIC METHODS

We can use Kaplan-Meier method to estimate the survival function $S(t)$, its variance and the confidence interval. We can also use Nelson-Aalen method to estimate the cumulative hazard function, its variance and the confidence interval. For survival function $S(t)$, hazard function $h(t)$ and cumulative hazard function $H(t)$, if we know one of those three function, we can conduct the other two functions.

3 HYPOTHESIS TESTING

3.1 ONE SAMPLE TESTING

If we have some samples of n iid observations from a population and we want to test if the samples comes from a population with a per-specified hazard rate.

$$H_0 : h(t) = h_0(t) \quad for \quad all \quad t \leq \tau$$
$$H_1 : h(t) \neq h_0(t) \quad for \quad some \quad t \leq \tau$$

where $h(t)$ is the hazard rate function of observation samples and $h_0(t)$ is the hazard rate function of pre-specified. τ is the largest observed study times.

3.2 K-Sample Test

If we want to compare the hazard rates of K (K ≥ 2) populations, we need to use K-Sample test(Log-rank test). The null hypothesis and alternate hypothesis are showed below:

$$H_0 : h_1(t) = h_2(t) = \cdots = h_k(t) \quad for \quad all \quad t < \tau$$
$$H_1 : At \quad least \quad one \quad of \quad h_j(t) \quad is \quad different \quad for \quad some \quad t \leq \tau$$

This may be the most frequency used testing in survival analysis.

3.3 Tests For Trend

This test should be applied only when there is some prior information that the alternative are ordered like ordinal categorical factors. By using this test, we can test the ordered hazard rates of K samples. The null hypothesis and alternate hypothesis are showed below:

$$H_0 : h_1(t) = h_2(t) = \cdots = h_k(t) \quad for \quad all \quad t < \tau$$
$$H_1 : h_1(t) \leq h_2(t) \leq \cdots \leq h_k(t) \quad for \quad t < \tau \quad with \quad at \quad least \quad one \quad strict$$

3.4 Stratified Tests

The stratified log rank test is the test you'd use when analyzing the survival distribution of two samples which are divided into two or more groups or "strata" based on common criteria that affect the outcome. Before using the stratified log rank test you can check that your variable of interest is not a confounding variable.Using the stratified log rank test also assumes that the effect (survival) is similar across both strata of the confounder.

$$H_0 : h_{1s}(t) = h_{2s}(t) = \cdots = h_{ks}(t) \quad for \quad all \quad t < \tau, \quad s = 1, 2, ..., M$$
$$H_1 : At \quad least \quad one \quad of \quad h_{js}(t) \quad is \quad different \quad for \quad some \quad t \leq \tau$$

Where M is the number of stratified.

4 Semiparametric Proportional Hazards Regression

Cox-PH regression model is very useful in comparing two or more groups of time to event with many covariates. This model can help us to explain our covariates. As a usual, if a Local test shows that one variable is significant and the $\theta = exp(\beta_j)$ is bigger than 1, we can roughly know that this variable can increase the relative risk compared with other status of this variable under the other covariates are the same. If the θ is smaller than 1, this can indicate that this variable will decrease the relative risk compared with other status of this variable under the other covariates are the same. If the θ equal with 1 or the Local test shows that this variable is not significant, we can roughly know that this variable has no effect to the hazard rate.

4.1 Model selection

The first step we need to do is model selection. There are three main model selection methods, its are forward selection, backward selection and stepwise selection. In forward selection, if a variable has been added into the model, it would not remove. In the backward selection, if a variable has been removed from the model, the variable would not be added into the model again. In stepwise selection, variables can be removed or add. For each method, we must set the P-Value for entry or out the model.

4.2 Model Diagnostics

In model diagnostics, we can use the Cox-Snell residuals to check the overall fit of the model. If the model fits the data well, the plot should follow the 45 degree line. We can use Martingale residuals to check nonlinearity of a continuous variable. If the plot is linear, it indicates no transformation. We can also use the Deviance residuals to identify outliers. There are two ways to check the PH assumption, the first is used Schoenfeld residuals for a variable against time to evaluate the PH assumption. A plot shows a non-random pattern against time is evidence of violation of the PH assumption for this variable. The second way is to use ZPH correlation test to test the correlation between time and the Schoenfeld residuals for each variable.

4.3 Stratified Cox-PH Model

If PH assumptions is not met, it can use the Stratified Cox-PH Model on that variable and employ the PH model within each stratum for the other covariates. The key assumption for Stratified Cox-PH Model is the coefficients is same for each stratum. We can use likelihood ratio test to test if the assumption holds. The test statistical is

$$T = -2(LL(\overrightarrow{\beta}) - \Sigma_{j=1}^{S} LL_j(\overrightarrow{\beta})) \sim \chi^2((S-1) \times P)$$

where S is the number of stratums, P is the number of covariates and $LL(\overrightarrow{\beta})$ is the partial log likelihood of stratified model that fitted with all of data. The $LL_j(\overrightarrow{\beta})$ is the partial log likelihood of stratified model that fitted with $j-th$ stratum data.

5 Conclusion

By using those knowledge, we can analyze any survival data set by ourselves, can solve most of problems that occurs in reality and give a reliable interpretation for those problems.