

Survival Analysis with Applications in Medicine: Take-home examination

2025-03-17

A. Weibull regression models

Q1. For a proportional hazards model with $S_0(t)$ that is a Weibull distribution, show that survival $S(t)$ is also from a Weibull distribution.

Given a proportional hazards model with survival function $S(t|x) = S_0(t)^{\exp(\beta x)}$

for time t and a given covariate x , $S_0(t)$ is the baseline survival function, and β is the log hazard ratio.

We have the Weibull survival function $S_0(t) = \exp(-\lambda t^k)$ for a scale parameter λ and a shape parameter k .

Substituting $S_0(t)$ in the proportional hazards model, we get the survival function as

$$S(t|x) = [\exp(-\lambda t^k)]^{\exp(\beta x)}$$

$$\Rightarrow S(t|x) = \exp(-\lambda t^k \exp(\beta x))$$

which is in form $S(t|x) = \exp(-\tilde{\lambda}_a t^k)$ where $\tilde{\lambda}_a = \lambda \exp(\beta x)$ is the new scale parameter for the same shape parameter k .

This means that the survival function $S(t|x)$ for a proportional hazards model is also from a Weibull distribution.

Here, $S(t) = E_X[S(t | X)]$, which will also be from a Weibull distribution.

Q2. For an accelerated failure time model with $S_0(t)$ that is a Weibull distribution, show that survival $S(t)$ is also from a Weibull distribution.

Given an accelerated failure time model with survival function $S(t|x) = S_0(t \exp(-\tilde{\beta} x))$

for time t and a given covariate x , $S_0(t)$ is the baseline survival function, and $\tilde{\beta}$ is the log time ratio.

Substituting $S_0(t)$ in the accelerated failure time model, we get the survival function as

$$S(t|x) = \exp(-\lambda (t \exp(-\tilde{\beta} x))^k)$$

$$\Rightarrow S(t|x) = \exp(-\lambda t^k \exp(-k\tilde{\beta} x))$$

which is in form $S(t|x) = \exp(-\tilde{\lambda}_b t^k)$ where $\tilde{\lambda}_b = \lambda \exp(-k\tilde{\beta} x)$ is the new scale parameter for the same shape parameter k .

This means that the survival function $S(t|x)$ for an accelerated failure time model is also from a Weibull distribution.

Similarly, $S(t) = E_X[S(t | X)]$, which will also be from a Weibull distribution.

Q3. What is the relationship between β and $\tilde{\beta}$ if both models have a Weibull baseline survival function?

For the proportional hazards model, we have $S(t|x) = \exp(-\lambda t^k \exp(\beta x))$.

For the accelerated failure time model, we have $S(t|x) = \exp(-\lambda t^k \exp(-k\tilde{\beta}x))$.

Comparing the two models, we get $\exp(\beta x) = \exp(-k\tilde{\beta}x)$.

Taking the natural logarithm of both sides, we get $\beta x = -k\tilde{\beta}x$.

Therefore, the relationship between β and $\tilde{\beta}$ is $\beta = -k\tilde{\beta}$ if both models have a Weibull baseline survival function.

B: Interval-censored likelihood

For a data tuple (t_i, u_i, v_i) where t_i is the (left truncated) delayed entry time, and the event is observed in the interval $(u_i, v_i]$ for an individual i .

Q1a. Express the log-likelihood in terms of Survival function $S(t)$ at time t :

The Likelihood for the interval-censored data: $u_i < T \leq v_i$ for an entry time t_i is given by:

$$L_i = \frac{S(u_i) - S(v_i)}{S(t_i)}$$

where $S(t)$ is the Survival function.

Hence, the log-likelihood in terms of Survival function $S(t)$ at time t is given by $\log(S(u_i) - S(v_i)) - \log(S(t_i))$.

Q1b. Express the log-likelihood in terms of the hazard function $h(t)$ at time t :

Now, for the derivation of the log-likelihood in terms of the hazard function $h(t)$ at time t , we need to express the Survival function $S(t)$ in terms of the hazard function $h(t)$.

We know that, the survival function $S(t)$ is given by $-\log(S(t)) = H(t)$,

where $H(t)$ is the cumulative hazard function, and $H(t) = \int_0^t h(u)du$ for the hazard function $h(t)$.

Therefore, the log-likelihood in terms of the hazard function $h(t)$ at time t is:

$$\begin{aligned} \log L_i &= \log(S(u_i) - S(v_i)) - \log(S(t_i)) \\ \Rightarrow \log L_i &= \log(\exp(-H(u_i)) - \exp(-H(v_i))) - \log(\exp(-H(t_i))) \\ \Rightarrow \log L_i &= \log(\exp(-H(v_i)) * (\exp(H(v_i) - H(u_i)) - 1)) + H(t_i) \\ \Rightarrow \log L_i &= \log(\exp(-H(v_i))) + \log(\exp(H(v_i) - H(u_i)) - 1) + H(t_i) \\ \Rightarrow \log L_i &= H(t_i) - H(v_i) + \log(\exp(H(v_i) - H(u_i)) - 1) \\ \Rightarrow \log L_i &= \int_0^{t_i} h(t)dt - \int_0^{v_i} h(t)dt + \log(\exp(\int_{u_i}^{v_i} h(t)dt) - 1) \\ \Rightarrow \log L_i &= -\int_{t_i}^{v_i} h(t)dt + \log(\exp(\int_{u_i}^{v_i} h(t)dt) - 1) \end{aligned}$$

Hence, the log-likelihood in terms of the hazard function $h(t)$ at time t is given by $-\int_{t_i}^{v_i} h(t)dt + \log(\exp(\int_{u_i}^{v_i} h(t)dt) - 1)$.

Q2. Can you express these data using the `Surv` function from the `survival` package? If so, show an example; if not, explain why.

Yes, we can express the interval-censored data using the `Surv` function from the `survival` package.

The `Surv` function is used to create a survival object that represents the survival time of an individual. It takes the form `Surv(time, event)` where `time` is the survival time and `event` is the event indicator.

For interval-censored data, we can use the `Surv` function as `Surv(time, time2, type = "interval2")` where `time` is the start of the interval, `time2` is the end of the interval and type `interval2` is used to indicate interval-censored data effectively.

Although, the `Surv` function doesn't support left truncation directly, we can filter out the left truncated data by taking the maximum of the entry time and the left truncation time.

Here is an example of how to express left-truncated interval-censored data using the `Surv` function:

```
library(survival)

# sample data for the given data tuple structure (t_i, u_i, v_i)
d = data.frame(left_truncation_time = c(1,1,3,3,3),
               entry_time = c(0,0,2,2,2),
               exit_time = 1:5,
               event = c(1,0,1,0,1))

# filter out left truncated data
d$entry_time = pmax(d$entry_time, d$left_truncation_time)

# interval-censored data
with(d, Surv(entry_time, exit_time, type="interval2"))
```

```
## [1] 1      [1, 2] 3      [3, 4] [3, 5]
```

C: Truncated distributions

D: Cox's partial likelihood with a time-varying effects

E: Data analysis of a randomised controlled trial for hormonal treatment of breast cancer patients in Germany

F: Analysis plan for a randomised controlled trial

This assignment took me approximately 4 hours to complete.