

*BIOS 522: Survival Analysis Methods*

**Homework - Weeks 9-11**

*Homework assignments contain a mix of theoretical, applied, and computing questions. Students are encouraged to work with classmates but must prepare their answers individually. Copying and pasting between students is not allowed. Assignments are graded on presentation, and points will be deducted from messy or illegible answers.*

Problem 1. Relationship between the AFT and proportional hazards model (5 points)

In class, we asserted that the Weibull AFT model is a proportional hazards model, but the log-logistic AFT model is not a proportional hazards model. In this question, we will demonstrate this.

*Exponential AFT model*

To see the process, we will start with an exponential AFT model. We will use a single covariate, although the approach generalizes to multiple covariates. The survival distribution for individual with covariate is:

where we define a new rate parameter for individual . Recall that, in an exponential model, the hazard function is constant over time and equal to the rate parameter.

Our AFT model implies the following hazard ratio for a one-unit increase in :

We see that the hazard ratio for a one-unit increase in meets the proportional hazards assumption, i.e., it is constant over time and does not depend on . Therefore, the exponential AFT model is also a proportional hazards model, and the hazard ratio is , i.e., the inverse of the acceleration factor.

*Weibull AFT model*

Now consider a Weibull AFT model with a single covariate.

1. (0.5 points) Write out the survival distribution for individual with covariate .

Since for the Weibull distribution, to find the survival distribution for individual , we replace with :

1. (0.5 points) Write out the Weibull rate parameter for individual .

where we define a new rate parameter for individual .

1. (1 point) Recall that the hazard function for a Weibull distribution is . Write out the hazard ratio for a one-unit increase in implied by our Weibull AFT model. Simplify this hazard ratio to show that it is NOT a function of time.

We see that, in our AFT model, the hazard ratio for a one-unit increase in is , and this is constant over time. Therefore, the Weibull AFT model implies a proportional hazards model.

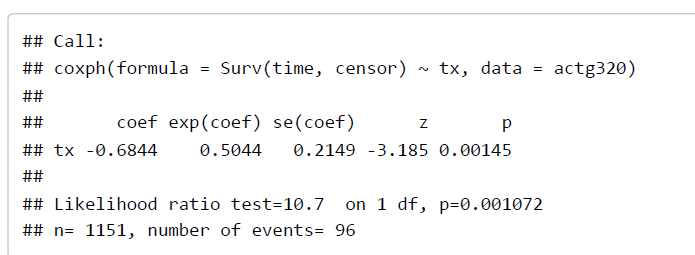
1. (0.5 points) Consider the Weibull AFT model fit to the ACTG 320 data summarized in the Week 10: R Tutorial. The estimated baseline rate parameter is   days-1, the estimated shape parameter is  , and the estimated coefficient for the single covariate for treatment is  .

Use these results and the expression in the previous problem to calculate the hazard ratio implied by our Weibull AFT model.

Note that *the acceleration factor is not the inverse of the hazard ratio* . This is because the shape parameter appears in the hazard ratio. But we can estimate the hazard ratio as

1. (0.5 points) Consider the Cox proportional hazards model fit on page 3 of the Week 7: R Tutorial. What was the hazard ratio for the covariate of treatment? (*Hint: it should be similar to what you just calculated…*)

HR = Exp(coef)= 0.5044. They are very similar.



*Log-logistic AFT model*

In contrast to the exponential and Weibull models, the log-logistic AFT model and log-logistic proportional hazards models are not equivalent.

Now consider a log-logistic AFT model with a single covariate.

1. (0.5 points) Write out the survival distribution for individual with covariate .

Since for the log-logistic distribution, to find the survival distribution for individual , we replace with :

1. (0.5 points) Write out the Weibull rate parameter for individual .

where we define a new rate parameter for individual .

1. (1 point) Recall that the hazard function for a Weibull distribution is . Write out the hazard ratio for a one-unit increase in implied by our log-logistic AFT model. It is not necessary to simplify your expression.

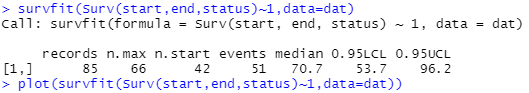
*Even after some simplification, the hazard ratio would still be a function of time . Because it is not constant over time, the log-logistic AFT model does not imply a proportional hazards model. Thus, these models are different.*

Problem 2. Staggered entry example (2 points)

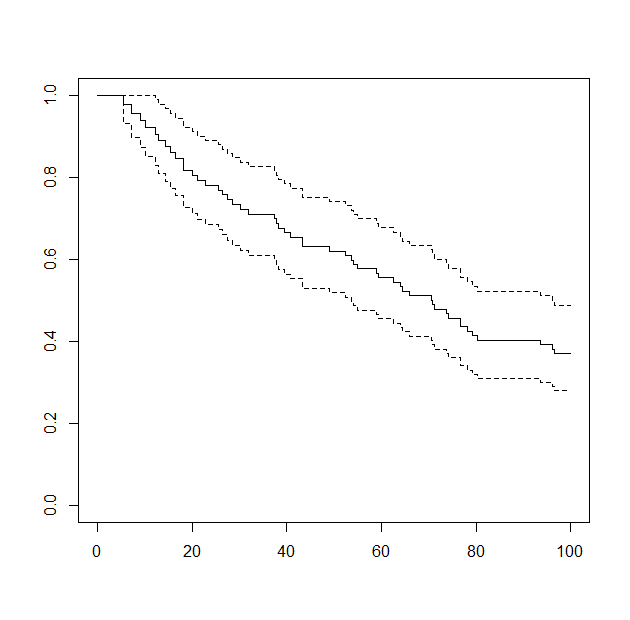
The file dat.Rdata includes study entry and follow-up times for a hypothetical cohort. For individuals, start times and event times were randomly generated from a gamma distribution. Individuals were censored at the study end time units as appropriate. The data set includes observations because 15 individuals were *left-truncated*.

1. (1 point) Fit a Kaplan-Meier curve fitted to the data accounting for the staggered entry times. Report the median survival time and a plot.

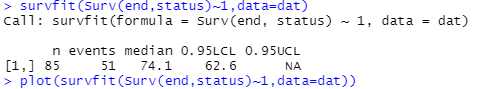
*Hint: Ensure that the Surv() object includes the entry times (start).*



Median is 70.7 units.

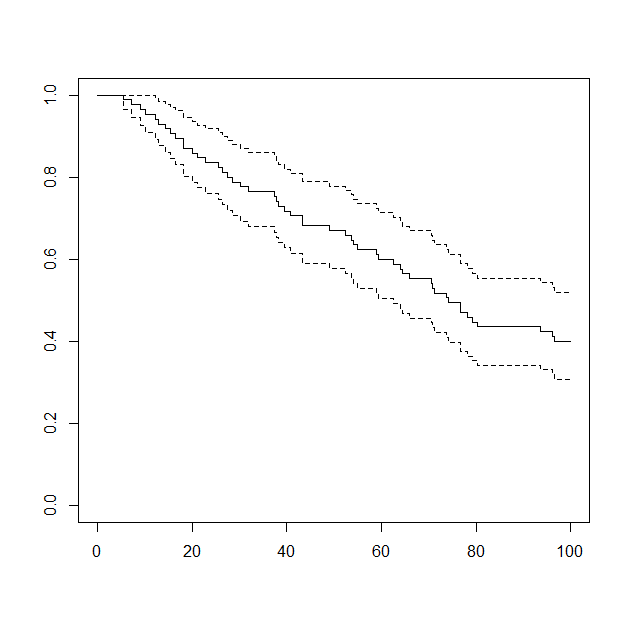


1. (1 point) Fit a Kaplan-Meier curve fitted to the data IGNORING the staggered entry times. Report the median survival time and a plot. *Which has higher survival and why?*



Median is 74.1 units. This inappropriate analysis has higher survival because it disregards the potential for LEFT TRUNCATION… people who fail before they enter into the study. As we saw in this example, there were 15 such people.

plot(survfit(Surv(end,status)~1,data=dat))



Problem 3. Breast cancer treatment trial (4 points)

Goss et al. (2003) summarizes the results of a double-blind, placebo-controlled, Phase 3 clinical trial evaluating the breast cancer treatment letrozole.

Women with hormone-dependent breast cancer are frequently prescribed long-term treatment with tamoxifen, an estrogen-receptor modulator, to reduce the risk of tumor recurrence. This type of treatment is referred to as “adjuvant chemotherapy.” Earlier clinical research has suggested that long-term treatment with tamoxifen is not beneficial because the patient may eventually develop resistance to the drug. For this reason, tamoxifen is not prescribed for more than five years. For patients completing their tamoxifen regimen, early evidence suggests that aromatase inhibitors, including letrozole, may reduce cancer recurrence.

The trial enrolled post-menopausal, amenorrheic women with confirmed hormone-receptive breast cancer who had received approximately 5 years of adjuvant tamoxifen therapy. Participants were randomized to letrozole or placebo.

The primary endpoint is “disease-free survival,” and it is defined in the first and last paragraphs of the methods section.

In the methods section, the authors describe how they conducted their sample size calculations. They determined that they needed to observe 515 events and needed 4800 women to have 80% power to detect a 2.5% difference in four-year disease-free survival, with Type 1 error of 0.05.

1. (1 point) The authors assumed that the four-year disease-free survival rate in the placebo population was 88%. Calculate the hazard rate in the placebo group, assuming that it is constant over time.

*Hint: Look at the example given in Step 5 of the lecture notes.*

If survival at 4 years is 88%, we can estimate the hazard rate as the following.

years-1

1. (1 points) The authors determined that an increase in four-year survival of 2.5% was clinically meaningful. Thus, the authors want to detect four-year survival in the letrozole group above 90.5%. Calculate the hazard rate in the letrozole group that corresponds to four-year survival of 90.5%.

If survival at 4 years is 90.5%, we can estimate the hazard rate as the following.

years-1

1. (1 point) Calculate the hazard ratio implied if four-year survival in the letrozole group is 90.5% and four-year survival in the placebo group is 88%.

The hazard ratio is: .

This is also reported in the statistical analysis section. “Hazard ratio for local or metastatic recurrence of the disease or the diagnosis of contralateral breast cancer, 0.78).

1. (1 point) The authors report that they want 80% power at a two-sided alpha level of 0.05 to detect the hazard ratio from (c). Calculate the required number of events to achieve this power.

*Hint: Look at the example given in Step 4 of the Week 11 Reading. To recreate their calculations, round the hazard ratio to 3 significant digits and then round the total number of events up.*

Which we then round up to 515.

This is the number of events reported in the statistical analysis section. “These assumptions necessitated the enrollment of 4800 women, over a four-year period with two years of follow-up, accounting for 515 events.”