Applied Survival Analysis
Solutions to Lab 4: Cox Proportional Hazards Model

1. *Interpretation of Cox Model: Prognosis with Breast Cancer*

(a) From the following proportional hazards model:

$$\lambda(t, X) = \lambda_{0X_S}(t)\exp(-0.783X_E + 0.007X_A)$$

The estimated hazards for a woman with ER positive tumor and for a woman with ER negative tumor, holding all other values constant, are given by:

$$\lambda_+(t) = \lambda_{0X_S}(t)\exp(-0.783 \cdot 1 + 0.007X_A) = \lambda_{0X_S}(t)\exp(-0.783)\exp(0.007X_A)$$
$$\lambda_-(t) = \lambda_{0X_S}(t)\exp(-0.783 \cdot 0 + 0.007X_A) = \lambda_{0X_S}(t)\exp(0.007X_A)$$

Therefore the hazard ratio for ER positive woman versus ER negative is given by:

$$HR(+:-) = \frac{\lambda_+(t)}{\lambda_-(t)} = \exp(-0.783) = 0.457$$

This implies that women with ER positive tumors have an approximately 54% ((1-0.46)100%) lower risk of dying compared to women with ER negative tumors. So women with ER positive tumors have a more favorable prognosis.

(b) The two hazards that we are interested in are the following:

$$\lambda_{62+}(24) = \lambda_{02}(24)\exp(-0.783 \cdot 1 + 0.007 \cdot 62) = \lambda_{02}(24)\exp(-0.783)\exp(0.007 \cdot 62)$$
$$\lambda_{67-}(24) = \lambda_{02}(24)\exp(-0.783 \cdot 0 + 0.007 \cdot 67) = \lambda_{02}(24)\exp(0.007 \cdot 67)$$

Therefore the corresponding hazard ratio is:

$$HR(62+:67-) = \frac{\lambda_{62+}(24)}{\lambda_{67-}(24)} = \exp(-0.783)\exp\{0.007(62-67)\} = \exp(-0.783 - 0.035) = 0.441$$

So the hazard of death for a woman 62 years old at diagnosis with a localized ER positive tumor, 24 months beyond diagnosis is approximately 56% less than a woman 67 years old at diagnosis with localized ER negative tumor, 24 months beyond diagnosis.

(c) In this case the two hazards of interest are the following:

$$\lambda_{55+}(36) = \lambda_{01}(36)\exp(-0.783 \cdot 1 + 0.007 \cdot 55) = \lambda_{01}(36)\exp(-0.783)\exp(0.007 \cdot 55)$$
$$\lambda_{55-}(24) = \lambda_{01}(24)\exp(-0.783 \cdot 0 + 0.007 \cdot 55) = \lambda_{01}(24)\exp(0.007 \cdot 55)$$

Therefore the corresponding hazard ratio is:

$$HR(55+:55-) = \frac{\lambda_{55+}(36)}{\lambda_{55-}(24)} = \frac{\lambda_{01}(36)}{\lambda_{01}(24)} \cdot \exp(-0.783) = \frac{\lambda_{01}(36)}{\lambda_{01}(24)} \cdot 0.457$$

So the hazard of death for a woman 55 years old at diagnosis with a localized ER positive tumor, 36 months beyond diagnosis is approximately $\frac{\lambda_{01}(36)}{\lambda_{01}(24)} \cdot 0.457$ times the hazard for a woman with the same age at diagnosis and with localized ER negative tumor, 24 months beyond diagnosis.

(d) The two hazards that we are interested in are the following:

$$\lambda_{60+}(24) = \lambda_{03}(24)\exp(-0.783 \cdot 1 + 0.007 \cdot 60) = \lambda_{03}(24)\exp(-0.783)\exp(0.007 \cdot 60)$$
$$\lambda_{60-}(24) = \lambda_{01}(24)\exp(-0.783 \cdot 0 + 0.007 \cdot 60) = \lambda_{01}(24)\exp(0.007 \cdot 60)$$

Therefore the corresponding hazard ratio is:

$$HR(60+:60-) = \frac{\lambda_{60+}(24)}{\lambda_{60-}(24)} = \frac{\lambda_{03}(24)}{\lambda_{01}(24)}\exp(-0.783) = \frac{\lambda_{03}}{\lambda_{01}}0.457$$

Note that we need the ratio of the baseline hazards of the regional and *in situ* stage, because stratified analysis assumes that the coefficients are the same across strata but baseline hazards are unique to each stratum. So the hazard of death for a woman 60 years old at diagnosis with a localized ER positive tumor, 24 months beyond diagnosis is approximately $\frac{\lambda_{03}}{\lambda_{01}}0.457$ times the hazard for a woman 60 years old at diagnosis with localized ER negative tumor, 24 months beyond diagnosis.

2. *Fitting Cox Model and handling of ties: Nursing Home Data*

(a) The estimate of β is $\hat{\beta} = 0.310$ and the corresponding Cox PH model is:

$$\lambda(t, X) = \lambda_0(t) \cdot \exp(0.310 \cdot X), \text{ where } X = \begin{cases} 1 & \text{if Married} \\ 0 & \text{if Not Married} \end{cases}$$

HR(Married : Not Married) = 1.36

So the risk of discharge for married individuals was approximately 36% greater than single individuals. In other words, the length of stay for married individuals was shorter than for single individuals.

(b) The estimate of β is identical in all four methods which implies that the tied failure times is not a big issue in this dataset. The computing time required for the discrete method (**exactp**) wasn't that great and since it is the best (exact) estimation of the hazard function we might as well use it.

(c) The test statistic that we obtained in the last lab was for the log-rank 18.74 (p<0.001) and for the Wilcoxon 16.91 (p<0.001) and these statistics don't match with any of the above. Another option that you can use with the **`sts test`** command is the **`cox`** this is the corresponding <u>Likelihood Ratio test</u> with the *Breslow* tied option in the **`stcox`** command.

```
sts test married, cox

       failure _d:  fail
 analysis time _t:  los


Cox regression-based test for equality of survival curves
---------------------------------------------------------

             |  Events                         Relative
married      |  observed       expected         hazard
-------------+--------------------------------------
Not Married  |      1032        1085.95         0.9562
Married      |       237         183.05         1.3041
-------------+--------------------------------------
Total        |      1269        1269.00         1.0000

                LR chi2(1) =      17.31
                   Pr>chi2 =     0.0000
```

During the lecture (p.24) we saw that the <u>linear log-rank test</u> (you get this in SAS from PROC LIFETEST using the TEST statement) is the same as the <u>score test</u> in the Cox model with *Breslow* tied option. We additionally saw (p.40) that the <u>log-rank test</u> is equivalent to the <u>score test</u> in the Cox model with the *discrete* tied option. To get the Wald test in STATA you square the z-test of the coefficient e.g. in this dataset the z-test of β was 4.299 if you square it $(4.299)^2 = 18.481$ = Wald test.