# HW 3

*Steph Holm and Shelley Facente*

*3/8/2019*

## Parametric Survival Analysis

### Question 1: Assuming ***Weibull*** distributed event times,

**a) Write out the *general* expression, not substituting any estimated values, and clearly defining any parameters and distributions for any random terms, the log-hazard function, including the complete baseline hazard.**

*Assuming $T \sim$ Weibull, then the log-hazard function is expressed as:*

$$log[h_0(t|\boldsymbol{x})] = log(p * t^{p-1}) + \beta_0 + \boldsymbol{x}\beta$$

*where $p$ is the shape parameter, $log(p * t^{p-1}) + \beta_0$ is the log-baseline hazard, and $\boldsymbol{x}\beta$ is the log hazard ratio, with Weibull distribution.*

**b) Repeat the above for the log-time function (in the accelerated failure time metric).**

*Again assuming $T \sim$ Weibull, the log-**time** function is expressed as:*

$$log(T) = \alpha_0 + \boldsymbol{x}\alpha + \sigma \times \epsilon^*$$

*where $\alpha_0$ is the baseline time-to-event, $\boldsymbol{x}\alpha$ is the log-time ratio, $\sigma$ is the variance-type constant term (also known as $p$, unrestricted with Weibull specificiations), and $\epsilon^*$ is the error term, which with a Weibull model approximates a G distribution (0,1).*

**c) For an arbitrary covariate, show the expression of the hazard ratio from the proportional hazards Weibull model as a function of parameters from the accelerated failure time expression of the same model.**

$$h^{AFT}(t|\boldsymbol{x}) = \frac{t^{\frac{1}{\sigma}} - 1}{\sigma} exp(-(\alpha_0 + \boldsymbol{x}\alpha)/\sigma)$$

*Note that this hazard function will take the ratio of $h^{AFT}$ but will not necessary be proportional as in the log-hazard function expressed in part (a).*

### Question 2: Complete the following table using the results from your analyses.

| BMI | Cox HR (95% CI) | Cox $SE_\beta$ | Exponential HR (95% CI) | Exp $SE_\beta$ | Weibull HR (95% CI) | Weibull $SE_\beta$ |
|---|---|---|---|---|---|---|
| <18.5 | 1.26 (0.79, 2.01) | 0.24 | 0.83 (0.52, 1.32) | 0.24 | 0.88 (0.66, 1.16) | 0.14 |
| 18.5 - <25 | *ref* | - | *ref* | - | *ref* | - |
| 25 - <30 | 1.06 (0.95, 1.19) | 0.06 | 0.93 (0.83, 1.05) | 0.06 | 0.96 (0.9, 1.03) | 0.03 |
| $\geq$ 30 | 1.54 (1.32, 1.78) | 0.24 | 0.69 (0.59, 0.8) | 0.08 | 0.78 (0.71, 0.85) | 0.05 |

## Question 3: Answer the following questions

### a) Which model estimates the relationships of interest most precisely? Justify your answer.

*The Weibull model estimates the relations of interest most precisely, because as a parametric model the Weibull provides better statistical efficiency than the Cox model (evidenced by tighter CIs in the table from Question 2); the Weibull also allows the baseline hazard to vary over the entire time period (the scale of the shape parameter $p$ is allowed to vary), making it more flexible than the exponential model, which restricts $p = 1$.*

### b) Based on the likelihood ratio test, what parameter from the model you outlined in Q1 is being evaluated? Based on the results of this test, would you select the exponential or Weibull model? Justify your answer.

*The likelihood ratio test produces the following output:*

```
##                                                     Terms Resid. Df    -2*LL
## 1 as.factor(bmi_cat) + age + male + factor(cursmoke)      4408 14695.21
## 2 as.factor(bmi_cat) + age + male + factor(cursmoke)      4407 14292.94
##   Test Df Deviance     Pr(>Chi)
## 1      NA       NA           NA
## 2    =  1 402.2752 1.760543e-89
```

*From this we can see that the p-value for $H_0 : p = 1$ (exponential) vs. $H_A : \sigma \neq 1$ is essentially zero; therefore we reject the exponential model and select the Weibull model.*

### c) Using the output from the Weibull model, calculate the time ratio comparing individuals with BMI > 30 to those with BMI 18.5-<25 (no need for confidence interval). Interpret this parameter. Does this agree or not with the corresponding hazard ratio?

*The time ratio (TR) for individuals with BMI > 30 compared to those with BMI 18.5-<25 is $e^{-0.248} = .781$. This means that being obese (BMI > 30) multiplies survival time by by .781, or decreases survival time by about 22%.*

# Competing Risks

## Question 4: Conceptually, what is the difference between the KM failure estimate for CVD death and the estimated CIF for CVD death? What does the comparison of these curves tell you about the risk of the competing event?

*The KM failure estimate for CVD death is the risk of all-cause mortality at time $t$, for people with CVD (compared to those without). The estimated CIF for CVD death represents the joint probability of mortality* ***from CVD****, by time $t$. Comparison of the curves tells you that as days of follow-up increase, subjects begin to die from causes other than CVD in higher proportion (i.e. the curves begin to diverge around 50 days of follow-up, with the KM failure function continuing to demonstrate greater probability of mortality than the CIF for the remainder of follow-up).*

## Question 5: In a competing risks analysis, briefly (1-2 sentences each) define in words the following terms:

### Cause-specific hazard

*A cause-specific hazard is the instantaneous (very short-term) rate of failure* ***for event**** *$j$* among those who have not yet experienced the event of interest* **or** *a competing event prior to $t$.*

**Subdistribution hazard**

*The subdistribution hazard is the instantaneous (very short-term) rate of failure for event $j$ among those alive at time $t$ or who experienced a competing event before time $t$.*

**Question 6: Complete the following table for the cause-specific hazard ratios (csHRs) and subdistribution HRs (sHRs) you estimated.**

| BMI | CVD csHR (95% CI) | Other csHR (95% CI) | CVD sHR (95% CI) | Other sHR (95% CI) |
|---|---|---|---|---|
| 18.5 - <25 | 1 | 1 | 1 | 1 |
| 25 - <30 | 0.89 (0.31, 2.53) | 0.54 (0.19, 1.51) | 1.17 (.40, 3.5) | 0.41 (.15, 1.1) |
| $\geq$ 30 | 2.36 (0.79, 7.02) | 0.25 (0.06, 0.96) | 3.3 (1.2, 8.9) | 0.18 (.07, .45) |

**Question 7: From the above table, explain how the pattern you observe in the csHRs is consistent with the pattern in the sHRs. (Consider the relationship between the csHR and sHRs.)**