Biostatistics 140.641 Survival Analysis
First Term, 2022

FINAL EXAM YOUR NAME:____________________________
!!!QUESTIONS ON BOTH SIDES!!!

1. (20%) Let T represent a failure time variable with the hazard function $\lambda(t)$. Answer the following questions:

   (a) Suppose $\lambda(t) = \theta \cdot I(t > 0)$, $\theta > 0$.

   - The probability density function of T is $\theta \cdot e^{-\theta t} \mathbb{1}(t>0)$.

   - The survival function of T is $e^{-\theta t} \mathbb{1}(t>0)$.

   - Suppose the survival data are observed subject to independent censoring. The likelihood function based on survival data $(3.3^+,\ 0.4,\ 2.5,\ 1.7^+)$ is

   $e^{-3.3\theta} \cdot \theta e^{-0.4\theta} \cdot \theta e^{-2.5\theta} \cdot e^{-1.7\theta}$.

   (b) Suppose $\lambda(t) = \beta e^{\theta t} \cdot I(t > 0)$, where $\beta > 0$ and $-\infty < \theta < \infty$. Identify the range of values of (β, θ) so that

   - the hazard function $\lambda(t)$ has a constant value: $\beta > 0$, $\theta = 0$.

   - the value of hazard function $\lambda(t)$ increases in t: $\beta > 0$, $\theta > 0$.

   - $\lambda(t)$ is a hazard function for "cure models" (that is, there are individuals who never experience the failure event): $\beta > 0$, $\theta < 0$.

2 (15%) Consider a two-sample (treatment v.s. control) study and answer the following questions:

(a) Describe the two-sample proportional hazards model. (Elaborate and describe model components and model assumption.)

① Let $X = 0, 1$ be the group indicator, the two-sample PHM assumes that $\lambda(t; X) = \lambda_0(t) e^{\beta}$, where $\beta \in \mathbb{R}$ and $\lambda_0(t)$ is the unspecified baseline hazard.

② Let $\lambda_0(t)$ be the hazard for the control group, and $\lambda_1(t)$ be that for the treatment group.
then two-sample PHM assumes that
$$\lambda_1(t) = \lambda_0(t) e^{\beta}$$
(both are correct).

(b) Suppose the two-sample proportional hazards model is used with treatment indicator, $x = 0$ or 1, as the covariate. Describe the independent censoring assumption which is required for use of the partial likelihood method under the two-sample proportional hazards model.

Censoring mechanism is independent of time-to-event outcome within control group and within treatment group.
(Conditional independent censoring)

(c) Describe a method to verify the validity of the two-sample proportional hazards model assumption. (Hint: model diagnostics for constant hazard ratio.)

Use Kaplan-Meier estimator to estimate the survival function for the treatment and control group respectively. Calculate the ratio of log estimated survivals, and plot ratio against time. If the calculated ratio fluctuates around a positive constant, then the two-sample PHM assumption likely holds.

3. (25 points) Suppose T is a continuous failure time and $x = 0, 1$ indicates treatment/control assignment. Consider the following proportional hazards models with an arbitrary baseline hazard function $\lambda_0(t)$: For $t > 0$,

(1) $\lambda(t; x) = \lambda_0(t)e^{\beta x}$, $-\infty < \beta < \infty$.

(2)$\lambda(t; x) = \lambda_0(t)e^{\beta x+\gamma(1-x)}$, $-\infty < \beta < \infty, -\infty < \gamma < \infty$

(3) $\lambda(t; x) = \lambda_0(t)e^{(\beta t)x}$, $-\infty < \beta < 0$

(4) $\lambda(t; x) = \lambda_0(t)e^{\alpha x+\beta xI(t\geq 5)+\gamma xI(t\geq 10)}$, $-\infty < \alpha < \infty, -\infty < \beta < \infty, -\infty < \gamma < \infty$

(5) $\lambda(t; x) = \lambda_0(t)e^{(\beta_0+\beta_1 t)x}$, $-\infty < \beta_0 < \infty,\ \beta_1 > 0$

(6) $\lambda(t; x) = \lambda_0(t)e^{(\beta_0+\beta_1 t+\beta_2 t^2)x}$, $-\infty < \beta_0 < \infty,\ -\infty < \beta_1 < \infty, -\infty < \beta_2 < \infty$

Please answer the following questions:
(Multiple choices; no explanation required.)

(a) Which models imply constant relative hazard between the treatment and control groups?

1, 2

______________________________________.

(b) Which models allow for the possibility of cross-over(s) in hazards for the two groups?

4, 5, 6

______________________________________.

(c) Identify those models that the log rank test can be used for testing the equivalence in hazards between the treatment and control groups.

1, 2, 3

______________________________________.

(d) Identify models where the log rank test is the most powerful test for testing the equivalence in hazards between the treatment and control groups.

1, 2 (Log rank test is the most powerful when hazard ratio is constant over time)

______________________________________.

(e) For Model (4), find the relative hazard between the treatment and control groups for the following 3 time periods: $0 < t < 5$, $5 \leq t < 10$, and $t \geq 10$. (Just provide the relative hazard; no explanation required.)

e^{α} for $0<t<5$, $e^{\alpha+\beta}$ for $5\leq t<10$, $e^{\alpha+\beta+\gamma}$ for $t\geq 10$.

______________________________________.

(also considered correct if you use $e^{-\alpha}$, $e^{-\alpha-\beta}$, $e^{-\alpha-\beta-\gamma}$)

4. (20%) Choose correct answers (multiple choices): a, c, d, f

The log-rank test is a statistic which is used to test survival time difference between two groups based on survival data. Please choose correct statements:

(a) Log-rank test is a nonparametric test.

(b) The log-rank test assumes a proportional hazard model.

(c) The log-rank test should not be used when a cross-over in the hazard functions of the two groups occurs.

(d) The log-rank test should not be used when a cross-over in the survival functions of the two groups occurs.

(e) The log-rank test is the most powerful hypothesis testing for testing the equivalence in hazards betweeen two groups.

(f) It is possible to have a cross-over between two hazard functions but there is no cross-over between the corresponding survival functions.

(g) It is possible to have a cross-over between two survival functions but there is no cross-over between the corresponding hazard functions.

(h) The calculation of p-value of log-rank test (or weighted log-rank tests) depends on large-sample normal approximation property. Thus, to use this test statistic, the sample size should be as large as possible. this is false.

5. (20%) Suppose a certain population consists of 55% males and 45% females. Let X be an indicator where $X = 0$ represents female and $X = 1$ represents male. Let T denote the failure time of interest and C denote the potential censoring time. Let $S_X(t)$, $X = 0, 1$, be the survival function for females ($X = 0$) and males ($X = 1$) separately. The failure time data are observed subject to the usual right-censoring. Assume that conditional on X, C follows the Uniform$(0, \theta_0 + x\theta_1)$ distribution, with $\theta_0 > 0$, $\theta_0 + \theta_1 > 0$, and $\theta_1 \neq 0$.

(a) Give the survival function, $S(t)$, for the whole population. (Hint: the whole population is a mixture of females and males.)

$$S(t) = P(T \geq t) = P(X=0)\,P(T \geq t \mid X=0) + P(X=1)\,P(T \geq t \mid X=1)$$
$$= 0.45 \cdot S_0(t) + 0.55\, S_1(t)$$

(b) Can you use the Kaplan-Meier estimator, on the basis of all the observed survival data (including data for both males and females), to estimate $S(t)$? Explain.

No. Because generally when S_0(t) is not assumed to be equal to S_1(t), the independent censoring assumption required for K-M to be used is violated. T and C both depend on gender (gender induced dependency between T and C, or gender confounds T and C).

(c) Can the Kaplan-Meier estimator be used to estimate $S_0(t)$ based on female survival data. Explain.

Yes, if given X, i.e., within a gender group, T and C are independent. If T and C are dependent even within a gender group, then no.

(You get points if you make arguments along this line of reasoning, assuming or not assuming that T and C are independent within gender group)

(d) Further assume that the hazard ratio between males and females is constant over time. Can the proportional hazard model be used to model the effect of X? Explain.

Yes, because the constant hazard ratio assumption is satisfied.

(This question does not specifically ask this, but the partial likelihood estimation can be used if the conditional independence assumption holds.)

Useful formulas or models (Not all are necessarily used in this exam.)

- Suppose T is a continuous survival time with survival function $S(t)$, hazard function $\lambda(t)$, and cumulative hazard function $\Lambda(t)$. Then, for $t > 0$, $S(t) = e^{-\Lambda(t)} = e^{-\int_0^t \lambda(u)du}$
- Suppose T follows Exponential(θ) distribution, $\theta > 0$. Then, the pdf is $f(t) = \theta e^{-\theta t} I(t > 0)$, the survival function is $S(t) = e^{-\theta t}$ for $t > 0$.
- Suppose T follows Weibull(θ, γ) distribution, $\theta > 0$, $\gamma > 0$. Then, the survival function is $S(t) = e^{-(\theta t)^\gamma}$ for $t > 0$, and the hazard function is $\lambda(t) = \gamma\theta(\theta t)^{\gamma-1}$ for $t > 0$.
- Accelerated Failure Time Model (AFTM):
  $\log(T_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \sigma\epsilon_i = \beta_0 + \boldsymbol{\beta}'\boldsymbol{x_i} + \sigma\epsilon_i$, where $\beta_0, \cdots, \beta_p$ are the regression coefficients of interest, σ is a scale parameter and ϵ_i is the random disturbance variable with unit variance, usually assumed to be iid, mean 0, and with pdf $f(\epsilon)$.

  An alternative way to formulate the AFTM is
  $T_i = T_{0i} \cdot e^{\boldsymbol{\beta x_i}} \Longleftrightarrow \log T_i = \boldsymbol{\beta}'\boldsymbol{x_i} + \log T_{0i}, \quad \log T_{0i} \sim S_0,$
  where $\{T_{0i}\}$ are iid, $\mathrm{E}[\log T_{0i}] = \beta_0$ and S_0 is the survival function for $\sigma\epsilon_i + \beta_0$.
- Proportional hazards model (PHM): $\lambda(t; \boldsymbol{x_i}) = \lambda_0(t)e^{\beta_1 x_{i1} + \beta_2 x_{i2} + \ldots \beta_p x_{ip}} = \lambda_0(t)e^{\boldsymbol{\beta}'\boldsymbol{x_i}}$, where $\boldsymbol{x_i}$ is a $p \times 1$ vector of covariates, $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters, and $\lambda_0(t)$ is an arbitrary baseline hazard function.

  The PHM can also be expressed as

  $$S(t; \boldsymbol{x_i}) = S_0(t)^{\exp(\boldsymbol{\beta x_i})}$$
- The function $I(B)$ is the indicator/characteristic function which has value 1 when event B is satisfied, and has value 0 otherwise. Example: $I(t > 0) = 1$ if $t > 0$, and $I(t > 0) = 0$ if $t \leq 0$.