Department of Mathematics and Statistics
King Fahd University of Petroleum & Minerals

AS476: Survival Models for Actuaries

Dr. Ridwan A. Sanusi

Major 2 Exam Term 251

December 23, 2025

Name:

ID#:

Instructions.

1. Please turn off your cell phones and place them under your chair. Any student caught with mobile phones on during the exam will be considered under the cheating rules of the University.
2. If you finish the test earlier and want to leave the room, please do so quietly so not to disturb others taking the test. No two person can leave the room at the same time. No extra time will be provided for the time missed outside the classroom.
3. Only materials provided by the instructor can be present on the table during the exam.
4. Do not spend too much time on any one question. If a question seems too difficult, leave it and go on.
5. Use the blank portions of each page for your work. Extra blank pages can be provided if necessary. If you use an extra page, indicate clearly what problem you are working on.
6. Only answers supported by work will be considered. Unsupported guesses will not be graded.
7. While every attempt is made to avoid defective questions, sometimes they do occur. In the rare event that you believe a question is defective, the instructor cannot give you any guidance beyond these instructions.
8. Mobile calculators, I-pad, or communicable devices are disallowed. Use regular scientific calculators or financial calculators only. Write important steps to arrive at the solution of the following problems.
9. Record your final answers to the multiple-choice questions on the OMR sheet. And write important steps to arrive at the solution of the last two problems.

The test is 110 minutes, GOOD LUCK, and you may begin now!

INTENTIONALLY LEFT BLANK

(Use this page for your work if needed)

Questions

1. (5 points) **Kaplan-Meier Estimator**

   The claim payments on a sample of ten insurance policies are:

   $$2,\ 3,\ 3,\ 5,\ 5^+,\ 6,\ 7,\ 7^+,\ 9,\ 10^+$$

   where $^+$ indicates that the loss exceeded the policy limit (right censored).

   Using the Kaplan-Meier Product-Limit estimator, calculate the probability that the loss on a policy exceeds 8.

   A. 0.20 B. 0.24 C. 0.30 D. 0.36 E. 0.40

   Final answer (1 point) ________________________________

   Work shown (4 points)

2. (7 points) **Cox Regression Analysis**

   A computer analysis on leukemia patients ($n = 42$) shows the following results:

Variable name	Coef.	Std. Err.	$p > \|z\|$	Haz. Ratio	[95% Conf. interval]
Log WBC	1.170	0.499	0.019	3.222	1.213 to 8.562
Rx	0.267	0.566	0.637	1.306	0.431 to 3.959
Sex × Log WBC	0.469	0.720	0.515	1.598	0.390 to 6.549
Sex × Rx	1.592	0.923	0.084	4.915	0.805 to 30.003

   No. of subjects = 42 Log likelihood = −55.835 Stratified by Sex

   (a) (3 points) Write the Cox model represented by the computer printout above.

   (b) (4 points) If you know that a reduced model with only Log WBC and Rx as predictors results in a log likelihood of −57.560, at a 5% significance level, what can you conclude about the full model?

3. (10 points) **Cox Proportional Hazards Models with R Data**

Use the R datafile "ovarian" to fit Cox proportional hazards models on futime as the dependent survival time. The edited output of computer results for this analysis are given as follows:

Model 1

Variable	coef	exp(coef)	se(coef)	z	p
age	0.162	1.18	0.0497	3.25	0.0012

Likelihood ratio test = 14.3 on 1 df, p = 0.000156
n = 26, number of events = 12

Model 2

Variable	coef	exp(coef)	se(coef)	z	p
rx			0.6320	-1.27	
age	0.147	1.159	0.0461	3.19	0.0014

Likelihood ratio test = 15.9 on 2 df, p = 0.000355
n = 26, number of events = 12

(a) (6 points) Using model 2, give an expression for the estimated survival curve for persons with rx = 1, adjusted for AGE. Also, give an expression for the estimated survival curve for persons with rx = 2, adjusted for AGE.

(b) (4 points) What is your overall conclusion about the effect of rx on survival time based on the computer results provided from this study?

4. (11 points) **General Insurance Claims Analysis**

For the claims on general insurance data (Gray and Pitts, 2012), there are $n = 140$ claims for each of four different possible types of policies. The variables are:

Variable #	Variable name	Coding
1	Claim amounts	loss are nonnegative
2	C60	60% coinsurance = 1, other = 0
3	Deductible	franchise Deductible ($250) = 1, other = 0
4	Ground up	Ground up = 1, other = 0
5	Limit	Limit ($10000) = 1, other = 0
6	Payment status	Paid at loss amount = 1, censored at policy limit = 0

For this actuarial survival model, claim amounts is treated as the survival variable and the payment status of 1 is regarded as the event.

The following is an edited R survival analysis results on the claim amounts. $n = 544$, missing = 16, number of events = 536

Variable	Coef.	Std.Err.	z	p $>\|z\|$	exp(coef)	[95% Conf. Interval]
Deductible	-0.4918	0.1241	-3.962	7.43×10^{-5} ***	0.6115	0.4795 to 0.7800
Ground up	-0.3768	0.1203	-3.132	0.00174 **	0.6861	0.5420 to 0.8685
Limit	-0.3645	0.1218	-2.992	0.00277 **	0.6945	0.5470 to 0.8818

Log likelihood = -2855.656
Likelihood ratio test = 17.31 on 3 df, p = 0.0003394

(a) (2 points) Write the Cox PH model represented by the output.

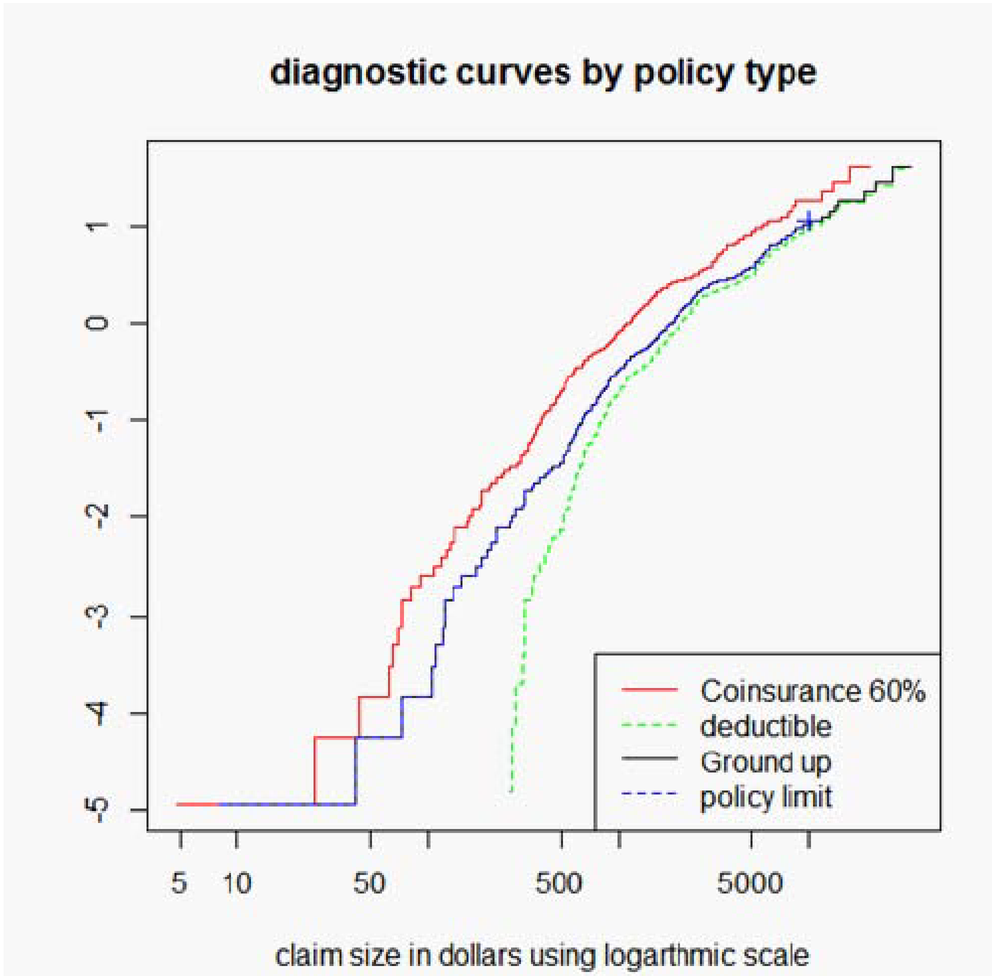
(b) (3 points) State the null hypothesis and carry out a test for the overall significance of the model?

(c) (3 points) Test for the variable "Deductible". What is your conclusion at the 0.05 significance level?

(d) (3 points) Obtain the Hazard Ratio for "Deductible" and interpret this Hazard Ratio.

5. (5 points) **Cox PH Model Assumptions**

For the general insurance claim data by Gray and Pitts (2012) described in Question 4, you created the following graph to investigate the assumptions of the Cox PH model:



In addition, you conducted a summary correlation analysis between the *Schoenfeld residuals* and the survival variable (*claim amounts*) for each predictor as follows:

Variable	Correlation	Test Statistic	p-value
Deductible	0.1234	7.95	0.0048
Ground up	0.0552	1.62	0.2038
Limit	0.0604	1.93	0.1646
GLOBAL	NA	7.98	0.0464

(a) (3 points) What should be the likely label for the y-axis in the above graph? And what is the graph indicating?

(b) (2 points) Using the information provided, what can you conclude regarding the *PH assumption* for the variables used in the model? Explain briefly.

6. (11 points) **Stratified Cox Model - Veteran's Administration Lung Cancer Trial**

The dataset "vets.dat" considers survival times in days for 137 patients from the Veteran's Administration Lung Cancer Trial (cited in Kalbfleisch and Prentice, 2002). The exposure variable of interest is treatment status. Other variables of interest as control variables are cell type (four types, defined by dummy variables), performance status, and prior therapy status. Failure status defined by the status variable (0 if censored, 1 if died). A complete list of the variables is given below.

Variable	**Description**
Treatment	(standard = 1, test = 2)
Cell type 1 (large)	large = 1, other = 0
Cell type 2 (adeno)	adeno = 1, other = 0
Cell type 3 (small)	small = 1, other = 0
Cell type 4 (squamous)	squamous = 1, other = 0
Survival time	Survival time (days)
Performance status	0 = worst, ..., 100 = best
Prior therapy	none = 0, some = 10
Status	0 = censored, 1 = died

Based on your earlier analysis, you have decided to stratify on the "Prior therapy" variable as given by the edited R output below:

$n = 137$, number of events = 128

Variable	Coef.	Std.Err.	z	p $>\|z\|$	exp(coef)	exp(-coef)	[95% Conf. In
Treatment	0.360776	0.232978	1.275	0.202230	1.434442	0.6971	0.9238 to 2.
Large cell	0.273083	0.201358	1.356	0.175034	1.314069	0.7610	0.8355 to 1.
Adeno cell	1.111310	0.296788		0.000181 ***	3.038327	0.3291	
Small cell	0.801184	0.268932	2.974	0.002935 **		0.4488	1.3142 to 3.
Perf status	-0.030671	0.009245	-5.844	5.09×10^{-4} ***	0.969795	1.0911	0.9599 to 0.

Stratified by "Prior therapy"

Signif. codes: 0 = *****, 0.001 = ****, *** = 0.01, ** = 0.05, * = 0.1

R-squared = 0.342 (max possible = 0.998)

Likelihood ratio test = 57.34 on 5 df, $p = 4.31 \times 10^{-11}$

Wald test = 55.46 on 5 df, $p = 2.928 \times 10^{-11}$

Score (logrank) test = 61.61 on 5 df, $p = 5.635 \times 10^{-12}$

Answer the following questions.

(a) (2 points) What reason from the previous analysis would have driven you to do the above analysis by stratifying on "Prior therapy"?

(b) (6 points) Complete the missing blank cells above. Be sure to show your work.

(c) (3 points) Write the stratified Cox model and write this model for (i) No Prior therapy (None) and (ii) Some Prior therapy (some).

7. (7 points) **Extended Cox Model with Time-Dependent Covariate**

The dataset "vets.dat" from Question 6 is analyzed again. This time, you introduced the variable $\ln t \times ps$ where ps = "performance status". Below are edited R outputs for different Cox models:

Model 1

Variable	Coef.	Std.Err.	z	p >$\|z\|$	exp(coef)	exp(-coef)	[95% Conf. In
Treatment	0.257213	0.206259	1.282	0.1997	1.293450	0.7721	0.8729 to 1
Large cell	0.393959	0.292232	1.392	0.1628	1.491539	0.6751	0.8520 to 2
Adeno cell	1.147672	0.293922	3.905	9.97×10^{-5} ***	3.150532	0.3174	1.7676 to 5
Small cell	0.855614	0.265909	3.218	0.0013 **	2.353924	0.4248	1.3971 to 3
Perf status	-0.031112	0.004557	-6.829	8.66×10^{-12} ***	0.969927	1.0310	0.9609 to 0

Log likelihood = -475.676

Model 2 (with interaction)

Variable	Coef.	Std.Err.	z	p >$\|z\|$	exp(coef)	exp(-coef)	[95% Conf. In
Treatment	0.117201	0.165974	0.707	0.47972	1.124458	0.8892	0.8122 to 1.5
Large cell	0.353490	0.245124	1.442	0.14934	1.424167	0.7021	0.8802 to 2.3
Adeno cell	1.132695	0.261054	4.339	1.43×10^{-5} ***	3.104910	0.3220	1.8604 to 5.1
Small cell	0.898314	0.273413	3.285	0.00102 **	2.455036	0.4073	1.4365 to 4.1
Perf status	-0.069272	0.019185	-3.611	0.000306 ***	0.933864	1.0709	0.8992 to 0.9
$\ln t \times ps$	0.013277	0.004890	2.715	0.00663 **	1.013367	0.9868	1.0037 to 1.0

Log likelihood = -470.4999

(a) (2 points) For Model 2, give an expression for the hazard ratio for the effect of the treatment variable adjusted for all other predictors.

(b) (3 points) Using Model 2, compute the estimated hazard for patients with large cells and performance status of 40 at 100 days. Also compute the hazard for patients with large cells and performance status of 50 at 120 days.

(c) (2 points) Carry out an appropriate test of hypothesis to evaluate whether there is any significant interaction in Model 2. What is your conclusion?

8. (5 points) **Cox PH Model Maximum Likelihood Estimation**

You are given:

- A Cox proportional hazards (PH) model was used to study claim amounts on two groups of general insurance policies.
- A single predictor variable Z was used with $Z = 0$ for a policy in Group A and $Z = 1$ for a policy in Group B.
- A sample of three policies was taken from each group. The losses were:
  - Group A: 75, 125, 200
  - Group B: 15, 50, 100
- The baseline survival function is $S_0(s) = \left(\frac{200}{200+s}\right)^{\alpha}, \ s > 0, \ \alpha > 0.$

Calculate the maximum likelihood estimate of the Cox PH model coefficient β.

A. -0.92 B. -0.40 C. 0.40 D. 0.92 E. 2.51

Final answer (1 point) ______________________________

Work shown (4 points)

9. (5 points) **Given Data**

The following survival data is given:

ID	TIME	STATUS	SMOKE
Barry	2	1	1
Gary	3	1	0
Harry	5	0	0
Larry	8	1	1

Where:

- TIME = Survival time (in years)
- STATUS = 1 for event, 0 for censorship
- SMOKE = 1 for a smoker, 0 for a nonsmoker

Question

Derive an expression for the Cox Likelihood for this dataset.

Solutions

1. **Kaplan-Meier Estimator** (5 points)

   Ordered data with censoring:

   Time: 2 3 3 5 5^+ 6 7 7^+ 9 10^+

   Event: 1 1 1 1 0 1 1 0 1 0

   Kaplan-Meier calculation:

t_i	n_i	d_i	$1 - \frac{d_i}{n_i}$	$\hat{S}(t)$
2	10	1	$\frac{9}{10}$	0.900
3	9	2	$\frac{7}{9}$	$0.900 \times \frac{7}{9} = 0.700$
5	7	1	$\frac{6}{7}$	$0.700 \times \frac{6}{7} = 0.600$
5^+	6	0	-	0.600
6	5	1	$\frac{4}{5}$	$0.600 \times \frac{4}{5} = 0.480$
7	4	1	$\frac{3}{4}$	$0.480 \times \frac{3}{4} = 0.360$
7^+	3	0	-	0.360
9	2	1	$\frac{1}{2}$	$0.360 \times \frac{1}{2} = 0.180$
10^+	1	0	-	0.180

   The probability that the loss exceeds 8 is $\hat{S}(8)$. Since the last event before 8 is at time 7, and the next event is at time 9:

   $\hat{S}(8) = \hat{S}(7) = 0.360$

   Alternative calculation:

$$\hat{S}(8) = \frac{9}{10} \times \frac{7}{9} \times \frac{6}{7} \times \frac{4}{5} \times \frac{3}{4} = \frac{9 \times 7 \times 6 \times 4 \times 3}{10 \times 9 \times 7 \times 5 \times 4} = \frac{6 \times 3}{10 \times 5} = \frac{18}{50} = 0.36$$

   Final answer: $\boxed{D}$ (0.36)

   Source: SOA C exam (2014 infiniteActuary Exercise ch2 Q6 [4.F02.25] – Kaplan-Meier Estimator) *chap 1 KK*

2. **Cox Regression Analysis** (7 points)

   (a) The Cox model represented by the computer printout is:

$$h_g(t, X) = h_{0_g}(t) \exp\left[\beta_1 \cdot \text{Log WBC} + \beta_2 \cdot \text{Rx} + \beta_3 \cdot (\text{Sex} \times \text{Log WBC}) + \beta_4 \cdot (\text{Sex} \times \text{Rx})\right]$$

   where g indicates stratification by Sex.

(b) We compare the full model (with interactions) against the reduced model (without interactions) using the likelihood ratio test:

$$\begin{aligned} LR &= -2\ln L_R - (-2\ln L_F) \\ &= -2(-57.560) - (-2(-55.835)) \\ &= 115.120 - 111.670 = 3.45 \end{aligned}$$

Comparing against $\chi^2_{0.95}(2) = 5.9915$ (2 degrees of freedom for the two interaction terms), we find:

$$3.45 < 5.9915$$

Therefore, we do not reject the null hypothesis $H_0 : \beta_3 = \beta_4 = 0$. We conclude that the reduced model (without interactions) is acceptable for our data at the 5% significance level.

Source: KK Survival Analysis Chapter 5 Example pg 212-218

3. **Cox Proportional Hazards Models with R Data** (10 points)

(a) From Model 2, we have the Cox model:

$$h(t) = h_0(t)\exp(\beta_1 \cdot rx + \beta_2 \cdot \text{age})$$

where $\beta_1 = -0.804$ and $\beta_2 = 0.147$.
The estimated survival function is:

$$S(t, X) = [S_0(t)]^{\exp(\beta_1 \cdot rx + \beta_2 \cdot \text{age})}$$

We need to recode rx values: rx of 1 as rx = 0 and rx of 2 as rx = 1.
For rx = 1 (coded as 0):

$$S(t, X) = [S_0(t)]^{\exp(\beta_1 \cdot 0 + \beta_2 \cdot \text{age})} = [S_0(t)]^{\exp(0.147 \cdot \text{age})}$$

For rx = 2 (coded as 1):

$$S(t, X) = [S_0(t)]^{\exp(\beta_1 \cdot 1 + \beta_2 \cdot \text{age})} = [S_0(t)]^{\exp(-0.804 + 0.147 \cdot \text{age})}$$

(b) For the effect of rx on survival time:

- The p-value for rx is 0.20, which is greater than the typical significance level of 0.05
- We test $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$
- Since p-value $= 0.20 > \alpha = 0.05$, we do not reject the null hypothesis
- Conclusion: rx is not statistically significant in predicting survival time after adjusting for age

Source: R ovarian dataset analysis

4. **General Insurance Claims Analysis** (11 points)

   (a) The Cox PH model represented by the output is:

$$h(t, X) = h_0(t) \exp\left(\beta_1 \cdot \text{Deductible} + \beta_2 \cdot \text{Ground up} + \beta_3 \cdot \text{Limit}\right)$$

$$= h_0(t) \exp\left(-0.4918 \cdot \text{Deductible} - 0.3768 \cdot \text{Ground up} - 0.3645 \cdot \text{Limit}\right)$$

   (b) **Null hypothesis:** $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ (all coefficients are zero)
   Test: Likelihood ratio test = 17.31 on 3 degrees of freedom
   Critical value: $\chi^2_{0.95}(3) = 7.8147$
   Conclusion: Since 17.31 >7.8147, we reject H_0 at $\alpha = 0.05$. The model with three predictors (Deductible, Ground up, and Limit) is statistically significant.

   (c) **Test for "Deductible":**

   - **Null hypothesis:** $H_0 : \beta_1 = 0$
   - **Wald test statistic:** $z = -3.962$
   - **Critical value:** $z_{0.025} = -1.96$ (two-tailed test)
   - **Conclusion:** Since $-3.962 < -1.96$, we reject H_0 at $\alpha = 0.05$. The variable "Deductible" is statistically significant.

   (d) **Hazard Ratio for "Deductible":**

$$HR = \exp(\beta_1) = \exp(-0.4918) = 0.6115$$

   Interpretation: The hazard ratio of 0.6115 indicates that policies with a franchise deductible have about 61.15% of the hazard (risk of claim payment) compared to the reference group (60% coinsurance policies). In other words, deductible policies have approximately 38.85% lower hazard of claim payment compared to coinsurance policies.

Source: Gray and Pitts (2012) general insurance data analysis

5. **Cox PH Model Assumptions** (5 points)

   (a) The likely label for the y-axis in the graph is **"Scaled Schoenfeld Residuals"** or **"Beta(t) for [Variable Name]"**.
   The graph is showing the **scaled Schoenfeld residuals plotted against time** for each variable in the Cox PH model. This type of graph is used to check the proportional hazards assumption. If the PH assumption holds, the residuals should show no systematic pattern and fluctuate randomly around zero.

   (b) Based on the Schoenfeld residuals correlation analysis:

   - **Deductible:** p-value = 0.0048 <0.05, so the PH assumption is violated for this variable

- **Ground up:** p-value = 0.2038 >0.05, so the PH assumption is satisfied
- **Limit:** p-value = 0.1646 >0.05, so the PH assumption is satisfied
- **GLOBAL:** p-value = 0.0464 <0.05, indicating overall violation of PH assumption

Conclusion: The PH assumption is violated for the "Deductible" variable and for the model overall. The variables "Ground up" and "Limit" individually satisfy the PH assumption. Further analysis (such as using time-dependent covariates or stratified analysis) may be needed for the "Deductible" variable.

Source: KK Survival Analysis Chapter 4

6. **Stratified Cox Model - Veteran's Administration Lung Cancer Trial** (11 points)

(a) The reason for stratifying on "Prior therapy" is that this variable **violates the proportional hazards assumption**. Previous diagnostic tests (such as Schoenfeld residuals test) likely showed that the effect of prior therapy changes over time, making it necessary to use stratification rather than including it as a regular covariate in the Cox model.

(b) **Missing values completion:**

- **Adeno cell z-value:** $z = \frac{\text{Coef}}{\text{Std.Err}} = \frac{1.111310}{0.296788} = 3.744$
- **Adeno cell 95% CI:**

$$\text{Lower bound} = \exp(1.111310 - 1.96 \times 0.296788) = \exp(0.529605) = 1.6982$$
$$\text{Upper bound} = \exp(1.111310 + 1.96 \times 0.296788) = \exp(1.693015) = 5.4388$$

- **Small cell exp(coef):** $\exp(0.801184) = 2.228178$

Completed table:

Var	Coef.	Std.Err.	z	p $>\lvert z\rvert$	exp(coef)	exp(-coef)	[95% Conf. Inter
Trt	0.3607	0.2330	1.275	0.202230	1.4344	0.6971	0.9238 to 2.497
L C	0.2731	0.2014	1.356	0.175034	1.3141	0.7610	0.8355 to 1.949
A C	1.1113	0.2968	3.744	0.000181 ***	3.0383	0.3291	1.6982 to 5.438
S C	0.8011	0.2689	2.974	0.002935 **	2.2282	0.4488	1.3142 to 3.777
P S	-0.0307	0.0092	-5.844	5.09×10^{-4} ***	0.9698	1.0911	0.9599 to 0.979

(c) **Stratified Cox model:**

General form: $h_g(t, X) = h_{0g}(t) \exp(\beta_1 \cdot \text{Treatment} + \beta_2 \cdot \text{Large cell} + \beta_3 \cdot \text{Adeno cell} + \beta_4 \cdot \text{Small cell} + \beta_5 \cdot \text{Perf status})$

Where g indicates the stratum (prior therapy group).

- **(i) No Prior therapy (None):** $h_1(t, X) = h_{01}(t) \exp(0.360776 \cdot \text{Treatment} + 0.273083 \cdot \text{Large cell} + 1.111310 \cdot \text{Adeno cell} + 0.801184 \cdot \text{Small cell} - 0.030671 \cdot \text{Perf status})$

- **(ii) Some Prior therapy:** $h_2(t, X) = h_{02}(t)\exp(0.360776 \cdot \text{Treatment} + 0.273083 \cdot \text{Large cell} + 1.111310 \cdot \text{Adeno cell} + 0.801184 \cdot \text{Small cell} - 0.030671 \cdot \text{Perf status})$

Note: The coefficients are the same for both strata, but the baseline hazard functions $h_{01}(t)$ and $h_{02}(t)$ are different, allowing the hazard functions to have different shapes for the two prior therapy groups.

Source: KK Survival Analysis Chapter 5, pg 192

7. **Extended Cox Model with Time-Dependent Covariate** (7 points)

(a) For Model 2, the hazard ratio for the treatment variable is:

$$HR(\text{treatment}) = \exp(\beta_1) = \exp(0.117201) = 1.124458$$

This represents the hazard ratio for treatment effect, adjusted for all other predictors including the time-dependent interaction term.

(b) **For patients with large cells and performance status 40 at 100 days:**

$$h(t, X) = h_0(t)\exp(0.117201 \cdot \text{Treatment} + 0.353490 \cdot 1 - 0.069272 \cdot 40 + 0.013277 \cdot \ln(100) \cdot 40)$$

$$= h_0(t)\exp(0.117201 \cdot \text{Treatment} + 0.353490 - 2.77088 + 0.013277 \cdot 4.60517 \cdot 40)$$

$$= h_0(t)\exp(0.117201 \cdot \text{Treatment} - 2.41739 + 2.445)$$

$$= h_0(t)\exp(0.117201 \cdot \text{Treatment} + 0.02761)$$

For patients with large cells and performance status 50 at 120 days:

$$h(t, X) = h_0(t)\exp(0.117201 \cdot \text{Treatment} + 0.353490 \cdot 1 - 0.069272 \cdot 50 + 0.013277 \cdot \ln(120) \cdot 50)$$

$$= h_0(t)\exp(0.117201 \cdot \text{Treatment} + 0.353490 - 3.4636 + 0.013277 \cdot 4.78749 \cdot 50)$$

$$= h_0(t)\exp(0.117201 \cdot \text{Treatment} - 3.11011 + 3.178)$$

$$= h_0(t)\exp(0.117201 \cdot \text{Treatment} + 0.06789)$$

(c) **Likelihood ratio test for interaction:**

$$LR = -2\ln L_{\text{reduced}} - (-2\ln L_{\text{full}}) = -2(-475.676) - (-2(-470.4999)) = 10.3522$$

Comparing against $\chi^2_{0.95}(1) = 3.841$, since $10.3522 > 3.841$, we reject $H_0 : \beta_6 = 0$.
Conclusion: The interaction term $\ln t \times ps$ is statistically significant at the 0.05 level, indicating that the effect of performance status changes over time.

Source: KK Survival Analysis Chapter 6

8. **Cox PH Model Maximum Likelihood Estimation** (5 points)

The baseline survival function is $S_0(s) = \left(\frac{200}{200+s}\right)^\alpha$ with corresponding density $f_0(s) = \alpha 200^\alpha (200+s)^{-\alpha-1}$.

For Group B ($Z = 1$), the survival and density functions are:

$$S_1(s) = [S_0(s)]^{\exp(\beta)} = \left(\frac{200}{200+s}\right)^{\alpha \exp(\beta)}$$

$$f_1(s) = \alpha \exp(\beta) 200^{\alpha \exp(\beta)} (200+s)^{-\alpha \exp(\beta) - 1}$$

The log-likelihood function for the six observations is:

$$\ell(\alpha, \beta) = \sum_{i=1}^{3} \ln f_0(x_i) + \sum_{i=1}^{3} \ln f_1(y_i)$$

$$= 3\ln\alpha + 3\alpha \ln 200 - (\alpha+1)\sum \ln(200+x_i) + 3\ln(\alpha \exp\beta) + 3\alpha \exp\beta \ln 200 - (\alpha \exp\beta + 1)\sum \ln(200+y_i)$$

After simplification and numerical computation with the given data:

$$\sum \ln(200+x_i) = \ln(275) + \ln(325) + \ln(400) = 17.6544$$
$$\sum \ln(200+y_i) = \ln(215) + \ln(250) + \ln(300) = 16.5959$$

Solving the maximum likelihood equations gives $\hat{\beta} = 0.92$.

Final answer: $\boxed{D}$ (0.92)

Source: SOA 2007 Question 22

9. **Cox Likelihood** (5 points)

Step 1: Order the Event Times

First, we order the event times in increasing order. The distinct event times are:

- Time 2: Barry (event)
- Time 3: Gary (event)
- Time 8: Larry (event)

Note: Harry is censored at time 5, so he doesn't contribute an event but remains in the risk set until his censoring time.

Step 2: Risk Sets

At each event time, the risk set consists of all individuals who are still under observation (not yet censored and not yet had an event):

- **At time 2**: Risk set = {Barry, Gary, Harry, Larry}
- **At time 3**: Risk set = {Gary, Harry, Larry} (Barry had event at time 2)
- **At time 8**: Risk set = {Harry, Larry} (Gary had event at time 3, Harry is censored at time 5 but still in risk set at time 3)

Note: Harry is censored at time 5, so he is NOT in the risk set at time 8.

Step 3: Cox Partial Likelihood

The Cox partial likelihood is given by:

$$L(\beta) = \prod_{i=1}^{k} \frac{\exp(\beta' z_{(i)})}{\sum_{j \in R(t_{(i)})} \exp(\beta' z_j)}$$

Where:

- k = number of events (3 in this case)
- $t_{(i)}$ = ordered event times
- $z_{(i)}$ = covariate vector for the individual who fails at time $t_{(i)}$
- $R(t_{(i)})$ = risk set at time $t_{(i)}$
- β = regression coefficient for SMOKE

Step 4: Likelihood Expression for Given Data

Let β be the coefficient for SMOKE. Then the likelihood is:

$$L(\beta) = \frac{\exp(\beta \cdot \text{SMOKE}_{\text{Barry}})}{\sum_{j \in R(t=2)} \exp(\beta \cdot \text{SMOKE}_j)} \times \frac{\exp(\beta \cdot \text{SMOKE}_{\text{Gary}})}{\sum_{j \in R(t=3)} \exp(\beta \cdot \text{SMOKE}_j)} \times \frac{\exp(\beta \cdot \text{SMOKE}_{\text{Larry}})}{\sum_{j \in R(t=8)} \exp(\beta \cdot \text{SMOKE}_j)}$$

Substituting the SMOKE values and risk sets:

$$L(\beta) =$$

$$\frac{\exp(\beta \cdot 1)}{\exp(\beta \cdot 1) + \exp(\beta \cdot 0) + \exp(\beta \cdot 0) + \exp(\beta \cdot 1)} \times \frac{\exp(\beta \cdot 0)}{\exp(\beta \cdot 0) + \exp(\beta \cdot 0) + \exp(\beta \cdot 1)} \times \frac{\exp(\beta \cdot 1)}{\exp(\beta \cdot 1)}$$

Simplifying:

$$L(\beta) = \frac{e^\beta}{e^\beta + 1 + 1 + e^\beta} \times \frac{1}{1 + 1 + e^\beta} \times \frac{e^\beta}{e^\beta}$$

Final expression:

$$\boxed{L(\beta) = \frac{e^\beta}{2e^\beta + 2} \times \frac{1}{e^\beta + 2}}$$