IAQS | INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

Subject: SRM 1

Chapter: Unit 2

Category: Practice Questions

Proportional Hazard Models

1. CT4 April 2015 Q3

(i) Explain what is meant by a proportional hazards model.

(ii) Outline three reasons why the Cox proportional hazards model is widely used in empirical work.

2. CT4 September 2015 Q6

(i) Describe what is meant by a proportional hazards model.

A pharmaceutical company is interested in testing a new treatment for a debilitating but non-fatal condition in cows. A randomised trial was carried out in which a sample of cows with the condition was assigned to either the new treatment or the previous treatment. The event of interest was the recovery of a cow from the condition. The results were analysed using a Cox regression model.

The final model estimated the hazard, h(t, x) as:

$$h(t, x) = h_0(t) \exp(\beta_0 z + \beta_1 x + \beta_2 xz),$$

where:

$h_0(t)$ is the baseline hazard;

z is a covariate taking the value 1 if the cow was assigned the new treatment and 0 if the cow was assigned the previous treatment;

x is a covariate denoting the length of time (in days) for which the cow had been suffering from the condition when treatment was started; and t is the number of days since treatment started.

β0, β1 and β2 are parameters.
Their estimated values were β0 = 0.8, β1 = 0.4 and β2 = − 0.1.

(ii) Determine the characteristics of the baseline cow.

For a particular cow, the new treatment and the previous treatment have exactly the same hazard.

(iii) Calculate the number of days for which that cow had the condition before the initiation of treatment.

Under the previous treatment, cows whose treatment began after they had been suffering from the condition for three days had a median recovery time of 14 days once treatment had started.

(iv) Calculate the proportion of these cows which would still have had the condition after 14 days if they had been given the new treatment.

Ans: (ii) A cow who started the previous treatment immediately the condition appeared. (iii) 8 days. (iv) 0.319

3. CT4 April 2016 Q11

An energy provider is worried about the number of its customers who transfer to other companies within the first two years of their contract and is trying to direct its advertising towards the most loyal section of the population.

The company has looked at its records over recent years and has fitted a Cox proportional hazards model to those who have transferred within the first two years using the factors which appear to have the most impact on early transfer rates.

The following figures have been derived from the data:

	Factor	*Parameter Estimate*	*Variance*
Gender	Male	–0.25	0.015
	Female	0	0
Volume of energy consumed	High	0.32	0.008
	Low	0	0
Area of Residence	City Centre	0.19	0.012
	City (not centre)	0	0
	Rural	–0.35	0.005

(i) Give the hazard function for this Cox proportional hazard model defining all the terms and covariates.

(ii) State the features of the person to whom the baseline hazard applies.

(iii) Calculate symmetric 95% confidence intervals for the parameters based on the standard errors.

(iv) Test the suggestion that women change energy providers more frequently than men.

There is a 70% probability that a male customer who is a low consumer of energy and lives in a rural area has transferred providers before the end of two years.

(v) Calculate the probability that a male customer who is a high consumer of energy and lives in a city center remains with the company for at least two years.

(vi) Set out how you would determine whether the effect of any of the factors depends upon any of the other factors.

Ans: (ii) Female, low energy consumption, lives in a city but not in the city centre. (iii) Male (−0.4900, −0.0100), High consumption (0.1447, 0.4953), City Centre (−0.0247, 0.4047), Rural (−0.4886, −0.2114), (v) 0.058124

4. CT4 September 2016 Q10

A researcher is investigating the contributing factors to the speed at which patients recover from a common minor surgical procedure undertaken in hospitals across the country. He has the questionnaires which each patient completed before the surgery and the length of time the patient remained in hospital after surgery and is attempting to fit a Cox proportional hazards model to the data.

He has fitted a model with what he assumes are the most common contributing factors and has calculated the parameters as shown in the table below:

Covariate	*Category*	*Parameter*
Gender	Male	0
	Female	0.065
Smoker	Non Smoker	−0.035
	Smoker	0
Drinker	Non Drinker	−0.06
	Moderate Drinker	0
	Heavy Drinker	0.085

(i) Give the hazard function for this Cox proportional hazards model, defining all the terms and covariates.

A male moderate drinker who does not smoke has a hazard of leaving hospital after three days of 0.6.

UNIT 2

PRACTICE QUESTIONS

(ii) Calculate the probability that a female heavy drinker who smokes and who is still in hospital after three days is NOT discharged at that point.

A colleague suggests that, in his experience, gender has no material impact on the length of time in hospital after surgery.

(iii) Explain how the researcher could test this suggestion statistically.

Another colleague suggests that the original model is good but could be improved by including an additional factor as to whether a patient is married or not.

(iv) Set out how the researcher could establish whether an additional factor representing marital status would improve the model.

Ans: (ii) 0.2781

5. CT4 April 2017 Q7

(i) Describe the essential feature of a proportional hazards model.

A study was made of the impact of drinking beer on men aged 60 years and over. A sample of men was followed from their 60th birthdays until they died or left the study for other reasons. The baseline hazard of death, μ, was assumed to be constant, and a proportional hazards model was estimated with a single covariate: the average daily beer intake in standard-sized glasses consumed, x.

The equation of the model is: $h(t) = \mu \exp(\beta x)$
where h(t) is the hazard of death at age 60 + t.

The estimated value of μ is 0.03, and the estimated value of β is 0.2.

(ii) Explain how μ and β should be interpreted, in the context of this model.

(iii) Calculate the estimated hazard of death of a man aged exactly 62 years who drinks two glasses of beer a day.

A man is aged exactly 60 years and drinks three glasses of beer a day.

(iv) (a) Calculate the estimated probability that this man will still be alive in 10 years' time.

(b) Calculate the expectation of life at age 60 years for this man. [2]

Another man is aged exactly 60 years. He drinks beer only in his local bar. He drinks all the beer he buys and is expected to continue drinking the same amount of beer

every day until he dies. The owner of the bar is interested in selling as much beer as possible.

(v) Determine the average number of glasses of beer a day the owner must sell the man in order to maximize the total amount of beer the man buys over his remaining lifetime.

Ans: (iii) 0.0448, (iv)a) 0.579, b)18.3 years, (v) 5 glasses

6. CT4 September 2017 Q6

A pharmaceutical company is undertaking trials on a new drug which, it claims, cures a particularly uncomfortable but not life threatening condition. It has conducted extensive testing of the drug on a large group of people suffering from the condition and has noticed that the drug is much more effective in some groups of patients than others.

It has fitted a Cox regression for the hazard of symptoms disappearing h(t) with three parameters;

$$h(t) = h_0(t)\exp\left(S\beta_S + A\beta_A + G\beta_G\right)$$

where βS, βA, and βG are parameters and

• S represents the sex of the patient and takes a value of 1 if the patient is female, 0 if male.

• A represents the age, in years minus 20, of the patient when the drug was administered.

• G takes the value 1 if the patient attended a gym, 0 otherwise.

The company has discovered the following, where the age given is the age when the drug was administered:

• a 25 year old female who attended a gym had a hazard of symptoms disappearing equal to twice that of a male of the same age who did not attend a gym;

• a 45 year old male who did not attend a gym had a hazard of symptoms disappearing half that of a 43 year old male who attended a gym; and

• a 32 year old female who attended a gym had a hazard of symptoms disappearing 60% greater than that of a 45 year old female who did not attend a gym.

(i) Calculate the values of the parameters βS, βA, and βG.

(ii) Determine for which group of people the drug is most effective.

The probability that a woman who attended a gym and was aged 38 years when she was given the drug still had symptoms of the condition after 28 days was found to be 0.75.

(iii) Calculate the probability of still having symptoms after 28 days for a male aged 26 years when given the drug who did not attend a gym.

Ans: (i) βS = -0.04057, βA = 0.02029, and βG = 0.73372, (ii) male, person to be as old as possible when the drug is administered, someone attends a gym. (iii) 0.89337

7. CS2A April 2019 Q8

(i) State two advantages of the Cox regression model for assessing the impact of risk factors on a hazard.

An exercise company called FlexPexApps is developing a computer program to investigate the effect of certain factors on the incidence of a common medical condition which affects millions of people in early middle age. It has identified three factors which appear to have a large impact on the onset of the disease and has set up a Cox regression model for the hazard as follows:

$$h(t) = h_0(t) \exp(\beta_A A + \beta_E E + \beta_D D)$$

where:
• A is the age of the individual minus 40 years
• E is an exercise indicator and takes the value of 1 if the person exercises, which in this case means they follow a set regime for 30 minutes each day, and 0 otherwise
• D is a diet indicator and takes the value of 1 if the person diets, which in this case means they consume fewer than 2,000 calories per day, and 0 otherwise
• β_A, β_E and β_D are the parameters to be estimated.

From the data FlexPexApps has managed to acquire, it has established that:

• a 53 year old who exercises but does not diet has a hazard of contracting the condition half that of a 48 year old who does not exercise but diets
• a 55 year old who does not exercise but diets has a hazard of contracting the condition 1.5 times that of a 55 year old who neither exercises nor diets
• a 58 year old who diets but does not exercise has a hazard of contracting the condition double that of a 43 year old who neither diets nor exercises.

(ii) Calculate β_A, β_E and β_D.

(iii) Explain what the values you have calculated in part (ii) say about the relative impact of age, diet and exercise on contracting this affliction.

FlexPexApps has created an advertisement based on the above findings, but the Advertising Regulator has contacted them on the grounds that their model was not sufficiently complex to take into account all the relevant factors. They have suggested four additional factors which might materially impact the hazard of contracting the condition.

(iv) Explain how FlexPexApps could extend their model to see if any one of the suggested additional factors materially impacts the hazard of contracting the condition.

Ans: (ii) $\beta D = 0.405465$, $\beta A = 0.019179$, $\beta E = -0.383576$

8. CS2A September 2020 Q8

A computer manufacturing company is monitoring the quality of the electronic circuit boards it produces. It has two factories, which are located in Oxford and London. The Oxford factory was opened only 12 months ago whereas the London factory has been producing circuit boards for 10 years.

The owner of the company wants to know whether the quality of production differs materially between the two factories. In addition, the owner wants to test whether a new manufacturing process improves the production quality. Over the last 6 months, half of the circuit boards in both factories have been made using the new process.

An analyst has been employed to test whether the location and the manufacturing process affect the quality of the circuit boards. The quality is measured by recording how long each circuit board lasts before developing a fault. The analyst decides to use the Cox proportional hazards model in the analysis.

The two covariates used are:
Location: London = 0, Oxford = 1, Process: Old = 0, New = 1.

(i) State the type of circuit board that represents the baseline hazard.

During the investigation, 6,000 boards were tested, equally split between the four possible permutations of location and process. The first board to fail was one made in Oxford using the new process. No previous censoring events had occurred.

(ii) State the term in the partial likelihood expression that relates to this first failure, clearly defining all notation used.

The following coefficients were obtained from the investigation using the Cox proportional hazards model:

Location = 0.01, Process = –0.30.

(iii) Comment on the coefficient values obtained.

(iv) Identify the additional information that would be required before being able to reach a conclusion on the effects of location and process on the quality of the circuit boards produced.

Prior to the introduction of the new process, the probability that a circuit board made in London failed in the first year was estimated to be 20%. The company plans to make a batch of 10,000 circuit boards next year in London.

(v) Which one of the following options represents the best estimate of the probability that a circuit board made in London using the new process does not fail in the first year?
A 0.5927
B 0.6965
C 0.7982
D 0.8476

(vi) Which one of the following options represents the best estimate of the expected absolute difference in the number of circuit board failures in the first year if the new process is used, compared to if the old process is used?
A 18
B 476
C 1,035
D 2,073

Ans: (i) Circuit boards made in “London” using the “Old” process. (v) Answer: D
(vi) Answer: B

9. CS2A April 2022 Q4

An investigation was carried out into the effects of a newly developed medication to treat a potentially fatal illness, including an analysis of the effects of delaying the treatment after symptoms are first reported. A placebo that was designed to look like the real medication but that contained no active treatment was also used among the patients in the study to act as a control to test the effectiveness of the newly developed medication.

A Cox proportional hazards model was used to model the rate at which patients recovered from the illness.
The following two covariates were used in the model:

X1 = Medication indicator (0 = Placebo administered, 1 = Medication administered)

X2 = Treatment delay indicator
(0 = Treated on first day after symptoms developed
1 = Treated on second day after symptoms developed
2 = Treated after second day after symptoms developed)

The investigation followed a sample of 600 patients for 10 days after treatment. The number of patients in each of the six covariate groupings was as follows:

	$X_1 = 1$	$X_1 = 0$
$X_2 = 0$	100	100
$X_2 = 1$	100	100
$X_2 = 2$	100	100

The first two patients to recover were both treated with medication on the second day after symptoms developed. The first individual recovered 2 days after treatment and the second individual recovered 3 days after treatment. There were no censoring events prior to the second individual recovering.

(i) State the term in the partial likelihood expression that relates to the second individual recovering, clearly defining each of the terms you use.

Following completion of the investigation, the coefficients for the two covariates were estimated as follows:

Covariate	*Coefficient*	*Standard error*
X_1	+0.15	0.02
X_2	–0.02	0.02

(ii) Comment on the impact of the two covariates implied by these results.

Age dependent transition intensities

1. CT4 April 2015 Q5

(i) State the principle of correspondence as it applies to death rates.

A nightclub opens at 10.00 p.m. and closes at 2.00 a.m. It admits only people aged over 21 years on the production of an identity card giving date of birth.

The table below shows the number of people entering in various intervals between 10.00 p.m. and 2.00 a.m. on 30 June 2013. No-one was admitted after 1.00 a.m., and you may assume that all those who enter the premises stay until 2.00 a.m.

Year of birth	*10.00–11.30 p.m.*	*11.30–12.00 p.m.*	*12.00 p.m.–1.00 a.m.*
1989	100	300	200
1990	200	400	350
1991	150	400	300
1992	100	250	200

During the period of opening, 40 people aged 22 last birthday required medical attention for heat exhaustion.

(ii) Calculate the rate per person-hour at which those attending the night club aged 22 last birthday required medical attention for heat exhaustion, stating any assumptions you make.

Ans: 0.02045 per person hours

2. CT4 September 2015 Q4

Company A and Company B are two small insurance companies which have recently merged to form Company C. Company C is reviewing its premium rates for a whole of life product and so is conducting an analysis of mortality rates experienced.

Company A recorded the number of policies in force every 1 January using a definition of age next birthday whereas Company B recorded the number of policies in force every 1 April using an age definition of age last birthday. Both companies recorded deaths as they happened using an age definition of age last birthday.

This is the data for the most recent years.

	Company A		
Age next birthday	*Number of policies 1 Jan. 2012*	*Number of policies 1 Jan. 2013*	*Number of policies 1 Jan. 2014*
51	8,192	6,421	8,118
52	7,684	8,298	7,187
53	9,421	8,016	9,026

	Company B		
Age last birthday	*Number of policies 1 April 2012*	*Number of policies 1 April 2013*	*Number of policies 1 April 2014*
51	4,496	3,817	4,872
52	5,281	5,218	3,812
53	4,992	5,076	5,076

In the calendar year 2013 Company A recorded 28 deaths of those aged 52 last birthday and Company B recorded 17 deaths of those aged 52 last birthday.

(i) Estimate the force of mortality for the combined company for age 52 last birthday, stating all assumptions that you make.

(ii) Explain the exact age to which your estimate applies.

Ans: (i) 0.0034, (ii) 52.5 exact

3. CT4 April 2016 Q3

(i) State the principle of correspondence in the context of a mortality investigation.

A mortality investigation collects the following data:

n_x(t) = total number of policies under which death claims are made when the policyholder is aged x last birthday for each calendar year t.

P_x (t) = number of in-force policies where the policyholder was aged x nearest birthday on 1 January in year t.

(ii) (a) Derive an expression, in terms of P_x (t), for the central exposed to risk, E_x^c, corresponding to the claims data which may be used to estimate the force of mortality in year t at each age x, μ_x.

(b) State any assumptions that you make, indicating at which point in your derivation each assumption is relevant.

4. CT4 September 2016 Q8

An analysis of the number of term assurance policies in force for three companies has revealed the following information:

	Year	*Company A*	*Company B*	*Company C*
Age 50	2013	6,728	2,643	4,132
	2014	6,189	2,548	
	2015	5,962	2,496	4,630
Age 51	2013	5,987	2,333	4,012
	2014	6,002	2,417	
	2015	5,056	2,213	4,500
Age 52	2013	5,359	2,155	3,895
	2014	5,600	1,992	
	2015	4,906	2,006	4,367

- Company A has reported the number of policies in force on 1 January each year using age nearest birthday.
- Company B has reported the number of policies in force on 1 November each year using age last birthday.
- Company C has reported the number of policies in force on 31 December each year using age next birthday, but failed to provide data for 2014.

(i) Calculate the contribution to the central exposed to risk for lives age 51 last birthday for the calendar year 2014 for each company individually.

(ii) (a) State the assumptions you have made in order to perform your calculations.

(b) Explain why these assumptions were required.

Ans: (i) Company A - 5,391. Company B - 2,385. Company C - 4,013

5. CT4 September 2017 Q4

A study was conducted into the mortality of persons aged between exact ages 85 and 86 years. The study took place from 1 April 2015 to 31 March 2016. The following table shows information on 10 lives observed in the study:

Life number	*Date of 85th birthday*	*Date of death*
1	1 August 2014	–
2	1 November 2014	–
3	1 January 2015	1 February 2016
4	1 February 2015	–
5	1 March 2015	–
6	1 April 2015	1 January 2016
7	1 June 2015	1 November 2015
8	1 July 2015	–
9	1 September 2015	1 March 2016
10	1 January 2016	–

(i) Calculate a central exposed to risk for the 10 lives in the sample, working in months.

(ii) Give the maximum likelihood estimate of the mortality hazard at age 85 last birthday.

(iii) Estimate q_{85}.

Ans: (i) 73 months. (ii) 0.04110 (monthly) or 0.49315 (working annually).
(iii) CONSTANT FORCE = 0.3893, ACTUARIAL ESTIMATE = 0.3956 OR EXACT EXPOSURE = 0.4045.

6. CT4 April 2018 Q6

The National Statistics Office of a small, low income country wants to estimate recent death rates. A death registration system has allowed the National Statistics Office to estimate deaths by age nearest birthday for the ten-year period 1 January 2005 – 31 December 2014.

Censuses of this country are infrequent. A successful census was completed on 1 January 2015, but the previous reliable census was on 1 January 2002. Both censuses collected data on the population aged x last birthday by single years of age.

(i) Explain why the National Statistics Office should adjust the age definition in the census data to correspond with that of the deaths data.

Let the census population at age x last birthday on 1 January in year t be $P_{x,t}$.

(ii) Derive an expression, in terms of the $P_{x,t}$, for the exposed to risk for the period covering the years 2005 to 2014 inclusive which the National Statistics Office could use to estimate the overall death rate at age x nearest birthday.

(iii) Set out any assumptions you make in your derivation in part (ii), indicating where in the derivation they are needed.

The death registration system in this country is being maintained, but the next census is not planned until 2025.

(iv) Discuss how you might estimate death rates at age x nearest birthday for the calendar years 2015 and 2016.

7. CT4 September 2018 Q5

The following data are available for a sample of lives that were alive for at least some time between 1 January 2017 and 31 December 2017:
- date of birth;
- date of entry into observation (if entering after 1 January 2017);
- date of death (if died between 1 January 2017 and 31 December 2017);
- date of exit from observation while still alive (if leaving before 31 December 2017).

(i) Derive a maximum likelihood estimator of the hazard of death which could be used with these data and which uses all the information available on the timing of death.

(ii) Explain how this data can be used to estimate a life table.

8. CS2A April 2019 Q5

(i) State the principle of correspondence as it applies to death rates.

A country uses an administrative data system which shows the estimated population aged x last birthday on 1 January each year. Deaths are registered on an age nearest birthday basis.

(ii) Derive a formula which may be used to estimate an appropriate exposed to risk for calculating the average death rate at age x exact in the two-year period 1 January 2016 – 1 January 2018, defining all the terms you use. Assume that birthdays are evenly distributed across the calendar year.

UNIT 2

PRACTICE QUESTIONS

Someone remarks that, in this country, only one third of births take place in the first half of the calendar year.

(iii) Discuss the implications of this for the formula you have derived in part (ii).

9. CS2A April 2021 Q5

A life insurance company has offices in Towns A and B. The company writes 25-year term assurance policies. Below is data from the two offices relating to policyholders of the same age, x. Policies in force and deaths are on an 'age last birthday' basis.

	Town A	*Town B*
Policies in force on 1 January 2020	3,000	1,770
Policies in force on 1 January 2021	3,300	1,674
Deaths in calendar year 2020	63	26

(i) Estimate the force of mortality for the calendar year 2020 in respect of each of the offices in Towns A and B.

A detailed examination of the records shows that 50% of the policies in force in Town A at both dates were in respect of smokers, and 20% of policies in force in Town B at both dates were in respect of smokers.

The national forces of mortality at age x for smokers in 2020 were 50% higher than those for non-smokers.

(ii) Estimate the force of mortality for smokers and for non-smokers in each of the Towns A and B, clearly stating any assumptions that you make.

The life insurance company charges policyholders in Towns A and B the same premiums. It charges smokers in both towns 50% more than non-smokers.

(iii) Comment on the company's pricing structure in light of your results from parts (i) and (ii).

Ans: (i) Town A - 0.02, Town B - 0.0151. (ii) Town A - NS - 0.016, S - 0.024, Town B - NS - 0.01373, S - 0.02059

10. CS2A April 2022 Q2

An international pensions provider is interested in quantifying the force of mortality at certain ages for a particular country, for the period from 1 January 2017 through to 1 January 2020.

In this country, deaths are recorded by calendar year and classified by age at last birthday on the date of death, and the annual population censuses are completed on 30 June each year, classifying the population by age at last birthday.

Derive an expression for the crude force of mortality, $\hat{\mu}_{x+1/2}$, for a given age interval [x, x + 1], for the period from 1 January 2017 to 1 January 2020 using the population and death data available, clearly stating all assumptions and defining each of the terms you use.

11. CS2A September 2022 Q4

When considering rates of mortality, actuaries generally use the initial rate of mortality, q_x, whereas demographers often use the central rate of mortality, m_x. In order to estimate mx for each age x, a demographer collects the following data:

Age at last birthday	*Observed number of deaths*	*Central exposed to risk in years*
60	34	4,267
61	36	4,037
62	42	4,269
63	47	4,299
64	46	3,929

(i) Explain the difference between the two definitions of rate of mortality, stating which rate has the higher value at a given age, x.

(ii) Estimate m_x for x = 60, 61, 62, 63 and 64 years using the data above.

(iii) A pension scheme starts paying pensions on a member's 65th birthday. Calculate the probability that the pension scheme member aged 60 exactly will start receiving a pension, using the estimates in part (ii) above and carefully explaining all the steps and assumptions needed.

Ans: (ii)

x	m_x
60	0.007968127
61	0.008917513
62	0.00983837
63	0.010932775
64	0.011707814

(iii) 0.951834

UNIT 2

PRACTICE QUESTIONS