



Subject: Statistic & Risk Modelling

Chapter: Unit 2

Category: Practice Question

1. CT4 April 2010 Question 9

(i) Write down the hazard function for the Cox proportional hazards model defining all the terms that you use.

A farmer is concerned that he is losing a lot of his birds to a predator, so he decides to build a new enclosure using taller fencing. This fencing is expensive and he cannot afford to build a large enough area for all his birds. He therefore decides to put half his birds in the new enclosure and leave the others in the existing enclosure. He is convinced that the new enclosure is an improvement, but has asked an actuarial student to determine whether the new enclosure will result in an increase in the life expectancy of his birds. The student has fitted a Cox proportional hazards model to data on the duration until a bird is killed by a predator and calculated the following figures relating to the regression parameters:

		Parameter estimate	Variance
Bird	Chicken	0	0
	Duck	-0.210	0.002
	Goose	0.075	0.004
Enclosure	New	0.125	0.0015
	Old	0	0
Sex	Male	0.2	0.0026
	Female	0	0

(ii) State the features of the bird to which the baseline hazard applies.

(iii) For each regression parameter:

(a) Define the associated covariate.

(b) Calculate the 95% confidence interval based on the standard error.

(iv) Comment on the farmer's belief that the new enclosure will result in an increase in his birds' life expectancy.

(v) Calculate, using this model, the probability that a female duck in the new enclosure has been killed by a predator at the end of six months, given that the probability that a male goose in the old enclosure has been killed at the end of the same period is 0.1 (all other decrements can be ignored).

2. CT4 September 2011 Question 7

A study is made of the impact of regular exercise and gender on the risk of developing heart disease among 50-70 year olds. A sample of people is followed from exact age 50 years until either they develop heart disease or they attain the age of 70 years. The study uses a Cox regression model.

(i) List reasons why the Cox regression model is a suitable model for analyses of this kind. The investigator defined two covariates as follows:

- $Z_1 = 1$ if male, 0 if female
- $Z_2 = 1$ if takes regular exercise, 0 otherwise.

The investigator then fitted three models, one with just gender as a covariate, a second with gender and exercise as covariates, and a third with gender, exercise and the interaction between them as covariates. The maximised log-likelihoods of the three models and the maximum likelihood estimates of the parameters in the third model were as follows:

Null model	-1,269
Gender	-1,256
Gender + exercise	-1,250
Gender + exercise + interaction	-1,246

Covariate	Parameter
Gender	0.2
Exercise	-0.3
Interaction	-0.35

(ii) Show that the interaction term is required in the model by performing a suitable statistical test.

(iii) Interpret the results of the model.

3. Subject CT4 September 2012 Question 4

- State one advantage of a semi-parametric model over a fully parametric one.
- Write down a general expression for the Cox proportional hazard model, defining all the terms you use.

A life office is trying to understand the impact of certain factors on the lapse rates of its policies. It has studied the lapse rates on blocks of business subdivided by:

- Sex of policyholder (Male or Female)
- Policy type (Term Assurance or Whole Life)
- Sales channel (internet, Direct Sales Force or independent Financial Adviser.)

The office has fitted a Cox proportional hazards model to data and has calculated the following regression parameters:

Covariate	Regression parameter
Female	0.2
Male	0
Term Assurance	-0.1
Whole Life	0
Internet	0.4
Independent Financial Adviser	-0.2
Direct Sales Force	0

- (iii) State the sex/sales channel/policy type combination to which the baseline hazard relates

A Term Assurance is sold to a Female by an independent Financial Adviser.

- (iv) Calculate the probability this Term Assurance is still in force after five years giving that 60% of Whole Life policies bought on the Internet by Males have lapsed by the end of year five.

4. Subject CT4 April 2012 Question 9

- (i) List four factors other than age and smoker status by which life insurance mortality statistics are often subdivided.

Two offices in different towns of the same life insurance company write 25-year term assurance policies. Below are data from these two offices relating to policyholders of the same age. Both deaths and policies in force are on an age last birthday basis.

	<i>Gasperton</i>	<i>Great Hawking</i>
Policies in force on 1 January 2009	2,000	1,770
Policies in force on 1 January 2010	2,100	1,674
Deaths in calendar year 2009	25	21

(ii) Calculate the central death rate for the calendar year 2009 at this age for the offices in Gasperton and Great Hawking.

A detailed examination of the records shows that 50% of the policyholders in Gasperton at both censuses were smokers, and 20% of policyholders in Great Hawking at both censuses were smokers. National death rates at this age for smokers in 2009 were 40% higher than those for non-smokers.

(iii) Estimate the central death rates for smokers and non-smokers in Gasperton and Great Hawking.

The life insurance company charges policyholders in Gasperton and Great Hawking the same premiums for the 25-year term assurance policies. It charges smokers in both towns 40% more than non-smokers.

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

5. Subject CT4 April 2013 Question 6

- (i) State the form of the hazard function for the Cox regression model, defining all the terms used.
- (ii) State two advantages of the Cox regression model.

Susanna is studying for an online test. She has collected data on past attempts at the test and has fitted a Cox regression model to the success rate using three covariates:

Employee	$Z_1 = 0$ if an employee, and 1 if self-employed
Attempt	$Z_2 = 0$ if first attempt, and 1 if subsequent attempt
Study time	$Z_3 = 0$ if no study time taken, and 1 if study time taken.

Having analyzed the data Susanna estimates the parameters as:

Employment	0.4
Attempt	-0.2
Study time	1.15

Bill is an employee. He has taken study time and is attempting the test for the second time. Ben is self-employed and is attempting the test for the first time without taking study time.

(iii) Calculate how much more or less likely Ben is to pass, compared with Bill.

Susanna subsequently discovers that the effect of the number of attempts is different for employee and self-employed.

(iv) Explain how the model could be adjusted to take this into account.

6. CT4 April 2014 Question 5

An investigation has been performed into risk factors for liver disease in persons currently resident in the United Kingdom (UK) and aged over 50 years. It considered the impact of three covariates: age at the start of the investigation, weekly alcohol consumption and previous residence in a tropical country. The investigation used a Cox regression model for the hazard of developing the disease, $h(t)$, with three parameters, β_A , β_C and β_T , as follows:

$$h(t) = h_0(t) \exp(\beta_A A + \beta_C C + \beta_T T)$$

A was defined as exact age at the start of the investigation less 50 years.

C represented weekly alcohol consumption, and took the value 1 if the person consumed more than the recommended maximum per week (a heavy drinker) and 0 otherwise.

T represented previous residence in a tropical country, and took the value 1 if the person had lived in a tropical country for more than 12 months and 0 otherwise.

(i) State the characteristics of a person to whom the baseline hazard $h_0(t)$ applies.

The results of the investigation revealed that the hazard was:

- twice as high for a heavy drinker aged 60 years exact at the start of the investigation than for a person aged 50 years exact at the start of the investigation who was not a heavy drinker, where neither had previously lived in a tropical country.

- four times as high for a heavy drinker who had previously lived in a tropical country for more than 12 months than for a non-heavy drinker of the same age who had not previously lived in a tropical country.
- three times as high for a person who had lived in a tropical country for more than 12 months than for a person of the same age and drinking habits who had always lived in the UK.

(ii) Calculate β_A , β_C and β_T

The probability of a person aged 50 years exact at the start of the investigation, who does not drink heavily and has always lived in the UK remaining free of the disease for 10 years is 0.8.

(iii) Show that the probability of a person of the same age and drinking habits, who has lived for more than 12 months in a tropical country, remaining free of the disease for 10 years is slightly over one half.

7. CT4 September 2015 Question 6

(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
- t is the number of days since treatment started. $\beta_0, \beta_1, \beta_2$ are parameters. Their estimated values were $\beta_0 = 0.8$, $\beta_1 = 0.4$ and $\beta_2 = -0.1$.

(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.



INSTITUTE OF ACTUARIAL
& QUANTITATIVE STUDIES

8. CT4 October 2015 Question 6

(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. β_0 , β_1 and β_2 are parameters.

Their estimated values were $\beta_0 = 0.8$, $\beta_1 = 0.4$ and $\beta_2 = -0.1$.

(ii) Determine the characteristics of the baseline cow. [1] 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.

9. CT4 April 2015 Question 3

(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.

10. CT4 October 2016 Question 10

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:

<i>Covariate</i>	<i>Category</i>	<i>Parameter</i>
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.

(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.

11. CT4 April 2016 Question 6

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:

	<i>Factor</i>	<i>Parameter Estimate</i>	<i>Variance</i>
<i>Gender</i>	Male	-0.25	0.015
	Female	0	0
<i>Volume of energy consumed</i>	High	0.32	0.008
	Low	0	0
<i>Area of Residence</i>	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 centre 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.

12. CT4 September 2017 Question 6

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

$$h(t) = h_0(t) \exp(S\beta_S + A\beta_B + G\beta_G)$$

where β_S , β_B and β_G 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
 - 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 β_S , β_B and β_G

(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.

13. CS2A April 2019 Question 8

(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
- $\beta_A, \beta_E, \beta_D$ 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 $\beta_A, \beta_E, \beta_D$.

(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

14. CS2A September 2020 Question 8.

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

15. CS2A April 2022 Question 4

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:

X_1 = Medication indicator (0 = Placebo administered, 1 = Medication administered)

X_2 = 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:

<i>Covariate</i>	<i>Coefficient</i>	<i>Standard error</i>
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.