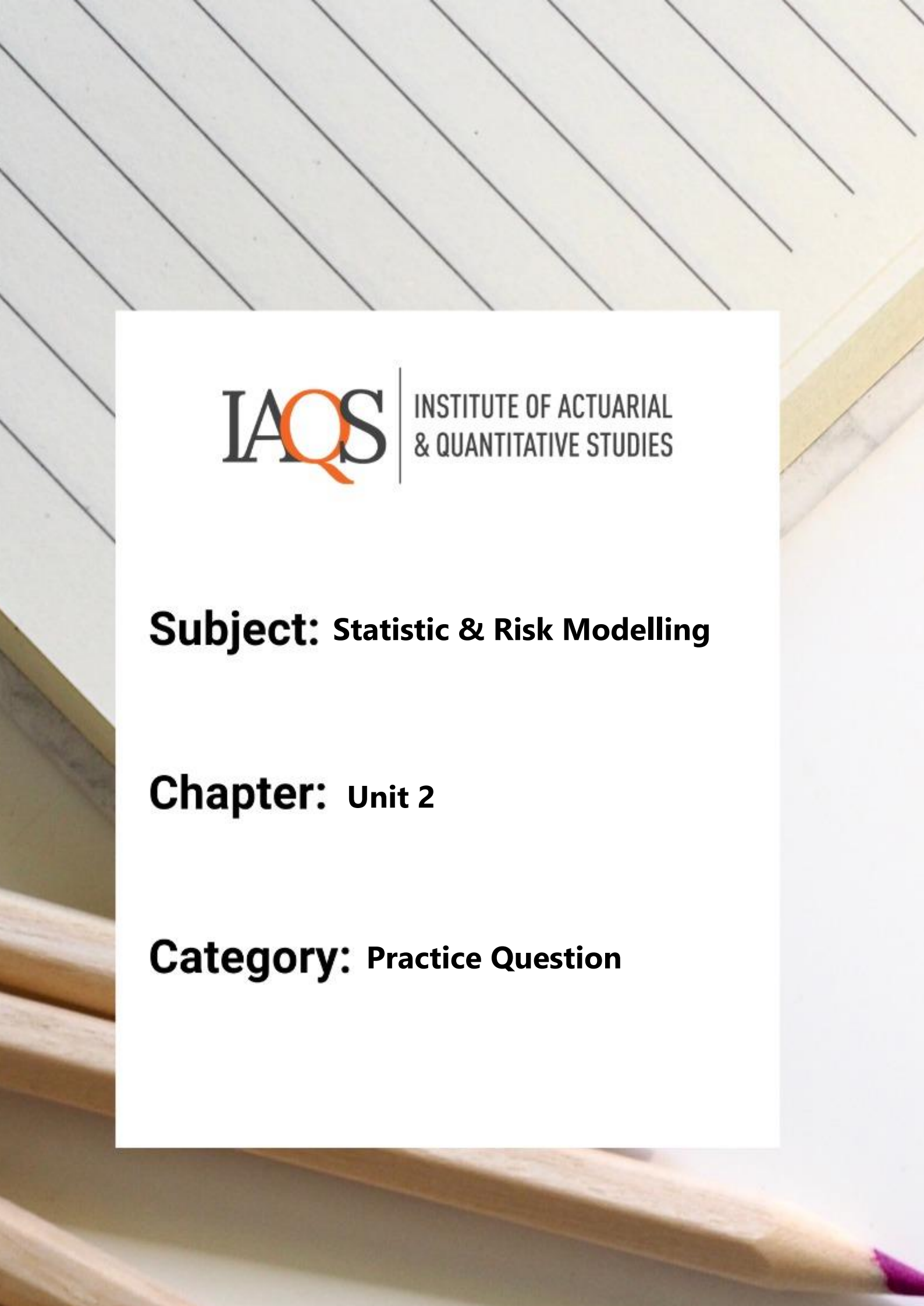

INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

Subject: Statistic & Risk Modelling

Chapter: Unit 2

Category: Practice Question

1. Subject 104 April 2004 Question

(i) A random variable X measure the duration until some event occurs.

Write down a definition of the hazard, *h(t)* of the event occurring at duration t,in terms of probabilities relating to the random variable *X* And state what the definition means in words.

(ii) You wish to investigate the effect of two factors, Z_1 and Z_2, on the duration until the event occurs. Someone suggests that you use a proportional hazards model. Explain what is meant by the term proportional hazards model.

(ii) The Weibull distribution has a survival function given by the formula

$S(t) = \exp[-(\lambda t)^{\alpha}]$

Where λ and α are parameters.

Show that, by choosing λ so that it depends on Z_1 and Z_2 only, the Weibull distribution can be used as proportional hazards model to investigate the effects of Z_1 and Z_2 on the duration.

2. Subject 104 September 2004 Question 3

A study has been undertaken into the effect of a new treatment on the survival times of patients suffering from a tropical disease. The following model has been fitted:

$$h_i(t) = h_o(t)\ exp\ (\underline{\beta}^T\ \underline{z})$$

Where:

$h_i(t)$ is the hazard at time t, where t is the time since treatment $h_o(t)$ is the baseline hazard at time t

$\underline{z}$ is a vector of covariates, where

z_1 = period from diagnosis to treatment in year
z_2 = 0 if existing treatment given, 1 if new treatment given
z_3 = 0 if female, 1 if male

$\underline{\beta}$ is a vector of parameters, where

β_1 = 0.5
β_2 = 0.01
β_3 = 0.05

(i) State the group of lives to which the baseline hazard applies.

(ii) For a male who was given the new treatment 6 months after diagnosis:

a) Write down the hazard function, in terms of $h_o(t)$ only.

b) Express the survival function, in term of $h_o(t)$ only.

(iii) For a female given the new treatment at the time of diagnosis, the probability of survival for 5 years is 0.75. Calculate the probability that the male in

(iv) (ii) will survive 5 years.

3. Subject CT4 April 2005 Question B1

(i) Write down the equation of the Cox proportional hazards model in which the hazard function depends on duration *t* and a vector of covariates *z.*
You should define all the other terms that you use.

(iii) Explain why the Cox model is sometimes described as "semi-parametric".

4. Subject CT4 September 2005 Question B5

An investigation was carried out into the effects of lifestyle factors on the mortality of people aged between 50 and 65 years. The investigation took the form of a prospective study following a sample of several hundred individuals
Form their 50th birthdays until their 65th birthday and collecting data on the following covariates for each person:
X_1 Sex (a categorical variable with 0 = female, 1 = male)

X_2 Cigarette smoking (a categorical variable with 0 = non-smoker, 1 = smoker)

X_3 Alcohol consumption (a categorical variable with 0 = consumes fewer than 21 units of alcohol per week, 1 = consumes 21or more units of alcohol per week)

In addition, data were collected on the age at death for persons who died during the period of the investigation.

In order to analyze the data, it was decided to use a Gompertz hazard,
$\lambda_x = Bc^x$, where x is the duration since the start of the observation.

(i) Explain why the Gompertz hazard might be appropriate for analyzing the mortality of persons aged between 50 and 65 years.

(ii) Show that the substitution:

$$B = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)$$

In the Gompertz model (where $\beta_0, \beta_1, \beta_2, \beta_3$ are parameters to be estimated), leads to a proportional hazard model for this particular analysis.

Using the Gompertz hazard, the parameter estimates in the proportional hazaeds model were as follows:

Covariate	*Parameter*	*Parameter estimate*
Sex	β_1	+0.40
Cigarette smoking	β_2	+0.75
Alcohol consumption	β_3	-0.20
	β_0	-5.00
	c	+1.10

(a) Describe the characteristics of the person to whom the baseline hazard applies in this model.

(b) Calculate the estimated hazard for a female cigarette smoking aged 55 years who does not consume alcohol.

(c) Show that, according to this model, a cigarette smoker at any age has risk of death roughly equal to that of a non-smoker aged eight years older.

5. Subject CT4 April 2006 Question B1

A Cox proportional hazards model was estimated to assess the effect on survival of a person's sex and his or her self-esteem (measured on a three point scale as "low", "medium" or "high"). The baseline category was males with "low" self-esteem.

Write down the equation of the model, using algebraic symbols to represent variables and parameters and defining all the symbols that you use.

6. Subject CT4 September 2006 Question B3

An investigation was undertaken into the effect of a new treatment on the survival times of cancer patients. Two groups of patients were identified. One group was given the new treatment and the other an existing treatment.

The following model was considered:

$$h_i(t) = h_o(t) \ exp\ (\underline{\beta}^T \underline{z})$$

Where: $h_i(t)$ is the hazard at time t is the time since the start of treatment

h_0 is the baseline hazard at time t
$\underline{z}$ is a vector of covariates such that:
z_1 is sex (a categorical variable with 0 = female, 1= male)
z_2 is tretment (a categorical variable with 0 = existing treatment, 1 = new treatment)

and $\underline{\beta}$ is a vector of parameters, (β_1, β_2)

The results of the investigation showed that, If the model is correct the risk of death for a male patient is 1.02 times that of a female patient; and the risk of death for a patient given the existing treatment is 1.05 times that for a patient given the new treatment

(i) Estimate the value of the parameters β_1 and β_2.

(ii) Estimate the ratio by which the risk of death for a male patient who has been given the new treatment is greater or less than that for a female patient given the existing treatment.

(iii) Determine, in terms of the baseline hazard only, the probability that a male patient will die within 3 years of receiving the new treatment.

7. Subject CT4 September 2007 Question 10

(i) Compare the advantage and disadvantages of fully parametric models and the Cox regression model for assessing the impact of covariates on survival.

You have been asked to investigate the impact of a set of covariates, including age, sex, smoking, region of residence, education attainment and amount of exercise undertaken, on the risk of heart attack. Data are available from a prospective study which followed a set of several

thousand persons from an initial interview until their first heart attack, or until their death from a cause other than a heart attack, or 10 years had elapsed since the initial interview (which of these occurred first).

(ii) State the types of censoring present in this study, and explain how each arises.

(iii) Describe a criterion which would allow you to select, those covariates which have a statistically effect on the risk of heart attack, when controlling the other covariates of the model.

Suppose your final model is a Cox model which has three covariates; age (measured in age last birthday minus 50 at the initial interview), sex (male = 0,
Female = 1) and smoking (non-smoker = 0, smoker = 1), and that the estimated parameters are;

Age	0.01
Sex	-0.4
Smoking	0.5
Sex x smoking	-0.25

Where "sex x smoking" is an additional covariate formed by multiplying the two covariates "sex" and "smoking".

(iv) Describe the final model's estimate of the effect of sex and of smoking behavior on the risk of heart attack.

(v) Use the results of the model to determine how old a female smoker must be at the initial interview to have the same risk of heart attack as a male non-smoker aged 50 years at the initial interview.

8. Subject CT4 April 2008 Question 8

An education authority provides children with musical instrument tuition. The authority is concerned about the number of children given up playing their instrument and is testing a new tuition method with a proportion of the children which it hopes will improve persistency rates. Data have been collected and a Cox proportional hazard model has been fitted for the hazard of giving up playing the instrument. Symmetric 95% confidence intervals (based upon standard errors) for the regression parameters are shown below.

Covariate	Confidence Interval
Instrument	
Piano	0
Violin	[-0.05, 0.19]
Trumpet	[0.07, 0.21]
Tuition method	
Traditional	0
New	[-0.15, 0.05]
Sex	
Male	[-0.08, 0.12]
Female	0

(i) Write down general expression for the Cox proportional hazards model, Defining all terms that you use.

(ii) State the regression parameters for the fitted model.

(iii) Describe the class of children to which the baseline hazard applies.

(iv) Discuss the suggestion that the new tuition method has improved the chances of children continuing to play their instrument.

(v) Calculate, using the results from the model, the probability that a boy will still be playing the piano after 4 years if provided with the new tuition method, given that the probability that a girl will still be playing the trumpet after 4 years following the traditional method is 0.7.

9. Subject 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 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 new enclosure and leaves the others in the existing enclosure. He 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 life expectancy of his birds. The students has fitted a Cox

proportional hazards model to data on the duration until a birds is killed by a predator and calculated the following figures relating to the regression parameters:

		Parameter estimate	Variance
Birds	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 birds 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 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 decrement can be ignored).

10. Subject CT4 September 2012 Question 4

(i) State one advantage of a semi-parametric model over a fully parametric one.

(ii) 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 bloke of business subdivided by:

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

The office has fitted a Cox proportional hazards models model to data and has calculates 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.

11. 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.

12. Subject CT4 April 2005 Question B6

An investigation into mortality collects the following data:

θx = Total number of policies under which death claims are made when the Policyholder is aged x last birthday in each calendar year

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

(i) State the principle of correspondence .

(ii) Obtain an expression, in terms of the Px (t), for the central exposed to risk E_x^c, which corresponds to the claim data and which may be used to estimate the force of mortality in year t at each age $x_1\mu_x$. State any assumptions you make.

(iii) Comment on the effect on the estimation of the fact that the θx relate to claims, rather than deaths, and the Px (t) relate to policies, not lives

13. Subject CT4 April 2005 Question B7

An investigation took place into the mortality of pensioners . The investigation began on 1 January 2003 and ended on 1 January 2004. The table gives the data collected In this investigation for 8 lives.

Date of birth	Date of entry into observation	Date of exit from observation	Whether or not exit was due to: death (1) or other reason (0)
1 April 1932	1 January 2003	1 January 2004	0
1 October 1932	1 January 2003	1 January 2004	0
1 November 1932	1 March 2003	1 September 2003	1
1 January 1993	1 March 2003	1 June 2003	1
1 January 1993	1 June 2003	1 September 2003	0
1 March 1993	1 September 2003	1 January 2004	0
1 June 1993	1 January 2003	1 January 2004	0
1 October 1993	1 June 2003	1 January 2004	0

The force of mortality μ_{70}, between exact ages 70 and 71 is assumed to be constant.

(i) (a) Estimate the constant force of mortality, μ_{70}, using a two-state model and the data for the 8 lives in the table.
(b) Hence or otherwise estimate q_{70}.

(ii) Show that the maximum likelihood estimate of the constant force μ_{70}, using a Poisson model of mortality is the same as the estimate using the two-state model.

(iii) Outline the difference between the two- state model and the Poisson model when used to estimate transition rates.

14. Subject CT4 April 2006 Question B2

i. (a) Explain why it is important to sub-divide data when carrying out mortality investigations.
(b) Describe the problems that can arise with sub-dividing data.

ii. List four factors which are often used to sub-divide life assurance data.

15. Subject CT4 September 2006 Question B2

A national mortality investigation is carried out over the calendar years 2002, 2003 and 2004. Data are collected from a number of insurance companies.
Deaths during the period of the investigation θx, are classified by age nearest at death.
Each insurance company provides details of the number of in-force policies on the 1 January 2002, 2003, 2004 and 2005, where policyholders are classified by age nearest birthday, Px (t).

i. (a) State the rate year implied by the classification of deaths.
(b) State the ages of the lives at the lives at the start of the rate interval

ii. Derive an expression for the exposed to risk, in terms of Px (t), which may be used to estimate the force of mortality in year t at each age. State any assumptions you make.

iii. Describe how your answer to (ii) would change if the census information provided by some companies was Px (t), the number of in-force policies on 1 January each year, where policyholders are classified by age last birthday.

16. Subject CT4 September 2006 Question B4

An investigation took place into the mortality of persons between exact ages 60 and 61 years. The table below gives an extract from the results. For each person it gives the age at which they were first observed, the age at which they ceased to be observed and the reason for their departure from observation.

Person	Age at entry Years	Months	Age at exit Years	Months	Reason for exit
1	60	0	60	6	Withdraw
2	60	1	61	0	Survived 61
3	60	1	60	3	Died
4	60	2	61	0	Survived 61
5	60	3	60	9	Died
6	60	4	61	0	Survived 61
7	60	5	60	11	Died
8	60	7	61	0	Survived 61
9	60	8	60	10	Died
10	60	9	61	0	Survived 61

i. Estimate q_{60} using the Binomial model.

ii. List the strengths and weaknesses of the binominal model for the estimation of empirical mortality rates, compared with the Poisson and two-state models.

17. Subject CT4 April 2007 Question 4

The actuary to a large pension scheme carried out an investigation of the mortality of the scheme's pensioners over the two years from 1 January 2005 to 1 January 2007.

i. List the data required by the actuary for an exact calculation of the central exposed to risk for lives aged x.

The Following is an extract from the data collected by the actuary.

Age x nearest birthday.	Number of pensioners at:			Deaths During:	
	1 January 2005	1 January 2006	1 January 2007	2005	2006
63	1,248	1,312	1,290	10	6
64	1,465	1,386	1,405	13	15
65	1,678	1,720	1,622	16	23
66	1,719	1,642	1,667	22	19
67	1,686	1,695	1,601	19	25

ii. (a) Derive an expression that could be used to estimate the central exposed to risk using the available data. State any assumptions you make.

(b) Use the data to estimate μ_{65}. State any further assumptions that you make.

18. Subject CT4 September 2007 Question 8

i. Describe the difference between the central exposed to risk and the initial exposed to risk

The following data come from an investigation of the mortality of participants in a dangerous sport during the calendar year 2005.

Age X	Number of lives aged X last birthday on:		Number of deaths during 2005 to persons aged X last birthday at death
	1 January 2005	1 January 2006	
22	150	160	20
23	160	155	25

ii. (a) Estimate the initial exposed to risk at ages 22 and 23.

(b) Hence estimate q_{22} and q_{23}.
Suppose that in this investigation, instead of aggregate data we had individual-level data on each persons date of birth, death of death, and date of exit from observation (if exit was for reasons other than death).

iii. Explain how you would calculate the initial exposed-to-risk for lives aged 22 years last birthday.

19. Subject CT4 April 2008 Question 1

List four factors in respect of which life insurance mortality statistics are often subdivided.

20. Subject CT4 September 2008 Question 3

Compare the advantages and disadvantages of the binomial and the multiple-state model in the following situation:

a) Analysing human mortality without distinguishing between cause of death.

b) Analysing human mortality when distinguishing between cause of death.

21. Subject CT4 September 2008 Question 3

i. List the data needed for the exact calculation of a central exposed to risk depending on age.
An investigation studied the mortality of persons aged between exact ages 40 and 41 years. The investigation began on 1 January 2008 and ended on 31 December 2008. The following table gives details of 10 lives involved in the investigation.

Life	Date of 40th Birthday	Date of death
1	1 March 2007	--
2	1 May 2007	1 October 2008
3	1 July 2007	--
4	1 October 2007	--
5	1 December 2007	1 February 2008
6	1 February 2008	--
7	1 April 2008	--
8	1 June 2008	1 November 2008
9	1 August 2008	--
10	1 December 2008	--

Persons with no date of death given were still alive when the investigation ended.

ii. Calculate a central exposed to risk using the data for the 10 Lives in the sample.

iii. (a) Calculate the maximum likelihood estimate of the hazard of death at age 40 last birthday.
(b) Hence, or otherwise, Estimate q_{40}

22. Subject CT4 September 2008 Question 3

i. In the context of mortality investigations describe the principal of correspondence and given an example of a situation in which it may be hard to adhere to this principle.

On 1 January 2005 a country introduced a comprehensive system of death registration, which classified death by age last birthday on the date of death.
The government of the country wishes to obtain estimates of the force of mortality, μ_x, by single years of age x for the period between 1 January 2005 and 1 January 2008. Annual population censuses have been taken on birthday. However the only copy of the data from the population census of 30 June 2006 was lost when the computer disc on which it was stored was being transferred between government departments.
Let the population aged x last birthday on 30 June in Year t be denoted by the symbol Pxt, and the number of death during the period of investigation of persons aged X be denoted by the symbol d_x

ii. Derive an expression in terms of Pxt and dx which may be used to estimate μ_x.

23. Subject CT4 April 2010 Question 1

List four factors often used to subdivide life insurance mortality statistics.

24. Subject CT4 April 2010 Question 6

An Oil company has discovered a vast deposit of oil in an equatorial swamp. The area is extremely unhealthy and inhabited by venomous spider. There is an antidote to bites from these spider but it is expensive the antidote acts instantly but does not provide future immunity. The company commissions a study to estimate the rate of being bitten by the spiders among its employees, in order to determine the amount of antidote to provide.

Employees of the company are posted to the swamp for six months tours of duty starting on 1 January, 1 April, 1 July or 1 October. The first employees to be posted arrived on 1January 2008. The swamp is so inaccessible that no employees are allowed to leave before their six month tours of duty are completed.
Accidental deaths are common in this dangerous location.

The table below gives some data from the study

Quarter beginning	Number of new arrivals at start of quarter	Number of accidental deaths during quarter	Number of spider bites during quarter
1 January 2008	90	10	15
1 April 2008	80	8	25
1 July 2008	114	10	30
1 October	126	13	40

i. Estimate the quarterly rate of being bitten by a spider for each quarter of 2008, stating any assumptions you make.

ii. Suggest reasons why the assumptions you made in (i) might not be valid.

25. Subject CT4 October 2010 Question 7

See Question Analysis (1)

26. Subject CT4 April 2011 Question 1

Give three advantages of the two-state model over the Binominal model for estimating transition intensities where exact dates of entry into and exit from observation are known.

27. Subject CT4 April 2011 Question 8

i. Explain the difference between the central and the initial exposed to risk, in the context of mortality investigations.

An investigation studied the mortality of infants aged under 1 year. The following table gives details of 10 lives involved in the investigation. Infants with no date of death given were still alive on their first birthday.

Life	Date of birth	Date of death
1	1 August 2008	--
2	1 September 2008	--
3	1 December 2008	1 February 2009
4	1 January 2009	--
5	1 February 2009	--
6	1 March 2009	1 December 2009
7	1 June 2009	--
8	1 July 2009	--
9	1 September 2009	--
10	1 November 2009	1 December 2009

ii. Calculate the maximum likelihood estimate of the force of mortality, using a two-state model and assuming that the force is constant .

iii. Hence estimate the infant mortality rate q_0.

iv. Estimate the infant mortality rate q_0 using the initial exposed to risk

v. Explain the difference between the two estimates.

28. Subject CT4 October 2011 Question 9

i. State the principle of correspondence as it applies to the estimation of mortality rates.

ii. Explain why it might be difficult to ensure the principle of correspondence is adhered to, and give a specify example of an investigation where this may be the case.

An actuary was asked to investigate the mortality of lives in a particular geographical area. Data are available of the population of this area, classified by age last birthday, on 1 January in each year, classified by age nearest birthday at death, are also available.

iii. Derive a formula which would allow the actuary to estimate the force of mortality at age $x + f$, μ_{x+f}, in a particular calendar year, in terms of the available data, and derive a value for f.

iv. List four factors other than geographical location which a government statistical office might use to subdivide data for national mortality analysis.

29. Subject CT4 April 2012 Question 2

i. Explain the reasons why data are subdivided when conducting mortality investigation.

ii. Describe the problems which can arise with subdividing data

30. Subject CT4 April 2012 Question 9

See Question Analysis (2).

31. Subject CT4 September 2012 Question 3

i. State the principle of correspondence as it applies to mortality rates.

A life insurance company has the following data:

	Number of policies in force on			
Age last birthday	1 January 2009	1 January 2,010	1 July 2010	1 January 2010
49	2,000	2,100	2,300	2,500
50	2,100	2,200	2,300	2,400
51	2,300	2,400	2,300	2,600

Number of deaths classified by age next birthday and calendar year

Age next birthday	2009	2010
49	175	200
50	200	225
51	225	235

ii. Estimate, using these data, the force of mortality at age 50 next birthday for the period 1 January 2009 to 1 January 2011.

iii. State the exact age to which your answer to part (ii) relates.

32. Subject CT4 September 2013 Question 5

Population census in a certain country are taken each year on the president's birthday, provided that the president's astrological advisor deems the taking of a census favourable. Censuses record the age of every inhabitant in completed years (that is, curtate age). Deaths in this country are registered as they happen, and classified according to age nearest birthday at the time of death.

Below are some data from the three most recent census.

Age in completed years	Population 2006 (thousands)	Population 2009 (thousands)	Population 2010 (thousands)
64	300	320	350
65	290	310	330
66	280	300	320

Between the censuses of 2006 and 2009 there were a total of 3,000 deaths to inhabitants aged 65 nearest birthday, and between the censuses of 2009 and 2010 there were a total of 1,000 deaths to inhabitants aged 65 nearest birthday.

i. Estimate, stating any assumptions you make, the death rate at aged 65 years for each of the following periods:
- The period between the 2006 and 2009 censuses
- The period between the 2009 and 2010 censuses

ii. Explain the exact age to which your estimate apply.