Unit 3 & 4

Assignment - 2

Statistical and Risk Modelling - I

Q1. The mortality rates, as evident from the large experiences, are believed to vary smoothly with age; therefore the crude estimate of mortality at any age carries information about the mortality rates at adjacent ages.

By smoothing the experience, we can make use of data at adjacent ages to improve the estimate at each age. This reduces the sampling (or random) errors.

This mortality experience may be used in financial calculations. Irregularities, jumps and anomalies in financial quantities (such as premium rates under life insurance contracts). are hard to justify to customers.

The main limitation in mortality investigations that graduation ~~will~~ will not be able to overcome is to remove any bias in the data arising from faulty data collection or otherwise.

Page No.
Date

Q3.

(1) The crude mortality rates derived from the data will progress more or less roughly. However, it is expected that mortality rates would be smooth functions of age. Hence it follows that a crude estimate at any age 'x' also carries information about the mortality rates at age 'x-1' and 'x+1'.

Graduating the crude rates will enable this information to be captured, thus resulting in the mortality rate progressing smoothly over age.

The mortality rates would be used in computing premium rates / annuity rates. If the underlying mortality rates do not progress smoothly, then the derived premium rates etc would also fluctuate randomly, which would be difficult to understand / explain to potential customers.

The purpose of graduation is

- To ~~pro~~ produce a smooth set of rates that are suitable for a practical purpose.
- To remove random sampling errors.
- To use the information available from adjacent ages.

Page No.
Date

(2) The methods used for graduation are

- Graduation by parametric formula
- Graduation by reference to a standard table.
- Graphical graduation.

(3) The null hypothesis is that the graduated rates represent the mortality of the members in pension scheme.
To test the appropriateness of the graduation using chi-squared test, we compare
$\sum_x z_x^2$ with χ^2 with 'm' degrees of freedom.

z_x^2	Age-group	Pensioners	Actual deaths	Exp deaths	Std. devn
1.9005	50-54	37259	248	222.28	1.3786
1.64850	55-59	28057	392	367.55	1.2839
0.09451	60-64	25654	680	672.13	0.3074
0.10823	65-69	20475	987	997.13	-0.3290
0.29651	70-74	16219	1380	1360.77	0.5445
1.18951	75-79	11843	1625	1584.59	1.0906
4.8933	80-84	7532	1564	1487.57	2.2121
1.7409	85-89	3294	925	891.36	1.3195
6.37761	90+	450	130	155.48	-2.5254

The number of age-groups is 9. As the graduation has been done with reference to a standard table, some degrees of freedom are lost.

$m < 9$, $m = 8$

$\Sigma z_x^2 = 18.2497$, which is higher than the critical value of chi-squared distribution with 8 degrees of freedom at 5% which is 15.51.

We reject the null hypothesis.

(4)

Signs test

The signs test checks for any overall bias, whether the graduated rates are too high or too low.

The number of positive signs (or negative) is a Binomial distribution $B(9, 0.5)$. There are 7 positive deviations and the probability of obtaining 7 or more positive signs is 0.01953 and since this is lower than 2.5% we reject the null hypothesis and conclude that the graduated rates do not adhere to the data.

Cumulative deviations test

The cumulative deviations test detects overall bias or long runs of deviations of the same sign.

The test statistic :-

$\Sigma(\theta_x - E_x \mathring{q}_x) / \sqrt{\Sigma E_x \mathring{q}_x}$ is $N(0,1)$

$= 187.14 / 87.99 = 2.1265$

Since this value is higher than 1.96, the critical value at the 2.5% level, we reject the null hypothesis.

(5) The graduated rates are below the crude mortality rates for most of the ages, the exceptions being 65-69 and 90+ age group.

Since the graduated rates ~~ta~~ under-estimate the actual mortality experience, use of the graduated rates may results in over-provisioning of the annuity liabilities under the pension scheme.

Q4.

Cumulative deviations test

The null hypothesis is :-
H_0 = the standard table rates are the true underlying mortality rates for the term assurance policyholders.

We first calculate the individual deviations using the formulas:-

$$Z_x \approx \frac{O_x - E_x q_x^s}{\sqrt{E_x q_x^s (1-q_x^s)}}$$

$$\approx \frac{O_x - E_x q_x^s}{\sqrt{E_x q_x^s}}$$

Since, $q_x^s \approx 0$, we can use the approximation that $1-q_x^s \approx 1$

x	E_x	d_x	$q_x(s)$	$E_x \cdot q_x(s)$	$d_x - E_x \cdot q_x$
40	50000	87	0.2053%	102.65	-15.65
41	48560	84	0.2247%	109.11	-25.11
42	47190	101	0.2418%	114.11	-13.11
43	44100	112	0.2602%	114.75	-2.75
44	43600	123	0.2832%	123.48	-0.48
45	40400	110	0.3110%	125.64	-15.64
46	37280	108	0.3438%	128.17	-20.17
47	35370	122	0.3816%	134.97	-12.97
48	32100	150	0.4243%	136.20	13.80
49	29000	139	0.4719%	136.85	2.15
50	26200	131	0.5244%	137.39	13.61
				1363.32	-76.32

Test statistic = $\dfrac{\sum (O_x - E_x q_x^s)}{\sqrt{\sum E_x q_x^s}}$ = - 2.07

Under the null hypothesis, this has a Normal distribution. Since this is a two-tailed test, we compare the test statistic, at 95% confidence level $|-2.07| > 1.96$.

We reject the null hypothesis and that the standard table rates do not adequately reflect the true mortality rates for the term assurance policyholders.

Page No.
Date

Q5.

1. (a) Maximum smoothness would be achieved by ignoring the plotted estimated rates and drawing a graduation curve which is very smooth, eg. straight line. The deviations between the rates read from such a curve and the observed rates are likely to be very large. The graduation curve will be very smooth but have poor adherence to data. On the other hand joining up a plot of the estimated values will give perfect adherence to data but is likely to produce a curve with rapidly changing curvature which would not satisfy the smoothness criteria.

(b) Graduation aims to resolve these conflicts by striking a fair balance between the adherence to data and smooth progression of rates.

Graduated rates can be obtained by many methods, some ensure smoothness, eg. graduation by a mathematical from, reference to a standard table. In this case the graduated rates just need to satisfy tests of adherence to data.

Graphical graduation does not ensure smoothness, so graduated rates must be checked for smoothess and adherence to data. The graduation process must be repeated until both criteria are satisfied.

Q5.

(2) By rearranging we get

$$\log(qx/px) = a + bx$$

$$(qx/px) = \exp(a + b^{*}x)$$

$$qx/(1-qx) = \exp(a + b^{*}x)$$

$$qx = 1/[1+\exp[-(a+bx)]]$$

Age	a+bx	qx	E	A
30	-7.0140	0.00089	224.60	245.00
31	-6.8978	0.00100	252.25	276.00
32	-6.7816	0.001133	283.29	313.00
33	-6.6654	0.00127	318.16	323.00
34	-6.5492	0.00142	357.30	339.00
Total			1435.60	1496.00

	A − E	$(A-E)^2/E$
	20.40	1.85
	23.75	2.24
	29.71	3.12
	4.84	0.07
	-18.30	0.94
Total	60.40	8.22

Chi-square value - 8.22

Page No.	
Date	

However we had used two estimated parameters so the number of degrees of freedom to use is 5 − 2 = 3. The upper 5% point of chi square (3) distribution is 7.815. So on the basis of chi square test, our hypothesis of good fit is rejected.

Q6.

(a) The low volume of data is the principal problem in the given situation, hence:-

- The crude rates may not be suitable for the purpose.
- There are random sampling errors.
- The rate at a particular age can not be linked to the rates at adjacent ages.
- Rates will not progress smoothly from age to age which allows a practical smooth set of premium rates to be produced.

(b)

Age	Z_x	Z_x^2	$\mathring{\mu}_{x+1/2}$	$\Delta \mathring{\mu}_{x+1/2}$	$\Delta^2 \mathring{\mu}_{x+1/2}$	$\Delta^3 \mathring{\mu}_{x+1/2}$
18-22	0.5249	0.27552	0.0061	0.0070	0.0061	0.0033
23-27	0.3615	0.13068	0.0131	0.0131	0.0094	0.0033
28-32	1.0266	1.05391	0.0262	0.0205	0.0127	0.0020
33-37	-1.0912	1.19072	0.0487	0.0352	0.0147	-0.0009
38-42	1.2394	1.53611	0.0839	0.0499	0.0138	-0.0044
43-47	0.59449	0.35391	0.1338	0.0637	0.0094	-0.0076
48-52	0.6346	0.40271	0.1975	0.0731	0.0018	
53-57	0.1787	0.03193	0.2706	0.0749		
58+	-0.0367	0.001349	0.3455			
		4.97683				

Page No.
Date

Test Statistics $X = \Sigma z_i^2$

Null Hypothesis

H_0 – X has a chi square distribution
The observed value of X is 4.98.

Degree of freedom = 9
This is one sided test. Upper 5% point of Chi Square distribution exceeds the observed value 4.98 for all degrees of freedom (≤ 9) ~~expe~~ except one.
There is insufficient evidence to reject H_0.
The third difference of $\mathring{\mu}_{x+1/2}$ is not very insignificant and hence the graduated rates are not very smooth.

Q7.

One of the most widely used models is that developed by Lee and Carter in the early 1990's.

The Lee-Carter model has two factors, age and period and may be written as follows:-

Lee - Carter model.

$$\log_e m_{x,t} = a_x + b_x k_t + \varepsilon_{x,t}$$

where:

- $m_{x,t}$ is the central mortality rate at age x in year t.

- a_x describes the general shape of mortality at age x (more exactly it is the mean of the time - averaged logarithms of the central mortality rate at age x).

- b_x measures the change in the rates in response to an underlying time trend in the level of mortality k.

- k_t reflects the effect of the time trend on mortality at time t, and

- $\varepsilon_{x,t}$ are independently distributed random variables with means of zero and some variance to be estimated.

So, ignoring the error term, the mortality rate at age x in projection year t is:-

$$m_{x,t} = \exp\left(a_x + b_x k_t\right)$$

As written above, the Lee-Carter model is ~~no~~ not 'identifiable'. To obtain unique estimates of the parameters a_x, b_x and k_t constraints need to be imposed.

Q9.

(i) H_0 - Standard mortality table represents the true underlying mortality rates of the experience.

Age (x)	Initial exposed to risk (E_x)	Standard Mortality rate (q_x)	Expected Deaths E	Actual deaths A	A-E	Chi-squared
50	2305	0.0064	14.75	15	0.25	0.00417
51	2475	0.0069	17.08	16	-1.08	0.06798
52	2705	0.0075	20.29	22	1.71	0.01022
53	2900	0.0081	23.49	23	-0.49	0.01216
54	3170	0.0087	27.58	27	-0.58	0.11850
55	6730	0.0094	63.26	66	2.74	0.08556
56	6875	0.0101	69.44	67	-2.44	0.0856
57	8190	0.0109	89.27	88	-1.27	0.01810
58	8200	0.0117	95.94	102	6.06	0.38278
59	7680	0.0119	91.39	80	-11.39	1.42001
60	7160	0.0121	86.64	85	-1.64	0.0309
Total	58390		599.12	591.00	-8.12	2.295

(a) If H_0 is true, the Chi-squared statistic will follow the χ^2 sampling distribution.

Since we are an experience with a standard table, the degrees of freedom = 11.

Observed value of χ^2 = 2.295

Tabular value of χ^2 for upper 95% point = 19.68

The observed value of test statistic is less than 5% significance point.

Hence, there is no evidence to reject the null hypothesis.

(b) If H_0 is true statistic $Z(c) \sim N(0,1)$

Here, observed value of $Z(c) = \frac{-8.12}{\sqrt{599.12}}$

$= -0.332$

This is a two-tailed test. We compare the value of statistic with the upper and lower 2.5% points of $N(0,1)$.

As $-1.96 < -0.332 < 1.96$, there is insufficient evidence to reject the null hypothesis.

(c) If H_0 is true :-

Age	Z_x	Z_x^2
50	0.0646	0.00417
51	-0.2607	0.06798
52	0.3802	0.14455
53	-0.1011	0.01022
54	-0.1103	0.01216
55	0.3442	0.11850
56	-0.2925	0.08556
57	-0.1345	0.01810
58	0.6187	0.38278
59	-1.1916	1.42001
60	-0.1758	0.03089

Number of groups of positive signs = 4

Number of positive signs = 4
Number of negative signs = 7

From the table, we can find that the value for k for which

$$\sum_{t=1}^{k} \frac{\binom{n1-1}{t-1}}{\binom{m}{n1}} \binom{n2+1}{t} < 0.05$$

where $n1 = 4$, $n2 = 7$ & $m = 4+7 = 11$
The tabular value of k is 1 which is less than number of groups of signs and hence there is no evidence to reject the null hypothesis.

Q10.

(i) The hypothesis requires the expected claims to be calculated based on the observed mortality experience in th last year.

E(A_LY) = observed mort rate (LY) * Exp_risk_CY

= Deaths_LY / Exp_risk_LY * Exp risk CY

This requires the exposed to risk at each age. As the exposed to risk is not directly available the expected deaths based on actual experience in last year can be derived as

E(A_LY) = Deaths_LY / [Exp risk LY * std mort rate] * (Exp_risk_CY * std mor rate)

= Deaths_LY / exp_deaths_LY * Exp_deaths CY.

$\chi^2 = (A-E)^2/E$	Age.	Deaths LY	Exp deaths LY	Exp deaths CY	E(A_LY) E	Act claim A
0.01	25	808	810	724	722	719
34.46	30	1851	1708	1433	1553	1322
9.28	35	1400	1084	444	573	500
4.15	40	1562	1705	1397	1280	1207
7.24	45	1366	1572	1465	1273	1177
2.48	50	1296	1643	1209	954	905
1.00	55	2200	2911	2436	1841	1798

There are 7 ages. Hence the degrees of freedom is 7. The upper 95% for χ^2_7 for 7 degrees of freedom is 15.5. The observed value is 58.6 which is much higher than this, hence the hypothesis is rejected.

Page No.			
Date			

Q11.

The null hypothesis is that the graduated rates are the same as the true underlying rates for the block of business.

The test statistic -

$$\chi_n^2 = \sum_x^{\text{all ages}} Z_x^2$$

where

n is the degree of freedom.

Z_x^2	Age	Exposed to risk	Observed death	Graduated Rates (q_x)	Expected Death	Z_x
0.1179	20	1060	12	0.0125	13.25	-0.3434
0.0381	21	1250	14	0.0118	14.75	-0.1953
0.0028	22	1210	16	0.0134	16.214	-0.0531
0.4845	23	700	9	0.0102	7.140	0.6961
0.1707	24	875	11	0.0111	9.7125	0.4131
0.3027	25	950	15	0.0137	13.015	0.5502
0.0020	26	805	10	0.0126	10.143	-0.0449
0.1548	27	1390	16	0.0127	17.65	-0.3934
1.1098	28	1080	17	0.0122	13.176	1.0535
0.5822	29	1310	14	0.0131	17.161	-0.7631
2.9657			Total			0.9197

Observed Test Statistic 2.9657.

The number of age groups (n) is 10, but we lose at least one number of degrees since we are comparing our data ~~with~~ with a set of graduated data derived from our data, perhaps 2. So the number of degree of freedom is 8.

The critical value of the chi-squared distribution with 8 degrees of freedom at the 5% level is 15.51.

Since,

$2.9657 < 15.51$

We do not reject the null hypothesis

Q 12.

Age	E	A	Z_x	Z_{x+1}	$A = Z_x - \bar{Z}$	$B = Z_{x+2} - \bar{Z}$
60	38.3	36	-0.37165	-1.01415	-0.40343	-1.04593
61	40.45	34	-1.01415	-0.69317	-1.04593	-0.72496
62	42.52	38	-0.69317	-0.68311	-0.72496	-0.7149
63	44.56	40	-0.68311	0.6332	-0.7149	0.601412
64	47.63	52	0.6332	-0.45398	0.601412	-0.48577
65	51.25	48	-0.45398	0.221781	-0.48577	0.189993
66	55.35	57	0.221781	0.585212	0.18999	0.553424
67	60.45	65	0.58521	0.726943	0.553424	0.695155
68	63.22	69	0.72694	0.755509	0.695155	0.723721

A2	AB
0.162758	0.421964
1.093978	0.758262
0.525569	0.518275
0.511083	-0.42995
0.361697	-0.29215
0.23597	-0.09229
0.036098	0.105147
0.3006278	0.384716
0.483241	0.503099

69	67.78	74	0.755509	0.643078	0.723721	0.61129	0.52772
70	71.56	77	0.643078		0.61129		0.37676
							0.442403
							0

Zx is calculated using the ~~too~~ formula
$Zx = (A-E) / SQRT(E)$ derived from the Poisson Distribution assumption.

Now,

$$r(j) = \frac{\Sigma_{i=1}^{m-j} (Zx - \bar{Z})(Z_{x+1} - \bar{Z})}{\{(m-j)/m * \Sigma_{i=1}^{m} (Zx - \bar{Z})^2\}}$$

$r_1 = 0.55296$

And,

the T ratio is $r_1 * Sqrt(11) = 1.83$.

The T Ratio is positive which suggest that the rates are over graduated and value $1.83 < 1.96$ and hence we do not reject the Null Hypothesis that the graduated rates are the true rates underlying the observed data.

Q13.

(a) Standardised deviations test:- The observations of Zx are divided into 6 groups
[- (-∞, -2), (-2, -1) (-1, 0) (0, 1) (1, 2) (2, ∞.]
The expected number of Zx that should be in each group is compared with the actual numbers.

Range	(-(∞, -2))	(-2, -1)	(-1, 0)	(0, 1)	(1, 2)	(2, ∞)
Expected	0.4	2.8	6.8	6.8	2.8	0.4
Actual	2	1	9	4	2	2
$(A - E)^2/E$	6.4	1.2	0.7	1.2	0.2	6.4

The χ^2 statistic will have a χ^2 distribution with 5 degrees of freedom (as 6 intervals are used and we have constrained the total to 20).

The observed value of 12.3109 is higher than the critical value of 11.070 and hence the graduation is not a good fit.

(b) Signs test :-

The number of positive signs is 8 and negative is 12. The number of signs is ~~Binomi~~ Binomial (20, 1/2). The probability of obtaining 8 positive signs in 20 observation is (2)(0.2517), which is higher than 5%. Hence there is no evidence to reject the graduated.

(c) Grouping of signs test :

The number of positive deviations is 8 and negative is 12. From the tables, the critical value when n_1 is 8 and n_2 is 12 is 2. Since we have observed 2 groups of positive deviations, we reject the graduated rates and conclude there is evidence of grouping of deviations of the same sign.

Q14.

(i) "Undergraduation" occurs when too much emphasis is given to goodness of fit. Undergraduated rates adhere closely to the crude rates, but the resulting rates do not show a smooth progression from age to age.

"Overgraduation" occurs when too much emphasis is given to smoothness. Overgraduated rates show a smooth progression from age to age; but the resulting rates do not adhere closely to the crude rates.

(ii) The chi-squared test is for the overall fit of the graduated rates to the data. The test statistics is $\sum Z_x^2$, where

$$Z_x = \frac{(O - E_x \mathring{q}_x)}{\sqrt{E_x \mathring{q}_x (1-\mathring{q}_x)}}$$

However, considering the fact that q_x is very small, we redefined the Z_x as below

$$Z_x = \frac{(O_x - E_x \mathring{q}_x)}{\sqrt{E_x \mathring{q}_x}}$$

Page No.
Date

The calculations are given in the table below:-

Average Age	O_x	q_x	Expected Deaths	Z_x	Z_x^2
23	2	~~0.00889~~ 0.00220	1.98	0.01421	0.00020
28	4	0.00240	2.88	0.65997	0.43556
33	5	0.00260	3.38	0.88116	0.77645
38	7	0.00360	5.40	0.68853	0.47407
43	8	0.00540	5.94	0.84523	0.71441
48	9	0.00960	7.20	0.67083	0.45000
53	9	0.01500	9.75	-0.24019	0.05769
58	5	0.02300	8.05	-1.07498	1.15559

$$\Sigma Z_x^2 = 4.06397$$

The test statistic has a chi- squared distribution with degrees of freedom given by number of age groups less 1 for the parametric function and further reduction for using the standard table.

The critical value of chi-squared distribution with 6 degree freedom at 5% level is 12.59

Since $4.06397 < 12.59$, there is no evidence to reject the null hypothesis that the graduated rates are the true rates underlying the crude rates.

Page No.
Date

(iii) Signs test:

(a) The Signs test looks for overall bias.

(b) If the null hypothesis is true, the number of positive signs is distributed Bionomial (8, 0.5). From the table, we observe that there are 6 positive signs.

$$\text{Prob}\left(\begin{array}{l}\text{observed number}\\ \text{of positive signs} <= 6\end{array}\right) = 1 - \text{Prob}\left(\text{positive signs} > 6\right)$$

$$= 1 - \left\{\binom{8}{7} + \binom{8}{8}\right\} * (0.5)^8$$

$$= 1 - 0.035156$$

$$= 0.965$$

This greater than 0.025 (two tailed test).

(c) We cannot reject the null hypothesis and conclude that the graduated rates are not systematically higher or lower than the crude rates.

Grouping of Signs test

(a) the grouping of signs test looks for run or clumping of deviations with the same sign for overgraduation.

(b) We have total 8 age groups with 6 positive signs 2 and negative signs. There is only one run in this analysis

Pr (one positive run) =

$$\frac{\left\{\binom{5}{0}\binom{3}{1}\right\}}{\binom{8}{6}} = 3/28 = 0.10714.$$

This is greater than 0.05 (using one-tailed test).

(c) We accept the null hypothesis that graduated rates are ~~true~~ true underlying the crude rates.