

Statistical and Risk Modelling 1

Assignment 2

1. There are 2 avenues of thought that support graduation:

i. Theoretical Argument:

The crude estimates move erratically with age, which is not in line with the intuitive idea that μ_x varies smoothly with age. It follows the argument that a crude estimate for μ_x carries information about the values of μ_{x-1} and μ_{x+1} . By smoothing we can make use of the data at adjacent ages to improve the estimate at each age.

ii. Practical Argument:

We generally use life tables to compute financial quantities, it is desirable that quantities such as insurance premiums progress smoothly with age. It would be difficult to explain irregularities in practice. In practice one never applies any data to a financial problem without investigating its suitability.

However, care must be taken while data collection as graduation relies on already available data and cannot remove any bias from faulty data collection.

2.

The aims of graduation are:

- i. To produce a smooth set of rates that are suitable for a particular purpose.
- ii. To remove random sampling errors
- iii. To use the information available from adjacent ages.

The features of graduation are:

- i. Smoothness of data.
- ii. Adherence to data.
- iii. Suitability for the purpose on hand.

3.

- i. The company will want to perform graduation as:
They will want a smooth set of rates that are suitable for the pensioners mortality projection.
Random sampling errors can be removed.
They would want to use information available from adjacent ages to price their policies more fairly.

ii.

- a. Graduation with respect to parametric formula.
- b. Graduation in with reference to a standard table.
- c. Graduation using spline functions.

iii. Using the statistic: $\sum \frac{(O_i - E_i)^2}{E_i} \sim \chi_{n-1}^2$

We get a statistic value of 18.24967.

Comparing with critical point, at 7 degrees of freedom

There is sufficient evidence to reject H0 at the 5% level.

Hence, it is reasonable to assume that the graduation does not fit the data well.

iv.

a. 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 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.

b. Cumulative Deviations Test

The cumulative deviations test, tests for overall bias of the data.

Under H0: $D_x \sim N(E_x^o * \mu_x^o, E_x^o * \mu_x^o)$

$$\therefore \sum D_x - E_x^o * \mu_x^o / \sqrt{\sum E_x^o \mu_x^o} \sim N(0,1) = 187.14 / 87.99 = 2.1266$$

There is sufficient evidence to reject H0 at the 5% level. Hence, it is reasonable to assume that bias exists.

v.

4.

H0: The standard table rates are the true underlying rates for the term assurance policy holders.

$$\text{Statistic: } \sum (D_x - E_x^c \mu_x^s) / \sqrt{\sum E_x^c \mu_x^s} \sim N(0,1)$$

Age	Exposed to risk	Number of deaths	Standard table mortality rates	Zx
40	50000	87	0.21%	-0.42385
41	48560	84	0.22%	-0.68018
42	47190	101	0.24%	-0.35494
43	44100	112	0.26%	-0.07443
44	43600	123	0.28%	-0.01287
45	40400	110	0.31%	-0.42369
46	37280	108	0.34%	-0.54623
47	35370	122	0.38%	-0.35132
48	32100	150	0.42%	0.373741
49	29000	139	0.47%	0.058202
50	26200	151	0.52%	0.368527
Statistic value				-2.06704

Hence there is sufficient evidence to reject H0 at the 5% level and it is reasonable to assume that the true mortality rate is not reflected by the standard table rates.

5.

i.

a. Maximum smoothness would be achieved by ignoring the plotted estimated rates and drawing a graduation curve which is very smooth, e.g., 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, e.g., graduation by a mathematical form (the chosen functional form will ensure smoothness), reference to a standard table (a simple relationship with an already smooth set of standard table rates will ensure smoothness). 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 smoothness and adherence to data. The graduation process must be repeated until both criteria are satisfied.

ii.

a. Provided a formula with a small number of parameters is chosen the resulting graduation will be acceptably smooth.

b. The graduation should be tested for smoothness using the third differences of the graduated rates which should be small in magnitude and progress regularly. A further iterative process, which involves manual adjustment of the graduation (called 'hand-polishing') is sometimes necessary to ensure smoothness.

iii. Using the statistic:

$$\sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{n-1}$$

Age	Exposed to risk	Number of deaths	a + bx	qx	Expected deaths	Chi Square Statistic
30	250,000.00	245	- 7.01400	0.000898397	224.5991811	1.853049549
31	250,000.00	276	- 6.89780	0.001008986	252.2464792	2.236819131
32	250,000.00	313	- 6.78160	0.001133173	283.2931937	3.115127222
33	250,000.00	323	- 6.66540	0.001272625	318.1562967	0.073741936
34	250,000.00	339	- 6.54920	0.001429215	357.3036427	0.937643212
Test Statistic						8.21638105

5% critical value is 7.815 at 3 degrees of freedom. Hence, we have sufficient evidence to reject H_0 at the 5% level, and it is reasonable to assume that the estimates do not fit the data well.

6.

- i. Using age bands might not be the best idea to determine the premiums charged as there can be significant variations within each group. There would be a risk of overcharging or under charging particular age groups as a result. Additionally, we cannot determine any undue fluctuations within age bands which would also impact the graduation that we may undertake.

- ii. Using the test statistic:

$$\sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{n-1} \Rightarrow \sum z_x^2 \sim \chi^2_{n-1}$$

Required test statistic is: 4.976846

5% critical value at 7 degrees of freedom is 14.07. There is insufficient evidence to reject H0 at the 5% level, hence it is reasonable to assume that the graduated rates fit the data well.

Test for smoothness:

Age	μ_x^o	$\Delta\mu_x^o$	$\Delta^2\mu_x^o$	$\Delta^3\mu_x^o$
18-22	0.0061	0.0070	0.00610	0.00330
23-27	0.0131	0.0131	0.00940	0.00330
28-32	0.0262	0.0225	0.01270	0.00200
33-37	0.0487	0.0352	0.01470	-0.00090
38-42	0.0839	0.0499	0.01380	-0.00440
43-47	0.1338	0.0637	0.00940	-0.00760
48-52	0.1975	0.0731	0.00180	
53-57	0.2706	0.0749		
58+	0.3455			

The 3rd order difference for this data is not that far off from the initial values of the data themselves. Additionally, here is an anomaly at the age range 33-37 . This points to the graduated rates note being smooth enough.

7. The Lee-Carter model has 2 factors, age and period and is defined as follows:

$$\ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}$$

Where, $(m_{x,t})$ is the central rate of mortality for age x at time t .

a_x defines the general shape of mortality.

b_x measures the change in rates with respect to an underlying trend in the level of mortality k .

$\varepsilon_{x,t}$ are independently distributed error terms with mean 0 and some variance.

The applicable constraints are:

$$\sum_x b_x = 1 \text{ and } \sum_t k_t = 0$$

8.

The mortality projection by expectation method involves the use of reduction factors, $R_{x,t}$, which measure the proportion by which the mortality rate at age x, q_x , is expected to be reduced by future year t.

It can be written as: $R_{x,t} = \alpha_x + (1 - \alpha_x)(1 - f_{(n,x)})^{\frac{t}{n}}$

Where, α_x is the ultimate reduction factor and $f_{n,x}$ represents the proportion of the total decline expected to occur in 'n' years. Expert opinion is used to set the targets α_x and $f_{n,x}$.

9.

i.

a. H0: The estimates of crude rates are similar to standard mortality rates

H1: The estimates of crude rates are significantly different from standard mortality rates.

Age	Initial Exposed to Risk	Standard mortality rates	Expected deaths	Actual deaths	z_x $(z_x = \frac{d_x - E_x q_x^s}{\sqrt{E_x^c q_x^s}})$	z_x^2
x	E_x	q_x	E	A		
50	2305	0.0064	14.75	15	0.065094	0.004237
51	2475	0.0069	17.08	16	-0.261324	0.0682902
52	2705	0.0075	20.29	22	0.379625	0.144115
53	2900	0.0081	23.49	23	-0.101101	0.0102214
54	3170	0.0087	27.58	27	-0.1104411	0.0121972
55	6730	0.0094	63.26	66	0.344497	0.1186782
56	6875	0.0101	69.44	67	-0.2928094	0.08573735
57	8190	0.0109	89.27	88	-0.134416	0.01806766
58	8200	0.0117	95.94	102	0.6186895	0.3827767
59	7680	0.0119	91.39	80	-1.1914461	1.4195438
60	7160	0.0121	86.64	85	-0.1761914	0.0310434
Total	58390		599.12	591.00		2.2949079

Since, the mortality experience was compared with a published standard mortality table

The degrees of freedom will be maximum 11

$$X_{11,5\%}^2 = 19.98$$

Since, $2.2949079 < 19.98$, we do not have sufficient evidence to reject out null hypothesis at 5% level of significance.

Thus, we conclude that the estimates of crude mortality rates are similar to that of standard mortality rates.

b. H0: There is no overall bias present in the standard mortality rates

H1: There is an overall bias present in the standard mortality rates

Under null hypothesis

$$\frac{\sum_x d_x - \sum_x E_x^c \mu_x^o}{\sqrt{\sum_x E_x^c \mu_x^o}} \sim N(0,1)$$

$$Z = \frac{591 - 599.12}{\sqrt{599.12}}$$

$$Z = -0.331741$$

Since, cumulative deviations test is a two-sided test,

$$-1.96 < -0.331741 < 1.96$$

Thus, we do not have sufficient evidence to reject our null hypothesis.

Hence, we conclude that there is no overall bias present in the standard mortality rates.

c. H0: The rates are not over graduated

H1: The rates are over graduated

Here there are total of 11 Z_x values out of which 4 are positive and 7 are negative

And there are 4 positive groups

$$n_1 = 4, n_2 = 7 \text{ and } G = 4$$

Since $4 > 1$,

We do not have sufficient evidence to reject our null hypothesis

Hence, we conclude that the rates are not over graduated.

10.

Age	Actual claims (Last year)	Expected claims (last year)	Expected claims (current year)	$\frac{\text{Actual claims (LY)}}{\text{Expected claims (LY)}} * \text{Expected Claims (CY)}$	Actual claims (CY)	Z_x	Z_x^2
25	808	810	724	722.2124	719	-0.11954	0.0142888
30	1851	1708	1433	1552.976	1322	-5.861172	34.35334
35	1400	1084	444	573.432	500	-3.06651	9.403484
40	1562	1705	1397	1279.832	1207	-2.03585	4.144685
45	1366	1572	1465	1273.022	1177	-2.69124	7.242785
50	1296	1643	1209	953.6604	905	-1.57572	2.482891
55	2200	2911	2436	1841.017	1798	-1.002562	1.005131

$$\sum_x z_x^2 = 58.646605$$

$$X_{7,5\%}^2 = 14.07$$

Since, $58.646605 > 14.07$, we have sufficiently strong evidence to reject the null hypothesis at 5% level of significance.

11. H0: The graduated rates are similar to the estimates of the crude rates.

H1: The graduated rates are significantly different from the estimates of the crude rates.

Age	E_x	$q(circle)_x$	$E_x * q(circle)_x$	d_x	Z_x	Z_x^2
20	1060	0.0125	13.25	12	-0.343401	0.11793
21	1250	0.0118	14.75	14	-0.19528	0.038136
22	1210	0.0134	16.214	16	-0.05315	0.002825
23	700	0.0102	7.14	9	0.69609	0.484538
24	875	0.0111	9.7125	11	0.41313	0.170673
25	950	0.0137	13.015	15	0.550223	0.302745
26	805	0.0126	10.143	10	-0.044901	0.0020161
27	1390	0.0127	17.653	16	-0.393427	0.1547844
28	1080	0.0122	13.176	17	1.05348	1.109819
29	1310	0.0131	17.161	14	-0.763051	0.582246

$$\sum_{all\ x} z_x^2 = 2.9657125$$

$$\chi_{10,5\%}^2 = 18.31$$

Since, $2.9657125 < 18.31$, we do not have sufficient evidence to reject our null hypothesis.

Hence, we conclude that the graduated rates are similar to the estimates of the crude rates.

12.

- i. The role of graduation in producing the national life table is as follows:
- To smoothen the estimates of crude rates which can then be used for specific purposes.
 - To remove any sampling errors.
 - To obtain information from adjacent ages to estimate the rate at a particular age.

- ii. Procedure to use for graduating the crude rates:

Step 1: To select a method of graduation

Step 2: Select a statistical method to provide the graduated rates

Step 3: Test the graduated rates using several statistical tests

Step 4: The process will be repeated until the desirable level of smoothness and adherence to data is achieved.

- iii.

Consistency between graduated rates and a published standard table implies two concepts: the shape of the mortality curve over the range of ages and the level of mortality rates.

- iv.

Serial correlations test is one of the tests that is performed while graduation. The purpose of serial correlations test is to test for clumping of deviations of the same sign (an issue that will not be picked up by a chi-squared test). If clumping is present, then the graduation has the wrong shape.

v.

Age	Expected deaths	Actual deaths	z_x	z_{x+1}	$z_x - \bar{z}$	$z_{x+1} - \bar{z}$	$(z_x - \bar{z})^*$ $(z_{x+1} - \bar{z})$	$(z_x - \bar{z})^2$
60	38.3	36	- 0.3716 451	-1.014146	-0.403433	-1.045934	0.421964	0.162758
61	40.45	34	- 1.0141 46	-0.693173	-1.045934	-0.724961	0.7582614	1.093978
62	42.52	38	- 0.6931 73	-0.6831125	-0.724961	-0.7149	0.5182746	0.525569
63	44.56	40	- 0.6831 125	0.6332004	-0.7149	0.6014127	-0.42995	0.511082
64	47.63	52	0.6332 004	-0.4539797	0.6014127	-0.485767	-0.2921464	0.361697
65	51.25	48	- 0.4539 797	0.22178141	-0.485767	0.189994	-0.0922928	0.235967
66	55.35	57	0.2217 8141	0.585212	0.189994	0.5534243	0.1051473	0.036098
67	60.45	65	0.5852 12	0.7269434	0.5534243	0.6951557	0.384716	0.306279
68	63.22	69	0.7269 434	0.755509	0.6951557	0.7237213	0.503099	0.483241
69	67.78	74	0.7555 09	0.6430781	0.7237213	0.6112904	0.44240388	0.523773
70	71.56	77	0.6430 781		0.6112904		0	0.373676

$$\bar{z} = 0.0317877$$

Using the formula,

$$r_j = \frac{\left[\left(\frac{1}{m-j} \right) * \sum_{x=1}^{m-j} (z_x - \bar{z}) * (z_{x+1} - \bar{z}) \right]}{\left[\left(\frac{1}{m} \right) * \sum_{x=1}^m ((z_x - \bar{z})^2) \right]}$$

Here m = 11 and j = 1

$$r_1 = \frac{\left[\left(\frac{1}{11-1}\right) * 2.319477\right]}{\left[\left(\frac{1}{11}\right) * \sum_{x=1}^m 4.614118\right]}$$

$$r_1 = 0.5529604$$

We know,

$$r_j * \sqrt{m} \sim N(0,1) \text{ approximately}$$

$$Z = r_1 * \sqrt{11} = 0.5529604 * \sqrt{11}$$

$$Z = 1.8339622$$

Since $1.8339622 < 1.96$, we do not have sufficient evidence to reject our null hypothesis.

Thus, we conclude that the graduated rates are similar to the estimates of the crude rates.

13.

i. Individual standardised deviations test

Range	$(-\infty, -3)$	$(-3, -2)$	$(-2, -1)$	$(-1, 0)$	$(0, 1)$	$(1, 2)$	$(2, 3)$	$(3, \infty)$
Actual	0	2	1	9	4	2	2	0
Expected	0	0.4	2.8	6.8	6.8	2.8	0.4	0

Here ideally should be 10 values between 0 and ∞ . However, there are 8 values actually lying between 0 and ∞ .

There are a total of 4 values in the ranges $(-3, -2)$ and $(2, 3)$ indicating the presence of outliers.

$$\sum z_x^2 = 16.05042$$

$$X_{7,5\%}^2 = 14.07$$

Since, $16.05042 > 14.07$, we have sufficient evidence to reject the null hypothesis.

ii. Signs test

There is a total of 20 z_x values out of which 12 are negative and 8 are positive.

Under null hypothesis,

Assuming $P \sim \text{Binomial}(20, 0.5)$ where P is the number of positive values

$$2 * P(P \leq 8) = 2 * 0.2517 = 0.5034$$

Since, $0.5034 > 0.05$, we do not have sufficient evidence to reject our null hypothesis at 5% level of significance.

14.

- i. Under graduation refers to a situation in graduation wherein the graduated rates adhere more to data relative to smoothness.

Over graduation refers to a situation in graduation wherein the graduated rates are smoother relative to adherence to data.

- ii. H0: The graduated rates are similar to the estimates of the crude rates.

H1: The graduated rates are significantly different from the estimates of the crude rates.

Age group	Average age	E_x	θ_x	q_x^s	q_x^o	$E_x q_x^o$	$\theta_x - E_x q_x^o$	z_x	z_x^2
20-25	23	900	2	0.0011	0.0022	1.98	0.02	0.014213	0.000202
26-30	28	1200	4	0.0012	0.0024	2.88	1.12	0.659966	0.435556
31-35	33	1300	5	0.0013	0.0026	3.38	1.62	0.881164	0.776450
36-40	38	1500	7	0.0018	0.0036	5.40	1.60	0.688530	0.474074
41-45	43	1100	8	0.0027	0.0054	5.94	2.06	0.845228	0.714411
46-50	48	800	9	0.0045	0.0090	7.20	1.80	0.670820	0.450000
51-55	53	650	9	0.0075	0.0150	9.75	-0.75	-0.24019	0.057692
56-60	58	350	5	0.0115	0.0230	8.05	-3.05	-1.07498	1.155590

$$\sum z_x^2 = 4.063974$$

$$\chi_{6,5\%}^2 = 12.59$$

Since, $4.063974 < 12.59$, we do not have sufficient evidence to reject the null hypothesis at 5% level of significance.

Thus, we conclude that the graduated rates are similar to the estimates of the crude rates.

iii. Signs test

The signs test is a simple test for overall bias.

It is designed to identify the second deficiency of the chi-squared test, i.e., failure to detect where there is an imbalance between positive and negative deviations.

There are a total of 8 values of z_x out of which 6 are positive and 2 are negative

Under null hypothesis,

$P \sim \text{Binomial}(8, 0.5)$ where P are the number of positive values

$$2 * P(P \geq 6) = 2 * [1 - P(P \leq 5)] = 2 * [1 - 0.8555] = 0.289$$

Since, $0.289 > 0.05$, we do not have sufficient evidence to reject our null hypothesis at 5% level of significance.

Thus, we conclude that there is no overall bias in the graduated rates.

iv. Grouping of signs test

The grouping of signs test is a test for over graduation

Here, there are a total of 8 values of z_x out of which 6 are positive and 2 are negative

Also, there is 1 group of positive values

Since this is greater than 0, we do not have sufficient evidence to reject the null hypothesis at 5% level of significance.

Thus, we conclude over graduation is absent.

15.

i. H0: The graduated rates are similar to the estimates of the crude rates

H1: The graduated rates are significantly different from the estimates of the crude rates

t	$\hat{q}_t$	$q(circle)_t$	E_t	$\hat{q}_t * E_t$	$q(circle)_t * E_t$	$1 - q(circle)_t$	z_t	z_t^2
0	0.1515	0.1324	3706	561.45900	490.67440	0.86760	3.43070	11.76968
1	0.0963	0.1051	3262	314.13060	342.83620	0.89490	-1.63884	2.68579
2	0.0746	0.0832	5171	385.75660	430.22720	0.91680	-2.23917	5.01388
3	0.0683	0.0654	3243	221.49690	212.09220	0.93460	0.66799	0.44621
4	0.0553	0.0521	5147	284.62910	268.15870	0.94790	1.03306	1.06722
5	0.0396	0.0438	3850	152.46000	168.63000	0.95620	-1.27341	1.62157
6	0.0432	0.0383	3080	133.05600	117.96400	0.96170	1.41694	2.00773
7	0.0514	0.0381	2094	107.63160	79.78140	0.96190	3.17916	10.10707
8	0.0381	0.0423	1083	41.26230	45.81090	0.95770	-0.68672	0.47158
9	0.0437	0.0515	1511	66.03070	77.81650	0.94850	-1.37184	1.88195

$$\sum z_t^2 = 37.072675$$

$$X_{7,5\%}^2 = 14.07$$

Since, $37.072675 > 14.07$, we have sufficiently strong evidence to reject the null hypothesis.

Thus, we conclude that there graduated rates are significantly different from the estimates of the crud rates.

ii. Standard deviation test

H0: The individual standardised deviations follow standard normal distribution

H1: The individual standardised deviations do not follow standard normal distribution

Here, as we can see that there are two values that are greater than 3 and one value that is less than -2. This indicates the presence of outliers.

Thus, by the test of eye, we can conclude that we reject the null hypothesis and that the individual standardised deviations do not follow standard normal distribution.

iii. H0: There is overall bias in the graduated rates

H1: There is no overall bias in the graduated rates

There are 10 z_t values out of which 5 are positive and 5 are negative.

Since 50% are positive and 50% are negative values, we do not have sufficient evidence to reject our null hypothesis.

Thus, we conclude that there is no overall bias in the graduated rates.