6 Graduation and Statistical Tests

6.1 Introduction

By postulating a model of mortality and then incorporating some actual data we will have obtained estimates of \( \mu_x \) or \( q_x \) at successive ages. These estimates are called crude mortality rates and are subject to:

1. Random Fluctuations - since on statistical considerations any estimate is just a single sample from the sampling distribution of the estimator.

2. Irregularity - since we have random sampling, the crude estimates of mortality will not proceed smoothly over the years of age.

Intuitively we expect a smooth progression of mortality rates given large enough samples; we do not expect mortality rates to move in discrete steps since the main factor influencing human mortality is the gradual aging process.

In particular, in small studies with low levels of exposed to risk the estimates obtained can be unreliable. This would also be the case in larger studies at high ages where exposed to risk is low and mortality rates are high.

We prefer to use smoothly progressing rates in actuarial work, not just from a theoretical viewpoint, but because smooth rates of decrement imply a smooth progression of premium rates as ages rise. This is important for a variety of practical reasons including marketing.

6.2 The graduation process

In Figure 1 the dots represent the crude rates and the line represents the graduated rates. The graduated rates are then our estimates of the ‘true underlying’ rates of mortality. The purpose of graduation is, therefore:

- To provide a smooth set of rates for practical use
- To reduce random sampling errors (principally by making use of estimates at ages \( x - 2, x - 1, \ldots \) and \( x + 1, x + 2, \ldots \) to improve our estimate at age \( x \)).

It is important to note that the graduation process cannot remove non-random error or bias in our estimates which might arise from poor data collection methods or an inappropriate statistical model.

There are two basic steps in a graduation which we will cover in this course:

1. The construction of a set of graduated rates or forces of mortality from a set of crude estimates. This is a curve fitting problem.

2. The testing of graduated estimates to determine whether they are
   
   (a) acceptably smooth, and
   
   (b) an acceptable fit to our original data.
Figure 1: A graduation of $q_x$ for $70 \leq x \leq 80$

We will be concerned with a variety of statistical tests for goodness of fit and the detailed examination of residuals.

**Note:** The two aims of smoothness and goodness of fit may be in conflict with each other and an element of compromise is often needed in a graduation. We do not want to smooth out true features of the data. An element of judgement and experience is needed in most graduation exercises.

### 6.3 Examples of poor graduation

In the left hand panel of Figure 2 the graduated rates are a poor fit to the crude rates; mortality rates tend to be overestimated at young ages and underestimated at older ages. In the right hand panel the graduated rates adhere too closely to the graduated rates, and there is no smooth progression of rates from age to age.

### 6.4 Testing smoothness

We wish to obtain graduated rates that do not smooth out intrinsic features of the data or adhere to the crude rates too closely. A mathematical definition of smoothness might be for derivatives of some high order to exist. This might however permit graduations which vary too much (particularly within short age ranges and at extreme ages). In practice, actuaries tend to use the following criterion:

- The third differences of the graduated estimates should be small in comparison with the estimates themselves and should progress regularly. This is equivalent to saying that the graduated rates should be roughly quadratic.
Definition: Let \( \hat{q}_x \) and \( \hat{\mu}_x \) denote the graduated estimates of \( q_x \) and \( \mu_x \) respectively, i.e. \( \hat{q}_x \) and \( \hat{\mu}_x \) represent the smoothed values of \( q_x \) and \( \mu_x \).

The differences are denoted as follows:

- First difference: \( \Delta \hat{q}_x = \hat{q}_x - \hat{q}_{x-1} \)
- Second difference: \( \Delta^2 \hat{q}_x = \Delta \hat{q}_x - \Delta \hat{q}_{x-1} \)
- Third difference: \( \Delta^3 \hat{q}_x = \Delta^2 \hat{q}_x - \Delta^2 \hat{q}_{x-1} \)

Example (in rates per 1000)

<table>
<thead>
<tr>
<th>( x )</th>
<th>( q_x )</th>
<th>( \Delta \hat{q}_x )</th>
<th>( \Delta^2 \hat{q}_x )</th>
<th>( \Delta^3 \hat{q}_x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>1.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>31</td>
<td>1.2</td>
<td>0.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>32</td>
<td>1.6</td>
<td>0.4</td>
<td>0.2</td>
<td>-</td>
</tr>
<tr>
<td>33</td>
<td>2.3</td>
<td>0.7</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>34</td>
<td>3.2</td>
<td>0.9</td>
<td>0.2</td>
<td>-0.1</td>
</tr>
</tbody>
</table>

Note that when you are comparing two alternative graduations (with different \( \Delta^3 \hat{q}_x \) values) there are no definite rules about which one to choose. A criterion sometimes used is to choose the one which gives the smaller value of \( \sum_x |\Delta^3 \hat{q}_x| \) or \( \sum_x |\Delta^3 \hat{\mu}_x| \). Other tests are however also required, as will be discussed later.

As discussed, the smoothing should never remove intrinsic features of the data. Where we know from previous studies and from the consideration of related tables that some particular feature is likely to occur then a graduation method or formula can reasonably be chosen to accommodate the anticipated effect.
Consistency with other mortality tables is often an attractive feature of a graduation. This will be discussed when we consider graduation with reference to some standard table. Standard tables can be used to improve estimates of mortality (or some other decrement) where uncertainty is high. This might be needed at:

- ages where there is greater uncertainty (e.g. low exposed to risk).
- in an ‘assured lives’ investigation where few policies are sold at certain ages.

Other methods of coping with sparse data include:

- extrapolation (which can be very dangerous, particularly if a complex mathematical formulae is being used).
- grouping of data to increase exposure (which can invalidate statistical tests, particularly variance estimates.)

When a graduation has been completed, it is important to check its main features against other tables of interest. This is particularly important where estimates have been based on sparse or grouped data or, even worse, where extrapolation has been used. The following comparisons might be made:

A proposed table of male assured lives mortality might be considered in the light of the following experiences:

- the previous table developed for similar lives
- female mortality (past/present)
- select and ultimate rates
- current annuity rate mortality tables
- current population mortality tables.

Differences which are seen should be capable of rational explanation e.g. we expect female mortality to be lighter than male mortality, rates to improve over time, for annuitant mortality to be lighter than assured lives mortality, etc.

We may also examine the impact of the new graduated rates on financial functions of interest e.g. $A_x$, $a_x$, etc.

### 6.5 Testing adherence to data

In this section we will consider the Binomial Model for $q_x$ and the Poisson approximation to the two state model for $\mu_x$.

**Binomial Model**

We assume we have a measure of the initial exposed to risk $E_x$ (at age $x$). This corresponds to the ‘number of lives’, say $N$, in the simple Binomial Model. We consider $E_x$ to be the initial population from which $d_x$ deaths will occur over the year of age $x$ to $x+1$.

**Note:** In practice $E_x$ is often estimated from the central exposed to risk as $E_x \approx E_x^c + \frac{1}{2}d_x$. Preferably $q_x$ is estimated from $\mu_x$ without ever having to calculate $E_x$.

The estimator $\hat{q}_x = \frac{D_x}{E_x}$ has mean $q_x$ and variance $\frac{q_x(1 - q_x)}{E_x}$.
Poisson Model

In this case the number of deaths is assumed to have a Poisson Distribution with parameter \( \mu_x E_x^c \). We assume that an estimate of the constant \( \mu \) for the age interval \([x, x+1]\) is an estimate of \( \mu_{x+\frac{1}{2}} \). Then \( \tilde{\mu}_{x+\frac{1}{2}} = \frac{D_x}{E_x^c} \) has

\[
\begin{align*}
\text{Mean: } & \mu_{x+\frac{1}{2}}, \\
\text{Approximate variance: } & \frac{\mu_{x+\frac{1}{2}}}{E_x^c}.
\end{align*}
\]

We wish, amongst other things, to test the adherence to data of the graduation. Are the number of deaths in each age group ‘close’ to those ‘expected’ under graduation assumptions? We study the deviations

- Binomial: \( D_x - E_x \hat{q}_x \)
- Poisson/2 State Markov: \( D_x - \hat{E}_x \tilde{\mu}_{x+\frac{1}{2}} \)

We often write the expected number of deaths as \( \hat{D}_x \).

We also define the standardised deviations (or residuals)

- Binomial: \( Z_x = \frac{D_x - E_x \hat{q}_x}{\sqrt{E_x \hat{q}_x(1 - \hat{q}_x)}} \approx \frac{D_x - \hat{D}_x}{\sqrt{D_x}} \) if \( \hat{q}_x \) is small.

- Poisson/2 State Markov: \( Z_x = \frac{D_x - \hat{E}_x \tilde{\mu}_{x+\frac{1}{2}}}{\sqrt{\hat{E}_x \tilde{\mu}_{x+\frac{1}{2}}}} = \frac{D_x - \hat{D}_x}{\sqrt{\hat{D}_x}}. \)

If the expected number of deaths is reasonably large then we invariably use the normal approximation and assume that \( Z_x \sim N(0,1) \).

Note: If we do not have a homogeneous group of lives aged \( x \) then \( q_x \) will vary within the group and this will distort the estimate of variance, and may invalidate the test. This may occur, for example, when

- age ranges are too wide, or
- we include males and females in the study, or
- we have duplicate policies, i.e., one individual has several policies. This can affect assured lives tables. See later.
6.6 $\chi^2$ test

This is a test for overall ‘goodness of fit’. There are two cases.

1. We can test whether our data are consistent with a standard table with

$$X = \sum \frac{(D_x - \hat{D}_x)^2}{\hat{D}_x} = \sum Z_x^2$$

where $D_x$ are the actual numbers observed and $\hat{D}_x$ are the expected number observed under the tested model. The higher the value of $X$ the greater are the differences between our observed data and our graduated values, i.e., we have a one-sided test whose null distribution is approximately:

$$X = \sum Z_x^2 \sim \chi^2_N$$

where $N$ is the number of age groups.

2. However, if a graduation method has made use of the $D_x$ data in the construction of $\hat{q}_x$ or $\hat{\mu}_{x+\frac{1}{2}}$ (as is usually the case) then the $\chi^2$ statistic will have fewer than $N$ degrees of freedom. The reduction in the degrees of freedom will depend on the method of graduation (see example below).

**N.B.** It is important in practical calculations to group data where expected numbers are low (say < 5) as $\frac{(D_x - \hat{D}_x)^2}{\hat{D}_x}$ is sensitive to small changes in $\hat{D}_x$. (Try some examples as an exercise). We often group ages until $\hat{D}_x > 5$ whilst accepting that this reduces the sensitivity of our test.

**Note:** Group ages until $\hat{D}_x > 5$, and **not** $D_x > 5$.

**Example: Degrees of Freedom**

Mortality rates over ages 30 – 34 were estimated by fitting the binomial model with

$$\log \left( \frac{q_x}{1 - q_x} \right) = \alpha + \beta x$$

to the following data.

<table>
<thead>
<tr>
<th>Age $x$</th>
<th>Deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>335</td>
</tr>
<tr>
<td>31</td>
<td>391</td>
</tr>
<tr>
<td>32</td>
<td>428</td>
</tr>
<tr>
<td>33</td>
<td>436</td>
</tr>
<tr>
<td>34</td>
<td>458</td>
</tr>
</tbody>
</table>

The initial exposed to risk for each year of age was 700,000. The parameters were estimated to be $\hat{\alpha} = -10.945$ and $\hat{\beta} = 0.110406$. (Checking this as an exercise with R.)

Does this model provide a good fit to the data?

**Solution** (check details as exercise)

$$\log \left( \frac{q_x}{1 - q_x} \right) = \alpha + \beta x \Rightarrow q_x = \frac{1}{1 + e^{-(\alpha + \beta x)}}$$

Substituting the **estimated** values for $\alpha$ and $\beta$, we can then calculate the expected numbers of deaths. For example, $\hat{q}_{30} = 0.000484 \Rightarrow \hat{d}_{30} = 338.84$. 
<table>
<thead>
<tr>
<th>Age $x$</th>
<th>$d_x$</th>
<th>$\hat{d}_x$</th>
<th>$\frac{(d_x-\hat{d}_x)^2}{\hat{d}_x}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>335</td>
<td>338.84</td>
<td>0.044</td>
</tr>
<tr>
<td>31</td>
<td>391</td>
<td>378.37</td>
<td>0.421</td>
</tr>
<tr>
<td>32</td>
<td>428</td>
<td>422.51</td>
<td>0.071</td>
</tr>
<tr>
<td>33</td>
<td>436</td>
<td>471.80</td>
<td>2.717</td>
</tr>
<tr>
<td>34</td>
<td>456</td>
<td>526.83</td>
<td>9.524</td>
</tr>
</tbody>
</table>

$X = 12.78$

We have five ages but have estimated two parameters so we refer to the upper 5% point of the $\chi^2$ distribution i.e. $\chi^2_{30.95} = 7.815$. We reject the null hypothesis that the model describes mortality over ages 30 to 34 since our test statistic of 12.3 > 7.815. Note the very high contribution to $\chi^2$ at age 34; a plot of the data will reveal the problem at $x = 34$.

The linear logistic model is often used to analyse data in the form of proportions:

$$\log\left(\frac{q_x}{1 - q_x}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k.$$  

This model is a member of the class of generalised linear models and is often helpful in exploring the relationship between a binary response variable and one or more explanatory variables. In our basic example the only explanatory variable was age but others can be added fairly easily - see examples on the Cox model.

In general, we would reject the graduation if the $\chi^2$ statistic exceeded the 95% point of the $\chi^2_{n-t}$ distribution (where $t$ is the reduction in the degrees of freedom). Note that this is not sufficient on its own as a decision criterion because:

- Too small a statistic might suggest overgraduation.
6.7 The standardised deviations test

If the \( \{ Z_x \} \) are indeed \( N(0, 1) \) and independent then we can derive the following proportions from statistical tables (check as an exercise). The proportions expected in various ranges are shown in Figure 4 and in the table.

Figure 4: Proportions in a normal distribution: left panel shows equal cell widths, right panel shows equal cell areas.
We can then compare the expected numbers in each range with those which we have observed. We may need to group data at the ends of the range so we have expected numbers \( \geq 5 \) in each range. This procedure enables us to:

- quickly identify outliers, e.g. \( Z_x \) values outside \((-3, 3)\)
- examine the symmetry of the residuals, e.g. an excessive number of positive \( Z_x \) values would indicate bias
- examine the size of the absolute deviations, e.g. too many small \( Z_x \) values would suggest over-adherence to the data.

By calculating the test statistic

\[
\sum \frac{(\text{Actual no. } Z_x - \text{Expected no. } Z_x)^2}{\text{Expected no. } Z_x}
\]

and comparing this with the upper 5% point on the \( \chi^2_{k-1} \) distribution, we can further test the hypothesis that the \( Z_x \) values come from the \( N(0, 1) \) distribution.

**Note:** We may have \( k \) groups but have constrained the sum of the actual numbers to be equal to the sum of the expected numbers so the degrees of freedom are reduced by 1.

### 6.8 A test for bias - the sign test

Suppose that \( S \) denotes the number of positive \( Z_x \) and that there are \( n \) age groups. If there is no positive or negative bias in the \( \{Z_x\} \) then each \( Z_x \) is equally likely to be positive or negative and so \( S \sim \mathcal{B}(n, \frac{1}{2}) \). Since we wish to identify positive or negative bias, a **two-sided test** is required. Suppose we observe \( S = s \). There are two cases.

- **Small** \( n \): if \( s > n/2 \) then the significance probability is \( 2 \Pr(S \geq s) \) while if \( s < n/2 \) then the significance probability is \( 2 \Pr(S \leq s) \).

- **Large** \( n \): we can use the Normal Approximation with

\[
E(S) = \frac{n}{2}, \quad \text{Var}(S) = n \times \frac{1}{2} \times \frac{1}{2} = \frac{n}{4}
\]

or \( S \sim \mathcal{N}(\frac{n}{2}, \frac{n}{4}) \).

The next test examines non randomness of the signs of the deviations on a smaller scale. These might be ‘runs’ (i.e. a long sequence of positive \( Z_x \)) or ‘groups’ or ‘clumps’ (i.e., where sequences are shorter but not consistent with randomness.

<table>
<thead>
<tr>
<th>Range</th>
<th>% of ( Z_x ) in range</th>
<th>Range</th>
<th>% of ( Z_x ) in range</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-\infty, -3)</td>
<td>0</td>
<td>(-\infty, -0.9674)</td>
<td>16% 16%</td>
</tr>
<tr>
<td>(-3, -2)</td>
<td>2</td>
<td>(-0.9674, -0.4307)</td>
<td>16% 16%</td>
</tr>
<tr>
<td>(-2, -1)</td>
<td>14</td>
<td>(-0.4307, 0)</td>
<td>16% 16%</td>
</tr>
<tr>
<td>(-1, 0)</td>
<td>34</td>
<td>0, 0.4307</td>
<td>16% 16%</td>
</tr>
<tr>
<td>0, 1</td>
<td>34</td>
<td>0.4307, 0.9674</td>
<td>16% 16%</td>
</tr>
<tr>
<td>1, 2</td>
<td>14</td>
<td>0.9674, ( \infty )</td>
<td>16% 16%</td>
</tr>
<tr>
<td>2, 3</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3, ( \infty )</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6.9 Change of sign test

If the $Z_x$ are independent and distributed $N(0, 1)$ then with probability $\frac{1}{2}$ the sign of the $(x + 1)^{th}$ deviation will be the same as the sign of the $x^{th}$ deviation. If $C$ denotes the total number of sign changes then $C \sim B(n - 1, \frac{1}{2})$. If $c$ is the observed value of $C$ then in most cases we are interested in small values of $c$, i.e., **too few sign changes**. We calculate $p = \Pr(C \leq c)$ and reject the graduation if $p \leq 0.05$, say. Again a normal approximation can be used if numbers are sufficiently large.

6.10 Cumulative deviations test

There are two cases: the binomial, and the Poisson/Markov models.

- **Binomial** It is assumed that the deviations $D_x - E_xq_x = D_x - \hat{D}_x$ are approximately normal with mean 0 and variance $E_x\hat{q}_x(1 - \hat{q}_x) \approx \hat{D}_x$. Then

$$W = \sum_x (D_x - E_x\hat{q}_x) = \sum_x (D_x - \hat{D}_x) \sim \mathcal{N}(0, \sum_x E_x\hat{q}_x(1 - \hat{q}_x)) \approx \mathcal{N}(0, \sum_x \hat{D}_x)$$

where the summation may be over some sub-range, for example, of financial importance.

**Note:** The sub-range **must not** be chosen with reference to the data.

- **Poisson/Markov** In the two state model we use

$$W = \sum_x \left( D_x - E_x\hat{u}_{x+\frac{1}{2}} \right) = \sum_x (D_x - \hat{D}_x) \sim \mathcal{N}(0, \sum_x E_x\hat{u}_{x+\frac{1}{2}}) = \mathcal{N}(0, \sum_x \hat{D}_x)$$

**Example:** An assured lives investigation ranged from age 20 to age 100. We might apply this test to the age ranges

- The whole table, i.e., 20 to 100.
- Endowments & Term Assurance, say 20-60.
- Whole life, say 60-100.

This test will detect large positive or negative deviations over some of, or the whole of, the age range. The test is **two-sided** and so a graduation is rejected if the observed cumulative deviation is in the $2\frac{1}{2}\%$ tails of the relevant normal distribution.

**Warning**

Certain methods of graduation (e.g. maximum likelihood and least squares) automatically result in a cumulative deviation $= 0$ over the whole age range in which case the test cannot be applied to the whole age range. For this reason, the test generally performs rather poorly. The test will show up bias in the graduated rates or higher than predicted variance, but only for the age range considered.
6.11 The grouping of signs test (or Stevens test)

This test looks at the number of ‘groups’ or ‘runs’ of positive signs amongst the residuals. We define the test statistic:

\[ G = \text{Number of groups of positive residuals}, \]

and let \( g \) be the observed value of \( G \). Suppose we have \( m \) residuals of which \( n_1 \) are positive and \( n_2 \) are negative. The null hypothesis is that the given \( n_1 \) positive deviations and \( n_2 \) negative deviations are in random order. We are on the lookout that we have seen too few groups since this will indicate non-randomness. The test is one-sided and so we compute

\[ \Pr(G \leq g) \]

and reject the null hypothesis if this is less than 5%, say.

Let \( t \leq g \). We argue

- There are \( \binom{n_1 - 1}{t - 1} \) ways to arrange \( n_1 \) positive signs into \( t \) positive groups.
- There are \( \binom{n_2 + 1}{t} \) ways to arrange \( t \) positive groups among \( n_2 \) negative signs.
- There are \( \binom{n_1 + n_2}{n_1} \) ways to arrange \( n_1 \) positive and \( n_2 \) negative signs.

Thus, the probability of obtaining exactly \( t \) positive groups is

\[ p_t = \frac{\binom{n_1 - 1}{t - 1} \binom{n_2 + 1}{t}}{\binom{n_1 + n_2}{n_1}}. \]

The required significance probability is

\[ \Pr(G \leq g) = \sum_{t=1}^{g} p_t. \]

Alternatively we can use a set of tables, or if \( m \) is large, say \( m > 20 \), we can use a normal approximation

\[ G \sim \mathcal{N}\left(\frac{n_1(n_2 + 1)}{n_1 + n_2}, \frac{(n_1n_2)^2}{(n_1 + n_2)^3}\right). \]

6.12 Serial Correlation Test

This test again addresses the limitations of the \( \chi^2 \) test in detecting groups of deviations of the same sign. It also takes into account the size of such deviations. We examine the tendency of deviations at adjacent ages to be positively correlated.

Consider the first \((n - 1)\) deviations \( \{Z_1, \ldots, Z_{n-1}\} \) as one set of data and the last \((n - 1)\) deviations \( \{Z_2, \ldots, Z_n\} \) as another set. Let

\[ \bar{Z}_1 = \sum_{j=1}^{n-1} \frac{Z_j}{(n - 1)}, \quad \bar{Z}_2 = \sum_{j=2}^{n} \frac{Z_j}{(n - 1)} \]
and define $r_1$ as the correlation between the two sets, i.e.

$$r_1 = \frac{\sum_{x=2}^{n} (Z_{x-1} - \bar{Z}_1)(Z_{x} - \bar{Z}_2)}{\sqrt{\sum_{x=1}^{n-1} (Z_{x} - \bar{Z}_1)^2 \sum_{x=2}^{n} (Z_{x} - \bar{Z}_2)^2}}.$$ 

If $n$ is large, we often use the approximation

$$\bar{Z}_1 \approx \bar{Z}_2 \approx \frac{\sum_{j=1}^{n} Z_j}{n} = \bar{Z}$$

and then take

$$r_1 \approx \frac{\sum_{x=2}^{n} (Z_{x-1} - \bar{Z})(Z_{x} - \bar{Z})}{\sum_{x=1}^{n} (Z_{x} - \bar{Z})^2}$$

which, under the null distribution of randomness of the residuals, has asymptotic distribution

$$R_1 \sim \mathcal{N}\left(0, \frac{1}{n-1}\right).$$

Students of time series will recognise the lag 1 serial correlation coefficient. Since we are concerned with the tendency of deviations at adjacent ages to be positively correlated, the test is one-sided with large positive values of $r_1$ providing evidence against $H_0$; the significance probability is $\Pr(R_1 > r_1)$.

The lag 1 serial correlation coefficient can easily and obviously be generalised to a lag $k$ correlation, with $\mathcal{N}(0, 1/(n-k))$ as its asymptotic null distribution.

### 6.13 Exam tips

In exam/tutorial questions, the data sets are necessarily small. If you use certain approximations these will generally be acceptable if you remember to indicate that, in practice, a large number of observations would be required and emphasise any approximations you are making or additional assumptions.

### 6.14 Practical points

Also, in practice, the actuary must examine his or her results critically for obvious problems. Examples which have occurred in the past are as follows:

1. In censuses, ages reported can be unreliable with age multiples of 5 more common than would be expected! This introduces bias which graduation cannot remove.

2. When graduating $a(90)$, $q_{94}$ was 30 times higher than expected. This was found to be because the information collected was on policies and not on lives and a particular individual held a large number of policies then died whilst under observation. See later discussion on duplicates.

3. The ‘accident hump’ used to be removed.

Finally, the actuary must consider whether differences exist between the observed population and the population to which graduated rates are applied (and the financial consequences of any errors).
Figure 5: Top: Crude mortality rates $q_x$ against age, $x$. Bottom: Logit of crude mortality rates $\logit(q_x/(1-q_x))$ against age, $x$. 
6.15 Example: Testing a Graduation

Suppose we graduate the data in tutorial 6, question 2. Figure 5 (top) gives a plot of the crude rates \( \hat{q}_x \) against age. The plot is rather unsatisfactory for a number of reasons:

- the functional form of \( q_x \) is rather unclear, and
- the values of \( q_x \) are very low for \( x < 60 \), say, after which they take off.

This suggests we should try a transformation of \( q_x \). A first try is the logistic transformation

\[
\text{Logit}(q_x) = \log \left( \frac{q_x}{1 - q_x} \right).
\]

Figure 5 (bottom) looks much better and a simple linear, quadratic or cubic function is indicated. We fit the quadratic function

\[
\log \left( \frac{q_x}{1 - q_x} \right) = a_0 + a_1 x + a_2 x^2
\]

where \( x \) is age. We find \( \hat{a}_0 = -6.148874, \hat{a}_1 = -0.001511027 \) and \( \hat{a}_2 = 0.0006966005 \). The data, fitted values and standardised residuals are in the table

<table>
<thead>
<tr>
<th>Age</th>
<th>( E_x )</th>
<th>( d_x )</th>
<th>( \hat{d}_x )</th>
<th>( \frac{d_x - \hat{d}_x}{\sqrt{E_x\hat{q}_x(1 - \hat{q}_x)}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>8119</td>
<td>14</td>
<td>18.78</td>
<td>-1.104</td>
</tr>
<tr>
<td>17</td>
<td>7750</td>
<td>20</td>
<td>19.68</td>
<td>0.071</td>
</tr>
<tr>
<td>22</td>
<td>6525</td>
<td>22</td>
<td>18.83</td>
<td>0.731</td>
</tr>
<tr>
<td>27</td>
<td>5998</td>
<td>23</td>
<td>20.37</td>
<td>0.584</td>
</tr>
<tr>
<td>32</td>
<td>5586</td>
<td>26</td>
<td>23.10</td>
<td>0.603</td>
</tr>
<tr>
<td>37</td>
<td>5245</td>
<td>28</td>
<td>27.35</td>
<td>0.124</td>
</tr>
<tr>
<td>42</td>
<td>4659</td>
<td>32</td>
<td>31.70</td>
<td>0.054</td>
</tr>
<tr>
<td>47</td>
<td>4222</td>
<td>37</td>
<td>38.77</td>
<td>-0.286</td>
</tr>
<tr>
<td>52</td>
<td>3660</td>
<td>44</td>
<td>46.92</td>
<td>-0.429</td>
</tr>
<tr>
<td>57</td>
<td>3012</td>
<td>54</td>
<td>55.70</td>
<td>-0.229</td>
</tr>
<tr>
<td>62</td>
<td>2500</td>
<td>68</td>
<td>68.81</td>
<td>-0.098</td>
</tr>
<tr>
<td>67</td>
<td>2113</td>
<td>87</td>
<td>89.10</td>
<td>-0.226</td>
</tr>
<tr>
<td>72</td>
<td>1469</td>
<td>100</td>
<td>97.26</td>
<td>0.287</td>
</tr>
<tr>
<td>77</td>
<td>883</td>
<td>95</td>
<td>93.36</td>
<td>0.179</td>
</tr>
<tr>
<td>82</td>
<td>418</td>
<td>70</td>
<td>70.87</td>
<td>-0.113</td>
</tr>
<tr>
<td>87</td>
<td>181</td>
<td>49</td>
<td>48.40</td>
<td>0.100</td>
</tr>
</tbody>
</table>

It is now a straightforward matter to check our graduation by running through our battery of tests.
\( \chi^2 \text{ test} \)

We compute

\[
\chi^2 = \sum \frac{(d_x - \hat{d}_x)^2}{E_x \hat{q}_x(1 - \hat{q}_x)} = 3.003.
\]

We have \( n - p = 16 - 3 = 13 \) degrees of freedom and so there is no suggestion that the graduation is unsatisfactory.

**Standardised Deviations Test**

There are very few observations for this test. With cells of equal width we calculate

<table>
<thead>
<tr>
<th>Range of ( z )</th>
<th>(-\infty, 1)</th>
<th>(-1, 0)</th>
<th>(0, 1)</th>
<th>(1, \infty)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>1</td>
<td>6</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Expected</td>
<td>2.56</td>
<td>5.44</td>
<td>5.44</td>
<td>2.56</td>
</tr>
</tbody>
</table>

which gives

\[
\chi^2 = \frac{1.56^2}{2.56} + \frac{0.54^2}{5.44} + \frac{3.54^2}{5.44} + \frac{2.56^2}{2.56} = 5.87
\]

with 3 degrees of freedom. Now \( \Pr(\chi^2 > 5.87) \approx 0.12 \) when \( \chi^2 \sim \chi_3^2 \) so there is weak evidence that the graduated data do not fit the original data. If we are worried that the expected values (of 2.56) are rather small we can combine cells to give

\[
\chi^2 = \frac{2 \times 1^2}{8} = 0.25
\]

where \( \chi^2 \sim \chi_1^2 \). Any evidence of a poor fit has disappeared.

A better approach might be to use cells of equal area. With 16 observations we could use 4 (or 3) cells. We find

<table>
<thead>
<tr>
<th>Range of ( z )</th>
<th>(-\infty, 0.6745)</th>
<th>(-0.6745, 0)</th>
<th>(0, 0.6745)</th>
<th>(0.6745, \infty)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>1</td>
<td>6</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Expected</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

and \( \chi^2 = 9.5 \). Now \( \Pr(\chi^2 > 9.348) = 0.025 \) so we have quite strong evidence that the graduation should be rejected. The equal area approach has given a more powerful test, in this case at least.

**The Sign Test**

We have 7 negative and 9 positive signs. This is clearly quite unexceptional. If \( S \) is the number of positive signs then \( S \sim B(16, \frac{1}{2}) \), and, for example, \( 2 \times \Pr(S \leq 3) = 0.021 \) and \( 2 \times \Pr(S \leq 4) = 0.077 \) so we are nowhere near a significantly low (or high) value of \( S \).

**The Change of Sign Test**

We have 5 changes of sign. If \( C \) is the number of sign changes then \( C \sim B(15, \frac{1}{2}) \), and \( \Pr(C \leq 5) = 0.15 \), so there is nothing to make us suspect non-randomness here.
The Cumulative Deviations Test
This is usually a very poor test, particularly as here, when we will test over the whole age range, and we have fitted by maximum likelihood. We find $\sum_x (D_x - E_x q_x) = -4.7 \times 10^{-7}$, i.e., almost zero, as expected.

The Grouping of Signs Test
We have $n_1 = 9$ positive signs, and $n_2 = 7$ negative signs. Let $G$ denote the number of groups of positive signs. We observe $g = 3$. Then

$$p_t = \Pr(G = t) = \frac{(n_1 - 1)(n_2 + 1)}{(n_1 + n_2)} = \frac{(8)(8)}{16} = \frac{(8)(8)}{11440}.$$  

Thus,

$$p_1 = \frac{8}{11440}, \quad p_2 = \frac{224}{11440}, \quad p_3 = \frac{1568}{11440}$$

and so $\Pr(G \leq 3) = 0.16$. Once more there is nothing untoward here.

We can also use the normal approximation with $\text{E}(G) = n_1(n_2 + 1)/(n_1 + n_2) = 9 \times 8/16 = 4.5$ and $\text{Var}(G) = (n_1n_2)^2/(n_1 + n_2)^3 = 0.9689941$. Thus, $\Pr(G \leq 3) \approx \Pr[Z \leq (3.5 - 4.5)/\sqrt{0.9689941}] = 0.15$, in agreement with the exact calculation.

The Serial Correlation Test
We compute $r_1 = 0.461$. As a random variable, $r_1 \sim \mathcal{N}(0, \frac{1}{15})$ so the observed value of $r_1$ has a $z$-score of 1.78. The test is a 1-sided test (large positive values of $r_1$ are significant) so the observed value of 1.78 does provide significant evidence against $H_0$ (the 5% point is 1.645).

One further plot we could do is that of the standardised residuals against age.

Figure 6: Standardised deviations against age, $x$
The serial correlation coefficient test is probably picking up the (apparent) tendency for the deviations at young ages to be positive and at old ages to be negative.

**Conclusion:** The graduation passes all the tests except the serial correlation test. We could fit a cubic logistic regression and see whether this gave an improved fit.

**R**

Here is a short program in R which produces the plots and does some of the above calculations.

```r
# Data entry
# Age <- seq(12,87, by = 5)
E.x <- c(8119,7750,6525,5998,5586, 5245,4659,4222,3660,3012, 2500,2113,1469,883,418, 181)
d.x <- c(14,20,22,23,26, 28,32,37,44,54, 68,87,100,95,70,49)
q.x <- d.x/E.x
cbind(Age, E.x, d.x, q.x)
# Exploratory plots
# plot(Age, q.x, ylab = "Raw q.x")
plot(Age, log(q.x/(1-q.x)), ylab = "Raw log(q.x/(1-q.x))")
# Model fitting
# options(contrasts= c("contr.treatment","contr.poly"))
D.and.A <- cbind(d.x, E.x - d.x)
Age.2 <- Age^2; Age.3 <- Age^3
M.1 <- glm(D.and.A ~ Age, family = binomial)
M.2 <- glm(D.and.A ~ Age + Age.2, family = binomial)
M.3 <- glm(D.and.A ~ Age + Age.2 + Age.3, family = binomial)
summary(M.1)
summary(M.2)
summary(M.3)
# M.2 looks a good bet
# a.0 <- coefficients(M.2)[[1]]; a.1 <- coefficients(M.2)[[2]];
a.2 <- coefficients(M.2)[[3]]
q.x.fitted <- 1/(1 + exp(-(a.0 + a.1*Age + a.2*Age.2)))
d.x.fitted <- E.x * q.x.fitted
res <- (d.x - d.x.fitted)/sqrt(E.x * q.x.fitted * (1 - q.x.fitted))
# Or you can get these as the Pearson residuals as
# res.1 <- residuals(M.2, type = "p")
# and res = res.1
```

plot(Age, res, ylab = "Standardised residuals")
#
# This is the table of data, fitted values and standardised
# deviations
#
cbind(Age, E.x, d.x, d.x.fitted, res)
#
# Chi^2 test
#
Chi.2 <- sum(res^2)
#
# Cumulative deviations test
#
CDT <- sum(d.x - d.x.fitted)
CDT.var <- sum(E.x * q.x.fitted * (1 - q.x.fitted))
#
# Lag 1 serial correlation test
#
cor(res[1:15], res[2:16])
1. The following table shows the actual deaths $d_x$, graduated mortality rates $\hat{q}_x$ and expected deaths $E_x\hat{q}_x$ on the basis of these graduated rates, from an investigation of life office assured lives mortality at policy durations 3 years and over. Some further information is also provided.

<table>
<thead>
<tr>
<th>Ages Attained</th>
<th>$d_x$</th>
<th>$\hat{q}_x$</th>
<th>$E_x\hat{q}_x$</th>
<th>$d_x - E_x\hat{q}_x$</th>
<th>$\frac{d_x - E_x\hat{q}_x}{\sqrt{E_x\hat{q}_x(1 - \hat{q}_x)}}$</th>
<th>$E_x\hat{q}_x(1 - \hat{q}_x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.5–19.5</td>
<td>63</td>
<td>0.00224</td>
<td>65</td>
<td>2</td>
<td>-0.248</td>
<td>64.9</td>
</tr>
<tr>
<td>20.5–24.5</td>
<td>668</td>
<td>0.00239</td>
<td>656</td>
<td>+12</td>
<td>+0.469</td>
<td>654.4</td>
</tr>
<tr>
<td>25.5–29.5</td>
<td>1505</td>
<td>0.00232</td>
<td>1533</td>
<td>-28</td>
<td>-0.716</td>
<td>1529.4</td>
</tr>
<tr>
<td>30.5–34.5</td>
<td>2325</td>
<td>0.00261</td>
<td>2317</td>
<td>+8</td>
<td>+0.166</td>
<td>2311.0</td>
</tr>
<tr>
<td>35.5–39.5</td>
<td>3725</td>
<td>0.00331</td>
<td>3759</td>
<td>-34</td>
<td>-0.555</td>
<td>3746.6</td>
</tr>
<tr>
<td>40.5–44.5</td>
<td>5733</td>
<td>0.00466</td>
<td>5591</td>
<td>+142</td>
<td>+1.904</td>
<td>5564.9</td>
</tr>
<tr>
<td>45.5–49.5</td>
<td>7618</td>
<td>0.00622</td>
<td>7734</td>
<td>-116</td>
<td>-1.323</td>
<td>7685.9</td>
</tr>
<tr>
<td>50.5–54.5</td>
<td>9953</td>
<td>0.00952</td>
<td>9930</td>
<td>+23</td>
<td>+0.232</td>
<td>9835.5</td>
</tr>
<tr>
<td>55.5–59.5</td>
<td>11801</td>
<td>0.01522</td>
<td>11806</td>
<td>-5</td>
<td>-0.046</td>
<td>11626.3</td>
</tr>
<tr>
<td>60.5–64.5</td>
<td>12491</td>
<td>0.02516</td>
<td>12391</td>
<td>+100</td>
<td>+0.909</td>
<td>12079.2</td>
</tr>
<tr>
<td>65.5–69.5</td>
<td>13526</td>
<td>0.04045</td>
<td>13746</td>
<td>-220</td>
<td>-1.916</td>
<td>13190.0</td>
</tr>
<tr>
<td>70.5–74.5</td>
<td>15607</td>
<td>0.06732</td>
<td>15570</td>
<td>+37</td>
<td>+0.307</td>
<td>14521.8</td>
</tr>
<tr>
<td>75.5–79.5</td>
<td>14159</td>
<td>0.10341</td>
<td>14161</td>
<td>-2</td>
<td>-0.018</td>
<td>12696.6</td>
</tr>
<tr>
<td>80.5–84.5</td>
<td>9183</td>
<td>0.15317</td>
<td>9149</td>
<td>+34</td>
<td>+0.386</td>
<td>7747.6</td>
</tr>
<tr>
<td>85.5–89.5</td>
<td>3919</td>
<td>0.21223</td>
<td>3931</td>
<td>-12</td>
<td>-0.216</td>
<td>3096.7</td>
</tr>
<tr>
<td>90.5–94.5</td>
<td>999</td>
<td>0.28085</td>
<td>1008</td>
<td>-9</td>
<td>-0.334</td>
<td>724.9</td>
</tr>
<tr>
<td>95.5–99.5</td>
<td>122</td>
<td>0.30847</td>
<td>145</td>
<td>-23</td>
<td>-2.297</td>
<td>100.3</td>
</tr>
<tr>
<td>$\geq 100.5$</td>
<td>9</td>
<td>0.28800</td>
<td>15</td>
<td>-6</td>
<td>-1.836</td>
<td>10.7</td>
</tr>
</tbody>
</table>

You may assume that the graduation process required 4 parameters to be determined.

Perform the following tests of the adherence to the data of this graduation (some values of the test statistics are given in brackets).

(a) the $\chi^2$ test (20.1, 14 df),
(b) the standardized deviations test,
(c) the sign test,
(d) the change of sign test,
(e) the cumulative deviations test,
(f) the grouping of signs test, and
(g) the lag–1 serial correlation test ($r \approx -0.168$, $z = -0.694$)

State, with reasons, whether you consider the graduated rates suitable for the calculation of premium rates and reserves for endowment assurances and whole life assurances.
2. The following data relate to a graduation which has been carried out by the graphical method.

<table>
<thead>
<tr>
<th>Ages</th>
<th>$E_x$</th>
<th>$1000 \times \hat{q}_x$</th>
<th>$E_x\hat{q}_x$</th>
<th>$d_x$</th>
<th>$d_x - E_x\hat{q}_x$</th>
<th>$\frac{d_x - E_x\hat{q}_x}{\sqrt{E_x\hat{q}_x^2P_x}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-14</td>
<td>8119</td>
<td>1.747</td>
<td>14.18</td>
<td>14</td>
<td>-0.18</td>
<td>-0.048</td>
</tr>
<tr>
<td>15-19</td>
<td>7750</td>
<td>2.112</td>
<td>16.37</td>
<td>20</td>
<td>3.63</td>
<td>+0.898</td>
</tr>
<tr>
<td>20-24</td>
<td>6525</td>
<td>3.028</td>
<td>19.75</td>
<td>22</td>
<td>2.25</td>
<td>+0.506</td>
</tr>
<tr>
<td>25-29</td>
<td>5998</td>
<td>3.698</td>
<td>22.18</td>
<td>23</td>
<td>0.82</td>
<td>+0.174</td>
</tr>
<tr>
<td>30-34</td>
<td>5586</td>
<td>4.796</td>
<td>26.79</td>
<td>26</td>
<td>-0.79</td>
<td>-0.153</td>
</tr>
<tr>
<td>35-39</td>
<td>5245</td>
<td>5.248</td>
<td>27.52</td>
<td>28</td>
<td>0.48</td>
<td>+0.091</td>
</tr>
<tr>
<td>40-44</td>
<td>4659</td>
<td>6.806</td>
<td>31.71</td>
<td>32</td>
<td>0.29</td>
<td>+0.052</td>
</tr>
<tr>
<td>45-49</td>
<td>4222</td>
<td>8.915</td>
<td>37.64</td>
<td>37</td>
<td>-0.64</td>
<td>-0.105</td>
</tr>
<tr>
<td>50-54</td>
<td>3660</td>
<td>11.679</td>
<td>42.74</td>
<td>44</td>
<td>1.26</td>
<td>+0.193</td>
</tr>
<tr>
<td>55-59</td>
<td>3012</td>
<td>19.644</td>
<td>59.17</td>
<td>54</td>
<td>-5.17</td>
<td>-0.678</td>
</tr>
<tr>
<td>60-64</td>
<td>2500</td>
<td>28.725</td>
<td>71.81</td>
<td>68</td>
<td>-3.81</td>
<td>-0.456</td>
</tr>
<tr>
<td>65-69</td>
<td>2113</td>
<td>39.955</td>
<td>84.43</td>
<td>87</td>
<td>2.57</td>
<td>+0.286</td>
</tr>
<tr>
<td>70-74</td>
<td>1469</td>
<td>69.252</td>
<td>101.73</td>
<td>100</td>
<td>-1.73</td>
<td>-0.178</td>
</tr>
<tr>
<td>75-79</td>
<td>883</td>
<td>110.803</td>
<td>97.74</td>
<td>95</td>
<td>-2.84</td>
<td>-0.304</td>
</tr>
<tr>
<td>80-84</td>
<td>418</td>
<td>180.866</td>
<td>75.60</td>
<td>70</td>
<td>-5.60</td>
<td>-0.711</td>
</tr>
<tr>
<td>85+</td>
<td>181</td>
<td>292.293</td>
<td>52.90</td>
<td>49</td>
<td>-3.90</td>
<td>-0.638</td>
</tr>
</tbody>
</table>

(a) Explain the difficulties in carrying out a $\chi^2$-test of a graphical graduation.

(b) The $\chi^2$-test may fail to detect certain defects in a graduation. For each of the following suggest one test which would detect the defect, and carry it out.

(i) Fails to detect an abnormally large number of small deviations offset by a few large deviations.

(ii) Fails to detect positive or negative bias in the residuals.

(iii) Fails to detect clumping or non-random grouping of positive or negative signs.

(c) Test the smoothness of the graduation by considering third differences $\Delta^3\hat{q}_x$. What conclusion do you come to?

On the basis of all the tests you have performed state with reasons whether or not you consider the graduation to be acceptable.

