INSTITUTE AND FACULTY OF ACTUARIES

Audit Trail

September 2015

CA2: Model Documentation, Analysis and Reporting

Paper 1 (L20 P1)

Exam simulations

Overview

The purpose of the spreadsheet is to use a set of Uniform [0,1] random numbers to simulate the progress of 200 new students through a series of exams (up to a specified maximum number of exams, N) over ten years, using half year projection periods.

For each scenario, the following are calculated:

- The number of students who are simulated to have passed X exams at the end of the period, for $X = 0$ to N.

- The proportion who are simulated to have qualified (i.e. passed N exams) by the end of the period.

The scenarios modelled are:

- Base scenario, ignoring the possibility of students leaving the profession.

- Withdrawal scenario, allowing for withdrawal rates which depend on whether the student passed or failed the exam in the previous half year time period.

- Alternative pass rate scenario, ignoring withdrawals but modelling a different pass rate.

Data

This worksheet includes the raw data provided internally, which comprise 200 rows × 20 columns = 4,000 random numbers generated from a continuous Uniform distribution on [0,1].

Data validation

In this worksheet the raw data are validated as follows:

- Count of numbers provided = 4,000.

- None of the numbers is less than 0 or greater than 1.

- The mean is as for a $U[0,1]$ distribution, i.e. 0.5 (=1/2).

- The standard deviation is as for a $U[0,1]$ distribution, i.e. 0.289 ($=1/12^{.5}$).

- The actual split of data between the ten ranges of 0.1 is consistent with that expected (i.e. 400 per range).

As the numbers are a randomly generated sample, the mean and standard deviation will not be exact. Therefore a permitted (low) tolerance level has been included (cell C11) in order to set automatic checks.

For the fifth check in the list above, the frequency of actual observations within each range has been determined using COUNTIF (first to determine the cumulative frequency up to the upper limit of each range, and then by differencing to obtain the frequency within each range).

This has been compared with the expected frequencies graphically (and a check has been included to ensure that total actual = total expected). *[The graph shows a reasonable fit.]*

A chi-squared test has also been performed, by first calculating $(A - E)^2 / E$ for each range and then summing these across all ranges. This total has been compared with the chi-squared test statistic (CHIINV) at the appropriate degrees of freedom (cell G30 = number of ranges – 1) and a given test level (cell G31, e.g. 5%). If the total is less than the test statistic, then we cannot reject the hypothesis that the random numbers come from a $U[0,1]$ distribution at the chosen confidence level. *[The test shows that we cannot reject the hypothesis at the 95% confidence level, i.e. the distribution is a good fit.]*

Assumptions

- All information provided by AIS is correct.
- Pass rates do not vary over time.
- Pass rates are independent over time periods / number of previous attempts, i.e. the probability of a student passing an exam is independent of previous success rates / whether they have sat it before.
- Every student sits an exam at each possible opportunity / no sickness absence.
- Every student has to pass all N exams; there are no exemptions granted.
- Results are known before the half year end, so a student passing the N^{th} exam at the final half year will count as qualified by the end of the projection period.
- No students withdraw in the first half year.
- Withdrawal rates are constant over the projection period.
- Withdrawal rates are not affected by other factors, e.g. employment opportunities.
- Membership lapses due to deaths can be ignored (or can be assumed to be included in the withdrawals).

Base no wds

In this worksheet the base scenario pass/fail indicators are determined for each simulated student and each future time period, ignoring withdrawals.

The base scenario pass rate (R) is input to cell H5 and the maximum number of exams (N) to cell H6.

Table 1 simulates for each student n (from n = 1 to 200, in rows) and for each half year time period ending at time t (from t = 0.5 to 10, in columns) an indicator of 1 to represent an exam pass and an indicator of 0 to represent an exam fail. This table ignores the maximum number of exams (N).

These indicators are simulated as follows:

- For n = 1 to 200 and T = 1 to 20, let $U(n,T)$ represent the random variable from worksheet "Data" in row n, column T.

- For each simulation n and each time period ending at time $t = T/2$, an IF statement performs the following:

  - If $U(n,T) \leq R$ then the formula will return the value 1.
  - Otherwise it will return the value 0.

Column W sums the indicators across the rows, i.e. gives the total number of simulated passes across the full projection period for each student.

Check: The validation in cell O7 checks that the average total number of passes (across students) is close to the pass rate (R) multiplied by the number of possible exam sittings. [This check is unlikely to give an exact answer, so a tolerance is used; this is input to cell O9.]

To the right of the first table (column Y onwards), Table 2 reproduces Table 1 through direct cell referencing, but replaces the indicator with a 0 in the time period to t if the sum of indicators (i.e. total number of exam passes) for that student across all previous time periods is N or more. This table has therefore allowed for the restriction that it is not possible to pass more than N exams.

The totals across each row are calculated in column AU.

Check: The validation in cell AA5 checks that the total number of exam passes does not exceed N for any of the students. The validation in cell AA6 checks that the average number of exam passes with the restriction (i.e. maximum of N) is less than that without the restriction.

Base with wds

In this worksheet the base scenario is adjusted to allow for withdrawals.

The assumed withdrawal rates (per half year period) allowing for whether the student has passed or failed in the previous sitting are input to cells H4 and H5 respectively.

Table 1 determines for each student n and for each half year time period ending at time t the relevant persistency rate over that single time period.

This is determined as 1 – withdrawal rate, where the withdrawal rate is set as follows:

- For t = 0.5 (first period), the withdrawal rate is 0.
- Thereafter, the withdrawal rate for the time period to t equals input cell H4 (withdrawal rate given previous pass) if the pass/fail indicator for that student in the time period to $t - 0.5$ is 1. [This indicator is referenced from Table 2 in the worksheet "Base no wds".]
- Otherwise, it equals input cell H5 (withdrawal rate given previous fail).

Check: Cell U6 checks that all persistency rates are in the range [0,1] and U7 checks that the average persistency rate across the simulations is equal to the expected average (i.e. persistency rate on pass × pass rate + persistency rate on fail × fail rate) to 4dp.

To the right of the first table (column X onwards), Table 2 determines the *cumulative* persistency rates. The cumulative persistency rate for student n up to time t is calculated as the product of the single time period persistency rates from Table 1 in respect of student n, from t = 0.5 up to and including the rate for the time period to t. This is done using Excel function PRODUCT.

Table 3 (column AU onwards) multiplies the cells of Table 2 (cumulative persistency rates) with the indicators in Table 2 of the "Base no wds" worksheet. This therefore provides the required expected exam passes, allowing for withdrawals.

The rows of Table 3 are totalled in column BP, giving the expected total number of exam passes allowing for withdrawals. These totals are rounded in column BQ to the nearest integer (i.e. whole number of exams passed) using Excel function ROUND, to facilitate the later summarising.

Alt no wds

In this worksheet the base scenario is re-run using a different pass rate, ignoring withdrawals.

This worksheet is a copy of "Base no wds" with the following changes:

- New pass rate input to cell H5.
- The maximum number of exams (N) is linked to the original input cell rather than being input again, since it is unchanged.
- The check tolerance is similarly linked back rather than input again.
- Calculations now refer to the new pass rate rather than to the original R.

Summary of results

In this worksheet the simulated frequency of number of exam passes and proportion qualified at the end of the full projection period are determined for each of the three scenarios, and charts are produced.

For each of the three scenarios in turn, the following summary results are determined using the calculations in the previous three worksheets:

- Frequency of number of exam passes, from 0 up to *N*, performed by applying COUNTIF to the total number of passes for each simulation in the previous worksheets.

- Probability of qualification (see column G) = number with *N* passes (from the frequency table described above) divided by the total number of students.

[The calculation approach is identical for each scenario, but the simulated numbers of passes are referenced from “Base no wds”, “Base with wds” and “Alt no wds” respectively for each of the three modelled scenarios.]

The frequencies as determined above are illustrated using a bar chart for each of the scenarios. For the withdrawal and alternative pass rate scenarios, the base scenario results are also shown for comparison purposes.

Checks:

- Total of frequencies = total number of student simulations (i.e. 200).

- Reconciliation of averages calculated using SUMPRODUCT on the frequency tables is the same as determined in the previous worksheets.

- In the base scenario, the frequencies show a hump at the expected number of passes (= number of time periods × pass rate *R*)…
- … but a more significant number reaching *N* since this includes all those who were effectively “truncated” by the restriction and would have passed more than *N* if more exams were available.

- In the withdrawal scenario, the number of passes and proportion qualified are lower than in the base scenario as fewer are modelled to sit each of the exams.

- The shape of the graph has also changed, due to there being a higher withdrawal rate following failure than following success.

- Setting the withdrawal rates in worksheet “Base with wds” to zero gives the same results as under the base scenario.

- In the alternative pass rate scenario, the number of passes and proportion qualified should differ relative to the base scenario according to whether the pass rate has increased or reduced. Here, the pass rate has increased so the number of passes and proportion qualified have similarly increased, as would be expected.

- The very high frequency at *N* reflects the increased impact of the “truncation”.

- The proportion qualified now exceeds 50%, which is reasonable since the expected number of passes is *N* and the simulation of passes and fails is now symmetric on a pass rate of 0.5. Therefore it is expected that more than half will have *N* or more passes.

- Setting the pass rate in worksheet “Alt no wds” to the same as in the base scenario gives the same results as under the base scenario.

END OF AUDIT TRAIL