**REPORT COMPUTATIONAL STATISTICS**

**1ST PROJECT WORK**

***Michele Cibelli***

**CODE EXPLANATION**

In relation to the task received, the code on R is divided into 3 different points, such as the requests of the task. To solve all 3 points we use Monte Carlo simulations and then we analyze the results obtained.

Let's start by entering all the probabilities of death that affect us from year x to year x + 1, between 30 and 49 years.

After doing this we insert our population, through a sequence from 30 to 49 and with a normal distribution equal to 5000 individuals for each age between 30 and 49.

From here on we start our real Monte Carlo simulation. For greater accuracy, we will see how the results change between 1000, 10000, and as many as 100000 different simulations. By simulation we mean a computation calculation that we perform on R, in such a way that, always following the previously entered probabilities, we see on 1000, 10000 and 100000 cases, how many people have died and in what years and how much we have gained or lost in these 5 years all the people who survived. The program is able each time to simulate and extrapolate different individuals, offering us a very broad overview of cases, which can give us greater confidence and accuracy with regard to our final conclusions on the goodness or otherwise of the investment in question

We then open a matrix of null results, which we will fill later, with the results subsequently obtained.

From here a for loop opens with a sequence ranging from 1 to 1000 (then 1: 10000 and then 1: 100000) to be able to operate through the simulation. Within the loop, we let the machine choose our sample of people using the "rbinom" function, of living or dead individuals. Obviously we do not make sure that all people of all ages are chosen because it would not coincide with the real situation, in which not all individuals would accept to subscribe to this type of investment. So the machine has as a setting a maximum of 20 ages, but consistently with reality choose only a total of individuals of only some ages and not all. Then we start the count with tot <-0, we insert the living ones, which at time zero coincide with the total chosen by the machine for the simulation, and we start another for loop from 0 to 4 (that is our years). Within this other for loop, the machine will assign a value between 0 and 1 for each individual in our population, in relation to whether that specific individual is dead (0) or alive (1), always taking into account the probabilities of surviving inserted at the beginning of the code. From here, if 0 is assigned, there is a break that blocks the count, if instead a 1 is assigned, you move on to counting how much has been monetized through the investment. This is possible thanks to another for loop that counts the total amount earned by each surviving individual, using the formula (year \* alive + (year \* dead/alive)), as explained in the text of the exercise.

At this point we are able to fill in the result matrix we called earlier and which was empty at the time.

After doing this, using the aggregate function of R, we see on average how much for each age has been earned according to the investment and insert the graph with the age on x and the expected return on y.

For point 1 we limit ourselves to seeing the difference in results between 1000, 10000, and 100000 simulations, taking into account that especially for 100000 simulations, the calculation time can be high and may involve a wait of a few minutes. So we are able to answer the first question, whether we would recommend this type of investment or not.

For point 2, the code remains the same, except at the beginning, in which the population for each age is not more than 5000 individuals, but changes, according to a "runif", that is a random extraction of value ranging from 1 to 5000. We will do 5 different extractions and see the results, trying to answer the question whether or not by changing the number of individuals for each age (people participating in the investment for each age), the result changes or not. We will always keep 10000 as the number of simulations, the right compromise between the number of simulations and the calculation waiting time.

Also for point 3 the code remains almost the same, but this time we will change the last part, that of the calculation of what has been earned, adding an interest rate for each year, therefore 5 different ones, which follow an exponential trend and have an average of 2%. We will see if this changes the results or not and will re-evaluate the entire investment if necessary.

**RESULTS OBTAINED**

***POINT 1***

Immagine che contiene testo

Descrizione generata automaticamente 

Immagine che contiene testo

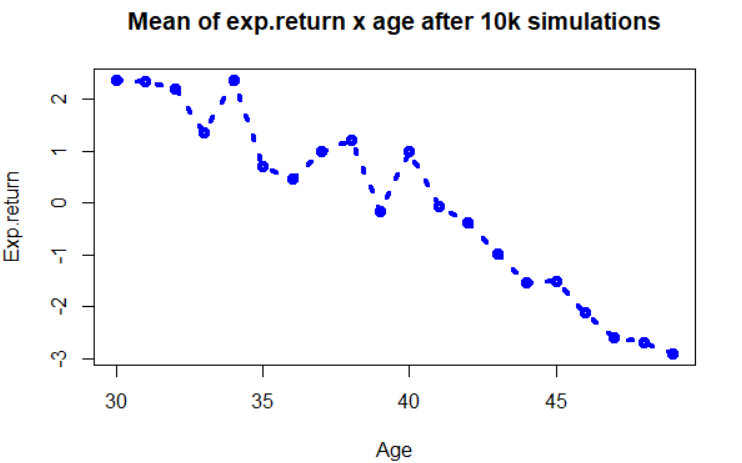
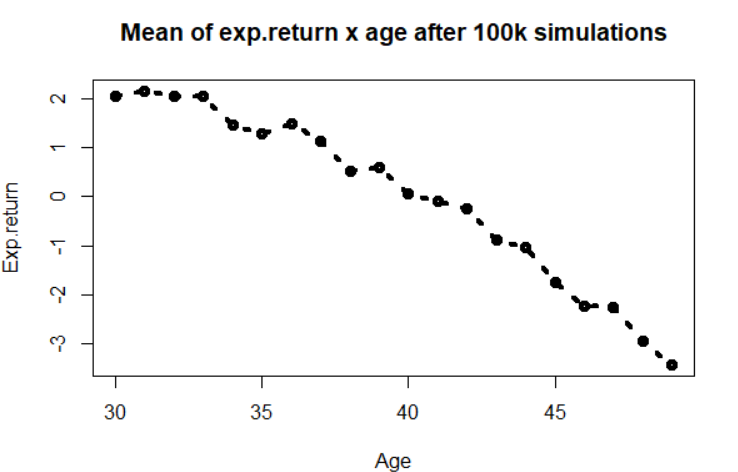
Descrizione generata automaticamente 

Immagine che contiene testo

Descrizione generata automaticamente 

As we can see, on average what we already expected occurs, namely that as we age, the probability of dying increases and therefore the value of the expected return becomes even negative (in fact, those who live longer should have more money in the end). However there are big differences depending on the number of simulations we are going to operate. As can be seen after 1000 simulations, some outliers can be identified such as those that are 37 and 39 years old, which in the subsequent 10000 and 100000 simulations, have always obtained a positive exp.return, while in this case it is negative. Other outiliers can be considered the observations with 1000 simulations aged 41 to 43, which are positive and instead have negative exp.return with the increase of the simulations. We can also observe how as the number of simulations increases, there is more and more a direct relationship between age and what is gained (the line from 1000 to 100000 simulations tends more and more to flatten and become a direct linearity relationship between the two components). In general, we cannot consider this type of investment as profitable given the very low yield within 5 years and only for some ages, not for all. We can say that we would not recommend this type of investment under these conditions. Obviously the best and most reliable and complete would be that of the 100,000 simulations, in which the values ​​and intra-class differences are reduced, but due to the great computational effort and the calculation time, we will operate with 10,000 simulations.

***POINT 2***

In this specific case we observe how the return on this investment can change in percentage terms, if the number of people of each age changes and our sample is no longer based on 5000 observations for each age, but a different number, from time to time.

To do this we use a random extraction of numbers from 1 to 5000.

What we get are the following results for each population:

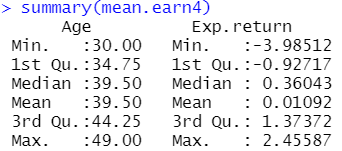
1. mean.earn4 = population made up of 2680 individuals of each age;

2. mean.earn5 = population made up of 2975 individuals of all ages;

3. mean.earn6 = population made up of 702 individuals of each age;

4. mean.earn7 = population made up of 890 individuals of each age;

5. mean.earn8 = population made up of 4446 individuals of all ages;

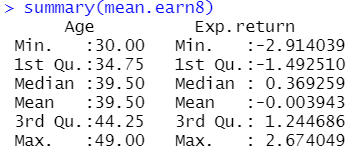
Immagine che contiene testo

Descrizione generata automaticamente

Immagine che contiene testo

Descrizione generata automaticamenteImmagine che contiene testo

Descrizione generata automaticamente



As we can see the results do not seem to change much, the oscillations between the various values ​​are very small (we go a 1% variation in the maximum loss, a 0.3% average gain and a 0.3% maximum gain), and furthermore, the trend is confirmed that it is younger people who can potentially earn more, as expected in the first point. The variations therefore do not seem to be significant and we continue to think that this type of investment is not advisable.

***POINT 3***

In this last point we try to introduce a series of interest rates (one per year), which modify the remunerative share of our investment. Our 5 interest rates follow an exponential trend and have an average of 2%. To do this in r we use the “rexp” function which generates these different rates with the characteristics listed above. They correspond to:

year 1 = 1.1%

year 2 = 2.4%

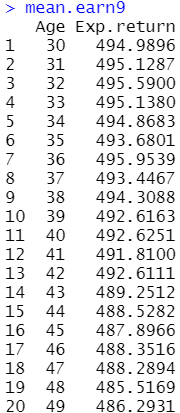
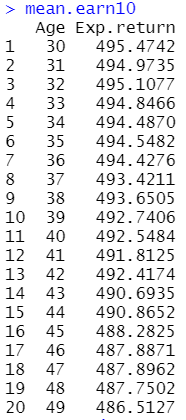
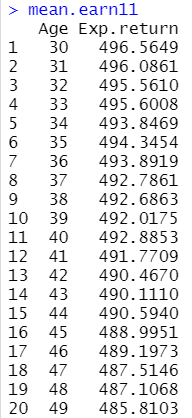
year 3 = 1.8%

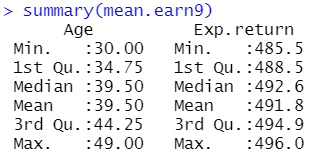
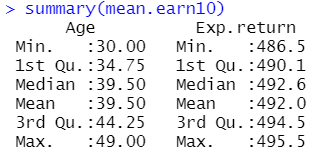
year 4 = 0.2%

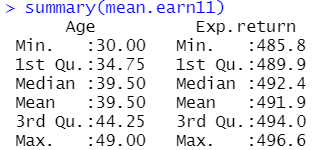
year 5 = 1.6%

We then repeat the simulation not only for 5000 individuals for each age (“mean.earn9”), but also considering 2500 and 1000 individuals for each age (respectively “mean.earn10” and “mean.earn11”).

We get the following results:



Now the situation seems very different: according to our simulations, the people who survived saw their investment almost fivefold, thanks to the inclusion of annual interest rates. In fact, the interest rate of the investment becomes about 490%, totally changing the prospects. Now we feel we can recommend this type of investment because it turns out to be profitable. However, the rule still applies that those who are younger are more likely to stay alive and therefore earn more, in fact between the ages of 30 and 49 there is a 10% drop in the rate of return on investment. The observation of point 2 also remains unchanged in which we stated that the number of people for each age was not a determining factor. In fact, the oscillations between the three different types of simulation are almost irrelevant, as we can see from the summaries above.

This is the final plot with mean.earn9 in red, mean.earn10 in orange and mean.earn11 in green: we can see how the trend of the lines does not change much and how they differ only in some points, such as for example at both 36 and 48 years old.

