Series of notes on Randomization – Part II

Techniques to generate random numbers

In selecting random samples for analysis, it is necessary to generate random numbers. Random numbers also are used for simulations and can be used to create sample datasets.

Random numbers can be generated in a number of different ways. Let's discuss a few common methods to generate random numbers.

1. Using a random number table

Some textbooks and statistical handbook do include random number tables as appendices. The best way to obtain a random sample from a population is by the use of a random number table. Figure 1 below is an example of a random number table taken from "Statistics and Chemometrics for Analytical Chemistry" (6th edition) by James N Miller and Jane C Miller, printed by Pearson Education Ltd. This table is good for not more than 100,000 random samples of a population.

Figure 1: A table of random numbers

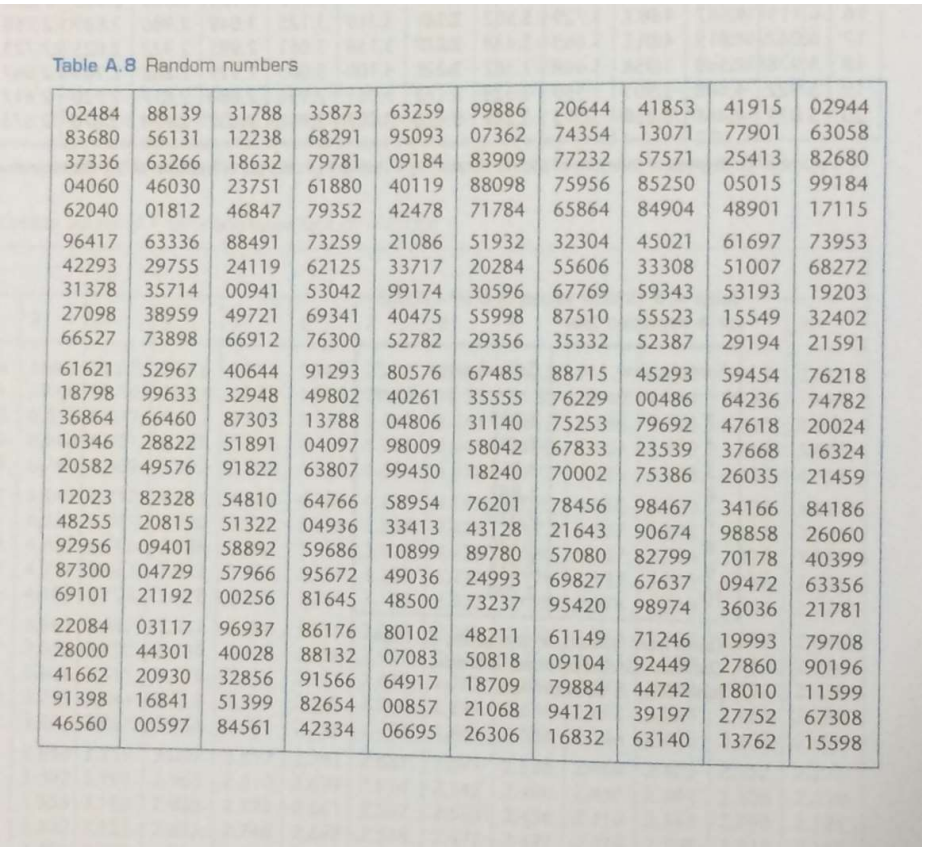

Table A.8 Random numbers

02484	88139	31788	35873	63259	99886	20644	41853	41915	02944
83680	56131	12238	68291	95093	07362	74354	13071	77901	63058
37336	63266	18632	79781	09184	83909	77232	57571	25413	82680
04060	46030	23751	61880	40119	88098	75956	85250	05015	99184
62040	01812	46847	79352	42478	71784	65864	84904	48901	17115
96417	63336	88491	73259	21086	51932	32304	45021	61697	73953
42293	29755	24119	62125	33717	20284	55606	33308	51007	68272
31378	35714	00941	53042	99174	30596	67769	59343	53193	19203
27098	38959	49721	69341	40475	55998	87510	55523	15549	32402
66527	73898	66912	76300	52782	29356	35332	52387	29194	21591
61621	52967	40644	91293	80576	67485	88715	45293	59454	76218
18798	99633	32948	49802	40261	35555	76229	00486	64236	74782
36864	66460	87303	13788	04806	31140	75253	79692	47618	20024
10346	28822	51891	04097	98009	58042	67833	23539	37668	16324
20582	49576	91822	63807	99450	18240	70002	75386	26035	21459
12023	82328	54810	64766	58954	76201	78456	98467	34166	84186
48255	20815	51322	04936	33413	43128	21643	90674	98858	26060
92956	09401	58892	59686	10899	89780	57080	82799	70178	40399
87300	04729	57966	95672	49036	24993	69827	67637	09472	63356
69101	21192	00256	81645	48500	73237	95420	98974	36036	21781
22084	03117	96937	86176	80102	48211	61149	71246	19993	79708
28000	44301	40028	88132	07083	50818	09104	92449	27860	90196
41662	20930	32856	91566	64917	18709	79884	44742	18010	11599
91398	16841	51399	82654	00857	21068	94121	39197	27752	67308
46560	00597	84561	42334	06695	26306	16832	63140	13762	15598

We may wish to generate our own random numbers in a table form by using the random RAND function of the MS Excel® spreadsheet or the R statistical programming language. We shall discuss them later.

Let's see how we can use such random number table.

For example, there is a shipment of 500 cartons of cocoa butter and we need to sample 10% of the consignment for inspection, i.e. we have to take 50 carton samples randomly from this population of 500 cartons.

We start the process by allocate a number for each carton of the shipping consignment in such a way that all the numbers have an equal number of digits, e.g. 001, 002, 010, 100, *etc.* from 001 to 500.

Random numbers are then read off from the above random number table (see Figure 1), starting at an arbitrary point to give, say, 484 from the very first number of the random number table in Figure 1: 02484. When we move down the column to the second number 83680. The last digit 680 is beyond the highest number of 500 assigned for this shipment population. Hence, we omit this number and move on to the third number under the same column, i.e. 37336, from where we pick up the second random carton sample with corresponding number 336. This process is repeated until all 50 cartons are sampled for inspection.

2. Using spreadsheet software such as MS Excel®

Most spreadsheets include random number generators. In Excel®, the relevant provided function is RAND(). These are usually adequate for small experiments, though they do not perform as well as the best random number generators found in good statistical software.

The Excel® function RAND() creates random numbers between 0 and 1. If we put "=RAND()" into a worksheet cell and press Enter key, we will see a random value which is less than 1.

If we want random numbers larger than 1, we multiply the random number function by a constant. For example, "=RAND()*100" gives random numbers between 0 and 99. It is interesting to note that the largest value is 99 rather than 100 because RAND( ) will not give a value equal to 1.0, the largest being .9999….. Therefore, multiplying by 100 gives a maximum value of 99.9999… and the integer portion is 99. However, when we use a function "=RAND()*100+1", it gives integers from 1 to 100.

The Excel® RAND function generates so-called pseudo-random numbers from the uniform probability distribution (or rectangular probability distribution) where there is equal probability for all values that a random number can take on. An obvious example of the uniform distribution is tossing a single uniformly made die. The outcome of the toss is six possible values (1,2,3,4,5,6) and each of these values is equally likely to occur.

It may be noted that when using the random number generator in Excel®, we cannot get a repeatable sequence with the RAND function, nor with VBA, nor by any other means. This is because Excel® RAND function, like another function RANK(), does not have a 'seed' or a starting point for generating a sequence of random numbers. Hence, in order to 'fix' the random numbers generated, we have to copy and paste the range of values on the other spreadsheet cells.

3. Using R program to generate random numbers

See: https://consultglp.com/wp-content/uploads/2016/12/r-techniques-in-generating-random-numbers.pdf