CONTAMINATED SITES STATISTICAL APPLICATIONS GUIDANCE DOCUMENT NO. 12-16

RANDOMIZATION

A guide for data analysts and field staff on how to select random samples and randomize sample locations

This guidance document is one of a series that outlines important basic statistical concepts and procedures that are useful in contaminated sites studies. BC Environment recommends that these suggestions be followed where applicable, but is open to other techniques provided that these alternatives are technically sound. Before a different methodology is adopted it should be discussed with BC Environment.

April 2001

THE GENERAL IDEA

Randomization is a recurring aspect of sampling and data analysis in contaminated site studies. For example, during *in situ* characterization, if we identify a particular value as an outlier and need to replace it with another nearby sample, it is recommended that this replacement sample be located at random within a circle of radius 1 m from the original sample location. Other examples of situations that involve randomization are:

- For monitoring internal heterogeneity of composite samples, one in every ten composite samples should be chosen at random to have all of its discrete samples analyzed.
- In designing a sampling grid for a contaminated site, it may be necessary to randomize the origin of the grid.

With these and other similar situations that call for randomization, it is not appropriate to make the decision haphazardly; lack of forethought does not produce good random samples. Nor is it appropriate to leave the choice of a random sample to someone's guesswork; the "random" choices that people make usually turn out not to be suitable for statistical purposes. Proper randomization is a systematic and repeatable procedure that can be checked and verified.

This guidance document addresses the issue of randomization and presents procedures for making random choices. It begins with a discussion of uniform random numbers and then describes how a table of such numbers can be used for randomization. Other guidance documents in this series that make specific references to randomization are *COMPOSITE SAMPLES, SAMPLING PLANS, STOCKPILING* and *OUTLIERS.*

UNIFORM RANDOM NUMBERS

The cornerstone of randomization is a sequence of uniform random numbers between 0 and 1. Table 1 on the next page shows 500 uniform random numbers; information on how to generate such a table, or on other similar tables, can be found in the references listed at the end of this guidance document. The values shown in Table 1 are "uniform" in the sense that a histogram of the values, such as the one shown in Figure 1, will show that the numbers in the sequence are as likely to come from one particular class in the histogram as from any other class — all values are equally probable. The values are "random" in the sense that the next value in the sequence is always unpredictable. Whether we read across the rows or down the columns of Table 1, there is no pattern or clue that tells us what the 501st value might be; regardless of the past sequence, all values between 0 and 1 remain equally probable. One way of demonstrating the lack of predictability in the values is to plot consecutive pairs on a scatterplot, such as the one shown in Figure 2. In this figure, the first random number in each of the 250 possible pairs from Table 1 is plotted as the x value and the second in each pair is plotted as the y value. If the sequence is truly random then such a plot should show no correlation. Kennedy and Gentle (1980) and Bratley, Fox and Schrage (1980) discuss several other statistical criteria for sequences of random numbers; the two most important, however, are uniformity of the histogram (as shown in Figure 1) and lack of correlation on a scatterplot of consecutive pairs (as shown in Figure 2).

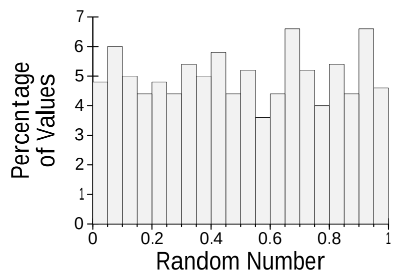

Figure 1 Histogram of the 500 values shown in Table 1.

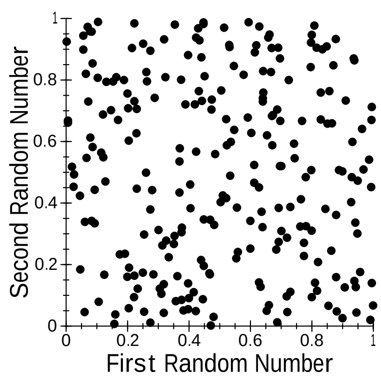

Figure 2 Scatterplot of the 250 consecutive pairs from Table 1.

USING UNIFORM RANDOM NUMBERS

A sequence of uniform random numbers, such as the one shown in Table 1, can be used as the basis for randomization procedures that are both systematic and verifiable. Two common randomization problems are discussed below; the first is the random selection of one of several samples, the second is the selection of a random location.

Selecting a random sample

The following procedure can be used to select one sample at random from a group of N samples:

1. Assign numbers 1 through N to the samples; which sample gets which number is unimportant, but each sample should have a unique index from 1 to N. The assignment of these unique indexes should be recorded for future reference in case another random sample is needed or in case the randomization needs to be checked.
2. Generate U, a uniform random number between 0 and 1. This can be done by taking the available next number from a table of random numbers, such as the one given in

Table 1 500 uniform random numbers between 0 and 1.

.534	.489	.800	.947	.641	.829	.719	.287	.432	.764
.222	.164	.640	.730	.600	.252	.936	.870	.086	.582
.305	.122	.188	.800	.744	.546	.405	.383	.820	.208
.701	.309	.556	.385	.466	.172	.503	.403	.377	.320
.829	.672	.230	.706	.397	.139	.559	.241	.200	.756
.233	.122	.310	.105	.434	.929	.653	.050	.531	.914
.404	.460	.612	.524	.870	.848	.086	.854	.480	.030
.093	.443	.878	.933	.104	.989	.128	.470	.993	.452
.485	.559	.471	.002	.124	.167	.263	.796	.469	.346
.274	.011	.636	.372	.848	.909	.439	.215	.121	.549
.146	.702	.718	.097	.357	.081	.660	.068	.995	.712
.274	.379	.064	.820	.354	.980	.800	.094	.888	.507
.536	.599	.423	.938	.687	.012	.261	.826	.730	.387
.696	.870	.654	.620	.214	.904	.274	.895	.999	.067
.673	.686	.578	.817	.397	.881	.670	.683	.334	.224
.521	.416	.175	.233	.474	.736	.780	.484	.879	.361
.318	.043	.628	.142	.374	.801	.521	.673	.692	.274
.810	.141	.799	.310	.370	.578	.163	.809	.160	.038
.774	.271	.362	.162	.546	.846	.473	.704	.692	.384
.938	.147	.451	.812	.547	.638	.447	.982	.070	.973
.854	.066	.591	.678	.701	.520	.510	.425	.230	.447
.388	.721	.986	.541	.114	.564	.060	.046	.369	.535
.945	.044	.002	.925	.671	.588	.072	.730	.205	.190
.083	.957	.280	.442	.153	.795	.990	.020	.720	.046
.941	.336	.762	.324	.254	.298	.045	.424	.131	.794
.318	.136	.949	.473	.103	.807	.106	.079	.930	.484
.900	.026	.684	.249	.006	.668	.369	.434	.066	.547
.928	.403	.169	.670	.447	.987	.448	.347	.662	.948
.415	.110	.796	.842	.376	.305	.422	.049	.725	.800
.026	.493	.301	.312	.229	.627	.284	.168	.351	.267
.079	.613	.006	.661	.482	.329	.899	.503	.222	.731
.603	.628	.857	.764	.202	.708	.808	.977	.382	.051
.532	.907	.056	.899	.690	.905	.191	.235	.157	.007
.642	.759	.448	.196	.419	.721	.083	.342	.251	.918
.742	.593	.056	.944	.619	.913	.554	.220	.612	.466
.994	.670	.667	.826	.907	.126	.815	.905	.250	.174
.120	.688	.696	.520	.326	.278	.222	.984	.774	.228
.844	.381	.078	.958	.376	.085	.199	.160	.851	.658
.658	.938	.514	.970	.204	.058	.639	.322	.879	.159
.019	.518	.204	.603	.260	.499	.829	.759	.288	.742
.024	.453	.397	.055	.061	.339	.697	.667	.633	.128
.963	.641	.523	.588	.781	.325	.442	.732	.323	.809
.817	.115	.764	.412	.863	.245	.440	.874	.505	.766
.910	.733	.353	.293	.968	.509	.948	.301	.614	.890
.400	.091	.669	.904	.430	.968	.628	.451	.957	.176
.254	.047	.881	.048	.093	.334	.221	.094	.629	.974
.468	.168	.594	.988	.798	.507	.600	.342	.768	.667
.319	.932	.730	.111	.995	.140	.932	.599	.445	.087
.423	.567	.313	.262	.865	.659	.046	.184	.938	.864
.943	.127	.640	.741	.834	.900	.687	.704	.797	.922

Table 1. As the tabulated random numbers are used, they should be crossed off so that it is immediately obvious which one to use when we next consult the table.

3. Turn U into an integer from 1 to N by multiplying it by N, adding 1 and dropping any digits after the decimal point:

$$\text{Random integer} = \text{Integer part of } [U \times N + 1]$$

4. Select as the random sample the one that was assigned this calculated random integer as its unique index.

As an example of the use of the procedure, let us go through the exercise of deciding which of the ten composite samples shown in Table 2 should be selected for separate analysis of all of its discrete samples. The first column in this table gives the sample number for the composite samples, the next four columns identify the discrete samples that make up each of the composite samples and the last column gives the unique index from 1 to 10 that we have assigned according to step 1 of the procedure given above. Table 3 shows our table of random numbers; in this example we are supposing that we have been using it for similar randomization exercises and have been crossing off the numbers as we use them; the next number on our list, 0.466, is the value of U. We take this number and turn it into a random index from 1 to 10:

$$\begin{aligned}\text{Random index} &= \text{Integer part of } [0.466 \times 10 + 1] \\ &= \text{Integer part of } 5.66 \\ &= 5\end{aligned}$$

Our random sample from the group of ten would therefore be C130 since this is the one that was designated as number 5 when we assigned unique indexes from 1 to N.

Table 2 Ten composite samples and their discretes.

Composite Sample No.	Discrete Sample Numbers				Unique Index
C103	D562	D563	D564	D565	1
C104	D567	D568	D569	D570	2
C105	D572	D573	D574	D575	3
C129	D709	D710	D711	D712	4
C130	D714	D715	D716	D717	5
C131	D719	D720	D721	D722	6
C149	D831	D832	D833	D834	7
C150	D836	D837	D838	D839	8
C151	D841	D842	D843	D844	9
C172	D902	D903	D904	D905	10

Table 3 Uniform random numbers between 0 and 1.

.534	.489	.800	.947	.641	.829	.719	.287	.432	.764
.222	.164	.640	.730	.600	.252	.936	.870	.086	.582
.305	.122	.188	.800	.744	.546	.405	.383	.820	.208
.701	.309	.556	.385	.466	.172	.503	.403	.377	...

Selecting a random location

The following procedure can be used to select a random location within a rectangular area whose width is X_{width} and whose height is Y_{height} (see Figure 3):

1. Generate U_1 and U_2, two uniform random numbers between 0 and 1. This can be done by taking the next pair of numbers from a table of random numbers, such as the one given in Table 1. As the tabulated random numbers are used, they should be crossed off so that it is immediately obvious which one to use when we next consult the table.

2. Turn U_1 into an x-coordinate from 0 to X_{width} by multiplying it by X_{width} and turn U_2 into a y-coordinate from 0 to Y_{height} by multiplying it by Y_{height}:

$$X = U_1 \times X_{width} \qquad Y = U_2 \times Y_{height}$$

3. Using the corner of the rectangular area as the origin (0,0) select as the random location the point whose coordinates are (X,Y).

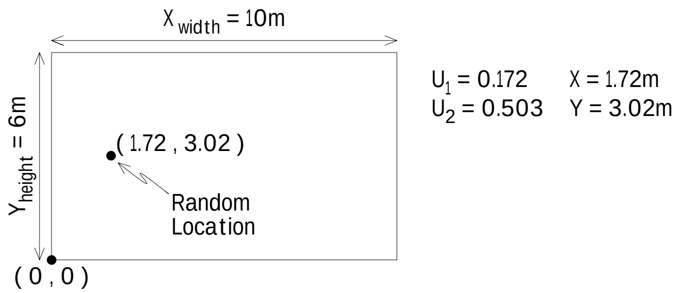

Figure 3 Random sampling from a rectangular area.

If the area within which we want a random sample is circular, rather than rectangular, a similar procedure can be used, with the first uniform random number being converted to a random azimuth from 0° to 360° and the second uniform random number being converted to a radius from 0 to R, the radius of the circular area (see Figure 4):

$$\text{Azimuth} = U_1 \times 360^\circ \qquad \text{Radius} = U_2 \times R$$

As an example of the use of this procedure, let us go through the exercise described at the beginning of this document. At one of our existing sample locations there is an outlier value that we believe to be erroneous and we need to collect a replacement sample from a random location within 1 m of the existing sample. Table 4 shows the random numbers left after we have crossed off the one we used in the first example. The next two values are 0.172 and 0.503. Multiplying the first one by 360 gives us a random azimuth of 62° (N62°E). Since the radius of our circular area is 1 m, the second random number can serve directly as our radius. The replacement sample would therefore be taken at a distance of 0.503 metres from the location of the outlier sample in a direction of N62°E.

Table 4 Uniform random numbers between 0 and 1.

.534	.489	.800	.947	.641	.829	.719	.287	.432	.764
.222	.164	.640	.730	.600	.252	.936	.870	.086	.582
.305	.122	.188	.800	.744	.546	.405	.383	.820	.208
.701	.309	.556	.385	.466	.172	.503	.403	.377	...

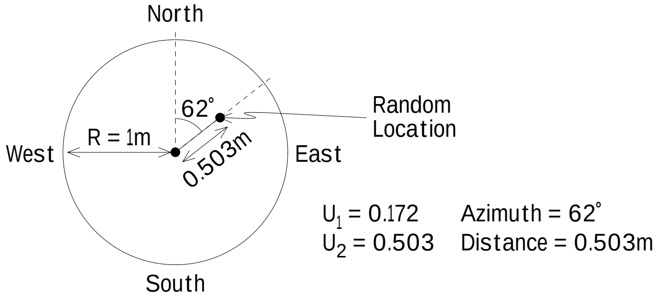

Figure 4 Random sampling from a circular area.

RECOMMENDED PRACTICE

1. When selecting a single random sample from a larger group or when selecting a random location within a specified area, use a systematic and verifiable procedure that is based on uniform random numbers.

2. Use a published table of random numbers or, if a computer or calculator is being used to create the random numbers, print an actual table of the random numbers it produces so that the procedure can be checked and verified.

REFERENCES AND FURTHER READING

The following references provide useful additional information on the specific problem of generating uniform random numbers and on the general problem of randomization.

Abramowitz, M. and Stegun, I.A., (eds.), *Handbook of Mathematical Functions*, Dover, New York, 1970.

Bratley, P., Fox, B.L. and Schrage, L.E., *A Guide to Simulation*, Springer-Verlag, New York, 1983.

Kennedy, W.J. and Gentle, J.E., *Statistical Computing*, Marcel Dekker, New York, 1980.