Random

Thomas R. Knapp

"The generation of random numbers is too important to be left to chance."
(Robert R. Coveyou)

Preface

What is the meaning of the word "random"? What is the difference between random sampling and random assignment? If and when a researcher finds that some data are missing in a particular study, under what circumstances can such data be regarded as "missing at random"? These are just a few of the many questions that are addressed in this monograph. I have divided it into 20 sections--one section for each of 20 questions--a feeble attempt at humor by appealing to an analogy with that once-popular game. But it is my sincere hope that when you get to the end of the monograph you will have a better understanding of this crucial term than you had at the beginning.

To give you some idea of the importance of the term, the widely-used search engine Google returns a list of approximately 1.4 billion web pages when given the prompt "random". Many of them are duplicates or near-duplicates, and some of them have nothing to do with the meaning of the term as treated in this monograph (for example, the web pages that are concerned with the rock group Random), but many of those pages do contain some very helpful information about the use of "random" in the scientific sense with which I am concerned.

I suggest that you pay particular attention to the connection between randomness and probability (see Section 2), especially the matter of which of those is defined in terms of the other. The literature is quite confusing in that respect.

There are very few symbols and no formulas, but there are LOTS of important concepts. A basic knowledge of statistics, measurement, and research design should be sufficient to follow the narrative (and even to catch me when I say something stupid).

Thanks for stopping by, and enjoy!

Table of Contents

Section 1: What is the meaning of the word "random"

According to Random House Webster's Dictionary [forgive me, but I just had to cite that source!], "random" is an adjective that means "occurring or done without definite aim, reason, or pattern". But that is a layperson's definition. A currently popular scientific definition of a random phenomenon, as given by Starnes, Yates, and Moore (2012) ---the authors of the textbook used in many Advanced Placement statistics courses in high schools--is one for which individual outcomes cannot be specified but there is a mathematical distribution of outcomes when the number of repetitions is very large. However, Liu and Thompson (2002) claimed that interpretations of such a definition are often circular, and in his article "What is random?" Kac (1983) argued that most scientists never even bother to define "random"; they just take for granted that everyone knows what it means. He went on to say that the notion of something being "random" is actually very complicated.

[There is an article by May (1997) that is also entitled "What is random?", in which he refers to Kac's article. And there is a book entitled What is random?, by Beltrami (1999), which bears the subtitle Chance and order in mathematics and life. He (Beltrami) said that Kac's article "prompted the title of this book." (p. 146)]

My personal preference is that something is random if it is the result of a process in which chance is permitted to operate and plays a key role in the outcome.

Associated with the adjective "random" is the rather awkward noun "randomness". (At the amazon.com website a book-search of the word "randomness" results in 534 hits.) Wallis and Roberts (1962) included a chapter on randomness in their statistics book; it was defined as a property of a probabilistic process. In her book that bears the one-word title, Randomness, Bennett (1998) provided several definitions of randomness, but all were concerned with the notions of uncertainty and unpredictability. She also pointed out that even the experts have different views of it. In his book, The jungles of randomness, Peterson (1998) provided "a set of mathematical X rays that disclose the astonishing scope of randomness" (Preface, p. xii). The late 1990s seems to have been a productive time for books about randomness!

[TheThesaurasize website lists 71 synonyms for the word "randomness", but precious few of them have anything to do with chance.]

Perhaps equally as important as what the word "random" DOES mean is what the word DOES NOT mean (in science):

1. It does not mean "haphazard". Although it sounds like a contradiction in terms, something that is random is subject to the "laws" of probability.

2. It does not mean "inconsequential". Although random outcomes often balance out, they can seriously attenuate certain matters such as the relationship between two variables--see, for example, Muchinsky (1996).

3. It does not mean "hopeless". There are many tried and true methods for understanding random phenomena and coping successfully with them.

4. Interestingly, and most importantly, it does not (necessarily) mean "representative". A small random sample, for example, may not reflect very well the population from which it is drawn. As a matter of fact, it is not possible for a sample to be perfectly representative unless it consists of the entire population itself.

In the previous paragraphs I have used expressions such as "random phenomenon", "random sample", and "random outcomes". Those matters will be treated in greater detail in the remainder of this book, along with such terms as "random assignment", "random error", and many others (see Table of Contents). There are a few other statistical terms with the word “random” in them that are NOT treated in this book, but the online Pocket Dictionary of Statistics is very good for defining most if not all of them. (The first 34 entries under the letter "R" are all concerned with randomness.) There is also the term, “random access memory (RAM)”, which is an important concept associated with computers (I’m not very good at computers).

For some interesting answers to the question “How can there be such a concept as "random"? Surely everything has a structure if you look deeply enough. What does random actually mean?”, see the webpage http://www.fortunecity.com/emachines/e11/86/random.html that is associated with Ian Stewart’s (1989) book, Does God play dice?.

A good source for discussions of randomness of various kinds that is very funny, yet at the same time very instructive, is the book by Larry Gonick and Woollcott Smith entitled The cartoon guide to statistics (1993). I will refer to that book several times throughout this monograph.

Section 2: Which comes first, randomness or probability?

On page 161 of her book, Bennett (1998) discussed the attempts by vonMises (1957 and elsewhere) to define probability based upon the concept of randomness. In his theory of randomness, Kendall (1941) defines a “suite” (his term for an infinite sequence of a finite number of different symbols) as random if it satisfies certain probabilistic requirements. The first of these sources seems to imply that probability depends upon randomness; the second source seems to imply that randomness depends upon probability. What in the world is going on?

What is going on is a confusion as to which comes first--randomness or probability. Different authorities should be free to define randomness and probability in any way they choose (recall from Section 1 Kac's claim that most people don't define "random" at all), but it seems to me that you must take a stand one way or the other regarding which is the more basic concept and proceed from there. In the remainder of this chapter I would like to summarize a few of the various positions that have been taken concerning randomness and probability, tell you what mine is, and then ask you to decide for yourself who's "right" and who's "wrong".

Let's start with Bennett (1998). She claimed (page 9): "Probability is based on the concept of a random event...". I don't think that is always the case, if "a random event" is the tossing of a "fair" coin, the rolling of a perfectly balanced die, the drawing of a card from a well-shuffled deck, and the like. Probability applies to situations like those, to be sure--the probability of "heads" is 1/2; the probability of a "four" is one-sixth; the probability of an ace is 1/13; etc. But it also is applicable to other situations such as whether it will rain tomorrow or whether I'll win my tennis match, which may have nothing at all to do with random events. It would seem, therefore, that probability should either not be defined in terms of randomness or if it is it should be clear that only certain kinds of probabilities are so defined.

In David Moore's very popular statistics book, <u>Statistics: Concepts and controversies</u> (1979), Part III is entitled Drawing Conclusions From Data and consists of two chapters, one of which is “Probability: The Study of Randomness” (sound familiar?) and the other is “Formal Statistical Reasoning”. In the fifth edition (2001) Part III is entitled Chance and consists of four chapters: “Thinking About Chance”; “Probability Models”; “Simulation”; and “The House Edge: Expected Values”.

I agree with Kendall (1941) that randomness should be defined in terms of probability. I further attest that there are different kinds of randomness and each kind of randomness should be defined in terms of some kind of probability.

There are essentially three competing definitions of probability. The first definition, sometimes called the "a priori" or "deductive" definition, applies to symmetrical situations, and goes something like this:

The probability of a result A is equal to the ratio of the number of ways A can take place to the total number of equally likely results.

For example, we say that the probability of "heads" in a single toss of a "fair" coin is equal to 1/2, because there is one way that heads can take place, out of two equally likely results ("heads" and "tails").

There are two problems with that definition. The first problem has already been referred to: it only works for symmetrical situations. The second problem is that it is circular; probability is defined in terms of "equally likely" results, which is itself a probabilistic concept. It has at least one compensating feature, however. You don't have to actually toss the coin in order to talk about its probability of landing heads or tails!

The second definition, sometimes called the "relative frequency" or "empirical" definition, applies to any repeatable situation, and is:

The probability of a result A is equal to the limiting value of the ratio of the number of times A took place to the total number of results.

For example, in order to talk about the probability that a tossed thumbtack will land on its head, i.e., with its point up (note that the a priori definition wouldn't work here, since landing with its point up and landing on its side are not expected to be equally likely), you would toss the thumbtack a large number of times, count how many times it landed point up, count the total number of tosses, and divide the former by the latter. This definition also works for coins, whether fair or unfair; but there is a crucial change in the tense of the verb--from "can take place" to "took place".

There are also two problems with this definition, however: (1) What do we mean by "limiting value"? (How many times is a "large" number of times?); and (2) There is a rather strange time element entailed [bad pun]; you have to actually do something lots of times in order to get a fix on the probability of a particular result, which "from then on" (future tense?) gets associated with it.

The third definition is sometimes called the "subjective" or "personal" definition, and is the most controversial definition of the three (it is an integral part of the Bayesian approach to probability and statistics):

The probability of a result A is a number between 0 (impossibility) and 1 (certainty) that reflects a person's strength of conviction that A will take place.

For example, if you were interested in the probability that you would choose to order steak at a restaurant you had never been to before, you could assign to that eventuality a small number such as .01 if you don't like steak and/or can't afford it; or you could assign a large number such as .99, if you love steak and have lots of money.

The most serious objection to this third definition is that it is too subjective, with the probability of certain results having the potential to vary widely from one person to another. Some people argue vociferously that science in general, and mathematics in particular, should be objective. Proponents of the definition reply that complete objectivity is not possible in science or in anything else, and some point out that the selections of the .05 significance level or the 95% confidence interval in traditional statistical inference, for example, are decidedly subjective, albeit often reasonable, choices. Note also that the verb tense for this definition is "will take place".

[Interesting aside: Starnes, Yates, and Moore, the authors of The practice of statistics, all have first and middle initials D.S. What is the probability of that?!]

What about randomness? None of the above definitions of probability explicitly refer to random events or random phenomena. Where do they fit in? To answer that question, let me now turn to a couple of non-coin examples from typical everyday research.

Example #1 (classical significance testing): Do boys do better in mathematics than girls do? In order to get some evidence regarding this question you need to give the same math test to a group of boys and a group of girls. You decide to take a random sample of 50 boys and a random sample of 50 girls from a very large public school district (say, New York or Los Angeles). When you go to sample the first person do you first think about the probability of a particular person being drawn and then worry about whether the draw is random; or do you first think about whether the draw is random and then worry about the probability of a particular person being drawn? I think it's the former; don't you? You want to define the probability that a particular person will be drawn and then you want to use some sort of randomizing device that will operationalize that probability.

Example #2 (survey research): What percentage of nurses smoke cigarettes? You decide to take a random sample of 1000 nurses from the population of all nurses who are members of the American Nurses Association (about two million people). You get access to a list of the R.N. license numbers of all two million of the nurses (the so-called "sampling frame"). You want each nurse in the population to have an equal chance of being drawn into your sample, so you might put each one of those numbers on a piece of paper, put the pieces of paper in a huge barrel, mix them up, and draw out 1000 of them. Although the purpose of the mixing is to accomplish randomness of selection, it is the equal likelihood that takes precedence.

For more on various meanings of probability and their applications I recommend that you read my favorite little probability book, Probabilities and life, by Emile Borel (1962); Shafer's (1993) chapter in the Keren and Lewis handbook of methodological issues; and Salsburg's (2001) fascinating book, The lady tasting tea. I couldn't find one mention of the word "random" or any of its derivatives in Borel's book (but he does talk a lot about chance...see the following section). Shafer has a brief, but very nice, section on randomness (he talks about it with respect to numbers and an observer of those numbers). Salsburg doesn't say very much about randomness, but he talks a lot about probability and statistics. See if you can figure out whether they would regard probability as dependent upon randomness or randomness as dependent upon probability (or neither?).

For two interesting philosophical discussions of randomness, see Keene (1957) and Spencer Brown and Keene (1957). And for several cartoons that explain probability about as well as I've ever seen it (to paraphrase a popular saying, "A cartoon is worth a thousand words"), see Chapter 3 in Gonick and Smith (1993).

Section 3: Are randomness and chance the same thing?

As far as I'm concerned, the answer is "yes". As far as Google is concerned, the answer is apparently "no". It returns about 1.5 billion web pages for "chance" but only a little over 26 million pages for "randomness" (about 1.4 billion for "random", as already cited in the Preface to this monograph).

Here are some examples. Consider first the tossing of a fair coin. The process of tossing such a coin usually possesses the property of randomness, because the result of any particular coin toss (heads or tails) usually cannot be pre-determined; it is a matter of chance, or “luck”-- Rescher's (2001) term for "randomness"--as to whether it will land face up (heads) or face down (tails) on any given toss. [I say "usually" because there is at least one person, statistician Persi Diaconis, who CAN pre-determine the results of the coins he tosses, by controlling the manner in which the toss is made--see McNamara (2003); Diaconis is an amateur magician. And see Ford (1983) regarding the randomness of a coin toss.]

The rolling of a fair die is a second example. It also possesses the characteristic of randomness for the same reason: The result of any particular roll (1, 2, 3, 4, 5, or 6) cannot be pre-determined. [As far as I know, Diaconis does not claim to be able to pre-determine the result of a die roll. But see Watson & Moritz (2003) regarding children’s judgments of the fairness of dice.]

A third example is the drawing of a card from a well-shuffled deck of cards. The card might be a black card or a red card; a spade, a heart, a diamond, or a club; an ace, 2, 3, 4, 5, 6, 7, 8, 9, 10, jack, queen, or king. Chance and chance alone (by virtue of the shuffling) determines what the actual result will be. [Card drawing by Diaconis is another exception here; he is VERY GOOD at pre-determining what card is drawn--see Mackenzie (2002) regarding Diaconis and see Gardner (1975a) regarding card shuffling in general.]

At the beginning of this section I said that chance and randomness are the same thing. But in Section 1 I claimed that probability and randomness are not the same thing. Aren't chance and probability the same thing, and by the transitive law shouldn't probability and randomness also be the same thing? My answer is "no" (chance and probability are not the same thing; Freedman, Pisani, & Purves, 1998, however, equate the two), but I'd better explain. I think of the results of events as "having" certain probabilities; I don't think of the results of events as "having" chance. For example, if the result in question is "fair coin lands heads", that result has an associated probability of .5; it doesn't have an associated chance of .5. "By chance" it may land heads, but "by chance" it may not. This may seem to you like a semantic distinction without a difference, but such a distinction between probability and chance is an important one (at least to me).

I recently came across a sentence that read "...[a particular statistic] has only a 5% chance of occurring by chance alone." I know what was meant by that: If the value for the population parameter that was stipulated in the hypothesis being tested was the true value, the probability was less than .05 that the given statistic would be as discrepant or more discrepant from the hypothesized value. But the "...chance...chance" wording blew my mind. The first usage (the 5% chance) is concerned with probability; the second usage (by chance alone) is concerned with randomness. I suggest that all of us avoid using expressions such as "there is a 5% chance" and concentrate on "there is a 5% probability" instead.

[It doesn't really fit here, but have you noticed that people often confuse "probability" and "odds"? When talking, for example, about the drawing of a spade in a single draw from a well-shuffled deck they say "the odds are 1 in 4". No, the probability is 1 in 4; the odds are 1 to 3 "in favor" and 3 to 1 "against".]

A fascinating example of the notion of randomness was provided by Peterson (1999), who discussed Divakar Viswanath's work regarding Fibonacci Numbers: a sequence of positive integers for which each integer is equal to the sum of the previous two integers (1, 1, 2, 3, 5, 8, 13, ...). He (Viswanath) wondered what would happen if you introduced an element of randomness in such a sequence (e.g., flipping a coin to decide whether to add or to subtract the previous two numbers). In so doing he discovered a new mathematical constant (1.13198824...). A constant resulting from randomness?! Neat, huh? (See Bennett, 1998, pp. 144-148, for a good discussion of Fibonacci Numbers.)

Beltrami (1999) discussed the equally fascinating example of "Janus-Faced" sequences of 0s and 1s (due to Bartlett, 1990) that are said to be random in one direction and deterministic in the other!

For an especially readable discussion of the role of chance in scientific arguments, see Chapter 2 of Abelson's (1995) book. For an equally readable discussion of randomness and chance, see Levinson (1963), esp. pp. 184-186. Chance is also a favorite topic for courses, newsletters, and serious academic journals. See, for example, information regarding the quantitative literacy course developed at Dartmouth College (http://www.dartmouth.edu/~chance/), their Chance News, and the highly-regarded journal Chance. It's all wonderful stuff!

Section 4: Is randomness a characteristic of a process or a product?

In Section 1 I made reference to randomness as a characteristic of a process. That is actually a rather controversial matter, as Bennett (1998) explained in her chapter on "Randomness as Uncertainty" (see esp. pp. 165-172). Those (like me) who claim that something is random if it has been determined by a chance process appeal to some property of the object used in the process (e.g., the balance of a coin and a die) and/or the mixing mechanism for the process (e.g., the shuffling of a deck of cards). Others (like most missing-data experts and most measurement theorists--see Sections 11, 17, and 18) treat randomness as a characteristic of a product, and claim that it is the product that must be judged to be random or non-random. In the Preface to his book, Beltrami (1999, p. xiii) stated that there has been a general shift in the last several decades from randomness-as-process to randomness-as-product. [That may be true (alas) but I have at least two other people on my side. In the first of their two-volume handbook on data analysis, Keren & Lewis (1993) say: “Randomness is a property of the generating process rather than the outcome.” (p. 310)]

The consideration that separates the "product" advocates from the "process" advocates is whether or not you need any data in order to claim randomness. The product folks insist that data are necessary (but perhaps not sufficient), whereas the process folks insist that you can argue for or against randomness data-free, by appealing to one or more features of an alleged randomizing device.

A "middle-ground" approach is to argue that randomness pertains to the basic <u>product</u> of the alleged randomness-generating <u>process</u>. That is, it is those processes that must be judged to be random or non-random by virtue of the basic products that they generate. Once a process has been deemed to be random, any further data gathered as a result of the use of such a process need not be judged to be random or non-random.

Consider the following example: You would like to estimate the standard deviation of the heights of adult males. You have identified a target population of 1000 adult males, with associated ID numbers of 000 to 999, and you plan to draw a sample of size 50. You decide to use Minitab's random sampling routine (the process). You give it the proper commands and ask it to print the 50 sampled ID numbers. You indicate whether each of those numbers is less than or equal to 499 (and call those 0) or greater than 499 (and call those 1). You subject that string of 1s and 0s (the product) to one or more "tests of randomness" (see Section 7). Let's say that the string passes that (those) test(s), i.e., it is declared "random" and your sample is likewise declared "random". You then measure the heights of the 50 adult males, record them, calculate their standard deviation, and make whatever inference to the population of 1000 adult males is warranted. The heights of the 50 men in the sample need not be subject to any tests of randomness (Siegel & Castellan's [1988] argument to the contrary notwithstanding)--neither the randomness of their actual magnitudes nor the randomness of the order in which they were drawn--because the process has already been judged to be random. You may wind up with a lousy estimate of the standard deviation of the heights of all 1000 adult males in the sampled population, but that is a separate matter! (Do you follow that? Does it make sense?)

Section 5: What is a random-number generator?

A random-number generator is a device (usually a computer program of some sort) that creates sequences of single digits that are alleged to be "random" (in its product sense) and are useful in a variety of applications ranging from drawing a single sample from a population to establishing military codes that are very difficult to break. Some authors, e.g., Whitney (1984), insist that they be called "pseudo-random-number generators" because, they argue, if a sequence of digits can be generated it must have a deterministic, non-random force behind it.

Probability and randomness have often been defined in such a way that the result is circular (i.e., probability defined in terms of randomness and randomness defined in terms of probability) or is of infinite regress (A is defined in terms of B, which is defined in terms of C,...). The situation for random (and/or pseudo-random) number generators is analogous. How do we know that the numbers produced by a random-number generator are random? Can we test them for randomness? Perhaps (there are several such tests--see Section 7); but do we have to carry out a test of randomness for every sequence of numbers that is generated by an alleged random-number generator? (In my opinion, no, as discussed in the previous section.) Can we ever be sure that the numbers we're particularly interested in are random? (Alas, also no, in my opinion. We can never be "sure" of anything that has a stochastic, i.e., non-deterministic, component.)

It is important to understand that there is no such thing as "a" random number (Craw, 2003), even though there is currently a website that will give you "the random number of the day". Any number (1, 23, 617, whatever) can be one of the members of a set of numbers that are claimed to be random, with the randomness designation associated with the set and not with a particular member of that set. [Some numbers that are of special interest to mathematicians, for example “Omega” (see Gardner, 1979) have been given the designation of a “random” number or a “normal” number, but that designation has a particular and idiosyncratic meaning.]

There are lots of random-number generators in use today, both free-standing devices and routines built in to various statistical packages such as SAS, SPSS, and Minitab. The most common applications for random-number generators are in so-called "Monte Carlo" studies and in "bootstrap" inferences (Efron & Tibshirani, 1993). Monte Carlo studies are often undertaken in order to construct empirical sampling distributions for various statistics whose theoretical sampling distributions are too difficult to derive mathematically. (Well-defined population distributions having various properties are repeatedly sampled a very large number of times.) Bootstrap non-parametric methods are also used in the approximation of sampling distributions. A set of randomly sampled observations of size n is itself randomly sampled a very large number of times, with replacement both within sample and between samples, a particular statistic of interest is calculated for each sample, and a frequency distribution for that statistic is obtained. The actual value of the statistic for the given sample is compared to "all possible values" of the statistic for the given sample size, and an appropriate statistical inference is made (a test of a null hypothesis regarding, or a confidence interval for, a particular parameter of interest).

For an interesting discussion of the generation of random sequences of numbers I recommend the chapter by Pashley (1993). For further reading on random-number generators in general and for brief discussions of Monte Carlo and bootstrap methods, I recommend Chapter 8 in Bennett's (1998) book [entitled "Wanted: Random Numbers"); Chapter 9 in Peterson's (1998) book [he points out that many "seeds" for random-number generation lead to repeating cycles]; pp. 28-32, 243-245, and 289-291 of Salsburg's (2001) book; and the University of Utah's random-number-generators webpage. There is also the clever cartoon on page 65 of the Gonick and Smith (1993) book that shows a person pressing buttons to get random (actually pseudo-random) numbers.

Section 6: Where can you find tables of random numbers?

In his delightful essay, "Randomness as a resource", Hayes (2001) discussed the history and the ubiquity of random numbers and pseudo-random numbers, ranging from the results of 26,306 die rolls made by W.F.R. Weldon and his wife that were analyzed by Karl Pearson in 1900, to the 4.8 billion random bits recently made available to the public by Marsaglia (1995).

One of the most accessible and most frequently used tables is the RAND Corporation's table (1955, 2002), A million random digits with 100,000 normal deviates ["normal deviates"--how's that for an oxymoron?!]. That table has been subjected to all sorts of tests of randomness (see following section), and there have even been published Errata concerning it--provided by the statistician I.J. Good and available on the web. Brief excerpts taken from various portions of the table have been published in the backs of countless numbers of statistics textbooks. The numbers are alleged to be in random order when read in any direction (horizontally, vertically, diagonally, or whatever). Gardner (1975b) made some interesting observations concerning the RAND table, including his reference to a claim by Bork (1967) that such a table is strictly a twentieth century phenomenon: "A rational nineteenth-century man would have thought it the height of folly to produce a book containing only random numbers." (p. 40 of Bork)

Other "classic" random-number tables are those generated by Tippett (1927), by Kendall and Babington-Smith (1938), and by Peatman and Schafer (1942). Kendall (1941) also propounded a theory of randomness, to which reference has already been made (see Section 2). In her book, Bennett (1998) pointed out that Tippett's alleged random numbers were soon (within ten years) found to be inadequate for many sampling applications (see Yule, 1938).

With the development of computers having ever-increasing speed and capacity, accompanied by the ever-increasing expertise of people to program them, it is far more common, however, for researchers to use random-number routines that are included in popular statistical packages such as Minitab. In even the older versions of Minitab (which I personally prefer to the newer ones), all you need do is give commands such as "Set 1:100 C1" and "Sample 10 C1 C2" and Shazam! you get a random sample of 10 numbers in Column 2 out of the 100 numbers from 1 to 100 in Column 1. You need to exercise extreme caution when using some computer-generated numbers that are alleged to be random, however. In a devastating but humorous article, Marsaglia (1968) showed that one popular (at the time) random-number generator, RANDU, was not very random at all.

You can also get random numbers on the internet. The website, random.org, for example, "offers true random numbers to anyone on the internet" [to quote from its website]. It also provides information regarding how those numbers are generated. And for a very rich source for the generation of random numbers and for all sorts of other statistical calculations I recommend the Interactive Stats section of the http://www.statpages.net website. It is REALLY nice.

Reference has already been made to the RAND book of single-digit random numbers, which also contains random decimal numbers that range from approximately -3.00 to +3.00 and have a normal distribution rather than a rectangular distribution. (See Box & Muller, 1958 and the Taygeta website concerning the generation of random normal deviates.)

Believe it or not, you can even get random names (first names--both male and female--and last names) on the internet. The website that provides them (for free) is www.kleimo.com. (They have over 21 million of them.) It's great fun. You can also choose an "obscurity factor" from 1 to 20 (the higher the number, the more obscure the name). I asked for five names that could be either male or female and had an obscurity factor of 20. I got as output the following names:

1. Minaya
2. Penelope Kornreich
3. Katy Pattillo
4. Jessie Bolten
5. Zelma Whitesides

You can't get much more obscure than that! Try it sometime. You'll like it.

How about random words? Yes, you can get those too--on the internet from Gammadyne software. The examples they give of words created by their Random Word Generator are: Asprari, Cropoli, Eclon, Enthyme, Flun, Lycand, Mofra, Nespresco, Nokamu, Shrunt, Strelm, and Vermack. They don't have any meanings (necessarily) and some of them are hard to pronounce, but think of all the possibilities! And you can get random "passphrases" that have been found to be useful in developing security codes. (See the Random Passphrase and Diceware Passphrase websites.)

And to properly use random words you need random sentences, don't you? No problem; see Charles Kelly's "Fun with Randomly-Generated Sentences" webpage, http://www.manythings.org/rs/.

Random poems? Try the plagiarist.com website. Random mathematical quotations? You can get them at the math.furma.edu website.

There is a website that will give you random facts (although I don't think they're generated by a random process) and another website that creates a random password (for free) that provides greater security protection than a password based upon lucky numbers, birthdates, or whatever.

Do you like baseball? I do (I'm what they call a "baseball nut"). There is a website called baseball-reference.com where you can call up a randomly-selected former major league baseball player and get all of his lifetime statistics.

If that isn't enough for you, you may want to try the random birthday generator that is included in John Pezzullo's marvelous collection of statistics stuff on the statpages.net website, and convince yourself that the probability of at least two out of n people having the same birthday is really quite high for relatively small n such as 25 or 30.

Finally, not to be outdone, another enterprising website (http://www.noentropy.net) with tongue in cheek provides NON-RANDOM, deterministic numbers. You can request from 1 to 10,000 of such numbers, but they're all 1.

Section 7: What are tests of randomness?

Most tests of randomness are tests of statistical significance applied to the product of a process (not the process itself) that is hypothesized to be random. The products are usually sequences of letters or numbers whose random or non-random pattern is of concern. Such tests are thought to be necessary by those who claim that randomness is strictly a property of a product; they are thought to be unnecessary and irrelevant by those who argue that randomness is strictly a property of a process.

The simplest of these is the so-called "runs test". (See Siegel & Castellan, 1988 and McKenzie et al., 1999.) Consider one sequence of the results of tossing a coin 30 times:

HHHHHHHHHHHHHHHTTTTTTTTTTTTTTT.

That sequence has two runs--an uninterrupted run of 15 Hs and an uninterrupted run of 15 Ts. Intuitively that seems like <u>too few</u> runs for a sequence to be judged to be random. It could have happened "by chance" with a fair coin, but most people would doubt it and would reject the hypothesis that the coin was a fair coin. The runs test would also reject that hypothesis.

Next consider another sequence of the results of tossing a different coin 30 times:

HTHTHTHTHTHTHTHTHTHTHTHTHTHTHT.

That sequence has 30 runs (of one symbol each). Intuitively that seems like <u>too many</u> runs for a random sequence. The runs test would also judge that coin to not be a fair coin.

Now consider the sequence HTHHTTHHHTTTHHHHTTTTHHHHHTTTTT. That has 10 runs--a run of one H, a run of one T, a run of two Hs, a run of two Ts, a run of three Hs, a run of three Ts, a run of four Hs, a run of four Ts, a run of five Hs, and a run of five Ts. That sequence "looks" more random, and the runs test would not reject the hypothesis that a fair coin produced it. But note the "non-random" regularity of that sequence.

Finally, consider the sequence TTHTHTHHHTHTTTHHTHHHTHTTTTHHHTH. That has 18 runs (count 'em), looks much more random than any of the previous sequences, would be judged to be random by the runs test, and has no apparent regularity.

[Are you bothered by the fact that all four of the sequences were "too good" in the sense that each yielded exactly 15 heads and exactly 15 tails? Hmmm. Seems like there's two kinds of randomness or non-randomness going on here--one involving the relative numbers of heads and tails and the other involving the order in which they appear.]

For most tests of randomness all that can really be determined is a sort of "relative randomness", i.e., whether a sequence of digits is "more random than regular" or "more regular than random" (Griffiths & Tenenbaum, 2001; Pincus & Kalman, 1997; Pincus & Singer, 1996).

A particularly interesting approach to testing for randomness that does NOT involve a test of statistical significance is due to Chaitin (1966; 1975; 1990; 2001; 2002), who argued (drawing upon the previous work of Kolmogorov) that a sequence of digits is random if the shortest computer program for generating it is at least as long as the sequence itself. (See also Devlin, 2000.) By that rule the first example in this section (with Hs and Ts replaced by 1s and 0s, respectively) would be judged to be non-random, because a command "print 15 1s followed by 15 0s" is shorter than the sequence 111111111111111000000000000000. That particular sequence would also be judged to be non-random by the runs test.

The second example would also be judged to be non-random by Chaitin's definition (and by the runs test), because "print 15 pairs of alternating 1s and 0s" (or some such command) is shorter than 101010101010101010101010101010.

The third example is the most interesting. Unlike the runs test decision of "random", Chaitin's rule would claim "non-random" because the command "print patterns of pairs of alternating 1s and 0s, starting with one of each" (or, again, something to that effect--you can tell that I'm not a computer programmer!) can be made to be shorter than 101100111000111100001111100000.

The fourth example is easy. There is apparently no command other than "print 001010111010001101110100001101" that is shorter than 001010111010001101110100001101 itself.

I recommend that you "play around" with various sequences of 1s and 0s such as the above, "eyeball" each of them to make a judgment concerning which of the sequences you personally would regard as random and which you would not, and then subject each sequence to one or more of the tests of randomness and to Chaitin's rule, and see how the formal tests agree with your personal judgments.

Two of the most interesting sequences of digits are the mathematical constants e (the base for natural logarithms) and π (the ratio of the circumference of a circle to its diameter). The first of these, e, is equal to 2.71828... (it never cuts off) and the second of these, π, is equal to 3.14159...(it doesn't either). The question has been asked: Are the digits of π random? [It has also been asked about the digits of e.] Pathria (1961), for example, tested the first 10,000 digits of π for randomness; more recently, Bailey and Crandall (2001) subjected the first six billion digits of π (yes, it's known to at least that many places--see the article in the October, 1989 issue of Focus) to a test of randomness and found that each of the digits 0-9 appeared about six hundred million times. On the basis of that test they claimed that the six-billion-digit sequence for π is random. [By the way, the sequence of digits for π has also been put to music--see the MuSoft Builders website.]

There are a few other tests of randomness. For information regarding how to use them I suggest that you go to Chris Wetzel's website (just give Google the prompt "Wetzel randomness" and click on the first entry), or check out Marsaglia's (1995) series of such tests (called DIEHARD tests). And if you would like to study the results of coin tosses and are too lazy to actually toss the coins yourself, Ken White's Coin Flipping Page (on the web) will do it for you!

There is also an interesting literature in psychology regarding the extent to which people are accurate in their judgments about randomness. (See, for example, Bar-Hillel & Wagenaar [1991, 1993], Falk [1975, 1981], and Falk & Konold [1997]. The first four of those references all bear the title “The perception of randomness”.)

Section 8: What is "random sampling" and why is it important?

(Simple) random sampling is a type of sampling from a population which is such that every sample of the same size has an equal chance of being selected. The word "simple" has been written in parentheses before "random sampling" because there are other types of random sampling (e.g., stratified random sampling--see Section 15), but whenever "random sampling" is not further modified it is assumed to be simple random sampling.

Consider a small population consisting of five persons A, B, C, D, and E. If you would like to draw a random sample of two persons from that population you must utilize a process such that the ten combinations A&B, A&C, A&D, A&E, B&C, B&D, B&E, C&D, C&E, and D&E are equally likely to be drawn. For example, you could write each of those ten combinations on a separate piece of paper, put those pieces of paper in a hat, stir them up, and draw out one piece of paper. The two letters that are written on that piece of paper would identify the two people who are to constitute your sample.

There are several reasons why random sampling is important:

(1) It is fair. If one of those combinations, say A&E [the TV network?] had a greater likelihood of being drawn than the others, any results based upon those two persons would be biased in their favor.

(2) It is objective. If you and I agree to draw a random sample of two persons from that population of five persons, your sample might be different from mine, but neither of us would be introducing any subjectivity into the process.

(3) Random sampling is an assumption that underlies just about every procedure for making statistical inferences from samples to populations. The formulas and tables that are commonly used for such inferences are not appropriate for non-random samples. As Bennett (1998) wrote: "The only way we can legitimately rate or compare the value of an observed sample statistic, such as a mean or a sum, within the hypothesized sampling distribution is to select a sample randomly." (p. 110)

Tables of random numbers (see Section 6) are the principal sources for carrying out random sampling. Rather than using letters, pieces of paper, and a hat for drawing that sample of two persons from a population of five persons, you could "code" them 1, 2, 3, 4, and 5 rather than A, B, C, D, and E; open up a table of single-digit random numbers to a "random" page (oh,oh! sounds circular already, doesn't it? but let's press on); close your eyes; drop your finger on a "random" spot on that page (that's worse?); and read off the digit that your finger landed on, along with the one to its immediate right (why the right? that's worst?). If both of those digits are different and/or are not 1, 2, 3, 4, or 5 you'll have to read a little further to the right (even "come around the bend" to the next row of digits?) until you encounter two different digits in the 1-5 range. Strange business, this randomness, isn't it?

I like to make the distinction between sampling persons and sampling already-existing measurements taken on persons. [Most people don't make this distinction, but you may have already figured out that I'm a loner when it comes to certain things.] Consider again this same simple example of persons A, B, C, D, and E, and an interest in measuring their heights. If we draw two persons from the five persons and then measure their heights in inches, should we be concerned about the randomness of the various identifying combinations A&B, A&C, etc., or should we be concerned about the randomness of their heights, say 67& 72, 67 & 64, etc.? As you can gather from my remarks in Section 4, I would argue that we should only be concerned with the former, because the randomness--or non-randomness--occurs before the measurement. On the other hand, if their heights have already been measured, are stored in a file, and are sampled, some (for example, Siegel & Castellan, 1988) would be inclined to argue that it is the randomness of the heights (more specifically, the order in which the heights are drawn) that is of concern. I wouldn't, although I think it is important to make the distinction between sampling persons and sampling measurements. It is the process of selecting the persons that is to be judged to be random or non-random (in this case, how we choose the persons from the file), not the product of the process (in this case their heights).

There are those who do not actually draw samples at random but "regard" their samples as having been randomly drawn, for purposes of inferring from sample to population. Or they say something like: I'm using inferential statistics to generalize from the sample of subjects that I have to a population of subjects "like these". I think both of those positions are indefensible.

The statistcal literature is replete with a number of excellent discussions of random sampling. But once again I especially recommend that you see some of the cartoons in Gonick and Smith (1993) that provide visual representations of the essential aspects of random sampling. The best of these, in my opinion, appear on pages 92 and 138.

Section 9: What is the difference between random sampling and random assignment?

One of the most bothersome (to me, anyhow) shortcomings in the methodological literature is the failure to properly distinguish between random sampling and random assignment. [Edgington (1995 and elsewhere) and Ludbrook & Dudley (1998) are notable exceptions.] The confusion is understandable since the two have much in common, but they have different purposes.

As indicated in the previous section, the purpose of random sampling is to provide the basis for making a statistical inference from a sample to the population from which the sample is drawn. We summarize the sample data by calculating some <u>statistic</u> (which we then know) for those data (a mean, a variance, a correlation coefficient, whatever) and infer something about the corresponding <u>parameter</u> (which we don't and may never know) for the sampled population. In terms of the jargon of the popular Campbell and Stanley (1966) book on experimental design, the matter is one of <u>external validity</u> (generalizability).

The purpose of random assignment (sometimes called "randomization") is quite different . First of all, it applies only to experimental research in which the independent variable will be "manipulated", i.e. some sort of "intervention" is to take place. We randomly assign subjects (participants in human research; animals in infra-human research; iron bars in metallurgical research; whatever) to "treatments" (the interventions) so that each subject has the same chance of being assigned to each treatment. Why? We want the subjects in the various treatment groups to be comparable at the beginning of the experiment, so that if they differ at the end of the experiment we can be reasonably assured that it is the treatments that "did it". That, again in the jargon of Campbell and Stanley, is a matter of <u>internal validity</u> (causality). For a particularly good discussion of causality see the article by Holland (1986), the accompanying comments, his rejoinder, and his later chapter (1993). There is also the paper by Freedman (2002). That paper gets a bit technical in spots, but he (Freedman, the senior author of the popular Freedman, Pisani, & Purves, 1998 statistics textbook) writes extremely well.

Ideally, we would like an experiment to possess both features (generalizability and causality), i.e., we would like to employ both random sampling and random assignment. For example, if we were to compare two different methods of teaching subtraction (take it from me that there are at least two methods), we would draw a random sample from the population of interest (say, all second graders in Los Angeles) and randomly assign half of them to be given Method A and the other half of them to be given Method B. If we were able to carry out such an experiment, we would be justified in using the traditional t-test of the significance of the difference between two independent sample means, provided that we were willing to make the usual assumptions of normality and homogeneity of variance. It is the random sampling that provides the justification.

Suppose, however, that we had random assignment but not random sampling, i.e., we had a non-random ("convenience") sample of the population to which we would like to generalize. In that case the appropriate analysis would not be the traditional t-test (which assumes random sampling), but a randomization test (not to be confused with a test of randomness)--sometimes called a permutation test (see Edgington, 1995 for the details)--which would provide the basis for the generalization not to the population itself (because of the lack of random sampling) but from the particular way that the subjects in the sample happen to have been allocated to the treatments to all of the possible ways that those same people could have been allocated to the treatments. That would in turn provide a basis for claiming causality for those subjects, but any generalization to the full population would have to be a non-statistical one (based upon the researcher's judgment of the representativeness of the non-random sample).

For non-experimental research you might have random sampling but not random assignment (since there are no "treatments" to "assign" subjects to), in which case you would have the statistical basis for generalizability to the population, but an insufficient basis for assessing causality.

Finally, you might find yourself in a position of not having the luxury of either random sampling or random assignment. That doesn't necessarily mean that you should not carry out the study and report its results. But it does mean that you are restricted to the use of descriptive statistics only, with any sort of causal or generalizable interpretation being necessarily subjective and inadvisable. The best approach, in my opinion, to approximating causality in such studies is to use the technique called "propensity score analysis" (PSA)--see Rosenbaum and Rubin (1983) and Pruzek and Helmreich (in preparation). There is also the method advocated by Copas and Li (1997), whose long article bears the unfortunate title "Inference for non-random samples" (it is concerned solely with observational studies that have non-random assignment). And for some kinds of medical investigations with genetic aspects there is a recent approach that incorporates so-called "Mendelian randomization" (see Davey Smith & Ebrahim, 2003).

Associated with random assignment is the matter of "blocking". Rather than simply randomizing n subjects to two treatments A and B, with n/2 people assigned to each treatment, one might first create "blocks" of, say, four people each according to some variable, e.g., age, and randomly assign, within blocks, two people to Treatment A and two people to Treatment B. The four oldest persons would wind up two/two; the next four oldest persons also two/two; etc., with the total numbers in each of the treatments still n/2. The process could be refined even more by creating within each block what are sometimes called "random sandwiches", with the first oldest and the fourth oldest persons constituting the bread and the middle two persons constituting the meat. This is similar to the scheme used in a sport such as doubles in tennis, where the strongest player and the weakest player are pitted against the two players who are intermediate in strength.

[I can't resist pointing out that there are websites where you can purchase "real" random sandwiches. Just give the order and a sandwich consisting of a strange combination of ingredients will be delivered to your door!]

Although the objectives of random sampling and random assignment are different, you can use the same random-number tables for accomplishing both. One such source is the "Research Randomizer" website (www.randomizer.org). It will provide you, at no charge, a random sample of any n numbers out of any N numbers (where n is less than or equal to N) and/or a random assignment of n numbers into subsets of n_1 , n_2 , ..., n_k where the sum of those subscripted n's is equal to n, and k is the number of "treatments".

For a fine article on the importance of randomization in experiments, see Boruch (2002). For an interesting exchange concerning random sampling vs. random assignment, see Shaver (1993) and Levin (1993). (See also Levin, 2002.) For a later "debate" on the necessity for randomization in clinical trials, see the article in Research in Nursing & Health by Sidani, Epstein, and Moritz (2003) and the commentary regarding that article by Ward, Scharf Donovan, and Serlin (2003).

Section 10: What is the randomized response method in survey research?

One of the most common problems in survey research is the refusal by many people to respond to sensitive questions that deal with emotionally charged matters, e.g., sexual behavior and religious beliefs. Even if they are promised anonymity ("nobody will ever be able to associate your response with your name") or confidentiality ("I can but I won't tell anyone else"), they will refuse to answer those questions when posed to them in a face-to-face interview and they will leave blank all such questions on a written questionnaire.

Several years ago Stanley Warner (1965) devised an ingenious procedure for trying to minimize non-response to sensitive questions, a method he called "randomized response". It goes something like this:

Let's say you were interested in estimating the percentage of college students who smoke marijuana (a sensitive matter that also has legal ramifications). Each respondent could be asked to use a table of random numbers to select a random number from 00 to 99 and if the number selected is, say, 70 or above the student would be asked to respond to a sensitive question such as "Do you smoke marijuana at least once a week?" If the number is 69 or below the student would be asked to respond to an unrelated question such as "Is the last digit of your student ID number odd?" [This example is described in considerable detail in the delightful article by Campbell & Joiner (1973) entitled "How to get the answer without being sure you've asked the question".] Nobody need know who answered which question (the responses consist of a simple "yes" or "no" for each student), but by making certain reasonable assumptions (that the percentage of students who have answered the sensitive question is 30 and the percentage of students who have an odd ID number is 50) and using a standard formula for conditional probabilities, the percentages of "yes" answers to the sensitive question can be calculated.

There are several variations of this technique that have been devised over the years (see, for example, Fox & Tracy, 1986) but all have the same objective of estimating the percentage of respondents who hold certain views on various sensitive issues or who engage in various sensitive practices.

Section 11: Under what circumstances can data be regarded as missing at random?

One of the most frustrating problems in data analysis is the absence of one or more pieces of data. Researchers usually go to great lengths in designing their studies, choosing the appropriate measuring instruments, drawing their samples, and collecting their data, only to discover that some of the observations are missing, due to a variety of reasons (an item on a questionnaire left blank, a clerk's neglect to enter an important piece of information, a data-recording instrument's failure, etc.). How do you cope with such a problem? The literature suggests that there are essentially two strategies--deletion or imputation. You can delete all or some of the non-missing data for the entities for which any data are missing; or you can try to impute (estimate) what the missing data "would have been". If you choose the imputation strategy, the literature further suggests that you need to determine whether or not the data are "missing at random". But when are data missing at random?

My personal opinion is "never", but I am apparently in the minority. One of the gurus of missing data, Donald B. Rubin (1976), defined three kinds of "missingness" (see also Little & Rubin, 2002):

1. Missing at random (MAR). Data are said to be MAR if the distribution of missingness does not depend upon the actual values of the missing data (i.e, what they would have been).

2. Missing completely at random (MCAR). Data are said to be MCAR if the distribution of missingness also does not depend upon the actual values of the other data that are not missing.

3. Missing not at random (MNAR). Data are said to be MNAR if the distribution of missingness does depend upon the actual values of the missing data.

The article by Schafer and Graham (2002) is particularly good for further clarifying those three kinds of missingness, along with examples of each. Rubin's definitions of random missingness are "product-oriented" rather than "process-oriented", i.e., one needs to make certain asumptions and/or analyze some actual evidence in order to determine whether or not, or the extent to which, data might be missing "at random". That view, although perhaps correct, is contrary to mine. It also appears (at least to me) to be contrary to most people's concept of a random phenomenon, where chance should play an essential role. People don't flip coins, roll dice, or draw cards in order to determine whether or not they will respond to a particular item on a questionnaire. Data entry clerks don't employ such devices in order to determine whether or not they will enter a participant's response in a data file. And recording instruments don't choose random times to break down. Do they?? And how can you analyze non-missing data to draw conclusions regarding the randomness or non-randomness of missing data?

In the spirit of Rubin's product-dependent concept of randomness, Schmitz and Franz (2001) even provide a method (the popular bootstrap technique) for testing whether or not the data from study dropouts are missing at random! Amazing.

The actual size of the total sample and the proportion of missingness need to be taken into consideration no matter what you decide about the "randomness of the missingness". If the sample is large and the proportion of missingness is small, it doesn't really matter what you do. But if, say, you have at least one missing observation for each member of your sample, and it is a different observation for each subject, you have a major, major problem. (In that case so-called "listwise deletion", where all of the data for a subject are deleted if the subject has any missing data, is not an option, since you would then have an n of 0.)

Section 12: What is a random variable?

"Random variable" is a term originating in mathematical statistics, and refers to a variable that can take on any value from some given probability distribution. For example, "outcome of a single toss of a fair coin" is a random variable (a so-called Bernoulli variable) that can take on the value H (1) or T(0) for any particular toss of such a coin with equal probability.

The most commonly discussed, but in my opinion the most over-rated, random variable is a variable that is distributed according to the normal, bell-shaped, Gaussian form. [I'll probably get into trouble for saying it's over-rated, but I (and Micceri, 1989) claim that most variables are NOT normally distributed--income, for example, is not normally distributed in any population---and the primary reason for the popularity of the normal distribution is that the mathematical statisticians know all about it!]

Traub (1994) based his development of classical reliability theory upon the concept of a random variable, as had Lord and Novick (1968) in their very popular textbook. I personally prefer Gulliksen's (1950) approach, which did not explicitly employ the concept of a random variable.

There is an interesting parallel between a random variable and the concept of an "event" in probability theory, as recently pointed out to me by my colleague Ronald Serlin (personal communication, March 10, 2003). Mathematical statisticians talk about a random variable Y taking on the value of y and an event E taking on the value of result e. For example, it is well-known that the probability of a "1" (= y) for a Bernoulli variable (= Y) is equal to P. Similarly, the probability of a "1" (= e) for a (fair or unfair) coin toss (= E) is equal to its P.

An interesting "take" on the concept of a random variable is that of Van Lehn (2002), who claims that a random variable is neither random nor a variable, but is best thought of as simply a function on a sample space.

For an entire collection of cartoons that illustrate the concept of a random variable better than any section of a traditional statistics textbook, see Chapter 4 in Gonick and Smith (1993).

Section 13: What is the difference between random effects and fixed effects in experimental research?

Random effects are effects in an experiment that can be generalized to a "population" of treatments from which a subset of treatments has been randomly sampled. For example, if in a particular study you randomly selected three drug dosages out of ten drug dosages to actually be tested against one another, and you found a statistically significant difference among the three, you would be justified in inferring that there was a difference among all ten. [But you could of course be wrong.]

Fixed effects, on the other hand, are effects that pertain only to the treatments that are specifically employed in the experiment itself. For example, if you were only interested in the effect of 100 milligrams of a particular drug vs. 200 milligrams of that drug, the research subjects were randomly assigned to one or the other of those two doses, and you found a statistically significant difference between those two dosages, you would have no basis for generalizing the findings to other dosages such as 50 milligrams or 500 milligrams.

Some variables that are employed as covariates in experimental research are necessarily fixed, e.g., sex. There are just two sexes, so if and when sex is a variable in a research design generalizations to "other sexes" wouldn't make any sense. (It is of course possible for sex to be held constant, rather than to be a variable, if one were interested solely in an experimental effect for males only or for females only.)

Section 14: Can random sampling be either with replacement or without replacement?

Yes (but see Hoffman, 2002, who claims otherwise). If the first object is not replaced in the population before the second object is drawn, the second object admittedly has a different selection probability, but the process is nevertheless still random. The better approach to the problem, however, is to refer to the formal definition of random sampling (as given by Wallis & Roberts, 1962, for example) and think of the drawing of a random sample as a procedure for selecting <u>a combination</u> of n things from a population of N things, with each <u>combination</u> being equally likely to be the one actually drawn.

The matter of sampling with replacement or without replacement is rather fascinating. Virtually all of traditional statistical inference is based upon sampling with replacement, e.g., all of the well-known parametric significance tests for which a normal population distribution is known or assumed. But most procedures for actually drawing samples are based upon sampling without replacement. When using a table of random numbers (see Section 6), if an ID number is encountered that is the same as an ID number that has already been drawn, the researcher is told to ignore that number and go on to the next. The reason for that is to avoid having a subject being selected more than once; if one or more subjects were in the data two or more times, the sample observations would not be independent.

The saving grace in all of this is that most sample sizes are much smaller than most population sizes, so for all practical purposes it doesn't really matter whether the sampling is with or without replacement. It is extremely unlikely in such cases that a given ID number would be sampled more than once.

Section 15: What is stratified random sampling and how does it differ from stratified random assignment (blocking)?

Stratified random sampling is a two-step process. The population to be sampled is first divided into two or more parts called "strata" (singular: "stratum") and then a simple random sample is drawn from each stratum. For example, if you wanted to take a sample of 30 people from a population that consisted of 100 men and 200 women, and you wanted your sample to have the same proportion of men and women as the population had, you would draw a simple random sample of 10 men from the male stratum and a simple random sample of 20 women from the female stratum.

Stratified random <u>assignment</u>, better designated as <u>blocking</u>, pertains to the sample and not to the population, but is a similar process. That is, you first divide the sample (no matter how it has been drawn) into two or more strata and then you randomly assign to treatments within strata. For example, if you were carrying out an experiment to test the effect of an experimental treatment vs. a control treatment for a sample that consisted of 10 males and 20 females, and you wanted to be sure that you had the proper proportions of males and females in each of the treatments, you would randomly assign 5 of the males to the experimental treatment and the other 5 of the males to the control treatment, and you would randomly assign 10 of the females to the experimental treatment and the other 10 of the females to the control treatment. This would permit you to test the “main effect” of treatment, the “main effect” of sex, and the sex-by-treatment “interaction effect”.

Blocking also plays a key role in the statistical method for testing causality in non-experimental research called propensity score analysis (PSA)--see Section 9. What is entailed is the calculation of a propensity (to fall into one intact group vs. the other) score, the rank-ordering of those scores, the creation of a small, but not too small, number of blocks of those scores, and a comparison of the two groups within each of those blocks. (See Rosenbaum & Rubin, 1983 and Pruzek & Helmreich (in press) for more specific details.)

The best way to remember the difference between stratified random sampling and stratified random assignment (blocking) is to repeat the mantra "stratify the population; block the sample".

For a clever cartoon that illustrates stratified random sampling, see Gonick and Smith (1993), page 95.

Section 16: What is the difference between stratified random sampling and quota sampling?

The principal difference is the presence of a chance element in one (stratified random sampling) and the absence of a chance element in the other (quota sampling). For example, if you wanted to conduct a survey and you insisted that there be the same number of men and women, and of younger and older adults, in your survey, you could stratify the population to be sampled into, say, four strata (male, 18-40; female 18-40; male, over 40; female, over 40) and take a simple random sample of, say, 50 people from each of those strata. OR you could go out into the highways and byways and sample until you got 50 people (any 50 people...your "quota") for each of the four categories. Which approach do you think is better?

Section 17: What is random error?

Random error is any error that can be attributed to chance. The most common use of the term is in various theories of measurement where the emphasis is upon "random measurement error" due to the unreliability of measuring instruments.

The type of error that is associated with having a sample of a population rather than the entire population is usually called, logically enough, "sampling error", but the modifier "random" is taken to be understood. Random measurement error is often called "non-sampling error", but that term is confusing in educational and psychological research, where much of measurement error is actually attributable to the sampling of test items from populations of test items.

There is one other kind of error that is unique to educational and psychological testing. We could arrive at a total score on an achievement test, for example, by counting the number of wrong responses rather than the number of right responses. But the number of right responses and the number of wrong responses are both types of obtained scores, not error scores in the measurement-theoretic sense of that word (see next section), and in any event are not random.

Speaking of right responses and wrong responses, there is a vast literature in educational and psychology on the matter of chance success for certain kinds of tests (multiple-choice, true/false, and matching) in which answers are selected from various given alternatives, i.e., they are not supplied by the test-taker. Some measurement experts claim that scores on such tests should be corrected for chance success by subtracting some fraction of the wrong answers from the total number of right answers. The formulas for so doing assume that (a) some examinees do guess (whether you tell them to or not to) and (b) if and when they guess they guess randomly. Both (a) and (b) are questionable assumptions. I personally feel that the whole matter of correction for chance success has been highly over-rated. (For opposite points of view you might want to read Plumlee, 1952,1954; Mattson, 1965; Zimmerman & Williams, 1965; or the "correction-for-guessing" sections of most introductory textbooks in educational or psychological measurement.) However, I must admit that I once wrote a long article in which I provided formulas for estimating the reliability of a single test item where chance success is possible (Knapp, 1977).

There is an equally vast literature regarding the correction for chance agreement of various inter-rater and intra-rater reliability coeficients--see Cohen's (1960) discussion of his kappa coefficient and any other source that talks about that particular statistic. The formulas for doing that, like the correction for chance success formulas, assume that some raters provide some ratings randomly and some portion of the percent agreement between or within raters is chance agreement. I think that is also questionable and there is little or no need for Cohen's kappa. If there is ever any reason to believe that raters rate randomly all you need to do is raise the standard as to what constitutes "good" percent agreement.

For more on random measurement error, see any or all of the following: Cureton (1931), Cochran (1968), Grubbs (1973), Jaech (1985), Schmidt and Hunter (1996), the "Standards" for educational and psychological tests (AERA, APA, & NCME (1999), and Dunn (2003).

Section 18: Does classical reliability theory necessarily assume random error?

The answer is an emphatic "no", as shown by Gulliksen (1950) in his excellent textbook on the theory of mental tests and as I have tried to demonstrate in my reliability book (Knapp, 2009). Most other sources, however, insist that the kinds of errors that are associated with reliability are random measurement errors. (They also claim that non-random measurement errors are “constant” errors and are associated with invalidity.)

Gulliksen (1950) explained how all of the theorems of classical reliability theory can be derived EITHER by first defining error score as random (mean equal to zero, zero correlation between error for one instrument and error for another instrument, etc.) and then defining true score as the difference between obtained score and error score; OR by first defining true score as the mean of an infinite number of obtained scores on parallel forms of the instrument and then defining error score as the difference between obtained score and true score.

Defining random measurement error first and then letting true score fall out as the difference between obtained score and error score is by far the more popular approach and, as you can see, its definition (alas) is "product-oriented" rather than "process-oriented" because it is based upon the assumption of, and/or evidence for, zero mean, zero correlations with other measurement errors, and the like.

Reference was made above to "an infinite number of obtained scores on parallel forms". The matter of parallel forms has been a controversial one ever since its introduction to measurement theory almost a century ago. It is a reasonable (albeit ethereal) notion in certain contexts such as spelling tests, but not in other contexts such as the measurement of length. For a spelling test it is at least possible to imagine a perhaps not infinite but very large number of parallel forms that could be constructed by randomly sampling words from an unabridged dictionary (without replacement within form but with replacement between forms). But for a yardstick or other instrument for measuring length it is almost impossible to define parallel forms, much less imagine an infinite number of them.

Parallel forms can be "randomly parallel" (as alluded to in the previous paragraph) or "rigorously parallel" (satisfying a variety of conditions). If you are interested in that distinction please see the textbooks written by Kelley (1927), by Gulliksen (1950), by Lord and Novick (1968), and by Nunnally and Bernstein (1994) and/or the articles written by Lord (1964) and by Novick (1966). Kelley was a strong advocate of the parallel form approach to reliability. Another measurement expert, Louis Guttman, questioned the entire concept of parallel forms. He favored the test-retest approach. Others, e.g., Lee Cronbach, contributed to the development of methods for determining “internal consistency” reliability (which Kelley refused to acknowledge as “reliability” at all; he referred to them as procedures for assessing the homogeneity of a number of variables--usually test items--that were alleged to measure the same construct.)

Section 19: What do you do if one or more subjects who are randomly selected for a research study refuse to participate and/or refuse to be randomly assigned to a particular treatment?

In Section 11 I discussed the frustration associated with missing data and whether or not the absent data could be considered as being randomly missing. What is equally frustrating is to draw a random sample of subjects for a research study and have one or more of them refuse to participate at all or, in the case of an experiment, refuse to be assigned to a particular treatment. If that happens, what should you do?

You have several choices:

1. You can augment the cooperating segment of the sample with another random sample equal in size to that of the non-cooperating segment. For example, if you had selected a random sample of 50 subjects from a particular population and 10 of them refused to participate, you could select a random sample of an additional 10 subjects from that same population to replace the 10 refusals. This would restore your original sample size (unless some of the new 10 also refuse to participate!), but the final sample would not be nearly as "clean" as the originally drawn sample, as far as any inference to the population is concerned, since non-cooperators would not be represented in that sample.

2. You could go ahead and carry out the study on the basis of the sub-sample of cooperating respondents only, but that would constrain the generalizability to the sub-population of cooperators.

3. You could use a complicated analysis called “intent to treat” analysis (see, for example, Green, Benedetti, & Crowley, 2002), wherein the data for the “refuseniks” are counted in with the group to which they were assigned, no matter what they decided to do (or not do) subsequent to that assignment.

4. You could (but don't) try to estimate all of the data that those non-cooperating subjects would have provided had they been cooperators. There is a variety of techniques for estimating data when you have some non-missing data for subjects who do agree to participate in the study but do not provide full data (again see Section 11), but if subjects don't provide <u>any</u> data those techniques don't work.

If you are comparing “pretest” and “posttest” scores and there is some attrition between pretest and posttest, there is one thing you should not do, and that’s to display in a table the summary descriptive statistical information (means, standard deviations, etc.) for the full sample at pretest time and for the reduced sample at posttest time. It’s a classic case of “apples and oranges”. If you feel the need for such a table you must display <u>three</u> sets of summary data: (1)

pretest scores for the “dropouts”; (2) pretest scores for the “non-dropouts”; and (3) posttest scores for the non-dropouts. Summaries (2) and (3) are comparable; summary (1) provides a basis for determining whether the dropouts and the non-dropouts are sufficiently similar for the (2) vs. (3) comparison to be meaningful.

Section 20: What is a random walk?

A random walk is a series of steps that have a point of origin and are such that the direction each step takes is a matter of chance, i.e., it is a random process. The most widely-discussed application is to the drunk who starts at a bar and walks (stumbles, really) in one direction or the other, with each direction having a known probability (the probabilities are usually assumed to be equal) of being followed. (See the cartoons on page 215 of Gonick & Smith, 1993 and on the Physics Random Walk website.) The typical research questions for such a situation are things like "What is the expected location after n steps?" or "What is the probability that the drunk is at a certain distance x, from the bar, after n steps?" (See, for example, the delightful paper by Monte, 2003 for an analysis of the case of x = 0, i.e., the drunk arrives right back at the bar). But there are many more serious applications of random walks to important problems in the physical and the social sciences. (See, for example, the online Training Handbook for Anthropometric Surveys.htm for a fascinating illustration of the use of a random walk for collecting household survey data.)

Is it too much of a stretch for me to claim that I have taken you on sort of a random walk through this monograph? If so, I hope your final destination has been an understanding of the concept of a random phenomenon and not back at "the bar" where you started out.

References

[Note: The numbers within the parentheses at the ends of the references are the numbers of the sections in which the references are cited.]

Abelson, R.P. (1995). Statistics as principled argument. Hillsdale, NJ: Erlbaum. (3)

American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). (1999). Standards for educational and psychological testing. Washington, DC: American Educational Research Association. (17)

Bailey, D.H., & Crandall, R. E. (2001). On the random character of fundamental constant expansions. Experimental Mathematics, 10 (2), 175-190. (7)

Bar-Hillel, M., & Wagenaar, W.A. (1991). The perception of randomness. Advances in Applied Mathematics, 12, 428-454. [Reprinted as Chapter 14 in G. Keren & C. Lewis (Eds.) (1993), A handbook for data analysis in the behavioral sciences: Methodological issues. Hillsdale, NJ: Erlbaum.] (7)

Bartlett, M. (1990). Chance or chaos? Journal of the Royal Statistical Society, Series A, 153, 321-347. (3)

Beltrami, E. (1999). What is random? Chance and order in mathematics and life. New York: Copernicus (Springer-Verlag). (1, 3)

Bennett, D.J. (1998). Randomness. Cambridge, MA: Harvard University Press. (1, 2, 3, 4, 5, 6, 8)

Borel, E. (1962). Probabilities and life. New York: Dover. (2)

Bork, A.M. (1967). Randomness and the twentieth century. Antioch Review, 27, 40-61. (6)

Boruch, R. (Fall, 2002). The virtues of randomness. Education Next, pp. 37-41. Stanford, CA: Hoover Institution. (9)

Box, G.E.P., & Muller, M.E. (1958). A note on the generation of random normal deviates. Annals of Mathematical Statistics, 29, 610-611. (6)

Campbell, C., & Joiner, B.L. (1973). How to get the answer without being sure you've asked the question. The American Statistician, 27 (5), 229-231. (10)

Campbell, D.T., & Stanley, J.C. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand McNally. (9)

Chaitin, G.J. (1966). On the length of programs for computing finite binary sequences. Journal of the Association for Computing Machinery, 13, 547-569. (7)

Chaitin, G.J. (1975). Randomness and mathematical proof. Scientific American, 232, 47-52. (7)

Chaitin, G.J. (1990). A random walk in arithmetic. New Scientist, 125 (1709), 44-46. (7)

Chaitin, G.J. (2001). Exploring randomness. London: Springer-Verlag. (7)

Chaitin, G.J. (2002). Paradoxes of randomness. Complexity, 7 (5), 14-21. (7)

Cochran, W.G. (1968). Errors of measurement in statistics. Technometrics, 10, 637-666. (17)

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46. (17)

Copas, J.B., & Li, H.G. (1997). Inference for non-random samples. Journal of the Royal Statistical Society, Series B, 59 (1), 55-95. (Includes discussions by several statisticians and a reply by the authors.) (9)

Cureton, E.E. (1931). Errors of measurement and correlation. Archives of Psychology, 19, No. 125. Pp. 63. (17)

Davey Smith, G., & Ebrahim, H. (2003). ‘Mendelian randomization’: Can genetic epidemiology contribute to understanding environmental determinants of disease? International Journal of Epidemiology, 32, 1-22. (9)

Dunn, G. (2003). Statistical evaluation of measurement errors: Design and analysis of reliability studies (2nd. ed.). New York: Edward Arnold. (17)

Edgington, E.S. (1995). Randomization tests (3rd. ed.). New York: Marcel Dekker. (9)

Efron, B., & Tibshirani, R.J. (1993). An introduction to the bootstrap. New York: Chapman and Hall. (5)

Falk, R. (1975). The perception of randomness. Unpublished doctoral dissertation, The Hebrew University (in Hebrew). [Reprinted in English in 1981 in

Proceedings of the Fifth International Conference for the Psychology of Mathematical Education (pp. 222-229). Grenoble, France: Laboratoire I.M.A.G.] (7)

Falk, R., & Konold, C. (1994). Random means hard to digest. Focus on Learning Problems in Mathematics, 16, 1-12. (7)

Ford, J. (1983). How random is a coin toss? Physics Today, 36, 40-47. (3)

Freedman, D. (2002). On specifying graphical models for causation, and the identification problem. Technical Report No. 601. Berkeley, CA: Statistics Department, The University of California. (9)

Freedman, D.A., Pisani, R., & Purves, R. (1998). Statistics (3rd. ed.) New York: Norton. (3)

Fox, J.A., & Tracy, P.E. (1986). Randomized response: A method for sensitive surveys. Newbury Park, CA: Sage. (10)

Gardner, M. (1975a). Card shuffles. In M. Gardner, Mathematical carnival. New York: Knopf. Pp. 123-138. (3)

Gardner, M. (1975b). Random numbers. In M. Gardner, Mathematical carnival. New York: Knopf. Pp. 161-172. (6)

Gardner, M. (1979). The random number Omega bids fair to hold the mysteries of the universe. Scientific American, 241, 20-34. (5)

Gonick, L., & Smith, W. (1993). The cartoon guide to statistics. New York: HarperCollins. (1, 2, 5, 8, 12, 15, 20)

Green, S., Benedetti, J., & Crowley, J. (2002). Clinical trials in oncology (2nd. ed.). New York: Chapman & Hall. (19)

Griffiths, T.L., & Tenenbaum, J.B. (2001). Randomness and coincidences: Reconciling intuition and probability theory. Paper presented at the 23rd Annual Conference of the Cognitive Science Society. (7)

Grubbs, F.E. (1973). Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments. Technometrics, 15, 53-66. (17)

Gulliksen, H. (1950). Theory of mental tests. New York: Wiley. (12, 18)

Hayes, B. (2001). Randomness as a resource. American Scientist, 89 (4), 300-304. (6)

Hoffman, H.S. (2002). Definition of a random sample. Internet glossary of statistical terms. The Animated Software Company. (14)

Holland, P.W. (1986). Statistics and causal inference. Journal of the American Statistical Association, 81, 945-970. [Includes comments by other statisticians and Holland's rejoinder.] (9)

Holland, P.W. (1993). Which comes first, cause or effect? Chapter 9 in G. Keren & C. Lewis (Eds.), A handbook for data analysis in the behavioral sciences: Methodological issues. Hillsdale, NJ: Erlbaum.

Jaech, J.L. (1985). Statistical analysis of measurement errors. New York: Wiley. (17)

Kac, M. (1983). What is random? American Scientist, 71, 405-406. (1)

Keene, G.B. (1957). Randomness II. Proceedings of the Aristotelian Society, Supplement, 31, 151-160. (2)

Kelley, T.L. (1927). The interpretation of educational measurements. New York: World Book. (18)

Kendall, M.G. (1941). A theory of randomness. Biometrika, 32, 1-15. (2, 6)

Kendall, M.G., & Babington-Smith, B. (1938). Randomness and random sampling numbers. Journal of the Royal Statistical Society, Series B, 6, 51-61. (6)

Keren, G., & Lewis, C. (Eds.). (1993). A handbook for data analysis in the behavioral sciences: Methodological issues. Hillsdale, NJ: Erlbaum.

Knapp, T.R. (1977). The reliability of a dichotomous test item: A "correlationless" approach. Journal of Educational Measurement, 14, 237-252. (17)

Knapp, T.R. (2009). The reliability of measuring instruments. Accessible free of charge at the following website: www.StatLit.org/Knapp.htm (18)

Levin, J.R. (1993). Statistical significance testing from three perspectives. The Journal of Experimental Education, 61, 378-382. (9)

Levin, J.R. (2002). Random thoughts on the (in)credibility of educational psychological intervention research. Paper presented at the annual meeting of the American Psychological Association, August 2002, Chicago. (Division 15 Thorndike Award invited address) (9)

Levinson, H.C. (1963). Chance, luck, and statistics. New York: Dover.

Little, R.J.A., & Rubin, D.B. (2002). Statistical analysis with missing data. (2nd. ed.) New York: Wiley. (11)

Liu, Y., & Thompson, P.W. (2002). Randomness: Rethinking the foundation of probability. Athens, GA: Proceedings of the 24th Annual Meeting of the International Group for the Psychology of Mathematics Education. (1)

Lord, F.M. (1964). Nominally and rigorously parallel test forms. Psychometrika, 29, 335-346. (18)

Lord, F.M., & Novick, M.R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley. (12, 18)

Ludbrook, J., & Dudley, H. (1998). Why permutation tests are superior to t and F tests in biomedical research. The American Statistician, 52 (2), 127-132. (9)

Mackenzie, D. (2002). The mathematics of shuffling: The Stanford flip. Discover, 23 (10), 22-23. (3)

Marsaglia, G. (1968). Random numbers fall mainly in the planes. Proceedings of the National Academy of Sciences, 61, 25-28. (6)

Marsaglia, G. (1995). The Marsaglia random number CDROM, including the DIEHARD battery of tests of randomness. Tallahassee, FL: Department of Statistics, Florida State University. (6, 7)

Mattson, D. (1965). The effects of guessing on the standard error of measurement and the reliability of test scores. Educational and Psychological Measurement, 25, 727-730. (17)

May, M. (May-June, 1997). What is random? American Scientist, 85 (3). (1)

McKenzie, D.P., Onghena, P., Hogenraad, R., Martindale, C., & MacKinnon, A.J. (1999). Detecting patterns by one-sample runs test: Paradox, explanation, and a new omnibus procedure. Journal of Experimental Education, 67 (2), 167-179. (7)

McNamara, C. (2003). Heads or tails? It really is a tossup. Chicago Tribune, January 26th. (3)

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin, 105 (1), 156-166. (12)

Monte, R.A. (2003). The random walk for dummies. Accessible at http://www-math.mit.edu/phase2/UJM/vol1/RMONTE-F.PDF, pp. 143-148. (20)

Moore, D.S. (1979). Statistics: Concepts and controversies. New York: Freeman. (2)

Moore, D.S. (2001). Statistics: Concepts and controversies (5^{th} ed.) New York: Freeman. (2)

Muchinsky, P.M. (1996). The correction for attenuation. Educational and Psychological Measurement, 56, 63-75. (1)

Novick, M.R. (1966). The axioms and principal results of classical test theory. Journal of Mathematical Psychology, 3, 1-18. (18)

Nunnally, J.C., & Bernstein, I.H. (1994). Psychometric theory (3rd. ed.). New York: McGraw-Hill. (18)

Pashley, P.J. (1993). On generating random sequences. Chapter 15 in G. Keren & C. Lewis, A handbook for data analysis in the behavorial sciences: Methodological issues. Hillsdale, NJ: Erlbaum.

Pathria, R.K. (1961). A statistical study of randomness among the first 10,000 digits of π. Mathematics of Computation, 16, 188-197. (7)

Peatman, J.G., & Schafer, R. (1942). A table of random numbers from Selective Service numbers. The Journal of Psychology, 14, 295-305. (5)

Peterson, I. (1998). The jungles of randomness. New York: Wiley.

Peterson, I. (1999). Fibonacci at random: Uncovering a new mathematical constant. The weekly newsmagazine of science, 155 (24), 376. (3)

Pincus, S., & Kalman, R.E. (1997). Not all (possibly) “random” sequences are created equal. Proceedings of the National Academy of Sciences, 94, 3513-3518. (7)

Pincus, S., & Singer, B.H. (1996). Randomness and degrees of irregularity. Proceedings of the National Academy of Sciences, 93, 2083-2088. (7)

Plumlee, L.B. (1952). The effect of difficulty and chance success on item-test correlations and test reliability. Psychometrika, 17, 69-86. (17)

Plumlee, L.B. (1954). The predicted and observed effect of chance success on multiple-choice test validity. Psychometrika, 19, 65-70. (17)

Pruzek, R. M., & Helmreich, J. (in preparation). Comparing treatments in observational studies: Propensity score methods. (9, 15)

RAND Corporation (1955). A million random digits with 100,000 normal deviates. Glencoe, IL: Free Press. [Reprinted in 2002] (6)

Rescher, N. (2001). Luck: The brilliant randomness of everyday life. Pittsburgh, PA: University of Pittsburgh Press. (3)

Rosenbaum, P., & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70, 41-55. (9, 15)

Rubin, D. B. (1976). Inference and missing data. Biometrika, 63, 581–592. (11)

Salsburg, D. (2001). The lady tasting tea. New York: Freeman. (2, 5)

Schafer, J.L., & Graham, J.W. (2002). Missing data: Our view of the state of the art. Psychological Methods, 7 (2), 147-177. (11)

Schmidt, F.L., & Hunter, J.E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. Psychological Methods, 1, 199-223. (17)

Schmitz, N., & Franz, M. (2002). A bootstrap method to test if study dropouts are missing randomly. Quality & Quantity, 36, 1-16. (11)

Shafer, G. (1993). Can the various meanings of probability be reconciled? Chapter 5 in G. Keren & C. Lewis (Eds.), A handbook for data analysis in the behavioral sciences: Methodological issues. Hillsdale, NJ: Erlbaum. (2)

Shaver, J.P. (1993). What statistical significance is, and what it is not. The Journal of Experimental Education, 61, 293-316. (9)

Sidani, S., Epstein, D.R., & Moritz, P. (2003). An alternative paradigm for clinical research: An exemplar. Research in Nursing & Health, 26, 244-255. (9)

Siegel, S., & Castellan, N.J. (1988). Nonparametric statistics for the behavioral sciences (2nd. ed.). New York: McGraw-Hill. (4, 7, 8)

Spencer Brown, G., & Keene, G. B. (1957). Randomness I. Proceedings of the Aristotelian Society, Supplement, 31, 145-150. (2)

Starnes, D.S., Yates, D.S., & Moore, D.S. (2012). The practice of statistics (4th ed.), New York: Freeman. (1, 2)

Stewart,I. (1989). Does God play dice?: The new mathematics of chaos. New York: Penguin. (1)

Tippett, L.H.C. (1927). Random sampling numbers. Tracts for computers, No. 15. With a foreword by Karl Pearson. London: Cambridge University Press, 1950. [Reprinted in 1959.] (6)

Traub, R.E. (1994). Reliability for the social sciences: Theory and applications. Thousand Oaks, CA: Sage. (12)

vonMises, R. (1957). Probability, statistics, and truth. London: Allen & Unwin. (2)

Wallis, W.A., & Roberts, H.V. (1962). The nature of statistics. New York: The Free Press. (1, 14)

Ward, S., Scharf Donovan, H., & Serlin, R.C. (2003). An alternative view on “An alternative paradigm”. Research in Nursing & Health, 26, 256-259. (9)

Warner, S.L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, 60, 63-69. (10)

Watson, J.M., & Moritz, J.B. (2003). Fairness of dice: A longitudinal study of students’ beliefs and strategies for making judgments. Journal for Research in Mathematics Education, 34 (4), 270-304.

Whitney, C.A. (1984). Generating and testing pseudorandom numbers. Chance 9 (11), 128-129, 442-464. (5)

Yule, G.U. (1938). A test of Tippett's random sampling numbers. Journal of the Royal Statistical Society, 101, 167-172. (6)

Zimmerman, D.W., & Williams, R.H. (1965). Effect of chance success due to guessing on error of measurement in multiple-choice tests. Psychological Reports, 16, 1193-1196. (17)