STATISTICAL DILEMMAS:

Factors contributing to the uncertainty of conclusions from statistical methodology

Or, why

" … no finding made statistically can ever be certain."

by
Dr. Simon R Mouer, PE, PhD
Professor, College of Engineering
University of Southeastern Philippines

TABLE OF CONTENTS

LIST OF FIGURES

CONTENTS

INTRODUCTION

Statistics can be fairly divided into two constructs: *descriptive* statistics and *inferential* statistics.

Descriptive statistics merely describes, without implying anything. Descriptive statistics employs counting occurrences (frequencies) and relative occurrences (frequencies of one classification relative to some other classification; calculating the *mean*, *median* , or *mode*; and calculating the *variation* and *standard deviation* of the set Such *group* descriptions may be by number or percentages. Not much else can be accomplished in *descriptive* statistics without involving *inferential* statistics.

Inferential statistics takes the *group* descriptions and goes a step or two further, verifying that classifications are *significant*, implying or inferring *relationships* and *correlations*, establishing the *strength* or *importance* of relationships, pondering *causations* and/or making *predictions*. Inferential statistics is often taught as if it were deterministic and conclusive to the same degree that well-established and tested laws of science are assumed to be. While the mathematics of inferential statistics is determinant, i.e., data input to the statistical procedures and algorithms yield answers, **the routine application of inferential statistics in the real world too often leads to frivolous, irrelevant, and/or erroneous conclusions**.

The investigator employing inferential statistics must understand that it is a mathematical construct – an *ideal concept* of perfect *randomness* and perfect *probability* distributions applied to completely-defined and perfectly independent variables which are perfectly measured. In such ideal world, the set is *homogeneous*, *randomness* is the only cause of irregularity, and variance of a measure from the random function is smooth and tantamount to systemic cause.

In the *real* world, however, **nothing is perfect**. Many non-random operators cause irregularities, such as chaos and catastrophe, confounding our analysis of real systems. Identifiable variables are seldom truly independent, and many important variables are unseen and unknown. Coincidental and spurious variables and relations appear that suppress, mask or enhance the real actors which often remain elusive. Measurement of variables, if it can be done at all, is imprecise, irregular, and very often indirect. Often we must measure *marker* or *proxy* variables because we cannot directly access or see the causal variables. *Correlation* is too often only coincidental, masking spuriousness and confused by the investigator with *causation*. Measurement of real variables is subject to error, pre-judgment, and bias. Random distributions are not smooth, and differentiation of randomness from irregular systemic causes is difficult or impossible. A serious and seemingly unbridgeable gap exists between the *ideal* concept and the *real* world.

Despite the above shortcomings**, statistical analysis is a powerful and useful tool** for the scientific investigator and the engineer – provided we understand its limitations and caveats, and the knowledge that **no finding made statistically can ever be certain**. Thus, **every statistically-based finding should be subject to repeated re-examination**.

This paper discusses the relationships and identifying characteristics of different, but related, concepts and causes of *uncertainty*, such as *disorder*, *randomness*, *probability*, *catastrophe*, *chaos*, *approximation, precision*, and *order*, and their implications on *prediction* using *ideal* systems to predict *real* behavior. The implications of *sampling* real behavior and *estimating* real populations using *ideal* constructs is discussed. Considerable discussion is given to the limitations of *randomness* because it is an essential underlying concept in statistical analysis, and to the effects of *spuriousness* because it is the single most prominent cause invalidating statistical investigations.

SHORT DEFINITIONS

Approximation – a less than complete representation of a measurement or a system.

Bias – the occurrence of a *systemic* actor to influence the outcome of an otherwise *random* event.

Catastrophe – a sudden and usually unexpected event, or series of events emanating from a system, wherein the event(s) interrupts, collapses, or destroys the system.

Chaos – a series of irregular *systemic* events wherein the precise predictability of a specific event is difficult or impossible with any degree of confidence.

Coincidental – happening at the same time, but otherwise unrelated to another happening.

Conceptual – a *model* or *ideal* of a *real* event or system; a mental or mathematical *representation* of reality.

Discrete – individual, or entity, as in a *discrete* member of a set; not continuous, as in a discrete mathematical function.

Disorder – unpredictability of a series of events, lack of a discernable pattern. Complete disorder implies a total lack of a system in a series of events. Partial disorder may be characterized by varying degrees of *randomness, catastrophe,* or *chaos*.

Event – the thing that happens; that which we observe and measure, and desire to predict.

Gaussian – a family of mathematical functions which characterize randomness by a probability density function, wherein the sum or integral of the area under the density function is equal to one (one hundred percent probablitity.)

Homogeneous – having the same properties throughout.

Homogeneity – the degree to which the members of a set are homogeneous.

Ideal system – a *conceptual* system with completely known initial conditions and predictable behavior.

Irregular – lacking uniformity in periodicity. *Irregular* may apply to be a systemic reoccurring event, and is not necessarily synonymous with randomness.

Marker – a non-causal variable, usually easily measured or observed, that is believed to co-vary with a causal variable, and is often difficult to measure or observe.

Model – an *ideal representation* of a *real system;* an *ideal* system.

Normal – a mathematical Gaussian function of a continuous probability distribution otherwise known as the "bell" curve, because its shape resembles the outline of a bell. (Also see "Gaussian," above.)

Occurrence – the happening in time of an event.

Order – a series of events with predictable or repetitive patterns.

Population – a *discrete* set whose members share some common *attribute*, or, in a study, are thought to share some common *attribute*.

Precision – the relative accuracy in real systems or real events in which an item is manufactured, a series of events executed, or a measurement made.

Prediction – the ability to predetermine the occurrence of an event by means of evaluating the initial or known conditions, relations, and variables.

Probability – a measure of the *opportunity* of occurrence of a specific event based on the number of times the specific event exists in a set of all events possible in the set.

Proxy – a measured variable, usually easily observed, that stands in place of an unmeasured variable, usually difficult to observe.

Random – a non-repetitive pattern that has a Gaussian or *probability* distribution (see *Randomness*).

Randomness – the property of a set or series of non-chaotic events wherein the events within the set occur without a repetitive pattern, and each event in the set has a specific probability of occurrence (usually an equal chance), but the timing or occurrence of a specific event in the set or series is unknown.

Real system – an *actua*l system, wherein observer control and or identification of all the contributing conditions, relationships, and variables are inherently incomplete.

Representation – the degree that a *model* or *ideal system* conforms to a *real system*, or that a *sample* conforms to a *population.*

Sample – a subset of a *population* that infers some degree of *representation* of the *population.*

Sample frame – a subset of the *population* from which the *sample* is drawn.

Spurious – unrelated, coincidental.

System – a set of related events. A system implies a degree of order and relationship among the events of the system set.

Uncertainty – the degree that the predictability of the occurrence of an event is unknown. In statistical analysis and probability, uncertainty of an *ideal system* is the calculated value of one minus the *probability* (e.g., if the *probability* of an event is 80%, then the *uncertainty* is 20%). However, a broader definition for a *real system* would include the degree that *existence* and *representation* are undetermined, which would cause the uncertainty to be much greater in real systems.

NON-RANDOM ACTORS

In an *ideal* statistical system *randomness* is the only cause of irregularity, and all else is regular and systemic. However, in the real world **there are many causes of irregularity that are not random**. Such irregularities may even be systemic. Some of the better known *non-random* actors, such as *chaos* and *catastrophe* and the concepts *order* and *disorder* are discussed below.

THE NATURE OF ORDER

Order implies predictability. The relationships and variables in an ordered set of events may be simple, such as a set of integers - each of which is one more than the previous (e.g., 1, 2, 3, 4….), or quite complex – such as harmonic modes of response for a slender elastic column buckling under a compressive load. Ordered sets may be discrete events – such as water drops falling on a surface, or continuous events, such as the deflection of a beam under a continuously increasing load. *Order* is often classified as to how we measure it, such as rank, class, size, color, or physical state; or how we might classify the relationships, such as discrete, continuous, arithmetic, geometric, transcendent, logarithmic, and so forth.

System. *Order* is synonymous with *system*. *Systems* have *variables*, *relationships*, and *events*. *Variables* are the *inputs* and *outcomes* of systems. The term *variable* may include system actors with zero or imperceptible change, as well as actors that change. Sometimes we can directly view the variables of a system, and sometimes they are difficult to discern. *Relationships* between variables are seldom obvious and require considerable effort to identify and model. Relationships are stated in the form of equations and formulas. *Events* are distinct happenings in a *system* that we can observe or discern. In an *ideal* system, *events* are associated with a *variable* as and *input* or an *output*, but in *real* systems, clearly demarked events may be difficult to discern, or to firmly establish an association with known variables.

Ideal system. An *ideal* system in one we construct in our head, on paper, or in a computer, purportedly representing a *real* system. *Ideal* systems are primarily a human construct, although a few other life forms exhibit some degree of reflection, contemplation, and strategizing in their interactions with their environment. The neophyte, emerging investigator-scientist must develop a keen appreciation for the profound difference that can exist between the idea, or *model*, in the human brain (or computer), and the *real* system. The potential for distortion and error to creep into the *model* is great. To appreciate the difference in *real* and *ideal* systems, one only has to imagine the tortuous path that a visual image takes to travel from the *real* system to the representation in the brain:

Light shines on the real thing from multiple sources, such as the sun, and from reflection and refraction off other surfaces, which may cause the occurrence of Moiré or other effects to be present in the light beams. Such light rays hit the surface of the real thing, where further reflection and/or refraction may introduce phase shifts and patterns into the light rays occluding the nature of the thing. From there the reflected light rays travel through the atmosphere, suffering some scatter and distortion along the way until the light ray reaches the eye(s) of the observer and is imperfectly focused onto a finite number of photo-receptor cells that convert the light energy into chemical-electrical impulses which travel through a neural network to a processing center in the brain where the chemical-electrical impulses are interpreted. All along the path to the brain lies considerable opportunity for signal interruption and distortion to occur from damaged or dysfunctional receptors *or nerves, the nature of which may not be detectable to, or correctable by, the interpreting center of the brain. Indeed, the interpreting center of the brain itself may be damaged from disease or accident, or otherwise dysfunctional through hardship, abuse, or even other misrepresentations in the brain, so that its attempts at interpretation and reconstruction introduce more distortion. Probably these distortion effects are present in all humans to some degree, and to some perhaps so much that complete misrepresentation occurs to the point it is fatal to the individual or disabling to the human group to which the observer belongs.*

Examples of ideal systems. Mathematics is an *ideal* construct. We impose absolute consistency and strict logical algorithms to it. In many cases it is a very accurate representation of reality, such as when we count discrete objects – for example individual humans. Let's say we can count the number of humans in a building: one, two three, … up to 15 people total at the time of counting. That is as representative of *reality* as it gets, but is still an *ideal* construct. Now perhaps we have four rooms in the building, and the individuals are dispersed rather uniformly throughout the building. So we now calculate with our mathematical *division* algorithm that there must be then an average of 15/4 or 3 and ¾ person per room. Now we are definitely into an *ideal* construct, as there is no such thing in *reality* as ¾ of a person. A person is either "1" or "0" (zero) – no fractions allowed. Does that mean that the concept of 3-¾ personas is invalid or useless? Not at all! The concept of a partial person is useful in describing *groups*, but meaningless in describing an *individual*. Statistics is method of describing a *group* and not an *individual* of the group. We will emphasis this point many times.

Real system. *Real* systems exist in reality, while *ideal* systems reside in our brain as an approximate representation of reality. *Real* systems are never completely static, although they may change very slowly – perhaps over millions of years, so we may represent them *ideally* as static without appreciable error. Other real systems are more observably dynamic, the change from input to output occurring over a time span that may range from nanoseconds to centuries.

Order may also be a subset of *disorder* - that is, within a disordered series of events, such as *randomness*, there may appear pseudo-orders, or even real patterns which are unrelated to the observed phenomena or system. It is partially from this principle that *spurious* events and relations appear in measuring *real* systems.

THE NATURE OF DISORDER

The cyclic nature the seasons and phases of the moon has been recognized for thousands of years. Most religions perceive an ordered universe, generally imagined to be far more orderly than it actually is. In some philosophical speculation or conjecture, *disorder* is believed to be the rule, yet because it is without discernable pattern, *disorder* is seldom recognized directly, and usually discovered in the unexpected disturbances of observed systems with expected behaviors. Most events we commonly think as *disordered* are actually *chaotic* systems. The evolution of our senses to detect the presence of events may be necessarily dependent on the systematic reoccurrence of events in order to serve a predictive function and enhance our survivability. True and complete *disorder*, then, may be largely invisible to us.

Disorder may also be a subset of order, i.e., within a set of ordered events may exist a disordered set. There are some important subsets of disorder: *chaos*, *catastrophe*, and *randomness*, which seem to be a blend of *order* and *disorder*, but often have distinctive and recognizable characteristics that impart a limited degree of predictability.

CHAOS

Chaos is a developing field in mathematics which deals with chaotic systems. Chaotic systems are extremely sensitive to initial conditions or small disturbances, but appear to have definite bounds or limits within which system events occur. Such systems are systemic and determinant in a sense that there are relationships and variables that determine outcome, but the initial conditions, and small perturbations that can occur may so drastically alter outcomes that precise event prediction in real systems may be impossible. Chaotic systems may have both regular and irregular periodicities. Some ideal chaotic systems display graphically in unexpected patterns that seem to revolve around "strange attractors" as if influenced by some invisible force.

It is thought that most real systems are *chaotic* to some degree. An example is weather patterns, such as rainfall. Many cyclic, systemic factors influence rainfall patterns, and the actual rainfall experienced is a result of the combination of such factors.

EXAMPLE:

To illustrate how the combination of a few simple cyclic relationships can result in extremely complex outcomes, consider the following two graphical plots which depicted a fictitious rainfall over a one-hundred-month period. Figure 1 is a plot of the original initial conditions, and plot 2 is a plot of a slight change in the initial conditions

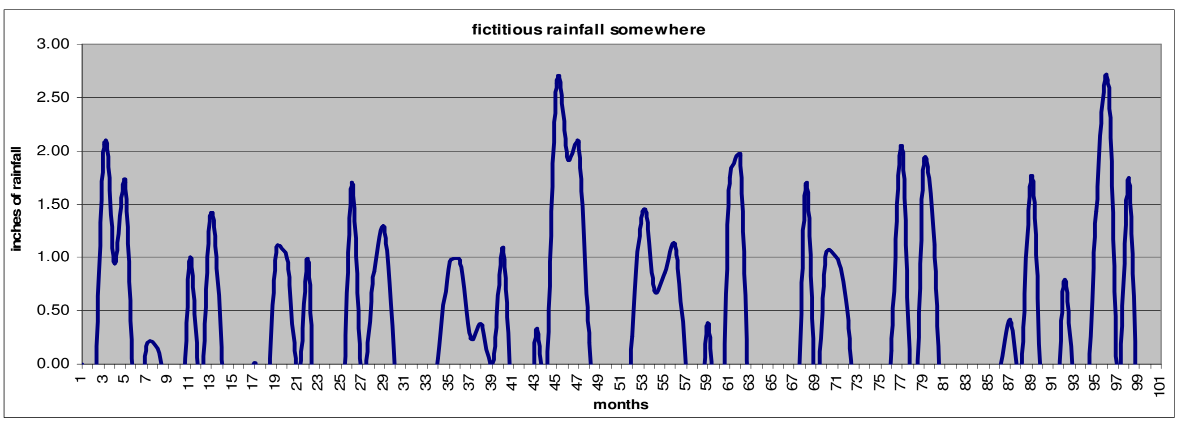

Figure 1. Fictitious rainfall, original initial conditions

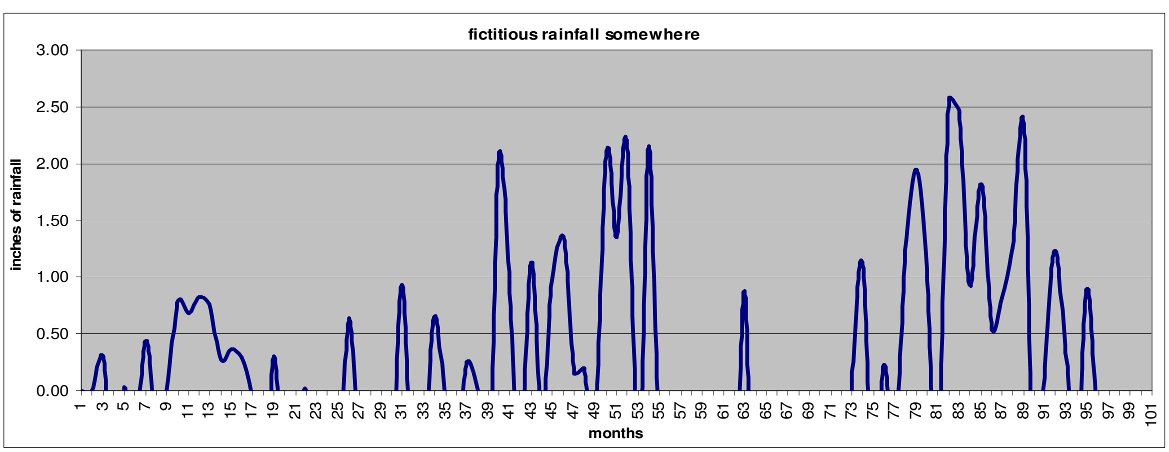

Figure 2. Fictitious Rainfall, one (of 6) initial condition changed 1%

An inspection of either of the above plots might lead us to conclude that the magnitude and timing of the rainfall might be *random* because of the irregularity of the timing, duration, and magnitudes of the rainfall. However, the truth is that the plot is a sum of three relatively simple trigonometric functions, and not random at all. There are only six initial input values. A change in only one of the six initial variables, from a value of 1.00 to 0.99 (a 1% change) makes a drastic change in the output, as plotted in the second graph, depicted in Figure 2.

This is an example of a *chaotic* system and how small changes in initial conditions drastically changes outcome. It is also an example that ***irregular* events do not necessarily imply *randomness*.**

The above fictitious example was an *ideal* system. In *real* systems, "initial" conditions may be input irregularly, or even randomly, so that precise outcome is impossible to predict (although bounds may be obvious.) The point in this example is that *chaos* is probably a significant actor in the real world, as much or more so than *randomness*, and the implications of that on statistical analysis of real systems, and the *ideal* assumption that all *irregularity* is *randomness*, may be profound, and is yet to be evaluated. ***Chaotic* events are irregular *systemic* events, and not *random* events**.

CATASTROPHE

Catastrophe is an emerging mathematical field that explores systems, including chaotic systems, where triggers or thresholds under certain conditions cause sudden and often unrecoverable or irreversible collapse of the system. An example of a *real* catastrophic system would be the buckling of a slender column under specific critical compressive loads. Current non-linear structural theory would predict an increasingly divergent lateral deflection over a very short loading change near the critical load, and at the buckling point a singularity (i.e., an infinite lateral deflection) at the critical load. The *ideal* conventional system would predict a recovery after the critical load was exceeded. In real systems, the slender column would collapse catastrophically - and not recover. Catastrophe theory would seek to form an ideal relation which mirrored the real systems inability to recover. Catastrophe is very common in the real world, and difficult to model in the ideal world.

RANDOMNESS

Randomness is a special mathematical concept that appears to have significant representation in real systems. *Random* events are those that occur in a *probabilistic* manner, wherein the occurrence of many such events tends to follow a Gaussian distribution. The classical Gaussian distribution is the so-called "normal," or bell curve so predominant in statistical problems. However there are many other probability distributions that exhibit randomness. *Randomness* is an extremely important concept and construct for the study and application of statistical procedures. It is important to realize that the concept of randomness is an <u>*ideal*</u> construct, and therefore only approximately representative of <u>*real*</u> systems.

A classic example of *randomness* is the distribution of heights or weights in adult men or women. Within any *homogeneous* population of humans of like attributes, such as race, ethnicity, or gender, the variation of weight or height among that population appears to be *randomly* distributed. The apparent occurrence of randomness in real populations allows a certain amount of characterization and predictability of the group, although any particular member may possess attributes at either extreme of the distribution curve, or anywhere in between. Therefore, **the predictability is for the group or set**, and not for a particular individual or member of the set.

In the following discussion, we will use as an example adult human height. Adult human height varies worldwide from 2^{-} feet to nearly 9 feet tall. Yet at the extremes, very few examples can be found, while at the mean, and within one standard deviation, the majority of the individuals fit within a relatively narrow range.

POPULATION

A population is a set of members, each member having, or believed to have one or more common attributes. Members may be animate beings or inanimate objects. In any statistical analysis, the *ideal* population is *homogeneous* in the attribute, with all variation in the attribute being caused either by *randomness*, or by a *systemic* cause. No other alternatives exist in the ideal statistical system. *Irregularities* caused by *chaotic* or *catastrophic* systemic cause will be interpreted as *random* events. In fact, the happenstance occurrence of a systemic-appearing subset of *random* events will be interpreted as *systemic*, when in reality it would be *spurious*.

In the *real* world, populations are seldom *homogeneous*, and random-appearing *irregularities* caused by *chaotic* or *catastrophic* causes may exist along with regular systemic causes. Moreover, systemic-appearing subsets of *random* events may occur in our measurements, giving rise to spuriousness. Such spuriousness would be in addition to that occasioned by the inclusion of variables unrelated to the population. Statistical procedures cannot discern spuriousness, and will interpret systemic irregularity as randomness.

Sub-populations. Within a larger population may exist groups with significant systemic differences in the measure of one or more attributes of the entire population. In extracting random samples from a population, the existence of significant sub-populations should be considered, as otherwise the calculation of statistical estimators for the larger population may be irrelevant and meaningless.

An example is the measurement of height in an adult population of humans. For a particular ethnic group, the *average* height of an adult human might be calculated as 5'-6". Yet, very few such adults may actually exist at that height. In reality two sub-groups exist – adult males, and adult females.

Humans display *gender dimorphism* – that is, males are typically taller than females of the same ethnic group by about six inches, while variation within one standard deviation for each gender is only about two inches (see example under Significance, below.)

Limits. In theory, a *random* distribution has an *infinite* number of members in the population with distributions ranging between minus infinity (- ∞) to plus infinity (+ ∞), or for some distributions, from zero to infinity. However, in *real* systems, random distributions have *finite* populations and finite limits to the distribution of variance. For example, in human adult height, a finite and positive height greater than zero always exists, and rarely, if ever, exceeds the upper bounds approaching nine feet.

Homogeneity. *Ideal* populations are *homogeneous* – i.e., the members all have identical attributes that vary only in *random* assignment. *Random* sampling of an ideal *homogeneous* population produces *unbiased* estimators of the population's characteristics. *Real* populations, however, are almost never *homogeneous*, containing sub-populations with significant systemic differences. The overall statistical characteristics of such real populations (with significant systemic differences among sub-populations) may be misleading or meaningless. In such cases, the statistical characteristics of each sub-population need to be investigated, and the sampling rate needs to be set so that each sub-population sample is of sufficient size.

SAMPLING

An important adjunct to *randomness* is the concept of random *sampling*. In real populations, measuring the characteristics of every individual entity within the population is difficult, impossible, or very expensive – so measurement of a *representative* sample is attempted.

Sample Frame. A *sample frame* is that portion of the total population from which a *sample* is actually extracted. Several different factors exist in the real world to cause a sample frame to be significantly different from the total population. In addition, the cost of sampling may be great enough that the investigator or researcher deliberately limits the scope of the investigation. When such occurs, as is most often, then any **generalization of the sample characteristics can only be legitimately done for the *sample frame***, and not the total population.

Most *real* populations are not homogeneous - so much so, in fact, that some sub-populations are deliberately, or de facto, excluded from consideration for sampling. For example, any researcher bent on extracting a random sample from a population of humans would first need a listing of the entire population. Such a listing may not exist. Two sources may be useful – a census or a telephone listing. However, many countries do not have accurate census information, and in many countries, telephones are a luxury item of the few and privileged. Moreover, human populations are somewhat mobile, and by the time the researcher has developed a comprehensive list, some of the members listed may have moved or become unavailable. Even in very affluent countries, such as the US, not every member of the population has a telephone or is listed in the directory. If random participants were drawn from the names in the telephone book, very significant portions of the population would be excluded – those without telephones, and those persons with unlisted numbers. For example, it is unlikely that significant numbers of teenagers and children would have listings in the telephone directory, nor would prisoners, mentally-challenged adults, itinerants, visitors, or the long-term unemployed. Compounding the problem of random sample extraction in this example is that the affluent may have several phones, while those at or below the poverty line may have none.

Bias. The characteristics of a sample can be an *unbiased estimator* of the population characteristics only if it is a *randomly*-selected sample from a *homogeneous* population. For samples not selected randomly, and/or where the population is not homogenous, the sample characteristics may be severely *biased*. There are several sources of sample bias, some of which are listed below:

1) by the investigator deliberately or unknowingly excluding listing sources,
2) by deliberately or unknowingly excluding sub-populations,
3) by non-responsive sample members (self-selection bias),
4) by erroneous measurements,
5) by an imperfect randomizing process,
6) by the investigation or measurement methodology employed, and
7) by using indirect measurement or the use of proxies or markers.

There is considerable speculation that **true random selection of sample members is impossible in real systems**, and some investigator, process and/or self-selection bias always exists. It is therefore incumbent on the researcher to adopt procedures which minimize sample bias, and to be aware that the sample estimators are never certain. Despite the difficulty to obtain completely random sampling, the investigator should strive to the maximum degree possible to minimize the degree of *bias* in the sample estimators.

SIGNIFICANCE

Another important adjunct to randomness is the concept of *significance*. ***Significance* is a measure of the probability that an event is caused by *random chance***, and not by some ordered system. For example, if human adult weight or height is separated into males and females of a particular race or ethnic group, then gender difference will be obvious – human adult males, as a group, are heavier and taller than human adult females as a group. This group characteristic difference, however must be evaluated as to whether there is a *systematic* (i.e., ordered) cause, or whether the difference is due to *random chance*. If we estimate the difference to be caused by random chance, then we would conclude that there is no *significant* difference. However if we estimate that the difference is due to a *systematic* or *ordered* cause, then we conclude that the differences are *significant*. Significance will always remain a fuzzy concept, however, because **in *real* systems we are never completely certain of our estimate or decision**, and generally always assign a *probability* or *confidence level* to our decision. A general rule of thumb is to estimate *significance* to a 95% confidence level or higher. That means that there would remain a 5% chance or probability that the differences are not significant. It would be quite erroneous to assume the 5% probability of error will never happen. In fact, it might happen to you on the very first trial. That is why **repeated trials are required to assure ourselves that we have indeed differentiated between systemic or random factors**.

EXAMPLE

Ideal System. In this example we apply the actual measured parameters of the population of Great Britain to a normal Gaussian distribution. The actual data may have a skewed distribution, which we will ignore for the moment. In the initial survey measuring the heights of individuals, we might have a plot of the data that looks like that in figure 3, below.

Note: the frequency associated with each height depends on the included interval of each measurement (e.g., for the height of 5.0 feet, the interval in this example is from 4.95^+ *to 5.05.) For very small intervals approaching zero (e.g., exactly 5.000000 feet) the frequency would approach zero, as very few adults would measure exactly that height.*

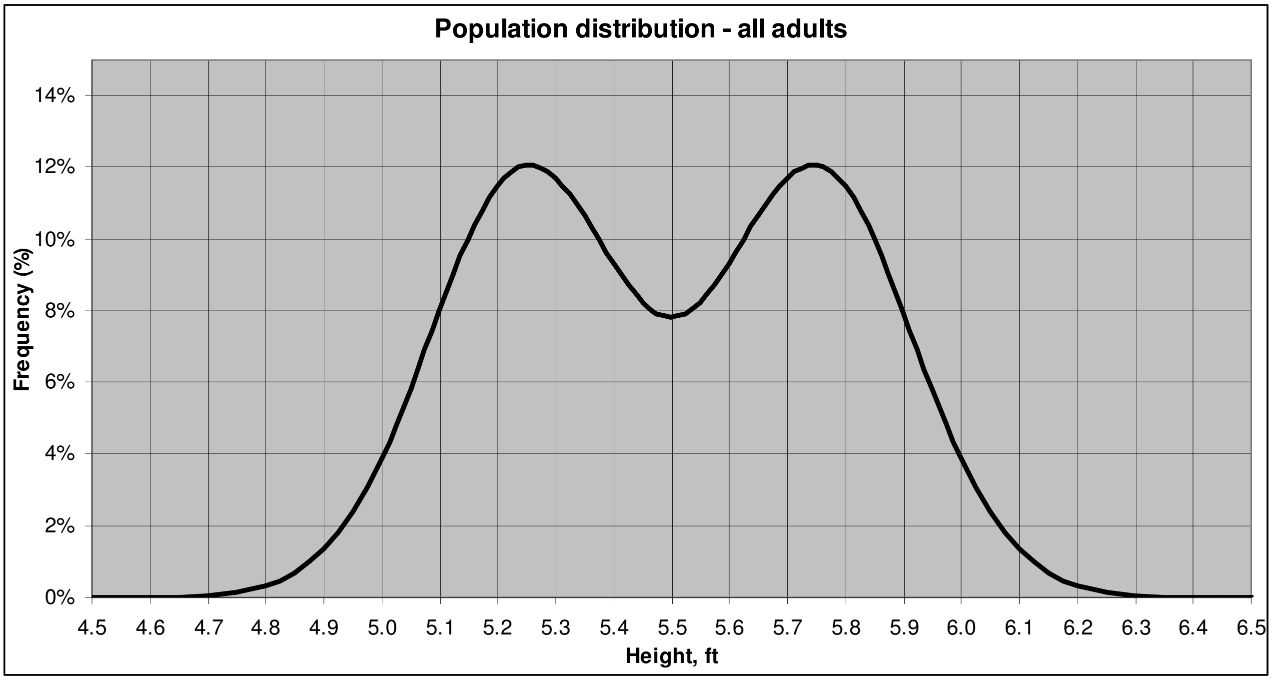

Figure 3. Distribution of height in a homogeneous adult British population

From Figure 3 we can view the mean (average) height as 5'-6" (5.5 ft), with a standard deviation of 5". However, the existence of two peaks in the distribution of adult heights in the above plot should cause us to doubt that our population above is randomly distributed, and wonder if perhaps there is a *systemic* cause for the two peaks. On a hunch, we segregate our data into males and females, and re-plot the distributions, depicted in Figure 4

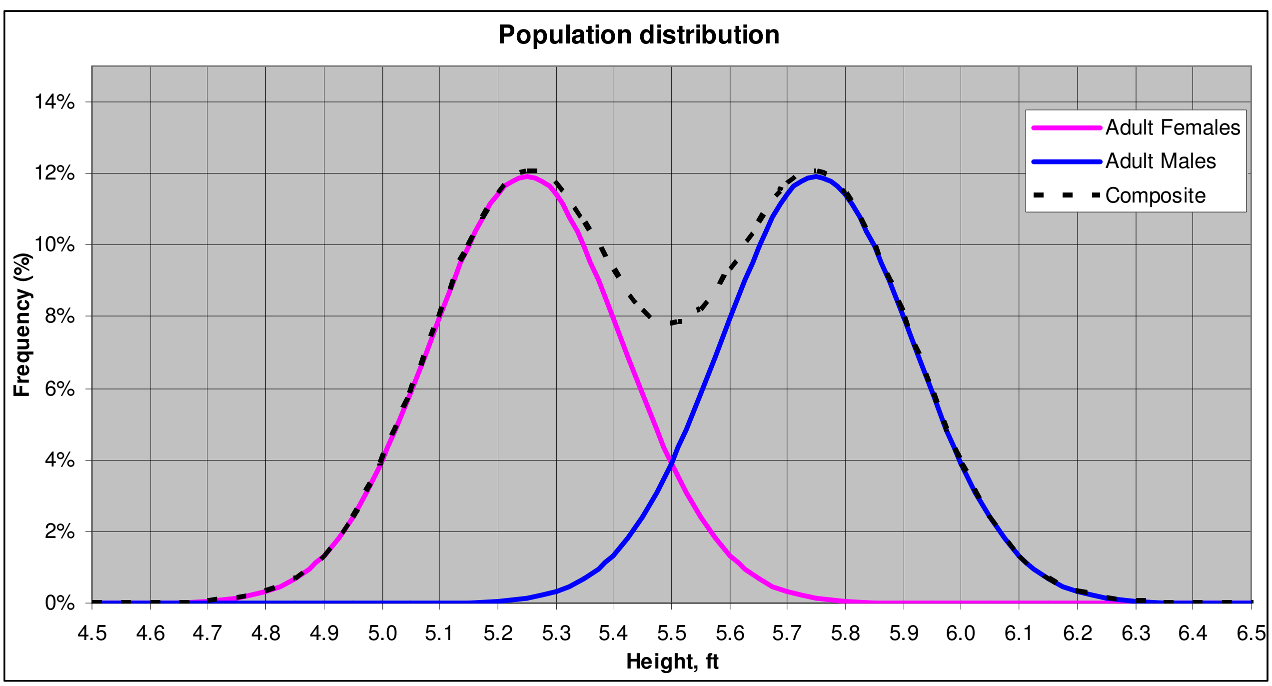

Figure 4. Distribution of height by gender in a homogeneous adult British population

The two separate bell-shaped random distribution curves in Figure 4 overlap each other. The pink curve for adult females has a mean (average) height of 5’-3” (5.25 ft), while the blue curve for adult males has a mean (average) height of 5’-9” (5.75 ft). The standard deviation for either gender is 2” (0.167 ft). There is a 6” (0.5 ft) difference between the average heights of males and females. The separation of the measurement data into males (blue-colored curve) and females (pink-colored curve) explains the two peaks. For this specific example the explanation is complete because our ideal assumptions allowed no other possibility. Such explanation is a *systemic* cause, and not a random factor, and we were able to separate the systemic cause from randomness because of the perfect distribution of randomness in our ideal example.

Also note that the mean (average) height 5’-6” and standard deviation of about 5” in the composite distribution of all adults in Figure 3 is largely meaningless. We can rightly conclude that before we can meaningfully establish population descriptors, we should account for significant systemic explanations first.

Real system. In the real world we would not obtain the perfect distribution depicted in figures 3, 4, and 5, above, which, to the extent that they are perfect, are contrived (we contrived them so as to better illustrate the difference between randomness and systemic cause.) Instead, we would find that the measure of randomness in any actual finite population would deviate from perfect randomness by a considerable margin. That is the real nature of randomness – occurrences are quite irregular – and there is no guaranty that greater probabilities will occur before lesser probabilities, or be fairly distributed. The question we would have to ask ourselves about our finite measurements would be how distorted can the results be and still be considered random? Consider the two random samples below, in Figures 5 and 6.

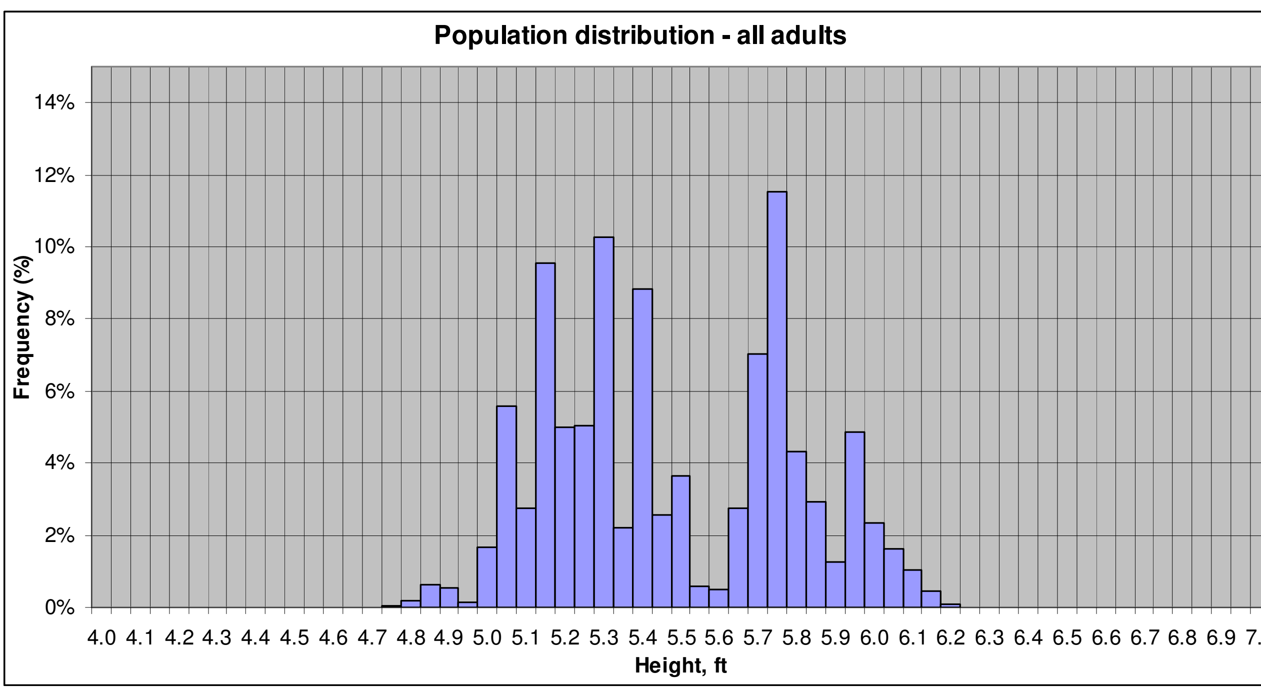

Figure 5. Sample 1, British Population Distribution in a Random Sample of N=1,000

Figure 5, above, is a random sample of 1000 persons (both sexes). This is typical of a fairly representative distribution, and the two modes appear relatively well defined. However, consider the distribution in Figure 6, below, which is another random sample of 1,000 person (both sexes). This time, in Figure 6, one might be hard pressed to recognize the existence of two modes. (*However, a good statistician would investigate gender dimorphism automatically, even if the initial data did not indicate any differences.*)

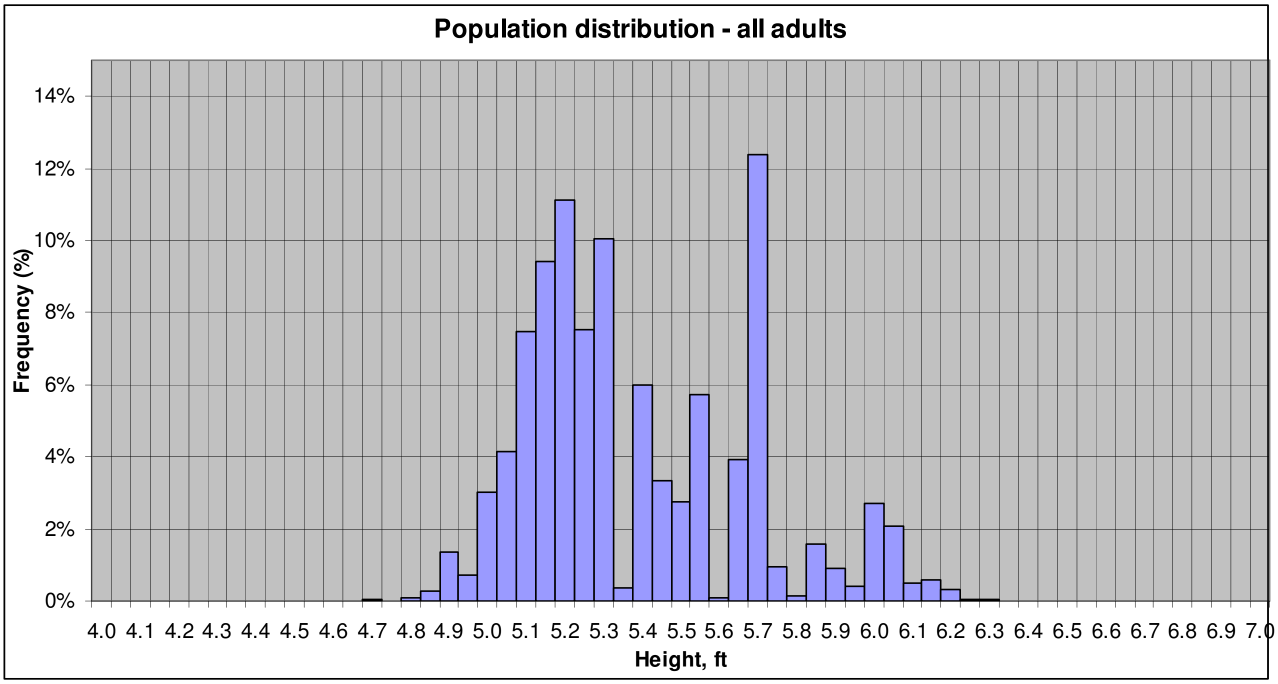

Figure 6. Sample 2, British Population Distribution in a Random Sample of N=1,000

With these real-world examples, we can now consider a few important pointers to note about actual population sampling.

1. Real populations are generally not homogeneous, which adds to the uncertainty of the calculated sample population estimators;
2. Repeated random samples (as in Figures 5 and 6) may vary considerably from each other, adding more uncertainty to the accuracy of the sample population estimators. (However, we may be able to decrease the uncertainty by combing separate random samples into a single larger sample group.)

Note: The population of Great Britain is about 60 million persons. A random sample of 1,000 persons would theoretically be expected to yield fairly representative population estimators. The only way to establish this, however, is to repeatedly sample the population, and measure the distortions from sample to sample.

So the answer to our question about how much distortion can appear in a random sample is that the distortion can be considerable, which greatly complicates our investigation into systemic cause. The irregular variation due to randomness can easily mask systemic, non-random factors, especially irregular systemic ones such as chaos. Further complicating the situation, regular-appearing data can exist as a subset of randomness, which can mislead us into seeing a systemic cause where there is none.

Thus it is that we make a great unsubstantiated leap of faith when we measure a real population or a sample, and assign all irregularity to randomness, and all regularity to systemic cause. Part of the error is caused by having finite, non-homogeneous populations, but more often it is due to the unpredictable nature of randomness. This is but one of many reasons why **conclusions reached statistically can never be absolutely certain**.

Note: Theoretically, a random distribution should become smoother and more symmetric as the number of members in the set become larger, approaching perfection as N approach infinity. Practically, however, even in populations numbering in the tens of millions, considerable deviation from smoothness and perfection occurs, and should be expected.

MODES AND DIMORPHISM

Dimorphism. The *bi-modal* shape displayed in Figure 3, above, for our ideal distribution is also known as *dimorphism* ('*di*' – meaning 'two,' and '*morph*' – meaning 'shape'.) *Gender dimorphism*, also called *sexual dimorphism*, is a hot and often controversial topic in sociological circles and among radical feminists groups who rally to obliterate any suggestion that gender inherently indicates any performance advantage for men. Of course, much of the radical feminist ranting is nonsense – as a cursory review of almost every human *physical* characteristic indicate a significant performance difference between gender. Males are as a group taller, heavier, stronger, faster, more aggressive, and suffer less injuries in similar sports activities than do females as a group.

Before any male crows too loud about his being physically superior as an individual, he might consider two things:

1) There is a high probability that <u>some</u> females will exist who can best him, and

2) The majority of men devote all their physical advantage just to provide sustenance, shelter and service to their female mate and offspring.

(Note: There is some suspicion among males that females possess a power and process not easily recognized by males or measured by sociologist to subjugate a male without his realizing it so that all his efforts are for her benefit. Naturally, the females aren't talking. We will talk more this and about other nuances causing difficulty in measuring the less-obvious human traits.)

But, not every human endeavor requires extreme physical exertion. Socialization, communication, and domestication skills are also necessary for a functional society. Many of these non-physical skills are difficult to define and measure, or have yet to be measured, although some studies indicate women are clearly superior in many such areas. So not every human attribute is slanted in favor of males, and many may have no gender-influenced performance bias.

The overview of the gender performance dimorphism issue is that overall, the total score for men and women are probably quite even, but on specific attributes, gender is a *systemic* factor. The difficulty facing the sociologist, educational psychologists, and other gender-related research is that the easy to measure attributes, (primarily the physical ones) generally favor males, while the attributes that would probably favor females are difficult ones to identify and measure, being largely invisible to male investigators, and, strangely enough, many so-called "liberated women" apparently fail to distinguish their natural advantages.

Modes. Modal issues may also be present in engineering studies, such as traffic counts. Daily traffic flow might be expected to peak in the early morning from the working population trudging (or driving) off to work, and then peaking again at the end of the day as the workers drive home. More than two modes may be present in any distribution. It is important to recognize that such **modes are**

a strong indicator of a systemic cause, and should be investigated. In the gender example on height, the two modes were obvious; however, the presence of modes is not always so obvious. Any time the distribution curve deviates significantly from a Gaussian or probability distribution, systemic cause should be evaluated. This is so because, **the Gaussian distribution itself is the definition of randomness** (and as we demonstrated <u>in an ideal system</u> it will be *perfectly* random.)

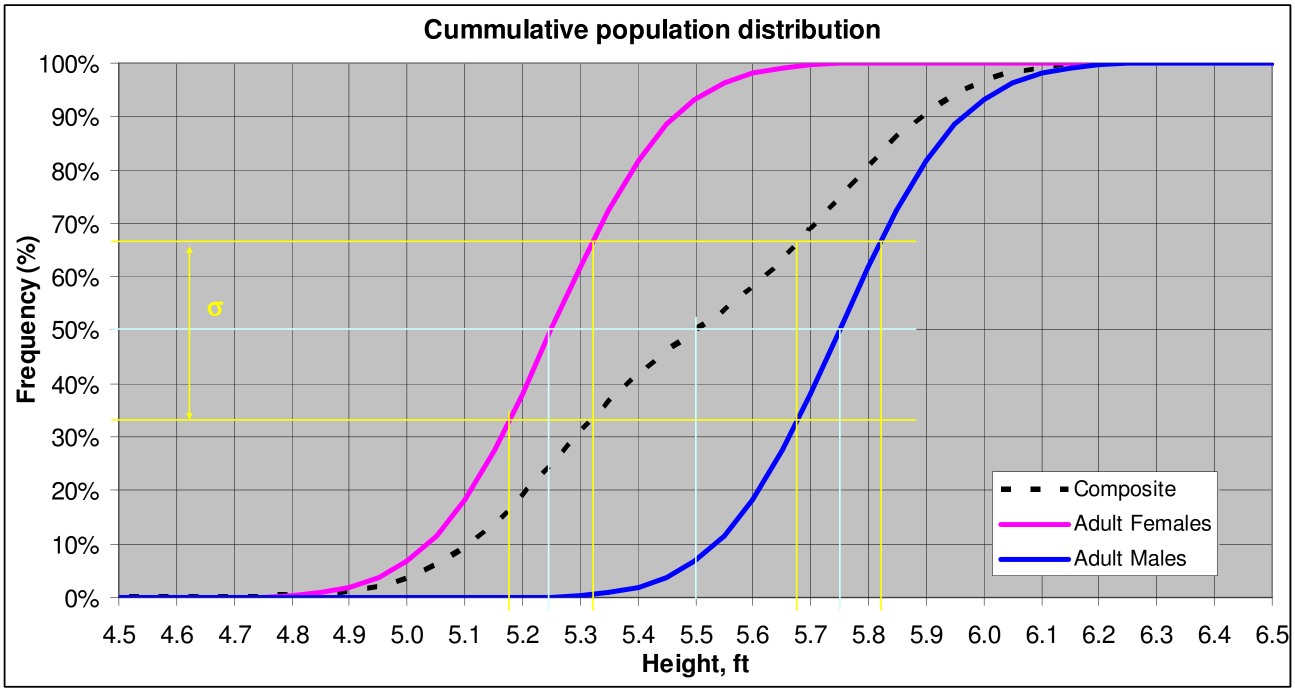

Figure 7. Distribution of Height by Gender – cumulative population distribution

Deriving information from the cumulative population distribution graph. In the graph depicted in Figure 7, above, of *cumulative population distribution*, which is a summation of the *population distribution c*urves shown earlier in Figures 3 and 4, we can visually observe and determine the percentage of males or females of a given height. The *mean* of the population of each curve lies directly under the intersection of each curve with the 50% frequency line. Thus we can visually determine that males as a group are on the average six inches taller than females (for the British population). The mean height shown for males is 5.75 feet (5'-9"), while the mean height for females is 5.25 feet (5'-3"). The difference between those two means is 0.5 feet (6").

We can also determine the mean heights in the population distribution graph (Figure 4) by observing the height corresponding to the two peaks of the graph for men and women separately (the pink line and the blue line) because each distribution curve is symmetrical in our ideal system. However, if the distribution was not symmetrical, the *population distribution* could not be used. The *cumulative distribution* graph can always be used to find the *mean.*

Also, knowing that *one standard deviation*, **σ**, always occurs between 31.73% and 68.27% cumulative frequencies, we can determine standard deviation for each curve graphically, as it is the differences of the heights under the intersections of the curves at those frequencies (the yellow lines in Figure 7). Finally, the frequency values expressed as a percentage is equal to the probability of occurrence.

Other systemic factors. In our example of human adult height, actual statistical evaluations have determined the height differences of male and female populations to be *significant*, i.e., due to a systematic cause, and not due to random chance. The systematic cause would be *gender*, and whatever causal system determines gender in humans. Such a finding for humans appears to be true for all human races and ethnic groups. In our *ideal* British example, we have provided no other perturbations to disturb the randomness of the distribution curve. In the *real* world, however, the females of some ethnic groups are taller on the average than the males of other ethnic groups. For example, both males and females of the Watutsi tribe of Africa are *significantly* taller on the average than Caucasian or Asian males. In contrast, African pygmy males and female are significantly shorter on the average than Caucasian or Asian males. Thus we would be aware that gender may not the only systematic cause of height and weight differences in humans, but that ethnicity or race could also be a factor.

Individual versus group characterization. It is important to note again that issues of classification, such as human adult height and weight by gender, are *group* characterizations, and not necessarily *individual* characterizations. In Figure 7, we can determine the relative frequency of males and females for any given height. For any particular height, say 6'-0", we observe that less than 0.00034% of the females reach or exceed this height (we can't actually read that fine on the graph displayed, because the graph is plotted to coarsely, but the tabulation in Appendix A yields that figure), while about 6.68% of the males reach or exceed 6'-0". The *ratio* of 6' males to 6' females is then about 19,645 to 1 for this particular example. From this not-so-far-fetched example, we can derive a few valuable insights:

1) Even at 6'there are still <u>some</u> women who are taller than a vast majority of men; and
2) In competition where tallness is a legitimate competitive advantage (e.g., basketball), the ratio of successful women competitors to successful men competitors in mixed-gender activities will be small, at the same ratio (almost 20,000 to 1 in our example), all other factors held constant.

The first insight should provide female aspirants with the knowledge that the a few individual women equaling or exceeding the selection criteria ($\geq$6'-0" in our example) may be able to compete with men successfully in those areas where their personal attributes equal or exceed the selection criteria.

The second insight should provide women with the knowledge that mixed gender competition dependent only on a tallness advantage (e.g., basketball) will of natural course tend be dominated by men, competition will be very stiff, and only a very few women will have a realistic chance to make the grade (for the basketball example of 6', only 1 out of 20,000).

The third insight is that where teams are selected independently, such as franchise with the five primary players, and perhaps ten backup players, <u>no</u> women would b likely to be selected, as the 15 tallest candidates are likely to all be men. A 6' female from this example population desiring to compete for a position would theoretically be faced with almost 20,000 taller males as competition for the position, and thus highly unlikely to be selected to a 15-man franchise.

Note: This basketball example is somewhat contrived, because tallness would not be the only selection criteria for basketball, but also spatial orientation (shooting baskets in 3-dimensional space) and aggressiveness. All these traits are male-dominated, and thus the odds of a female selection based only on these traits (i.e., no gender-equality laws altering the selection process) will be trebly magnified. That is, instead of nearly a 20,000 to 1 odds against a female selection, the odds *would be more likely 1000,000,000,000 to 1, or approximately never. This in fact is what happens (or, more correctly, doesn't happen) in professional sports. There is no male-female competition in any professional competition, and none in the nominally amateur competition of the Olympics, except for mandated mixed-double teams in tennis.*

OCCURRENCE. The very smooth-appearing Gaussian distribution of *random* variance depicted in Figures 3 and 4 in the example of height distribution in adult British males or females should not be taken to imply that random events occur in any regular pattern in time. For example, rainfall intensity patterns appear to be random (more likely they are chaotic), but in the prediction of the probability of occurrence, say of a 20-year storm, we cannot imply that such a storm will occur once every twenty years or so. In fact such occurrence would not be random but systematic. Rather, we imply that over a rather long period, say 100 years, the 20-year storm will occur 5 times. Now, those five times may be sometimes back to back storms, may occur in separate years, or may all occur in the same year. In any case, **if the events are truly random, we will not be able to predict <u>when</u> they will occur**. Indeed, if we had a thousand year period of data to study, we might find some 100-year sub-periods where no 20-year storms occurred, and other 100-year sub-periods where more than five 20-year storms occurred. But in the thousand-year period, the 20 year storm would be defined by that storm intensity that occurred 50 time ($1000 \div 20 = 50$).

SPURIOUSNESS

Spurious events, also called *confounding* factors, are those that occur coincidentally in our observations and measurements, but in fact are unrelated to the system being observed or measured.

Spuriousness can arise from several sources. In the Introduction we alluded to the possibility that *order* may haphazardly appear as a subset of *disorder*. An example would be the appearance of an *apparently* systemic factor in a set of events. If in fact the factor appeared repeatedly in several trials, then the factor may indeed to systemic. However, if the factor never appears again, then we may conclude it was spurious, perhaps arising as a randomly occurring temporarily ordered-looking subset.

Spuriousness also occurs in studies in which unrelated variables are included for measurement. In true scientific experiments, where variables are manipulated, spurious variables can be identified and removed, because manipulating them will not cause a change in outcome or a change in other variables. But in quasi-experimental investigations, where variables are not manipulated, spurious variables can not be easily identified. Thus, for quasi-experimental designs, it is not a good practice to "measure everything" and let the statistical process "sort it out," because mathematically the statistical process cannot differentiate between a spurious event and a related event- and will not be able to "sort it out" correctly.

Spurious variables are a very serious factor in invalidating studies. It is important for the researcher to address spuriousness and how it is controlled. The existence of **spurious variables can give rise to alternate explanations to hypothesis, and thus invalidate any conclusions**.

EXPERIMENTAL DESIGN

In a true experimental scientific exploration or study, the variables are manipulated. Such manipulation is the best way to eliminate spurious variables, and to observe causal relations. In the best of worlds, the actual causal variables are directly manipulated and the results directly observed while all other factors are held constant. Repeating such an experiment several times allows spurious occurrences and spurious variables to be identified and removed from the results. Spurious occurrences will not be repeated in subsequent trials, and spurious variables, when manipulated, will not cause systemic variation in the dependent variable. In this 'best-scenario' case, the contributing variables are *independent*, meaning that they can be held constant (i.e., unchanging) while each one of the variables is manipulated separately, thus facilitating the measurement of the manipulated variable and the measurement of the effect on the dependent variable, and establishing the magnitude, direction, and type of relationship of each variable. While some uncertainty may still exist, even in very carefully controlled experiments, the level of uncertainty is very small, and for all practical purposes, nonexistent.

More often than not one or more of the following conditions persist in an experiment:

1) the variables available to be measured are not independent,
2) significant and important causal variables are omitted,
3) the variables available to be measured are proxies or markers, and/or
4) the observations are *indirect*.

Independence. Independent variables are desirable to effectively establish relationships in experimental trials. Employing variables that co-vary in a model introduces confusion in the interpretation of results, and to a certain degree is tantamount to measuring the same thing twice. For example, if human *height* is measured for a standing person from the barefoot to the top of the head, and we again measure the *length* of the person laying down from the sole of the barefoot to the crown of the skull, we may get a different measure, but it will be very close to the standing measure, and the *length* so measured can be expected to strongly co-vary with the *height*. One is not a cause of the other, but one *is* the other, in a different light. Employing both measures in a model would serve little purpose. Rather the researcher should choose one and forgo the other as redundant.

Of course, nothing is absolute. One could conceive of an experiment as to the compressibility of the human spine and joints under the influence of gravity or acceleration, in which case the standing and prone measurements would be relevant (actually, the difference between them.)

Omission. Often it is difficult or unwieldy to include all independent variables in the final model intended for application, as many variables may have only a small or even negligible contribution in practical applications. However, the researcher needs to establish that such is in fact the case, and not depend solely on logic or his model to exclude variables. Despite the investigator's best attempt, important variables often go missing because the phenomenon under investigation is poorly understood. In such cases, the study should be labeled as "exploratory."

Indirect. Often *direct* measurement in an investigation is not possible, and *indirect* measurements are necessary. Indirect measurement includes the use of proxies and markers as substitutes for actual variables. In such cases it is more difficult to be certain that the indirect measurement is meaningful. And, in such case, spurious events may be harder to detect

Proxies and markers. Many phenomena do not readily reveal their causal factors to the human investigator. Out of necessity, and sometimes for convenience, a *proxy* variable is measured which is thought (or hoped) to fairly represent the actual causal variable. A *marker* is a variable that is measured because it is thought to be strongly associated with a causal variable. In many cases we will not know for certain whether a variable measured is a direct causal factor or a proxy or a marker. In those cases, where prediction only is desired, such proxies and markers may be sufficient. But in those cases where change is desired, the deliberate or inadvertent employment of proxies and markers may yield disappointing and contrary results.

Control groups. A control group is one that is not manipulated, and serves as a base to compare against a manipulated group. Control groups greatly enhance the credibility and certainty of a study. When control groups are comprised of humans, measurement error can be exacerbated if the subjects know they are in the control group, and may report anomalous and misleading results, either consciously or subconsciously.

Blind and double-blind studies. To eliminate the possibility of erroneous subject feedback, the blind study does not allow the participants to know whether they are in the control group or in the manipulated group. A common example of a blind study is the study of the efficacy of a medicine for an illness, or for side effects, where some of the participants are given the medicine, and others are given a placebo, but neither can tell which is which. In a double blind study, the administers of the test also do not know who is getting what, and thus cannot inadvertently clue the participant.

QUASI-EXPERIMENTAL STUDIES

In *quasi-experimental* studies, the variables are not manipulated. An investigator may not be able to manipulate variables for several reasons: it may be impossible, difficult, illegal, unethical, immoral, or perhaps the relational variables have not been identified sufficiently to even know what to measure. Most sociological and psychological studies involving human input or outcome fall into this category, and the method is sometimes used when more experimental methods could be employed were it not for legal, ethical or moral considerations.

There is some feeling that quasi-experimental designs are not truly scientific. Indeed, there is too often an over-reliance on statistical processing, particularly the use of the mathematical processing of *partial correlation coefficients*, to arrive at questionably "independent" correlation coefficients. Despite these shortcomings, however, in many cases, there are no other acceptable alternatives to assess the nature of the relationships that may exist. So a quasi-experimental design is employed out of necessity. Quasi-experimental designs suffer all the vagaries of experimental designs, plus additional vagaries peculiar to quasi-experimental approach.

The investigator employing quasi-experimental methodology should be fully cognizant that the certainty of the study results is adversely affected thereby. Whenever possible, the study design should employ techniques to reduce the inherent uncertainty of the quasi-experimental approach. The following techniques should be incorporated as much as possible.

1) Develop a keen familiarity and insight of the phenomenon to be investigated. Consult many specialists to get their insight on agreement and on controversies in that area.
2) Develop a well-defined model that includes all significant variables and allows direct identification of spurious variables whenever possible.
3) Directly eliminate spurious variables and their measurements as input whenever possible. The elimination of spuriousness from a statistical study is one of the requisites for inferring causality.
4) Conduct many studies from many angles, trying to isolate individual variables to determine their nature and relevance.
5) Refine the model and start over again when investigation results are inconclusive.

Cross-sectional studies. Cross-sectional studies are those that gather data at one point in time. Most quasi-experimental studies fall into this category. As such, there is no way for the investigator of quasi-experimental study to determine whether the thing measured is constant or varies with respect to time.

Repeated cross-sectional studies. One method to try to gain appreciation of a variable over time is to make repeated cross-sectional studies at subsequent time intervals. Such studies, however, have the inherent flaw that the subjects, or members of the study set, are different for each new cross-section. This discontinuity in set membership introduces the possible alternative explanation that any differences might be due to the change of members, and not the change in time.

Longitudinal studies. To compensate for the inherent flaw in repeated cross-sectional design, the *longitudinal* study keeps the same members and measures them repeatedly at subsequent time intervals. The biggest problem facing longitudinal design, especially long-term designs over a period of years, is that the members leave the study, become unavailable, die, or otherwise do not participate in all the measurements. An additional problem with this design is that it is usually very expensive, and the number of participating individuals small in number. Even for the few well-funded studies with a large number of participants, the issues of spuriousness and/or omission of significant variables are not addressed by the longitudinal design alone.

Partial coefficients. In statistical analysis, most obvious in regression analysis, a common problem is that the variables measured are not truly independent, and too often the supposedly independent variables co-vary with one another. This happens most often because a model of the system is poorly defined, and/or the truly independent variables are not easy to define and/or difficult to measure, so the available or easy-to-measure variables (which may not be independent, and may even be spurious) are measured. Such covariance distorts the true description of the system, and a means to remove the covariance is desirable. The best method to remove co-variance between "independent" variables is to develop or refine the model so that only truly independent variables are measured. However, this is not always possible, especially in sociological models.

In a true and properly controlled experiment, each variable would be manipulated separately with all other variables held constant. This process would be repeated for each variable. The result is of such a controlled experiment is a set of *partial* descriptors, wherein *partial* takes on a mathematical definition – meaning that it is the descriptor relating the independent and dependent variable with all other variables held constant. If in fact none of the other alleged 'independent' variables co-vary during the manipulation of the target variable, then the variables are truly independent, and the model is well specified, at least as in regard to the included variables.

In studies of relationships between physical inanimate objects and systems, such as in the "hard" sciences and engineering, it is in most cases actually possible to hold conditions fairly constant and manipulate one variable at a time, and see what changes occur in the system due to such manipulation. If in such experiment the variation between repeated trials is very small, and the strength of the relation near unity, then a causal relationship can be strongly inferred between system variables. In many such relations, the designation of which variable is dependent is immaterial and interchangeable.

There are some important considerations for variables in experiments seeking to establish causal relationships between variables:

1) the model of the system should employ truly independent variables,

2) all important independent variables should be accounted for, and

3) all spurious variables must be eliminated.

Spurious variables are also called *confounding* factors for good reason. Their presence can completely invalidate any experiment or quasi-experiment if their effects are not identified and removed from consideration. A researcher may not know initially if a variable is spurious or not in a single experiment, but repeated experiments should show the variable to be erratic, indicating spuriousness. Spuriousness is particularly hard to identify in quasi-experimental designs common in sociological and psychological investigations. The best defense against spuriousness is a well-defined model and well specified experiment.

In many statistical studies, particularly sociological, manipulation of variables is difficult, impossible, illegal or unethical. Such investigations are called quasi-experimental because variables are not manipulated, and spuriousness is too often not directly controlled. Instead the unaware researcher goes about measuring all that can be measured, whether related or not, and attempts to filter out the relevant from the irrelevant by a process of calculating partial coefficients, purporting to remove the effect of the covariance between not-so-independent variables. This process has been resoundingly criticized as just as apt to introduce additional distortion as to remove it. Nevertheless, it remains widely practiced, and for no other reason alone, such studies should be viewed with more than the usual grain of salt. If spurious variable are present, the effect on partial coefficients may be great enough to severely distort and misrepresent the true influence.

BIAS

Bias is usually a *systemic* actor on otherwise random events. A classic example is a loaded dice, wherein a weight is placed in the dice so as to cause one face to be favored to show when the dice is rolled. In some loaded dice, the weight is movable, so that by tapping the dice in the right way, the weight will slide to the opposite side, or be neutral. Even with a fair, unbiased dice, some individuals are sufficiently adept at rolling the dice to control to some measure the surfacing face. Such control is particularly evident in flipping a coin. Most individuals, with some practice, can learn to flip a coin so that one side habitually lands face up.

Bias may also be a natural subset of randomness. In the totality of possibilities, the ordered subset and the biased subset exist. The probability may be low, but the occurrence sooner or later is a certainty. Such occurrences might be termed a ‘phantom’ actor. In an experimental design, with repeated manipulations, we will be able to identify the phantom actor and eliminate it, as it is unlikely to be repeated. But in quasi-experimental designs, such phantom actors cannot be identified with any certainty, if at all, and remain a source of spuriousness.

Humans that we are - who employ statistical methodology, our own perverse nature can induce a bias in the investigation. Humans are subject to justifying presupposed outcomes, having hidden agendas, ignoring the facts before us, deleting measurements that don’t support our preconceived answers, and suppressing findings contrary to our economic interests. Even reputable scientists are occasionally unmasked as having compromised the purity of an investigation, and become tainted by revelation of a vested interest in a particular outcome.

We should look with a natural suspicion on outcomes of investigation made by, or funded by, those entities with a vested interest in the outcome. A classic example is cigarette company researches into the dangers of smoking. New court trials and expert witnesses reveal that adversarial findings were systematically suppressed by the cigarette companies by non-disclosure clauses in the contracts for research. But cigarette companies are not the only entities with vested interests. So do politicians, and advocacy groups. Not a few such groups engage in fabricated studies and deliberately skewed results to support their cause.

REPRESENTATION, APPROXIMATION AND PRECISION

In the *idea*l world, everything is certain and defined precisely, and all measurements are exact. Not so in the real world. **Every *ideal* representation is an *approximation* of the real world**. Every ideal or empirical formula and equation is only an approximate representation of the real-world system or phenomenon described. Every measurement we make in the real world is imprecise. Every observation is incomplete and to some extent distorted.

Despite this rather negative revelation, we humans go about making our approximate models, and simplifying assumptions without much concern for imperfection. And in the great majority of cases, we can predict well enough to be useful. And if our processes become too inaccurate, why, we'll just refine them enough to be useful again. In fact, this process of progressive refinement is probably our only avenue to keep the gap between our ideal representations and the real world to a manageable level.

But let not the investigator be blasé about it. There are today groups of humans who live and function much as their ancestors did thousands of years ago, even prehistoric. And they may continue to do so. They are not in charge of their destinies, and make no contribution to progress. Indeed, it is believed by some that only a small percentage of humans in any generation actually contribute to learning and progress in knowledge, application, or organization, and the rest are simply beneficiaries (or victims) of an evolutionary process they had little cognizance of. Those of us on the frontiers of refining, expanding and disseminating human knowledge have an opportunity as well as an obligation to understand the processes by which we progress or perish. We, who desire to maximize our control of our own destinies and that of our progeny, must be very cognizant of the need to ever sharpen our idealistic tools and processes so that our understanding and manipulations of the real world bring us benefit, and not harm.

REPRESENTATION

Representation pertains to the ideal model, and how well it mimics the real system we seek to represent. Representation can also pertain to the sample that we extracted from the population – how well do the sample predictors represent the population parameters

In the case of the ideal model, it is usually adequate to include only a few of the most important factors, and omit the less important ones. So inherently the model is approximate. Also, in many cases, especially quasi-experimental models, it is usually not known which factors are important, which are not, and which are spurious – so the model representation is even more ambiguous.

APPROXIMATION

Approximation is commonly used loosely to be interchangeable with representation and precision. But in this discussion, *approximation* refers to the *ideal* formulas and equations used to describe the relations of variables in *real* systems. While undergraduate students in the sciences and in engineering occasionally receive glimpses of the approximations and simplifications made in the derivation of formulas and equations, many never realize that <u>every</u> such formula and equation is an approximation to the real phenomena described.

PRECISION

Precision is a definition of measurement and manufacturing. In the real world, precision of measurement is a matter of degree that can always be improved. Often our measuring instruments are intrusive and disturb the thing being measured so that inaccuracy is inherent.

Inanimate objects. In the ideal world, we can measure exactly, but in the real world, all of our instrumentation is imprecise to some degree. For example, a surveyor's 100-foot chain shrinks and elongates according to the temperature, requiring a temperature correction. That same chain sags when stretched out for measurement, and requires a spring-loaded tension at specific force to compensate for the sag. The chainman's hand trembles a bit has he struggles to hold the end of the chain precisely over the pin on the ground as he holds a plumb-bob in one hand and the chain end in the other. The crewman reading the measure from the chain must interpolate a reading as the marker on the chain undulates a fraction here or there. When it all done, there is a small error, hopefully negligible. But as the survey crew traverses the outline of a property, or extends the geodetic position, such error accumulates, and must be arbitrarily distributed. In the end, if the accumulated errors exceed the contracted precision, it must be done again.

Advances in measurement accuracy more or less keep pace with the need for more accuracy. However, as longer distances are projected, such as in predicting orbits for satellite placement, sufficient accuracy in initial measurements are impossible (remember chaos?) and self-adjusting mechanisms are employed to make necessary corrections in the initial trajectory and final orbit.

Animate objects. The above illustrates problems in precision in measuring inanimate objects. When living organisms are being measured, especially when human observation is employed, which is almost always, many other factors begin interplay with severe effect on the precision of the measurement.

Deceit and camouflage are a necessary and essential part of living organisms. They are part of the survival mechanisms each life form employs to ensure survival and replication. The methods employed range from faking aggression to faking death. And if that were not enough, many life forms *are* aggressive, with sharp fangs and/ or claws, as well as poisonous stingers. The point is that the investigator must be aware of this normal deceit and camouflage, and be able to look beyond it.

In the case of humans, elaborate social rituals and actions have evolved to mask our true intents and feelings from prying, spying eyes and ears. We may not answer truthfully or at all if the question threatens our well-being, or puts us at risk of harm or exposing our secrets. In addition, we may have pre-conceived notions of the outcome, and the research is nothing more than a ritual to formalize our position – whether or not it is the truth.

PREDICTION

Prediction seems to be a basic biological need. All life forms appear to need a prediction mechanism to answer the questions of where the sustenance is, how to get to it, how to capture it, how to avoid being eaten by another organism, when to replicate, what is dangerous, what is symbiotic, etc. We study order and disorder, and their offspring chaos and randomness, precisely because their occurrences in real systems appear to thwart our ability to predict events accurately or at all by our ideal systems.

There are two types of prediction that interest us – the first to predict *outcome*, and the second to cause a *change*. In predicting outcome, especially of natural events, the use of proxies and markers may be quite adequate. However, when predicting change, causal factors must be used if any semblance of precision is desired. And causation is a very difficult thing to establish using statistical processes. In true experiments, of course, causal relations can be established with great certainty for carefully designed, executed and repeated experiments. But in quasi-experimental designs, causation is much more difficult, if it is possible at all. Social designers, generally with no other choice but the quasi-experimental design, are more often than not disappointed in the perverse outcomes of their attempts to cause social change. In many cases there never existed any study to verify the social model, and where there was, causal factors were not established, and/or spurious factors were not eliminated.

Within ideal systems *time* may be reversible or irrelevant, but within real systems, *time* is irreversible and usually always relevant. Many observational problems plague our ability to detect, discern, or measure real systemic mechanisms. The relative timing of events is often difficult to establish because our observations are indirect and/or imprecise.

CLOSING

The student of statistical methodology is taught, and too often routinely employs, mathematical analogies and processes that give the appearance of determinacy and accuracy to the formulaic output. What this paper has tried to present is a feel for the error and indeterminacy inherent in the statistical process, and especially the quasi-experimental design.

Forearmed with the knowledge that statistical answers are only indicative, however weakly or strongly, and not certain, it is hoped that the statistical investigator will more cautiously state their conclusions.

It is also hoped that this paper is simple enough to help the general public understand the inherent limitations of statistical studies. Sufficiently forearmed, the ordinary layperson is better able to discount the many questionable findings so frequently (mis)reported in popular magazines, newspapers, and periodicals.

ANNOTATED BIBLIOGRAPHY

1. Babbie E. 1990. *Survey research methods*. 2nd Ed. Belmont (CA). Wadsworth Publishing Co. 395 p.

   Synopsis: A comprehensive text on preparing and executing statistical surveys.

2. Davis JA. 1992. *The logic of causal order*. Newbury Park. Sage Publ. 71 p.

   Synopsis: A treatise on making causal inference in statistical analysis.

3. Davis JA, Smith TW. 1992. *The NORC general social survey*. Newbury Park. Sage Publ. 95 p.

   Synopsis: A treatise on accessing, using General Social Survey (GSS) databases for social studies.

4. Glenn, ND. 1998. *Survey research methods*. University of Texas at Austin. Unpublished.

   Synopsis: Class lecture notes and handouts and discussions for graduate level course on controversial issues and problems with social survey research methods and conclusions.

5. Graziano A, Raulin M. 1996. *Research methods - a process of inquiry*. 3rd Ed. NY. Longman. 464 p.

   Synopsis: A text on research techniques, scientific methodology, and statistical analysis.

6. Hamilton RF, Wright JF. 1986. *The state of the masses*. Hawthorne. Aldine Publ. 470 p.

   Synopsis: A lengthy critique of the findings and errors of contemporary social studies in the satisfaction or discontent of American family life and the American work force.

7. Kalton, G. 1983. *Introduction to survey sampling*. Newbury Park. Sage Publ. 96 p.

   Synopsis: A treatise on designing statistical surveys.

8. Kiecolt KJ, Nathan LE. *Secondary Analysis of survey data*. Newbury Park. Sage Publ. 88 p.

   Synopsis: A treatise on the advantageous and disadvantages of using existing statistical databases for social studies.

9. Knoke D, Hohrnstedt GW. 1994. *Statistics for social data analysis*. 3rd Ed. Itaska (IL). F.E. Peacock. 574 p.

   Synopsis: A comprehensive text on the theory and techniques of statistical analysis.

10. Lieberson S. 1985. *Making it count: the improvement of social research and theory*. Univ. of California. Berkley. 257 p.

    Synopsis: A text on the limitations of the quasi-experimental method, the need for well-specified models, and the inherent problems in using partial correlations in quasi-experimental designs.

11. Menard S. 1991. *Longitudinal Research*. Newbury Park. Sage Publ. 80 p.

    Synopsis: A treatise on repeated cross-sections, and longitudinal research and analysis.

12. Speigel, M. 1975. *Theory and problems of probability and statistics*. Schaum's Outline

Series, McGraw Hill, NY. 372 p.

Synopsis: A concise reference of probability and statistical theory with numerous completed examples and problems.

13. Sudman S, Bradburn NN. 1982. *Asking questions - a practical guide to questionnaire design*. San Francisco. Jossey-Bass Publ. 397 p.

Synopsis: A text on formulating and validating questions to be used in survey research.

APPENDIX A

	Adult Females			Adult Males			Composite		
	δx	0.05		δx	0.05		δx	0.05	
	σ	0.166667		σ	0.166667		σ		
	μ	5.25		μ	5.75		μ		
	height	***cdf***	***pdf***	**height**	***cdf***	***pdf***	**height**	***cdf***	***pdf***
	4.00	0.000000	0.000000	4.00	0.000000	0.000000	4.00	0.000000	0.000000
	4.05	0.000000	0.000000	4.05	0.000000	0.000000	4.05	0.000000	0.000000
	4.10	0.000000	0.000000	4.10	0.000000	0.000000	4.10	0.000000	0.000000
	4.15	0.000000	0.000000	4.15	0.000000	0.000000	4.15	0.000000	0.000000
	4.20	0.000000	0.000000	4.20	0.000000	0.000000	4.20	0.000000	0.000000
	4.25	0.000000	0.000000	4.25	0.000000	0.000000	4.25	0.000000	0.000000
	4.30	0.000000	0.000000	4.30	0.000000	0.000000	4.30	0.000000	0.000000
	4.35	0.000000	0.000000	4.35	0.000000	0.000000	4.35	0.000000	0.000000
	4.40	0.000000	0.000000	4.40	0.000000	0.000000	4.40	0.000000	0.000000
	4.45	0.000001	0.000001	4.45	0.000000	0.000000	4.45	0.000000	0.000001
	4.50	0.000003	0.000005	4.50	0.000000	0.000000	4.50	0.000002	0.000005
	4.55	0.000013	0.000019	4.55	0.000000	0.000000	4.55	0.000007	0.000019
	4.60	0.000048	0.000063	4.60	0.000000	0.000000	4.60	0.000024	0.000063
	4.65	0.000159	0.000192	4.65	0.000000	0.000000	4.65	0.000080	0.000192
	4.70	0.000483	0.000536	4.70	0.000000	0.000000	4.70	0.000242	0.000536
	4.75	0.001350	0.001370	4.75	0.000000	0.000000	4.75	0.000675	0.001370
	4.80	0.003467	0.003200	4.80	0.000000	0.000000	4.80	0.001734	0.003200
	4.85	0.008198	0.006838	4.85	0.000000	0.000000	4.85	0.004099	0.006838
	4.90	0.017864	0.013364	4.90	0.000000	0.000000	4.90	0.008932	0.013364
	4.95	0.035930	0.023883	4.95	0.000001	0.000001	4.95	0.017966	0.023885
	5.00	0.066807	0.039037	5.00	0.000003	0.000005	5.00	0.033405	0.039042
	5.05	0.115070	0.058351	5.05	0.000013	0.000019	5.05	0.057542	0.058370
	5.10	0.184060	0.079768	5.10	0.000048	0.000063	5.10	0.092054	0.079831
	5.15	0.274253	0.099728	5.15	0.000159	0.000192	5.15	0.137206	0.099920
	5.20	0.382089	0.114027	5.20	0.000483	0.000536	5.20	0.191286	0.114563
5'-3"	**5.25**	**0.500000**	0.119235	5.25	0.001350	0.001370	5.25	0.250675	0.120605
	5.30	0.617911	0.114027	5.30	0.003467	0.003200	5.30	0.310689	0.117227
	5.35	0.725747	0.099728	5.35	0.008198	0.006838	5.35	0.366972	0.106566
	5.40	0.815940	0.079768	5.40	0.017864	0.013364	5.40	0.416902	0.093132
	5.45	0.884930	0.058351	5.45	0.035930	0.023883	5.45	0.460430	0.082234
5'-6"	5.50	0.933193	0.039037	5.50	0.066807	0.039037	**5.50**	**0.500000**	0.078073
	5.55	0.964070	0.023883	5.55	0.115070	0.058351	5.55	0.539570	0.082234
	5.60	0.982136	0.013364	5.60	0.184060	0.079768	5.60	0.583098	0.093132
	5.65	0.991802	0.006838	5.65	0.274253	0.099728	5.65	0.633028	0.106566
	5.70	0.996533	0.003200	5.70	0.382089	0.114027	5.70	0.689311	0.117227
5'-9"	5.75	0.998650	0.001370	**5.75**	**0.500000**	0.119235	5.75	0.749325	0.120605
	5.80	0.999517	0.000536	5.80	0.617911	0.114027	5.80	0.808714	0.114563
	5.85	0.999841	0.000192	5.85	0.725747	0.099728	5.85	0.862794	0.099920
	5.90	0.999952	0.000063	5.90	0.815940	0.079768	5.90	0.907946	0.079831
6.68%	5.95	0.999987	0.000019	5.95	0.884930	0.058351	5.95	0.942458	0.058370
19645	6.00	0.999997	0.000005	6.00	0.933193	0.039037	6.00	0.966595	0.039042
0.00034%	6.05	0.999999	0.000001	6.05	0.964070	0.023883	6.05	0.982034	0.023885
	6.10	1.000000	0.000000	6.10	0.982136	0.013364	6.10	0.991068	0.013364
	6.15	1.000000	0.000000	6.15	0.991802	0.006838	6.15	0.995901	0.006838
	6.20	1.000000	0.000000	6.20	0.996533	0.003200	6.20	0.998266	0.003200
	6.25	1.000000	0.000000	6.25	0.998650	0.001370	6.25	0.999325	0.001370
	6.30	1.000000	0.000000	6.30	0.999517	0.000536	6.30	0.999758	0.000536
	6.35	1.000000	0.000000	6.35	0.999841	0.000192	6.35	0.999920	0.000192
	6.40	1.000000	0.000000	6.40	0.999952	0.000063	6.40	0.999976	0.000063
	6.45	1.000000	0.000000	6.45	0.999987	0.000019	6.45	0.999993	0.000019
	6.50	1.000000	0.000000	6.50	0.999997	0.000005	6.50	0.999998	0.000005
	6.55	1.000000	0.000000	6.55	0.999999	0.000001	6.55	1.000000	0.000001
	6.60	1.000000	0.000000	6.60	1.000000	0.000000	6.60	1.000000	0.000000
	6.65	1.000000	0.000000	6.65	1.000000	0.000000	6.65	1.000000	0.000000
	6.70	1.000000	0.000000	6.70	1.000000	0.000000	6.70	1.000000	0.000000
	6.75	1.000000	0.000000	6.75	1.000000	0.000000	6.75	1.000000	0.000000
	6.80	1.000000	0.000000	6.80	1.000000	0.000000	6.80	1.000000	0.000000
	6.85	1.000000	0.000000	6.85	1.000000	0.000000	6.85	1.000000	0.000000
	6.90	1.000000	0.000000	6.90	1.000000	0.000000	6.90	1.000000	0.000000
	6.95	1.000000	0.000000	6.95	1.000000	0.000000	6.95	1.000000	0.000000
	7.00	1.000000	0.000000	7.00	1.000000	0.000000	7.00	1.000000	0.000000
			1.000000			1.000000			2.000000