Why Sampling Doesn't Seem to Work

Edward J. Stanek III
Department of Biostatistics and Epidemiology
University of Massachusetts
Amherst

Abstract

The idea of random sampling a finite population has broad application in statistics. Response is recorded on a subset (i.e. sample) of the population with a goal of estimating the population average response. By picking the members of the subset in a random manner, a statistical argument is used to draw conclusions about the population average based on the sample response. The same conclusions do not follow if the subset is chosen on purpose. Since the same responses will be observed in each setting, how can the inference from sampling be so different? This is the 'magic' of sampling which may have motivated Mark Twain to say: 'There are three kinds of lies: lies, damned lies, and statistics.' There have been doubters about the validity of inference from probability sampling for a long time. We illustrate using a geometric framework how the added insight attributed to probability sampling appears to be false. We suggest that apart from using probability sampling as an approach to provide face validity for making unbiased data collection decisions, a preference should be given to statistical inference approaches other than probability sampling.

Keywords: Probability Sampling; Estimation; Nonparametric Inference; Sampling theory

1. Introduction

A basic idea often introduced in a first discussion of statistics is simple random sampling from a finite population. Sampling has a long history dating from Laplace (1814) and has been developed into a systematic science embraced by the US census, and popularized by books such as Hansen, Hurwitz and Madow 1952, Kish 1966, Cochran 1977, Lohr 1999, Valliant et al. 2013)and many others. To summarize, by selecting a simple random sample of subjects from a finite population and observing response on the selected subjects, the theory is that we can estimate the population mean. This approach is thought to be much superior to a purposeful (or volunteer) sample, as discussed by Neyman (1934) with respect to Gini and Galvani's use of a purposeful sample of the Italian 1921 census to predict birthrates. We examine this theory via two examples, highlighting the role played by the model for a subjects' response, and the subject-response link. The first example illustrates the practical importance of the problem and raises a doubt as to whether or not sampling is up to the task. The second example illustrates the problem in a simple geometric setting, showing that probability sampling appears to provide no more insight than purposeful sampling. These examples are followed by a critical summary of several arguments that claim to show the superiority of probability sampling.

2. Example 1: Complication Rate for Cateterizations

Fifty patients ($N = 50$) enrolled in a group practice are considered by physicians to be 'good candidates' for a new catheterization procedure. The new procedure is thought to have low risk of complication (measured as presence/absence of one or more complication for a patient), with the risk varying from patient to patient. The patients are listed, and forty five patients ($n = 45$) are selected at random from the listing for the new procedure. After performing a catheterization on the selected patients, a complication occurred for only one patient, yielding $\hat{p} = 1/45$, or a sample complication rate of 2.2%. Based on sampling, this is the estimated complication rate in the population, and could be used to estimate the probability of a complication among the remaining patients. When each patient has his own risk of complication, the complication rate in the population is in fact the mean of the risk of complication overall patients. One physician argues that the results of the 45 patients do not help in estimating response for the remaining five patients.

She argues that since the five patients may have different risks of complication, while the sample response does estimate the overall mean risk for the 45 patients, it doesn't predict the probability of a complication among the remaining patients. We return to this example later, and discuss the role that probability sampling has in drawing such a conclusion.

3. *Example 2: A Population With* $N = 3$ *and A Sample Of* $n = 2$

We consider a simple example to illustrate inference from probability sampling. Let a population of $N = 3$ households (HH), $\Omega_P = \left\{\lambda : \lambda_{Daisy}, \lambda_{Lily}, \lambda_{Rose}\right\}$, be listed using the name of the HH head in Table 1. Our goal is to estimate the average # of HH members per household in the population.

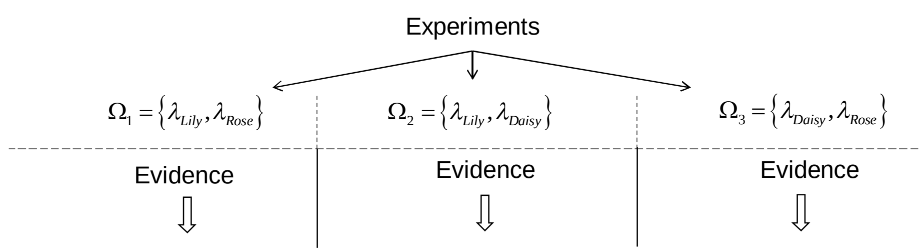
Table 1: Listing of Head of Household in the Sample Frame for $N = 3$ Households (HH)

Head of Household λ	# of HH Members y
Rose	?
Lily	?
Daisy	?

We plan to do so by selecting a simple random sample (without replacement) of $n = 2$ households, and recording the # of HH members, y, in each household. These numbers are the values of two random variables, Y_1 and Y_2. Each is a function of the households in the set, Ω_P, i.e. $\lambda_{Daisy} \rightarrow f\left(\lambda_{Daisy}\right) = y_{Daisy}$ where y_{Daisy} is the number of HH members in $Daisy's$ household. We assume that the # of HH members is not known in advance, and can be observed without error. Our estimate is given by the sample mean, $\bar{Y} = \frac{1}{2}\left(Y_1 + Y_2\right)$, where Y_i, $i = 1,...,n = 2$, is the number observed for aHH in the sample set.

This is a simple problem where probability sampling may be used. Consider what can be learned about the population mean, $\mu = \frac{1}{N}\sum_{\lambda\in\Omega} y_\lambda$, i.e., $\mu = \frac{1}{3}\left(y_{Rose} + y_{Lily} + y_{Daisy}\right)$, via probability sampling when we randomly select $n = 2$ HHs and observe the # of HH members for each. We illustrate the possible sample sets, Ω_d, $d = 1,...,D$, in Figure 1, using terminology introduced by Birnbaum (1962), where evidence, i.e. the # of HH members, could potentially be collected from different possible experiments. Figure 1 separates the 1^{st} step in sampling, choosing a sample set (above the horizontal dashed line), from the 2^{nd} step, observing the random variables for the HHs in the set.

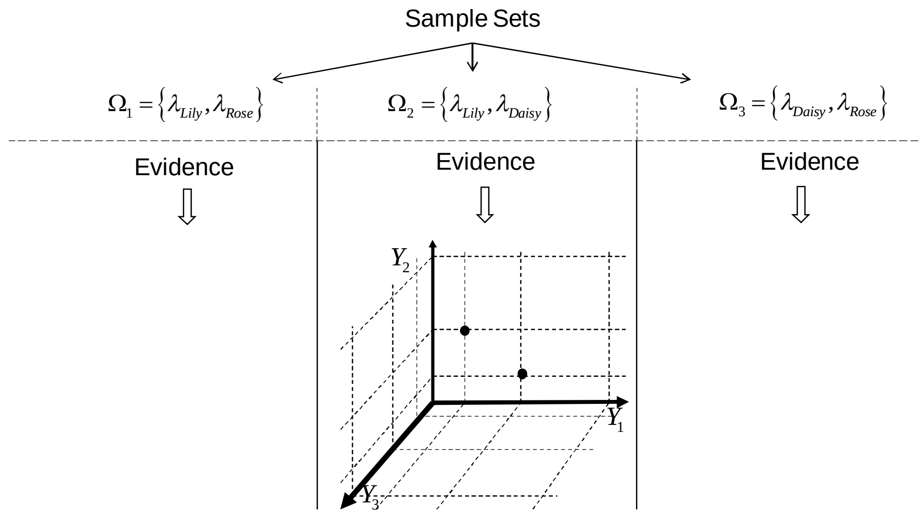
Figure 1. Possible Experiments (Sample Sets)



Birnbaum (1962) concluded that when considering a group of possible experiments to conduct, we only learn from the experiment conducted. In Figure 1, each sample set corresponds to a particular experiment, with mean given by $\mu_d = \frac{1}{n}\sum_{\lambda\in\Omega_d} y_\lambda$. Although simple random sampling associates an equal probability, 1/3, with each of the possible experiments, while purposeful sampling assigns all the probability toa single sample set, Birnbaum claim's that nothing is learned by knowing (or specifying) the probabilities associated with the experiments.

What is learned, i.e. the evidence, comes from observing the values of the random variables from the sample set, which corresponds to the 'sample space'. If Ω_2 is the set of HHs where response is observed, the responses provide evidence only for HHs in Ω_2.There is no additional insight gained for HH in Ω_P , but not in the sample set observed. An illustration of this is given in Figure 2, where the evidence for the sample set, Ω_2, is diagramed relative to a 3-dimensional parameter space for the population, and the sample space for Ω_2.

Figure 2. The Reality of Sampling



The sample space is the $Y_1 \times Y_2$ plane; this is where the evidence occurs. We index the first random variable by $i=1$, i.e. Y_1 , and the second by $i=2$, i.e. Y_2. The two points in Figure 2 corresponding to $\mathbf{Y} = \begin{pmatrix} Y_1 & Y_2 \end{pmatrix}'$ are given by $\begin{pmatrix} y_{Lily} & y_{Daisy} \end{pmatrix}'$ or $\begin{pmatrix} y_{Daisy} & y_{Lily} \end{pmatrix}'$. The order of the random variables is not important, but the index distinguishes response for one HH from another.

Figure 2 can help guide interpretation of $\bar{Y}$ by focusing on the interpretation of the expected value of Y_i, $i=1,...,n$. The expected value of a random variable is $E(Y) = \sum_{\lambda_s \in \Omega} \pi_s y_s$, where π_s represents the probability assigned to the HH $\lambda_s \in \Omega$, $0 < \pi_s < 1$, and $\sum_{\lambda_s \in \Omega} \pi_s = 1$. The key to interpreting $E(\bar{Y})$ is determining whether we should consider Y_i, $i=1,...,n$, to be an element of Ω_P or an element of Ω_d. In Figure 2, the random variable Y_1 (or Y_2) represents a random variable for HHs in Ω_2. The fact that Y_1 cannot be interpreted as the number of HH members for $Rose's$ HH implies that once the experiment is selected, probabilities associated with HHs in Ω_P no longer apply. Assuming $\pi_\lambda = \frac{1}{n}$ for $\lambda \in \Omega_d$, $E_d(Y_i) = \sum_{\lambda_s \in \Omega_d} \pi_s y_s$ for $i=1,...,n=2$ is μ_d, so that when the sample space is defined by $\Omega_2, E(\bar{Y})$ is given by $E_2(\bar{Y}) = \mu_2$. The conclusion that $E_P(\bar{Y}) = \mu$ for Y_i defined over $\lambda \in \Omega_d$ is most likely false since some of the HH in Ω_P are not in Ω_d.

4. Re-Examining Sampling in Example 1

In Example 1, we learn that one complication resulted from performing a new catheterization procedure on a simple random sample of $n = 45$ patients selected from a population of $N = 50$ patients. One physician argued that since the probability of a complication is likely to differ from patient to patient, the results do not shed light on the chance of complication among the five remaining patients. We use the ideas in Section 3 to discuss why based on probability sampling, this conclusion appears to be correct.

First, we note that there are $d = 1,..., \binom{N=50}{n=45} = 2,118,760 = D$ possible distinct sets of $n = 45$ patients that could be formed from the $N = 50$ patients in the population. The sample space represented by $Y_1 \times Y_2$ in Figure 2 corresponds to a 45-dimensional space, $Y_1 \times Y_2 \times \cdots \times Y_{45}$, for each sample set, Ω_d, and will contain 45! points. The remaining space (corresponding to Y_3 in Figure 2) is a 5-dimensional space, $Y_{46} \times Y_{47} \times \cdots \times Y_{50}$, that is perpendicular (orthogonal) to the sample space. Only one of the $d = 1,..., D$ sets is selected, with response, i.e. Y_i, $i = 1,..., n = 45$, observed on patients in the selected set, i.e. $\lambda \in \Omega_d$. Response is not observed on the remaining patients. Since the remaining patients are not members of the sample set, there is zero probability that an observed response, $f(\lambda)$ for $\lambda \in \Omega_d$, is the response for a remaining subject, $\lambda^* \in \Omega_P$, where $\lambda^* \notin \Omega_d$. Thus, nothing is learned about response for the remaining five patients.

5. Why Sampling Appears to Work

Several arguments can be made as to why probability sampling may appear to work. Each alters in some way the subject-response link that defines a random variable as a function of an element of a set (Table 2). We discuss each of these arguments/theories that appear to show that the sample mean is an unbiased estimator of the population mean, and identify what we consider to be problems in each.

Table 2: Frameworks for Inference Relative to the Subject-Response Link

	Subject-Response Link Preserved	
	Yes	No
Population Sampling	Reality	A
Superpopulation Sampling		B
Model Based Approaches (with Response Error)		C
Bayesian Models		D

The framework in Table 2 summarizes different possible basic assumptions that accompany each argument. The cell labeled "Reality" corresponds to an assumption that there is a subject-response link for each subject in the population. For the example of households considered in Section 3, this assumption corresponds to representing a population $\Omega_P = \{\lambda : \lambda_{Daisy}, \lambda_{Lily}, \lambda_{Rose}\}$ via the set of pairs, $\{\lambda_{Daisy}, y_{Daisy}\}$, $\{\lambda_{Lily}, y_{Lily}\}$, and $\{\lambda_{Rose}, y_{Rose}\}$. The elements in a pair correspond to the household label, and the number of members in the household with that label. There is a link between these values that corresponds to reality. The cell identified by A in Table 2 does not require this subject-response link, and could include pairs such as $\{\lambda_{Daisy}, y_{Lily}\}$, where the household label and response label are not linked. The cell identified by B not only does not require a subject-response link, but can also include households in a super population that are not necessarily members of Ω_P. Finally, the model based approaches and Bayesian models, cells C and D, do not typically consider data as being labeled, with response associated with a unique label.

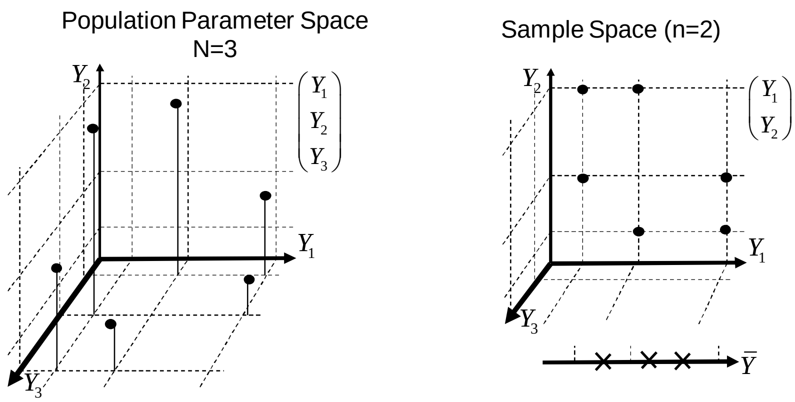
5.1Ignoring the Subject-Response Link (Table 2, Cell A)

The first way that probability sampling may appear to work is by ignoring the subject-response link, so that more points appear to be included in the sample space. To see how this occurs, consider the way that the parameters for HHs in Ω_P can be represented in a parameter space, where $y_{Rose} = 11$, $y_{Lily} = 5$, and $y_{Daisy} = 2$.

The points in Figure 3 on the left correspond to the population parameter when plotted with the HHs assigned to different axes in all possible ways. For each of these points, the population mean, $\mu = \frac{1}{N}\sum_{i=1}^{N} Y_i$ is identical. Notice that HH, i.e. subjects, are dropped in Figure 3.

The projection of each point onto the sample space formed by the $Y_1 \times Y_2$ plane is illustrated in Figure 3 on the upper right. By assuming the HHs have same chance of being associated with any axis on the left, each of the six points on the right is equally likely. Using these probabilities, the expected response is equal to the population mean, $E(Y_i) = \mu$ for $i = 1,...,n = 2$, and $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ is an unbiased estimator of the population mean.

Figure 3. Population Parameter Space and Sample Space Ignoring Labels



The flaw in this argument is that by dropping the HH from the diagram, the sample set is masked in the sample space. Knowing the sample set where response is to be observed, i.e. by picking a sample, a positive probability is associated with only two possible points. Other points have zero probability of occurring. The axis for $\bar{Y}$ in the lower right of Figure 3 plots $\bar{y}_d$ for the three possible sample sets. Let $\bar{Y}$ represent a random variable, i.e. $\Omega_d \to g(\Omega_d) = \bar{y}_d$, for a set in $\Lambda = \{\Omega_d; d = 1,..., D\}$. Assuming each set is equally likely, $E_\Lambda(\bar{Y}) = \mu$. However, when (i.e. given) the set Ω_d corresponds to the sample set, $\bar{y}_d$ is constant given by the parameter μ_d . The fact that probability sampling identifies the set Ω_d where evidence accrues limits interpretation of $\bar{Y}$ to the sample set.

5.2. Adding Artificial Response to form a Super population (Table 2, Cell B)

A different argument supporting probability sampling can be made by adding artificial response to form a super population (Table 2, cell B). First, let us represent the population via the set of pairs, $(\lambda \quad y)$ given by $\Omega_0^* = \{(\lambda_{Rose} \quad 11), (\lambda_{Lily} \quad 5), (\lambda_{Daisy} \quad 2)\}$, as in Cassel et al. (1977). The members of Ω_0^* retain the link between a HH, λ , and the # of HH members, $y = f(\lambda)$.We refer to a larger set of (subject, response) pairs given by $\Omega^* = \{(\lambda \quad y^*)_q; q = 1,..., Q\}$ as a super population. The super population is formed by adding pairs $(\lambda \quad y)^*$ as if any response could occur for each subject, such that

$$\Omega^* = \begin{Bmatrix} (\lambda_{Rose} \quad 11), (\lambda_{Rose} \quad 5), (\lambda_{Rose} \quad 2), \\ (\lambda_{Lily} \quad 11), (\lambda_{Lily} \quad 5), (\lambda_{Lily} \quad 2), \\ (\lambda_{Daisy} \quad 11), (\lambda_{Daisy} \quad 5), (\lambda_{Daisy} \quad 2) \end{Bmatrix}.$$

The argument for adding such pairs is that when the # of HH members for a HH is unknown, any HH could have any of the possible # of HH members that occur in the population. Since $y_{Rose} = 11$, the pairs $(\lambda_{Rose} \quad 5)$ and $(\lambda_{Rose} \quad 2)$ are artificial since Rose's HH does not have 5 or 2 HH members. Notice that $\Omega_0^* \subset \Omega^*$.

Probability sampling associates a probability with each subset Ω_d, $d = 1,...,D = 3$. Suppose the sample is $\Omega_1 = \{\lambda_{Lily}, \lambda_{Rose}\}$, and define Ω_d^* as a subset of Ω^* where $\lambda \in \Omega_d$, such that

$$\Omega_1^* = \begin{Bmatrix} (\lambda_{Rose} \quad 11),(\lambda_{Rose} \quad 5),(\lambda_{Rose} \quad 2), \\ (\lambda_{Lily} \quad 11),(\lambda_{Lily} \quad 5),(\lambda_{Lily} \quad 2) \end{Bmatrix}.$$

Since Ω_d^* includes only the pairs $(\lambda \quad y)$ for sample HH (where $\lambda \in \Omega_d$), we drop the HH labels, λ, and focus on the set of # HH members, $\Upsilon = \{y^* \mid y^* \in (\lambda \quad y^*) \in \Omega_d^*\} = \{2,5,11\}$. This set is the starting point for 'model based' inference (Valliant, et al 2000). The probability associated with the set Ω_d is not relevant, since Ω_d^* is defined conditional on Ω_d. Let the random variable Y_i^*, $i = 1,...,n = 2$ correspond to the value of $y^* \in \Upsilon$ observed for the sample HHs. Assigning equal probabilities to the elements of $\Upsilon = \{y_s^* \mid s = 1,...,N = 3\}$, $E_\Upsilon(Y_i^*) = \sum_{s=1}^{N} \frac{1}{N} y_s^*$ so that $E_\Upsilon(Y_i^*) = \mu$ for $i = 1,...,n$, and $E_\Upsilon\left(\frac{1}{2}\sum_{i=1}^{2} Y_i^*\right) = \mu$. By adding artificial response, even after conditioning on the HHs in the sample, Ω_d, sampling appears to work. The super population approach is appealing since it is conditional on the sample HHs. Response is represented by an element of Υ, which includes response for all HHs in the population. The problem with the logic behind these arguments is the assumption that positive probabilities are assigned to pairs $(\lambda \quad y^*)$ where $y^* \neq y = f(\lambda)$. This assumption violates the connection between theory and reality, since positive probabilities are assigned to pairs $(\lambda \quad y^*)$ that do not in reality exist.

5.3. Interpreting Response via Model Based Approaches (Table 2, Cell C)

A model based framework may be specified for settings where probability sampling is used, and may appear to justify the probability sampling inference (Table 2, cell C). A simple model based framework consists of 'random generator' that produces the value of a random variable, Y, with repetitions independent and typically identically distributed. In the long run, the expected value of Y is equal to μ, which we represent via the response error model, $Y = \mu + \mathrm{E}$. The 'random generator' produces response while at the same time ignores HH labels. There are no 'subjects' in the model-based framework, and hence no subject-response links. As an example, the random-generator could consist of independent selections of a response from Υ, where an equal probability is assigned to each possible response. Sampling without replacement of $n = 2$ can be mimicked by assigning equal probabilities to all possible pairs of response. Since $E_\Upsilon(Y) = \mu$, the sample mean is an unbiased estimator of μ. This framework is very similar to the probability sampling problem, and so it is tempting to assume that the same conclusions apply. However, the model is defined for response without any connection to HHs.

One way to relate a response Y to a HH is to assume that the response error model holds for each HH. With this assumption, the expected # of HH members in each HH is the same, and $Y - \mu$ is 'response error'. The flaw in the model is that usually the model is false. Households have different household sizes. An assumption that all HH have the same size doesn't match reality. However, in such a model perspective, the HHs are not relevant, and the model-based mean, μ, can be interpreted as the mean of any population the analyst suggests. The response-error model argument may appear more plausible for Example 1, where the internal working of the generator is to repeatedly toss a biased coin, with the probability of "complication" equal to μ.

All possible tosses of the coin form the theoretical 'population'. By running the generator 45 times, inference is made about μ for the 'random generator' in a theoretical world (Kass (2011)). Using this logic, we could assume the 45 runs of the 'random generator' are the results for 45 sample patients, and that 5 more 'runs' would produce the results for the remaining five patients. Alternatively, we could assume the 45 runs of the 'random generator' are the results of 45 procedures on a single patient. The only connection of the model to the real problem is the assumption that responses for the model are like responses for a real problem. Inference occurs in the theoretical world.

5.4Bayesian Models (Table 2, Cell D)

We briefly mention a Bayesian framework for inference, not because it is used to justify probability sampling, but rather to note that with such models, the subject-response link is typically not preserved. The prior distribution, which is often an exchangeable distribution, does not identify subjects and their response. Although in the data, a subject-response link can be recognized, the linking is rarely used to update the prior distribution since it isn't captured as part of the prior. By conditioning on the data, the Bayesian framework explicitly adopts Birnbaum's advice, and does not support a conclusion that probability sampling works.

6. Discussion

How has the problem with sampling escaped the scrutiny of researchers? There is evidence that it has not been so lucky. Fisher (1956, p33), in discussing use of relative frequency as a measure of probability noted that "it is essential that the sequence of events contains no recognizable subsets, it must be 'subjectively homogeneous and without recognizable stratification'". This comment can be seen to distinguish the use of probability in finite population sampling and model based frameworks and was noted by Holt and Smith (1979), who recognized similar difficulties with probability sampling. In finite population sampling, the labels create recognizable subsets, limiting the inference to the subjects observed. If the subjects in the population are distinguishable, then unless all the subjects have the same response (or the same expected response), connecting probability to response without breaking the subject-response bond limits inference from sampling. Since in human and biological population, the expected response nearly always differs for different subjects, for such applications, whether a sample set was obtained via probability sampling or purposeful sampling does not change the evidence contributed by observing response on the sample set. Arguments that support inference from probability sampling are misleading, usually because they break the subject-response link, or creation of a false reality. By ignoring the subject-response link, a single population parameter can be represented as illusionary cloud of points in a parameter space (Figure 3). Projecting these points onto a sample space, and associating a positive probability with each point makes it seem that there is added benefit to probability sampling.

Since the subject labels are not traced, it is not possible to distinguish which points represent response for sample subjects, and other points whose assigned probability conditional on the sample should be zero. Others, such as Godambe (1955) have challenged the logical basis of inference from probability sampling. Appreciation of these challenges led directly to a crises in statistics in the 1970s (Johnson and Smith, 1969, Cassel, Särndal, and Wretman (1977)). This has led to other approaches such as super population models. While seeking to justify conditioning on the sample set, super population models appear to create artificial responses. Alternative model-based approaches seek solutions in a theoretical framework, avoiding a formal connection to reality apart from connection by analogy. There is pressure to use inference based on probability sampling even in light of these problems. Sampling provides the theory behind sample surveys which are an established tool used by social scientists and governments to understand reality. Sampling theory is the foundation for a randomization based analysis of experimental design (Scheffé 1959, Kempthorne 1952), and underlies approaches for understanding missing data (Rubin 1976). Clinical trials have randomization and sampling theory at their core (Feinberg and Tanner 1996). Perhaps most importantly, generations of scientists have been trained in the distinction between inference from observational and experimental (sampling based) studies. Since probability sampling promises unbiased methods for answering questions that can be interpreted as extending beyond the data, the approach is compelling. The problems with inference based on probability sampling result in the same limitation faced by purposeful sampling: observing response on some subjects does not tell you the response on those who remain. An important advantage of a sampling framework is the clear definition of a problem in terms of identifiable subjects and the subject's response.

This advantage facilitates describing similarities and differences between subject with similar response, form groups of subjects with homogenous response, and develop hypotheses and theories as to why responses are different. A clear problem definition makes it easier to identify settings where not enough data are available to understand response differences, and accelerate the data collection-understanding process of science. Probability sampling has contributed to focused attention on a clear problem definition. Randomization and probability sampling has also contributed as a strategy for face validity of results that uncouples potential investigator bias or prejudices from the results. By reducing the focus on inference from probability sampling theory, the frequentist-Bayesian split in statistics dissipates. The impact may help shift the focus of statistics from comparison of population averages to understanding reasons for differences between subject's response, or predicting a subject's response. Such a shift will strengthen the paradigm of observation, inference, critical thinking, and then more observation, that is the original hallmark of science.

References

Birnbaum, A. (1962), "On the foundations of statistical inference (with Discussion)," Journal of the American Statistical Association, 57, 269-326.

Cassel, C.M., Sarndäl, C.E. and Wretman, J.H. (1977).Foundations of inference in survey sampling. New York: John Wiley & Sons.

Cochran, W.G. (1977), Sampling techniques, 3rd edition. New York: John Wiley & Sons.

Feinberg, S.E. and Tanur, J.M. (1996). "Reconsidering the fundamental contributions of Fisher and Neyman on experimentation and sampling," International Statistical Review. 64(3):237-253.

Fisher, R.A. (1956), Statistical Methods and Scientific Inference, Hafner Publishing Company, New York.

Godambe, V.P. (1955), "A unified theory of sampling from finite populations," Journal of the Royal Statistical Society, B**,** 17, 269-278.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953), Sample survey methods and theory. Vols I and II. New York: Wiley.

Holt, D. and Smith, T.M.F. (1979). "Post stratification," Journal of the Royal Statistical Society. A. 142(1):33-46.

Johnson, N. and Smith, H. editors (1969).New Developments in Survey Sampling. New York: John Wiley & Sons.

Kass, R.E. (2011). "Statistical Inference: the big picture (with Discussion)," Journal of the American Statistical Association, 26, 1-20.

Kempthorne, O. (1952). Design and analysis of experiements. New York: John Wiley & Sons.

Kish, L. (1965), Survey sampling. New York: John Wiley & Sons.

Laplace, Pierre Simon, (1902) A Philosophical Essay on Probabilities, translated from the 6th French edition by Frederick Wilson Truscott and Frederick Lincoln Emory. New York: John Wiley & Sons.

Lohr, S.L. (1999), Sampling: Design and Analysis, Duxbury Press, Boston.

Neyman, J. (1934). "On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection," Journal of the Royal Statistical Society, 97, 558-625.

Rubin, D.B. (1976). "Inference and missing data," Biometrika63(3);581-592.

Scheffé, H. (1959). The Analysis of Variance. New York: John Wiley & Sons.

Twain, M.attributed (North American Review, 1906) to Benjamin Disraeli (19th century British Prime Minister) but of uncertain origin, see (http://www.york.ac.uk/depts/maths/histstat/lies.htm).

Valliant, R., Dorfman, H., and Royall, R (2000).Finite population sampling and inference. New York: John Wiley & Sons.

Valliant, R., Dever, J.A., and Kreuter (2013).Practical Tools for Designing and Weighting Survey Samples. New York: Springer.