Sampling
Geography 450, Urban Analysis
Elvin Wyly

> "Sampling and generalizing are unavoidable practices because, even before being scientific, they are everyday life activities deeply rooted in thought, language, and practice."[1]

> "Sampling is a major problem for any kind of research. We can't study every case of whatever we're interested in, nor should we want to. Every scientific enterprise tries to find out something that will apply to *everything* of a certain kind by studying *a few examples*, the results of the study being, as we say, 'generalizable' to all members of that class of stuff. We need the sample to persuade people that we know something about the whole class."[2]

So much of what is interesting about urban analysis -- and indeed all fields of scholarly inquiry -- involves the sense of unbounded possibility, the infinite range of exciting processes, questions, and narratives. Cities invite us to revel in their rich complexity of social process and spatial form, their internal connections, contradictions and tensions, and their inter-relations and inter-dependencies with other cities, regions, and nation-states.

Nevertheless, there are limits. Urban research may be *inspired* by the sense of infinite possibility, but meaningful knowledge becomes elusive or impossible as we approach the limits imposed by time, resources, and the attention spans of writers and readers, speakers and listeners. If the urban world is infinitely complex, we need simplification, summarization, and synthesis to make sense of things. It's simply not possible to define or study "the urban" in its entirety without making choices. We can't consume it all. Understanding the most important facets of even a small- or medium-sized city often presents difficult choices of what to include and what to exclude; since everything is related to everything else, any comprehensive view would also require an analysis of a city's linkages to other places -- opening up a paralyzing range of new and difficult choices. Even if we were to limit ourselves to the world's largest cities, that would still put us in the position of trying to cope with the 414 cities in the world with populations of more than one million; we'd get dizzy even before we got acquainted with just a few of the realities of the Tokyo urban region's 35.2 million people, Mexico City's 19.4 million, or New York's 18.7 million.[3]

---

[1] Giampietro Gobo (2008). "Re-Conceptualizing Generalization: Old Issues in a New Frame." In Pertti Alasuutari, Leonard Bickmand, and Julia Brannen, eds., *The Sage Handbook of Social Research Methods*. Thousand Oaks, CA: Sage Publications, 193-213, quote from p. 198.
[2] Howard Becker (1998). *Tricks of the Trade: How to Think About Your Research While You're Doing It.* Chicago: University of Chicago Press, p. 67.
[3] Stanley D. Brunn, Maureen Hays-Mitchell, and Donald J. Zeigler (2008). *Cities of the World: World Regional Urban Development,* Fourth Edition. Lanham, MD: Rowman & Littlefield, p. 4.

Clearly, we need some way to translate infinite complexity and magnitude into human understanding. Sampling offers us one path out of the paralysis of infinite complexity. **Sampling** is the identification and selection of a limited number of individual phenomena -- events, people, places, etc. -- chosen to represent a much larger (and sometimes nearly infinite) class of phenomena. Since the universe of information and possibilities is much too vast for us to grasp or comprehend in all its multifaceted complexity, we must choose a small number of facets to measure for a manageable number of representative elements. Sampling is not simply a matter of choosing something for "illustration," and the plural of anecdote is not "data."[4] There are clear principles that govern how samples should be identified, and every decision taken in the sampling phase of a study will have important implications for the results and interpretations. Sampling is a craft, art, and science all rolled into one -- a set of methods and techniques shaped by philosophies of knowledge that give us guidance on how to acquire meaningful information about the world.

Across the natural and social sciences, the most common type of sample is referred to as a *simple random sample* -- a subset that is selected purely at random, from a *population* in which each member has a knowable, equal, and non-zero probability of being included in the sample. If these conditions are met, then we can use a wide range of tools of inference -- to observe the characteristics of the sample and to make reliable inferences about the characteristics of the entire population. These tools are extremely powerful, and they are used frequently in nearly all domains of contemporary social, economic, and political life.

Unfortunately, these conditions are often extremely difficult and costly to satisfy. The 'simple random sample' that is simple in theory becomes extremely difficult in practice, and thus most textbook discussions quickly dive into thick, technical discussions. Technical complexity isn't really necessary, however. We'll consider a few technical issues later, but we begin with a rather straightforward distinction of the two main approaches to sampling. **Strategic or non-probability sampling** involves subjective decisions about which things we choose to observe and study; strategic sampling is usually quite qualitative, and it is often associated with case study research (although the two are not exactly the same thing). Strategic sampling is highly individualized: when a researcher uses this approach to choose what to study closely, everything about the researcher's experience, background, expertise, and familiarity with the object of study comes into play. **Probability sampling**, by contrast, privileges objectivity and replication. Epistemologically, of course, we know that objectivity is unattainable; but probability sampling procedures attempt to minimize subjectivity, so that two different researchers, following the same steps, should obtain similar results. Probability sampling is usually quite quantitative, and the most familiar technique is the 'simple random sample.'

[4] "When a man fell into his anecdotage is was a sign for him to retire from the world." Benjamin Disraeli (1870), *Lothair*, Chapter 28; quoted in Una McGovern, ed. (2005). *Webster's New World Dictionary of Quotations*. Hoboken, NJ: Wiley Publishing, Inc., quote from p. 277.

Strategic Sampling

In a delightful book titled *Tricks of the Trade*, the sociologist Howard Becker includes a valuable chapter on sampling. Becker begins by describing the standard social-science reverence for the random sample (a probability approach we'll discuss below) and its utility in allowing us to estimate how well the part (the sample we've identified) represents the whole (the entire population) within a given margin of error. But Becker is quick to emphasize that there are other important and interesting questions. "The relation of a variable's value in the sample to its value in the population is a problem, but it isn't the only sampling problem..."[5] Many times, we have no choice but to make sense of a particular **case study** the best we can, or we have to devise creative sampling strategies to answer other kinds of questions. There are several examples.

1. **The machine trick**. Suppose we "want to know what kind of an organization could be the whole of the thing we have studied is a part." If we're studying a prominent institution, event, or place, it may be less important to try to generalize to an entire "population" than to understand "the way the parts of some complicated whole reveal its overall design.... Archaeologists and paleontologists have this problem to solve when they uncover the remnants of a now-vanished society. They find some bones, but not a whole skeleton; they find some cooking equipment, but not the whole kitchen; they find some garbage, but not the stuff of which the garbage is the remains. They know that they are lucky to have found the little they have, because the world is not organized to make life easy for archaeologists. So they don't complain about having lousy data. Instead, they work on getting from this thigh bone to the whole organism, from this pot to the way of life in which it played its small role as a tool of living. It's the problem of the Machine Trick, of inferring the organization of a machine from a few parts we have found somewhere."[6]

2. **The variance trick**. In the simple random sample, most of the things we choose for study will tend towards the 'middle' on most measures. This tendency becomes stronger as we include more things in our sample, and this is one of the most potent features of simple random sampling. But this is precisely what can obscure the full range of variation in something. There is thus a rich tradition in social research of sampling on the extremes -- trying to choose cases across the entire range of possibilities on a particular variable.

3. The trick of "**finding what doesn't fit**." As Becker notes, Thomas Kuhn's work offered a compelling argument that "Science can only make progress when scientists agree on what a problem and its solution look like -- when, that is, they use conventionalized categories." Without those conventional categories, everyone will be studying phenomena that cannot be compared, "and it won't add up to anything. This is the situation Kuhn describes as having plenty of scientists, but no science."[7] Unfortunately, conventional categories place constraints on the processes and

---

[5] Becker, *Tricks of the Trade*, p. 70.
[6] Becker, *Tricks of the Trade*, p. 70-71.
[7] Becker, *Tricks of the Trade*, p. 85.

relationships that can possibly be observed, making it all to easy for investigators to miss things that are unusual, or that don't fit the standard categories. One kind of strategic sampling, then, involves a deliberate search for events, places, institutions, or situations that stretch the bounds of possibility and defy conventional categories. "*The trick...is to identify the case that is likely to upset your thinking and look for it*."[8]

There are several other approaches in strategic, non-probability sampling, which is also sometimes called "purposive" sampling. **Convenience sampling** involves various kinds of institutional or social situations where there's a captive (or easily-persuaded) audience; students in introductory psychology classes, consumers signing up for loyalty or discount cards, and radio station listeners calling in or logging on to the Internet to take 'web polls' are some examples of convenience samples. None of these are representative of the general population. **Snowball sampling** involves recruiting a small number of willing participants, and then persuading them to refer others who may be willing to be interviewed; if you can get the snowball to roll down the hill with good momentum, you'll get a sufficient sample size. The sample will not be representative, but the snowball method can be a good way to gain the trust of potential participants -- very important if certain types of sensitive questions are to be asked. **Quota sampling** is an approach that is rarely used in scholarly research, but is widespread in commercial market research. A population is divided into various categories, and then individuals are recruited to ensure a specified number for each group; the individuals within each category are not chosen following the rules of random selection and probability, however (see 'stratified' sampling below), making it impossible to generalize from the sample to the population. But it's quick and cheap.

Once you've identified one or more cases with these strategic sampling approaches, then what? Becker recommends, paradoxically, that we first try to forget what sampling is all about, and try to include everything. This applies especially when we're including a very small number in our sample -- perhaps just a single case study. "When I teach field research, I always insist that students begin their observations and interviews by writing down 'everything.' That is, I claim that I don't want them to sample but rather to report the universe of 'relevant' occurrences. This generally leads to a good deal of foot dragging by them and nagging by me."[9] Becker's students object, quite reasonably, that it is impossible to capture everything; and he replies, just as reasonably, that it is possible to record quite a bit more than most casual observers will. Becker advises recording as much as possible, in as detached a manner as possible, with a minimum of interpretation or inference. Becker describes the approach as **massively detailed description**, and he identifies two landmark examples that set the standard -- the French novelist Georges Perec's entire day spent describing everything he saw at Mabillon Junction on May 19, 1978 for a radio station, and the 1941 book *Let Us Now Praise Famous Men: Three Tenant Families*, by James Agee and Walker Evans, who included no less than forty-five pages describing every detail of one Alabama sharecropper family's shack. Becker advocates, at least for the data-collection phase of research, a careful attempt to record a pure, unvarnished description -- with no explanations or interpretations. But when it

[8] Becker, *Tricks of the Trade*, p. 87.
[9] Becker, *Tricks of the Trade*, p. 76.

comes time to analyze the ‘raw’ information, many scholars find much value in “**thick description**.” This is “a term coined by the philosopher Gilbert Ryle and introduced into social science by Clifford Geertz, referring to ethnographic descriptions of informants’ actions and their interpretations of their own actions placed within a specific cultural context. Thick description is contrasted with ‘thin description’ based on the tenets of behavioralism or on survey research...where a detailed description of the informants’ meaning system and broader social context is not always considered necessary. ....”[10]

The limits to strategic sampling

Strategic sampling is powerful and inherently interesting, but it also entails significant limits. Most crucially, strategic sampling can make it difficult or impossible to *generalize*. One of the most pervasive mistakes in research is to present a rich, interesting case study that was selected on the basis of personal, idiosyncratic knowledge and expertise (or simply the convenience of location or access), and then to draw sweeping generalizations and conclusions about the broader class of phenomena. If you’ve chosen to study a particular city, neighborhood, or urban event because of strategic, subjective circumstances, then you give up the opportunity to draw general inferences about *other* cities, neighborhoods, or urban events -- unless, of course, you can show that these different phenomena share relevant characteristics.

There is some disagreement on whether strategic, qualitative sampling approaches allow generalization. We can identify at least five distinct positions.[11]

> 1. *Don’t generalize.* The first position holds that generalization is a misguided goal. “The interpretivist rejects generalization as a goal and never aims to draw randomly selected samples of human experience. For the interpretivist every instance of social interaction, if thickly described,” to use the anthropologist Clifford Geertz’s phrase, “represents a slice from the life world that is the proper subject matter for interpretative inquiry.”[12]
>
> 2. *Generalize only the “emblematic” case.* A second position argues “that the purpose of case studies is not so much to produce general conclusions as to describe and analyze the principal features of the phenomenon studied. If these features concern an emblematic case of political, social, or economic importance ... the ‘intrinsic case study’ will *per se* produce results of indubitable intrinsic relevance, even though they cannot be generalized in accordance with the canons of scientific induction.”[13]

---

[10] James Duncan (2000). “Thick Description.” In R.J. Johnston, Derek Gregory, Geraldine Pratt, and Michael Watts, eds., *The Dictionary of Human Geography*, Fourth Edition. Oxford: Blackwell, p. 827.
[11] Gobo, “Re-Conceptualizing...”
[12] Norman K. Denzin (1983). “Interpretive Interactionism.” In G. Morgan, ed., *Beyond Method: Strategy for Social Research.* Beverly Hills, CA: Sage Publications, 129-146, quote from p. 133-134.
[13] Gobo, “Re-Conceptualizing,” p. 197.

3. *Generalize by connecting 'intensive' and 'extensive' reasoning.* A third position "starts from the distinction between extensive vs. intensive studies. The aim of the former ... is to identify statistically significant and therefore generalizable causal relations; the aim of the latter is to reconstruct in detail the mechanisms that connect cause and effect."[14] This means that strategic, qualitative case studies of particular mechanisms can and should be generalized by comparing the results to the findings of 'extensive' studies done by other researchers.

4. *Generalize ideas, not specific cases.* A fourth approach is based on the recognition that case studies and other strategic sampling approaches can never yield samples that are "perfectly" representative of an entire population or universe. Over time, however, the cumulative growth of knowledge makes it possible to generalize key ideas and concepts: "case study after case study, in the course of time in a particular sector of research, there accumulates a repertoire or inventory of the possible forms that a particular object of study may assume."[15]

5. *Generalize through induction, not deduction.* This involves an approach described as "analytic induction." "The purpose of analytical induction is to uncover causal relations through identification of the essential characteristics of the phenomenon studied. To this end, the method starts with not a hypothesis but with a limited set of cases from which an initial explanatory hypothesis is then derived." Other cases are then studied, and "If the initial hypothesis fails to be confirmed, it is revised. Additional cases of the same class of phenomena are then selected. If the hypothesis is not confirmed by these further cases, the conceptual definition of the phenomenon is revised. The process continues until the hypothesis is no longer refuted and further study tells the researcher nothing new."[16]

Probability Sampling

The second main approach in sampling is rooted in theories of probability and inference that allow us to generalize from the characteristics of a sample to the characteristics of a larger population. The most powerful systematic sampling approach is the simple random sample, in which each element in the universe has a knowable, equal, and non-zero chance of being included in the sample. Choosing a simple random sample requires sufficient information about the entire population or universe of possibilities -- a comprehensive "sampling frame" or inventory of all members -- to ensure that each has an equal, positive chance of being selected. Many sampling frames do not meet the conditions required for simple random samples; random-dialed telephone calls, for example, miss everyone without telephones. The importance of obtaining simple random samples for so many social, political, and administrative functions -- so that the sample

[14] Gobo, "Re-Conceptualizing," p. 197.
[15] Gobo, "Re-Conceptualizing," p. 198.
[16] Gobo, "Re-Conceptualizing," p. 198.

characteristics can be generalized and inferred to the entire population -- justifies enormous public expenditures by nation-states to conduct periodic censuses.

Simple random sample

Suppose we're interested in measuring the social characteristics of a town of 10,000 people, and we only have enough time and money to interview 450 people. The probability of inclusion in the sample is thus

$\frac{450}{10,000}$, or about 1 in 22.

This is the sampling fraction, and it is typically expressed as

$$\frac{n}{N}$$

where *n* is the sample size, and *N* is the population size. For a simple random sample, choosing which people to interview involves several steps.

> a. Define the population; in our case, this is the entire population of 10,000 residents of our city.
>
> b. Select a comprehensive sampling frame -- an inventory of all persons who meet the criteria for inclusion; this requires an accurate and complete list of all 10,000 residents.
>
> c. Decide upon a sample size, *n*.
>
> d. List all members of the population, assigning each a unique number between 1 and N.
>
> e. Use a table of random numbers, or a computer program that generates random numbers, to select *n* different numbers within the range between 1 and N. Each member of the population corresponding to a chosen random number would be selected for an interview.

Systematic sampling

The steps above can be quite tedious (and they were particularly annoying before the development of certain types of computer algorithms). One common shortcut is the systematic sample, which replaces the random-number step with something else. If we had a comprehensive alphabetical listing of all 10,00 city residents, and if we counted down every 22 lines to choose another name, we would obtain the same sample size as the simple random sampling approach. In this approach, however, not every member has an equal chance of being included in the sample. Individuals whose names happen to be in certain places on the list have no chance of being included in the sample. Moreover, the decision of where to start counting on the list, and the idiosyncrasies of names rendered in a particular language with a particular alphabet, will affect the probabilities of inclusion in the sample. These idiosyncratic factors may not be entirely predictable, but they also do not meet the requirements of pure, random, chance variation.

Stratified random sampling

In many cases, we wish to ensure that a sample has sufficient representation of different categories or subdivisions of the entire population. If our study of the city with 10,000 residents focused on issues of housing, for instance, we might wish to ensure that we have the right proportion of <u>owners</u> and <u>renters</u> in the sample. One way to do this is to divide the population into strata, and then select a random or systematic sample within each strata.

Stratification can be designed in two different ways.

> <u>Proportionate stratified samples</u> ensure that the proportion of the population sampled is the same for each category. This approach "ensures that the resulting sample will be distributed in the same way as the population in terms of the stratifying criterion. If you use a simple random or systematic sampling approach, you *may* end up with a distribution like that of the stratified sample, but it is unlikely. Two points are relevant here. First, you can conduct stratified sampling sensibly only when it is relatively easy to identify and allocate units to strata. If it is not possible or it would be very difficult to do so, stratified sampling will not be feasible. Second, you can use more than one stratifying criterion."[17]
>
> <u>Disproportionate stratified samples</u> apply a different sampling fraction to the different strata. This approach is essential when resources are limited, and we wish to compare some variable between two strata that are very unequal in their prevalence. By "over-sampling" the small strata, we ensure that there is a sufficient sample size to yield reliable estimates to compare with the sample drawn from the larger strata. If we wish to obtain total estimates for the entire population, we will have to 'weight' the strata to adjust for over- and under-sampling.

Cluster sampling and multi-stage cluster sampling

While the purpose of probability sampling is to ensure a pure, random selection of things to observe and measure, we know that the world -- especially the social world -- is not completely chaotic and random. Our city of 10,000 residents is not just a chaotic assemblage of 10,000 people scattered across a featureless plain. People are organized in households, neighborhoods, schools, community groups, employers, and many other institutions. Multi-stage cluster sampling recognizes this organization, and takes

---

[17] Alan Bryman (2008). *Social Research Methods*, Third Edition. Don Mills, ON: Oxford University Press, p. 173.

advantage of it. In some cases, clusters are chosen because they make it easy to collect large sample sizes in vast, far-flung regions or complex environments. In other cases, the clusters are defined by logic or policy: if your target population is currently-enrolled students who attend class in elementary school, for instance, then it would be logical to choose a set of elementary schools, and then identify a simple or systematic random sample of students within each cluster. In the language of sampling methodology, the school is the "primary sampling unit," the first stage of selection, and then the students within each school are "population units."

Cluster sampling involves using a random selection procedure to choose primary sampling units (clusters), and then including every member of the population within each cluster. Multi-stage cluster sampling involves random selection both to identify the clusters, and to identify a sample of the population within each cluster.

The Rules of Inference

If a sample has been chosen through random sampling methods, this makes it possible to use certain principles to infer from the characteristics of the sample to the characteristics of the population. The simplest of these principles begins with the standard normal distribution.

For centuries, it was understood that repeated measurements of various phenomena -- human heights and weights, the sizes of various plants, etc. -- tended to reveal a remarkably similar pattern. When measures were plotted as histograms, the results resembled a smooth, bell-shaped curve, with most of the observations clustering around the mean or the average, and proportionately smaller shares of observations farther away from the mean in both directions. "It is remarkable how many times we end up with this same bell-shaped curve, irrespective of which variable we are studying... As the curve is produced so often it is called the normal distribution. The interesting and very useful feature of this curve is that it is actually quite simple to express mathematically and can be calculated using only the mean and standard deviation."[18] Not all phenomena conform to the normal distribution, but many do -- **so many, in fact, that it eventually became possible to work out the precise details and proportions of various parts of the bell curve, if we expressed all observations in terms of standard deviations from the mean -- z-scores.**

---

[18] Perry R. Hinton (1995). *Statistics Explained.* New York: Routledge, p. 27.

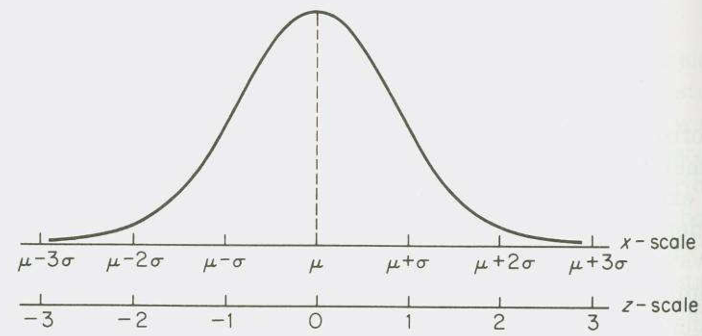

Figure 8.9. Change of scale.

Figure 2. The standard normal distribution. Source: Figure 8.9 from John E. Freund (1973). *Modern Elementary Statistics*, Fourth Edition. Englewood Cliffs, NJ: Prentice-Hall, p. 200.

Clearly, the vast majority of observations in the standard normal distribution are clustered around the mean -- within one or two standard deviations above or below the mean. Few observations are to be found farther away from the mean. Specifically, in the idealized, perfect normal distribution, only 2.5 percent of the observations will be more than 1.96 standard deviations above the mean, and only 2.5 percent of the observations will be more than 1.96 standard deviations below the mean. Put another way, 95 percent of all of the observations are in the range of -1.96 *z* to +1.96 *z*. (You can find these cut-offs and many others in the 'standard normal distribution' tables in the appendix of almost any statistics text.)

Not all phenomena will 'conform' to the standard normal distribution. Some things are 'skewed' one way or another, with the 'tails' of the distribution extending farther out to the right, or to the left. In many cases, if a variable is not normally distributed, it is possible to use a transformation (if applied consistently to all observations) to bring it closer to the normal curve shape. But here's what's crucial for sampling theory: **the one phenomenon that is always normally distributed is the variation that results from random sampling error.** In other words, if we conduct a small survey of our town of 10,000 people, and then we do another survey, and if we continued to do many, many surveys, each one will yield a slightly different result on whatever variable we're measuring (age, income, etc.). If we were to draw a histogram of the different results from all these repeated surveys, we would obtain a normal bell curve. This is called the <u>sampling distribution</u>. Just as the entire population has its own distribution, with a mean and a standard deviation, the sampling distribution has its own mean, and its own standard distribution -- which is called the *standard error of the sample means*, or sometimes the *standard error of the mean*. The standard error of the mean is the average deviation of any sample mean from the "true" population mean.

Since the sampling distribution conforms to the standard normal curve, probability theory offers several crucial principles for drawing inferences. Three issues are most important.

1. First, even if our variable of interest is does not conform to a perfect normal distribution, the sampling distribution will tend towards a normal distribution. In other words, if we're measuring something that is highly 'skewed' and non-normal, when we try to find its mean with repeated random samples, the distribution of sample means around the 'true' mean will be normal. This is known as the **central limit theorem**. Figure 3 illustrates the central limit theorem for a series of 60 random samples measuring discharge from the River Clyde.

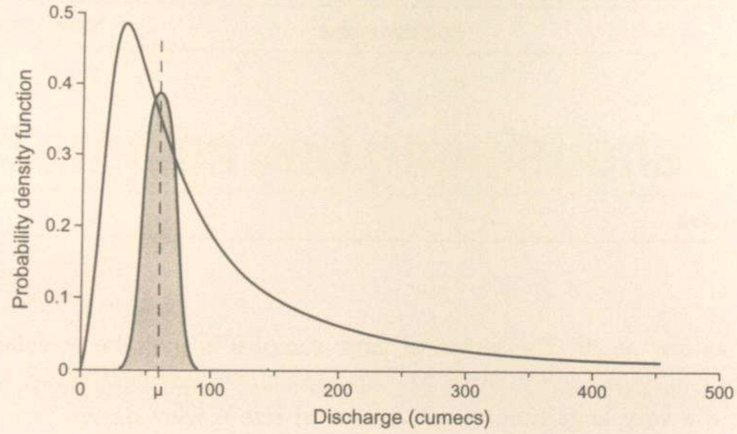

FIGURE 6.1 Summary distribution of daily discharge data for the River Clyde. The superimposed (shaded) area shows the nature of the sampling distributions of means for $n = 60$

Figure 3. An Illustration of the Central Limit Theorem. Source: Figure 6.1, from Dennis Wheeler, Gareth Shaw, and Stewart Barr (2008). *Statistical Techniques in Geographical Analysis*, Third Edition. London: Routledge, p. 118.

2. The central limit theorem applies only to 'large' samples. With repeated random samples, the sampling distribution comes closer to a 'perfect' normal distribution. How large is 'large?' There is some variation in the consensus of statisticians on this point -- perhaps it has its own sampling distribution! -- but the range is between 30 and 60 observations. As the sample size increases, the distribution becomes closer to a perfect normal curve, and the standard deviation becomes smaller -- such that the curve "tightens" around the "true" population mean. Specifically, the standard error of the mean decreases in proportion to the square root of the number of observations:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

In words, the standard error of the mean is equal to the 'true' standard deviation of the population, divided by the square root of the number of sample observations.

3. These principles make it possible to answer one of the most common questions in social research. You need to do a quick survey to find out an important piece of information: how large must your sample size be?

This question requires good judgment, and involves a number of considerations. The short, simple answer: for a non-parametric contingency table like we discussed last week, you need at least 25 or 30 observations, and more if any of the cells of your table will have fewer than 5 members; to satisfy the central limit theorem, an absolute, bare minimum is 30. A minimum of 60 is preferable. Increasing sample size beyond 60 will reduce the standard error of the mean, increasing the precision of the analysis. Above about 150, the improvements to the standard error begin to diminish considerably.

Now the more complicated answer: if you want to measure just one thing, or to create a very simple contingency table, then these simple criteria should suffice. But if you wish to do any comparisons amongst different subgroups, or if you wish to measure the relations among multiple characteristics, then sample sizes must be larger.

Why it Matters

If this all seems rather dry, boring, and technical, let's consider the importance of carefully designing a multi-stage cluster analysis. In the years after the U.S. invaded Iraq in March, 2003, intense protests and political debates erupted over the question of how many civilians were killed -- above and beyond those who died during the initial wave of combat. Many of these deaths were caused by civil war, while countless others died because the invasion had destabilized or destroyed significant parts of the nation's infrastructure. It was all too easy to ignore those who did not count precisely because they were not counted, and so for many years, activists, public officials, and many others fought over alternative estimates of civilian deaths. At one point, U.S. Army General Tommy Franks famously declared at a press conference, "We don't do body counts," apparently forgetting that the Geneva Conventions specifically *require* this activity among the leadership of armed forces in combat. A coalition of activist researchers responded by closely monitoring all available media sources for reports of violence, attempting to distinguish civilian deaths from military or insurgent casualties, and aggregating the results to obtain estimates over time.

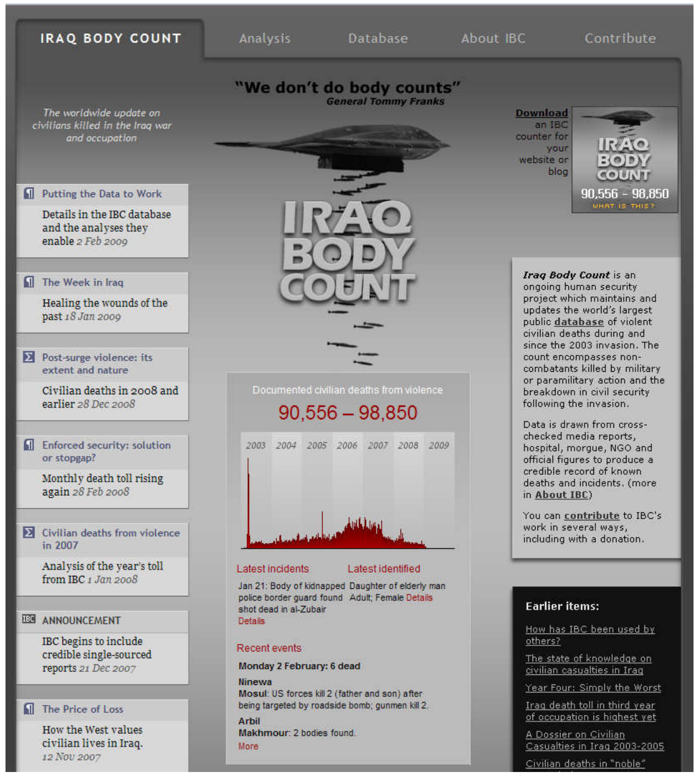

Figure 4. IraqBodyCount.org.

The results attracted considerable attention, but they relied on all the factors that shape the information that winds up in the press sources monitored for reports of deaths -- factors that include, among other things, the uneven geographical distribution of journalists in Iraq, and the greater propensity for stories to be published in cases of large-scale violence than in the more mundane, ordinary circumstances of destruction that became so commonplace. To find out how many people died, from all causes, who would not otherwise have died if the invasion had not taken place, requires identifying a representative sample of households and families, and obtaining very sensitive information from them -- asking them about their relatives, and whether any of these relatives died during a particular period. Here's an excerpt from the methodology section of a landmark study by researchers at the Johns Hopkins School of Public Health and the School of Medicine at Baghdad's Al Mustansiriya University:

> "As a first stage of sampling, 50 clusters were selected systematically by Governorate with a population proportional to size approach, on the basis of the 2004 UNDP/Iraqi Ministry of Planning population estimates .... At the second stage of sampling, the Governorate's constituent administrative units were listed by population or estimated population, and location(s) were selected randomly proportionate to population size. The third stage consisted of random selection of a main street within the administrative unit from a list of all main streets. A residential street was then randomly selected from a list of residential streets crossing the main street. On the

residential street, houses were numbered and a start household was randomly selected. From this start household, the team proceeded to the adjacent residence until 40 households were surveyed. For this study, a household was defined as a unit that ate together, and had a separate entrance from the street or a separate apartment entrance.

> The two survey teams each consisted of two female and two male interviewers, with the field manager ... serving as supervisor. All were medical doctors with previous survey and community medicine experience and were fluent in English and Arabic. A 2-day training session was held. Decisions on sampling sites were made by the field manager. The interview team were given the responsibility and authority to change to an alternate location if they perceived the level of insecurity or risk to be unacceptable. In every cluster, the numbers of households where no-one was at home or where participation was refused were recorded. In every cluster, queries were made about any household that had been present during the survey period that had ceased to exist because all members had died or left. Empty houses or those that refused to participate were passed over until 40 households had been interviewed in all locations.

> The survey purpose was explained to the head of household or spouse, and oral consent was obtained. Participants were assured that no unique identifiers would be gathered. No incentives were provided. The survey listed current household members by sex, and asked who had lived in this household on January 1, 2002. The interviewers then asked about births, deaths, and in-migration and out-migration, and confirmed that the reported inflow and exit of residents explained the differences in composition between the start and end of the recall period. Separation of combatant from non-combatant deaths during interviews was not attempted, since such information would probably be concealed by household informants, and to ask about this could put interviewers at risk. Deaths were recorded only if the decedent had lived in the household continuously for 3 months before the event. Additional probing was done to establish the cause and circumstances of deaths to the extent feasible, taking into account family sensitivities."[19]

Burnham and his colleagues estimated that between March, 2003, and July, 2006, there were more than 654 thousand "excess" Iraqi deaths -- deaths that would not have happened given prevailing mortality trends prior to the invasion -- and that more than 601 thousand of these deaths were due to violence.

[19] Gilbert Burnham, Riyadh Lafta, Shannon Doocy, and Les Roberts (2006). "Mortality After the 2003 Invasion of Iraq: A Cross-Sectional Cluster Sample Survey." *The Lancet*, October 11, 1-8.