IN THEIR OWN WORDS

Laboratory statistician **JOANNE WENDELBERGER** explains how the use of statistics is central to strong national security research.

WHEN I WAS A STUDENT AT OBERLIN COLLEGE, I especially enjoyed my math and science courses. I majored in math, but I wondered what I would do with a math degree, and I also worried that I would miss the excitement of the sciences. During my junior year, Joseph Kadane, a statistician from Carnegie Mellon University, visited my department and gave a seminar on statistical work he did for the United Nations related to negotiations of their Law of the Sea Treaty. I had the opportunity to meet with Dr. Kadane to talk about careers in statistics. Our conversation that day did two things for me: it convinced me I wanted to go to graduate school, and it also introduced me to the idea that statisticians can engage in interdisciplinary problem-solving. As part of a team of experts from different fields, they can tackle important issues, rather than working in isolation or simply acting as an outside consultant. That idea stuck with me and has helped to shape my career. This approach—taking on the substance of a complex question—was perfect for me because I didn't really want to choose between math and science. Although I specialize in statistics, I have been able to work on problems in many different scientific disciplines.

There's a popular saying in statistics and other fields: "All models are wrong, but some are useful." While there are varying viewpoints on this, my Ph.D. advisor at the University of Wisconsin-Madison, Professor George Box, to whom the quote is attributed, imprinted this idea on me during my graduate training. A model is a simplification or normalization of a complex and varied system, intended to aid in the study of that system. Most complex questions in science require the use of models, at least in part, to make them tractable. Even if it is not possible to completely and precisely capture all of the intricate details of a complex system, that's no cause for despair—models, as the saying goes, can be quite useful, despite shortcomings. Complex models can sometimes predict the behavior of a system quite accurately, whereas a simpler model might still yield valuable information by focusing on the essential features of a complex system.

Some models are physical, like a miniature replica or mockup, and some are visual, like a chart or diagram. The kinds of models I use are statistical models, which use mathematical notation to represent particular situations that include randomness, which can add to the complexity of predicting their behavior. For example, in material aging studies, anticipated observations can sometimes be modeled as the sum of low-order polynomial expressions plus an error term that accounts for variability in individual observations. Despite their simplicity, these often do a good job of capturing the main impacts of the experimental factors. This commonly occurs in experiments across many different fields and can be a good starting point when beginning to model complex phenomena. If needed, more complicated models can be developed that incorporate more of the mechanistic details and complex interactions between variables.

Some of the most important considerations when building a statistical model are: the question to be answered (what information is needed), sampling and experiment design (how to measure and collect data), and error (what types of variability are present, and how accurate are the measurements). Both physical and computational experiments can be used to test whether a model is valid. Such experiments benefit from the use of statistical methodologies in their design to help identify the best possible experimental settings. A well-designed experiment can be powerful and illuminating, whereas a poorly designed experiment can be essentially useless, a situation that is often avoidable. This is the part of model building where my research fits in.

THERE'S A POPULAR SAYING IN STATISTICS: "ALL MODELS ARE WRONG, BUT SOME ARE USEFUL."

At Los Alamos, I've worked on statistical aspects of many different groundbreaking research problems, including analysis of ocean-simulation results that contribute to our understanding of the earth's climate, and sampling and visualization of particles from simulations of the origins of the universe. These are big questions that can't be tackled without the unique confluence of math and science capabilities that national laboratories like Los Alamos have come to be known for. Thanks to that fateful seminar back in college, my educational experiences, and my Los Alamos colleagues, I have been fortunate to be a part of that confluence and have tremendously enjoyed it.

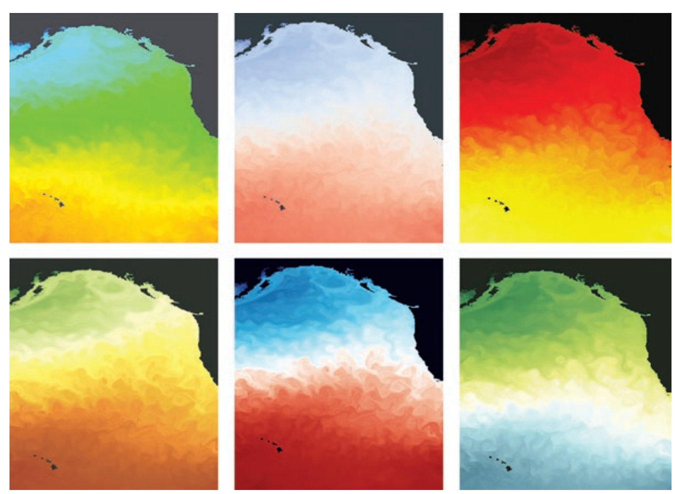
Nature and nurture (with a dash of good timing)

I live and breathe statistics on a daily basis. Both at work and outside of work, I think about everything in terms of distributions. For example, when I take an early morning walk, I automatically count how many rabbits I see. The number varies depending on the time of year and my mind tracks those changes as seasonal distributions. Another example is the experimentation inherent in cooking. Over the years I have amassed a personal catalog of high-altitude cake baking outcomes—some terrible, some fantastic, most somewhere in between—in which the average quality has gradually improved, and the distribution has steadily shifted. Thinking in distributions like this comes naturally to me, so I suppose it's not surprising that I've made a career out of it.

I wound up at Los Alamos through a combination of chance and courage. Near the end of my Ph.D. program, I came across an ad for a position at Los Alamos that looked like a perfect fit for me—and it even came with mountains! I wasn't really looking for a job yet, but my husband and I researched the town and the Laboratory and agreed that it looked like a great place to live and work, so I applied for the job. Because I was still several months away from graduating, it was the only job I applied for, and when it was offered to me, I took a leap. That was 25 years ago, and our initial impression has panned out—it's been an excellent match.

One of the most important things to me about working at the Lab is being a part of a broad science effort across many disparate fields. Computer scientists, mathematicians, chemists, physicists, engineers, and more all work together here. Being a statistician means I have a special set of skills that I can use to help them design experiments and analyze data. It can be very challenging to figure out how to do this in the presence of practical constraints. For example, experiments may need to be arranged in groups of runs that fit into a heating chamber, or scheduled to minimize the need to collect measurements on weekends or holidays. Or there may be limited quantities of special materials available with which to work. As challenging as these kinds of limitations are, it is that much more rewarding when we figure out the best solution.

With the current explosion of computing power, more and more applications rely on streaming data, such as images, video, or spectra. In the case of images, the data can be quite noisy and may require cleaning as part of the analysis process. I recently collaborated with statistics colleague Sarah Michalak and Los Alamos applied mathematician Laura Monroe to create a statistical solution to a noisy radio-astronomy data



Six different presentations of the same northern Pacific Ocean data using different color schemes. Standard color maps (top row) are sometimes unable to resolve differences or areas of interest. Improvements in the perceptual range of color were achieved by developing new color maps (bottom row), which provide scientists with more useful data images. The author helped design and analyze experiments where subjects looked at the old and new maps and reported how many colors they could distinguish, thus establishing a mathematical means of comparison and estimation of improvement.

CREDIT: Francesca Samsel

problem (discussed in more detail in "Two Roads to Next-gen Computing" on page 26). Dr. Michalak and I brought the necessary statistical expertise for this project, including the use of binomial models, theory of ranks and sorted observations, and experiment design for selecting different sets of parameter values. We're still looking at the performance of our solution algorithm and exploring ways to further optimize it. But that evaluation process, too, involves statistical concepts—we're using statistics both to solve the noisy-data problem and also to evaluate and improve the resulting solution.

Big data

Data is often expensive to generate, expensive to collect, and expensive to store, so it is important to get the most we can out of it and not waste resources on uninformative data. This is an excellent reason to work with a statistician. There are also challenges associated with the current trend toward working with extremely large datasets. The Laboratory has huge supercomputers for doing complex simulations, and many scientists and engineers are producing and collecting massive amounts of data. Whether the data are observational, experimental, or come from simulation models, statistics provides a rigorous framework for drawing inferences from these data. But our ability to generate data is outpacing our capacity to transfer and store data. Our data sets are so massive—cosmology simulations, for example, can exceed tens of billions of multi-dimensional points—that we can't realistically keep it all; we have to pick and choose which data to store. Which are the most informative or valuable data points, and which can be deleted or processed into a more compact form? How do we make that call?

One way to handle massive data sets is by using an *in situ* approach, in which sampling and analysis are embedded in a simulation while it is still running. One such method that I collaborated on was designed to perform statistical calculations while a simulation is still running, in order to identify which of the resulting data are the most important, and to transfer and store those data, along with sufficient information to reconstruct an approximate representation of the complete set.

This approach can be used for different types of simulations. For example, what if an asteroid were to collide with the earth? We would need to know what happens, when it happens, and what happens next. In this hypothetical scenario, there are some time points when not much, or maybe even nothing is happening, such as before the collision or after, once all the ejecta have been launched onto well-established trajectories. We don't need to waste resources storing all of those data. But at other time points, such as the moments immediately following the impact, the behavior of the simulation is changing rapidly, so we would want to collect and store data frequently. We've been able to implement basic algorithms that run in real time to explore computer models developed by other scientists that use samples to approximate whole processes. This method also stores data in a smaller number of bits. There are two benefits to our approach: smarter time steps and more efficient data capture.

The race to exascale computing drives an increasingly urgent requirement to streamline our data storage and analysis solutions. This next revolution in supercomputing refers to the global goal of developing computing platforms that conduct over a billion-billion calculations per second. As the size and complexity of computer simulations continue to grow, our agility and creativity has to keep pace, which requires an understanding of sampling, error analysis, and design of experiments. Statistical methods built on these concepts are ever-evolving and advancing over time to keep up with the increasingly complex challenges of modern science.

OUR ABILITY TO GENERATE DATA IS OUTPACING OUR CAPACITY TO TRANSFER AND STORE DATA.

Sampling schema

A general challenge in experiment design is sampling, or choosing which data to collect out of all the data that could be collected. Many different types of sampling procedures have been proposed to address different types of scenarios. Traditionally, attention has focused on straightforward sampling scenarios, such as random sampling, in which each item is selected with equal probability from the entire population of interest; stratified random sampling, in which items are selected from different subsets of the overall population; and systematic sampling, in which items are selected systematically, according to some preset rationale. For one project, I used a systematic sampling plan to obtain information from Laboratory historical records (which were on microfilm!) that involved sampling records at equal intervals.

Over the years, sampling theory has enabled many other, more complicated types of sampling to be developed, studied, and implemented. Sampling theory is a subfield of statistics that provides a mathematical structure for understanding and probing statistical populations. It describes many different sampling procedures as well as the resulting estimates and associated uncertainties involved with using those procedures.

Observations from large and complex data sets may take the form of curves, spectra, or more general types of functions. In order to employ effective sampling approaches on these types of data, it is important to understand the population to be sampled, the sampling objective, and any practical constraints. Sampling strategies have evolved as increasingly complex problems have arisen, leading to new advances in the field of statistics. Customized sampling plans and procedures have played an important role in addressing institutional issues at Los Alamos including environmental remediation, analysis of historical records, and assessment of security procedures.

When problems become larger and the data become more complex, it's important to develop and apply sampling, design, and analysis methods that can provide representative samples and effective analysis results. With the rise of computational models and sophisticated estimation schemes, substantial growth has occurred in the use of specialized statistical techniques. Several highly sophisticated techniques that are now standard tools of statisticians worldwide have historical origins here at Los Alamos (perhaps most famously, Monte Carlo methods, a class of computational algorithms that use a statistical approach to solve complex problems).

Understanding uncertainty

Because of the uncertainty present in data and decisions, I sometimes feel uncomfortable making decisions. I often say that being a statistician means never having to be sure about anything, because there's always some amount of uncertainty. But, evasive maneuvering aside, uncertainty and error quantification are critical to any statistical analysis, and ultimately, decisions must be made even in the presence of uncertainty.

While the field of statistics focuses on uncertainty, the field of metrology focuses on measurement. Measurement error creates uncertainty, and has long been a concern of both statisticians and metrologists, requiring careful analysis of the process used to collect the data. These two fields really go hand in hand, as they each provide perspective on the sources of variability. For example, if we measure a variable x, what we observe is actually the sum of x plus an associated measurement error (which could be positive or negative). The overall error includes an intrinsic variation in the underlying variable—some values naturally shift around, such as the precise location of eddies circulating in an ocean—as well as the variability associated with the instrument and method of measurement. This type of assessment quickly gets complicated in situations with multiple measurements and multiple measuring devices. Despite its importance, measurement error is often underemphasized or even ignored. Measurement processes and their associated errors will interact and propagate throughout the analysis, sometimes compounding one another, so failure to consider errors associated with the measurement process can result in misleading analyses, even for basic methods, such as pairwise comparisons and linear regression.

WHEN A VARIABLE IS MEASURED, THE RESULT IS ACTUALLY THE SUM OF THE MEASUREMENT AND THE ASSOCIATED MEASUREMENT ERROR.

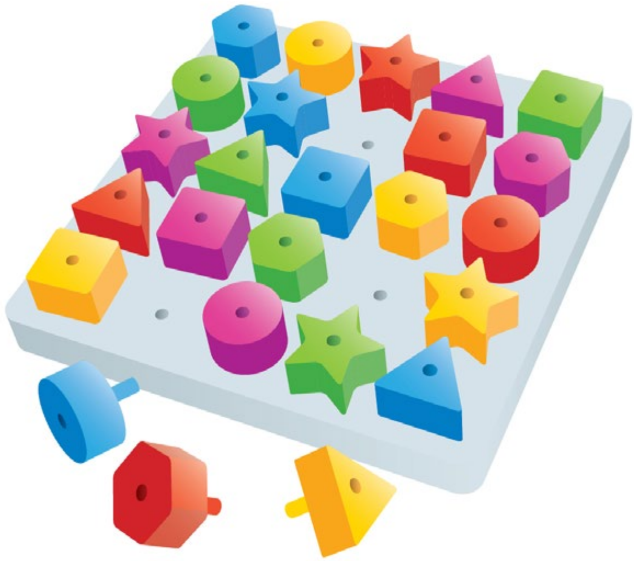
When carrying out statistical analyses, an important concept to understand is that sample variances—the typical plus or minus variation any one measurement might have with respect to the mean—are inherently much more variable than the means themselves. While I was working on my Ph.D. my advisor asked me to determine how many samples would be needed to obtain a coefficient of variation (standard deviation divided by mean) of 5 percent, for the variance of a normal distribution—the famous "bell curve" used to characterize many kinds of distributions, such as heights of different individuals in a class. The coefficient of variation is a measure of relative variability of a population, so the more samples there are, the lower the coefficient of variation. But I was shocked to find that 801 samples were needed to bring it down to 5 percent—although the number of samples required to estimate a mean depends on the specifics of the problem, a common rule of thumb is to take about 30 samples. This exercise made a huge impression on me as it so clearly illustrated the peril that lies in ignoring the inherent variability in sample variances, even for a distribution so well-defined and commonly used as the normal distribution.

The analysis of outputs from computer models is an increasingly important concern in scientific modeling. Ultimately, the uncertainty in measured data will have an impact not just on the analysis of the measured values themselves but also on subsequent analyses, where the measured data is used as input to models, and the associated variation is transmitted to the resulting model outputs. There are statistical techniques for approximating the behavior of complex computer models and associated discrepancy of these models from observed data. The traditional approach involves first fitting low-order polynomial expressions (as in the earlier example on aging materials), then subtracting the model predictions from the actual data values, and finally examining the discrepancies, or residual errors, left behind to see how the model can be improved.

New methods are being developed to fit more flexible types of curves, ones that may not be fit well by low-order polynomials. But as with the traditional approach, these more sophisticated methods can estimate an entire curve and reveal the error between the model predictions and the actual data points. This gives us a sense of how good our predictions are and how far off they might be from the true underlying values. Residual errors and discrepancy functions provide clues about what aspects of a model need to be adjusted to obtain a better fit between the model and the data.

Pursuit of patterns

My first introduction to experiment design, though I didn't know it at the time, was a puzzle I was given when I was about eight. The challenge of this puzzle was to arrange 16 colored shapes on a 4×4 grid such that each row and each column contained each color and each shape exactly once. I later learned that this puzzle was formally known as a 4×4 Graeco-Latin Square and represents an example of an experiment design. It is essentially a two-variable experiment where the variables are color and shape, each variable has four possibilities, and the solved puzzle contains every combination with no repeats. (The popular Japanese number puzzles Sudoku and Kakuro are also connected to experiment design



This puzzle toy illustrates an approach to experiment design. To solve the puzzle, each row and column must contain each color and each shape exactly once. The solution represents a two-variable experiment where the variables are color and shape, each variable has five possibilities, and the solved puzzle contains every combination.

methodologies. My family, which contains a higher-than-statistically-expected concentration of statisticians, named our two pet gerbils Sudoku and Kakuro.)

Advances in our understanding of design patterns support the development of new statistical designs to address increasingly complex problems. Agricultural experiments are an ancient and quintessential example of a simple, two-way experiment layout. The application of different treatments to rows and columns in a field, like different amounts of water, or different fertilizers, or variable spacing of seeds, has long been how growers learn to optimize their crops. The two-way layout that arises naturally from a field can be used first to represent two variables and then generalized to many variables. As we look back at how the discipline of statistical experiment design expanded beyond traditional two-way layouts like this, we see the evolution of the search for balanced patterns to obtain information about multiple experimental factors. This search for balanced patterns and the discovery of new or more complex patterns has always fascinated me.

My personal lifelong interest in patterns and arrangements has influenced and motivated much of my statistical work. When designing real-world experiments, there are often practical considerations and constraints that must be taken into account, which can, in turn, motivate new advances. For example, a couple years ago, I was involved in a project in which the need to generate a design for a physical experiment to learn about material properties eventually led to the development of a new algorithm. The algorithm generates matrices of experiment designs for studying computer models by looking at the outputs obtained by varying different parameters in the model. This design algorithm takes methods that can be carried out using traditional hand calculations and extends the functionality to computational experiments. This is particularly useful because these types of computational experiments, involving computer simulations that are run on very large computers, are used extensively for many different research projects at the Laboratory.

Women and statistics

With the rise of data science, interest in statistical concepts has expanded beyond traditional statisticians. To address the needs of this broader audience, I think some changes are needed in how statistics is taught. Instead of focusing on formulas, students should be taught to ask themselves, "What is the problem I'm trying to solve? What are the different ways I can look at this?" It's about statistical thinking—and often thinking about someone else's problem. But that's the fun part for me, digging into the worlds of other scientists, getting a taste of many different projects, being a part of all these cool ideas. Those are the things that sometimes go missing in the classroom. During the course of my career, I've been very involved in outreach activities aimed at getting students excited about math and science and helping them think like statisticians.

As a woman in science, and as the mother of three daughters with technical degrees, two of whom have chosen to become statisticians themselves, it came naturally to direct much of my outreach over the years toward activities sponsored by women's associations and professional groups. I've volunteered with ten-year-old girl scouts, participated in activities sponsored by the Los Alamos National Laboratory Women's Group, and enjoyed presenting workshops on statistics and careers in related fields. I value being able to support and encourage women, from kindergarten to early career, in pursuits both technical and practical.

It's no secret that women are underrepresented in technical fields. Some people find it surprising that statistics is one of the technical fields that has a higher proportion of women, about 30 percent, compared to 24 percent in computer and information science, 15 percent in engineering, and 11 percent in physics and astronomy, according to the National Science Board's Science and Engineering Indicators 2016 report. These numbers are climbing, with the proportion of women graduating in these fields much higher today than they used to be, but we're still the minority, and that can be difficult. I think that Los Alamos provides a stimulating work environment overall, but it's not without challenges. It's not uncommon, for instance, for me to find myself working in situations where I'm the only woman in the room. When an important decision is being made, technical or nontechnical, my opinion often comes from a minority perspective, and nearly always being outnumbered can be frustrating, but I don't let that dominate my experience. I have been fortunate throughout my career to work with great colleagues on exciting projects.

I believe the experience of women in science has improved significantly in the 25 years that I've been at Los Alamos, and I think that trend will continue. I'm seeing more women in management positions, more technical positions filled by women, and more opportunities for women to advance their technical careers. I think it's important to acknowledge the challenges women in science face while still fostering enthusiasm and a sense of discovery about science itself. It's that excitement after all, that draws us all here!

—Joanne Wendelberger