Customer Story

photo by Brandon Semel

Virginia Tech

Challenge

Graduate students in statistics need experience applying their knowledge to practical situations. Graduate students in other disciplines need help making their research more statistically robust.

At Virginia Tech, collaborative research leads to a better understanding of species ecology

An interdisciplinary research effort sheds light on lemur behavior with the aim of improving conservation strategy and captive rearing outcomes.

Scientists + statisticians = discovery. It's simple math. So says Jiangeng Huang, PhD candidate in statistics at Virginia Tech. And Huang would know; over the past three years, he has collaborated with researcher Brandon Semel, a PhD candidate in fish and wildlife conservation also at Virginia Tech, to explain the ecological function of soil consumption in wild lemur populations. Their combined efforts have revealed a number of biologically important variables representing an expansion of scientists' understanding of lemur behavior.

From fieldwork to robust statistical modeling

Primates have long been known to consume soil – a behavior known as geophagy – says Semel of his research; however, scientists have yet to explain why this behavior evolved and what purpose it now serves. Existing hypotheses take into account a combination of nutritional, geographic and demographic factors and range from parasite mitigation to mineral supplementation to flushing toxins. Cracking this question could be the key to understanding how conservationists can better direct their management decisions to increase the likelihood that the species will survive.

"Forests in Madagascar are rapidly disappearing, threatening many of the over 100 lemur species that depend on them with extinction," Semel says. "Learning more about lemur dietary needs will help us to focus conservation efforts on the best habitats that remain, provide the best care in captive breeding programs and guide reforestation efforts." While in the field in Madagascar, Semel records data such as bite rates, food quantity, food type and the time spent consuming different foods and soil. This data then gets converted to nutritional information that can be more easily studied in the lab: minerals, fat, calories, fiber, protein.

Returning to Blacksburg from one such trip three years ago, Semel found himself facing several data challenges for which he was ill-equipped; the data set was large, the variables multicollinear and the missing values many. In order to augment the quality and precision of his findings, Semel sought a partnership through the Statistical Application and Innovation Group (SAIG), a Virginia Tech Department of Statistics initiative.

A university initiative provides applied experience, better science

Offering statistical collaboration, consulting and support for research scientists in other disciplines across the university, SAIG is the embodiment of the scientists + statisticians = discovery formula. The group's objective is to assist researchers in designing more robust experiments, honing their data modeling and analysis skills, and teaching them to use the software they need to carry any newly acquired statistical skills with them into future research. In return, graduate students in statistics gain an opportunity to apply their expertise to real-world scenarios – valuable experience for any new graduate looking to pursue an industry career.

Since beginning his studies at Virginia Tech – and joining SAIG – Huang has worked on more than 60 projects with researchers from fields as disparate as forestry and civil engineering. "We listen to and learn about our colleagues' research context and then figure out the most appropriate algorithms to use to solve their problem," Huang says. "We just teach them how to fish, and they can then fish by themselves."

JMP makes it so easy for researchers to analyze their own data. I think that's the really cool thing about JMP.

Jiangeng Huang, PhD candidate in statistics

When Semel first turned up at SAIG, he knew a bit about fishing; that is, he knew about linear models but not much about how to apply them. Huang, on the other hand, knew nothing about lemurs. That soon changed, and the two have now become trusted colleagues and good friends. Collaboratively, they make a formidable duo.

JMP® simplifies the stats

After sitting down with Semel to better understand the goals and parameters of his research, Huang began exploring the data. Together, they hashed out an approach which eventually led them to build a series of regularized linear models that allowed them to, in one step, select important features and estimate their effects. By combining lasso and ridge regressions with the elastic net, they were able to address the multicollinearity issue. And to build these models, they used JMP.

"I like to use JMP because it's so interactive," Huang attests. "It's very user-friendly, and there are so many tools available." Also, he says, because users don't need to learn to write code, JMP makes statistical methods more approachable for scientists like Semel who need to free up time to focus on fieldwork. "Researchers from other disciplines all have different levels of programming skill and statistical knowledge. Many ongoing research projects are interdisciplinary in nature and require effective collaboration from both domain expertise and statistical computing," Huang says. And JMP makes this process easier. "We want to try many different methods to figure out which one works best for the problem at hand. I can't spend 10-plus hours coding things from scratch for every potential method we want to try, when I am also working on multiple research and consulting projects at once."

That's why the SAIG initiative aims to not only equip researchers with a knowledge of the basic principles, but to teach them how to use JMP. With the wide-ranging prepackaged options available in the software, researchers don't need any other tool, Huang says. "JMP makes it so easy for researchers to analyze their own data. I think that's the really cool thing about JMP."

Semel has an especially complex data set. Because zero values are recorded in any instance in which lemurs are observed eating food but not soil – and not all lemurs eat soil – this means that as much as half of any given data set can be zeros. "This data set has a special structure in its response variable," Huang says. "We call it zero-inflated. If you look at the response variable, there are counting data with many zeros. It's a non-normal data set, and we handled that in JMP by extending to a generalized linear model using our link function. We also need to be creative with the link function and in this case, we found a zero-inflated Poisson model works well." Dealing with "non-normal" data sets can be a challenge even for a trained statistician. But knowing that this scenario would certainly not be the last in Semel's career in the field, Huang worked with him to develop a standard method that would be easily replicable and adaptable in JMP.

Experience that prepares for the world ahead

"I believe this collaboration trend is a very strong part of our program," Huang says of SAIG. "Statistics is very interdisciplinary and SAIG allows us (statisticians) to see how these statistical methods are applied to real problems." It all comes back to the collaboration paradigm, he says. "You advance research, but you're also learning in this process; you're learning skills that you can apply when you move on to do something that's not pure statistics. And the research itself is interesting. Working together, you come up with a solution."

While Semel continues to wrap up his research on lemur geophagy, he is now looking at how lemur populations will be affected by climate change. "We know that climate change is already causing shifts in forest types around the world," he says. "By studying lemur populations across a range of forest types, I am hoping to predict future lemur abundance based on current protected area coverage and various climate change scenarios." Adds Huang: "We all have different talents. That's why I like to collaborate. I chose statistics because I can play in everyone's backyard."

Solution

Virginia Tech's Statistical Applications and Innovation Group (SAIG) pairs graduate-level statisticians with scientists in other fields to address research questions collaboratively. SAIG consultants use JMP® to make statistics more approachable.

Results

Two Virginia Tech doctoral students – a statistician and an ecologist – explored out-of-lab nutritional data in JMP to identify biologically meaningful variables that help explain why wild lemurs consume soil.

To contact your local JMP office, please visit: jmp.com/offices

THE POWER TO KNOW®