President Jeff Gold
Chancellor Rodney Bennett
Vice Chancellor Tiffany Heng-Moss
Academic Planning Committee Coordinator Karen Griffin

Dear President Gold, Chancellor Bennett, Vice Chancellor Heng-Moss, and Coordinator Griffin,

As a professor of statistics at Stanford University, I am writing to convey my unequivocal support for retaining the University of Nebraska Statistics Department, whose indispensable contributions to teaching, research, and interdisciplinary collaboration are especially vital in this era of rapidly expanding demand for data analysis and data science across every major university campus.

**Distinction of Department**

The University of Nebraska has a distinguished record, which I believe must not have been adequately captured in whatever statistics you may have been given; and their work has been valuable to me personally. In fact, during the pandemic, I spent a great deal of time studying publications from the University of Nebraska's Statistics Department. I later invited faculty members from University of Nebraska to speak at the Stanford  statistics seminar and also arranged meetings with people in the health sciences industry who were interested in  using group testing ideas from which the Statistics Department of the University of Nebraska is a pioneer to solve an urgent problem — scaling Covid tests to larger user populations when few PCR machines are available. This type of work will be essential the next time we have new public health emergencies.

**Growing Need for Data Science**

The biggest positive happening nationally in undergraduate instruction in the last decade is the growing demand among undergraduates for training in Data Science. Around 2010 we saw the arrival of "Data Science" as a known hiring category for tech company headhunters; in response universities all around the nation established data science master's programs. Over the last 15 years, well over 250,000 undergraduate and masters students have been employed as data scientists, at salaries exceeding even computer scientists.

I have been able to study the "Data Science Phenomenon" in depth, as I served on the National Academies panel on the Undergraduate Data Science Curriculum in 2017-2018, and have written a few well cited papers on Data Science including "50 Years of Data Science" (2017) and "Data Science at the Singularity" (2024). I teach at Stanford both in the Statistics department and in the Data Science program.

Statistics departments have been teaching Data Science long before there was this current enthusiasm; and have been inventing and systematizing many of the key innovations long before they were ever discussed in Computer Science or in Mathematics Departments. For example, they have been studying how to clean, wrangle and process messy data for decades and came up with many important principled ways to deal with messy and contaminated data long before Data Science existed as a term of art. All working data scientists will tell you that such issues are central to actually getting any work done with real data. On the other hand, Mathematics and CS departments tend to emphasize non-central aspects. For Mathematics the blind spot is often the emphasis of certain theorems at the expense of the core data analysis workflow and the various issues that come up in actually getting valid real-world answers – in particular choosing a model that actually matches the behavior of the data and wrangling with messy data. In CS the blind spot is often the emphasis of hyper-scaling of very simple algorithms, which is very attractive to CS Faculty, but is not a real concern for the vast majority of professional data scientists; in fact the vast majority of datasets that professional data scientists and academic scientists face are not of the super-huge variety that CS faculty imagine. There really can be no doubt that trained Statisticians are truly the only people on campus who have the full picture of data analysis as a human activity and of serving the actual needs of professional data analysts.

It seems totally uncalled-for to demote such a qualified department excelling at serving the actual needs of scientists on campus at such a time of booming student demand. Moreover, there is plenty of experience with the proposed new administrative alignment and it's well understood that this will be worse for the campus and worse for students.

Let me step back and take a broader view.

**The prior history – why the proposed arrangement is not in the campus' best interest.**

A central development in scientific publishing and scientific procedure over the last century has been the onset of data rich studies and the deployment of concepts and tools involving rigorous

statistical methodology for data analysis. These concepts and tools have largely been invented, carried out, promoted, and taught in statistics departments nationwide and worldwide. Today there are hundreds of thousands of papers yearly in clinical medicine and computational science that publish their results and clinch their arguments with statistical tests and procedures.

Although rigorous methods are demanding, difficult and in certain senses forbidding, they bring great benefits in terms of research efficiency. Instead of inconclusive research, in which we have no confidence, we get actual conclusions in which we have confidence. Still, some non-statisticians who have not been trained in these concepts and methods may not be aware of the threats to scientific validity which arise if the rigor is allowed to slip away. In fact, in the last decade we have become able to peruse the entire body of scientific literature as a dataset, and to recognize the uneven quality and rigor of studies, as well as the proliferation of poor studies, which contaminate the corpus and now begin to undermine public confidence and funding.

Practically a century ago, statistics departments began to be formed around the country, as it was understood that rigorous methodology for data analysis would be better developed and nurtured if there were an institutional home rather than a distributed, scattered effort across campus.

This Statistics Department model was eventually widely understood and propagated throughout the postwar era, creating great advances in data analysis for clinical medicine, social sciences and physical sciences. Over the last 80 years, there have been many advances that are fundamentally statistical in nature and that have been recognized at the highest level, including Nobel Prizes.

The effort now apparently underway at University of Nebraska—to disestablish the statistics department and disperse and/or scatter data analysis throughout campus—is a throwback to 100 years ago and will only land us in the same situation as we encountered then.

Historians have looked into the practice of statistical data analysis during that earlier era. For example, Stephen Stigler of University of Chicago has an excellent article called "The History of Statistics in 1933". He showed that in those days there were earnest and hard-working data analysts who were not schooled in the kinds of patterns one might see in noisy data that had explainable causes recurring again and again from study to study. Instead, there was a great deal of wasted time and effort in those days, when people just didn't understand what they were seeing. They either thought they were seeing something real that was only an artifact, or they were ignoring patterns and tendencies that were fundamental and very important but were overlooked through lack of the right tools or simple awareness.

In fact, the situation of 100 years ago is how we got Statistics Departments in the first place. Harold Hotelling wrote an article in the early 1940s in Annals of Mathematical Statistics in which he specifically called out the so-called distributed model and explained what the bad effects of that model on university instruction and research.

The lack of lifelong training of faculty devoted to rigorous data analysis and its methodology on campus means that it will be difficult for researchers seeking large interdisciplinary grants to find data analysis experts on campus and for graduate students, post docs etc. to find rigorous training. And that means that University of Nebraska will be at a disadvantage in competing for grants against other universities that continue to support strong, centralized statistics programs. Those programs produce an identifiable resource on campus that offers data analysis experts who can support work on campus funded by NIH, NSF and DOD.

Thanks very much for reading this far, and please think carefully about the fact that this model has been tried before on many campuses and found wanting, in numerous ways. The current "statistics department model" has proven itself time and again on campus after campus.

I fully support the retention of the University of Nebraska Statistics Department, a group of noted scholars doing important work for which I have respect and which I have directly engaged with. I think it's particularly important to maintain this department at this time of booming demand for data analysis and data science everywhere on every major university campus.

Sincerely,

David Donoho
Anne T. and Robert M. Bass Professor in the Humanities and Sciences
Professor of Statistics

**References**

Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745-766.

Donoho, D. (2024). Data Science at the Singularity. *Harvard Data Science Review*, 6(1). https://doi.org/10.1162/99608f92.b91339ef

Stigler, S. M. (1996). The History of Statistics in 1933. *Statistical Science*, 11(3), 244–252. http://www.jstor.org/stable/2246117

Hotelling, H. (1940). The teaching of statistics. *The Annals of Mathematical Statistics*, *11*(4), 457-470.