The University of Arizona

THE UNIVERSITY OF ARIZONA
Statistics & Data Science
Graduate Interdisciplinary Program

Academic Program Review

Self Study Report

January 2024

Statistics Program Self Study Committee

Helen Zhang
Edward Bedrick
Maria Sans-Fuentes
Joe Watkins
Jian Liu
Melanie Bowman

Contents

SECTION A: SELF-STUDY SUMMARY

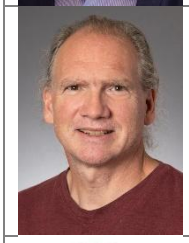
A.1 Administrative home of the Statistics and Data Science Graduate Interdisciplinary Program (GIDP)

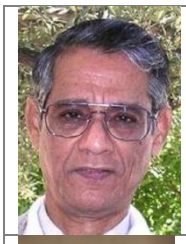
The Statistics and Data Science Program is one of 20 Graduate Interdisciplinary Programs that are housed in the Graduate College.

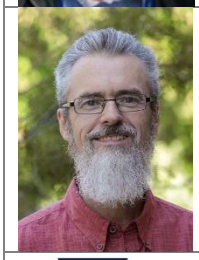
A.2 Faculty

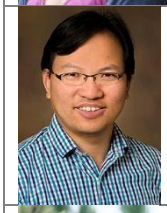
The GIDPs are designed to provide graduate training opportunities and consequently membership is only extended to those who can direct theses or dissertations. Thus, the GIDPs have no lecturers, adjunct instructors, and post-doctoral fellows.

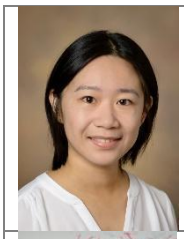
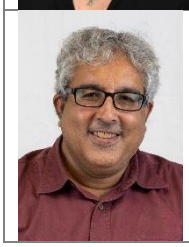
Statistics & Data Science Faculty Fall 2023

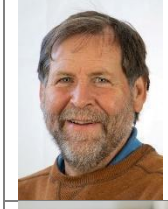
	Name	Title/Department	Research
	Lingling An	Member of Executive Committee Associate Professor, Biosystems Engineering	Statistical genomics/metagenomics; Single cell data analysis; Data mining and machine learning
	Satheesh Aradhyula	Associate Professor Agricultural-Resource Economics	Production Economics; Econometrics; Agricultural Policy
	Kobus Barnard	Professor, School of Mind, Brain and Behavior	Machine learning; Mathematical modeling of geometric form; Multi-modal data; Statistical applications in computer vision
	Edward Bedrick	Member of Executive Committee Professor, Public Health	Analysis of observational data; Bayesian methods; Generalized linear and mixed models

	Rabindra Bhattacharya	Professor Emeritus, Mathematics	Markov processes; Large sample theory; Statistical shape analysis; Economic theory of growth under uncertainty
	Dean Billheimer	Professor, Epidemiology and Biostatistics	Measurement and normalization; Quantitative proteomics; Statistical methods for compositional data
	Alexander Bucksch	Associate Professor School of Plant Sciences	Computational plant science; Morphological plant modeling; Plant shape; Plant simulation; Plant modeling
	Zhao Chen	Professor Epidemiology and Biostatistics	Research study design; Longitudinal data analysis; Risk assessment.
	Jianqiang Cheng	Assistant Professor Systems and Industrial Engineering	Stochastic Programming; Semidefinite programming and applications
	Michael Chertkov	Professor Mathematics	Energy systems; Data science/machine learning; Statistical physics; Control and optimization theory; Fluid mechanics
ARIZONA	Melinda Davis	Adjunct Research Assistant Professor, Psychology Research Assistant Professor, BIO5 Institute	Latent variable modeling; Measurement of change; Item response theory; Health outcomes research; Statistical consulting

	Hongxu Ding	Assistant Professor Translational Pharmacogenomics	Single-cell analysis; Nanopore sequencing; Computational biology approaches to interpret single-cell omics profiles and nanopore sequencing readouts
	Neng Fan	Associate Professor Systems and Industrial Engineering	Methodologies in Optimization; Applied Operations Research; Data Mining and Machine Learning
	Ryan Gutenkunst	Associate Professor Molecular and Cellular Biology	Computational biology, with focus on inferring history and natural selection from population genomic data and on understanding the function and evolution of complex biomolecular networks
	Ning Hao	Member of Executive Committee Associate Professor Mathematics	High dimensional data; Machine learning; Change point detection
	Jake Harwood	Professor of Communications	Applied statistics in the social sciences; Hypothesis testing; Moderator and mediator effects.
	Chiu-Hsieh Paul Hsu	Professor College of Public Health	Survival analysis; Missing data; Statistical modeling.
	Chengcheng Hu	Professor College of Public Health	High-dimensional data; Survival analysis; Longitudinal data; Missing data; Measurement error.
	Bonnie Hurwitz	Associate Professor Biosystems Engineering	Systems Biology; Functional Metagenomics; Big Data; Bioinformatics; Computational Biology

	Afrooz Jalilzadeh	Assistant Professor Systems and Industrial Engineering	Design, analysis, and implementation of stochastic approximation methods for solving convex optimization and stochastic variational inequality problems with applications in machine learning, game theory, and power systems
	Kwang-Sung Jun	Assistant Professor Computer Science	Machine Learning
	Nicole Kersting	Associate Professor College of Education	Measurement and educational assessment; Item response and generalizability theory; Value-added models; Random effects models
	Haiquan Li	Assistant Professor Medicine	Biomedical big data science, Translational bioinformatics; Bioinformatics; Clinical informatics
	Jingjing Liang	Assistant Research Professor, R Ken Coit College of Pharmacy	Statistical genetics and genomics, including computational methods for analyzing large-scale sequencing data; Rare variant association analysis; Genomics-driven drug target discovery; Precision medicine
	Kevin Lin	Associate Professor Mathematics	Stochastic nonlinear phenomena in biology and physics, especially problems from computational neuroscience and nonequilibrium statistical physics; Monte Carlo algorithms; Numerical methods for stochastic differential equations; Scientific computing
	Yifeng Lin	Associate Professor Epidemiology and Biostatistics	Meta-analysis; Network meta-analysis of multiple-treatment comparisons; Publication bias, Bayesian methods; Statistical applications
	Jian Liu	Associate Professor Systems & Industrial Engineering	Integration of manufacturing engineering knowledge, control theory and advanced statistics for product quality and productivity improvement; System prognostic/diagnostic modeling and analysis

	Yiwen Liu	Assistant Professor Epidemiology and Biostatistics	Dimension reduction and variable selection; Big data analytics; Data integration
	Clayton Morrison	Professor School of Information	Machine Learning, Causal Inference, Activity Recognition and Understanding, Automated Planning, Knowledge Representation, Computational Cognitive Science
	Selena Niu	Associate Professor Mathematics	Nonparametric statistics; Semiparametric modeling; Statistical genetics
	Jason Pacheco	Assistant Professor Computer Science	Statistical machine learning; Probabilistic graphical models; Approximate inference algorithms; Information-theoretic decision making
	Walter Piegorsch	Professor Mathematics	Data analytics; Genomics; Environmental statistics; Risk assessment; History of statistics.
	Denise Roe	Professor Public Health	Biostatistics. Clinical trials; Epidemiological studies; Pharmacokinetics.
	Liliana Salvador	Assistant Professor Animal & Comparative Biomedical Sciences	Multidisciplinary approach to study the dynamics of zoonotic infectious disease;. Computational, mathematical and data-driven models to understand the ecology and evolution of infectious diseases at the wildlife, livestock and human interface
	Sunder Sethuraman	Professor Mathematics	Probability and stochastic processes; Bayesian stick-breaking and applications; Models in statistical physics

	Robert Steidl	Professor Natural Resources	Quantitative ecology; Dynamics of animal populations; Conservation biology.
	Xiaoxiao Sun	Assistant Professor Epidemiology and Biostatistics	Medical Imaging; Nonparamatric Modelling; Computational Biology
	Xueying Tang	Assistant Professor Mathematics	High dimensional Bayesian statistics; latent variable models and their application in Education and Psychology
	Bruce Walsh	Professor Ecology and Evolutionary Biology	Biostatistics; Statistical genetics/genomics; Mixed models; Bayesian analysis; Resampling and MCMC methods
	Joseph Watkins	Vice-Chair Executive Committee Professor Mathematics	Limit theorems; Statistical applications in life sciences
	Erfan Yazdandost Hamedani	Assistant Professor Systems and Industrial Engineering	Methodologies in Optimization: Saddle point problems, Distributed Optimization, Bilevel Optimization; Applications: Machine Learning, Data Science, Artificial Intelligence
	Chicheng Zhang	Assistant Professor Computer Science	Machine Learning; Learning theory
	Hao (Helen) Zhang	Program Chair Professor Mathematics	Statistical machine learning; Nonparametric smoothing; Model selection; Statistical applications in biosciences and biomedicine

A.3 Academic programs

The Graduate Program in Statistics and Data Science has six distinct academic programs – the graduate certificate (available both as a residential and an online program), the Master's degree, the accelerated Master's degree, two tracks for the doctoral degree, the regular track and the informatics track, and a PhD minor. The accelerated Master's degree was established in 2014, the Online Certificate Program in 2013, and the doctoral informatics track in 2012. Enrollment numbers can be found in **Section I.3**.

Statistics and Data Science GIDP does not have undergraduate programs. However, it works closely with the Department of Mathematics for the undergraduate Statistics and Data Science degree program. Also, the new Professional Master's Degree in Data Science and Applied Statistics is hosted by the Departments of Mathematics and Computer Science with core courses offered by computer science, biostatistics, and mathematics. Several core courses will be developed and taught by Statistics and Data Science faculty. (For more, see **Section C.2: Major Changes in the past seven years.**)

SECTION B: UNIT DESCRIPTION AND GOALS

B. 1 Mission, role, and scope

Role and Scope

The University of Arizona has chosen to coordinate graduate education in statistics through a Graduate Interdisciplinary Program (GIDP) in Statistics and Data Science (SDS). This approach, unique to Research I Universities, has the advantage of flexibility in meeting the rapidly evolving research needs of a large land grant University. This structure enables students of Statistics and Data Science to design and complete research projects that not only apply the science of statistics to a broad array of data-intensive research questions, but also contribute to the creation of statistical theory and methods suitable for modern needs.

The Interdisciplinary Program in Statistics and Data Science meets the workforce and research needs through its Certificate Program, master's degree, an Accelerated Master's Degree, a doctoral minor, and a doctoral program that has both a regular track and an informatics track.

In this program, doctoral students complete their foundational training in Statistics and Statistical Informatics with a minor from many possible choices. The Graduate Committees play a highly interactive role in working with students. Each Committee is often led by co-mentors, one being core statistician and the other application area researcher.

This is summarized in the Program's mission:

Mission

> *The mission of the Graduate Program in Statistics and Data Science is to provide an environment whereby students become independent researchers and practitioners who make significant contributions at the forefront of knowledge across the disciplines that rely on statistical thinking. By merging data science approaches with practical innovation, our students are prepared to make fundamental advancements in both statistical theory and statistical methodology. Program members are dedicated to bringing theoretical, methodological, and applied expertise in statistics and data science through course offerings, student mentoring and advising, and research collaborations. This results in an extensive coordination across the campus' statistical and data science graduate curricula, course offerings, student involvement, and in the overall support of the University's mission to educate and train the next generation of data scientists.*

B. 2 Major goals or strategic directions for the next five years

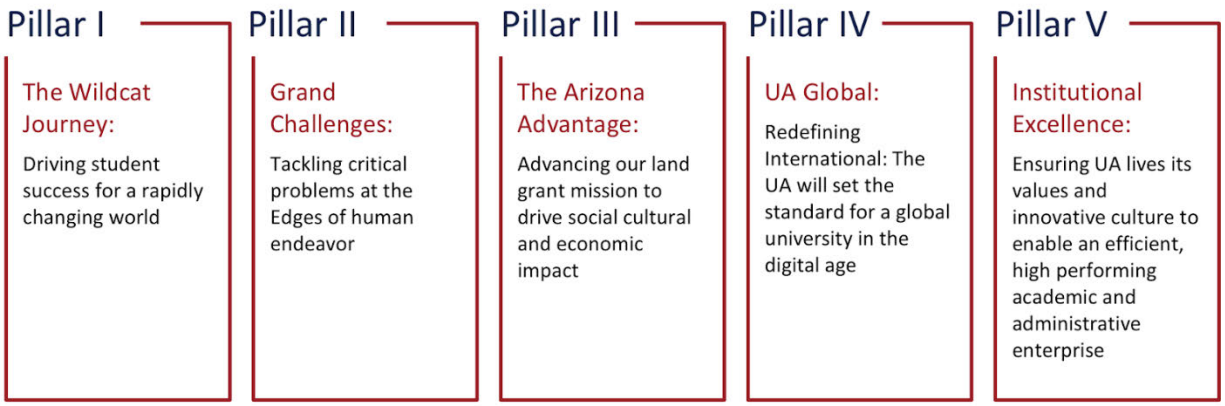
The Graduate Interdisciplinary Program in Statistics proposes the following six major goals

1. **Reach parity of peer institutions in funding for graduate students - both in funding amounts and in the time of guaranteed support.**
2. **Ensure financial robustness and adequacy for the operational budget, thereby enabling the strategic growth and scalable expansion of the GIDP program**
3. **Stabilize the existing graduate curriculum while working with key departments to continuously modernize the curriculum.**
4. **Continue to increase research capabilities in statistics and data science by bringing new faculty to the University and integrating them into the Graduate Interdisciplinary Program.**
5. **Promote the visibility and the participation of the Members of the Interdisciplinary Program in Statistics and Data Science at the University of Arizona campus in major activities involving statistics and data science.**
6. **Raise the national visibility of the Interdisciplinary Program in Statistics and Data Science through highlighting its unique positioning among Statistics and Data Science Graduate Programs.**

Before addressing in detail the goals and strategic directions for the next five years, we first look at the Statistics and Data Science Program in the context of the University's Strategic Plan.

B. 3 Relationship of program's goals to the University Strategic Plan

The current University of Arizona Strategic Plan was adopted in the fall of 2018. The Plan "is inspired by the Fourth Industrial Revolution – a time of augmented intelligence and the fusion of digital, physical and biological worlds." The Fourth Industrial Revolution is the notion that industrial changes will be powered by the joining of new technologies that blur the lines between the physical, digital, and biological worlds. Statistics and Data Science sit at the intersection of these big trends. The Interdisciplinary Program plays a significant role in this context. The diverse workplace environments that our alumni have entered are evidence of this. For example, our alumni now work in academia with tenure track positions (Oberlin College, Rochester, Cal State, Chico), instructional faculty (Northern Arizona, Wisconsin, Vermont), or postdoctoral fellows (Penn, NYU, Tennessee, UArizona), in biomedical science (AbbVie, Fred Hutchinson Cancer Center, Inari Medical, Allen Institute, Merck, Critical Path Institute, Johnson & Johnson, Food and Drug Administration), banking and finance (USAA, New York Community Bank, Quicken), entertainment (Cincinnati Reds, Arizona Diamondbacks, CBS Sports, Warner Brothers), and social media/computer services (TikTok, Zscaler)



The Strategic Plan for the University of Arizona is divided into 5 Pillars that represent broad focus areas. These pillars work in concert and collectively act as a robust roadmap for the UArizona's future (See figure above).

The Statistics and Data Science Program is evaluated in the context of the University's Strategic Plan, and its most visible places are found in Pillar II: Grand Challenges. As the Plan notes, the University of Arizona is an international leader in optics, space science, and environmental sciences. The University is both a land grant institution and a Hispanic Serving Institution (HIS). Our faculty either reside or collaborate with researchers in all of these three areas. Subsequent modifications to the plan led to the creation of the Arizona Space Institute, the Arizona Institute for Resilience, and the Data Sciences Academy.

The Data Sciences Academy is the initiative most directly related to the mission of the Interdisciplinary Program in Statistics and Data Science. As originally conceived, the Academy was designed to become a university-wide endeavor to coordinate and lead in all areas of the data sciences and all aspects of university research, education, and outreach. However, due to funding reductions induced by the COVID-19 pandemic, the scope of the Academy was limited, preventing it from fostering the level of activity envisioned in the Strategic Plan.

The rationale and action plan for each of the five goals are briefly outlined below. These goals - improving student support, stabilizing and modernizing curriculum, integrating program faculty into the life of the program, and raising visibility using the unique nature of the Arizona program - are not only interconnected but complementary. As a consequence, these goals should be addressed in a comprehensive manner by the members of the Statistics and Data Science Program in coordination the University's central administration, i.e., the Office of the President, the Provost's Office and the Office of Research, Innovation, and Impact and the Deans of the Colleges where the Program Members reside.

1. **Reach parity of peer institutions in funding for graduate students - both in funding amounts and in the time of guaranteed support.**

   Rationale The Statistics and Data Science Program is placed outside of the usual departmental structure and as an Interdisciplinary Program inside the Graduate College. This decision has advantages and disadvantages for students in the Program. The clear advantage is that students can easily enroll in statistics and data science courses and find research opportunities across many sectors of the University of Arizona. However, these sectors have little incentive to contribute to the funding of our graduate students, especially in the early years. Moreover, at peer institutions, the normal recruitment package for graduate students in Statistics has a guarantee of at least 4 years of 0.50FTE support for students who maintain adequate progress. Additionally, the typical level of financial support is >20% higher than the support available for UA students.

   For the 2015-2016 Self Study, we noted that the Statistics and Data Science Program had 7.5 guaranteed teaching assistantships with the 2 online TAships terminated shortly thereafter. Today, this number has increased to 13. (See table below.) A significant portion of this increase is attributable to the creation of the undergraduate degree in Statistics and Data Science within the Mathematics Department. This degree program now has ~50 graduates per year. In addition, the partnership with the Capital University of Economics and Business in Beijing, China hosts a University of Arizona undergraduate degree that is ramping up to >100 graduates per year.

   Action A sizable increase in the number and range of course offerings and degree programs in statistics and data science ought to be a more stable source of funding for our students. This increase in funding can occur through increased needs in existing graduate programs, the new professional master's degree, and the new partnership with the Capital University of Education and Business in Beijing (See **Section C.2 Major changes in the past seven years**) In addition, we will work with units across campus that teach introductory statistics. The SDS GIDP need to reach a long-term agreement on the number of TAs with the Mathematics Department that hosts the majority of the statistics courses as well as the undergraduate and Professional Master degree programs in data science. **The goal is to secure 20 0.5 FTE teaching assistantships positions over the next five years.**

   Our SDS faculty membership has been growing steadily, and faculty joining the program are presumably looking to become research advisors. Since a majority of our Ph.D. students are hired as Teaching Assistants for statistics courses by the Math Department, we do not require the faculty to support them as an RA before they agree to be a mentor. While support for students is not required, we highly encourage the SDS faculty advisors to provide RA support for senior students by working with the GIDP Executive Committee on 1) matching students to align with the faculty research interests, 2) providing partial funding to jointly support students, or 3) engaging students in grant applications for additional support. This approach with the goal to

support an increasing number of our students does not need to be limited to SDS faculty members. Many distinguished research groups on campus (for example, hydrology, precision agriculture, planetary science and astronomy, environmental science, and precision health) have big data projects that would benefit from having graduate students on their teams.

Further, collaboration with our companion graduate programs (Mathematics and Applied Mathematics) in the School of Mathematical Sciences can establish a coherent program for assistantship support in industry and the national labs.

Statistics & Data Science GIDP
Student Funding History

		Fall 2017	Fall 2018	Fall 2019	Fall 2020	Fall 2021	Fall 2022	Fall 2023
Math TAs	Number	6	7	9	7	10	11	14
	FTEs	2.75	3.5	4.5	3.5	4.5	5.5	6.75
RAs	Number	8	6	4	7	6	11	8
	FTEs	3.58	3.25	2	3.25	3	5.25	3.75
Online	Number	2	2	2	0	0	0	0
	FTEs	1	1	1	0	0	0	0
GIDP	Number	2	2	2	2	3	2	3
	FTEs	1	1	1	1	1.25	1	1.25
RA Sources		Medicine NIH Grant University Biostats Medicine	Biostats NIH Grant Medicine Tripods Internship MCB	Tripods Tripods Medicine Medicine	Medicine Biosys Eng RTG SIE Ed Psy	RTG SIE Sch of Info Ed Psy BIO5 Ed Psy	Cancer Math Health Svc Biosys Eng RTG Medicine Ed Psy BIO5 Pharmacy Biostats	Education Stat Lab BioStats Pharmacy HPC MCB Physics
Total Number		**18**	**17**	**17**	**16**	**19**	**24**	**25**
Total FTE		**8.33**	**8.75**	**8.5**	**7.75**	**8.75**	**11.75**	**11.75**

	Academic Year							
Budgeted Amounts	Proposed 24/25	23/24	22/23	21/22	20/21	19/20	18/19	17/18
Grad College Fellowship (GCF)	$30,000	$50,000	$20,000	$21,450	$22,000	$22,000	$22,000	$22,000
Grad Tuition Scholarship (GTS)	$55,000	$80,000	$65,000	$71,934	$71,934	$71,934	$75,000	$75,000
GIDP Funded TA Positions (FTEs)	1.25	1.25	1.00	1.25	1.00	1.25	1.00	1.00

Graduate College Funding History 2017 to 2023

The Statistics and Data Science program faculty members and especially its Executive Committee will establish a coherent plan for recruiting research assistantship (RA) and to look for creative ways to combine 0.25FTE TAships with fellowship (GCF) and tuition

(GTS) money as seed money for grants or to extend the funding period for our graduate students.

Student financial support at the University of Arizona is not competitive with our peer institutions. (See **Section I**.) We have seen this financial difference be determinative for the choice of graduate programs. Our administrative leadership is aware of this circumstance, and we will continue to work with the Graduate College to increase the GA base salary university-wide to make our program more attractive among the peer institutes. Marginal improvements can also be made through creative combinations of fellowship support and summer stipends to improve our financial offerings.

The SDS program is always exploring avenues to diversify its applicant pool. In this regard, the Associate Dean of Student Affairs, Diversity & Inclusion has been consulted in recruiting whenever a promising candidate applies. We are actively pursuing scholarships for underrepresented groups through an S-STEM Planning Grant. On the basis of information gathered for this grant, the Program is a part of the team in the School of Mathematics Sciences that will submit an S-STEM grant in March 2024. Our partners, the University of California, Merced, the University of Texas, Rio Grande Valley, and the University of Texas, San Antonio, are all HSIs. The main goal of the S-STEM program is to enable low-income students with academic ability, talent or potential to pursue successful careers in promising STEM fields, which in our case is the mathematical sciences.

With the first class of graduates with undergraduate degrees in Statistics and Data Science, we can now do *in house* recruiting and as a part of the S-STEM, recruiting at southwestern HSIs.

Additionally, we will continue to seek partnerships with GIDP affiliated faculty in other departments to seek federal funding for joint training programs that provide RA support. For example, the NSF TRIPODS program supported numerous undergraduates, graduates (as RAs) and three postdocs each year. The current NSF RTG (Research Training Groups in the Mathematical Sciences) program offers research training and support for students in structured research groups.

Using all of these approaches, we can anticipate a graduate enrollment of 40 doctoral students.

2. Ensure financial robustness and adequacy for the operational budget, thereby enabling the strategic growth and scalable expansion of the GIDP program

Rationale: The Graduate College directly allocates the yearly operational budget for the SDS program, which is pivotal for its day-to-day functions. This budget encompasses a range of activities, including organizing colloquiums, hosting visitors, facilitating student meetings and journal clubs, and providing travel funds for graduate students and faculty. Additionally, it supports the organization of community meetings, and the

sponsorship of local conferences and events. The consistent stability and sufficiency of this budget each year are crucial to ensuring the program's seamless operation and expansion, while also contributing to its increasing quality and improving its ranking.

Action: Recently, the Graduate College has revised its budget allocation formula, now calculating a significant portion of each program's budget based on the course revenue generated for the College. While there is a logical basis for this change, it presents challenges for our program. Notably, about 75% of the courses offered by the SDS are classified as mathematics courses, resulting in their revenue being attributed to the College of Sciences under the university's AIB model. Consequently, this new formula would considerably reduce the regular budget for SDS, leading to potential financial difficulties. This issue is not unique to our program; it is a common challenge faced by other Graduate Interdisciplinary Programs (GIDPs). Addressing this may require a college-level dialogue among the deans, aiming to resolve the issue in a way that supports and encourages the growth of the GIDPs, rather than imposing limitations.

3. Stabilize the existing graduate curriculum while working with key departments to continuously modernize the curriculum.

Rationale The Statistics and Data Science Program is part of the rapid advance in modern statistics and data science, leading to many more units offering valuable coursework for students. While the bulk of the offerings for Statistics and Data Science students were taught by faculty members whose home departments were either Mathematics or Biostatistics at the time of the previous Academic Program Review, many more units (e.g., Systems & Industrial Engineering, School of Information, Computer Science, Molecular and Cellular Biology, Hydrology and Atmospheric Sciences) are now offering valuable coursework (e.g., bioinformatics, graphical models, data visualization, data management, simulation modeling, neural networks, natural language processing). Moreover, the increased importance of Statistics and Data Science inside the Department of Mathematics has led to several new hires. Some of their teaching load, however, is being devoted to our undergraduate offerings.

Action Students face the consistent challenge that results from the instability of the course offerings from one year to the next. Consequently, the Statistics and Data Science Executive Committee will engage key departments (Mathematics, Computer Science, Systems and Industrial Engineering, Biostatistics, School of Information) to coordinate course offerings and call for broad collaboration across departments as new courses are developed. Each of these departments has membership in the Statistics and Data Science program and can serve as sentinels for their own departments.

In addition, the Program will undertake a survey of statistics and data science programs nationally to assess what is the "state of the art" curriculum, especially for doctoral students. This survey should be undertaken with a view of anticipating what new courses can be anticipated in the near future. This information can be shared with subject matter experts in application oriented computationally focused coursework so

that a thoughtful plan of action can be established. This approach will require buy in from department heads and deans to achieve consistent course offerings that may involve a rotation of teaching among departments. At present, Mathematics, Biosystems Engineering, and Biostatistics have such arrangements that should be expanded. Moreover, a carefully presented plan can minimize redundancies in course offerings. This is especially important given that the interest in such courses exceeds the ability of faculty to cover the demand. The new model creates more uncertainty in the funding stream for Graduate Interdisciplinary Programs.

4. **Continue to increase research capabilities in statistics and data science by bringing new faculty to the University and integrating them into the Graduate Interdisciplinary Program.**

Rationale The Statistics and Data Science Program has been aggressive in sponsoring partner hires. This program, which originates in the Provost's Office, offers, in most years, a competition among Interdisciplinary Programs. The Provost's Office provides two years of half salary support. This generally is sufficient to have a department raise in priority a hire in data science or to quickly pivot when a data scientist on the faculty retires or moves to another university. With this approach, Statistics and Data Science has recently sponsored hires in Mathematics, Computer Science, Geography and Development, and Hydrology and Atmospheric Sciences. Members of the Statistics and Data Science Program sit on the hiring committees and these departments look to us to help make a selection that will go far in integrating the department's goals with the mission of the Statistics and Data Science Program.

In the early years of the Program, this approach did not always result in a fruitful interaction of these partner hires into the Statistics and Data Science Program. In addition, some of the hires have left the University. More recently, both the realization among faculty in many disciplines of the need for data scientists and the increasing visibility and reputation of the SDS Program has led to better integration into the Program and better retention. More recent hires came with the expectation that these new faculty would play a role in expanding student offerings and will attempt to recruit SDS graduate students. As a consequence, both the home departments for the hires and the SDS Program see the advantages of this type of arrangement.

However, this initiative, as it is now constructed, does not allow for more senior hires. While young researchers bring fresh ideas, they naturally prioritize the demands associated with tenure in their own academic unit. Partner hires of more senior researchers would increase the ability to mentor young faculty, postdoctoral fellows, and graduate students.

Through this approach and other opportunities, the core faculty in Statistics and Data Science has grown from 32 at the last Program review to 45 today.

Action The Statistics and Data Science Program will continue this approach, working with deans and departments whose own strategic plans naturally align with Statistics and Data Science. The next important step is to improve the integration of new faculty into the life of the Program. This is more easily accomplished with faculty in Mathematics and Biostatistics. The TRIPODS grant (See **Section C.2: Major Changes in the past seven years**.) was the vehicle for a much richer collaborative structure, especially with computer scientists and systems engineers. Thus, we will actively seek out training grant or center grant opportunities, especially with domain scientists. The obvious targets are the department inviting partner hires in the past and the units participating in the professional master's degree, namely Computer Science, Biostatistics, Molecular and Cellular Biology, and Hydrology and Atmospheric Sciences. (Again, see **Section C.2: Major Changes in the past seven years**.)

Such approaches will create a more integrative faculty and assist in increasing student support as described in more detail in **Goal 1**.

5. **Promote the visibility and the participation at the University of Arizona campus of the Members of the Interdisciplinary Program in Statistics and Data Science in major activities involving statistics and data science.**

Rationale: With each coming decade, the sources of data sets become vastly larger and more complex. Research universities will by necessity face the fundamental goal of developing the systems and methods that will wrangle those data sets to produce meaningful information. Prominent examples at the University of Arizona are omics data, satellite data and telescope data from the solar system and deep space, and meteorological data. A decade ago, the use of machine learning and other artificial intelligence algorithms were not commonplace. Nowadays many high school science fair winners use neural nets or random forests as a tool to address the questions in their projects.

Each of these sources of massive data, along with the accompanying small data individual projects derived from these and other sources, will require new mathematical models, e.g. optimization, dynamical systems or stochastic processes, feasible computational strategies, and yet to be developed inferential and data discovery methods in need of solid theoretical underpinnings.

Statistics and Data Science play a central role in the data story - after the data are wrangled and a well-posed question is presented, a well-grounded theoretical model will call for certain data transformation and visualizations. Throughout the history of science, statisticians and data scientists have been essential collaborators in the study design to ensure that the research objectives can be answered with the data that has been mined or is to be collected. Moreover, these statisticians have the ability to frame the new and sophisticated procedures necessary to analyze the data. Oftentimes, the methods of analysis can become routine before a theoretical understanding exists and

theoretical data scientists have a long track record of enabling science to ground these new methods and carefully articulate their applicability.

The movement forward on these fronts was highlighted through the University of Arizona's TRIPODS program. TRIPODS, Translational Research in the Principles of Data Science, is a National Science Foundation endeavor and the University of Arizona was an awardee in the first round of competition.

Even before TRIPODS, mathematicians, statisticians, and computer scientists were frequent collaborators across dozens of university research projects. This interdisciplinary collaboration is a central aspect of the University's research culture. Surprisingly, the computer scientists and statisticians had never collaborated. TRIPODS catalyzed two important changes. First, large data intensive projects could now count on a team of teams approach with data scientists across the spectrum of specialties forming one of the teams. Second, the newly formed TRIPODS research working groups could tackle foundational questions that reached across traditional disciplinary boundaries. In the course of the three years of the grant, the newly formed TRIPODS community (about 30 faculty plus postdocs and graduates) jointly published > 100 papers and submitted > 30 grants.

Action: Several strategies can be employed to enhance the visibility and engagement of the GIDP members in Statistics and Data Science. Active participation in university-wide committees can boost the program's profile, while the institution of awards can motivate involvement in key activities related to the field. The organization of public lectures can serve as a platform to highlight the program's strengths and attract a larger student body. Collaboration with existing University-wide units can further advertise these strengths. Strengthening ties with interdisciplinary units can help draw in more students, while talks delivered by senior GIDP leaders to various data science groups and research centers can foster collaboration and inform about educational initiatives.

Lastly, staying updated with emerging data science efforts is crucial. These strategies collectively aim to increase the program's visibility, attract more students, and encourage active participation.

6. **Raise the local and national visibility of the Interdisciplinary Program in Statistics and Data Science through highlighting its unique positioning among Statistics and Data Science Graduate Programs.**

Rationale For the self study, we have identified research peers. All of these peers are either *Departments* of Statistics or *Departments* of Statistics and Data Science. Indeed, *All* of the University of Arizona's Board of Regents identified academic peers house their graduate education in statistics and data science inside an academic department.

Over the years, the number of applications to our graduate programs, particularly the number of international applicants, has been low compared to our peers. In spite of

that, generally speaking, over half of our offers of admissions with financial support are accepted. The students who decide to accept admission to the University of Arizona for graduate training recognize the special nature of the structure of the SDS GIDP. In particular, they find the interdisciplinary nature of our graduate education and mentoring very attractive.

Action The Statistics and Data Science Program can elevate its prominence both locally and nationally through a strategic action plan. This includes forging collaborations with other Arizona universities to highlight the program's unique positioning among Statistics and Data Science Graduate Programs. Increasing the nomination of fellows and serving on key committees of organizations like the American Statistical Association (ASA), Institute of Mathematical Statistics (IMS), and Western North American Region (WNAR) can further enhance the program's visibility. The recent reinvigoration of an Arizona chapter of ASA, with Helen Zhang serving as president and participation in events like the ASA DataFest can also serve to showcase the program's strengths and attract more students. Regular updates on a dedicated website highlighting faculty achievements, along with the publication of a GIDP newsletter detailing the program's recent activities and accomplishments, can keep interested parties informed. Hosting meetings such as the national ASA Women in Data Science (WiDS) fall meetings can raise visibility, while revising the Winter School Proposal can draw in more students. Engaging with media outlets to spotlight the program's achievements and organizing thematic workshops can also serve to attract a larger student body.

The ASA DataFest, SUnMaRC (Southwestern Undergraduate Mathematics Research Conference), and the proposed S-STEM grant are vehicles for raising the awareness among our neighboring institutions in Arizona and in the southwest, especially those with a focus on undergraduate and master's degrees. Continuing to apply for multi-university center grants will also contribute to raising the program's profile. At present, the Chair of the Statistics and Data Science Program is on the leadership team applying for an NSF-funded mathematical science research institute.

The SDS is also discussing the idea of creating an Advisory Board that includes members from industrial, academic, and community sectors for the program. The purpose of this Board is to offer strategic advice, insights and diverse perspectives, ensure the program aligns with industry trends, academic research, and community needs, and foster a collaborative environment for innovation and problem-solving. The Board members are invited to meet with the SDS members each year to receive updates on the program's progress, discuss challenges, and brainstorm ideas.

All of these efforts are certain to be the source of more research collaborations, increased candidate pool for graduate applications, and broader contacts with industry.

Executive Summary of Needs:

The SDS program faces a multifaceted set of needs crucial for its growth and excellence in the ever-evolving educational landscape. Firstly, sustainable and adequate support for students is paramount, encompassing financial aid, mentorship, and wellness resources to ensure their academic and personal success. Modernizing the curricula and training in data science is another critical need, reflecting the demand for skills relevant in today's data-driven world. Additionally, increasing the number of Teaching Assistants (TAs) and Research Assistants (RAs) is essential to enhance the learning and research experience, and provide necessary support for faculty in handling large, diverse classrooms. To advance academic research, there is a need to augment research capabilities and resources for the faculty, enabling them to conduct cutting-edge studies and contribute significantly to their fields. Expanding the size of the faculty is also crucial to accommodate growing student numbers and diversify academic offerings. Furthermore, raising national visibility through collaborations and high-impact research will position the program as a leader in education and attract more student applicants. Addressing these needs will not only enhance the quality of education and research but also solidify the program's reputation as a hub of innovation and excellence.

SECTION C: UNIT HISTORY

C.1 Beginnings of the Graduate Interdisciplinary Program in Statistics.

The University of Arizona hosted a Department of Statistics beginning approximately in 1980. Upon the advice of the University's central administration, the Arizona Board of Regents abolished the Department of Statistics in 1997. However, the Regents maintained the ability for the University to grant the PhD degree, majoring in Statistics.

In 2005, the Arizona Board of Regents gave approval for the Interdisciplinary Program in Statistics. The initial Program Chair, J. Bruce Walsh, Professor of Ecology and Evolutionary Biology, was tasked with shepherding the chartering process and recruiting an external Chair. Dr. Walter Piegorsch assumed the Chair's duties in early 2007. BIO5 seed money was used to support the hiring of a statistics faculty position in Agricultural and Biosystems Engineering, Geography and Development, Educational Psychology, and Epidemiology and Biostatistics. Through the Chair and the Executive Committee, the Program's mission, organizational structures, curriculum, and first contacts with industry were established. The early momentum provided by the BIO5 Institute led to the program's first certificate in 2009, Master's degree in 2010; and its first PhD in 2014.

The Statistics Program held its first academic review on November 9 and 10, 2015.

C.2 Major changes in the past seven years

The major changes in the activities in the past seven years are a direct response to the recommendations of the Academic Program External Review Committee or are initiatives that capture the spirit of their recommendations.

The major recommendation from the External Review Committee for the 2015 Academic Program Review (APR) was the creation of a new undergraduate degree program in Statistics & Data Science. This program was approved by the Arizona Board of Regents in April 2018. Because the GIDP can only host graduate programs, the undergraduate degree sits in the Department of Mathematics. The initial program was a bare bones addition, adding only three new courses - DATA 375 on statistical computing, DATA 467 on linear models, and DATA 498A, a capstone course that is largely focused on machine learning. The program has been very successful with the number of graduates ramping up to 50 for Spring, 2023.

The department took stock of the degree program in 2022 and made several revisions. The degree now has a data management course (ISTA 322) requirement. In addition, the capstone course has been split to one that focuses on machine learning (DATA 474). Moreover, several projects in designated courses are used to build a portfolio. DATA 498A is now devoted to more professional development activities - creating resumes and preparing for job searches or

graduate school, improving the portfolio, using modern tools for cloud computing, professional ethics, and improving both written and verbal communication skills.

Further, we have developed a new course DATA 201 that is meant to be a foundation course in data science for students across the sciences. The College of Science is promoting the idea of data science serving as a foundation for all sciences with a math/statistics/computer science core curriculum plus specialized courses? in the major (e.g., applied physics, neuroscience, hydrology) that encompass topics in machine learning, data visualization, and data management. All of these data science tracks will have a capstone experience.

Finally, we have partnered with the Capitol University in Economics and Business (CUEB) in Beijing to have a UA degree program on their campus. The initial cohort had ~70 students with the goal to ramp up to 120 entering each year. CUEB is proposing student exchanges and to encourage some of their students to enter our master's degree Program.

Shortly before the last APR, the GIDP created an Accelerated Master's Degree program. At present, the program has only a couple of students enrolling each year. The initial proposal was designed for Mathematics majors. Now that the SDS degree has become popular, we have established a pathway to the Accelerated Master's Degree through an SDS undergraduate major and we anticipate that the Accelerated Master's Degree program will grow.

As recommended in the 2015-2016 academic program review, we did not succeed in securing additional teaching assistantships from other academic units. However, the increase in statistics and data science activities inside the Department of Mathematics has brought revenue to the Department and with it an increase in 0.5 FTE TA support from 7.5 to 13 each year.

The APR External Advisory committee encouraged the Statistics Program to raise its visibility on campus. Our major new initiative to bring Statistics and Data Science to a more central role at the University was the securing of the NSF-sponsored Translational Research in the Principles of Data Science (TRIPODS) for the years 2017 to 2021. This is a collaboration among the foundation units - Computer Science, Mathematics, and Statistics. Helen Zhang (SDS) was the Principal Investigator for TRIPODS. The coPI Executive Committee consisted of Stephan Kobourov (CS), David Glickenstein (Math), and Joseph Watkins (SDS & Math). Before receiving the TRIPODS, a majority of the computer scientists and statisticians and a large fraction of the mathematicians had collaborated with domain scientists but rarely with each other. Through TRIPODS, these communities began extensive collaborations (over 100 papers over the four years of the grant). This is an important transition for data science at Arizona. Nowadays many data intensive research endeavors require a broad multidisciplinary team that includes a span of expertise in data science. Our central activities were the weekly TRIPODS seminar which had a regular attendance of 50 to 70 participants before the pandemic. In addition, the TRIPODS coordinated its activities through 8 research working groups (graphs and networks, data visualization, imaging, natural language processing, optimization, large point set data, and Bayesian methods). As with so many University activities, TRIPODS lost considerable momentum during the COVID pandemic.

In 2020, the School of Mathematical Sciences secured a five-year NSF-Funded Research Training Group (RTG) on the Data Driven Discovery. The grant supports one postdoctoral fellow and five graduate students. One of the graduate students is generally a student in the Statistics and Data Science Program.

TRIPODS served as one impetus to expand the size of the GIDP faculty notably into the Departments of Computer Science and Systems and Industrial Engineering and the College of Medicine. In recognition of this broader reach in departments having faculty members in the GIDP and in accordance with national trends, the program, in 2019, changed its name from Statistics to Statistics *and Data Science*.

The Provost's Office through the Graduate College has a program that allows Graduate Interdisciplinary Programs to apply for a sponsored hire. The Statistics and Data Science GIDP applies at every opportunity and has been able to secure a partner hire each time for every competition. Our approach is to work with the deans to have our proposal reinforce their college's strategic plan. We have completed successful partner hires with Computer Science, Mathematics, Geography & Development, and Hydrology & Atmospheric Sciences. This has proven to be a more effective strategy than trying to convince departments to replace retiring or exiting faculty.

In June 2023, the Arizona Board of Regents approved a new Professional Master's Degree Program in Data Science and Applied Statistics (DS&AS) hosted jointly by the Departments of Mathematics and Computer Science. Our professional Master's Degrees differ from traditional Master's Degrees in several ways. They are specifically designed for those presently in the workforce who look to advance and modernize their skills in data science. The course selection is more focused and the courses themselves are more structured consisting of core modules and assessed through individual and group projects, presentations, and exams. Like many professional master's degrees, the DS&AS program has a capstone experience, e.g., a long writing product or an internship of an applied nature.

The design of this program allows for both a rapid completion to the master's degree and the ability to advance by taking one course at a time. The curriculum is separated into smaller pieces consisting of a three-course sequence. Each of these sequences results in a graduate certificate.

The core courses sequences are **Foundational Theory** and **Application Tools**

Foundational Theory

MATH 509D	**Statistics for Data Science** (new)
STAT 675	**Statistical Computing**
MATH 574M	**Statistical Machine Learning**

Application Tools

CSC 501	**Advanced Programming** (new)
BIOS 576E	**Data Management** (new)
CSC 544	**Data Visualization**

Most top Data Science Professional Master's degrees have very similar core curricula. Their degrees can be completed by taking three additional data science elective courses. The University of Arizona offers the option of completing the Master's Degree with a subject matter emphasis. Our first two are **Computational Biology** and **Earth Science** with more emphases in the near future.

SECTION D: OVERVIEW OF ACADEMIC QUALITY

D.1. Reputational and outcome indicators

The Statistics and Data Science Graduate Program is unique among its peer institutions in that it is housed as a program rather than a department. We have an unusually high acceptance rate (generally >50% for financial supported doctoral students). Our candidates know when they apply that opportunities exist to have their dissertation work be guided by co-mentors, one who is a core statistician and another who has domain knowledge expertise. Indeed, our program is specifically designed to accommodate this kind of future statistician and data scientist.

Given the specific nature of this program, direct comparisons are tricky. However, the Graduate Program in Statistics and Data Science now has a long enough track record to be 54th in US News and World Report rankings.

D.2. Overall nature and breadth of the faculty's scholarly contributions

Membership is made of a group of core statisticians. Beyond that, the Program enjoys having its membership extend to include many who engage in a broad spectrum of questions involving data intensive science. This approach is motivated by our mission *to provide an environment whereby students become independent researchers and practitioners who make significant contributions at the forefront of knowledge across the disciplines that rely on statistical thinking.*

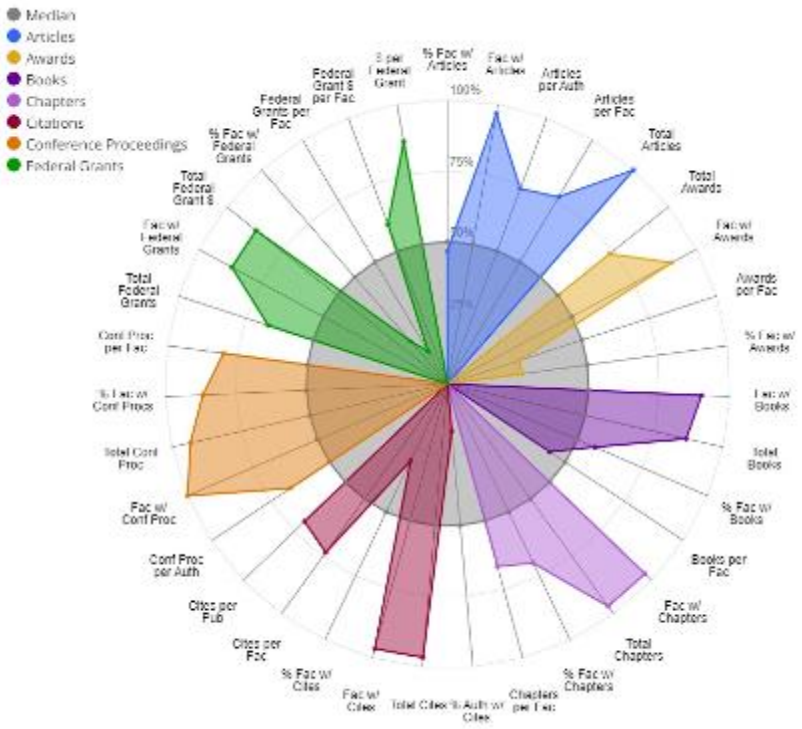
College			College	
Science	15		Medicine	3
Public Health	9		Social & Behav Sciences	1
Agriculture & Life Science	7		Information	1
Engineering	5		Education	1

Members have 15 different departments from eight colleges as their home. At present, the membership homes are the following home departments. *Agriculture, Life, and Environmental Sciences*: Animal & Comparative Biomedical Sciences, Biosystems Engineering, Agricultural-Resource Economics, Natural Resources, Plant Sciences, *Education:* Educational Psychology, *Engineering:* Systems and Industrial Engineering, *School of Information, Pharmacy, Public Health:* Biostatistics, *Science:* Computer Science, Ecology and Evolutionary Biology, Mathematics, Molecular and Cellular Biology.

Because the role and scope of the GIDP in Statistics and Data Science is substantially different from other Statistics and Data Science Programs, no easy to find comparison group is available. Different academic units, engineering, science, mathematics, and so on, have different traditions and varying access to funding. Likewise, the number of articles and the number of citations can

fluctuate wildly from among research endeavors represented by Program Members. These radar charts should be interpreted with this in mind. For example, the level of awards appears low compared to the quality of faculty. This may be a product of the databases used as a comparison. See also **Section B. 2: Major goals or strategic directions for the next five years, Goal 6.**

The first chart shows a comparison with all AAU public institutions.



AAU Public Statistics Departments	
Metric	**Percentile vs AAU Public Institutions**
% Fac w/ Articles	47%
Fac w/ Articles	97%
Articles per Auth	73%
Articles per Fac	77%
Total Articles	100%
Total Awards	73%
Fac w/ Awards	90%
Awards per Fac	27%
% Fac w/ Awards	27%
Fac w/ Books	90%
Total Books	87%
% Fac w/ Books	57%
Books per Fac	43%
Fac w/ Chapters	97%
Total Chapters	97%
% Fac w/ Chapters	70%
Chapters per Fac	67%
% Auth w/ Cites	17%
Total Cites	97%
Fac w/ Cites	97%
% Fac w/ Cites	30%
Cites per Fac	73%
Cites per Pub	70%
Conf Proc per Auth	67%
Fac w/ Conf Proc	100%
Total Conf Proc	93%
% Fac w/ Conf Procs	87%
Conf Proc per Fac	80%
Total Federal Grants	67%
Fac w/ Federal Grants	87%
Total Federal Grant $	87%
% Fac w/ Federal Grants	20%
Federal Grants per Fac	13%
Federal Grant $ per Fac	60%
$ per Federal Grant	87%

D.1.1. Major faculty and student honors and awards

Examples of faculty honors and awards

- Society Fellows
  - American Statistical Association (Bedrick, Piegorsch, Zhang),
  - International Statistical Institute (Piegorsch, Zhang)
  - Institute of Mathematical Statistics (Bhattacharya, Sethuraman, Zhang)
  - American Physical Society (Chertkov)
  - American Association for the Advancement of Science (Chertkov)
- Accredited Professional Statistician (Piegorsch)
- Galileo Circle Fellow (Watkins, Zhang)
- Guggenheim Fellow (Bhattacharya)
- Kavli Fellow (Gutenkunst)

- Hertz Foundation Fellowship (Lin)
- Lane Fellowship (Lin)
- Lifetime Achievement Award, Congress in Quantitative Genetics (Walsh)
- Medallion Lecturer, Institute of Mathematical Statistics (Zhang)
- NSERC postdoctoral fellow (Barnard, Watkins)
- NSF Graduate Research Traineeship (Gutenkunst)
- NSF Career Award (Barnard, Chen, Zhang)
- Oppenheimer Fellowship (Chertkov)
- SAMSI Fellow (Lin)
- Simons Foundation Fellow (Sethuraman)
- Best/Superior/Distinguished dissertation/paper awards (An, Hsu, Hu, Piegorsch, Steidl)
- Teaching award (Gutenkunst, Roe, Walsh, Watkins)

Examples of student honors and awards

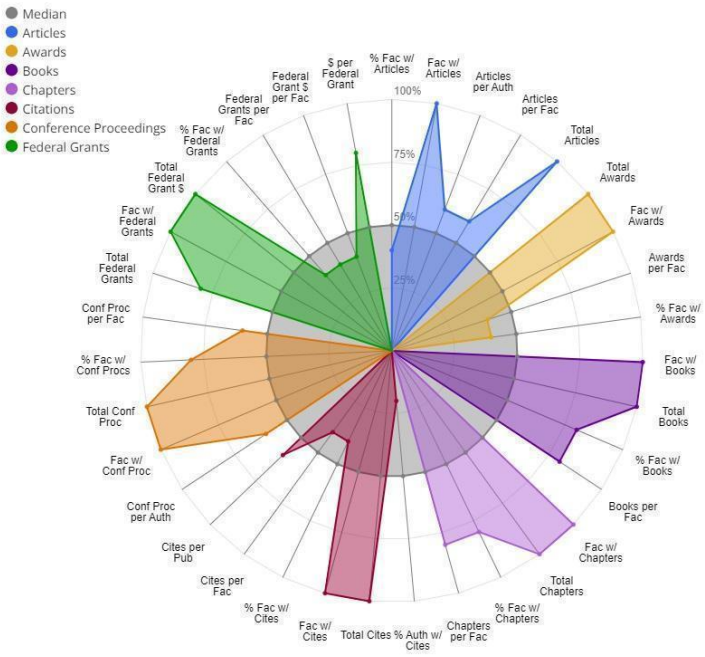
- Outstanding Poster Award (2022 Annual Conference of Korean Society for Bioinformatics), Excellence Award in the Poster presentation (2019 Conference of Korea Ecological and Environmental Science)
- Graduate program travel awards
- Honorable Mention in the 2021 Student Paper Competition. ASA. Section on Risk Analysis
- 2022 Galileo Scholarship

D.2 Aspirational Peer Programs

The self-study committee selected among its aspirational peer programs, three young programs, namely, Indiana University, the University of Texas at Austin, and the University of California, Los Angeles (although I think UCLA is having a 25th anniversary this year). The advancement in development of a statistics program at these four large public universities is meant to serve as a measure of the quality of the Statistics Program at the University of Arizona. Like the University of Arizona, the University of Texas and the University of California, Los Angeles call their unit Statistics *and Data Science.* This has become a national trend as units recognize that the scholarly interests of the faculty and the curricula are better described by including both terms - statistics and data science.

The fourth and fifth choices have mature programs – the University of Illinois at Urbana-Champaign and the University of Washington. The University of Washington has both a mature Department of Statistics in the College of Science and a mature Department of Biostatistics in the College of Public Health (as does UCLA).

This summary of the data for the aspirational peers is substantially the same as the comparison against all AAU public institutions.



Selected Peer Statistics Departments	
Metric	Percentile vs Selected Peer Institutions
% Fac w/ Articles	40%
Fac w/ Articles	100%
Articles per Auth	60%
Articles per Fac	60%
Total Articles	100%
Total Awards	100%
Fac w/ Awards	100%
Awards per Fac	40%
% Fac w/ Awards	40%
Fac w/ Books	100%
Total Books	100%
% Fac w/ Books	80%
Books per Fac	80%
Fac w/ Chapters	100%
Total Chapters	100%
% Fac w/ Chapters	80%
Chapters per Fac	80%
% Auth w/ Cites	20%
Total Cites	100%
Fac w/ Cites	100%
% Fac w/ Cites	40%
Cites per Fac	40%
Cites per Pub	60%
Conf Proc per Auth	60%
Fac w/ Conf Proc	100%
Total Conf Proc	100%
% Fac w/ Conf Procs	80%
Conf Proc per Fac	60%
Total Federal Grants	80%
Fac w/ Federal Grants	100%
Total Federal Grant $	100%
% Fac w/ Federal Grants	40%
Federal Grants per Fac	40%
Federal Grant $ per Fac	40%
$ per Federal Grant	80%

.

This summary of the data for the aspirational peers is substantially the same as the comparison against all AAU public institutions.

SECTION E: FACULTY

E. 1. Overall nature and breadth of the faculty's scholarly contributions

The faculty in the Program of Statistics and Data Science regularly make significant contributions across a range of issues in statistical theory and methodology, and in many important areas of activity that rely on modern and sophisticated data science approaches. Here is a sampling of our faculty's areas of expertise.

- large sample theory, Bayesian methods, random effect and latent variable models, spatiotemporal statistics, generalized linear and mixed models, longitudinal data analysis, survival analysis, item response theory, change point detection, compositional data, missing data, causal inference
- high dimensional data, nonparametric statistics, semiparametric modeling, dimension reduction and variable selection
- data mining, machine learning, probabilistic graphical models, approximate inference algorithms, and information-theoretic decision making
- computational biology and bioinformatics: statistical genomics/metagenomics/ proteomics/single-cell omics, population genomics, systems biology
- analysis of algorithms, parallel computing
- measurement and educational assessment, computational cognitive science and neuroscience, psychometrics, environmental statistics, conservation ecology, risk assessment
- clinical trials, statistical meta-analysis, pharmacokinetics, epidemiology
- computer vision, image analysis, multimodal data
- optimization under uncertainty, applied operations research, control theory, game theory
- stochastic approximation and simulation: Monte Carlo algorithms; numerical methods for stochastic differential equations
- probability theory and stochastic processes, Markov processes, Gaussian processes, martingales, stochastic interacting particle systems, nonequilibrium statistical physics, weak convergence, and limit theorems

1. Briefly describe the overall nature and breadth of the faculty's research and other scholarly contributions in the generation of knowledge, exemplary practice or creative performance with an appraisal of the most significant contributions to advancing the field or discipline.

E. 2. Grants, contracts, patents, and license agreements

A table of current and pending grants, contracts, patents and license agreements, list faculty and principal investigator names, funding source and amount, and funding period can be located in **Appendix 2.**

E. 3. Faculty participation in leadership activities
Journals and conferences (listed by journal and conference)

- Editor-in-chief (Stat)
- Journal Editor (Environmetrics, Journal of Alzheimer's Disease, Journal of the American Statistical Association, Journal of Methods and Measurement in the Social Sciences, Organelle)
- Associate Editor (Biometrics, BioScience , BMC Bioinformatics, BMC Ecology and Evolution, Energy Systems, Evolution, Frontiers in Genetics, Frontiers in Plant Science, Genetical Research, Genetics ,IEEE Transactions on Pattern Matching and Machine Intelligence, Journal of Computational Dynamics, Journal of Computational and Graphical Statistics, Journal of Educational and Behavioral Statistics, Journal of the American Statistical Association, Journal of Royal Statistical Society, Series B, Journal of Wildlife Management , Mathematical Medicine and Optimization, Letters, Plant Methods, PLOS Genetics, SIAM Journal on Applied Dynamical Systems, The American Naturalist, Theoretical Population Biology, Transactions of Machine Learning Research)
- Conference Program Chair (JSM)
- Conference Coordinator (Cold Spring Harbor, Gordon Conference, INFORMS, Workshop on Machine Learning for Cybersecurity, WNAR)

Leadership activities

- Society President
    - Western North American Region of Intl. Biometrics Society (Bedrick)
    - Western North American Region of the International Biometric Society (WNAR) (Roe)
- Committee on Scientific Freedom and Responsibility, AAAS (Watkins)
- Committee Human Rights of Mathematicians (Chair) (Watkins)
- Program Chair, 2023 Joint Statistical Meetings (JSM) (Zhang)
- Section Chair, ASA Section on Statistical Learning and Data Sciences (Zhang)
- President, ASA Arizona Chapter (Zhang)

E. 4. Membership teaching

All of our teaching assignments are made through the Members' respective departments. For the 6 required courses, the 4 first year core courses and the course Statistical Machine Learning are hosted by the Department of Mathematics. However, the instructor for Design of Experiments is a member of the Department of Biosystems Engineering. The sixth required course, Statistical Consulting is hosted by the Department of Epidemiology and Biostatistics. The instructor, Dean Billheimer, is a faculty member in that Department and the Director of the Statistical Laboratory.

Recent teaching evaluations for the core courses can be found in **Appendix 3.** The quality and adequacy of the teaching and the curriculum are assessed through the Annual Progress Review

(See **Appendix 09**.) and through student led focus groups. (See **Appendix 10** for the student input gathered for this self study.) This information is used to revise any aspect of the program.

E. 5. Faculty recruiting

The Graduate Interdisciplinary Programs, in general, have very limited opportunity to become involved in the recruitment and retention of faculty. The Statistics and Data Science Program has been active in participation in the sponsor hires opportunities through the Provost's Office, submitting a proposal for every competition over the past 7 years. The partner hire initiative provides half of the salary for a two-year period.

Over the past seven years, we have completed successful partner hires with Computer Science (2018), Mathematics (2022), Geography & Development (2019), and Hydrology & Atmospheric Sciences (2023) In each case, the hiring committee had one or two members of the Statistics and Data Science program on the hiring committee. This perspective was seriously taken into consideration by the entire hiring committee and the department. We will continue to pursue this avenue to recruit faculty and to build more genuine relationships between the Statistics and Data Science Program and the partnering departments.

E.6. Faculty Compensation

The American Statistical Society provides summary information on salaries. The table below gives this information for the 2021-2022 academic year.

Table 1. 2021-2022 Academic Faculty in Statistics Departments by Rank and Years in Rank, Based on 9-Month Salary

Rank	Years in Rank	N	1st Quartile	Median	3rd Quartile	90th Percentile
Professor	0-2	29	$116,694	$138,000	$169,700	$216,027
	3-5	29	$131,700	$152,100	$199,800	$266,250
	6-9	25	$150,200	$162,665	$195,951	$256,250
	10-16	23	$132,972	$172,545	$223,599	$236,060
	17+	50	$156,744	$176,090	$210,985	$255,054
	All	156	$137,043	$164,000	$201,425	$248,508
Associate Professor	0-1	15	$106,650	$118,242	$127,813	$135,000
	2-3	19	$105,000	$119,575	$136,953	$143,200
	4-6	23	$98,583	$106,590	$129,500	$159,850
	7+	19	$99,041	$109,696	$128,150	$137,300
	All	76	$101,153	$113,196	$128,150	$138,370
Assistant Professor	0-1	28	$95,990	$107,345	$111,108	$117,300
	2-3	40	$95,900	$104,475	$118,950	$131,950
	4-5	23	$100,589	$105,372	$115,474	$172,750
	6+	9	$110,000	$113,100	$120,950	$132,600
	All	100	$99,222	$106,575	$116,387	$131,300
Instructor	All	60	$63,000	$75,809	$89,388	$125,652
All Ranks	**All**	**392**	**$100,004**	**$118,662**	**$157,040**	**$201,500**

As indicated at the beginning of this section, our faculty have home departments across the University of Arizona campus. These academic units, in general, compete nationally not against Statistics and Data Science faculty but rather with faculty in their own field. As a consequence, competitive salaries may not match well with national salaries in Statistics and Data Science Departments. This will play a role in the distribution of salaries. Some colleges, e.g., business, engineering, and medicine, have higher salaries. Others, e.g., social and behavioral sciences and education, have lower salaries. Some units, e.g., public health, generally state 12-month salaries whereas most colleges use 9-month salaries. This can most easily be seen in the fluctuating nature of the assistant professor salary. (We should note that oftentimes a faculty member does not join the GIDP immediately upon hire. This may account for the low assistant professor count.)

With these caveats in place, the salary average of the University of Arizona matches closely with national averages. In particular, the data for both groups show a nearly 2 to 1 salary ratio between the most experienced faculty and recent hires.

We did not receive any data on the salaries of our designated peers and so cannot make that comparison.

	2017				2018			
Academic Rank	Employee Headcount	Base Salary MIN	Base Salary MAX	Base Salary AVERAGE	Employee Headcount	Base Salary MIN	Base Salary MAX	Base Salary AVERAGE
Professor	21	84,864	195,500	125,725	20	86,093	200,000	132,381
Associate Professor	11	75,750	98,386	88,805	11	75,750	98,386	89,093
Assistant Professor	1	125,500	125,500	125,500	1	125,500	125,500	125,500
No Academic Rank	1	184,775	184,775	184,775	1	184,775	184,775	184,775

	2019				2020			
Academic Rank	Employee Headcount	Base Salary MIN	Base Salary MAX	Base Salary AVERAGE	Employee Headcount	Base Salary MIN	Base Salary MAX	Base Salary AVERAGE
Professor	21	86,093	222,423	142,583	23	86,093	222,423	146,233
Associate Professor	11	80,000	101,174	92,341	11	86,000	104,854	95,335
Assistant Professor	1	90,000	90,000	90,000	1	90,000	90,000	90,000
No Academic Rank	1	194,525	194,525	194,525	1	194,525	194,525	194,525

	2021				2022			
Academic Rank	Employee Headcount	Base Salary MIN	Base Salary MAX	Base Salary AVERAGE	Employee Headcount	Base Salary MIN	Base Salary MAX	Base Salary AVERAGE
Professor	22	76,693	197,000	131,971	21	91,129	232,300	149,882
Associate Professor	11	75,388	102,375	87,921	11	89,579	114,382	101,297
Assistant Professor	1	76,500	76,500	76,500	1	93,150	93,150	93,150
No Academic Rank	1	194,525	194,525	194,525	1	197,443	197,443	197,443

	2023			
Academic Rank	Employee Headcount	Base Salary MIN	Base Salary MAX	Base Salary AVERAGE
Professor	22	98,373	251,681	165,957
Associate Professor	11	95,499	135,491	111,995
Assistant Professor	1	98,820	98,820	98,820

No Academic Rank	0							

E. 7. Faculty by gender and race/ethnicity

Here is the demographic make-up of University of Arizona Faculty in Statistics and Data Science as gathered from the University's database

Gender	2017	2018	2019	2020	2021	2022	2023
Female	29.6%	29.6%	32.1%	30.3%	32.4%	32.4%	30.6%
Male	70.4%	70.4%	67.9%	69.7%	67.6%	67.6%	69.4%
Grand Total	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**

IPEDS Race/Ethnicity	2017	2018	2019	2020	2021	2022	2023
Asian	44.4%	44.4%	39.3%	33.3%	35.3%	35.3%	41.7%
Hispanic or Latinx				6.1%			
International	3.7%	3.7%	10.7%	12.1%	20.6%	20.6%	13.9%
Not Reported	7.4%	7.4%	7.1%	6.1%	5.9%	5.9%	8.3%
White	**44.4%**	**44.4%**	**42.9%**	**42.4%**	**38.2%**	**38.2%**	**36.1%**
Grand Total	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**

Here is the comparison to our designated peers. This was gathered by inspection from departmental/program webpages

Gender	Arizona	Indiana	Illinois	UCLA	Washington
Female	30.6%	40.0%	35.9%	23.5%	33.3%
Male	69.4%	60.0%	64.1%	76.5%	66.7%

Race/Ethnicity	Arizona	Indiana	Illinois	UCLA	Washington
Asian	51.7%	50.0%	51.3%	41.2%	9.5%
Black	0.0%	0.0%	0.0%	2.9%	4.8%
Latin	0.0%	10.0%	5.1%	8.8%	4.8%
Middle East	4.7%	0.0%	0.0%	11.8%	9.5%
White	44.7%	40.0%	43.6%	35.3%	71.4%

The Program member demographics are similar in gender and in race/ethnicity except for the University of Washington which has a high percentage white faculty. The Program makes no direct hires and membership is open to any research active faculty who work with statisticians and data scientists on data intensive science. Thus, the Statistics and Data Science Program has little influence on the demographics of its faculty members beyond its efforts through partner hires. (See **Section E.5**)

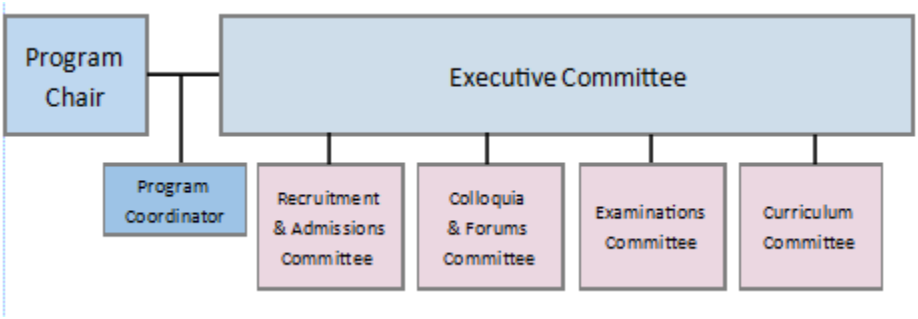
E. 8. Member Biographical Sketches

Biographical sketches of Statistics and Data Science Program faculty members are in **Appendix 4**.

SECTION F: UNIT ADMINISTRATION

F. 1. Organizational chart and governance structure

All Graduate Interdisciplinary Programs, including Statistics and Data Science, report to the Dean of the Graduate College through their respective Chairs. The Statistics and Data Science program receives a half time Program Coordinator, a budget for operations, fellowship support, graduate assistantship support (generally two 0.5FTE teaching assistantships) and tuition support from the Graduate College. Statistics and Data Science graduate students receive some support through teaching assistantships from the Department of Mathematics. Presently, Statistics and Data Science has an agreement with the Department of Mathematics for ~6.50 FTE per semester TA support.



Organization Structure of the Statistics Program

The Statistics Program Executive Committee (EC) is charged with administration of the academic program (See **Appendix 6** for the Program Bylaws). A three-quarter FTE Graduate Program Coordinator assists the Program Chair and Executive Committee. The Executive Committee consists of six faculty members, including the Chair who also serves as the Executive Committee chair. Committee members are elected in pairs from Program members to staggered, renewable three-year terms. Even though it is not stated in the Bylaws, Executive Committee membership is traditionally two from the Department of Mathematics, two from the College of Public Health and two from across campus (College of Agriculture, Life and Environmental Sciences, College of Social and Behavioral Sciences, College of Education, and College of Law in recent times). The Program Chair is elected separately and also serves a renewable three-year term.

The Executive Committee oversees membership in the Program and can approve or deny Regular or Affiliate Membership. The Executive Committee will assist the Chair in providing general oversight and direction for the GIDP, via the following activities:

- providing advice and counsel to the GIDP Chair regarding appointments to GIDP committees,
- preparing and submitting an annual review of Program activities and accomplishments to the Director of Graduate Interdisciplinary Programs,

- ruling on any curriculum matters brought by GIDP students,
- promoting interdepartmental awareness and supporting education and research related to the field of Statistics,
- advising the Director of Graduate Interdisciplinary Programs and the Vice President for Research on issues pertinent to the GIDP and to the field of Statistics, and
- reviewing these Bylaws annually to ensure that the GIDP structure remains modern, pertinent, and operable.

Four members of the Executive Committee serve as chairs for the Program's standing committees, (a) Recruitment and Admissions, (b) Curriculum, (c) Colloquia and Forums, and (d) Examinations Committee. The Program Vice Chair oversees Executive Committee elections, the annual progress report, and represents the program in place of the Chair as needed.

The duties of the Chair are to:

- convene and preside at meetings of the Executive Committee to be held not less than once a semester; (once a month is standard)
- help guide the efforts of the Executive Committee in acquiring University and other external resources to support the full functioning of the GIDP and manage effectively the budget and financial resources of the Program;
- convene and preside at meetings of the larger GIDP in Statistics at least once per year and otherwise as needed;
- prepare Promotion & Tenure and Continuing Review documentation for faculty members of the GIDP according to University policies and procedures;
- monitor the activities of the Statistical Consulting Laboratory and negotiate with the separate Consulting Director details pertinent to its operation, including budget, personnel, graduate student employment and duties, and any consulting activities external to the University;
- monitor and update catalog and other copy of all GIDP curricular and promotional materials; and
- with the advice and counsel of the Executive Committee appoint and supervise the Standing Committees of the GIDP. Article III below discusses how these committees will manage administrative matters of the GIDP, including course requirements and changes.

Executive Committee 2023-2024

- *Chair* – Helen Zhang Professor of Mathematics, College of Science
- *Vice Chair* – Joseph Watkins, Professor of Mathematics, College of Science
- *Recruitment and Admissions Committee* – Xiaoxiao Sun, Assistant Professor of Biostatistics, College of Public Health
- *Colloquia and Forums Committee* – Ning Hao, Associate Professor of Mathematics, College of Science
- *Curriculum Committee* – Edward Bedrick, Professor of Biostatistics, College of Public Health

- *Examinations Committee* – Lingling An, Professor Biosystems Engineering, College of Agriculture, Life, and Environmental Sciences

F.2 Program staff and adequacy of staff support

Typically, the Graduate College has provided funding for a single half time program administrator. As the Statistics and Data Science program continued to grow, the Program Chair found the need to take on some administrative responsibilities.

The Statistics and Data Science Program has a single staff person, a three-quarter-time (0.75 FTE) Program Coordinator, [redacted]. The remaining 0.25FTE is funded by the College of Science. [redacted] is 0.25FTE program administrator for the Data Sciences Academy. Thus, her two positions have some considerable synergy.

[redacted]. Staff salaries were raised substantially in 2017 and again in 2023. These increases have allowed the Program to recruit among a pool of candidates with broader program administrative experience. [redacted]

The additional 0.25FTE provided by the College of Science allowed [redacted] to become a part of the Department of Mathematics staff. This arrangement has resulted in a more efficient program administration. The most important change is the joint recruiting activities of the School of Mathematical Sciences. Candidates seeing the programs together give them a better sense of the breadth of activities at the University of Arizona. The graduate students in these three programs teach together and socialize together. The recruitment workshop provides this sense of unity from the beginning.

The Mathematics Department staff can take some responsibilities that are routine for them but rare for [redacted] In return, [redacted] is available to provide some support for Math Department activities. Importantly, the positioning in the Math Department has given [redacted]a sense of collegiality and a better job satisfaction.

The current staff support structure has evidently been effective, contributing to the success and efficiency of the program. The availability of a full-time staff person for both students and faculty has been a cornerstone for the program's smooth operation, daily activities, as well as its continued development and sustained expansion. Regarding the salary aspect, it is important to maintain the salary support of the program coordinator, comparable with the peers at the same administrative level. It serves not only as a recognition of the value and contribution of the staff member but also as a means to ensure continuity and stability in this role, which in turn affects the overall quality and consistency of the support provided to the program.

SECTION G: UNIT RESOURCES

G.1. Support services and resource needs

The students in the Statistics and Data Science Program have groups of natural colleagues – students in the School of Mathematical Sciences (Mathematics, Applied Mathematics, and Statistics and Data Science) and students in the Department of Epidemiology and Biostatistics in the College of Public Health. The students in these two academic units share with the Statistics and Data Science students participate in a number of joint activities – classes, colloquia, and recruitment.

The Mathematics Department provides space for the administration of the Statistics and Data Science Program – the offices for the chair and the coordinator along with a Data Science Community Room. It also provides office space for doctoral students who have teaching responsibilities in mathematics or statistics. These offices are primarily in the Mathematics Building, the Physics and Atmospheric Sciences Building, or in the Environmental and Natural Resources Building II. Classrooms, classroom support, technology support, and access to necessary office personnel are supplied by the Mathematics Department. About a third of our doctoral students receive support from the student's mentor. In these cases, the mentor's home academic unit accommodates office needs.

Research assistant positions can be found across campus. The Program Chair actively works with Program Members to find suitable research positions. The funding trends are contained in **Section I.3.c**.

In general, the Graduate Interdisciplinary Programs are unable to hold funding from external sources. However, the Departments with the Statistics and Data Science Program to obtain such funding. For example, GIDP Chair Helen Zhang was the PI for the TRIPODS grant and GIDP member Kevin Lin is the PI for a Research Training Grant (RTG). TRIPODS supported students in computer science and the mathematical sciences. RTG supports students in the mathematical sciences. In both cases the extramural funds supported Statistics and Data Science students. For more on TRIPODS, see **Section C.2 Major Changes in the past seven years.** We plan to look into more center type grant proposals. This is discussed under **Goal 5** in **Section B.**

The new initiative to establish a Statistics and Data Science undergraduate degree in Beijing will bring funding to the Mathematics Department that will, in part, be used to support Statistics and Data Science as teaching assistants. (For more, see **Section C.2 Major Changes in the past seven years**.)

G.2. Operations budget

The operations budget for the Program are provided by the Graduate College and the College of Science through the Mathematics Department. We use operational funds to pay for colloquium

expenses, to purchase office supplies and buy food for student-centered events, including community-building events – lunch with colloquia speakers and food and beverages for orientation, colloquia, and student-organized activities. For the graduate recruitment workshop we partner with the two other graduate programs in the School of Mathematical Sciences to bring our top applicants to Tucson to visit campus and interact with faculty and current students. The funds pay for airfares (for doctoral candidates), lodging and meals.

During the pandemic, when travel was severely curtailed, the funds were used to purchase a node on the University's High Performance Computer dedicated to for students' use.

SDS Budget 2023/2024 AY	
Recruitment funds from Grad College	$7,000.00
Operations funds from Grad College	$10,000.00
Operations funds from Math Department	$5,055.00
College of Science Funds	$6,500.00
Total budget available for operations	$28,555.00

G.3. Changes to increase efficiency

The Statistics and Data Science Program has a single member of staff. This position is now formally in the Department of Mathematics. Consequently, the Coordinator has access to administrative help in scheduling, program administration, technical support, and a generally supportive community of coworkers. This shift in the home unit has improved the quality of service to faculty and students. Tasks rarely performed by the GIDP coordinator are common for other staff members and their assistance improves efficiency. The weekly calendar of events, student recruitment, and other activities are naturally coordinated through the three units in the School of Mathematical Sciences. Moreover, as many of the doctoral students had teaching appointments in the Department of Mathematics, the placement of the SDS staff allows for easier coordination between the students' education and teaching responsibilities.

G.4. Projected changes in unit activities and quality outcomes

These changes are discussed in detail in the goals for the unit in **SECTION B: UNIT DESCRIPTIONS AND GOALS**. In broad terms, the new undergraduate degree, the new Chinese degree program, and the new professional master's degree hosted by the Departments of Mathematics and Computer Science further requires smooth interaction between the Program and the Mathematics Department.

SECTION I: GRADUATE STUDENTS, DEGREE PROGRAMS AND OUTCOMES

The goal of this section is to provide descriptions and supporting data on graduate programs. If the unit under review has no graduate programs, that fact should be noted in the self-study report. When reporting student data, please follow FERPA guidelines, available on the Office of the Registrar's website: https://registrar.arizona.edu/privacy-ferpa/ferpa-compliance.

I.1. Degree Programs

As noted in **Section A.3: Academic Programs**, the Graduate Program in Statistics and Data Science has six distinct academic programs – the graduate certificate (available both as a residential and an online program), the master's degree, the accelerated master's degree, two tracks for the doctoral degree, the regular track and the informatics track, and a PhD minor. The accelerated master's degree was established in 2014, the doctoral informatics track in 2012, and the Online Certificate Program in 2013.

Due to the constraints in being a Graduate Interdisciplinary Program and priorities of the central administration, the Statistics and Data Science Graduate Program has no current plans for new degree programs. Program members are closely associated with designing, implementing, and teaching the undergraduate degree program in Statistics and Data Science hosted by the Department of Mathematics and the Professional Master's Degree Program in Data Science and Applied Statistics hosted by the Departments of Mathematics and Computer Science. See **Section C.2: Major changes in the past seven years** for a more thorough description of these degree programs.

All of the Statistics and Data Science Graduate Programs have Classification of Instructional Programs (CIP) code 27.5001, Statistics, General. The CIP code for Data Science, General) is 30.7001. These designations are too broad for the current landscape in the data sciences.

For admission into graduate programs, the Recruitment and Admissions committee looks for several indicators.

- For a Certificate student, we look for the basic requirement for admission to the Graduate School plus a strong background in linear algebra and multivariate calculus along with some experience with statistics. Certificate students are often nontraditional students, several years removed from the university education possessing considerable work experience. Thus, acceptance into the Certificate Students requires a full review of the application material.
- For a master's student, we would like to see some background in linear algebra and some understanding of mathematical analysis. This can be obtained with a 2nd year linear algebra course and a 3rd year "proofs" course. Beyond that we like to see some evidence of interest in statistics and understanding of the nature of the subject. This can be achieved with a college algebra-based course in stats or some out of class activity that

shows that the student has worked on projects using formal statistical procedures. However, we prefer to see a calculus-based statistics course on their transcript. Background in programming is a plus.

- For a doctoral student, the requirement in linear algebra and mathematical analysis should be at the 4th year undergraduate level. We would like a strong mathematics background for admission to the regular track and, additionally, strong computational skills for the informatics track. Any grade below B in senior-level undergraduate courses needs to be explained to the satisfaction of the Admissions Committee.

Because practicing statisticians and data scientists are expected to have good communication skills as a part of their occupation, we expect adequate verbal GRE scores.

I.2. Graduate program - curriculum and courses

Given the wide range in the number of academic units, we are not able to provide detailed and up to date information on every course available to our students. We will provide details on required core courses and other frequently chosen courses.

The first-year core courses along with courses in Statistical Machine Learning and Statistical Consulting are required of all master's and doctoral students. (See **I.2.b. Core course offerings** below.) In August at the end of the first year of studies, both master's and doctoral students take a two-part written qualifying exam, theory (probability and statistics) and methodology (regression and experimental design) based on the first-year core courses. The students receive one of three assessments on each part: PhD pass, MS pass, or fail. Students may retake the exam once and the final result for each part is the better of the two results on each attempt. Students must receive the level appropriate to their degree program to continue towards graduation.

A doctoral minor student must complete 4 courses from a long list spanning many academic units. The single required course is the Theory of Statistics. Students must receive an A or B in this course to receive the minor.

A certificate student must complete the equivalent of a doctoral minor. In an effort to accommodate Certificate students in the workforce, all four first-year core courses are now a part of UA Online and thus the Certificate can be earned as an online student.

Master's students have the option of completing a thesis in place of the qualifying exam. In addition to the 6 required courses, master's students complete their degree by taking an additional 3 (with the thesis option) or 4 courses.

The Accelerated Master's Program is reserved for high achieving undergraduate mathematics majors. Students in the accelerated master's degree are recommended for admission to the program at the end of their third year of study. During their fourth year, they take the first-year core courses, finish their undergraduate studies, and attempt the qualifying exam.

During the third year, doctoral students are encouraged to nearly complete their coursework requirements and to form a Comprehensive Examination Committee. The Committee consists of at least three Program members and a representative of the student's minor program of study. The Comprehensive Examination Committee bears the responsibility for setting the written portion of the comprehensive exam. The format and the timing for the exam is flexible and left to the discretion of this Committee with the goal to structure the exam in the best interests of advancing the preparation of the candidate. The typical format for the written portion of the examination is a series of technical and conceptual questions put forth by the committee concerning the student's expected dissertation research. A variety of formats are acceptable and not limited to the following suggestions.

- A series of written questions prepared by the Committee under a specified time schedule.
- A review paper based on a specific set of background documents set by the Committee and related to the candidate's research topic.
- A literature review of the dissertation topic with an analysis of the shortcoming of previous research as they apply to the candidate's research topic.
- A dissertation proposal with preliminary analysis.

Upon successful completion of the written portion of the Comprehensive Examination, a student must sit for the oral portion of the exam. The Committee conducting the oral portion of the examination has both the opportunity and obligation to require the student to exhibit knowledge of

i. the specific questions/material posed during the written portion,
ii. general comprehension of the minor field(s) of study as it pertains to the student's research interests, and
iii. sufficient depth of understanding in the area(s) of the student's statistical specialization.

The ability to communicate effectively, both verbally and in writing, to audiences of varying levels of sophistication is essential to a successful career in industry, research, or teaching. Consequently, the Statistics program has a Communications Skills requirement. This requirement gives students an opportunity to develop their capabilities in a variety of directions to the public, to their peers, and to an audience of specialists. At present, we are broadening the Communications Skills requirement to a *Professional Preparation Requirement* that involves up to date resumes and webpages and building of a portfolio for use in job searches. The following two forms present tentative ideas and are under review by the Executive Committee. The Graduate Center at UA offers great resources to prepare students for their academic and professional careers through enhanced communication skills, career services, and networking. While we have regularly advertised the resources of the Graduate Center available to our students our field-specific Professional Preparation will allow us to complement our activities with the Center, in order to maximize students' preparation for their careers.

Professional Preparation Requirement
Doctoral Students

	Verbal	Written
Communications with General Audience	• K-12 classroom visit • Mentoring at DataFest • Speaker at events • Internships	• Extend previous projects • GIDP newsletter article
B. Communications with Statistical Audience	• Colloquium or seminar talk • Conference talk • GIDP showcase • Speak at student seminars • Job talk	• Research papers • Grant proposals • Posters • develop an R package
C. Community Activity/Service	• Meet speakers (at least twice) • Attend a conference (at least once) • Student Representatives • Organize journal clubs	
D. Webpage	Create a web page on professional activities (contact, education, research, teaching, experiences), submit the web address to the program's website	
E. Professional CV	biosketch, resume	

Master's Students

	Verbal	Written
. Communications	• Mentoring at DataFest • Job talk rehearsal at the student seminar • Internships	• Extend one course Report to be included for portfolio • Poster
B. Community Activity/ Service	• Meet speakers (at least once) • Student Representatives	
C. Webpage	Create a web page on professional activities (contact, education, research, teaching, experiences), submit the web address to the program's web site	
D. Portfolio	resume - projects, experiences, skills	

I.2.a. Dual enrollment courses

Dual enrollment courses are co-convened courses that are numbered 4xx for undergraduate students and 5xx for graduate students. The 5xx course is expected to have additional work and learning expectations for graduate students and the list is indicated on all syllabi of co-convened courses. At present, none of the core courses and other frequently chosen courses are dual enrollment. The Department of System Engineering and the School of Information offer dual enrollment courses that are used for the certificate and the graduate minor students. They are not a part of the Master's or doctoral degree program's accepted courses.

With the rapid growth of the Statistics and Data Science undergraduate degree, we may see the use of dual enrollment courses designed for master's degree students.

I.2.b. Core course offerings

The master's and doctoral degree programs share 6 three-unit core courses, two theory courses (MATH/STAT 564, Theory of Probability, and MATH/STAT 566, Theory of Statistics), two methodology (MATH/STAT 571A, Advanced Statistical Regression, and MATH/STAT 571B, Design of Experiments), a machine learning course MATH/STAT 574M and a consulting course BIOS/STAT 688A (1 unit) and BIOS/STAT 688B (2 units).

The first five are offered both face-to-face and online, synchronously and asynchronously. The lectures are captured and posted on the course page. All students have the same requirements - homework and exams. STAT 688A is a course on the basics of statistical consulting and STAT 688B is based on teams enrolled in the class working directly with clients and is face-to-face only.

I.2.c. Sufficiency of current courses

This information was gathered as part of a broader student and alumni input. See **Appendix 10** for the questions asked and a summary of responses.

While in general students were satisfied with the program and the course offerings, they did find some gaps in the curriculum.

Some of the requests were on the more application-oriented aspects of data science, e.g., more programming, data management, broader considerations of ethics, and data visualization. These are addressed in the new Applied Statistics and Data Science Degree (See **Section C.2.**) Some future students will prefer this program. Parts of this curriculum can be incorporated into our present master's program. The same could be said of applications of statistics and data science to domain science. Students can either choose that emphasis in the Applied Statistics and Data Science Degree or use these courses as a part of their doctoral minor.

Some of the requests are for courses like time series, non-parametric statistics, multivariate statistics, spatial statistics, and more advanced coursework in machine learning. These courses are in queue for new courses as more resources become available.

I.2.d. Classroom engagement strategies

We focus our classroom engagement strategies and the use of instructional technology on the six core courses. Four of the courses are the basis for the Program's Qualifying Exam. Consequently, these courses concentrate on the material tested by the exam and follow a more traditional approach.

Both the Regression class (STAT 571A) and the Machine Learning (MATH 574M) have group projects and presentations. The Consulting course (STAT 688A &688B) has meetings with statistical consulting clients and statistical consulting projects. This is preceded by a variety of skills building activities – written reports and oral communication, conducting a client meeting, negotiating, handling and resolving difficult situations, setting a research design, and using reproducible research methods.

Each teaching assistant undergoes a rigorous teacher training program by the Department of Mathematics. This entails a weeklong workshop before classes begin followed by classroom visits and close mentoring by an experienced lecturer. The undergraduate teaching workshop introduces many active engagement strategies – posing probing questions, encouraging collaboration in class, connecting material to the world outside the classroom, and the use of modern technology.

Starting 2021, we launched a series of "Professional Development Series" to help students develop important professional and interpersonal skills essential in research and practice of statistics and data science, to prepare students for their future careers as statisticians and data scientists. Our first series included lectures on resources for research, how to find research topics and advisors, introduction to Latex, jobs and internships, prepare CV and Interviews, career as a statistician and data scientist. We have brought faculty members to sit on the panels to meet the students, share experiences, and answer questions.

Our student representatives regularly organize student seminars each semester, providing a platform for students to share their valuable experiences with internships and job applications. Since 2022, the students have been organizing a journal reading group, a thriving forum for students to read and discuss papers on a regular basis throughout each semester. These activities not only enhance the students' academic experience but also strengthen the sense of community and collaborative learning among the students.

I.2.e. Use of instructional technology

Five of the six core courses (all but statistical computing) are taught in both a synchronous and asynchronous format. The lectures are recorded and archived using the D2L virtual learning environment. D2L is also used to receive student assignments and to provide other curricular material.

I.2.f. Online programs

The first four core courses (564, 566, 571A, and 571B) constitute a totally online certificate program. (Although other online options are available, primarily through Systems and Industrial Engineering.) In the near term, efforts for more online courses will generally be focused on the professional master's degree program. (See **C.2 Major changes in the past seven years**.)

I.2.g. Resources available for graduate students

The Mathematics Department provides space for the administration of the Statistics Program – the offices for the chair and the coordinator along with a Data Science Community Room. It also provides office space for Statistics and Data Science students primarily in the Mathematics Building, in the Environmental and Natural Resources Building II, or in the Physics and Atmospheric Sciences Building. Generally speaking, students who have research assistantships have space provided by the department that hosts the research faculty who is sponsoring the student.

I.2.h. Doctoral minors

All Statistics and Data Science doctoral students must complete a graduate minor and have a faculty member from the minor unit serve on the Comprehensive Examination Committee. The choice of minors is widespread and evolves over time. For example, Genetics minors were once more common and several of our recent graduates had minors in Mathematics, Applied Mathematics, or Computer Science.

We sometimes experience short coordination issues when a student elects a minor that had not been previously selected. This becomes more rare as the program continues. Many students across campus take our graduate courses and choose Statistics and Data Science as a doctoral minor. The table below shows the counts of declared minors. This is surely an undercount as several graduate students enrolled in statistics and data science courses have yet to declare a minor.

PhD Minor for those current SDS GIDP PhD students that have declared a minor.	
PhD Minor	Count
Biostatistics	1
Biosystems Analytics & Tech	1
Computer Science	1
Educational Psychology	1
Information	1
Optical Sciences	1
Systems & Industrial Engr	4

Graduate Minor Enrollees							
	Fall 17	Fall 18	Fall 19	Fall 2020	Fall 2021	Fall 22	Fall 23
PhD Minors	27	31	26	24	24	24	33

PhD Majors for those students with SDS Minor	
Fall 2023	
Major	Count
Aero & Mechanical Engineering	1
Applied Mathematics	4
Biostatistics	1
Civil Engr & Engr Mechanics	1
Computer Science	3
Electrical & Computer Engineering	2
Information	3
Linguistics	1
Management	3
Mathematics	4
Physics	1
SIE	9
Total	**33**

The breadth of interest in the core courses is extensive with student representation from anywhere between 5 to 15 academic programs taking the Statistics and Data Science core courses. A listing of those academic units for these courses can be found in Appendix 8

See **Appendix 8** for a complete list of majors represented in attendance at Statistics core courses.

I.2.i. Graduate student handbook

The graduate student handbook is **Appendix 5.**

I.3. Graduate Students

I.3.a. Recruitment

Recruitment is an around the year activity. The University provides leads for prospective students. In turn, the Program Coordinator sends out automated emails sent periodically to those that have expressed an interest in the program. These emails announce office hours for prospective students and all of the leads are encouraged to contact the Program Coordinator if they have any questions. Traffic is high during the fall semester with regular communication with prospective students. January 1st is the deadline for applications for the fall semester. With some limited exceptions, we do not accept applications for the spring semester.

The admissions committee reviews the applications and admissions decisions are made by the middle of February. The admissions committee ranks the PhD accepted students so that the program can make funding offers. About six accepted PhD applicants are invited to an in-person recruitment event in March. GRE quantitative scores average ~90th percentile and verbal scores ~75 percentile.

This in-person recruitment event is jointly sponsored by all three units in the School of Mathematical Sciences and the applicants in all three programs can see the breadth of the mathematical sciences and have time to socialize with other recruits and other students at meals and field trips to the Desert Museum or on a hike.

The recruits will have a small group meeting with the Program Chair, with Admissions Committee members, and with potential advisors based on the information provided in the Statement of Purpose. They will also attend two research blitzes. One is three faculty presentations from each of the three graduate programs. The other is a series of short presentations from Statistics and Data Science students chosen to align to the recruits' stated interests.

About half of the offers made are accepted. We generally compete well with closely ranked institutions for students. Candidates choose the University of Arizona for a variety of reasons – alumni recommendations, interaction with staff, personal reasons, and the quality of life in Tucson. The nature of an interdisciplinary program plays an important role – course selection and research directions are more open. For many students, our goal is to have a co-mentoring approach, a core statistician and a domain scientist advising together on a research question that promotes knowledge in both statistics and in the student's chosen scientific domain.

	Applicants				
`	2019	2020	2021	2022	2023
Certificate	10	5	3	3	7
M.S.	20	24	45	32	27
A.M.P.	2	1	3	8	3
PhD	27	26	24	20	35
PhD Info	4	6	4	5	2
Totals	63	62	79	68	74

	New Students				
	2019	2020	2021	2022	2023
Certificate	6	1	2	0	5
M.S.	12	8	12	4	10
A.M.P.	1	1	1	4	1
PhD & PhD Info	4	4	5	4	7
Totals	23	14	20	12	23

I.3.b. Demographics

Enrolled graduate students by sex							
Sex	2016	2017	2018	2019	2020	2021	2022
Female	13	14	19	18	18	21	18
Male	35	44	38	40	39	31	29
Grand Total	**48**	**58**	**57**	**58**	**57**	**52**	**47**

Enrolled graduate students by IPEDS race/ethnicity

IPEDS Race/Ethnicity	2016	2017	2018	2019	2020	2021	2022
American Indian or Alaska Native		1					
Asian	3	5	4	4	4	5	4
Black or African American				1	1		
Hispanic or Latinx	4	6	3	4	7	4	3
International	13	14	23	22	21	28	28
Not reported	5	3	2	5	7	4	3
Two or more races	1	1	1	1	1		
White	22	28	24	20	16	11	9
Grand Total	**48**	**58**	**57**	**57**	**57**	**52**	**47**

This distribution is not significantly different from national averages with roughly 65% of Statistics students are international and 61% of the domestic students identified as White. We have seen a recent decrease in White and Hispanic students. The new Statistics and Data Science undergraduate degree and the Data Science and Applied Statistics professional master's degree has given the graduate program the opportunity to recruit among the existing University of Arizona students. Success in this activity will increase domestic enrollment. With Arizona a Hispanic Serving Institution, we can expect more Hispanic student enrollment. We work with the Associate Dean, Student Affairs, Diversity & Inclusion, to offer financial and other incentives to underrepresented communities.

I.3.c. Stipends and assistantships

Every doctoral student in good standing has received both tuition remission and an assistantship (teaching or research) or fellowship support. We sometimes have money provided by the graduate college to support some master's students. Our stipends are often not competitive against our peer institutions and we lose some students due to the differences in financial support amounts.

Statistics & Data Science GIDP
Student Funding History

		2017	2018	2019	2020	2021	2022
Math TAs	Number	6	7	9	7	10	11
	FTEs	2.75	3.5	4.5	3.5	4.5	5.5
RAs	Number	8	6	4	7	6	11
	FTEs	3.58	3.25	2	3.25	3	5.25

Online	Number	2	2	2	0	0	0
	FTEs	1	1	1	0	0	0
GIDP	Number	2	2	2	2	3	2
	FTEs	1	1	1	1	1.25	1
RA Sources		Medicine NIH Grant University Fellows Biostats Medicine	Biostats NIH Grant Medicine Tripods Internship MCB	Tripods Tripods Medicine Medicine	Medicine Biosys Eng RTG SIE Ed Psy	RTG SIE Sch of Info Ed Psy BIO5 Ed Psy	Cancer Math Health Svc Biosys Eng RTG Medicine Ed Psy BIO5 Pharmacy Biostats
Total Number		**18**	**17**	**17**	**16**	**19**	**24**
Total FTE		**8.33**	**8.75**	**8.5**	**7.75**	**8.75**	**11.75**

Student Assistantship Salary History			
	2021	2022	2023
Level	**0.5 FTE Salary**	**0.5 FTE Salary**	**0.5 FTE Salary**
Grad Assistant I	$19,278	$20,000	$21,750
Grad Assistant II	$19,706	$20,500	$22,500
Grad Associate	$20,135	$20,500	$23,250

Level Descriptions:

- All new Graduate Assistants start as **Graduate Assistant I**
- After a year of satisfactory academic and teaching (or research) performance, students who maintain a minimum 3.5 GPA or who have passed the Qualifying Exam at the PhD level will be promoted to **Graduate Assistant II**.
- Students who have submitted their Comprehensive Exam Committee Appointment Form and Plan of Study in Grad Path and have earned at least 30 units towards their PhD program will be promoted to **Graduate Associate**.

  Additional Information
  - The TA/GA positions are generally 9 months.
  - First year students receive a $3,000 summer stipend at the end of their first year.
  - The GA/TA position comes with paid tuition and health insurance. Students must pay their fees which were $835 for the fall 2023 semester.

I.3.d. Faculty advising and mentoring

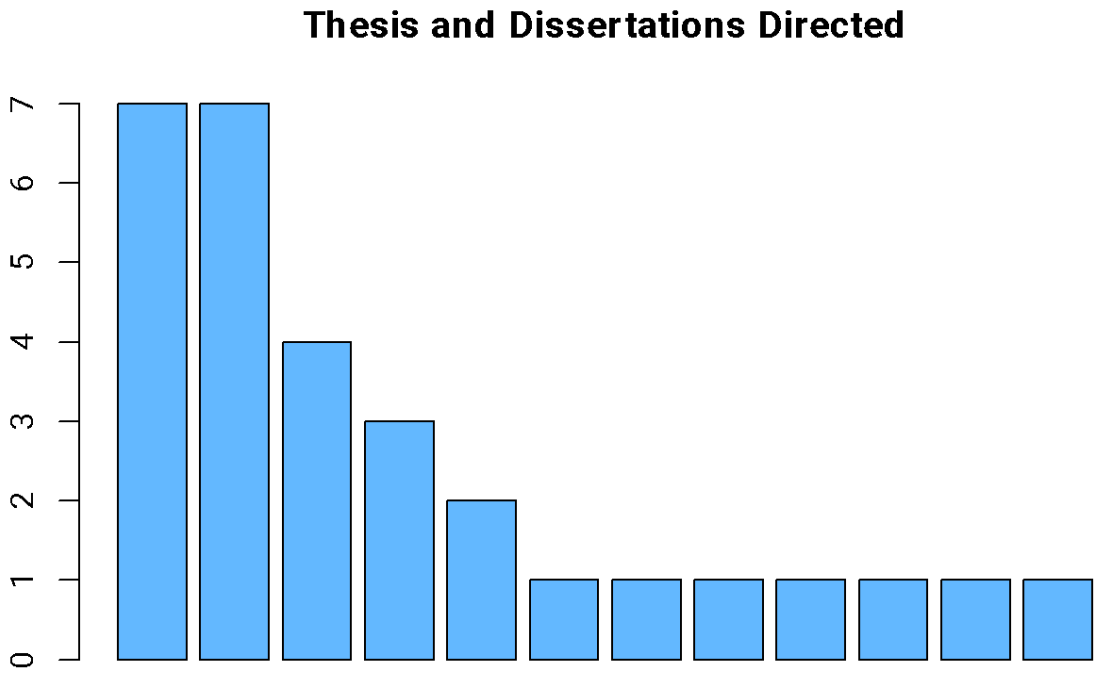
Every entering student is provided with an initial faculty advisor and a peer mentor. For master's students, these students are even split between the program chair and vice chair. The

entering doctoral students are advised by the member of the executive committee whose interests, as indicated by the personal statement, most closely aligns with the committee member. The peer mentors for the master's students are the two graduate representatives for the program. For doctoral students, the peer mentor is an advanced graduate student whose interests align with the student. The program provides a small budget for the peer mentors for a social activity (lunch or coffee).

The students Individual Development Plan is largely considered as a part of the annual progress report. The procedure begins with a meeting between the student and the advisor (initial for entering students, thesis or dissertation advisor for advanced students). For this report we gather basic information about the student (expected graduation date, funding source, academic minor, GPA, planned activities for the communications requirement, goal for the next year, increased understanding of the theoretical and methodological aspects of statistics and data science, self-assessment of progress based on goals for the past year, statement of goal for the next year). The advisor signs off on the report and the program vice chair review and coordinator review the responses to ensure that students are making good progress and to bring to the Executive Committee any issues. This is one the major avenues that the Program uses to effect change.

The second approach is to have the Program representatives hold focus sessions to assess the quality of the program. These two approaches help us prioritize new course offerings and assess the appropriateness of courses in other academic units. Thus, the Executive Committee works to add some courses and occasionally drop a course. This input also played an important role moving from a *communication requirement* to a *professional development requirement*. (See **Section I.2**). In all of these avenues, the Program endeavors to support the students' goals in career pathways and research directions along with the needed professional development to reach their goals.

Most of the thesis and dissertation chairs come from a small number of the Statistics and data Science faculty. The distribution of the last 30 theses/dissertations are displayed in the Pareto chart below. Note that only 12 members of the faculty directed a thesis or dissertation in this time frame. The student choices may be motivated by who they have as instructors in core courses or by the topics in statistics and data science that are currently gathering the most attention, e.g., machine learning or bioinformatics.



I.3.f. Scholarly activities

Students are strongly encouraged to join scholarly societies and present at conferences. We will assist them in applying for financial support to attend either through the sponsoring societies for the conferences or through special funds set aside for students in Interdisciplinary Programs. In addition, the Statistics and Data Science Program has ~$3000 in funding to support students.

We prioritize students in their final year as one way to help them make connections with their first post graduate job. During the pandemic, we substituted travel stipends with guaranteeing payment of conference fees. With reserved funding, the Statistics and Data Science Program purchased a node on the university's high performance computer.

Most doctoral students publish one or more papers based on their dissertation. Here is a sample based on student responses to the survey in **Appendix 10**.

[redacted]

[redacted]

I.3.f. Trends, time to degree and number completing the degree

The time to degree for both master's and doctoral programs has been relatively stable over the last eight years. We do have some master's students who are in the program part time and so take longer to complete the degree. The doctoral program had higher times to degree in the early years. As the quality of our students improved, the time to degree decreased. There is an uptick at the end that can likely be attributed to the pandemic.

MS Time to Degree			
	Average	Median	No of Grads
2014	3.44	2.94	3
2015	1.85	1.72	5
2016	2.32	1.92	5
2017	2.07	2.32	4
2018	1.64	1.66	4
2019	2.42	1.72	8
2020	2.20	2.03	12
2021	2.48	1.72	7
2022	2.99	2.72	7
2023	2.04	1.72	5
Total	2.34	1.92	60

PhD Time to Degree			
	Average	Median	No of Grads
2015	4.73	4.73	2
2016	6.34	5.97	3
2017	5.01	4.73	5
2018	5.97	5.61	3
2019	5.46	5.46	2
2020	5.57	6.00	3
2021	4.73	4.73	1
2022	4.85	4.61	3
2023	5.75	5.50	4
Total	5.44	5.16	26

MS Completion		
Average		2.44
6 Year Completion	47 of 48*	97.92%
8 Year Completion	48 of 48	100.00%
* The one exception was a student that was working on a PhD in another discipline concurrently with the MS in Statistics		
PhD Completion		
Average		5.32
6 Year Completion	21 of 26	80.77%
8 Year Completion	26 of 26	100.00%

Graduate Enrollment by Academic Plan							
	2017	2018	2019	2020	2021	2022	2023
Certificate	25	20	9	10	4	4	5
M.S.	21	26	27	31	25	21	21
Ph.D.	20	22	24	21	24	26	26
PhD Minors	27	31	26	24	24	24	24
Total	93	99	86	86	77	75	76

MS and PhD Total	41	48	51	52	49	47	47	
Graduate Completed Majors by Academic Plan								
	2017	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**Total**
Doctorate	4	4	2	3	3	0	4	**20**
MS	7	4	11	6	13	10	9	**60**
Minors	4	11	6	7	9	7	4	**47**

PhDs OUTSIDE of the discipline who minored in Statistics or Statistics & Data Science

	2017	2018	2019	2020	2021	2022	2023
PhD completers in other disciplines who minored in prompted plan	4	11	6	7	9	7	8
Total PhD completers in other disciplines	419	398	409	424	425	406	459
Proportion	0.95 %	2.76 %	1.47 %	1.65 %	2.12 %	1.72 %	1.74 %

There have been a few students that have not completed their SDS GIDP degree. One left for [redacted] and two for [redacted] There were an additional two students that moved to a different degree and there was one PhD student that did not pass the qualifying exam.

I.4. Graduate-Student Learning Outcomes Assessment

LO1 Theory of Probability and Statistics Understanding

Students demonstrate understanding of the key concepts in the theory of probability and statistics and can communicate that understanding through a well constructed theoretical argument.

Measure - The theory portion of the qualifying exam includes knowledge from two courses, STAT 564 (Theory of Probability) and STAT 566 (Theory of Statistics)

Direct Measure - *Theory of Probability: STAT 564, Target: 85% of PhD and MS students will pass the Theory portion of the qualifying exam.*

LO2 Statistical Methodology Understanding

Students demonstrate understanding of the key concepts in the statistical methodology and can communicate that understanding through effective experimental design and sophisticated use of statistical and computational tools.

Measure - The methodology portion of the qualifying exam includes materials from STAT 571A (Advanced Statistical Regression Analysis) and STAT 571B (Design of Experiments)

Direct Measure: *Advanced Statistical Regression Analysis: STAT 571A, Target: 80% of PhD and MS students will pass the methodology portion of the qualifying exam*

LO1 and LO2 Results:

M.S.	Theory			Methodology		
	Pass	Fail	Percent Pass	Pass	Fail	Percent Pass
August 2020	6	0	100%	6	0	100%
January 2021	0	1	0%	1	1	50%
August 2021	3	2	60%	4	1	80%
January 2022	2	5	29%	5	0	100%
August 2022	6	1	86%	5	2	71%
January 2023	0	1	0%	0	1	0%
August 2023	6	1	86%	6	0	100%
Total	**23**	**11**	**68%**	**27**	**5**	**84%**

PhD	Theory			Methodology		
	Pass	Fail	Percent Pass	Pass	Fail	Percent Pass
August 2020	4	0	100%	4	0	100%
January 2021	1	0	100%	1	0	100%
August 2021	3	0	100%	3	0	100%
January 2022	0	0		0	0	
August 2022	6	0	100%	3	3	50%
January 2023	0	0		0	0	
August 2023	3	0	100%	3	0	100%
Total	**17**	**0**	**100%**	**14**	**3**	**82%**

LO3 - Research Ideas

Students develops creative and innovative research ideas and approaches that can further the body of statistical knowledge and contribute to significant advances in the intended field of application. PhD Students develop creative and innovative research ideas and approaches that can further the body of statistical knowledge and contribute to significant advances in the intended field of application.

Measure - Thesis Topic or PhD Dissertation

MS students will either pass the qualifying exam with a score of "MS Pass" or "PhD Pass" or they can choose to do a thesis. Those that do a thesis will be graded on a scale of 0 to 5 for "Research Ideas".

Direct Measure: MS Students will achieve a score of at least 4 (In a 0 to 5 scale) for the topic of their MS thesis. PhD Student's dissertations are assessed by the faculty. If the student passes their dissertation defense, they meet this requirement.

Results: All PhD students have passed their dissertation defense and all MS students that do a thesis have passed their thesis defense.

LO4 Communication

Students clearly communicate statistical ideas, both written and oral, and adapts the presentation to be suitable for the intended audience.

All PhD and MS students are required to complete a set of communications skills requirements as outlined on the Statistics webpage (https://stat.arizona.edu/communications-skills-requirement). Faculty complete a survey when each of the requirements is met. It is important to note that the faculty do not fill out the survey unless the student meets the requirements so all of the students will meet or exceed the communications requirements.

Indirect Measure: The students are assessed on a 5 point scale. All students must meet or exceed 4 out of 5 or 80%.

Results: The average score on the following question for communications requirements was 4.2. Student communications activities are scored by a faculty member.

> *Student clearly communicates statistical ideas, both written and oral, and adapts the presentation to be suitable for the intended audience.*

Over the past seven years, the country has experienced a sharply increased demand for statistically well-trained data scientists. At the same time, both local and national trends for undergraduates in statistics and data science are still in an exponential growth phase. The expectations of graduates have evolved on at least two fronts – graduates are expected to have a solid background in machine learning and have more advanced computing skills. Moreover,

the field of data science is broadening and so, graduates need to be able to effectively articulate their capabilities to future employers. In addition, the COVID pandemic forced every program at every institution of higher education to rethink its mentoring profiles. As a consequence, the Statistics and Data Science Graduate Interdisciplinary Program made the following changes:

- Added Statistical Machine Learning (MATH 574M) to the list of required courses.
- Instituted a system of mentoring with the chair or vice chair assigned as initial advisors for master's student and an individual mentor usually selected from the Executive Committee to serve as initial advisors for doctoral students
- Created a "buddy" system that pairs an entering student with an advance students to provide peer mentoring opportunities
- Redesigned the communications requirements to a professional development plan
- Work with the programs across campus to broaden the course offerings available for our students

Going forward, we will

- Look more broadly, especially in regional undergraduate statistics, data science, and mathematics students as a way to broaden the participation in our graduate program.
- Cooperative effectively with the new Professional Master's Degree for strategies to improve the more practical aspects of data science, e.g., more advance computations or data management courses, more extensive internship, and professional development activities.

The needs of the present and future workforce for those who have command of both classical and morn statistically theory and methodology will certainly continue to grow rapidly. The Statistical and Data Science Graduate Program is committed to continue its self-examination to ensure that we are providing the best education possible in a rapidly evolving environment.

I.5. Post-Doctoral Fellows

Graduate Interdisciplinary Programs do not have a budget for postdoctoral fellows. The TRIPODS grant resulted in the hiring of three to four postdoctoral fellows. The Research Training Group in Data Driven Discovery also supports a postdoctoral fellow. These individuals are thoroughly integrated into the research lives of both TRIPODS and the RTG, leading discussion groups, serving as mentors to graduate students, and appearing in a number of publications. Each of these fellows serves for three years - some to four under the difficulties in moving jobs during the pandemic. These fellows now work in industry (Upstart, Sanofi) or academia (University of Texas-Arlington, Saint Louis University).We plan to continue to seek support for postdoctoral fellows through center grant proposals.

Appendices