ST 790 Navigating the PhD program and beyond: perspectives, skills, and strategies

Jonathan P Williams

North Carolina State University

Fall 2025

Transitioning from coursework to research

After passing the qualifying exam:

→ Developing your dissertation research is the most important aspect of your graduate studies

→ Insofar as receiving passing grades, courses are no longer the highest priority

→ Future employers will evaluate you based on the quality of your dissertation research

Note:

While many people with a PhD degree in statistics are choosing to work in industry, the purpose of a PhD degree in statistics is to train you as a researcher

A PhD is *not* a professional degree (e.g., Medical Doctor)

Transitioning from coursework to research

Timeline of next steps:

1. Narrow down your areas/types of potential research interest
   $\rightarrow$ Will overview areas later

2. Find 1-2 PhD advisors
   $\rightarrow$ Begin working on a first project
   $\rightarrow$ Might spend 6-12 months on background reading

3. Schedule written preliminary exam
   $\rightarrow$ Within $\approx$ 18 months of beginning research
   $\rightarrow$ Assemble your PhD committee
   $\rightarrow$ $\approx$ 5 faculty members, mostly from your department
   $\rightarrow$ Your advisor(s) are your PhD committee chair(s)

Transitioning from coursework to research

Timeline of next steps (continued):

4. Complete $\approx$ 75% of dissertation research
   $\rightarrow$ Schedule oral preliminary exam with committee
   $\rightarrow$ Present what you have already accomplished
   $\rightarrow$ Propose what the remaining 25% will look like

5. Complete $\approx$ 99% of dissertation research
   $\rightarrow$ Schedule oral final defense with committee
   $\rightarrow$ Present your dissertation work
   $\rightarrow$ Argue it is substantial enough to earn your PhD degree

6. Submit your dissertation manuscript to the university
   $\rightarrow$ Ask senior students for the university-compliant .tex file

Transitioning from coursework to research

Types of statistics research:

- $\rightarrow$ Theoretical or mathematical statistics
- $\rightarrow$ Machine learning or statistical learning
- $\rightarrow$ Statistics methodology
- $\rightarrow$ Applied statistics
- $\rightarrow$ Computational statistics
- $\rightarrow$ Statistical software

Note:

This list does not include statistical applications or collaborative research published in domain science journals

Transitioning from coursework to research

Theoretical or mathematical statistics:

$\rightarrow$ Investigations of theoretical or mathematical properties of estimators or computational tools

$\rightarrow$ Formulations/justifications for a paradigm of statistical inference. E.g., frequentist, Bayesian, fiducial

$\rightarrow$ etc.

$\rightarrow$ No immediate applications necessary

Top journals include:

Annals of Statistics (AoS)
Bernoulli

Transitioning from coursework to research

Machine learning or statistical learning:

- $\rightarrow$ Use data to train algorithms to perform tasks
- $\rightarrow$ Particular emphasis on prediction problems/tasks
- $\rightarrow$ Algorithm development
- $\rightarrow$ Theoretical and empirical performance metrics/evalaution
- $\rightarrow$ Unsupervised learning

Top journals include:

Journal of Machine Learning Research (JMLR)
Many prestigious conference proceedings (e.g., NeurIPS, ICML)

Transitioning from coursework to research

Statistics methodology (most common type):

→ Propose a new estimator/approach for making inference on population quantity of interest

→ Simulation study to investigate empirical properties of the proposed method

→ Formulate and prove theorems to guarantee consistency or other optimality properties of the proposed method, under certain assumptions

→ "Real data" implementations and proof of concept

Top journals include:

Journal of the Royal Statistical Society: Series B (JRSS B)
Journal of the American Stat Assoc: Theory and Methods
Biometrika

Transitioning from coursework to research

Applied statistics:

- → Method development/evaluation motivated by a real data set and/or questions of interest with considerable practical relevance in some application
- → Not necessarily methodologically novel
- → Illustration of important aspects of existing methods
- → Important case studies or comparisons

Top journals include:

Journal of the American Stat Assoc: Appl and Case Studies
Annals of Applied Statistics (AoAS)
Journal of the Royal Statistical Society: Series C (JRSS C)

Transitioning from coursework to research

Computational statistics:

→ Algorithms for implementation of estimation routines

→ Issues relating to computational efficiency versus statistical efficiency

→ Theoretical properties of algorithmic convergence

Top journals include:

Journal of Computational and Graphical Statistics (JCGS)

Transitioning from coursework to research

Statistical software:

$\rightarrow$ R package development

$\rightarrow$ Open-source statistical software development, more generally

$\rightarrow$ Demonstration/comparison of existing software

Top journals include:

Journal of Statistical Software

Transitioning from coursework to research

Areas of statistics research:

. . . very many.

Here are the "major areas" of research in our department:

`https://statistics.sciences.ncsu.edu/research/research-areas/`

Transitioning from coursework to research

Things to consider in choosing an advisor:

- → Type/area of research focus
  - → But be careful not to overemphasize this one...
- → Personal compatibility
  - → It is difficult to work with someone that you find difficult to interact with
  - → You'll meet $\approx$ weekly for the next 4 years
  - → You'll eventually need a strong letter of recommendation from them; so it's important they like you, as well

Transitioning from coursework to research

Things to consider in choosing an advisor (continued):

- → Their work ethic and intensity of expectations
  - → If you only want to work 30-40 hours per week, then you're never going to impress your advisor if she/he works around the clock
  - → Look for an advisor with a likeminded attitude about work–life balance
- → Feedback from current advisees
  - → So long as $n > 1$, this is perhaps the best calibrated source of information for a glimpse into what your experience with a potential advisor might be like

Transitioning from coursework to research

Things to consider in choosing an advisor (continued):

- → Advisor's network
  - → Do their students tend to get jobs in careers you are aiming for?
  - → Some faculty send almost all students to industry
  - → Some have better connections in academia or industry
- → Resources available from the potential advisor
  - → Can they fund you as an RA?
  - → Do they have funds for you to travel to present your research?
  - → Do they work with collaborators in domain sciences of interest to you?

Transitioning from coursework to research

Things to consider in choosing an advisor (continued):

- → Amount of interaction you need
  - → Some advisors meet with each student for 30 min/week
  - → Some advisors are willing to meet 4-5 hours/week
  - → In part, depends on how many other students are advised
    - → The number of students a faculty member chooses to advise in a given year gives an indication of how carefully they choose to think about research problems
    - → Also indicates how active the faculty member is

Transitioning from coursework to research

Things to consider in choosing an advisor (continued):

→ You are exclusively your own best advocate for you
  - → Don't expect that your advisor will make you aware of all that you need to be aware of
  - → Don't expect your advisor to always be correct
  - → Don't expect your advisor to always know best
  - → But you need to be able to trust their judgement
  - → Your advisor is as human as you are, proceed as such

Transitioning from coursework to research

Things to consider in choosing **to be an adult**:

- → Whatever choices you make:
  - → Sometimes you will have to work more hours in a day/week/month/year/etc. than you want to
  - → Oftentimes you will have to do work you don't want to
  - → Your work should be about more than how it benefits you; we live in a society
  - → Aiming for purpose, satisfaction, and fulfillment is more sustainable than aiming to feel happy, on any given day

Transitioning from coursework to research

Alas, don't worry too much about making the "correct choice" of advisor or research topic

→ A short list of top choices that are seemingly impossible to decide amongst, are *actually* impossible to decide amongst

  → There is not necessarily a single correct choice to make, even if the future was known

  → The only *incorrect* thing to do is to delay deciding

→ Most fundamentally, writing a dissertation is an exercise in *learning* to do research

  → The importance of the research outcomes are ancillary

  → Future directions are not limited by dissertation topics

Monte Carlo simulation studies versus mathematical proofs

A typical framework for statistical research is as follows.

1. Begins with a population and questions of interest
2. Population features are formulated and quantified in relation to the questions of interest
3. Data relevant to the population features of interest are collected
4. Statistics (i.e., functions of the data) are formulated to use the data to make inference on the population features of interest in a manner that is optimal in some way
   $\rightarrow$ e.g., least biased, most efficient, most powerful, most consistent, etc.

Monte Carlo simulation studies versus mathematical proofs

Research might be done to choose or formulate an estimator

As a research statistician, much of the work is to establish the properties of the chosen/formulated estimator

This work can be approached in a few ways:

→ Gold standard: properties established by mathematical proof

→ Simulation studies:

- → Helps to develop intuition for proofs
- → Drives intuition for reformulating/adjusting estimator
- → Can be used if proof is too complicated
- → Support arguments used in proof
- → Demonstrate concepts or strange phenomena

Monte Carlo simulation studies versus mathematical proofs

Consider a simple example:

$\rightarrow$ Population of measurements $\sim$ normal$(\mu, 1)$
$\rightarrow$ Unknown population feature μ
$\rightarrow$ Perhaps use a sample mean or median to make inference on μ

What are the properties of the sample mean $\overline{X}_n$ for estimating μ?

What are the properties of the sample median M_n for estimating μ?

Monte Carlo simulation studies versus mathematical proofs

Theorem

The sample mean of iid normal$(\mu, 1)$ *data follows the normal*$(\mu, 1/n)$ *distribution.*

Proof. If $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{normal}(\mu, 1)$, then each X_i/n has a moment generating function of the form $m_{X_i/n}(t) = e^{\mu t/n + (t/n)^2/2}$. By independence,

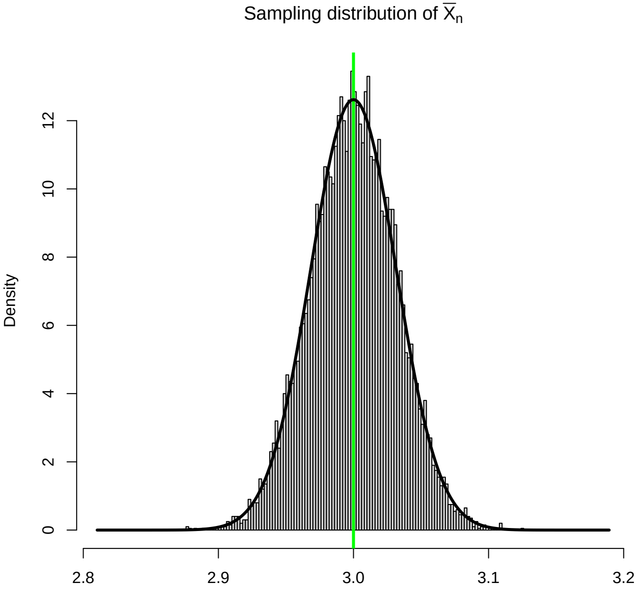
$$m_{\overline{X}_n}(t) = \prod m_{X_i/n}(t) = e^{\mu t + (1/n)t^2/2},$$

so that $\overline{X}_n \sim \text{normal}(\mu, 1/n)$. ■

Monte Carlo simulation studies versus mathematical proofs

```r
library(latex2exp)

mu = 3
sigma = 1
n = 30

# Simulate a large number of data sets and least squares estimators
num_sims = 300
x_bar_vec = rep( NA, n=num_sims)
for(k in 1:num_sims){
  y = rnorm( n, mean=mu, sd=sigma)
  x_bar_vec[k] = mean(y)
}

# Plot the sampling distributions of the estimator
upper = mu + 6*sigma/sqrt(n)
lower = mu - 6*sigma/sqrt(n)
grid = seq( lower, upper, by=.01)

hist( x_bar_vec, freq=F, main=TeX(r'(Sampling distribution of $\bar{X}_{n}$)'),
      xlab=NULL, xlim=c(lower,upper), breaks=floor(sqrt(num_sims)))
abline( v=mu, col="green", lwd=3)

lines(grid, dnorm( grid, mean=mu, sd=sigma/sqrt(n)), lwd=3)
```

Monte Carlo simulation studies versus mathematical proofs



Scientific writing: general principles

Rough outline of a typical statistics publication:

- Section 1. Introduction
- Section 2. Methods
  - Subsection 2.1. Algorithms
- Section 3. Theoretical results
  - Subsection 3.1. Proofs
- Section 4. Empirical results
  - Subsection 4.1. Numerical illustrations
  - Subsection 4.2. Simulation studies
- Section 5. Real data analyses
- Section 6. Concluding remarks and future work
- Appendix A. Additional proofs
- Appendix B. Additional figures, tables, algorithms etc.

Scientific writing: general principles

Things to consider when writing a title and abstract

Scientific writing: general principles

Link to TeX: `https://en.wikipedia.org/wiki/TeX`

Link to Overleaf: `https://www.overleaf.com/`

Scientific writing: general principles

The role of mathematical notation in writing about mathematical and statistical ideas

`https://jonathanpw.github.io/ST790/Marron1999.pdf`

Scientific writing (and reading): literature reviews

Generally, 4 levels of depth to reading a statistics research article:

1. Title + abstract
2. Title + abstract + introduction
3. Full manuscript
4. Full manuscript + appendices + proof details

Scientific writing (and reading): literature reviews

How to approach learning about new topics?

→ Usually start with a key reference(s) from your advisor, a colleague, a collaborator, etc.

→ Forward and backward citation search of key articles

→ Keyword search in a repository (e.g., Google scholar)

→ Decide on the reliability of a found article:

- → Do the authors have established credibility on the topic?
- → Is the article published in a relevant journal?
- → Should you trust preprints less then publications?

Scientific writing (and reading): literature reviews

How to approach learning about new topics? (continued)

→ Reach out to authors

- → Quick questions over email
- → Non-quick questions over Zoom or meet for a coffee, e.g., at a conference if non-local.
- → Most serious researchers enjoy having conversations about their work; I'm happy to talk about my work if anyone wants to come by my office

→ Find good literature review articles; usually titled:

- → "Survey of ...", "Primer on ...", "Tutorial on ...", etc.
- → Journal of the American Statistical Association: Reviews
- → Statistical Science

arXiv challenge

Start every workday by scrolling through all new submissions appearing in the Statistics topic section of arXiv:

`https://arxiv.org/list/stat/new`

$\rightarrow$ There are typically $\approx 30 - 40$ new articles each day

Scientific writing (and reading): literature reviews

arXiv challenge

→ As you scroll, read each article title and author list

→ Skim the abstract if:

  → the title sounds interesting

  → it's an author that you tend to appreciate

→ Of the abstracts read, if is compelling enough, open article:

  → maybe it's on a topic of interest

  → maybe it's relevant for a current/future literature review

  → maybe it's on a topic that you hadn't heard of

→ Of the articles opened, decide how much of them to read

  → recall the previously discussed levels of reading depth

Scientific writing: literature reviews

Things to consider in writing a literature review:

→ How broad is the audience?

→ Trying to establish credibility in an area?

→ Trying to establish relevance of an idea?

→ Trying to be informative?

→ Does it have to be exhaustive?

→ Scope versus depth of each article discussed in the literature review

Scientific writing: academic publishing and the purpose of journals

Challenge question:

What is the purpose of academic journals?

$\rightarrow$ Communication and discourse on discovery of knowledge?

$\rightarrow$ Archival and documentation of knowledge?

$\rightarrow$ Quality control?

Scientific writing: academic publishing and the purpose of journals

Other things to consider about academic journals:

→ Governance and censorship of knowledge dissemination?

→ Who has access to journal articles?

→ Who pays for the research leading to the journal articles?

- → Do they have free access?
- → Consider the story of sci-hub: link
- → Radiolab episode: link

→ What articles are excluded from journals?

- → Negative results?
- → Unpopular ideas? (…fiducial inference ideas?)

Scientific writing: academic publishing and the purpose of journals

Editorial board of a journal:

Editor(s)

→ Typically 1-2 serve on a limited term basis (e.g., two years)

→ Chosen by committees of other academics/researchers

→ Unpaid

Associate Editors (AEs)

→ Typically many dozens

→ Chosen by the Editor(s)

→ Unpaid

Scientific writing: academic publishing and the purpose of journals

Peer-review process:

→ Author submits manuscript to journal

→ Editor decides on further review or desk reject (∼1-2 weeks)

→ In the case of further review:

1. Editor determines/selects an appropriate AE
2. AE takes a closer look and decides:
   - 2.1 Reject with comments formulated (∼1-3 pages; ∼1-3 months)
   - 2.2 Further review by peers (∼3-8 months)
     - 2.2.1 AE identifies appropriate reviewers ("unbiased" experts)
     - 2.2.2 AE solicits recommendations from ∼2-3 appropriate reviewers
3. AE provides a recommendation to Editor
4. Editor makes a final decision:
   - 4.1 Usually reject
   - 4.2 Sometimes major revision; authors revise-and-resubmit

Scientific writing: academic publishing and the purpose of journals

Things to consider in navigating the peer-review process:

$\rightarrow$ Identify an appropriate journal for type/area of research

$\rightarrow$ Identify an appropriate journal for depth of presentation

$\rightarrow$ e.g., AOS versus Statistics and Probability Letters

$\rightarrow$ Journal prestige and impact factor

$\rightarrow$ Timeliness of the process for a given journal

Scientific writing: academic publishing and the purpose of journals

Things to consider in navigating the peer-review process (continued):

→ Known opinions/biases of the current Editor(s)

→ You can suggest reviewers & who not to review

→ If rejected, can always submit somewhere else

→ Post a preprint (e.g., arXiv or Researchers.One)

- → Immediate dissemination of your work
- → Timestamp on your work
- → Preprints are cited in the mathematical sciences

Open source research practices

→ Open source code repositories such as GitHub

→ Make a Google Scholar account

→ Get an ORCID

→ Make a personal academic website

→ Make research papers and code available on academic website

  → Idea of *time-capsule code*

ASSIGNMENT: create your Google Scholar profile (and a website if thinking about academia)

Academic presentations

- → Conferences and roles
  - → organizer, chair, presenter, discussant, panelist, etc.
- → When and how to go to conferences
- → Student paper competitions
- → How to get funding for travel
- → Department student seminar
- → Job talk

Academic presentations

→ Poster presentations

→ 10-15 minute talk

→ 20-30 minute talk

→ 45-60 minute talk

Academic presentations: Poster presentations

Usually as part of a conference

→ Invited session

→ Contributed session

Can be more fruitful than a talk because of one-on-one interactions

There's often a student poster session as part of every conference

→ Submit a poster if you want to attend

→ Many conferences also offer some funding to students

Use your time in grad school to travel and meet new people!

Academic presentations: 10-15 minute talk

→ Short talk

→ Usually as part of a contributed session in a conference

- → Approximately 90 minute session with 6 speakers
- → Common research theme of talks within session
- → "Contributed" usual means automatically accepted

Objectives:

- → Big picture overview of project; omit technical details
- → *Define the one thing you hope the audience takes away*
- → Hammer that point early and often
- → Rehearse a lot to get the timing down
- → **Avoid mathematical notation as much as possible**

Academic presentations: 20-30 minute talk

→ Typical-length talk

→ Usually as part of an invited session in a conference

- → Approximately 90 minute session with 3-4 speakers
- → Common research theme of talks within session
- → "Invited" usual means selected (solicited or unsolicited)

Objectives:

- → Big picture overview; nod to some technical details
- → *Define the one thing you hope the audience takes away*
- → Spend some time on motivating that point
- → Still rehearse, but you can breathe a few times during talk
- → **Avoid mathematical notation as much as possible**

Academic presentations: 45-60 minute talk

→ Full-length talk

→ Usually a seminar, keynote, or job talk

- → Sole session speaker
- → Important research topic
- → Typically in recognition of one or more contributions

Objectives:

- → Big picture overview; tell a story about the technical details
- → *Define the one thing you hope the audience takes away*
- → Lots of motivation, simple examples, and pictures
- → **Avoid mathematical notation as much as possible**

Academic presentations

Humans understand stories . . .

→ That's basically all we understand

→ Powerful stories are those that connect to bigger stories

→ You'll understand a story better if it connects to stories you already know; so will your audience

When the mathematician expresses frustration that something in mathematics is hard to understand, they are really admitting a lack of capacity for understanding

what the story is.

When a student expresses frustration in learning mathematics, it is because they have not yet understood that a story exists

Academic presentations

The *don't do's of presentations*:

- $\rightarrow$ Mathematical notation (to the greatest extent possible)
- $\rightarrow$ Over-crowded slides (students always do this)
- $\rightarrow$ Full sentences
- $\rightarrow$ Tiny or colored font
- $\rightarrow$ Distracting backgrounds
- $\rightarrow$ Going over time
- $\rightarrow$ Preparing your slides the day of

Academic presentations

How to handle questions:

- → Don't be afraid to pause and think
  - → good questions don't have quick answers
  - → wisdom comes from thinking, not speaking
- → You don't need to know the answer
  - → just have something insightful to say
- → Invite further dialogue
- → If the question doesn't make sense,
  - → it's *not* because you're an idiot
  - → it's because the asker is confused
  - → assume responsibility and try to clarify the confusion

Academic presentations

Finally, a quote I like:

> *"It takes courage to stand up and speak. It also takes courage to sit down and listen."*

You can learn a lot about your work in the feedback you get from those you present it to

Academic careers

Types of academic jobs:

→ Tenure-track research focus

→ Non-tenure teaching-track

→ Lecturers

→ Tenure-track teaching focus

Classification of colleges and universities:

→ E.g., R1 versus R2 institutions

→ Carnegie classification: link

Academic careers: Tenure-track research focus

Job description:

→ 40-50% research

→ 40-50% teaching (0-3 classes per year)

→ 5-10% service

Formal expectations:

→ Publish many papers in mid- and top-tier journals

→ Advise PhD students

→ Write *competitive* grant proposals

- → pay for summer salary
- → buyout of teaching
- → hire an RA or a postdoc
- → travel to conferences

Academic careers: Tenure-track research focus

Hard money versus soft money

Statistics departments versus biostatistics departments

Statistics departments in business school versus college of science

Academic careers: Tenure-track research focus

To prepare for a tenure-track research position:

→ Prioritize your dissertation research

→ Avoid non-academic internships

→ Teach for at least 2 semesters

→ Let your advisor know *as soon as possible* that your goal is a tenure-track research position

→ Find the websites of junior faculty to see how competitive they were on the market

→ that's how you know what the bar for entry is

→ If not competitive when you are graduating, then do a postdoc and try again in 2-3 years!

Non-academic careers

Alumni panel during final exam period:

15:30–18:00 ET on Monday, 8 December 2025 via Zoom