

PhD Training in (Bio)statistics: Challenges & Opportunities

Daniela Witten
Departments of Statistics & Biostatistics
University of Washington

Joint Statistical Meetings
July 29, 2018

Outline

- Why revisit the nature of the (bio)statistics PhD?
- Who should get a (bio)statistics PhD?
- What courses should (bio)statistics PhD students take?
- What dissertation research should (bio)statistics PhD students conduct?

This Talk

*Lots of questions;
no answers!*

Outline

- Why revisit the nature of the (bio)statistics PhD?
- Who should get a (bio)statistics PhD?
- What courses should (bio)statistics PhD students take?
- What should dissertation research look like?

Job Market Has Changed

- Most (bio)statistics PhDs do not get jobs proving theorems.
- *In industry:* demand for “data science” jobs, which merge statistical thinking with computing skills & common sense.
- *In academia:* demand for faculty who are able to collaborate on the analysis of large-scale and complex data sets, for which no hypothesis tests are available in Lehmann & Romano.
- Should PhD training reflect this new reality?

Job Market Has Changed

- Most (bio)statistics PhDs do not get jobs proving theorems.
- *In industry:* demand for “data science” jobs, which merge statistical thinking with computing skills & common sense.
- *In academia:* demand for faculty who are able to collaborate on the analysis of large-scale and complex data sets, for which no hypothesis tests are available in Lehmann & Romano.
- **Should PhD training reflect this new reality?**

Should PhD Training Change?

- Not a rhetorical question.
- We can be purists if we want to.
- But we will get left behind by more enterprising disciplines (computer science, “data science”).
- So transforming our PhD training is not an *imperative*, but it is an *opportunity*.

Outline

- Why revisit the nature of the (bio)statistics PhD?
- Who should get a (bio)statistics PhD?
- What courses should (bio)statistics PhD students take?
- What should dissertation research look like?

PhD Admissions Pre-Reqs

- Hard pre-reqs: multivariable calculus, real analysis (preferably with measure theory), linear algebra, introductory probability, and introductory statistics.
- Strong preference for students who have taken much more math/stat than that.
- So we are implicitly (or explicitly!) selecting for math majors, or at least math minors.
- Historically, a preference NOT to admit Stat majors to PhD programs, due to a lack of mathematical rigor in undergrad Stat programs (!!!)



Are These Pre-Reqs Appropriate?

Given the current design of (bio)statistics PhD programs,

Yes!

You do need to be a math major (or equivalent) to be successful in the coursework.

A Side Note: The GRE Exam

#GRExit

SHARE



12K



14



GREs don't predict grad school success. What does?

By [Beryl Lieff Benderly](#) | Jun. 7, 2017 , 8:30 AM

Summer is just beginning, but before we know it, the graduate school application and admission season will be upon us again. The ostensible goal of that arduous and anxiety-fraught procedure—and of the even more involved process of hiring and promoting faculty members at research universities—is to identify the next generation of productive scientists. But how can the relevant committees accomplish this when no one can really specify the qualities of mind, heart, character, spirit, and background that combine to produce great research?

This long-standing question has been bothering me since I read a pair of studies about grad school admissions practices, [one](#) from the University of North Carolina (UNC) in Chapel Hill and the [other](#) from Vanderbilt University in Nashville. (My colleague Maggie Kuo [wrote about them](#) when they were published in January.) These papers add to a growing body of information suggesting that widely used “objective” admissions measures, such as GRE test scores and GPA, are exactly the wrong way

Last month's Taken for Granted

Taken for Granted is a monthly column about training and career issues from scientific workforce expert Beryl Lieff Benderly.

[The downs and ups of corporate research](#)



[Read more Taken for Granted](#)

SHARE



12K



14



GREs don't predict grad school success. What does?

By **Beryl Lieff Benderly** | Jun. 7, 2017 , 8:30 AM

Summer is just beginning, but before we know it, the graduate school application and admission season will be up

of the even more to identify the ne accomplish this v background that

This long-standin bothering me sinc about grad schoo practices, **one** fro Carolina (UNC) in the **other** from Va Nashville. (My co **about them** when January.) These p of information su “objective” admis test scores and G

SHARE



14K



9



Student performance measures that don't perform

By **Maggie Kuo** | Jan. 11, 2017 , 5:00 PM

GRE scores and undergraduate GPA don't predict students' future graduate school productivity, but reference letters from previous research advisers may provide clues about whether they are going to publish well, according to two papers published today in *PLOS ONE*. These results add to the ongoing discussion about how graduate admissions decisions should be made, particularly in light of previous findings that the **GRE is biased against students from underrepresented groups**. The new studies emphasize that admissions committees should review applicants holistically and rely less on GRE scores in making decisions—a point that many acknowledge, but which requires significant time and energy to do well.

By tracking the research output of students who recently completed biological and biomedical doctorates at the University of North Carolina (UNC) in Chapel Hill and Vanderbilt University in Nashville, the researchers found that neither GRE scores nor GPA were associated with the number of first-author papers a student published. Moreover, these metrics did not predict other quantitative measures of productivity—the number of conference presentations and independent grants and fellowships received—or the student's progress through the program. **further analysis of the Vanderbilt cohort revealed**

SHARE



12K



14



GREs don't predict grad school success. What does?

By [Beryl Lief Benderly](#) | Jun. 7, 2017 , 8:30 AM

Summer is just beginning, but before we know it, the graduate school application and admission season will be up of the even more to identify the ne accomplish this v background that

This long-standin bothering me sinc about grad schoo practices, **one** fro Carolina (UNC) in

SHARE



14K



9

Student performance measures that don't perform

By [Maggie Kuo](#) | Jan. 11, 2017 , 5:00 PM

GRE scores and undergraduate GPA don't predict students' future graduate school productivity, but reference letters from previous research advisers may provide clues about whether they are going t

The Atlantic

Popular

Latest

Sections ▾

M

EDUCATION

The Problem With the GRE

The exam “is a proxy for asking ‘Are you rich?’ ‘Are you white?’ ‘Are you male?’”

VICTORIA CLAYTON | MAR 1, 2016

se results add to the ongoing particularly in light of previous **groups**. The new studies cally and rely less on GRE requires significant time an

ogical and biomedical anderbilt University in Nash with the number of first-au r quantitative measures of grants and fellowships rece **Vanderbilt cohort revealed**

Outline

- Why revisit the nature of the (bio)statistics PhD?
- Who should get a (bio)statistics PhD?
- What courses should (bio)statistics PhD students take?
- What should dissertation research look like?

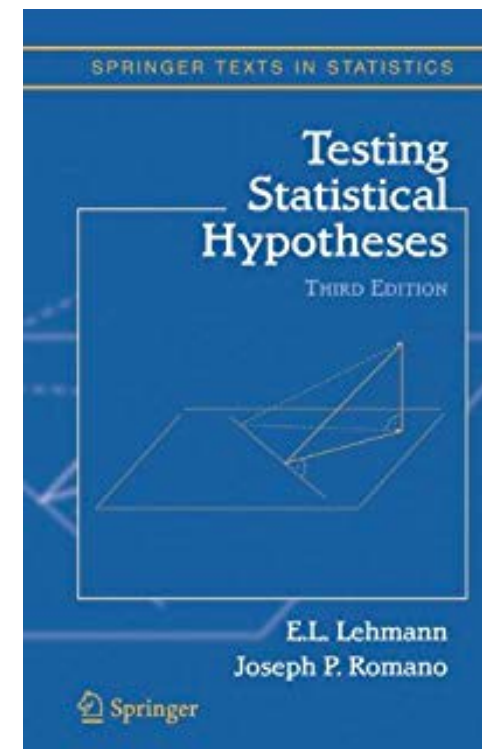
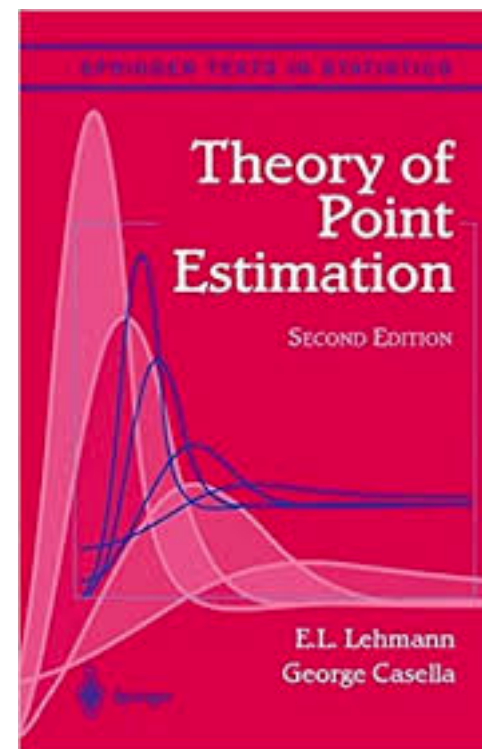
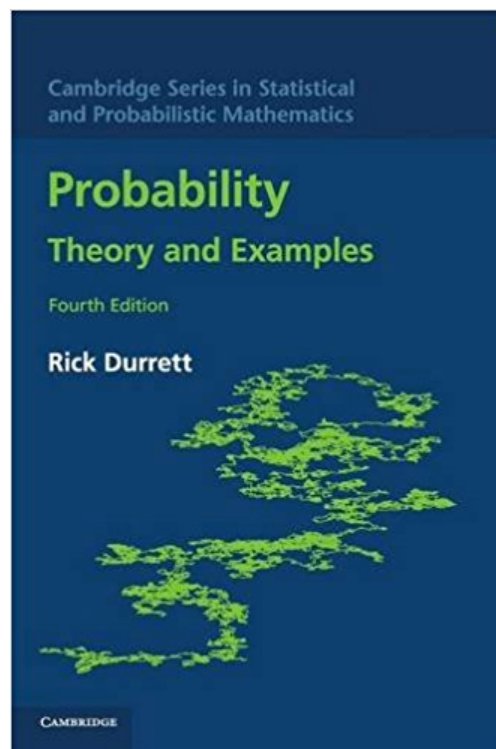
Core PhD Curriculum

- 1 year of measure-theoretic probability
- 1 year of theoretical statistics (hypothesis testing, point estimation, etc)
- 1 year of applied statistics
- Consulting
- A weed-out qualifying exam after 12-24 months
- Electives to “learn the things you need to know”

Core PhD Curriculum

- **1 year of measure-theoretic probability**
- **1 year of theoretical statistics (hypothesis testing, point estimation, etc)**
- 1 year of applied statistics
- Consulting
- A weed-out qualifying exam after 12-24 months
- Electives to “learn the things you need to know”

Probability & Theoretical Statistics



Probability & Theoretical Statistics

Does a PhD statistician working in tech really need a background in measure-theoretic probability?

Would a more applied treatment of hypothesis testing and point estimation be sufficient?

Core PhD Curriculum

- 1 year of measure-theoretic probability
- 1 year of theoretical statistics (hypothesis testing, point estimation, etc)
- **1 year of applied statistics**
- Consulting
- A weed-out qualifying exam after 12-24 months
- Electives to “learn the things you need to know”

Core PhD Curriculum

It is unrealistic to think that the skills needed to be a practicing applied/collaborative statistician will be learned in 1 year of an applied statistics class,

especially when the people teaching that class are teaching not the skills needed today, but rather, the material that they, themselves, learned 30 years ago.

Core PhD Curriculum

- 1 year of measure-theoretic probability
- 1 year of theoretical statistics (hypothesis testing, point estimation, etc)
- 1 year of applied statistics
- **Consulting**
- A weed-out qualifying exam after 12-24 months
- Electives to “learn the things you need to know”

Consulting

- Often 1 semester is required, during a 5-year PhD.
- This is the one chance that students have to perform statistics “in the wild”, or to develop “quantitative common sense”.
- Also an opportunity to develop the “soft skills” of statistics.
- Is one semester enough?

Core PhD Curriculum

- 1 year of measure-theoretic probability
- 1 year of theoretical statistics (hypothesis testing, point estimation, etc)
- 1 year of applied statistics
- Consulting
- **A weed-out qualifying exam after 12-24 months**
- Electives to “learn the things you need to know”

Qualifying Exam

- Quality-control is important
- Opportunity to consolidate knowledge
- But it often serves to weed out (and intimidate) members of underrepresented groups or students from non-traditional backgrounds
- How well does performance on a qualifying exam predict research or career success?

Core PhD Curriculum

- 1 year of measure-theoretic probability
- 1 year of theoretical statistics (hypothesis testing, point estimation, etc)
- 1 year of applied statistics
- Consulting
- A weed-out qualifying exam after 12-24 months
- **Electives to “learn the things you need to know”**

Electives

- Align with faculty's (academic) research careers
- Often (always?) not what our students will need to know over the course of their careers
- Different students will need different things, based on diversity of career paths
- An elective in “Higher-Order Asymptotics”: I’m not saying most students don’t need it, but I’m also not saying most students don’t not need it

Outline

- Why revisit the nature of the (bio)statistics PhD?
- Who should get a (bio)statistics PhD?
- What courses should (bio)statistics PhD students take?
- What should dissertation research look like?

Dissertation Research

- Designed to promote faculty careers: papers in journals, and preliminary results for grants.
- Are the skills learned (or even the CV line items) relevant for a career outside of academic research?

Dissertation Research, Re-Envisioned?

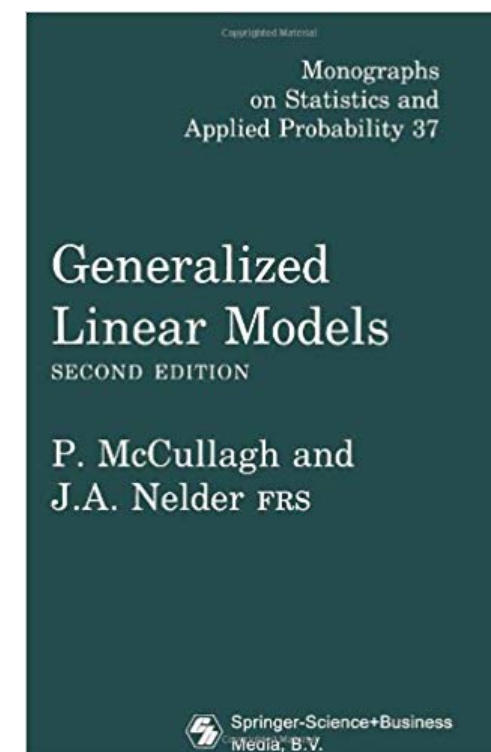
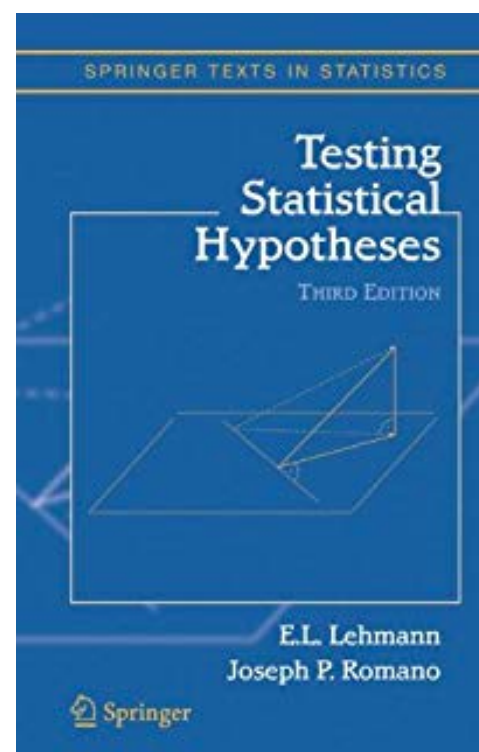
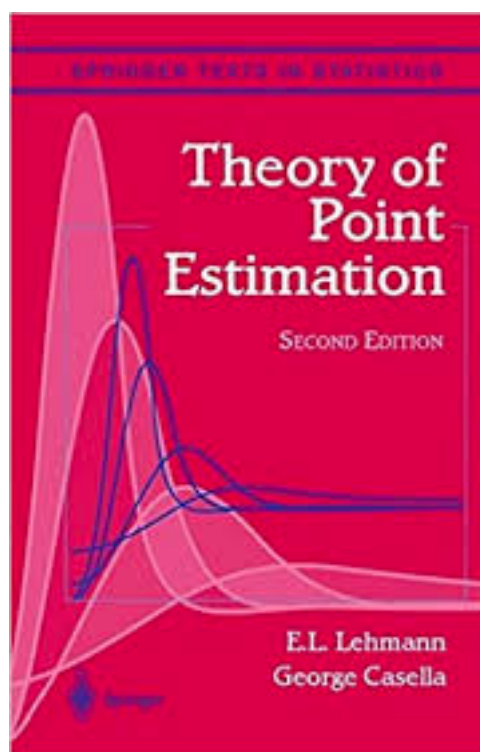
- Projects motivated by non-academic problems?
- Who will fund this? (Not my federal research grants!) Non-academic (industry) partners?
- What's in it for me to advise a PhD student whose research will not further my career?

What Are Our Options?

- If we continue as is, then computer science, data science, and related departments will continue to fill the void.
- To step into the void ourselves, here are some possibilities:
 - Parallel PhD tracks in theoretical statistics and applied statistics? Want to avoid a two-tier system.
 - Encourage (rather than tolerate) students who do summer internships, and bring this to bear in formulating their dissertation research.
 - Academic incentives to reward faculty who solve real-world problems.
 - New funding sources for PhD students outside of federal agencies, which will require infrastructure for industry collaboration.

What is Statistics?

- Ask a computer scientist what makes someone a computer scientist, and you'll get an answer encompassing everything from security to robotics to theory of computation to natural language processing.
- Ask a statistician, and they might answer with:



Conclusion

- The world around us has changed, but academic statistics hasn't.
- We can keep things the way they have always been, and risk getting left (further) behind.
- But making changes isn't so easy: which of our core values are we willing to compromise on?
- No easy answers.

Thank You!