Principles of the PhD Qualifying Exam Starting from Spring 2025

The aim of the Ph.D. Qualifying Exam is to assess the candidate's mastery of core statistical concepts, analytical techniques, and applied methodologies necessary for advanced research. The exam is designed to evaluate both theoretical understanding and practical problem-solving skills, ensuring that students are adequately prepared to undertake independent, original research at the doctoral level. Successful completion of the exam demonstrates the student's readiness to progress toward dissertation work and contribute meaningfully to the field of statistics.

The Ph.D. Qualifying Exam in the Department of Statistics consists of **three** parts: Part I, Part II, and Part III.

- Part I and Part II are written examinations.
- Part III is an oral examination.

Only students who pass both Part I and Part II are eligible to proceed to the Oral Examination (Part III). A candidate is considered successful upon passing all three parts of the exam.

The Ph.D. Qualifying Exam is administered twice a year, in November and May. Students wishing to take the exam must submit a formal letter of application to the department by the end of October for the November exam, and by the end of April for the May exam.

To qualify for the oral exam, students must achieve a score of 70 or above in the written exam (combined Parts I and II). In order to pass the overall qualifying exam, candidates must obtain at least 70 points in each part—written and oral—individually.

Students who do not pass the written exam but are eligible for a second attempt may retake only the parts in which they scored below 70. Failure to pass on the second attempt will result in dismissal from the Ph.D. program.

Part I: Advanced Probability Theory:

Aim/Objective of the Exam: This written examination intends to have students intensify their ability to follow the ideas given in Probability Theory.

Guidelines of the Exam: You may be asked to prove the specific results listed below as 'Theorem,' 'Lemma' or 'Law' or you may be asked to state results and summarize the proofs or you may be asked to solve or prove smaller exercises, which you may not have seen before.

Emphasized Topics:

- Probability measure, σ-algebras, measurability, random variables, distribution, expected value, Modes of convergence (almost sure (almost everywhere), in probability (stochastic convergence), in p L between them)
- Law of Large Numbers, in distribution (weak convergence) and the relationship
- λ-π systems, Independence and dependence of random variables and of σ-algebras, Monotone Class Theorem, and related topics
- Kolmogorov zero-one Law
- Characteristic Functions and the inversion formula, Infinitely divisible distributions, compound Poisson, general form of infinitely divisible
- Tightness and the method for showing convergence in distribution, Continuity Theorem
- Central Limit Theorem
- Conditional Expectations: Definition and Properties
- Martingales, filtrations, stopping times, Martingale Convergence Theorem, The Up-crossing Lemma
- Uniform integrability, theorems combining uniform integrability and conditional expectation, theorems combining martingales, uniform integrability and almost sure convergence
- Probability Concepts in Stat 501 and Stat 502

Related Courses: Stat 601, Stat 602, Probability Concepts in Stat 501 and Stat 502.

Suggested References:

- Richard Durrett (2017) Probability: Theory and Examples, Duxbury Press.
- Patrick Billingsley (2012) Probability and Measure, Wiley series in probability and mathematical statistics.
- Erhan Çınlar (2011) Probability and Stochastics, Springer
- Hayri Körezlioğlu, Azize Bastıyalı Hayfavi and Yeliz Yolcu Okur (2018) Elements of Probability Theory, METU Press.
- George Casella and Roger L. Berger, Statistical Inference, Duxbury

Part 1. Advanced Theory of Statistics Exam:

Aim/Objective of the Exam: This written examination intends to have students intensify their ability to follow the ideas given in the Theory of Statistics.

Guidelines of the Exam: The aim is to evaluate student's theoretical knowledge needed to develop a methodology for estimation and assessing the statistical properties of estimators. This part includes advanced theorems and methodologies related to the following topics:

- Techniques of estimation: moments method; likelihood-based estimation • Likelihood construction
- Conditional likelihood
- Information matrix calculations
- Statistical properties of estimators
- Hypothesis testing: Score tests, Wald tests, likelihood ratio tests
- Construction of confidence intervals
- Statistical properties of hypothesis tests: MP tests, UMP tests
- Statistical Inference Concepts in Stat 501 and Stat 502

Related courses: STAT 501, STAT 502, STAT 603, STAT 604

Suggested references:

- Boos, D.D. and Stefanski, L.A. (2013) Essential Statistical Inference: Theory and Methods
- Wasserman, L. (2003) All of Statistics: A Concise Course in Statistical Inference
- Bain, L.J. and Engeldhart, M. (1994), Introduction to Probability and Mathematical Statistics"
- Casella, G. and Berger, R.L. Statistical Inference, Duxburry
- Mood, A.M., Graybill, F.A., Boes, D.C. (1974), "Introduction to the theory of statistics", McGraw-Hill Publishing USA
- Sahoo, P. (2013), Probability and Mathematical Statistics
- Hogg, R.V., McKean, J., and Craig, A.T, Introduction to Mathematical Statistics

Part 1. Advanced Generalized Linear Models

Aim/Objective of the Exam: This written examination intends to have students intensify their ability to follow the ideas given in Advanced Generalized Linear Models.

Guidelines of the Exam: The aim of this part is to assess student's theoretical, applied, and computational knowledge of statistical models. **Topics include**

- Link functions, canonical links
- Likelihood and quasi-likelihood functions
- GLM for Gaussian, Binary, Polytomous, Poisson and Negative Binomial, and Continuous non-Gaussian responses: Parameter estimation, iterative approaches, interpretation, inferences, diagnostics. Detecting and fixing models of less than full rank. Extrapolation, Model selection, Model validation.
- Overdispersion
- Zero-truncated and zero-inflated models for count data
- Loglinear Models - Parameter estimation, interpretation, inferences. Connection to logit model. Models for matched data: conditional logistic regression.

Related courses: STAT 636

Suggested references:

- P. Dunn, G. Smyth. Generalized Linear Models With Examples in R, Springer, 2018.
- Hardin, J. W., & Hilbe, J. M. (2007). Generalized linear models and extensions. Stata press.
- P. McCullagh, J.A. Nelder. Generalized linear models, London ; New York : Chapman and Hall, 1989.
- C.R. Rao, Shalabh, H. Toutenburg, C. Heumann. Linear Models and Generalizations, 3rd edition. Springer New York, 2008.
- J. Neter, M.H. Kutner, C.J. Nachtsheim, W. Wasserman, Applied Linear Statistical Models, 4th edition, Irwin, 1996.
- D. C. Montgomery, E. A. Peck and G.G. Vining. Introduction to Linear Regression Analysis, 5th edition. Wiley & Sons, NY, 2012.

Part 1. Advanced Computational Statistics:

Aim/Objective of the Examination: This written examination intends to have students intensify their ability to follow the ideas given in the Advanced Computational Statistics.
Guidelines of the Exam: The aim of this part is to assess students' theoretical understanding and practical skills in data analytics, computational statistics, and their applications in predictive modelling, clustering, and text mining. The exam will be conducted in front of a computer.
Topics include

- Data Exploration and Preprocessing
    - Challenges of multidimensional data.
    - Data cleansing, transformation, and partitioning.
    - EDA techniques and projection methods.
- Data Mining Foundations
    - Objectives, task identification, and feature engineering.
    - Association rule mining and evaluation metrics.
- Statistical and Computational Techniques
    - Random number generation and nonparametric density estimation.
    - Nonparametric methods (e.g., GAMs, MARS).
    - Monte Carlo and bootstrapping methods.
    - Advanced MCMC techniques and diagnostics.
- Predictive Analytics and Classification
    - Supervised learning: decision trees, ensembles (e.g., boosting, random forests), SVMs, and neural networks.
    - Pattern recognition, and evaluation metrics.
- Clustering and Unsupervised Learning
    - Methods such as k-means, hierarchical clustering, Gaussian Mixture Models, DBSCAN.
    - Neural network-based and model-based clustering.
    - Evaluation metrics (e.g., silhouette score).
- Time Series and Sequential Data_Forecasting with ARIMA and deep learning (e.g., RNNs, LSTMs, GRUs).
- Text Mining and NLP
    - Preprocessing: tokenization, stemming, vectorization.
    - Models: RNNs, transformers, topic modelling.
    - Applications in sentiment analysis and classification.
- Deep Learning Architectures_CNNs for images, transformers for sequences, and emerging trends.

Related courses: STAT 635 and STAT 571
Suggested references:

- Martinez, W. L., & Martinez, A. R. (2001). Computational statistics handbook with MATLAB. Chapman and Hall/CRC.
- Gentle, J. E. (2009) Computational statistics (Vol. 308). New York: Springer.

- Givens G.H. and Hoeting J.A. (2013) Computational Statistics. New York: Wiley.
- Albert J and Rizzo. M (2012) R by Example. New York: Springer.
- Hastie T, Tibshirani R and Friedman J (2001), The Elements of Statistical Learning, Springer Series in Statistics, 2009. https://link.springer.com/book/10.1007/978-0-387-84858-7
- Han, J., Kamber, M, and Pei, J. (2012) Data Mining: Concepts and Techniques, 3rd Edition. San Francisco, CA: Morgan Kaufmann. ISBN-10: 0123814790
- Goodfellow I, Bengio y and Courville A, (2016) Deep Learning, The MIT Press. https://www.deeplearningbook.org/
- Gerrish S, Scott K (2018) How Smart Machines Think, The MIT Press
- Frankish K. (2014) The Cambridge Handbook of Artificial Intelligence
- Chollet F. (2021) Deep Learning with Python, Second Edition, Manning Pub.
- Prince S.J.D. (2023) Understanding Deep Learning, The MIT Press.

Part 2. Advanced Data Analysis

Aim/Objective of the Examination: To evaluate the student's understanding and application of advanced data analysis techniques, including statistical modelling, inference, and data-driven decision making, in preparation for independent research at the doctoral level.

Guidelines of the Exam: The Advanced Data Analysis exam will be conducted in front of a computer. Students will be provided with a dataset and are expected to perform a comprehensive analysis using appropriate statistical methods. The exam will assess both technical proficiency and the ability to communicate results clearly.

Students must submit a written report that includes:

A clear statement of the problem
Description of the methods used
Key results with appropriate visualizations and interpretations
Conclusions and recommendations

The report should be understandable to both statisticians and non-statisticians, demonstrating the ability to communicate complex ideas in a clear and accessible manner.

Use of statistical software (e.g., R, Python) is permitted and encouraged. All code and output should be included as an appendix.

Students are allowed to bring one of the following books:

- Gareth James. An Introduction to Statistical Learning with R
- Albert, J., & Rizzo, M. (2012). R by Example. Springer.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning, Springer Series in Statistics. https://link.springer.com/book/10.1007/978-0-387-84858-7
- Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques, 3rd Edition. Morgan Kaufmann. ISBN-10: 0123814790
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. https://www.deeplearningbook.org/
- Gerrish, S., & Scott, K. (2018). How Smart Machines Think. MIT Press
- Frankish, K. (2014). The Cambridge Handbook of Artificial Intelligence

- Chollet, F. (2021). Deep Learning with Python, Second Edition. Manning Publications
- Prince, S.J.D. (2023). Understanding Deep Learning. MIT Press

Part III: Oral Examination

Aim/Objective of the Oral Examination: The students will be examined on the Fundamental ideas of Theory and Application of Statistics and they will be encouraged to talk about their future research plan's and the obstacles they may encounter during the process. The student's ability and potential to conduct a doctorate level research will be measured.

Grading Policies of the Examination:

1. **Advanced Probability Theory, Advanced Theory of Statistics, Advanced Generalized Linear Models & Advanced Computational Statistics:** Candidates will be required to answer two questions in each exam. All exams are closed-book; however, if a question requires reference to textbook material, candidates will be informed in advance and allowed to bring one textbook. Each exam carries equal weight. Exams will be graded out of 100, and 70% of the overall score for Part I will be based on the results of these exams.
2. **Advanced Data Analysis.** In Part II of the exam, candidates will be provided with a dataset and are expected to apply appropriate analytical tools to thoroughly examine and interpret the data. This component will be graded out of 100, and it will constitute 30% of the total score for the Written Exam.

To be considered successful in the Written Exam (Parts I + II), candidates must achieve a total combined score of at least 70 out of 100.

If a candidate scores 70 or above in any individual part—either in Part I (any subsection) or in Part II—they will be exempt from that part in the second exam attempt. In subsequent sittings, the candidate will only need to retake and pass the part in which they did not meet the required threshold, with a minimum score of 70 out of 100.