

统计学院

2019 级攻读应用统计（大数据分析方向）专业学位研究生培养方案

一、适用专业学位

应用统计（大数据分析方向）

二、培养目标与培养方式

掌握马克思主义基本原理和中国特色社会主义理论体系，具有良好的政治素质和职业道德；系统掌握大数据采集、整理、分析及结果呈现的统计理论；具备熟练应用计算机集群进行大数据处理、统计分析的能力；能够独立完成对实际问题的统计分析并撰写规范的统计分析报告；掌握一门外国语。

三、学科专业研究方向

应用统计（大数据分析方向）

四、学习年限

基本学习年限 2 年。

五、课程设置和学分要求（见附表）

攻读硕士学位研究生期间，需要获得学位课程总学分不少于 36 学分。必修课不少于 18 学分，社会实践不少于 4 学分，选修课不少于 8 学分，公共课不少于 6 学分，具体课程及学分要求见“方案课程及学分要求”。

六、社会实践

参与和完成一项社会实际统计调查和数据分析的实践工作和实践报告。

七、论文撰写

硕士研究生修满学分并经考核合格后，进入学位论文写作阶段。学位论文在导师指导下，由硕士研究生本人按计划进度独立完成。学位论文应与实际问题、实际数据和实际案例紧密结合，可采用与数据收集、整理、分析相关的调研报告，数据分析报告，应用统计方法的实证研究等形式。硕士研究生写出硕士学位论文及其摘要，经指导教师推荐，研究生院审核批准，可进入硕士学位论文评阅和答辩阶段。

附：课程设置和学生课程学习的学分要求

1、公共课（6 学分）

（1）政治理论课

中国特色社会主义理论与实践研究	2 学分	1 学期
(The Theories and Practice of Socialism with Chinese Characteristic)		

马克思主义与社会科学方法论 1 学分 1 学期
(Marxism and method social sciences)

(2) 第一外国语

外语 3 学分 2 学期
(Foreign Language)

2、必修课 (18 学分)

大数据分析统计基础 3 学分 1 学期
(Basic Statistical Methods for Big Data Analysis)

(本课程主要讲解大数据分析的统计基础模型, 包括描述统计、估计与检验回归分析等。先修课: 数理统计)

非结构化大数据分析 3 学分 2 学期
(Statistical Methods for Non-Structure Big Data)

(本课程主要讲解处理非结构化大数据的方法, 包括文本挖掘、社交网络分析、数据流等。先修课: 大数据分析统计基础)

大数据分析统计建模 3 学分 2 学期
(Statistical Modeling for Big Data Analysis)

(本课程主要讲解大数据分析的统计模型, 包括多元统计、时间序列、空间统计等。先修课: 大数据分析统计基础本课程主要讲解大数据分析的统计模型, 包括多元统计、时间序列、空间统计等。先修课: 大数据分析统计基础)

大数据分析计算机基础 3 学分 1 学期
(Computing Skills for Big Data Analysis)

(本课程主要介绍 Linux 操作系统以及 Shell 编程命令, 结构化、非结构化数据库, 大数据清理、呈现等处理大数据的计算机基础知识。先修课: 计算机基础)

大数据分布式计算 3 学分 1 学期
(Distributed Computing for Big Data Analysis)

(本课程主要介绍处理大数据的 Hadoop 分布式平台, MapReduce 编程思想, Storm 实时计算平台。先修课: 计算机基础)

大数据挖掘与机器学习 3 学分 1 学期
(Data Mining and Machine Learning for Big Data Analysis)

(本课程主要讲解大数据分析的统计机器学习算法, 包括现代分类、回归方法, 聚类方法、高维数据处理方法等。先修课: 大数据分析统计基础)

学术规范和论文写作 1 学分 2 学期
(Academic Norms and Thesis Writing)

(【必修】讲授学术规范和论文写作规范及方法)

3、选修课 (不少于 8 学分)

习近平新时代中国特色社会主义思想研究 2 学分 2 学期
(Studies on The Xi Jinping Thought on Socialism with Chinese

Characteristics for a New Era)

(本课程主要是帮助学生把握习近平新时代中国特色社会主义思想的历史地位、时代背景、丰富内涵和实践要求,自觉用习近平新时代中国特色社会主义思想武装头脑,增强解决中国问题的能力。)

习近平教育思想研究 2 学分 1 学期

(Studies on Xi Jinping' s Education Thoughts)

(教育兴则国兴,教育强则国强。建设教育强国是中华民族伟大复兴的基础工程。党的十八大以来习近平总书记一系列重要讲话和治国理政新思想新理念新战略蕴含着十分丰富精辟的教育思想。党的十九大开创性提出了习近平新时代中国特色社会主义思想,是新时代我国教育奋进发展的指针和纲领。学习与研究习近平教育思想,要与学习习近平总书记系列重要讲话、学习十九大报告密切结合、融会贯通;与学习马克思主义、毛泽东思想、邓小平理论和教育学一般理论密切结合、融会贯通;与学习我国教育发展历史、当前全球教育发展和我国教育改革现状的现实实践、以及建设教育强国的发展方向密切结合、融会贯通。通过课程学习,帮助学生们全面理解和深刻领会习近平新时代中国特色社会主义思想事业的性质、宗旨、目标、主要矛盾、历史定位与使命、发展动力、发展战略、依法治国、高等教育建设规律、构建中国特色哲学社会科学、立德树人为根本的人才培育、教师职业发展等重要思想,从而更好地理解过去、把握当下、相信未来,提高政治理性、理论理性与实践理性,促进学生们成长成才。)

计量经济方法与应用 3 学分 1 学期

(Econometrics: Method and Application)

(本课程主要讲授计量经济方法的基本原理、方法以及前沿进展,重在通过实际应用来介绍计量经济分析中存在主要问题和解决思路。先修课程:数理统计、回归分析)

市场研究方法与实务 3 学分 2 学期

(Market Research: Method and Practice)

(通过案例分析展现市场研究过程的各个步骤,运用数据分析解决市场营销管理的决策问题)

金融统计案例研究 2 学分 2 学期

(Empirical Analysis for Finance and Banking)

(本课程从金融问题和统计分析方法入手,介绍我国金融体制和金融市场的基本框架,并就具体的金融问题展开问题、数据和实证的讨论,有侧重的讨论和研究中国金融改革过程中的实际问题 and 解决途径。先修课:宏观经济学、微观经济学、金融学)

社交网络数据分析专题 2 学分 2 学期

(Special Topics in Social Network Data Analysis)

(了解社交网络数据的基本特征,掌握社交网络数据的基本分析方法及其在应用统计领域中的应用。课程主要包括让学生深刻理解什么是关系数据,掌握网络结构数据的可视化方法,对网络结构数据的统计建模有初步的了解。先修课程:高等统计、回归分析等课程。)

贝叶斯统计 2 学分 1 学期

(Bayesian Methods)

(本课程将以比较通俗的方式, 结合具体问题和 R 语言实现, 介绍贝叶斯方法的主要概念和建模思想。基本内容包括: 概率的两种定义, 贝叶斯方法和频率派的异同; 先验分布、似然函数和后验分布; 常用的共轭先验分布; 层级贝叶斯模型和超参数; 随机模拟方法入门; 蒙特卡洛积分入门; 马尔可夫链和 MCMC 方法的原理; 用 Stan 实现 MCMC 建模; MCMC 的应用和解释。本课程使用的主要编程工具是 R, 比较简单的问题用 R 编程, 比较复杂的方法用 STAN。主要参考书是北美财险精算协会推荐使用的 McElreath(2016), *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, CRC Press)

商务大数据案例分析 2 学分 1 学期

(Big Data Analysis with Business Applications)

(本课程将通过来自不同数据源的商务大数据分析案例, 讲述实际数据的清理, 描述过程, 以及线性回归, 01 回归, 机器学习等方法的实际应用。目标是使得同学们能够独立完成实际商业数据分析全过程, 并形成最终的案例分析报告。)

统计案例 2 学分 2 学期

(Statistical Case Study)

(案例分析。)

调查组织与设计 2 学分 1 学期

(Survey Management and Design)

(通过教师讲授、师生讨论、学生实地收集数据、数据整理和分析、报告撰写、口头汇报等方式, 帮助学生掌握调查方案的设计和调查项目的组织管理方法。先修课: 数理统计)

应用抽样技术 2 学分 2 学期

(Applied Sampling Techniques)

(主要内容包括有限总体内的概率抽样方法, 复杂样本的方差估计, 非抽样误差以及抽样设计案例分析等。先修课程: 数理统计)

现代精算统计模型 2 学分 1 学期

(Modern Actuarial Models)

(该课程在非寿险费率厘定的背景下介绍了广义线性模型、广义可加模型和几种常用的机器学习算法, 包括神经网络、树模型、集成学习、支持向量机。通过在同一个数据集上使用这些不同的模型和算法, 我们比较了它们的优缺点。该课程不仅强调编写程序的能力, 还要求掌握和理解这些模型背后的数学原理和统计思想。在选这门课之前, 需要完成线性回归模型, 概率论等基础课程。)

可在全校研究生课程范围内选课, 鼓励在本院各专业硕博连读培养方案中所列课程选修。

4、社会实践 (4 学分)