**1.“大数据的统计学基础与分析方法”重大项目指南**

发展大数据技术是国家重大战略需求，也是统计学、数据科学、信息科学和管理科学等学科的国际前沿。在大数据环境下，数据的规模、类型、结构和增长速度发生了质的变化，传统数据分析和处理的统计学理论和分析方法已不能满足大数据时代下的种种需求。本重大项目拟围绕大数据分析与处理的统计学基础理论与分析方法的核心问题和难点问题，建立大数据统计分析的新理论和新方法，发展适应于大数据的统计算法，为大数据技术发展和大数据行业应用提供新的统计学基础理论和方法支撑。

　　一、科学目标

　　面向大数据技术的国际前沿，面向国家发展大数据技术和行业大数据应用的重大需求，针对大数据统计分析的若干重大瓶颈问题，开展大数据分析与处理的统计学基础理论与分析方法的研究，在大数据的统计学基础方面取得突破性或关键性进展，建立起若干新的理论，在大数据统计计算模式与算法方面提出适应异构计算环境下分布式统计计算模式的系列高效算法，形成大数据处理的具有独立自主知识产权的领先核心技术族，在结合领域的示范应用方面取得突破。

　　二、研究内容

　　（一）支持大数据分析与理解的统计学基础。

　　对于大数据而言，现有的许多统计学理论在大数据环境下不再成立，导致了统计建模的理论和方法需要重建，多源异构数据和超高维统计推断方法和理论诸多空白需要填补。主要研究内容包括：超高维数据和异源异构数据的统计推断；超高维统计问题的各类极限定理；大数据背景下罕见事件的稳健统计推断；生物医学大数据中统计建模等。

　　（二）大数据分析与处理的统计计算模式与核心算法。

　　大数据计算模式指的是根据大数据的不同数据和计算特征，从多样化的大数据计算问题和需求中提炼并形式化了的各种高层抽象或模型。大数据的出现为计算模式和基础算法提供了发展的新机遇，也提出了全新的基本科学问题。主要研究内容包括：分布实时统计计算问题；大数据统计算法的可行性理论、收敛性问题和误差可控性问题；多源异构数据融合的统计计算模式和高效算法；流数据的学习理论和方法。

　　（三）结合领域的大数据的科学发现及其方法论依据。

　　结合典型领域，探索科学的规律发现方法和决策方法，并建立科学的验证方法，验证并展示所发展的新理论与新方法的有效性。主要研究内容包括：大数据中将数据特征选择和表示、估计推断、决策分析融为一体的全过程一体化优化模型和方法；以目标任务为导向的规律发现和验证方法，在给定复杂度（能源、通信、时间、空间和规则结构约束下）的规律发现、决策形成及验证方法。

　　三、申请注意事项

　　（一）申请书的附注说明选择“大数据的统计学基础与分析方法”（以上选择不准确或未选择的项目申请不予受理）。

　　（二）申请人申请的直接费用预算不得超过1500万元/项（含1500万元/项）。

　　（三）本项目由数理科学部受理申请。