

# 2021 年全国大学生数据统计与分析竞赛

---

## 题目 A：硕士学位论文评价数据的统计与分析

从翟天临学位论文涉嫌学术不端被查，到教育部和科技部推出的“破四唯”，让学位论文质量成为了研究生学位授予质量的重要评价参考。从培养单位的角度看，学位论文质量的高低是衡量研究生学术水平的重要手段。从研究生个人层面看，学位论文的写作是一项“神圣”的使命。学位的获得应该既代表了科学知识的掌握，也代表了科学精神的具备。在这一意义上，研究生学位论文的写作过程不仅是提高研究、创新能力的过程，而且是一个铸造人格的过程。

根据《国务院学位委员会、教育部关于印发〈博士、硕士学位论文抽检办法〉的通知》（学位〔2014〕5 号）要求，中国北方某省教育厅 2019 年和 2020 年完成了全省每年 1.5 万余篇硕士学位论文的例行抽检工作，得到了大量的评审数据。该省教育厅采用了平均分和最低得分的方式对论文质量进行评价，提取 3 位评阅专家给出的总分最低分进行学科门类内排名，用“末位后 5%淘汰制”定出问题论文，用 3 位评阅专家给出的总分平均分反应论文的整体水平。为了更好的对学位论文进行科学评价，请参赛者根据部分抽检论文的评审信息（见附件 1），运用数据统计与分析方法，解决如下 5 个问题，并将问题解决的过程按照全国大学生数据统计与分析竞赛的要求撰写成论文，填写附件 2，一并提交。

**问题 1:** 请参赛者基于上述“末位后 5%淘汰制”，筛选出问题论文。填写附件 2。

**问题 2:** 计算每篇论文的各分项平均分和总分平均分。填写附件 2。按照学科门类，分别统计分析 3 位评阅专家给出的 3 个总分以及 1 个总分平均分之间的一致性和差异性，并分析对比各学科门类学位论文的水平。

**问题 3:** 按照学科门类统计分析论文评阅评语，采用文本分析方法，建立评阅专家观点评价模型，并论证文本分析结果与其各分项得分间是否存在一致性。

**问题 4:** 应用问题 2 和问题 3 得到的结论对附件 1 中的每篇论文进行综合评价，给出综合得分。填写附件 2。

**问题 5:** 以学科门类工学（08）为例，对比问题论文（被淘汰的论文）与优秀论文（综合得分排名前 10%）的典型特征。

#### 附录：

附件 1 提供了 1246 篇硕士学位论文的评审信息，涵盖 11 个学科门类。表中的 Tag 指的是学科门类编码（见后面释义），表中 bm 指的是论文编号，每个编号对应 1 篇硕士学位论文，每篇论文都有 3 位评阅专家评价信息。

评阅专家从选题与综述、创新性、论文价值、科研能力与基础知识、论文规范性，4 个维度对论文进行评价打分。表中的  $X_{k1}$  指的是第  $k$  位评审专家对论文选题与综述的评分； $X_{k2}$  指的是第  $k$  位评审专家对创新性、论文价值的评分； $X_{k3}$  指的是第  $k$  位评审专家对科研能力与基础知识的评分； $X_{k4}$  指的是第  $k$  位评审专家对论文规范性的评分； $X_k$  指的是第  $k$  位评审专家对论文的总体评价得分； $R_k$  指的是第  $k$  位评审专家对论文的评语， $k=1, 2, 3$ 。

表 1：学科门类编码释义表

Tag	释义	Tag	释义
01	哲学	08	工学
02	经济学	09	农学
03	法学	10	医学
04	教育学	12	管理学
05	文学	13	艺术学
07	理学		

表 2：评阅打分样表

评价指标	评价要素	各项得分数
选题与综述 (25 分)	研究的理论意义、实用； 对本学科及相关学科领域国内外发展状况和学术动态的了解程度	
创新性 及 论文 价值 (25 分)	论文提出的新见解、新方法所具有的价值； 论文成果对科技进步、经济建设、国家安全等方面产生的影响或作用	
科研能力与基 础知识 (25 分)	论文体现的理论基础的扎实程度；本学科及 相关学科领域专门知识的系统性； 分析问题、解决问题的能力； 研究方法的科学性，是否采用先进技术、设备、信息等进行论文研究工作。	
论文规范性 (25 分)	引文的规范性，学风的严谨性； 论文语言表达的准确性、逻辑的严密性、书写格式及图表的规范性	
专家评语（主要提出论文不足之处）：		总分：