

全过程非标准化考核改革 校级优秀案例申报材料

| | |
|------------|--------------------------|
| 课程名称及代码: | [63002CC0Q2]商务与经济统计 |
| 所属学院/专业团队: | 信息与商务管理学院 电子商务与供应链管理系 |
| 课程负责人: | 梁辰 |
| 主要参与教师: | 潘常虹、王文娣、连兰兰、霍飞飞 |
| 面向年级/专业: | 2024 级（大二年级）/供应链管理 |
| 课程学分/学时: | 4 学分/64 学时 |

| | | | |
|----------------|--|------------|---------|
| 改革亮点 (推荐理由) | 构建双线耦合、四阶递进、数智协同、量规导向的考核改革体系。内容上，以真实项目替代习题，落实高阶应用考核；评价上，AI 阶梯式赋能“学-评-创”一体化，实现动态评测。有效驱动学生从硬背公式向商业决策深度转型，达成能力培养与持续改进的主动闭环。 | | |
| 课程自评 得分 | 98 分 | 院级评审 得分 | 93.75 分 |

日期: 2025 年 12 月 31 日

一、改革背景与问题导向

(一) 课程原有考核痛点

本课程作为供应链管理专业的专业核心课及连接数理基础的应用桥梁，是构建数智化供应链管理能力的关键基石。课程旨在达成“问题定义—工具探究—系统决策”的核心能力目标，即培养学生在复杂商务场景中运用统计模型，合理使用数字化工具进行深度分析与实验性推演，并基于批判性思维实现科学决策的综合应用能力。但是，原有考核模式严重制约了目标达成，存在三大痛点。

1. 命题机械，高阶评价缺失

终结性考核题目高度模式化，试卷中存在定义辨析、套用公式计算等机械题目，学生无需理解统计原理在业务中的决策价值，凭死记硬背和“刷题”即可高分通过，严重缺乏对复杂商业问题分析与决策能力的考察。

3. 某城市酒店房间均值为 273 元（2024 年 11 月数据），假设此项估计是由 25 家酒店构成的样本得到的，样本标准差为 65 元。计算结果可保留 3 位小数。
- (1) 在 95% 置信度下，边际误差是多少？(4 分)
(2) 总体均值的置信度 95% 的区间估计为多少？(4 分)
(3) 三年前该市酒店房价均值为 229 元，分析酒店房价的变化。(2 分)

t 分布表

| 自由度 | 右尾面积 | | | | | |
|-----|-------|-------|-------|-------|-------|-------|
| | 0.20 | 0.10 | 0.05 | 0.025 | 0.01 | 0.005 |
| 24 | 0.857 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 |
| 25 | 0.856 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 |

4. 在一条生产线上，为每件产品平均装入 160g 食品，装得过多或过少都会产生严重的后果，一旦发生这种情况就要求操作人员及时关闭生产线并对机器进行重新调试，根据以往数据，假设总体标准差为 5g，质检员每小时检验 36 件产品，确定该生产线是否需要停产进行调整。取显著性水平 $\alpha=0.05$ 。 $(z_{0.005}=2.576, z_{0.025}=1.960, z_{0.05}=1.645)$
- (1) 对此质量监控问题，提出合理的假设。(2 分)
(2) 如果样本均值为 162.3g，检验的结论是什么？(4 分)
(3) 如果样本均值为 158.5g，检验的结论是什么？(4 分)

图 1 改革前试卷典型题目：脱离真实复杂情境的单一参数求解（2024-2025-1 A 卷）

2. 过程虚化，反馈机制滞后

考核呈现“重期末、轻过程”特征，平时项目评价缺乏即时反馈，学生无法通过“做中学”修正，能力成长断层。部分学生期末查分后才反馈对形成性成绩不满意。

3. 标准单一，产出同质严重

由于缺乏真实场景与个性化探究要求，项目千篇一律。例如在原项目 4 线性回归应用中，提供给学生的是统一数据集，指导书步骤一致，选用因变量、自变量要求相同，导致分析结果与解释高度雷同。

甚至出现不同变量分析中机械复制“价格每增加一美元”等低级逻辑错误，反映出学生处于“为了做作业而做作业”的被动状态，学习主动性低，难以激发其基于真实场景的主动探索与差异化创新意识。

The screenshot shows a Microsoft Word document titled "项目4-23002430124-罗欣萍.docx". The document contains a table of descriptive statistics for a car, followed by an ANOVA table and a regression coefficient table. The tables are identical to those in the previous screenshot.

【商务与经济统计单元项目】姓名：...罗欣萍...学号：...23002430124...

第3章描述性统计学：数值方法、第12章：简单线性回归

案例：汽车价值的合理评估

(1) 对重要的数据进行数值概述。

表1-数值概述。

| | 价格 | 费用/英里 | 道路测试分 | 预测的可靠性等级 | 评估分数 |
|-----|----------|---------|-------|----------|---------|
| 平均值 | 26886.20 | 0.6420 | 80.45 | 3.75 | 1.4580 |
| 标准差 | 3374.284 | 0.06118 | 9.902 | 0.639 | 0.19745 |
| 中位数 | 28067.50 | 0.6650 | 82.00 | 4.00 | 1.4250 |
| 众数 | 21800 | 0.67 | 80 | 4 | 1.20 |
| 全距 | 10560 | 18 | 41 | 2 | 0.7 |
| 最小值 | 21800 | 0.56 | 52 | 3 | 1.05 |
| 最大值 | 32360 | 0.74 | 93 | 5 | 1.75 |

(2) 回归分析1。

利用回归分析方法建立一个回归方程，当汽车价格已知时，能够用方程预测该车的评估分数。

输入除去的变量^a

模型摘要

| 模型 | R | R方 | 调整后R方 | 标准误差的估计 |
|----|------|-----|-------|---------|
| 1 | 572* | 328 | 290 | 16633 |

ANOVA^b

| 模型 | 平方和 | 自由度 | 均方 | F | 显著性 |
|-------|-----|-----|-----|-------|-------|
| 1 回归 | 243 | 1 | 243 | 8.775 | .008* |
| 残差 | 488 | 18 | 28 | | |
| Total | 741 | 19 | | | |

系数^c

| 系数 | Beta系数 | t | 显著性 |
|----|----------|-------|------|
| 截距 | | | |
| 价格 | 0.008206 | 0.798 | .000 |

The screenshot shows a Microsoft Word document titled "项目4-23002430122-张梦欢.docx". The layout is identical to the previous one, featuring a table of descriptive statistics, an ANOVA table, and a regression coefficient table.

【商务与经济统计单元项目】姓名：...张梦欢...学号：...23002430122...

第3章描述性统计学：数值方法、第12章：简单线性回归

案例：汽车价值的合理评估

(1) 对重要的数据进行数值概述。

表1-五数统计。

| | 价格 | 费用/英里 | 道路测试分 | 预测的可靠性等级 | 评估分数 |
|-----|----------|---------|-------|----------|---------|
| 平均值 | 26886.20 | 0.6420 | 80.45 | 3.75 | 1.4580 |
| 标准差 | 32067.50 | 0.6650 | 82.00 | 4.00 | 1.4250 |
| 中位数 | 21800 | 0.67 | 80 | 4 | 1.20 |
| 众数 | 3374.284 | 0.06118 | 9.902 | 0.639 | 0.19745 |
| 全距 | 10560 | 18 | 41 | 2 | 0.7 |
| 最小值 | 21800 | 0.56 | 52 | 3 | 1.05 |
| 最大值 | 32360 | 0.74 | 93 | 5 | 1.75 |

SPSS Statistics

模型摘要

| 模型 | 平方和 | 自由度 | 均方 | F | 显著性 |
|-------|-----|-----|-----|-------|-------|
| 1 回归 | 243 | 1 | 243 | 8.775 | .008* |
| 残差 | 488 | 18 | 28 | | |
| Total | 741 | 19 | | | |

ANOVA^b

| 模型 | 平方和 | 自由度 | 均方 | F | 显著性 |
|-------|-----|-----|-----|-------|-------|
| 1 回归 | 243 | 1 | 243 | 8.775 | .008* |
| 残差 | 488 | 18 | 28 | | |
| Total | 741 | 19 | | | |

系数^c

| 系数 | Beta系数 | t | 显著性 |
|----|----------|-------|------|
| 截距 | | | |
| 价格 | 0.008206 | 0.798 | .000 |

(4) 回归分析 3

利用回归分析方法建立一个回归方程，当道路测试分数已知时，能够用方程预测该车的评估分数。

样本量：20

因变量：评估分数

自变量：道路测评分数

SPSS 结果：

回归方程的估计式： $\hat{y}=0.798+0.008206x$

相关系数：r=0.412

可决系数：r²=0.169

回归系数的实际意义：0.008206，价格每增加一美元评估分数会减少0.008206

可决系数的实际意义：评估分数16.9%的波动和评估分数之间的线型关系所解释

图2 改革前考核痛点实证：项目报告的高度同质化与机械式复制错误（原项目4）

(二) 改革目标

本课程坚持以学生发展为中心的工作思路，针对上述痛点，确立以下改革目标：

1. 推行非标准考核，提升高阶认知落实

落实“从标准考试向非标准考核转变”的要求，摒弃死记硬背与单纯套公式的题目（如旧卷中的定义判断与参数计算），将非标准答案试题占比提升至核心地位。重点考察学生在复杂商业情境下解读 SPSS 结果、进行推断决策的高阶能力。

2. 重构形成性评价体系，注重持续改进

落实“从重结果向重过程转变”的要求，将形成性考核占比提升至 50%。建立全周期的过程管理机制，利用课堂互动与阶段性任务反馈，即时纠正学生逻辑错误，动态反馈改进结果，确保学生真实掌握数据分析技能，而非仅在期末突击。

3. 量化评价标准，根治产出同质

落实“从主观判断向客观评价转变”的要求，引入清晰的评价量规。针对项目报告设定个性化与开放性指标，明确要求商业结论的独创性与逻辑自洽性。通过精细化的评分标准引导学生独立思考，有效杜绝机械复制与“千人一面”的作业现象。

二、考核改革设计实施

本课程构建的“双线耦合、四阶递进、数智协同、量规导向”的考核体系，形成性、终结性考核占比各 50%。通过重构命题逻辑与项目评价标准，确保课程核心能力的培养与达成。

1. 双线耦合：构建“形成性积累、终结性升维”的连贯促学体系

课程打破“练考分离”壁垒，建立全周期连贯评价机制。形成性考核聚焦全流程实战，通过项目 1 至 4，积累数据处理与模型构建的“肌肉记忆”；终结性考核则紧扣试卷中的“异常值判断”、“P 值决策”及“回归系数解读”等核心考点，检验学生能否将实战经验转化为理论直觉，详见表 1。形成性考核为理论基础提供支撑、操作实践，终结性考核则进行逻辑验收，实现从计算操作到数据决策升维。

2. 四阶递进：实施“全景实战驱动”的真实项目考核模式

课程紧扣统计学核心知识体系，构建了由浅入深的能力螺旋上升考核路径。在描述性统计阶段，要求学生超越基础指标计算，从纷繁数据中提炼精准的商业洞察；在概率分布阶段，引导学生突破静态习题，通过动态参数建模挑战复杂的风险模拟；进入推断统计阶段，迫使学生走出单纯的 P 值判断，通过多样的假设验证实现严谨的逻辑重构；最终在回归分析阶段，要求学生打破对模型拟合度的盲从，结合业务逻辑剔除伪关联变量，达成批判性决策。这一纵向贯通的设计，实现了基础方法掌握与高阶问题解决能力的同步跃升。

表 1 “双线耦合”形成性与终结性连贯促学体系

| 项目 | 形成性考核内容 | 考核改革 | 与终结性考核的连贯性 |
|---|---|--|---|
| 项目 1 描述性统计应用 【销售/赛事数据 描述性统计】 (第 1-3 章) | 自主选择数据集, Kaggle 平台上零售、运动领域真实场景数据, 描述数据集特征, 合理划分变量类型, 根据变量类型, 挑选最合适的描述性统计方法, 发现业务规律。 | 数据处理实战: 由处理完美数据转向多源含噪数据清洗, 重点考核基于 IQR (四分位距) 规则的异常值识别与处理逻辑。 统计方法自主: 由指定图表指令转依据变量类型适配, 考核针对数值型/分类型变量自主选择恰当描述性统计方法的能力。 分析维度高阶: 由单纯数值计算转向统计指标解读, 考核利用集中趋势与离散程度挖掘商业洞察的能力。 | 试卷中题目考核了“异常值的影响”、“分位数逻辑”以及“分布特征描述”, 是项目 1 中训练的核心技能。 通过平时利用 IQR 规则与箱线图清洗数据, 深刻理解异常值对统计指标的干扰; 有效支撑期末考核中针对变异指标稳健性的判断(判断题 3) 及位置测度指标的计算(选择题 3) 从软件操作到理论内化的连贯升级 |
| 项目 2 概率分布应用 【Amazon 电子产 品需求预测与库存 决策分析】 (第 4-6 章) | 基于 Amazon 产品销售数据, 选择一款电子产品作为分析对象, 运用正态分布理论自主独立建立需求预测模型, 并通过缺货概率分析、利润情景模拟和服务水平优化, 为库存决策提供数据支持。 | 场景建模实战: 由已知分布做计算转向自主拟合真实分布。 参数设定自主: 由教科书给定均值方差转向基于历史数据估算参数, 考核从样本中提取信息估计总体参数的能力。 决策博弈高阶: 由静态概率求解转向动态情景模拟, 考核利用 Z 分数与服务水平在高库存积压与高缺货风险之间进行量化权衡的能力。 | 试卷中题目核心考察了期望计算(计算题 1.1)、正态分布概率求解(计算题 1.3) 以及基于置信水平的库存设定(计算题 1.4)。通过平时项目利用真实数据估算参数并进行标准正态变换; 有效支撑期末考核中在不确定性环境下利用期望值评估盈亏(计算题 1.2) 及利用逆向正态分布确定安全库存水平的复杂应用能力。 从简单概率计算到风险量化决策的连贯升级 |
| 项目 3 统计推断应用 【基于真实电商数 据的统计推断】 (第 7-9 章) | 基于在线零售平台 5,000 条真实交易数据, 模拟数据分析师决策流程, 自设检验问题, 并设立原假设、备择假设, 随机抽取交易样本, KPI 达标验证, 建立原假设与备择假设, 实施单样本 t 检验, 依据 P 值与显著性水平判断差异是否具有统计学意义, 并撰写商业决策报告。 | 推断逻辑构建: 由套公式计算统计量转向假设体系构建, 重点考核根据业务问题。 结果解读便捷: 由查表比对临界值转向 p 值决策法则, 考核直接阅读软件输出的 p 值, 并结合显著性水平 α 判定结果的能力。 风险认知挑战: 由单纯判定转向风险评估, 考核结合业务场景解释置信水平及第 I 类错误后果的能力 | 试卷综合题 2 全面考察了抽样分布特征(任务一)、单样本 t 检验(任务二) 及双样本差异分析(任务三)。通过平时项目亲手进行随机抽样, 直观感受样本量 n 对标准误的影响(任务一); 有效支撑期末针对 p 值的解读、“拒绝域”的判定, 帮助学生在面对复杂的供应商绩效对比(任务三)时, 能迅速匹配检验方法。 从理论假设到业务实证的连贯升级 |
| 项目 4 线性回归应用 【基于线性回归的 BoomBikes 共享单 车需求分析】 (第 10-13 章) | 基于 BoomBikes 共享单车需求数据集, 理解项目背景, 自主设定拟研究的回归问题, 构建简回归模型、多元回归模型, 自主选择自变量, 检验模型显著性, 解读判定系数与回归系数的商业含义, 量化各因素对需求的边际贡献。 | 多维建模实战: 由简单回归转向多元复杂建模, 重点考核处理多重共线性及筛选最优预测变量组合的能力。 变量处理进阶: 突破数值型限制, 考核利用虚拟变量将分类型数据纳入回归模型的高阶技巧。 模型解释高阶: 由求出方程转向解释方程, 考核将数学上的回归系数转化为具体商业策略的洞察力。 | 试卷综合题 1 重点考察了虚拟变量处理(1.1)、回归方程构建(1.2) 及系数与意义。通过平时项目处理含虚拟变量的复杂回归模型, 深刻理解截距与斜率的业务指代; 无缝衔接期末考核中要求解释“企业类型”如何影响利润以及判断模型拟合优度的考点。 从统计建模到商业策略制定的连贯升级 |

3. 数智协同：构建“AI 阶梯式赋能”的学-评-创模式

课程突破单一的技术应用模式，依据认知深度设计了 AI 角色随项目难度逐级演进的阶梯路径。从初阶（项目 1）作为提效助手辅助数据清洗，到中阶（项目 2、3）升级为开发伙伴与逻辑陪练，协同攻克交互建模与推断逻辑难关，直至高阶（项目 4）实战中转变为被审视对象，由学生主导辩证思考并自主完成项目结论。这种循序渐进的协同机制清晰界定了不同阶段的人机边界，确保数智技术始终服务于思维训练，有效引导学生完成从“工具依赖”向“自主驾驭”的根本跨越。



图 3 AI 阶梯式赋能 “学-评-创” 模式

4. 量规导向：确立基于量规的动态考核评价体系

课程细化评价颗粒度，针对每一具体评分项构建了初始级、发展级、示范级的三阶明细量规。这一可视化标准将抽象的质量要求具象化，帮助学生精准定位如“回归方程解释”、“AI 辅助及格式”等特定维度的现状。依托企业微信在线表格的实时反馈机制，量规转变为动态的改进路标，允许学生依据诊断结果进行多轮修正完善。这种透明化的迭代过程，使学生能够清晰感知自身在每一环节的进步与完善，真正实现了从“被动接受判分”到“主动补齐短板”的成长闭环。

表 2 项目评价量规（以项目 4 为例）

| 打分项 | 初始级 Capstone | 发展级 Milestone | 示范级 Benchmark |
|--------|----------------------------|------------------------------|--------------------------|
| 研究背景分析 | 缺少数据集变量介绍，未说明拟解决问题（0 分） | 变量介绍或拟解决问题描述不恰当（0.5 分） | 逻辑清晰，背景、变量、回归问题阐述清晰（1 分） |
| 线性回归方程 | 方程数量不足，因变量、自变量均错误（0-1 分） | 有 1-2 个方程正确，直接将分类变量纳入回归（2 分） | 模型构建严谨、正确，方程表达式无误（3 分） |
| 回归方程解释 | 回归方程关于系数、显著性等解释大量错误（0-1 分） | 回归方程关于系数、显著性等解释少量错误（2 分） | 回归方程解释均正确，且能阐述数据意义（3 分） |

| 打分项 | 初始级 Capstone | 发展级 Milestone | 示范级 Benchmark |
|----------|--|--|---|
| 回归方程应用 | 未进行任何应用分析，或仅停留于方程的展示，无法用于解决实际问题。（0分） | 能够代入具体数值进行简单的点预测（如“当气温为 20 度时的预估销量”），计算过程正确。（0.5 分） | 能够基于模型结果提出落地的策略建议。例如根据体感文档提出营销投放时间表，设计动态运维方案。（1分） |
| AI 辅助及格式 | 明显直接复制 AI 生成的代码或文字，排版混乱，存在代码报错或格式未对齐，缺乏人工复核与整理。（0.5 分） | 报告排版整洁，图表规范。合理使用 AI 生成示例，主要内容由学生主导，格式符合学术报告基本规范。（1分） | 利用 AI 生成了高质量的示例，且学生对 AI 结果进行了批判性修正，报告兼具专业性与美观度。（2分） |

25-26-1 商务与经济统计【项目1】姓名: ··许圣冉···学号: ··24002430105···

第2章描述性统计: 表格与图形、第3章描述性统计: 数值方法。

| 打分栏 | | | | |
|-----------|-----|-----|-----|-----|
| 打分项 | 分值 | 初始级 | 发展级 | 示范级 |
| 原始数据介绍 | 1分 | 0 | 0.5 | 1✓ |
| 位置指标与变异指标 | 2分 | 0.5 | 1 | 2✓ |
| 统计图表 | 2分 | 0.5 | 1✓ | 2 |
| 数据标准化处理 | 1分 | 0✓ | 0.5 | 1 |
| 相关性分析 | 2分 | 0.5 | 1 | 2✓ |
| 总结与结论 | 2分 | 0.5 | 1✓ | 2 |
| 总分 | 7.5 | | | |

25-26-1 商务与经济统计【项目2】姓名: ··许圣冉···学号: ··24002430105···

第5章离散型概率分布、第6章连续型概率分布: 概率分布在商业决策中的应用。

| 打分栏 | | | | |
|-----------|----|------|------|-----|
| 打分项 | 分值 | 初始级 | 发展级 | 示范级 |
| 产品选择与参数设定 | 1分 | 0✓ | 0.5 | 1 |
| 概率分布建模与应用 | 2分 | 0.5✓ | 1 | 2 |
| 缺货风险分析 | 2分 | 0.5 | 1✓ | 2 |
| 利润情境模拟 | 2分 | 0.5 | 1✓ | 2 |
| 决策优化与建议 | 2分 | 0.5 | 1✓ | 2 |
| 报告规范与展示 | 1分 | 0 | 0.5✓ | 1 |
| 总分 | 4 | | | |

25-26-1 商务与经济统计【项目3】姓名: ··许圣冉···学号: ··24002430105···

第7章抽样和抽样分布、第8章区间估计、第9章假设检验。

| 打分栏 | | | | |
|-------------|----|------|-----|-----|
| 打分项 | 分值 | 初始级 | 发展级 | 示范级 |
| 问题定义与假设建立 | 2分 | 0.5 | 1✓ | 2 |
| 抽样设计与执行 | 2分 | 0.5 | 1✓ | 2 |
| 假设检验分析 | 2分 | 0.5✓ | 1 | 2 |
| 结果解读与反思 | 2分 | 0.5 | 1✓ | 2 |
| 报告质量与 AI 协作 | 2分 | 0.5✓ | 1 | 2 |
| 总分 | 4 | | | |

25-26-1 商务与经济统计【项目4】姓名: ··许圣冉···学号: ··24002430105···

第10章总体均值的比较、试验设计及方差分析、第12章简单线性回归、第13章多元回归。

| 打分栏 | | | | |
|----------|----|-----|-----|-----|
| 打分项 | 分值 | 初始级 | 发展级 | 示范级 |
| 研究背景分析 | 1分 | 0 | 0.5 | 1✓ |
| 线性回归方程 | 3分 | 1 | 2 | 3✓ |
| 回归方程解释 | 3分 | 1 | 2 | 3✓ |
| 回归方程应用 | 1分 | 0 | 0.5 | 1✓ |
| AI 辅助及格式 | 2分 | 0.5 | 1✓ | 2 |
| 总分 | 9 | | | |

图 4 基于量规的学生项目打分栏（项目报告首页）

| 形成性考核项目成绩 | | | | | | | | | | | |
|-------------|-----|--------------------|--------------------|---------------------|-----|-----|--------------------|--|--|--|--|
| 项目4 | | | | | | | | | | | |
| 学号 | 姓名 | 4.1 研究背景分析 (1分) | | | | | 4.2 线性回归方程 (3分) | | | | |
| | | 4.3 回归方程解释 (3分) | 4.4 回归方程应用 (1分) | 4.5 AI辅助及格式 (2分) | | 小计 | | | | | |
| 24002430135 | 刘欣悦 | 1 | 3 | 3 | 1 | 2 | 10 | | | | |
| 24002430136 | 杜杰 | 1 | 3 | 3 | 1 | 2 | 10 | | | | |
| 24002430137 | 蔚璐楠 | 1 | 1 | 1 | 0.5 | 0.5 | 4 | | | | |
| 24002430138 | 王美惠 | 1 | 2 | 2 | 1 | 1 | 7 | | | | |
| 24002430201 | 孙铁楠 | 1 | 2 | 2 | 1 | 1 | 7 | | | | |
| 24002430202 | 张昊天 | 0.5 | 3 | 3 | 1 | 1 | 8.5 | | | | |
| 24002430203 | 梁智 | 1 | 3 | 3 | 1 | 1 | 9 | | | | |
| 24002430204 | 于相东 | 1 | 1 | 2 | 0.5 | 0.5 | 5 | | | | |
| 24002430205 | 谢桂东 | 1 | 2 | 3 | 1 | 0.5 | 7.5 | | | | |
| 24002430206 | 王富林 | 1 | 3 | 2 | 0.5 | 2 | 8.5 | | | | |
| 24002430207 | 孙博宇 | 1 | 3 | 3 | 1 | 0.5 | 8.5 | | | | |
| 24002430208 | 王子涵 | 1 | 2 | 2 | 0 | 1 | 6 | | | | |
| 24002430209 | 张其源 | 1 | 2 | 3 | 1 | 1 | 8 | | | | |
| 24002430210 | 刘佳宇 | 1 | 3 | 2 | 0.5 | 1 | 7.5 | | | | |

图 5 基于量规的在线表格动态反馈

三、考核改革成效亮点

(一) 量化与质性成效

1. 质性成效：深度学习与教学重构的双向奔赴

(1) 学生学习范式的根本性转变：从“被动应付”走向“主动研磨”

依托“学-评-创”一体化的迭代反馈机制，学生对考核任务的态度发生了质的飞跃，不再将作业视为一次性的任务终点，而是将其看作需要持续打磨的专业作品。在线协作表格与企业微信的高频互动、文件夹中多次修改的版本都见证了这一转变：绝大多数学生在收到针对性的量规反馈后，不再满足于及格分数，而是主动发起两轮甚至多轮的修正。这种“做中学、评中改”的良性循环，彻底激活了学生的学习内驱力，使其从关注分数的“讨价还价”转向了补齐短板的“精益求精”，展现出极高的自我效能感与专业认同感。

| 名称 | 修改日期 | 类型 | 大小 |
|----------------------------|------------------|-------------------|----------|
| 24002420308-王子涵（改进版）.docx | 2025/12/26 21:29 | Microsoft Word 文档 | 444 KB |
| 24002430101-宋宸博.docx | 2025/12/21 20:13 | Microsoft Word 文档 | 293 KB |
| 24002430102-王奥（改进版）.docx | 2025/12/23 14:43 | Microsoft Word 文档 | 975 KB |
| 24002430102-王奥.docx | 2025/12/21 20:17 | Microsoft Word 文档 | 584 KB |
| 24002430104-赵御屹（改进版）.docx | 2025/12/23 13:27 | Microsoft Word 文档 | 1,267 KB |
| 24002430104-赵御屹.docx | 2025/12/21 20:22 | Microsoft Word 文档 | 295 KB |
| 24002430105-许圣冉（改进版）.docx | 2025/12/23 15:38 | Microsoft Word 文档 | 426 KB |
| 24002430105-许圣冉（改进版2）.docx | 2025/12/23 17:32 | Microsoft Word 文档 | 442 KB |
| 24002430105-许圣冉（改进版3）.docx | 2025/12/24 9:26 | Microsoft Word 文档 | 428 KB |
| 24002430105-许圣冉.docx | 2025/12/21 20:26 | Microsoft Word 文档 | 424 KB |
| 24002430106-杨剑波.docx | 2025/12/21 20:28 | Microsoft Word 文档 | 1,113 KB |
| 24002430107-金瀛兴.docx | 2025/12/23 13:55 | Microsoft Word 文档 | 291 KB |
| 24002430110-王子超.docx | 2025/12/21 20:33 | Microsoft Word 文档 | 380 KB |
| 24002430111-徐铭莹（改进版）.docx | 2025/12/23 12:37 | Microsoft Word 文档 | 490 KB |
| 24002430111-徐铭莹（改进版2）.docx | 2025/12/23 15:24 | Microsoft Word 文档 | 481 KB |
| 24002430111-徐铭莹（改进版3）.docx | 2025/12/23 18:20 | Microsoft Word 文档 | 552 KB |

图 6 多轮修改文档（以项目 4 部分学生成果为例）

(2) 学生思维深度的阶梯式跃升：从“工具依赖”迈向“批判决策”

在 AI 阶梯式赋能的引导下，学生的认知水平突破了传统的唯结果论。在项目实战中，学生不再止步于跑通软件流程或盲从 AI 生成的通用结论，而是展现出深刻的商业洞察力与批判性思维。特别是在高阶的共享单车需求回归分析中，学生能够敏锐地识别出统计模型与业务逻辑的冲突，主动剔除 AI 未能识别的伪关联变量，并将回归系数转化为落地的运维策略。这种“人主导决策、AI 辅助验证”的思维模式，标志着学生真正完成了从单纯的数据计算者向具备复杂问题解决能力的决策者的蜕变。

(3) 教学设计维度的实战化重构：从“习题演示”升级“高阶挑战”

课程构建了校内教师、行业专家、AI 的多元协同命题机制，彻底摒弃了传统教科书中的僵化习题，转而营造具有高度复杂度、挑战度与契合度的实战场域。通过引入 Kaggle 等平台的真实数据与经专家校验的开放式商业命题，课程考核精准锚定电子商务与供应链管理岗位的核心技能需求。这种多主体共

创的高阶性项目设计迫使学生必须在不确定性环境中，综合调用统计原理、AI 编程技能与商业直觉来解决非标准化问题，不仅有效承载了“两性一度”的课程建设标准，更实现了考核内容与职业能力的无缝对接。

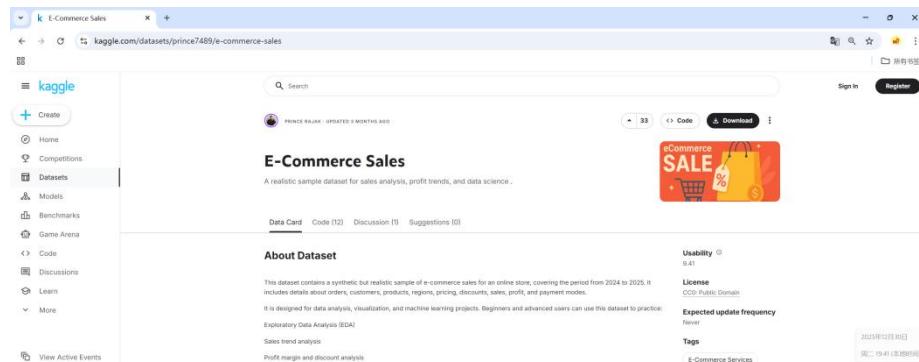


图 7 Kaggle 行业级命题（项目 3 数据集为例）

2. 量化成效：核心指标的结构性优化与达成

（1）考核内容的动态迭代率：彻底根除“题型固化、仅换数字”的僵化模式

课程改革坚决打破了传统统计学考核中题“换数不换题”的机械惯性，建立了高频的资源更新机制。本学期 A 卷考核内容深度融合了“十五五”规划建议、跨境电商供应链、冷链物流管理等最新行业热点，案例更新度 100%。试题借用“优购网”真实订单数据背景，引导学生在新颖的业务语境下重新建模，从源头上杜绝了死记硬背与套用旧模板的可能。

（1）任务一：抽样分布分析

公司计划从 15,000 笔订单中抽取一系列样本进行深入研究，随机抽取样本。假设已知订单履约时长近似服从正态分布，某次抽样分析中，履约时长样本均值的标准差（标准误差）为 0.82 小时，解释该数值在实际业务中的意义，描述该样本均值的抽样分布特征（形状、中心、离散程度），若增加样本量，标准误差会如何变化？为什么？（4 分）

（2）任务二：总体绩效检验

平台设定的履约时长目标为均值不超过 48 小时，检验实际均值是否显著高于目标，建立适当的检验假设（原假设与备择假设），说明检验的类型（单尾/双尾检验），选择检验方法（单样本 t 检验/独立样本 t 检验/配对样本 t 检验）。显著性水平 $\alpha=0.05$ ，检验的 p 值为 0.008，根据检验结果给出统计结论，结合业务背景解释其含义，分析潜在负面影响。（6 分）

（3）任务三：供应商差异检验

讯科智联： $n_1=165$, $\bar{x}_1=49.3$ 小时, $s_1=10.5$ 小时

星途数码： $n_2=144$, $\bar{x}_2=46.8$ 小时, $s_2=9.6$ 小时

比较讯科智联与星途数码的履约时长差异，建立适当的检验假设（原假设与备择假设），说明检验的类型（单尾/双尾检验），选择检验方法（单样本 t 检验/独立样本 t 检验/配对样本 t 检验），如果二者均值差异显著，假如你作为公司的供应链经理，会如何优先采取下一步措施（6 分）。

（4）任务四：区域均衡性分析

公司希望了解四个区域间的履约时长是否存在显著差异，建立单因素方差分析的原假设和备择假设，ANOVA 检验结果 $F=10.35$, $p<0.001$ ，解释统计结果含义，结合业务，提出对其供应链优化的可行性建议。（4 分）

图 8 基于真实供应链场景（优购网）的非标准化试题改革实证

(2) 终结性试卷结构的优化比：从“机械计算”向“高阶应用”的权重转移

在期末试卷的设计上，课程大幅压缩了单纯考察记忆与套公式计算的低阶题目。数据显示，综合案例分析题与开放式决策题在试卷中的分值占比由原本的不足 30% 跃升至 50%（如第四大题占比 30 分及第三大题部分）。试卷引入了 SPSS 软件输出结果解读题，替代了传统的手工复杂计算，重点考察学生“读懂数据”而非“制造数据”的能力，有效引导学生将精力从“刷题”转向对统计思想的深度理解。

(3) 预期学习效果 (ILO) 的覆盖度：全方位锚定能力进阶目标。

通过对试卷考核点的精准映射，课程实现了对核心预期学习效果的全面验收。针对 ILO-6 定义与功能与 ILO-5 数据获取的基础考察权重控制在 25% 左右，确保理论地基稳固；针对 ILO-2 实验性探索（工具选择与计算）的考察占比约为 25%，强化方法论的正确应用；而针对高阶的 ILO-3 问题发现与表达（将供应链痛点转化为假设检验）及 ILO-1 结论思考与 ILO-4 整体性思维（基于 ANOVA 结果提出供应链优化建议）的考察权重合计达到 50%。基础、应用与决策能力考核权重调整为 2:3:5，这一结构性变革客观量化并有力支撑了考核体系向数据驱动商业决策能力的深度转型。

(二) 改革前后对比

由于 2025-2026-1 终结性考核尚未进行，基于本学期已完成的 50% 形成性考核数据（涵盖 4 个项目），课程对 6 项预期学习效果进行了形成性考核的达成度测算。

改革前，学生的能力结构呈现显著的畸形特征，特别是 ILO-6（逻辑构建与参数界定）仅为 56.48 分，低于合格线，更成为制约决策能力的致命短板。这反映了传统考核下学生只算数、不懂逻辑的通病。成绩分布上，44.44% 的学生拥挤在中等分数段，而良好分数段仅占 7.41%，呈现出明显的中间大、两头小的橄榄型特征，说明大部分学生处于懂一点但不多的平庸状态。

改革后，雷达图呈现出显著的高位均衡特征，各项能力达成度均值紧密分布在 81% 左右。其中，代表理论建构能力的 ILO-6（逻辑模型构建）得分最高 82.00，与代表高阶商业决策能力的 ILO-1（批判性解读，81.99）及 ILO-2（实验性推演，81.98）形成了理论与应用的双强支撑。这一数据有力证实，通过“双线耦合、四阶递进、数智协同、量规导向”的考核改革，学生不仅夯实了基础统计技能（ILO-5, 81.33），更在核心的商业决策思维上实现了全面突围。相比于改革前“重记忆、轻应用”导致的各种弊端。成绩分布上，最显著的变化在于中等分数段的学生比例从 7.41% 增至 30.67%。与此同时，优秀率保持稳定，小幅提升。这表明，三级进阶量规的持续反馈机制，帮助原本处于中等水平的大量学生突破了能力瓶颈，达成优良。改革不是简单地让高分更多，而是让更多的学生通过努力够到了高阶能力的门槛，实现了从合格到优秀的自主性进步。

针对课程最核心、也是传统教学中最难攻克的痛点，ILO-6（基于经济理论的统计建模与逻辑界定能力），进行了三年的纵向追踪分析。数据显示，该指标的平均达成度呈现出显著提升，改革后绝大多数样本点向 80 分以上高分段的密集收敛，有力证明了通过持续的高阶项目实战，课程成功破解了无法

将统计方法联系业务实际的教学顽疾，实现了学生从浅层计算向深度建模逻辑的根本性质变。

各个课程预期学习效果平均达成度

各个课程预期学习效果平均达成度

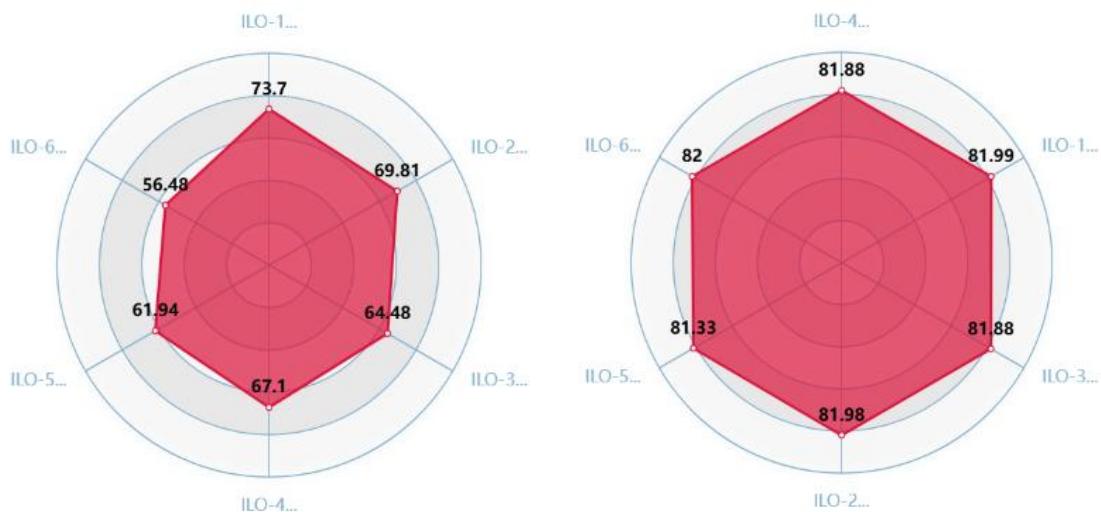
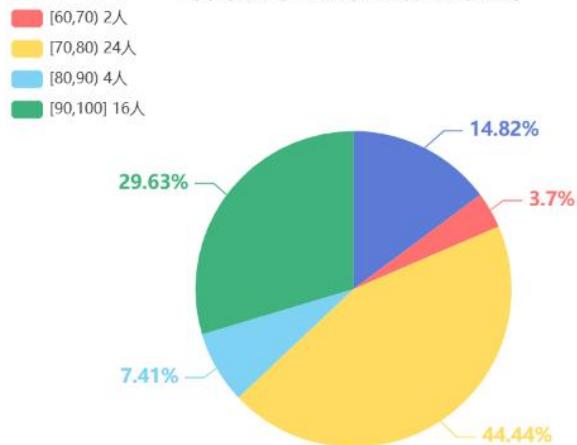


图 9 改革前后预期学习效果达成度对比（左图为 24-25-1、改革前，右图为 25-26-1、改革后）

课程预期学习效果达成度分布统计



课程预期学习效果达成度分布统计

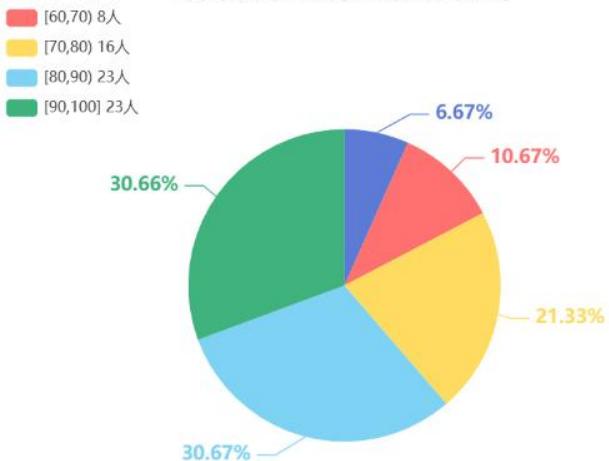
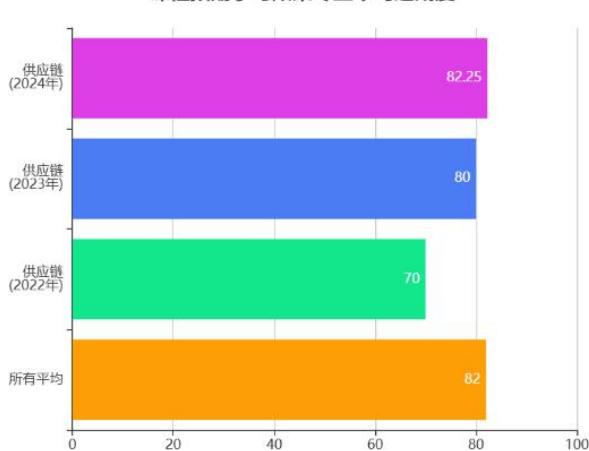


图 10 改革前后成绩分布对比（左图为 24-25-1、改革前，右图为 25-26-1、改革后）

课程预期学习效果专业平均达成度



课程预期学习效果达成度分布

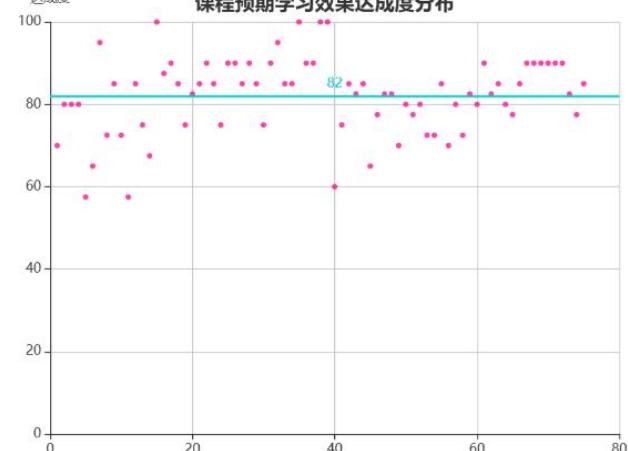


图 11 近三年 ILO-6 预期学习效果跟踪及改革后学生 ILO-6 达成度分布情况

(三) 学生优秀作答/成果

为了直观呈现量规导向、以评促改改革实效，课程选取了考核项目 2 概率分布在商业决策中的应用作为典型观测点。该项目要求学生基于 Amazon 真实销售数据，在不确定性环境下进行库存博弈与利润模拟。以下展示了学生孙博宇在收到初次诊断反馈后，如何利用教师个性化指导、三级进阶量规进行深度迭代。该案例清晰地记录了学生从关注计算正确转向追求商业决策优化的成长轨迹，有力印证了非标准化考核在培养学生高阶思维与解决复杂问题能力方面的显著成效。

案例对象：2024 级供应链管理专业 24002430207 孙博宇



图 12 教师基于初次评分的个性化指导

1. 基础能力：逻辑建构与知识应用（对应量规评分项：概率分布建模）

点评：孙博宇同学准确抓住了正态分布在库存管理中的核心应用逻辑。在初始版中，他已能利用 `pd_last_month(10000)` 设定均值，并正确设定了三种 CV（变异系数）情景下的标准差，完成了 Z 值与缺货概率的基础计算。

改进亮点：改进版中，他修正了对“服务水平”的理解，不再机械地罗列公式，而是将 Z=-2.5 到 2.5 的计算结果与具体的业务含义（保守策略 vs 激进策略）进行了精准匹配，展现了扎实的统计学理论功底。

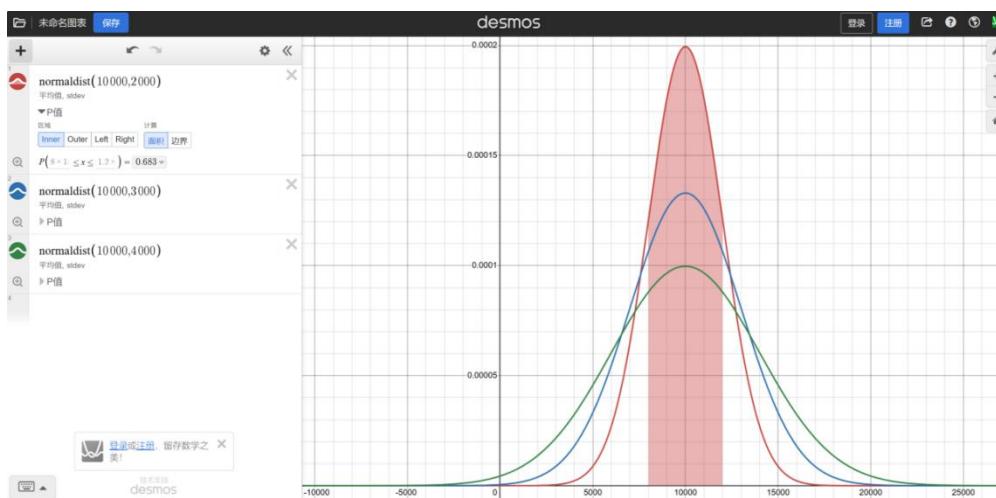


图 13 不同变异系数下需求概率分布

2.高阶思维：专业融合与思维深度（对应量规评分项：决策优化与建议）

点评：这是本案例最大的亮点。在初始版中，学生在推荐订货量时存在逻辑摇摆，既想推荐 8000 件（稳健），又发现 11573 件（70%服务水平）利润更高，结论模棱两可。

改进亮点（从“发展级”跃升至“示范级”）：在改进版中，学生展现了极强的批判性思维。他没有盲从 90% 的高服务水平，而是通过计算发现 90% 服务水平下预期利润反而下降（105 万降至 77 万）。他敏锐地指出对于高单价（\$280）、高贬值风险的电子产品，盲目追求高服务水平会导致库存积压成本压缩利润，从而坚定推荐 70% 服务水平方案。这一判断完美体现了统计学与供应链成本管理的跨学科融合。

3.实践创新：数智协同与个性特色（对应量规评分项：AI 辅助及格式）

点评：学生在附录中展示了利用 AI（豆包）辅助生成的交互式分析线程，并进行了人工修正。

改进亮点：报告未直接复制 AI 生成的通用建议，而是结合选定的具体产品（Apple Watch Series 10）进行了场景化落地。学生利用 AI 辅助模拟了“黑五”大促场景，提出了日常维持 11573 件、大促临时提升至 12525 件的动态策略。这种“人机协同”的作业模式，既利用 AI 拓宽了思路，又保持了人类决策的主导权。

The screenshot shows a web application titled "Amazon 电子产品需求预测与库存决策分析". The interface includes a sidebar with "需求分布", "利润分析", and "服务水平" tabs, and an "AI生成" button. The main content area displays a blue banner with the title and a subtitle: "基于Apple AirPods 4 Wireless Earbuds销售数据, 通过交互式工具优化库存决策". Below the banner are two buttons: "开始分析" and "了解更多". On the left, there is a code editor window showing HTML and JavaScript code for generating a normal distribution of demand data. The code includes functions for generating points and rounding values.

The screenshot shows a web-based tool titled "Amazon Apple Watch Series 10 订货量-利润计算工具". It has two main sections: 1. 选择需求变异系数 (CV) and 2. 推荐订货量及利润明细. In section 1, a dropdown menu for "变异系数 (CV)" is open, showing options: 0.3 (中等波动性), 0.2 (低波动性), 0.3 (中等波动性) (selected), and 0.4 (高波动性). A note below the dropdown states: "注: 基于文档设定, 0.3 (中等波动性) 为推荐值, 1.4 时需求分散, 默认推荐70%服务水平对应的订货量". In section 2, a table displays the recommended order quantity and profit details for three scenarios: 最差情景, 最可能情景, and 最好情景. The table columns include: 变异系数 (CV), 推荐订货量 (件), 需求情景, 总成本 (美元), 销售收入 (美元), 清仓收入 (美元), 利润 (美元). The table rows are:

| 变异系数 (CV) | 推荐订货量 (件) | 需求情景 | 总成本 (美元) | 销售收入 (美元) | 清仓收入 (美元) | 利润 (美元) |
|-----------|-----------|-------|----------------|----------------|--------------|----------------|
| 0.3 | 11,573 | 最差情景 | \$1,735,950.00 | \$1,400,600.00 | \$657,300.00 | \$321,950.00 |
| 0.3 | 11,573 | 最可能情景 | \$1,735,950.00 | \$2,801,200.00 | \$157,300.00 | \$1,222,550.00 |
| 0.3 | 11,573 | 最好情景 | \$1,735,950.00 | \$3,241,828.76 | \$0.00 | \$1,505,878.76 |



图 14 基于 AI 生产交互式网页制定订货量

4. 价值延伸：实践价值与应用潜力（对应量规评分项：利润情境模拟）

点评：报告超越了单纯的作业范畴，具备了商业计划书的雏形。

改进亮点：学生在文末提出的“动态参数更新机制”（建议建立月度复盘，根据实际销量动态调整 μ 与 CV 值）极具实战价值。这表明学生深刻理解了商业环境的动态非标准特性——没有一劳永逸的统计模型，只有持续迭代的数据决策。这一结论完全符合课程“培养数据驱动决策者”的改革目标。

孙博宇同学的案例生动诠释了“以评促改”的成效。通过量规的指引和多轮反馈，他成功从一个会算概率题的学生，成长为一名能权衡风险、成本与利润的初级数据分析师，契合考核改革中“重过程、重应用、重决策”的核心导向。

四、总结反思与持续改进

本次改革成功驱动了学生从机械计算向数据决策的能力跃迁，但高阶开放性项目的规模化评价仍面临挑战。反思发现，当前尝试引入的 AI 参与考核评价，虽在基础格式检查上效率显著，但在处理复杂的商业逻辑归因与创新性判断时，其信度与效度尚显不足，目前仍处于需要强人工干预阶段，难以独立承担精准评价任务。

后续改革将重点攻坚 AI 评价的可靠性。计划通过构建更精细的评分语料库、回复逻辑对模型进行定向训练，逐步提升其对非标准答案的语义理解力。同时，坚定确立“教师主导、AI 辅助”的评价原则，严守人工复核的最后一道防线，在探索数智化赋能的同时，确保考核结果的绝对公平与严谨。

附件：

附件 1：《商务与经济统计》全过程非标准化考核改革的评分表（院级评审）

| 一级指标 | 二级指标 | 三级指标 | 指标解读 | 院评得分 | 申报材料是否佐证 | |
|------------------|--------------------|---|--|------|----------|--|
| 考核设计(30分) | 1.1 目标适配性(10分) | 1.1.1 培养目标对齐度 | 考核任务与课程目标高度一致，覆盖知识、能力、素养三个维度，无单一维度偏废。 | 9.75 | 是 | |
| | 1.2 场景任务设计真实性(10分) | 1.2.1 场景真实性 | 跳出传统经验命题局限，至少1个任务的背景、数据、问题源于真实行业、社会或前沿案例，贴近实践。 | 5 | 是 | |
| | | 1.2.2 命题多元性 | 建立校内教师+AI+行业专家(至少1名)等多元主体协同命题机制。 | 4 | 是 | |
| | 1.3 命题高阶性(10分) | 1.3.1 任务高阶能力导向 | 记忆类题目≤30% (根据课程目标调整)，提升考查学生批判性思维、知识迁移与应用、创新实践、问题解决等高阶能力考查题目占比，问题具有灵活性、探究性和开放性。 | 10 | 是 | |
| | | 1.3.2 任务综合性与挑战度 | 题目具有整合性，难易梯度清晰，能有效区分不同能力水平学生；无超纲或无意义难题，挑战度贴合学生认知水平与课程要求。 | | 是 | |
| | 2.1 评价标准开放性(15分) | 2.1.1 指标明确性与可视度 | 无“千人一面”的标准化要求，建立分级评价量表，每个维度的指标可观察、可考量、可评价、可验证，有效降低评阅人主观偏差，避免主观臆断。 | 14 | 是 | |
| 评价实施(40分) | | 2.1.2 开放性与限制性 | 高阶试题答案具有开放性(无现成答案可找)，尊重学生思维和方法差异性，允许选择不同技术路径、研究方法或成果形式，预留创新空间；需在评价标准中明确AI工具的使用场景、使用边界及违规判定标准，避免AI抄袭。 | | 是 | |
| | | 2.1.3 标准公示及时性 | 评价标准(含评价维度、分级指标、评分权重等)在考核任务启动前向学生完整公示，明确告知“如何评、评什么”，避免学生学习方法偏差以及成果偏离要求，助力实现“评价即学习”。 | | 是 | |
| 2.2 过程培养有效性(15分) | 2.2.1 教学过程互动性与引导性 | 教学方法善于融入互动、引导探究、组织讨论，调动学生积极性，激发潜能；学习任务具有阶梯式，引导学生能力逐级提升，杜绝“教学空转、期末叠加难度”断层现象。 | 14.25 | 是 | | |
| | 2.2.2 过程支持与技术赋能 | 依托信息化平台、AI等工具，赋能学生自学自评与能力提升，通过过程跟踪、成果答辩核验等方式验证成果真实性，避免成果代做。 | | 是 | | |

| 一级指标 | 二级指标 | 三级指标 | 指标解读 | 院评得分 | 申报材料是否佐证 |
|-----------|-----------------|---------------|---|-------|----------|
| 改革成效(30分) | 2.3 评分公平性(10分) | 2.2.3 及时个性化反馈 | 在学习任务关键节点嵌入精准反馈与及时指导，而非仅给出分数或笼统评语，提供学生个体的具体改进建议和支持，解决学生能力短板。 | | 是 |
| | | 2.3.1 评分客观公信度 | 试批试评校准评分尺度；运用AI等技术工具赋能评阅，提升效率与精准度；建立健全多层次评分复核机制，学生成绩复查量低且复查结果零差错。 | | 是 |
| | | 2.3.2 多元评价融合 | 融合教师评价、学生互评、行业专家点评等多元主体；互评通过制定成果贡献度量化评分项、评价主体回避、异常值筛查等机制规则，避免“搭便车”、印象打分、成果归属或责任划分不清等现象。 | 8.5 | 是 |
| 改革成效(30分) | 3.1 学生能力提升(15分) | 3.1.1 学习主动性激发 | 学生参与过程性互动研讨、主动优化成果的比例显著提升，形成以考促学、以学促能的良性循环。 | | 是 |
| | | 3.1.2 高阶能力达成 | 以学生作答及成果为核心证据，有效印证学生在理论联系实际的实践应用、跨域知识整合、复杂问题拆解、创新方案设计与实践落地等方面成效，无死记硬背、抄袭拼凑现象及AI作弊行为。 | 14.25 | 是 |
| | 3.2 成果价值(10分) | 3.2.1 实践应用价值 | 学生成果具备明确的问题解决逻辑与实践价值，或具备落地潜力，而非单纯的任务作业。 | | 是 |
| | | 3.2.2 创新性与个性化 | 成果体现学生独特思考或个性化表达，具有原创突破，无同质化现象。 | 9.25 | 是 |
| | 3.3 持续改进(5分) | 3.3.1 教与学优化联动 | 学生满意度不低于80%，学生自我评估良好；改革有效解决原有考核问题，并基于考核结果调整课程教学策略，实现以考促教的闭环。 | 4.75 | 是 |
| 合计得分： | | | | 93.75 | - |

说明：1. 总分采用100分制，按二级指标细化评分，求和计算最终得分。
 2. 自评与评审给分：二级指标按三档评分——“达标”（满分）：完全符合指标要求，核心量化指标达标，无任何偏差、“基本达标”（满分的60%—80%）：符合指标核心要求，次要量化指标存在轻微偏差（不超过10%），无实质性影响；“不达标”（0—59%）：未满足指标核心要求，或触发相关否决项关联条件。

3. 核心否决项：若存在以下情形之一，直接判定为“不达标”（分数不高于59分）：

- ①命题开放性不足（如记忆类题目占比超过30%且未设置开放性或高阶考查题目）；
- ②评价标准缺失，或标准表述模糊、无明确评判依据、无高阶能力评价、不可操作；
- ③过程评价缺乏有效支撑材料证明过程性培养学生能力发展；
- ④学生成果存在抄袭、拼凑行为，或经核查确认存在AI作弊情况。