**2022全国应用统计专业学位研究生案例大赛企业选题**

**请根据所选选题自拟参赛题目**

**统计案例题目1**

  在当前全球能源变革中，新能源作为石油类能源的替代者，无论是在应对气候危机，碳中和及能源再生等问题上，都有战略地位。作为新能源之一的锂电池技术，业界一直关注电池产品可靠性问题。目前认为影响锂电池寿命及产品可靠性的因素主要有:充放电次数、温度、放电深度和不同材料等。请根据“**附件1”**所提供的数据集，按照统计案例写作要求，解决如下问题：

 1．根据所给信息和数据，包括不同电池组件、4个不同加速温度设置、2000个充放电循环，以及相应记录的电芯剩余容量，请建立非线性参数统计分析模型并进行参数估计，预测该电池产品在高温50度情况下，50%产品失效的时间点估计和区间估计（电池剩余容量<75%时，视为产品失效）。

 2．根据三种不同材料的充放电后的失效循环次数，判定哪个材料可靠性更好，并给出具体理由。

**统计案例题目2**

  单细胞RNA测序能帮助基因学者更好地探索单细胞水平的基因表达谱，了解在不同组织微环境下细胞表达及功能差异，也为发现罕有细胞，细胞谱系分化轨迹提供了前所未有的机会。目前，单细胞RNA测序已成为研究细胞异质性的关键生物学问题的首选。该类技术的进步为单细胞RNA数据分析提供大量数据的同时，也提出了数据分析上挑战，包括如何根据测序信息进行准确的细胞分类，并找到细胞差异化表达的基因等。请根据“附件2”所提供的数据集，按照统计案例写作要求，解决如下问题：

1．对数据进行清理和预处理，使用数据分析方法（如Sparse SVD，Clustering等）对采集的细胞进行分类。

2．在细胞分类后，使用统计分析方法（如Mixed Model ANOVA），找到显著差异化表达的基因。

**统计案例题目3**

  随着各大电商平台不断发展，其行业竞争也日益激烈。针对平台订单进行数据分析，可以提高客户忠诚度、稳定客流，不断调整运营方向。这类分析工作往往专注于解决以下几个问题：与前一个季度/年度相比销售情况变化如何，如果下降、是什么原因导致的？用户消费行为有怎样的特征，高净值客户的行为是否有据可循？如何基于“商品&客户”之间的关系进行个性化的销售预测，从而开发推荐系统等，请根据“附件3”所提供的数据集，按照案例写作要求，解决如下问题：

1． 针对商品销售情况，设计数据分析、特征提取的流程，使用聚类等方法（如k-means），从“是否畅销”、“是否具备高盈利能力”、“是否退货率较高”等方面评估商品商业价值。

2. 根据帕累托法则，即：20%的客户购买了80%的产品/贡献了80%的收入。企业中往往称这20%能够推动公司80%的收入的客户为“高价值客户”。请从客户留存（留存的衡量方式包括但不限于客户下单时间间隔、活跃时长）、消费金额与退单率等方向，设计分析筛选高价值客户的流程与方法。

3. 电商平台常见的“推荐系统”是一组根据从用户购买（/收藏/浏览…）历史向用户推荐项目的算法，其算法逻辑：根据客户购买历史寻找到与他最为相似的N个客户，依据他们的购买记录计算相似度分数，得分最高的M个商品将会被推荐给原客户。附件3中数据文件“ecommercedata-预测.xlsx”中包含第13个月下单的客户。请根据他们此前一年的购买历史，预测他们该月有可能购买的产品并评估算法的准确性。