
作品题目：肥料登记数据分析

摘要

肥料是农业生产的一种重要的生产资料，其生产销售必须遵循《肥料登记管理办法》，依法在农业行政管理部门进行登记。近年来，随着我国肥料产业的飞速发展，国内肥料产业的问题逐渐暴露，肥料市场诚信缺乏，坑农伤农时间屡屡发生。由于，肥料的生产领域问题突出，导致不法商家利用未经处理的工业废弃物矿石、矿渣等对土壤有害的物质生产肥料，对于土壤造成不同程度的污染和损坏，肥料的不合理施用造成土壤生产能力的下降，目前，肥料存在严重的名称不规范不科学，鱼目混珠，使农民无法辨别真伪，销售渠道不畅，价格混乱，虚假宣传误导消费，个别商贩暴力坑农，使农产品生产成本上升。因此，《肥料登记管理办法》在加强肥料管理，保护生态环境，保护人畜安全，促进农业生产等方面具有重大意义，同时利用数据分析技术对肥料登记数据进行多维度数据分析为有关部门对肥料的技术审查、质量检测和安全性评价试验等提供的极大便利。

目录

1、 问题分析.....	3
2、任务 1 数据的预处理	4
2.1 附件 1 产品通用名称不规范处理	4
2.1.1 产品通用名称不规范情况说明:	4
2.1.2 产品通用名称不规范数据处理过程描述.....	4
2.2 计算附件 1 中各肥料产品中氮、磷、钾养分百分比之和，称为总无机养分百分比	4
3、任务 2 肥料产品的数据分析	4
3.1 任务 2.1	4
3.1.1 从附件 2 中筛选出复混肥料的产品	4
3.1.2 将所有的复混肥料按照总无机养分百分比和有机质百分比等距分为 10 组并打上分组标签	4
3.1.3 绘制产品登记数量的直方图	4
3.1.4 分析复混合肥料的分布特点	5
3.1.5 列出登记数量最大的前三组	5
3.3 任务 2.3	6
3.3.1 从附件 2 中筛选出复混肥料的产品，按照氮、磷、钾养分的百分比，使用聚类算法将这些产品分为 4 类，根据聚类结果为每个产品打上聚类标签（标签用 1~4 表示）	6

图表目录

图表 2 产品登记数量的直方图.....	5
图表 3 task2_3 雷达图.....	6

1、问题分析

1. 对肥料登记数据进行预处理。
2. 根据养分的百分比对肥料产品进行细分。
3. 从省份、日期、生产商、肥料构成等维度对肥料登记数据进行对比分析。
4. 对非结构化数据进行结构化处理。

2、任务 1 数据的预处理

2.1 附件 1 产品通用名称不规范处理

2.1.1 产品通用名称不规范情况说明：

2.1.2 产品通用名称不规范数据处理过程描述

2.2 计算附件 1 中各肥料产品中氮、磷、钾养分百分比之和，称为总无机养分百分比

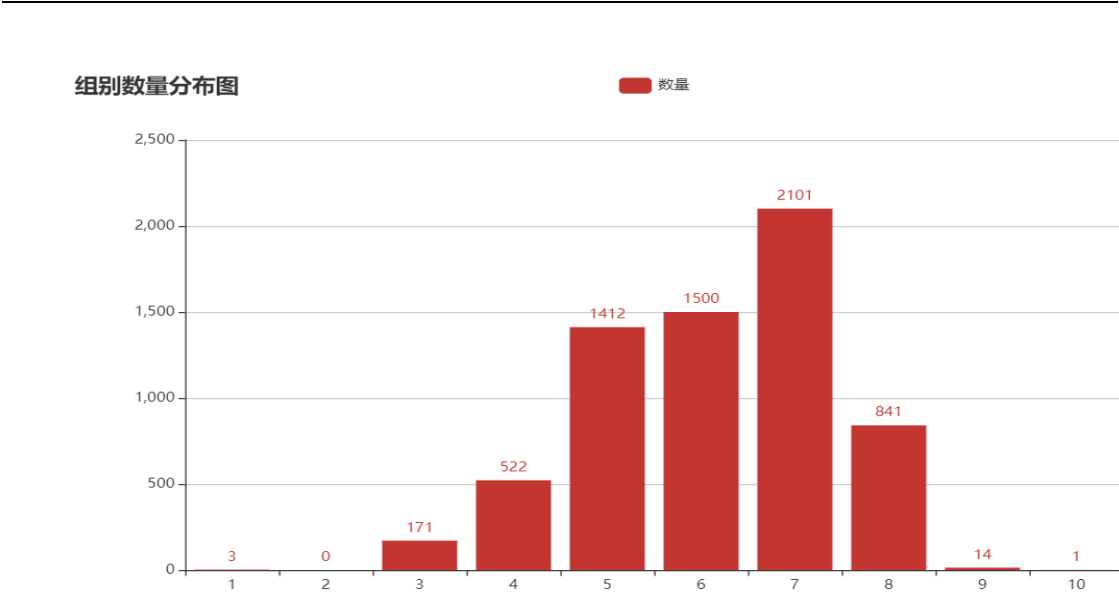
3、任务 2 肥料产品的数据分析

3.1 任务 2.1

3.1.1 从附件 2 中筛选出复混肥料的产品

3.1.2 将所有的复混肥料按照总无机养分百分比和有机质百分比等距分为 10 组并打上分组标签

3.1.3 绘制产品登记数量的直方图



图表 1 产品登记数量的直方图

3.1.4 分析复混合肥料的分布特点

由图表 17 可知复混合肥料主要集中在 3-8 组，复混合肥料产品登记数量主要集中在 3-8 组。其中组别数量最多的是 7 组有 2101，其次是 5 组和 6 组数量都接近 1500，然而 1 组、2 组、10 组的数量非常之少几乎没有。

因此，可以看出复混合肥料总无机养分百分比主要集中在 0.1442 到 0.5768 之间。

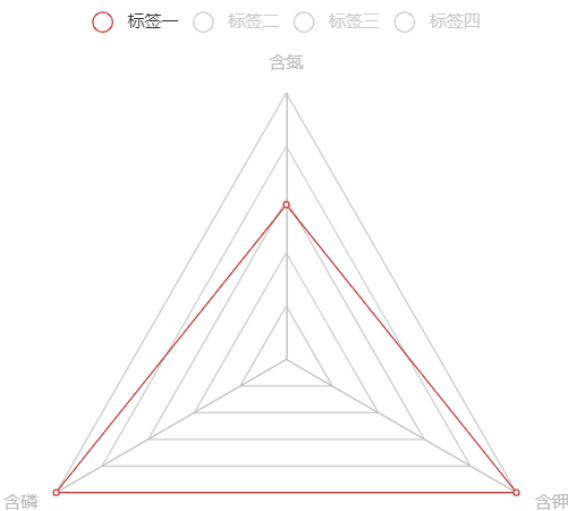
3.1.5 列出登记数量最大的前三组

排名	一	二	三
分组标签	7	6	5
产品登记数量	2101	1500	1412

3.3 任务 2.3

3.3.1 从附件 2 中筛选出复混肥料的产品，按照氮、磷、钾养分的百分比，使用聚类算法将这些产品分为 4 类，根据聚类结果为每个产品打上聚类标签（标签用 1~4 表示）

task2_3雷达图



图表 2 task2_3 雷达图

Task2_3 雷达图的聚类特征：

第一聚类的特征是含磷和钾较高，含氮量比价少。

第二聚类的特征含氮最高，磷 次之，钾与氮磷相比较少。

第三聚类的特征含磷量比含氮量要高，含钾量最低。

第四聚类的特征含磷较高，钾次之，含氮量最少。