

2024 年（第 7 届）“泰迪杯”数据分析技能赛

A 题 自动化生产线数据分析

一、背景

随着信息技术的快速发展，工业自动化领域的智能控制系统日益完善。自动化生产线能够独立完成从物料输送到元件抓取，再到产品安装和质量检验的各个环节，这不仅极大提升了制造效率和产品质量，也有效降低了生产成本。

为了使企业的管理层和工程技术人员及时掌握生产线的运行状况和潜在问题，需要对生产线产生的数据进行深入分析，从而提高生产线的生产效率，减少产品的次品率。

二、目标

- (1) 数据整理与统计。
- (2) 生产线运行情况的可视化分析。
- (3) 生产线影响因素分析。

三、任务

附件 1 中的文件 M101.csv、M102.csv 分别记录了某生产企业两条生产线全年加工处理过程中各个工序的生产线数据。请编程完成以下任务，根据附件 2 中的模板文件保存结果，并撰写报告，在报告中详细描述各项任务的处理思路、过程及必要的结果。所有实数类型的结果保留两位小数。

任务 1 数据整理与统计

任务 1.1 分别统计两条生产线每天的产品总数（包含不合格产品）、合格产品数、不合格产品数与合格率，将结果分别存放到 result1\_1.xlsx 的 M101 和 M102 工作表中（格式见表 1），并在报告中给出各生产线全年的产品总数、合格产品数、不合格产品数与合格率（格式见表 2）。

表 1

月份	日期	产品总数（件）	合格产品数（件）	不合格产品数（件）	合格率（%）
1	1				
1	2				
...	...				

表 2

生产线	产品总数（件）	合格产品数（件）	不合格产品数（件）	合格率（%）
M101				
M102				

任务 1.2 分别列出两条生产线每次故障的相关信息，按照月份、日期和开始时间升序排列，将结果分别存放到 result1\_2.xlsx 的 M101 和 M102 工作表中（格式见表 3），并在报告中给出各生产线每种故障一年中第 25 次发生的相关信息（格式见表 4）。

表 3

月份	日期	开始时间	故障类别	持续时长（秒）

表 4

生产线	故障类别	月份	日期	开始时间	持续时长（秒）
M101	A1				
M101	A2				
...	...	...	...	...	...
M102	A1				
...	...	...	...	...	...

**任务 1.3** 根据任务 1.2 的结果，分别统计两条生产线各类故障每天发生的总次数和平均持续时长，按照生产线、月份、日期、故障类别升序排列，将结果存放到 result1\_3.xlsx 中（格式见表 5；如果某天没有指定类别的故障，则该天该故障的总次数为 0，平均持续时长为 Null）。在报告中分别给出两条生产线各类故障发生的总次数、平均持续时长、故障发生频率，并进行汇总（格式见表 6）。

表 5

生产线	月份	日期	故障类别	总次数	平均持续时长（秒）
M101	1	1	A1	.....	.....
M101	1	1	A2	.....	.....
M101	1	1	A3	.....	.....
M101	1	1	A4	.....	.....
M101	1	2	A1	.....	.....
.....	.....	.....	.....	.....	.....
M101	12	31	A4	.....	.....
M102	1	1	A1	.....	.....
.....	.....	.....	.....	.....	.....

表 6

生产线 M101	故障类别 A1	故障类别 A2	...	汇总
总次数				
平均持续时长（秒/次）				
发生频率（次/天）				

**任务 1.4** 分别统计两条生产线每天的有效工作时长，按照月份、日期升序排列，将结果存放到 result1\_4.xlsx 中（格式见表 7），并在报告中给出各生产线的日平均有效工作时长（格式见表 8）。有效工作时长是当天开机时长减去因故障停机的时长。

表 7

月份	日期	M101（小时）	M102（小时）
1	1		
1	2		
...	...	...	

表 8

生产线	日平均有效工作时长（小时/天）
M101	
M102	

**任务 1.5** 对生产线 M101 每天推出的电路板数量、抓取的元件数量与抓取的故障次数做相关性分析，并结合问题背景解释所得到的结果。

**任务 2** 生产线运行情况的可视化分析

**任务 2.1** 根据各月的产品总数（包含不合格产品），以月份为横坐标，绘制两条生产线的堆叠柱状图。

**任务 2.2** 根据生产线每天的不合格产品数（取值范围为[0,35]）和不合格率（取值范围为[0,0.9%]），分别绘制两条生产线的双 Y 轴折线图。

**任务 2.3** 根据不同故障类别的全年发生总次数，绘制两条生产线各故障类别的占比双层环形图。

**任务 2.4** 根据不同故障类别，以持续时长（单位：秒）为横坐标，分别绘制两条生产线全年故障发生持续时长的叠加直方图。

**任务 2.5** 根据 4 月 26 日前 100 秒生产线数据，以时间（单位：秒）为横坐标，电路板推出次序为纵坐标，绘制生产线 M101 包含 4 个工序的甘特图。

**任务 3** 生产线影响因素分析

根据生产线数据，应用数据分析方法分析影响产品合格率和生产线产量的可能因素。

**四、数据说明**

**附件 1** 生产线数据

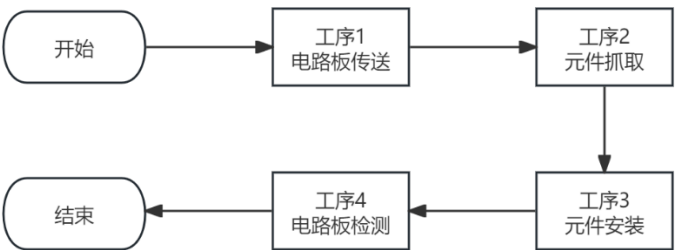
M101.csv、M102.csv：两条生产线一年的运行记录数据。

数据字段说明.xlsx：生产线数据中各字段的说明。

**附件 2** 结果的模板文件

**附录 生产线运行流程**

生产线由电路板传送、元件抓取、元件安装、电路板检测 4 道工序组成，运行流程如图 1 所示。



**图 1 生产线运行流程图**