

## Project2 方法文档

**说在前面：**本方法文档仅提供参考方法，各组方法可以与本文档描述有所不同，如有不同请在报告或者展示的时候进行解释，鼓励发散性思维。本文档由助教刘雨桐撰写，如有问题欢迎讨论，祝各位期末顺利。

### 任务一：

1. 十年的气象统计对应十张表，每个表中有多个子表，第一张子表是引航中断的情况统计，其中只需要考虑大风管制的统计即可。大风管制的统计表述为日期、管制次数、管制小时和对应的时间节点（2300 表示的是 23:00 即晚 11 点，如中断时间为 0500-0900 即对应天气记录里的“上午”）。则可以根据这个时间节点，找到后面子表中对应日期的对应时间段的天气预报，优先查找舟山沿海海面的天气，其次查找长江口区天气，如果这两个都没有对应的记录，可以用全市天气暂代。天气包含风级和风向，先把原对应时间段的风级和风向摘出来，有多条记录也都摘出来，然后再自行设计归纳的规则，规则不限，有理即可。

**提交要求：**本小题需要在报告中文字说明你们小组最终设计的数据处理具体规则，用可视化图表展示统计结果，并额外附上统计出来的大风管制日风向和风级表（.xlsx）。关于设计规则时候的思考过程，不强制要求写，如果对答案不够自信，可以附上，以便助教批改时更好地理解你们的想法。

**注意：**查找数据和提取数据的过程不强制要求使用 python，可以采用任意方法，甚至手工暴力搜索。

2. 根据第一问的数据抓取，可以看到引航中断往往是那么一些特定范围内的风向和风级的时候会发生，这个范围就是本题要考察的“阈值”。可以分别阐述，即风向在什么范围内时会触发引航中断，或风级在什么范围内时会触发。也可以共同阐述，即风向和风级一起在什么范围内时会触发。

**提交要求：**本小题需要在报告中文字说明你们小组最终得到的范围模型。关于得到范围时候的思考过程，不强制要求写，如果对答案不够自信，可以附上，以便助教批改时更好地理解你们的想法。

**注意：**本题仍然不强制要求使用 python，可以采用任意方法，具体采用了什么方法不做考察，主要看结果，其次是思考过程。

### 任务二：

1. 文件夹“任务二第一问-引航中断观测线 6 数据-xx”是这道题的分析目标，其中包含了五份警戒区船舶通行的详细统计数据，**每一个的文件名后面的时间段表述的就是引航中断的时间段**。可以发现，文件中包含引航中断时间内的统计，也有其他时间内的统计（可以认为是正常通航的统计），这个请查看“穿越时间”字段就会发现。这样首先就可以把一份数据集分为正常和中断两个情况来讨论。本题需要讨论的是东西向和南北向的流量，这里的东西和南北可以根据“船舶航向”字段进行判定，按 360 度思考，0-45, 135-225, 315-360 三个角度范围可以算作南北向，其余算作东西向（根据下面的航向图），这样可以统计出，正常或引航中断条件下，东西向和南北向各自的记录数量，即为流量。另外注意，如果航速为 0，不能算进流量中，因为压根就没有流动起来。根据最终得到的流量结果，可以自行做等级划分，划分标准不限，言之有理即可。



**提交要求:** 本小题需要在报告中文字说明 1) 正常和中断情况下的流量区别, 可以用具体数据来表示; 2) 你们划分中断等级的指标, 即什么等级对应什么范围的流量。本题不强制要求可视化和方法表述, 但建议这么做, 显示你们真的有自己在思考做法, 做法也可以不完全按照我的思路。

**注意:** 同样地, 本题也不要求 python, 任意方法即可, 答案不唯一。

2. 文件夹“任务二第二、三问-观测线数据-xx”是这道题的分析目标。对每一个子文件（每一条观测线）先找出集装箱船, 然后可以把他重新分成 9 个模块分别对应题目中所说的 9 种船长范围, 划分依据自然就是“船舶长度”字段。这样可以统计出每个模块里的记录条数（通行量）、航速范围（航速规律）、通行时间范围（通行时间规律）, 然后用分别的通行量除以总通行量就是通行量占比。

**提交要求:** 本小题需要在报告中文字说明每条观测线对应的每个船长范围下的通行量占比、航速规律和时间规律。可视化和方法描述不强制但建议这么做, 因为独立的数据并没有什么意义, 放在一起看反而能观察到一些可用于实际的规律, 但不这么做并不扣分。

**注意:** 抓数据和抓规律的方法不限制, 不强制 python, 任意均可。这 9 个模块的划分不强制要求, 只是这么做之后做规律统计的时候比较方便, 也可以作为附件提交上来, 不交不扣分。

3. 文件夹“任务二第二、三问-观测线数据-xx”是这道题的分析目标。本题虽然说要归纳同一条船舶的航行记录, 但是有些船舶只有一条记录甚至很少的记录, 因此本题只需要这么做:

- 第一, 保留上一问中每条观测线中集装箱船按船长分出来的 9 个类。
- 第二, 将所有观测线按类合并数据集, 也就是说 9 个类里每个类包含对应船长范围在所有观测线上的记录。
- 第三, 每个类里寻找有多条正常航行记录的船, 保留这些记录即可。
- 第四, 第一问就可以做了, 针对这样提取出来的记录, 对每条船给出其在警戒区中估计航速和时间。
- 第五, 第二问接着做, 对上一步中整理出来的结果按 9 个船长类做汇总, 得到一个总的通行规律即可。

**提交要求:** 本小题需要再报告中文字说明有多条航行记录的船舶其对应的估计航速和时间, 再说按类别分之后各个类的估计航速和时间（归纳的结果）。同样的, 可视化和方法描述不强制但可以用来补充解释。

**注意:** 同样地, 不强制 python, 任意方法均可。