

实验报告

目录：

目录： 2

实验目的： 3

实验过程： 3

 第一部分： 3

 第二部分： 4

 第三部分： 4

 杂项 5

误差分析 5

结语 6

实验目的：

本次实验共三个部分，第一部分是关于分析个专业的冷门、热门情况，第二部分是对录取分数线补充其对应的排名，第三部分是依据往年录取情况，对今年的考生提供志愿建议。

实验过程：

第一部分：

为了分析各个专业的冷门热门情况，我给出了两种指标的混合作为标准：

第一个的指标是源代码中的 **quota1**，其表达式为：

$$quota1 = \frac{1}{\ln\left(\frac{\text{录取最低排位} - \text{录取均值排名}}{\text{录取人数}}\right)}$$

具体实现为：

```
quota1 = 1.0 / log(_X: (it.avgMinRank - it.avgAvgRank) / enroll);
```

这一指标是基于录取人数和对应分段之间总的考试人数。不难看出，专业越热门，其录取的人数在对应分段中录取比例最大，最后得到的 **quota1** 的值就越大。

第二个指标是源代码中的 **quota2**，其表达式为：

$$quota2 = \frac{500}{\text{录取平均排位} - 500}$$

对应的实现为：

```
quota2 = 500 / (it.avgAvgRank - 500);
```

这一指标大体上是基于分数的高低，录取的平均分越高则得到的 **quota2** 就越大，而平均分的高低一定程度上也体现了专业的热门程度。

最后将两者相加即得到关于热门程度的指标，记录在专业的结构体中。再按

照该指标使用 STL 的 sort 函数即可排出最热门和冷门的十个专业。

第二部分：

这一部分只要获取各专业的分数对应的排位即可，所以只要通过 fstream 读取然后保存在 vector 中，再用录取平均分和最低分在该 vector 中查找即可。当然，这里为了查找方便，重载了分数储存容器 **score** 和 **int** 之间的相等运算符 **==**。

第三部分：

这一部分是本次实验的重头戏，在实现的过程中，我认为个专业的最终分数应当服从正态分布。显然，这一正态分布的参数应由该专业组包含的专业加权生成。考虑到更热门的专业会对大类的分数产生更大的影响，因此我选择了第一部分中生成的热度值作为权值。根据正态分布的有关公式：

$$\begin{aligned} X &= \sum_{i=1}^n a_n X_n \\ \mu &= \sum_{i=1}^n a_n \mu_n \\ \sigma^2 &= \sum_{i=1}^n a_n^2 \sigma_n^2 \end{aligned}$$

即可得到对应专业组的分布 $N(\eta, \sigma^2)$ 。最后输入分数即可通过正态分布的分布函数 (cdf) 公式： $F(x) = \frac{1}{2} \operatorname{erfc}(\frac{\eta-x}{\sqrt{2}\sigma})$ 得到分数线的分布，又因为我们希望分数比录取分数线更高，所以 $1-F(x)$ 即为所求概率。为了提高结果的直观性，我人为将概率划为三组，包括 0-20%的低概率（不显示）、20%-50%可冲刺、50%-80%较稳定、80%-100%可用作保底。除此之外，对于小于 800 或大于 10000 的排名，再做概率分析意义不大，所以我作为异常直接给出结果。

杂项

为了提高软件的易用性，我也使用了一些有效的交互，一定程度上也提高的美观性，同时 STL 的使用也精简了代码，提高了可读性。

误差分析

对于第一部分，主要的误差来源是指标的不准确性，以及两指标的系数关系，我在本次实验中仅给出了我认为合适的加权方式，所以或许还有更合适的加权方式。

可冲刺的专业组：	
专业名称	录取概率
经济学类（经管实验班）	42.4%
计算机类（广州-大数据实验班）	47.6%
较稳的专业组：	
专业名称	录取概率
计算机类（珠海-智能实验班）	58.5%
计算机类（深圳-新工科实验班）	58.9%
物理学类（物理实验班）	77.6%
可保底的专业组：	
专业名称	录取概率
数学类（数学实验班）	98.1%
生物科学类（广州-理工实验班）	98.9%

2500 名的推荐结果

可冲刺的专业组：	
专业名称	录取概率
经济学类（经管实验班）	34.2%
计算机类（广州-大数据实验班）	40.5%
数学类（数学实验班）	46.2%
较稳的专业组：	
专业名称	录取概率
计算机类（珠海-智能实验班）	51.4%
计算机类（深圳-新工科实验班）	51.9%
物理学类（物理实验班）	73%
可保底的专业组：	
专业名称	录取概率
生物科学类（广州-理工实验班）	97.5%

3000 名的推荐结果

对于第二部分，因为是直接查表的结果，所以主要误差是最低录取分数并不是同分中的最后一名。但是由于中大官网上没有更多信息，所以这一误差也无法弥补。

对于第三部分，因为涉及随机变量的估计，所以模型的准确性、权值的选取、样本量过小等因为都成为了导致误差的因素。而受限于样本量的大小，也不方便做假设模型的检验，所以可以会导致一定的误差。但是显然，对于软件的使用者

这一概率的大小比较比其本身的数轴大小更有意义。事实上, 从直观的角度上看, 这一模型还是有相当好的效率的。例如对于一考生:

假如他的排名为 2500, 那么他可以以几乎 100%的概率被数学类录取, 而当其排名达到 3000 时, 则被数学类录取的概率急剧下降, 而其他专业组的下降速率则没有这么快, 这体现了数学类历年专业录取的最低排位较其他专业要稳定得多。所以直观上, 这一推荐是合理的。

结语

本次实验考察了本人对排序和查找的运用能力, 是一次非常有意义的实验。在实验过程中也锻炼了个人的编程能力, 并逐渐适应 STL 的使用。希望以后还有更多机会提升自我。