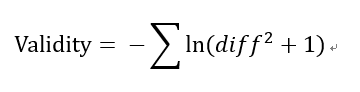
## 合理性评估

经过我们解第二题和第三题的经验，分组越少，每个评委评的论文越多，结果越合理。而题目的评审过程的步骤三和步骤四中，都有超过一半的评委对每个小组的论文打分，同时对于每个打分区间都可以看作是在同一个组内打分，因此步骤三和步骤四设置得较为合理。而步骤一和步骤二由于分组较多，评委评判文章较少，因此其合理性尚待验证。为此，我们采用排名差值的方法来评估步骤一和步骤二的合理性

### 排名差值

由于我们在问题三中的调分算法能很好地确保结果的合理性，因此在该模型中，我们以我们在问题三中提出的调分算法为基线，评估题目中原本的评分步骤。

对于相同的排名，原步骤的小组排名应与基线的小组排名相差越少越好，同时在基线中出现的小组要尽可能在原步骤的小组排名中出现。因此我们采用排名差值的方法来量化合理性，具体公式如下



### 求解步骤

#### 步骤一合理性验证

经过检验，按照原题的方式完成步骤一后，淘汰的小组排名与我们的方法淘汰的小组排名完全一致

#### 步骤二合理性验证

步骤二中，首先需要淘汰20个小组，原题方法淘汰的小组及其对应打分如下：

| 小组编号 |
| --- |
| 38 |
| 193 |
| 31 |
| 200 |
| 33 |
| 60 |
| 11 |
| 52 |
| 191 |
| 77 |
| 73 |
| 160 |
| 4 |
| 45 |
| 59 |
| 143 |
| 20 |
| 55 |
| 22 |
| 107 |

将该排名与我们的调分算法调整后的排名进行对比，得到差值情况如下：

| 小组编号 | 排名之差 |
| --- | --- |
| 38 | 0 |
| 193 | 2 |
| 31 | -1 |
| 200 | 未出现 |
| 33 | -2 |
| 60 | 2 |
| 11 | 2 |
| 52 | 6 |
| 191 | 4 |
| 77 | -3 |
| 73 | -1 |
| 160 | 4 |
| 4 | -8 |
| 45 | -2 |
| 59 | 未出现 |
| 143 | -10 |
| 20 | 0 |
| 55 | 未出现 |
| 22 | -8 |
| 107 | 未出现 |

使用评估公式计算得分为-33.98，显然步骤二是不合理的。

## 工作量评估

经过计算，原题打分标准下，平均每个评委至少需要给94篇论文打分，相较于原本每个评委给200篇论文打分的方法减少了一半，说明评委的工作量被原题的打分流程显著降低了。

## 改进的评分步骤

步骤（1）

打分：每个组的10篇论文由3位评委评审，分别用百分制给出评分。评委的分配方式采取问题一中第三小问所得的最优分配方案。

排序：每个组的10篇论文根据3位评委的给分进行平均，总的200篇论文排名后，淘汰后80篇。

步骤（2）

打分：未淘汰的120篇论文，再由没有评审过的2位评委进行评审（评审时采取圆桌模式），给出百分制得分。

排序：对120篇论文根据5个评委的平均给分进行总体排序，淘汰排名靠后的30篇文章，剩下的90篇论文为获奖论文。

90篇获奖论文中，排名31到40的论文，获二等奖；排名61到90的论文获三等奖；

步骤（3）

打分：排名1到30与排名41到60的50篇论文，再由2位评委进行打分（采取评委随机选论文的模式）

排序：原排名41到60的20篇论文，7位评委给分平均，新排名41-50的获二等奖，新排名51-60的获三等奖；

原排名1到30的30篇论文，7位评委给分平均，新排名26-30名获二等奖，新排名1-15名获一等奖。

步骤（4）

打分：7人平均分排名在16到25的论文，再由剩下的3位评委打分。

排序：最终排名16-20的获一等奖，最终排名21-25的获二等奖。