大学生科创项目统计

胡华平

2021-05-11

# 1 清洗数据

# 读取数据  
df <- openxlsx::read.xlsx("../data/training-proj-2017-2020.xlsx")  
  
# 清洗数据、计算得分  
df\_tidy <- df %>%  
 rename("负责人0" = "负责人学号") %>%  
 gather(key = "menmber", value = "id\_menmber", `负责人0`,`成员1`:`成员5`) %>%  
 arrange(desc(`立项年份`), `项目编号`) %>%  
 filter(!is.na(id\_menmber)) %>%  
 mutate(order\_menmber = as.numeric(str\_extract(menmber, "\\d{1}")),  
 score\_minus = order\_menmber \*0.5,  
 score\_base = ifelse(`项目等级`=="国家级", 3,  
 ifelse(`项目等级`=="省级", 2.5, 2)),  
 proj\_status = ifelse(is.na(`项目状态`), "未知", `项目状态`),  
 score\_status = ifelse(str\_detect(proj\_status,"中止"), 0, 1)) %>%  
 filter(score\_status ==1) %>%  
 mutate(score\_get = score\_status\*(score\_base - score\_minus))   
   
# 计算最高得分（不重复、不累加）  
df\_smry <- df\_tidy %>%  
 group\_by(id\_menmber) %>%  
 summarise(score\_max = max(score\_get)) %>%  
 mutate(score\_final = ifelse(score\_max < 1, 1, score\_max)) %>%  
 mutate(status\_ok = ifelse(score\_final >=2, "OK", "fail"),  
 ID = as.numeric(str\_extract(id\_menmber, "\\d{4}")))

# 2 计算步骤和计算公式

计算步骤：

1). 分析参与排序order\_menmber：主持人（0）；成员1（1）；成员2（2）；…

1). 计算基础得分score\_base：国家级3分；省级2.5分；校级2分。

1). 计算递减得分score\_minus：=0.5\*order\_menmber。参与排序乘以0.5的递减分数。

1). 分析项目状态score\_status：中止=0， 其他（结题、在研…）=1

1). 计算项目得分score\_get：=score\_status\*(score\_base - score\_minus )

1). 计算保底得分score\_final：如果项目得分<0.5，则记为1分。

1). 计算每个学生的最高得分score\_max：根据学号，用max()函数找到所有同学的最高得分。

项目得分的计算公式：

# 3 统计分析结果

下面表3.1给出了2017级学生科创项目参与不同**得分**水平下的人次和百分比情况。

Table 3.1: 2017级学生科创项目参与得分的人次和百分比

|  |  |  |
| --- | --- | --- |
| 得分 | 人数 | 百分比 |
| 1 | 992 | 32.4% |
| 1.5 | 672 | 22.0% |
| 2 | 842 | 27.5% |
| 2.5 | 421 | 13.8% |
| 3 | 131 | 4.3% |
| Total | 3058 | 100.0% |

**说明**： 1)项目状态为“中止”的学生没有统计在内（根据规定，主持人和参与人都不能获得分数）。 2)学号异常的学生没有包括在内（可能是访学学生）。

下面表3.2给出了2017级学生科创项目参与**得分状态**的人次和百分比情况。

Table 3.2: 2017级学生科创项目参与得分状态的人次和百分比

|  |  |  |
| --- | --- | --- |
| 得分状态 | 人数 | 百分比 |
| 小于2分 | 1664 | 54.4% |
| 2分及以上 | 1394 | 45.6% |
| 合计 | 3058 | 100.0% |