

# 计算偏相关性(Partial Correlation)

统计学中使用皮尔斯相关系数来度量两个变量的相关性. 但在有些情况下, 我们需要控制第三个变量的变化(排除第三个变量的影响), 来研究前两个变量的关系, 这就是偏相关性要解决的问题

例如:我们想要知道学生学习时间长短和考试成绩之间的关系, 因为不同学生在接受实验之前的起点成绩就不同, 所以需要排除当前成绩对实验的影响

```
• begin
•   using RCall
•   using StatsBase, DataFrames
• end
```

	currentGrade	hours	examScore
1	82	4	88
2	88	3	85
3	75	6	76
4	74	5	70
5	93	4	92
6	97	5	94
7	83	8	89
8	90	7	85
9	90	4	90
10	80	6	93

```
• begin
•   currentGrade = [82, 88, 75, 74, 93, 97, 83, 90, 90, 80]
•   hours = [4, 3, 6, 5, 4, 5, 8, 7, 4, 6]
•   examScore = [88, 85, 76, 70, 92, 94, 89, 85, 90, 93]
•   df=DataFrame(currentGrade=currentGrade, hours=hours, examScore=examScore)
• end
```

0.19062580099310358

- `partialcor(df.hours, df.examScore, df.currentGrade)` #固定当前成绩后计算学习时间和考试成绩相关性

0.7355672586526147

- `partialcor(df.examScore, df.currentGrade, df.hours)` #固定学习时间, 考察当前成绩和考试成绩的相关性

下面使用 R 软件包 *ppcor* 获取更多的信息

```
RObject{VecSxp}
$estimate
      currentGrade      hours examScore
currentGrade  1.0000000 -0.3112341 0.7355673
hours        -0.3112341  1.0000000 0.1906258
examScore     0.7355673  0.1906258 1.0000000

$p.value
      currentGrade      hours examScore
currentGrade  0.00000000 0.4149353 0.02389896
hours         0.41493532 0.0000000 0.62322848
examScore     0.02389896 0.6232285 0.00000000

$statistic
      currentGrade      hours examScore
currentGrade  0.00000000 -0.8664833 2.8727185
hours        -0.8664833  0.0000000 0.5137696
examScore     2.8727185  0.5137696 0.00000000

$n
[1] 10

$gp
[1] 1

$method
[1] "pearson"
```

- `begin`
- `@rput df`
- `R"""`
- `library(ppcor)`
- `res= pcor(df)`
- `"""`
- `#@rget res`
- `end`

*ppcor* 软件提供了统计显著性信息.

1. 固定当前成绩, 学习时间和最后考试成绩的相关性  $r = 0.190$ , 为弱正相关性, 随着学习时间增加, 考试成绩会少量提高.

$p - value = 0.62 > 0.05$ , 所以在  $\alpha = 0.05$  水平没有统计显著性

2. 如果学习时间固定, 当前成绩和最后的考试结果呈现强的正相关, 当前成绩越高, 最后的考试成绩越好.  $r = 0.735$ ,  $p - value = 0.002 < 0.05$ , 所以在  $\alpha = 0.05$  水平有统计显著性