

音乐对工作效率的方差分析

研究问题和假设

本文研究提供一些员工要求播放的背景音乐是否会提高生产力。

因此，这里收集数据，其中记录了150名员工。这150名员工被随机分为三组，每组50名参与者：(a)“对照组”不听音乐；(b)第一个“治疗组A”听音乐，但不能选择听什么；和(c)第二个“治疗组B”，不仅可以听音乐，还可以选择听什么音乐。

三组的“生产力”根据“每小时处理的平均包裹数量”来衡量。

本文假设，适当的娱乐放松会提高生产力，处于可选择音乐工作状态下的员工的生产力水平最高，其次是不可选择音乐工作状态下，最后是无音乐组。

研究问题

1. 这三组之间的生产力是否存在统计学上的显著差异？
2. 两两看来，组之间的生产力是否有区别？哪个组具有最高的生产力水平

选择的数据分析方法和理由

选择方法

这里采用方差分析作为组件生产力差异的统计分析的方法

选择理由

方差分析 (Analysis of Variance, 简称ANOVA) 是一种统计方法，是对离散的影响因素对多个总体的影响分析，它用于**比较两个或多个组或处理之间的均值差异是否显著**。方差分析可以帮助我们确定因素对于观察结果的影响，以及它们之间是否存在统计学上的显著差异。

方差分析的基本思想是将总体方差分解为不同来源的变异成分，例如组内变异和组间变异。组内变异是由于不同观察值之间的个体差异导致的，而组间变异则是由于所研究的因素（例如不同的处理或组别）导致的。

在本文的研究中，存在三个组(对照组、治疗组A、治疗组B)，因此我们可以利用方差分析进行三组之间的生产力差异的检测。

基础分析

```
# 导入数据
music <- read.csv("music.csv")
# 查看基本信息
str(music)
## 'data.frame': 150 obs. of 3 variables:
## $ ID : int 1 2 3 4 5 6 7 8 9 10 ...
## $ condition : chr "no_music" "no_music" "no_music" "no_music" ...
## $ productivity: num 188 196 194 190 157 ...
summary(music)
## ID condition productivity
## Min. : 1.00 Length:150 Min. :104.7
## 1st Qu.: 38.25 Class :character 1st Qu.:161.0
## Median : 75.50 Mode :character Median :185.0
## Mean : 75.50 Mean :184.9
## 3rd Qu.:112.75 3rd Qu.:205.0
## Max. :150.00 Max. :285.3
```

主要分析

整体显著性检验

```
# 方差分析
res <- aov(productivity ~ condition, data = music)

# 查看结果
summary(res)
## Df Sum Sq Mean Sq F value Pr(>F)
## condition 2 24734 12367 9.291 0.000159 ***
## Residuals 147 195661 1331
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

这里进行整体显著性检验，通过p值 $0.000159 < 0.05$ ，在置信度为0.95的情况下，结果表明有证据说明工作条件对生产力有显著影响。

两两比较显著性检验

```
# 进行Tukey's HSD测试
TukeyHSD(res)
##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = productivity ~ condition, data = music)
##
## $condition
##
##              diff          lwr          upr      p adj
## music_no_choice-music_choice -25.820579 -43.09679  -8.544367 0.0015539
## no_music-music_choice        -28.466400 -45.74261 -11.190188 0.0004246
## no_music-music_no_choice      -2.645821 -19.92203  14.630391 0.9301260
```

上述结果说明不同工作条件下对生产力均值之间两两比较：

1. music_no_choice 与 music_choice：

- **差异：** 均值差异为-25.82，95%置信区间为[-43.10, -8.54]。
- **显著性：** p值为 $0.00155 < 0.05$ ，在置信度为0.95的情况下，有证据说明在可选择音乐工作状态下的员工相比不可选择音乐工作状态下的员工，生产力有明显差异，并且可选择音乐员工生产力均值更高。

2. no_music 与 music_choice：

- **差异：** 均值差异为-28.47，95%置信区间为[-45.74, -11.19]。
- **显著性：** p值为 $0.00042 < 0.05$ ，在置信度为0.95的情况下，有证据说明在不听音乐的员工的生产力和可选择音乐工作状态下的员工有显著差异，并且可选择音乐员工生产力均值更高。

3. no_music 与 music_no_choice：

- **差异：** 均值差异为-2.65，95%置信区间为[-19.92, 14.63]。
- **显著性：** p值为 $0.93013 > 0.05$ ，在置信度为0.95的情况下，不能说明在不听音乐的员工和不可选择音乐员工生产力有显著差异。

分析结果和解读

- 方差分析的结果表明这三组之间的生产力存在统计学上的显著差异。
- 组之间的生产力之间的区别需要分别讨论
 - 对照组和治疗组A的生产力不能说明有显著差异
 - 对照组和治疗组B的生产力可以说明有显著差异
 - 治疗组A和治疗组B的生产力可以说明有显著差异
- 因此，治疗组B即可以选择音乐的员工生产力具有最高的水平