

Homework 4: 音乐对生产力提升的研究

2151294 马威

目录

1 研究问题和假设	1
1.1 背景	1
1.2 现有研究	1
1.3 研究问题	2
1.4 研究假设	2
2 数据分析方法选择	2
3 基础分析	2
4 主要分析	4
5 分析结果解读	5
5.1 音乐对于生产力影响:	5
5.2 三组生产力水平显著性比较:	5
5.3 三组生产力水平比较:	5
5.4 总结	6

1 研究问题和假设

1.1 背景

一个在线零售商希望提高员工的工作效率，同时改善他们的工作体验。目前，零售商订单管理中心的员工在工作时没有得到任何形式的娱乐，如背景音乐、电视等。零售商想知道提供一些员工要求的播放背景音乐是否会提高生产力，如果能提高，具体能提高多少。

1.2 现有研究

随机抽取 150 名员工。随机分为三组，每组 50 名参与者：

- 对照组 (control group): 不听音乐

- 治疗组 (treatment group): 听音乐, 但不能选择听什么
- 第二个治疗组 (treatment group): 听音乐, 可以自主选择听什么音乐。

在实验结束时, 三组的“生产力”是根据“每小时处理的平均包裹数量”来衡量。

1.3 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. 两两来看, 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?

1.4 研究假设

1. 音乐对生产力有显著影响, 且适当的娱乐放松会提高生产力。
2. 生产力水平预测为: 治疗组 B (可选音乐) > 治疗组 A (不可选音乐) > 对照组 (无音乐)

2 数据分析方法选择

1. 对于音乐这个因素对于生产力这个观测值的影响, 适合使用单因素方差分析法: 因变量是“生产力”, 而解释变量是“分组”。
2. 对于三组生产力水平的排序, 由于先前经过了方差分析, 可以使用 TokeyHSD 进行有修正的两两显著性检验, 并在有显著性的组对中进行单侧均值假设检验来进行比较。

3 基础分析

```
# 装载必要的包
library(ggplot2)

# 清除当前镜像中的数据
rm(list = ls())

# 读取数据
music.data <- read.csv("2 ANOVA_data_music.csv")

# 取三组的生产力数据
no_music <- music.data[music.data$condition == "no_music",][["productivity"]]
music_no_choice <- music.data[music.data$condition == "music_no_choice",][["productivity"]]
music_choice <- music.data[music.data$condition == "music_choice",][["productivity"]]
```

```
# 进行基本分析
```

```
sum1 <- summary(no_music)
sum2 <- summary(music_no_choice)
sum3 <- summary(music_choice)
```

```
# 显示分析结果
```

```
print("summary of no_music:")
print(sum1)
print("summary of music_no_choice")
print(sum2)
print("summary of music_choice")
print(sum3)
```

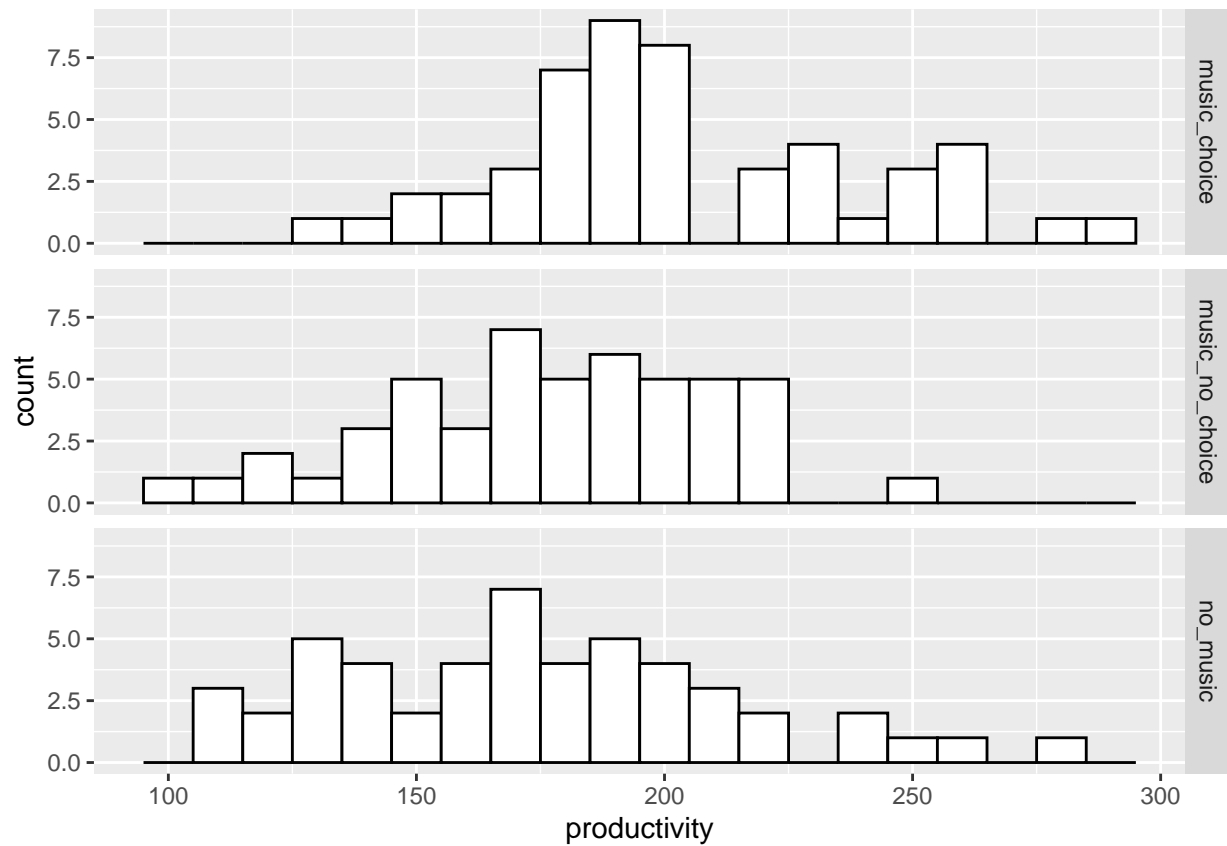
```
# 就平均数和中位数来看，治疗组 A 略微高于对照组，但治疗组 B 显著高于其余两组。
```

```
# 不过，音乐是否真的对生产力有显著影响，还需进行单因素方差分析。
```

```
## [1] "summary of no_music:"
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    110.7   143.1   171.7   174.5   196.7   276.6
## [1] "summary of music_no_choice"
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    104.7   152.4   179.0   177.1   201.0   252.8
## [1] "summary of music_choice"
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    130.9   180.4   195.0   203.0   230.8   285.3
```

```
# 图示三组分布情况
```

```
ggplot(music.data, aes(x = productivity)) +
  geom_histogram(fill = "white", color = "black", binwidth = 10) +
  facet_grid(condition~.)
```



4 主要分析

```
# 音乐对生产力影响的单因素方差分析
```

```
# H0: 影响不显著
```

```
# H1: 影响显著
```

```
fit <- aov(productivity~condition, data = music.data)
```

```
summary(fit)
```

```
##           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
```

```
# 两两之间比较显著性（有修正）
```

```
TukeyHSD(fit)
```

```
##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
```

```
## Fit: aov(formula = productivity ~ condition, data = music.data)
##
## $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

```
# 两两之间比较：对照组、治疗组 B
```

```
# H0: 治疗组 B <= 对照组
```

```
# H1: 治疗组 B > 对照组
```

```
comp.test1 <- t.test(music_choice, no_music, alternative = "greater")
print(comp.test1$p.value)
```

```
## [1] 0.0001652367
```

```
# 两两之间比较：治疗组 A、治疗组 B
```

```
# H0: 治疗组 B <= 治疗组 A
```

```
# H1: 治疗组 B > 治疗组 A
```

```
comp.test2 <- t.test(music_choice, music_no_choice, alternative = "greater")
print(comp.test2$p.value)
```

```
## [1] 0.0001648396
```

5 分析结果解读

5.1 音乐对于生产力影响：

$p\text{-value} = 0.000159 < 0.05$ ，拒绝 H_0 ，即认为音乐对于生产力影响显著

5.2 三组生产力水平显著性比较：

- 治疗组 A、治疗组 B： $p\text{ adj} = 0.00155 < 0.05$ ，拒绝 H_0 ，即认为两组有显著差异
- 对照组、治疗组 B： $p\text{ adj} = 0.00042 < 0.05$ ，拒绝 H_0 ，即认为两组有显著差异
- 对照组、治疗组 A： $p\text{ adj} = 0.93913 > 0.05$ ，无法拒绝 H_0 ，即无法认为两组有显著差异

5.3 三组生产力水平比较：

- 对照组、治疗组 B： $p\text{-value} = 0.000165 < 0.05$ ，拒绝 H_0 ，即认为治疗组 B 均值高于对照组均值
- 治疗组 A、治疗组 B： $p\text{-value} = 0.000165 < 0.05$ ，拒绝 H_0 ，即认为治疗组 B 均值高于治疗组 A 均值

5.4 总结

1. 音乐对于生产力影响显著
2. 三组的生产力水平：对照组 = 治疗组 A < 治疗组 B
3. 先前假设中关于音乐对生产力影响显著的假设正确，但没有足够证据证明关于三组生产力水平的假设，因为对照组和治疗组 A 的显著性没有足够的证据证明存在。
4. 但总而言之，治疗组 B 不但均值显著大于对照组和治疗组 A，而且经过假设检验，总体水平也高于这两个组，因此有理由认为治疗组 B 有最高的生产力水平。