

# 2151299\_苏家铭\_hw2

苏家铭

## Table of Contents

This is an [R Markdown](#) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

```
# Body piercing data
american.bp <- c(3, 5, 2, 1, 4, 4, 6, 3, 5, 4)
european.bp <- c(6, 5, 7, 7, 6, 3, 4, 6, 5, 4)
# Store data in a dataframe
ebp.survey <- data.frame("bp" = c(american.bp, european.bp),
  "group" = rep(c("American", "European"), each = 10), stringsAsFactors = FALSE)

# 导入pirates包
load("pirates.RData")

library(yarrn)

## 载入需要的程辑包: jpeg

## 载入需要的程辑包: BayesFactor

## 载入需要的程辑包: coda

## 载入需要的程辑包: Matrix

## *****
## Welcome to BayesFactor 0.9.12-4.5. If you have questions, please contact Richard Morey (richarddmorey@gmail.com).
##
## Type BFManual() to open the manual.
## *****

## 载入需要的程辑包: circlize

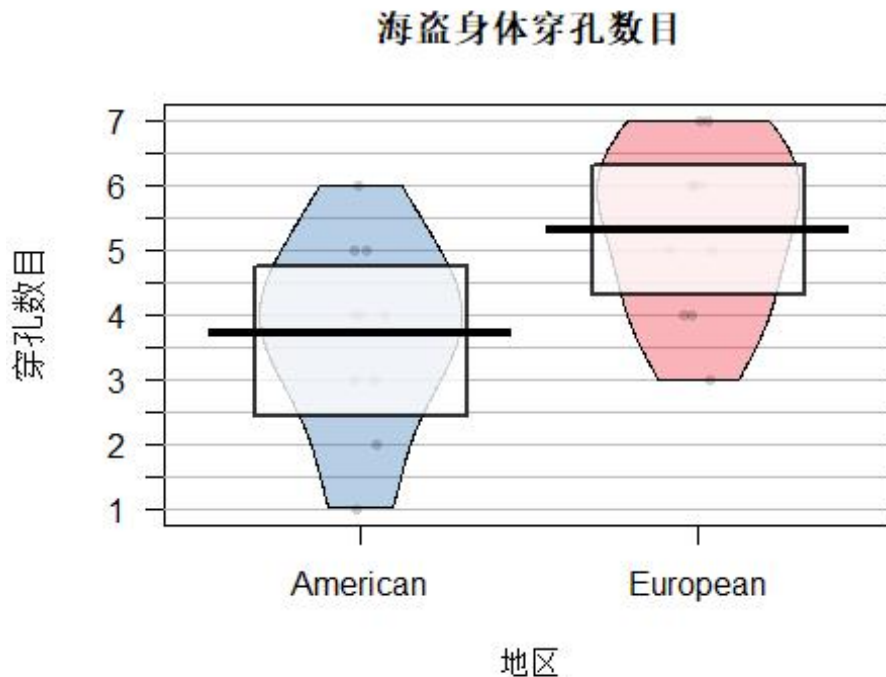
## =====
## circlize version 0.4.15
## CRAN page: https://cran.r-project.org/package=circlize
## Github page: https://github.com/jokergoo/circlize
## Documentation: https://jokergoo.github.io/circlize_book/book/
```

```
##
## If you use it in published research, please cite:
## Gu, Z. circlize implements and enhances circular visualization
##   in R. Bioinformatics 2014.
##
## This message can be suppressed by:
##   suppressPackageStartupMessages(library(circlize))
## =====

## yarr v0.1.5. Citation info at citation('yarr'). Package guide at ya
rrr.guide()

## Email me at Nathaniel.D.Phillips.is@gmail.com

pirateplot(bp ~ group, data = ebp.survey, main = "海盗身体穿孔数目", xlab
= "地区", ylab = "穿孔数目")
```



# 中位数 (Median)：箱线图的矩形框中线表示数据的中位数。观察中位数可以得知数据的中心趋势。欧洲的中位数比美国的高表示欧洲海盗穿孔中间值比美国的大。

# 四分位数 (Quartiles)：箱线图显示了数据的上下四分位数，即数据的前 25% 和后 25% 的范围。箱体的上下边缘表示第三四分位数 (Q3) 和第一四分位数 (Q1)，而箱子的高度表示数据的中间 50% 范围。欧洲的四分位数也比美国高。

# 箱子的长度：箱子的长度 (IQR，即四分位距) 表示数据的离散程度，即在四分位数范围内的数据分布情况。离散程度相差无几。

# 综合来看，欧洲海盗比美国海盗身体穿孔数目更多。

# 步骤 2: t-test

```
p.test_result <- t.test(bp ~ group, data = ebp.survey)
print(p.test_result)
```

```
##
## Welch Two Sample t-test
##
## data: bp by group
## t = -2.5228, df = 17.783, p-value = 0.0214
## alternative hypothesis: true difference in means between group American and group European is not equal to 0
## 95 percent confidence interval:
## -2.9335927 -0.2664073
## sample estimates:
## mean in group American mean in group European
##                3.7                5.3
```

# t-statistic (-2.5228): t-statistic 为负值，表示"American"组的样本均值较小。

# 自由度 (df = 17.783): 这是一个衡量我们有多少信息来估计总体方差的指标。由于 Welch's t-test 考虑到了两组方差不一致的情况，因此自由度被修正为一个小数。

# p-value (0.0214): 在这里，p-value 小于通常选择的显著性水平（通常是 0.05），因此我们拒绝了零假设。这意味着我们有足够的证据表明"American"组和"European"组的平均穿孔数目是显著不同的。

# 95 percent confidence interval (-2.9335927, -0.2664073): 在这里，置信区间不包括零，这进一步支持我们对显著性差异的结论。负的区间下限表明"American"组的平均值可能小于"European"组。

# 综合来说，根据这个 t-test 的结果，我们有理由相信"American"组和"European"组的平均穿孔数目是不同的。

# 提取 29 岁和 30 岁的海盗的文身数量

```
tattoos_29 <- pirates$tattoos[pirates$age == 29]
tattoos_30 <- pirates$tattoos[pirates$age == 30]
```

# 进行 t 检验

```
t.test_result <- t.test(tattoos_29, tattoos_30)
```

# 查看 t-test 结果

```
print(t.test_result)
```

```
##
## Welch Two Sample t-test
##
## data: tattoos_29 and tattoos_30
## t = 0.26552, df = 119.15, p-value = 0.7911
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.058586 1.386455
## sample estimates:
## mean of x mean of y
## 10.081967 9.918033
```

**# t-statistic (0.26552):** t-statistic 接近零，说明观察到的均值差异相对较小。

**# 自由度 (df = 119.15):** 由于 Welch's t-test 考虑到了两组方差不一致的情况，因此自由度被修正为一个小数。

**# p-value (0.7911):** 这是一个衡量观察到的样本均值差异是否由于随机变异引起的指标。在这里，p-value 很大，大于通常选择的显著性水平（通常是 0.05），因此我们没有足够的证据拒绝零假设。这意味着我们没有足够的证据表明 29 岁和 30 岁海盗的文身数量存在显著差异。

**# 95 percent confidence interval (-1.058586, 1.386455):** 这是对真实均值差异的估计区间。由于包含零，这意味着我们对真实差异的估计不显著。

**# 综合来看，根据这个 t-test 的结果，我们没有足够的证据来支持 29 岁和 30 岁海盗的文身数量存在显著差异的假设。p-value 大，置信区间包含零，都支持这一结论。**

```
# 创建一个列联表
cross_table <- table(pirates$eyepatch, pirates$college)

# 进行卡方检验
c.test <- chisq.test(cross_table)

# 查看卡方检验结果
print(c.test)

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: cross_table
## X-squared = 0, df = 1, p-value = 1
```

**# X-squared (0):** X-squared 为 0，表明观察到的频数与期望频数非常接近，没有显著的偏离。

**# 自由度 (df = 1):** 这是卡方分布的自由度。对于 Pearson's Chi-squared test，自由度是 (行数 - 1) \* (列数 - 1)。在这里，由于只有两个列 ('CCCC' 和 'JSSFP'

' ), 因此自由度为 1。

# *p-value* (1): 这是检验的 *p-value*, 用于判断观察到的频数与期望频数之间的差异是否显著。在这里, *p-value* 等于 1, 这说明我们没有足够的证据拒绝零假设, 即是否佩戴眼罩与所上大学之间没有显著关联。

# 总体上, 根据这个卡方检验的结果, 我们没有足够的证据支持是否佩戴眼罩与所上大学之间存在显著关联的假设。

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.