

hw1

YangTengchao

```
# 引入数据
```

```
load("pirates.RData")
```

```
# 问题调查
```

```
# 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),
```

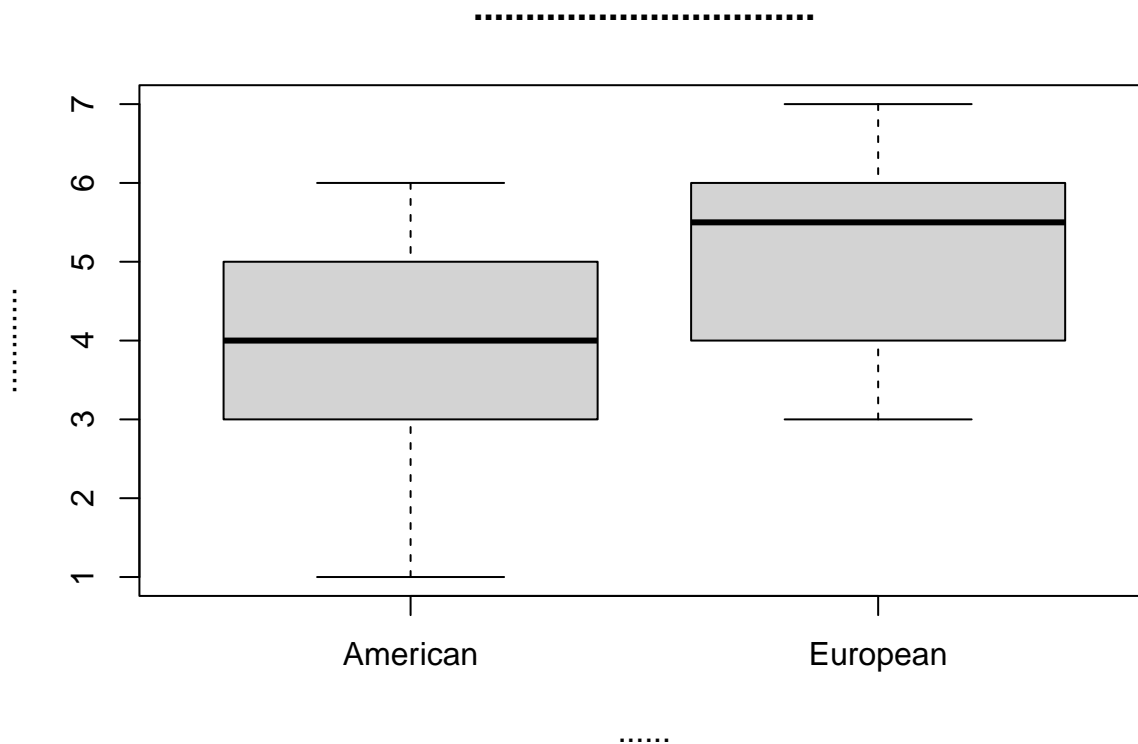
```
"group" = rep(c("American", "European"), each = 10), stringsAsFactors = FALSE)
```

```
# 箱线图判断
```

```
boxplot(bp ~ group, data = ebp.survey,
```

```
  main = "海盗身体穿孔数目箱线图",
```

```
  xlab = "地区", ylab = "穿孔数目")
```



对于中位数，欧洲地区的穿孔数目比美国高；同时说明欧洲地区穿孔数目多的人较多
 # 对于上下四分位数，可以看出欧洲地区穿孔数目主要集中在中间数目的水平；美国地区数目集中在中间偏大水平
 # 但从箱的长度可以看出其离散程度相差不大

*t.test*判断美国和欧洲穿孔数目差异

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

```
##
```

```
## 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
```

```
# 可以看出欧洲和美国穿孔数目有差异
# 可以看出 $p$ 值为 $0.0214 < 0.05$ ，说明在置信度 $0.95$ 的情况下
# 可以拒绝原假设，说明差异显著
```

```
# t.test判断29、30岁纹身差异
t.test <- t.test(tattoos ~ age, data = pirates, subset = age %in% c(29, 30))
t.test
```

```
##
## Welch Two Sample t-test
##
## data: tattoos by age
## t = 0.26552, df = 119.15, p-value = 0.7911
## alternative hypothesis: true difference in means between group 29 and group 30 is not equal to 0
## 95 percent confidence interval:
## -1.058586 1.386455
## sample estimates:
## mean in group 29 mean in group 30
##      10.081967      9.918033
```

```
# 没有足够证据证明29、30纹身情况存在差异
# 可以看出 $p$ 值 $0.7911 > 0.05$ ，在置信度为 $0.95$ 的情况下
# 没有足够证据拒绝原假设，因此不能说明有显著差异
```

```
# 卡方检验大学和佩戴眼罩是否有关联
c.test <- chisq.test(table(pirates$eyepatch, pirates$college))
c.test
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: table(pirates$eyepatch, pirates$college)
## X-squared = 0, df = 1, p-value = 1
```

```
# 没有足够证据证明大学和眼罩有关联
# 可以看出 $p$ 值 $1 > 0.05$ ，在置信度为 $0.95$ 的情况下
# 没有足够证据拒绝原假设，因此不能说明两者有关联
```