MockingTest

2118

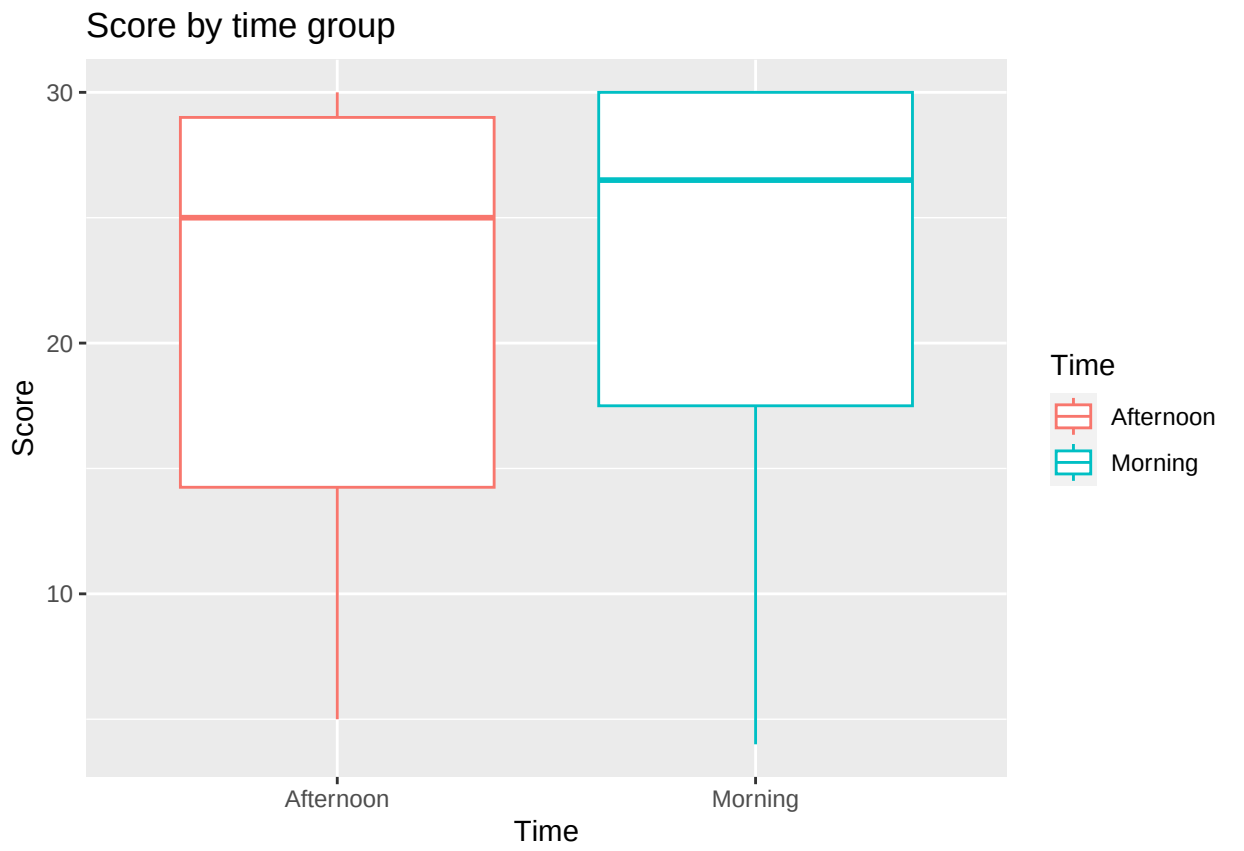
2023-12-21

Question 1: Stroop test

```
stroop=read.csv(file="stroop_test.csv")
stroop$Time=as.factor(stroop$Time)
```

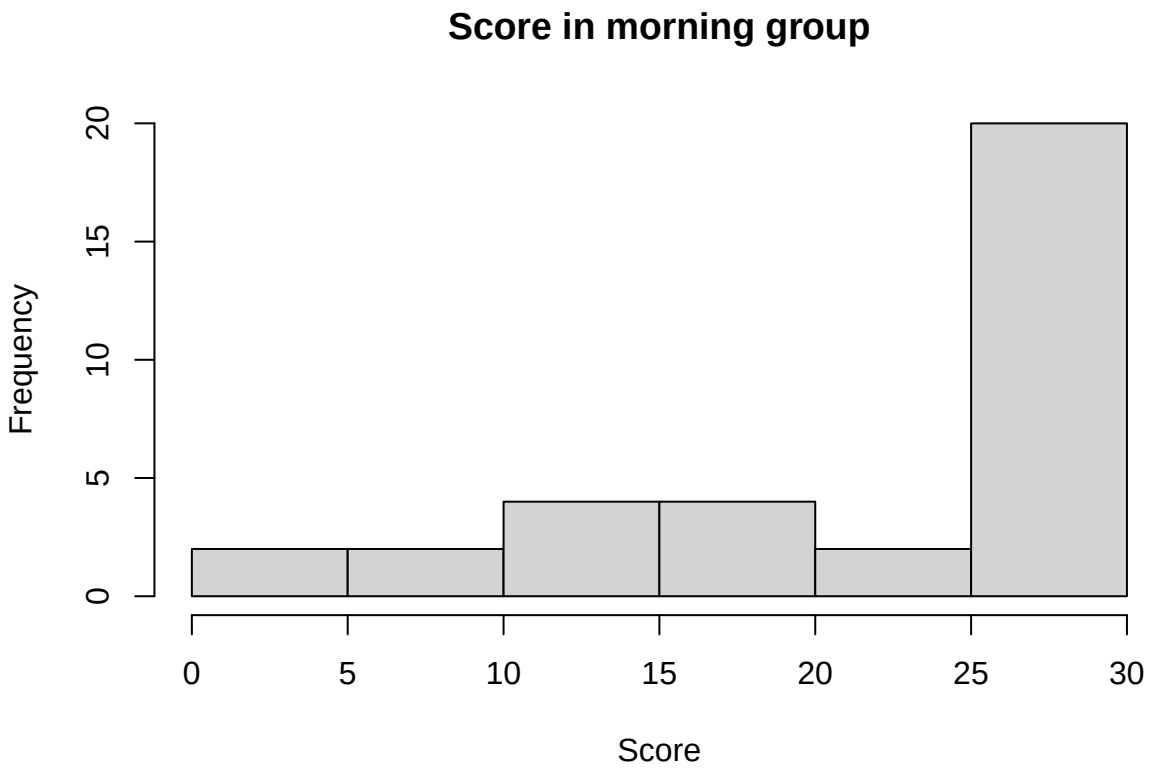
Score by time group

```
g = ggplot(data = stroop,
           mapping = aes(x = Time, y = Score, color = Time)) +
  geom_boxplot() + labs(title = "Score by time group")
g
```



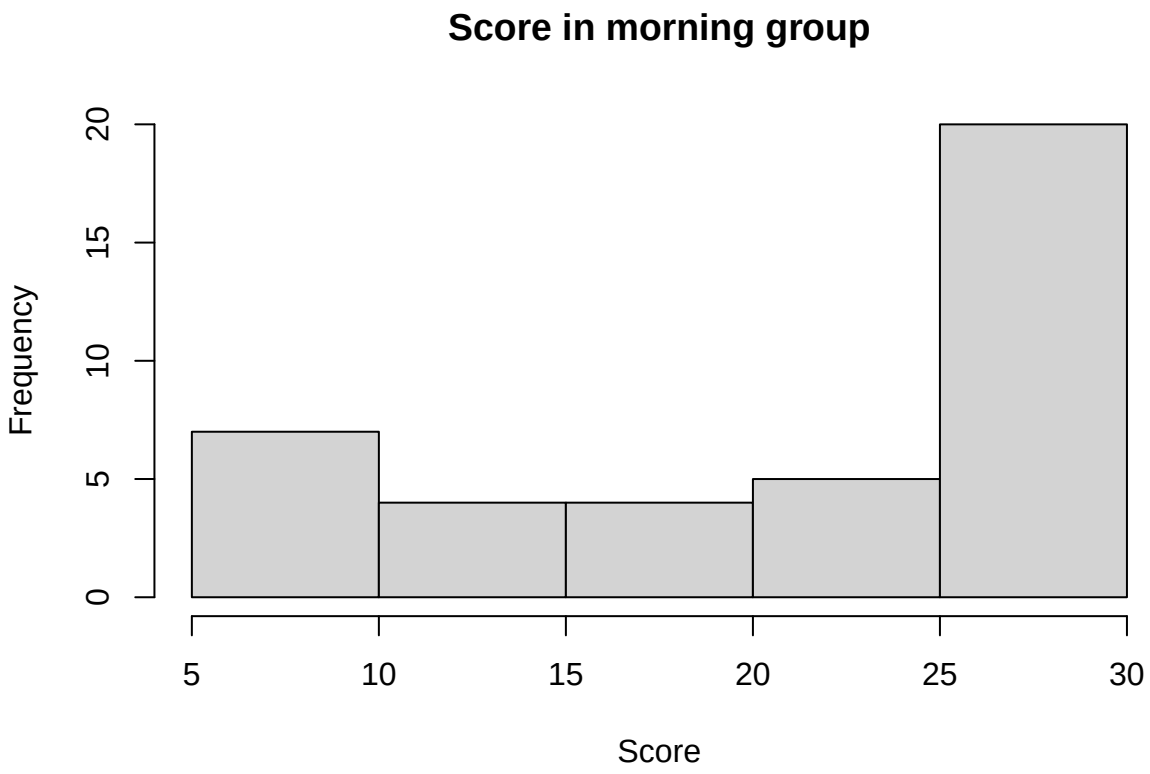
Score in morning group

```
hist(x=stroop[stroop$Time=="Morning",]$Score,xlab="Score",main="Score in morning group")
```



Score in afternoon group

```
hist(x=stroop[stroop$Time=="Afternoon",]$Score,xlab="Score",main="Score in morning group")
```



T-test analysis of permance difference

```
t.test(stroop$Score ~ stroop$Time, alternative = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  stroop$Score by stroop$Time
## t = -0.68147, df = 71.123, p-value = 0.4978
## alternative hypothesis: true difference in means between group Afternoon and group Morning is not equ
## 95 percent confidence interval:
##  -5.409583  2.653701
## sample estimates:
## mean in group Afternoon   mean in group Morning
##                21.47500                22.85294
```

Interpretation of t-test results

- H0: There is no difference in means between group Afternoon and group Morning.
- HA: There is difference in means between group Afternoon and group Morning.
- p-value=0.4978 > 0.05
- We cannot reject H0.

- There is insufficient evidence that there is difference in means between group Afternoon and group Morning.

Ways in which the study could be improved or followed up on

1. For one subject in the subject, test his or her score both in the morning and the afternoon.
2. Then, use paired t-test to test whether there is performance difference.

Question 2: Marathon finishing times

Time of amateur runners by age and gender

```
marathon = read.csv(file = "Chicago2013_random_finishers.csv")
marathon$Gender = as.factor(marathon$Gender)
#marathon$Age=as.factor(marathon$Age)
marathon$Age = cut(marathon$Age, breaks = c(18, 29, 40, 51, 63))
g = ggplot(data = marathon, mapping = aes(x = Age, y = Time, color = Gender))
g = g + geom_boxplot()
  #labs(title="")+
g = g + geom_jitter()
#g = g + geom_point(color = "classic", size=1.22)
g = g + theme(axis.title.x = element_text(size=15,
              margin = margin(t = 10)))
g = g + theme(axis.title.y = element_text(size=15,
              margin = margin(r = 10)))
g = g + theme(axis.text = element_text(size=12))
g = g + scale_color_discrete(labels=c("Female","Male"))
g = g + theme(legend.title = element_text(size = 12))
g = g + guides(color = guide_legend(overrides.aes = list(size = 30)))
g
```

Average finishing times for each gender

```
aggregate(Time~Gender,data=marathon,FUN=mean) %>%
  kable()
```

Gender	Time
F	4.883765
M	4.517882

Standard deviation of finishing times for each gender

```
aggregate(Time ~ Gender, data = marathon, FUN = sd) %>%
  kable()
```

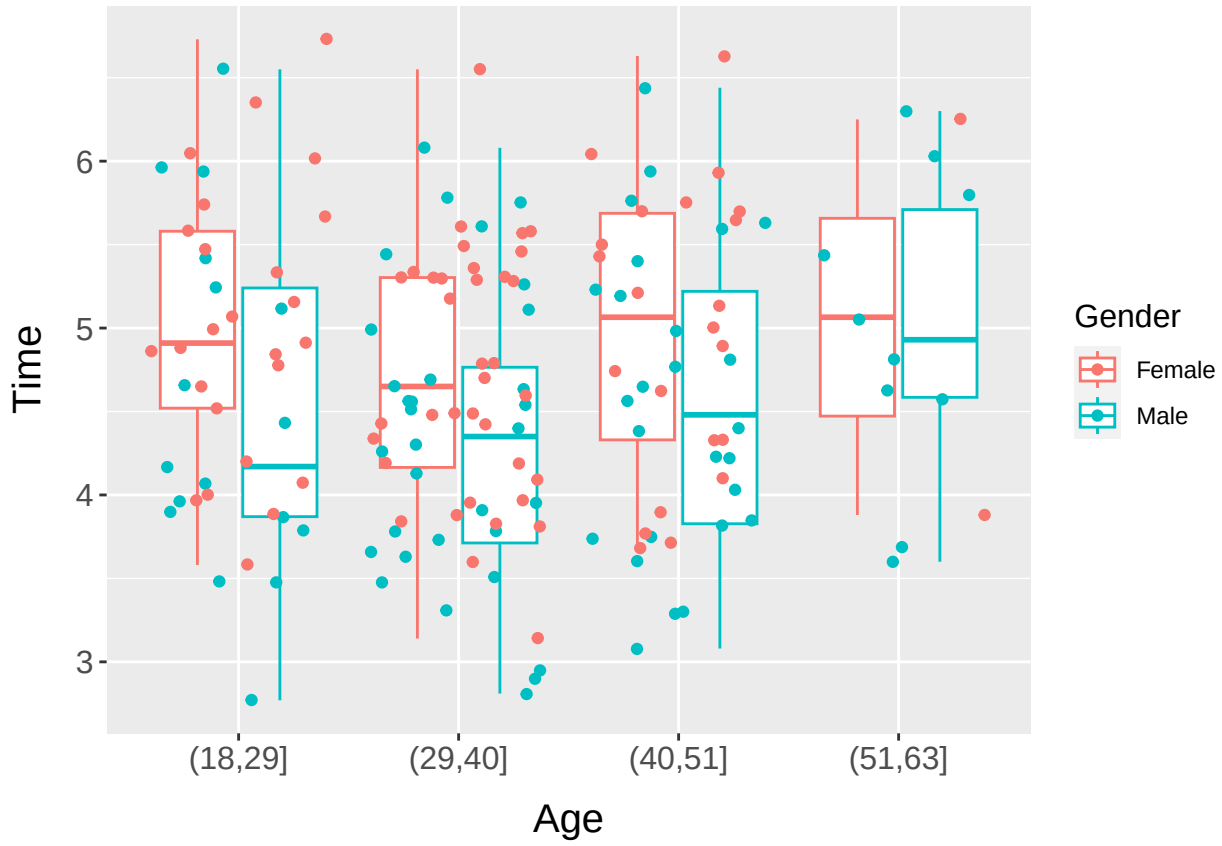


Figure 1: Figure. Time of amateur runners by age and gender.

Gender	Time
F	0.8073826
M	0.9243184

Average finishing times for each age quartile

```
kable(aggregate(Time~Age,data=marathon,FUN=mean))
```

Age	Time
(18,29]	4.812381
(29,40]	4.538088
(40,51]	4.757917
(51,63]	5.004167

Standard deviation of finishing times for each age quartile

```
aggregate(Time~Age,data=marathon,FUN=sd)
```

```
##       Age      Time
## 1 (18,29] 0.9309471
## 2 (29,40] 0.8213463
## 3 (40,51] 0.8927485
## 4 (51,63] 0.9741053
```

Determine the effect of age quartile and gender on marathon finishing time

1. To determine the effect of age quartile on marathon finishing time, we can use ANOVA.
2. To determine the effect of gender on marathon finishing time, we can use two-sample test.

Question 3: Antiviral drug

The Null Hypothesis and the Alternative Hypothesis

- H0: There is no difference in the full recovery time between the experiment group and the control group
- HA: There is difference in the full recovery time between the experiment group and the control group

Error relate to the the significant level

- Type 1 error: H0 is true in the true situation, but there is significant evidence for HA in the results of her statistical analysis.
- Type 2 error: HA is true in the true situation, but there is lack of significant evidence for HA in the results of her statistical analysis.

Meaning of P - value

If H0 is true, the probablity of the test statistic being at least as extreme as the value of the test statistic that was actually obtained.

Report of this study

p-value = $0.059 > 0.05$ We cannot reject H0. There is insufficient evidence that there is difference in means between the experiment group and the control group. Therefore, there is insufficient evidence that there is difference in treament efficacy between the experiment group and the control group.