MA256 Lesson 11 - Two Groups, Two Means (6.1-6.3)

Review for Single Mean:

Hypotheses (in symbols):

H_0 :

H_a :

Strength of Evidence:

Confidence Interval:

Two means:

Hypotheses (in symbols):

H_0 :

H_a :

Strength of Evidence:

Confidence Interval:

IOCT Tabbing

How fair are the "tabbing" IOCT times? Currently to tab the IOCT a male must have a time faster than 2:38 (158 seconds) and a female must have a time faster than 3:35 (215 seconds). This is a difference of 57 seconds. Is this a fair difference? How much do male and female's times differ on the IOCT? You think that the difference in average male and female times is not 57 seconds. The data we have is the IOCT times for all cadets who took MA206 in AY21-1 and have a valid IOCT time.

```
library(tidyverse)
ioct.data <- read.csv("https://raw.githubusercontent.com/jkstarling/MA256/main/data/IOCT_tab_data.csv",
                      stringsAsFactors = TRUE)
```

1. Identify the explanatory and response variables recorded and classify them as either categorical or quantitative.

2. In words, state the null and the alternative hypotheses to t est whether male or female's IOCT times differ from 57 seconds.

3. Define the parameters of interest and assign symbols.

4. State the null and the alternative hypotheses in symbols.

5. Calculate the five-number summary of IOCT time by group, calculate the IQR for each group, and create a graphical representation of the five-number summary in R Studio using example code from the course guide.

```
# ioct.data %>%
#   group_by(sex) %>%
#   summarize(Minimum = XXX,
#             LowerQuartile = XXX,
#             Median = XXX,
#             UpperQuartile = XXX,
#             Maximum = XXX)
#
# XXX-XXX #IQR-Females
# XXX-XXX # IQR-Males
#
# ioct.data %>%
#   ggplot(aes(x = as.factor(XXX), y = XXX)) + geom_boxplot() +
#   labs(x = "Biological Sex", y = "IOCT Time (Seconds)",
#        title = "IOCT Time (Seconds) by Biological Sex")
```

6. Do the validity conditions appear to be satisfied for these data? Justify your answer.

```
# ioct.data %>%
#   ggplot(aes(x=XXX)) +
#   geom_histogram() +
#   facet_grid(XXX~.)
```

7. Conduct the theory-based two-sample t-test:

(a) What is the standardized statistic?

```
# ioct.data %>%
#   group_by(XXX) %>%
#   summarize(avgtime=XXX,
#             sdtime=XXX,
#             size=XXX)

# # Calculate the Standardized Statistic
# xbar_M <- XXX
# xbar_F <- XXX
# s_M <- XXX
# s_F <- XXX
# n_M <- XXX
# n_F <- XXX
# sd <- XXX
# null <- XXX
# statistic <- xbar_F-xbar_M
# t <- (statistic-null)/sd
# c(statistic, t)
```

(b) In light of your standardized statistic, should you expect the p-value to be large or small? How are you deciding?

(c) What is your p-value? Provide the value (number) and an explanation in context of the problem.

```
# pval <- XXX
# pval
```

(d) Based on the p-value, evaluate the strength of evidence provided by the data against the null hypothesis. Do you reject or fail to reject your null hypothesis.

8. Determine the 95% confidence interval for the difference in means of male and female IOCT Times.

(a) What is the 95% confidence interval?

```
# multiplier <- XXX
# sd <- XXX
# c(statistic - multiplier * sd, statistic + multiplier * sd)
```

(b) Does the 95% confidence interval agree with your conclusion in #5?

(c) Interpret the 95% confidence interval.

9. Operating under the null hypothesis mentioned in # 3 above, create a simulation to simulate males and females taking the IOCT. Estimate the p-value. Use 1000 replications and plot your results.

```
# set.seed(256)
#
# M <- 1000
# ...
```