Training Intervention Analysis

your name, id=your student id

28 October, 2022

Before you start: if you are a Mac user, you will need to install Xquartz from https://www.xquartz.org (https://www.xquartz.org) so you can use the 'tolerance' package. You can delete this line from your final report.

Context: Celtic Study

A sample of 18 full-time youth soccer players from a Youth Academy performed high intensity aerobic interval training over a 10-week in-season period in addition to usual regime of soccer training and matches.

The aim of this study to find if this extra training improves V_IFT, the maximum velocity (km/hr) achieved in an intermittent fitness test (`VIFT_Pre` vs `VIFT_Post`)?

This is a **paired design:** each player's V_IFT measured before and after the training intervention (i.e. start and after 10 weeks)

A scaffold for the analysis with the response variable VO2 max is provided below. You need to rerun the analysis using the V_IFT variables (i.e. `VIFT_Pre` vs `VIFT_Post`) to answer the question of interest: is there, on average, an improvement in V_IFT? To assess the evidence, you will provide confidence intervals, and other statistical inference, for the mean improvement of players in the population (eg of future youth soccer players under the same training intervention).

To answer the question of interest, provide a detailed response for all of the tasks asked below using the V_IFT variables (i.e. `VIFT_Pre` vs `VIFT_Post`).

Task: State the appropriate null and alternative hypotheses for the V_IFT study.

Task: Define a Type I and Type II error and discuss the implication of making these errors in this study.

Read in the training intervention data

Read in the data and have a look at the variable names and structure of the data.

```
train.df <- read.csv("Training_intervention_data.csv")
glimpse(train.df)
```

```
## Rows: 18
## Columns: 5
## $ ID           <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17…
## $ VO2.max_Pre  <dbl> 66.4, 70.9, 64.9, 68.6, 76.7, 75.6, 78.1, 73.1, 74.4, 64.…
## $ VO2.max_Post <dbl> 67.8, 81.7, 70.1, 73.0, 84.5, 78.4, 80.5, 76.0, 78.7, 72.…
## $ VIFT_Pre     <dbl> 23.8, 28.3, 25.2, 26.9, 30.1, 29.9, 29.5, 30.2, 31.5, 22.…
## $ VIFT_Post    <dbl> 23.3, 33.3, 25.8, 30.2, 34.6, 32.7, 31.8, 30.6, 32.0, 27.…
```

Focus on the V_IFT response variables

Summary Statistics

```
train.df %>% select(VO2.max_Pre,VO2.max_Post) %>% summary()
```

```
##   VO2.max_Pre     VO2.max_Post
##  Min.   :58.70   Min.   :64.50
##  1st Qu.:63.30   1st Qu.:68.10
##  Median :66.40   Median :72.10
##  Mean   :67.66   Mean   :72.77
##  3rd Qu.:72.55   3rd Qu.:77.80
##  Max.   :78.10   Max.   :84.50
```

Task: Interpret!

Mean and Standard Deviation

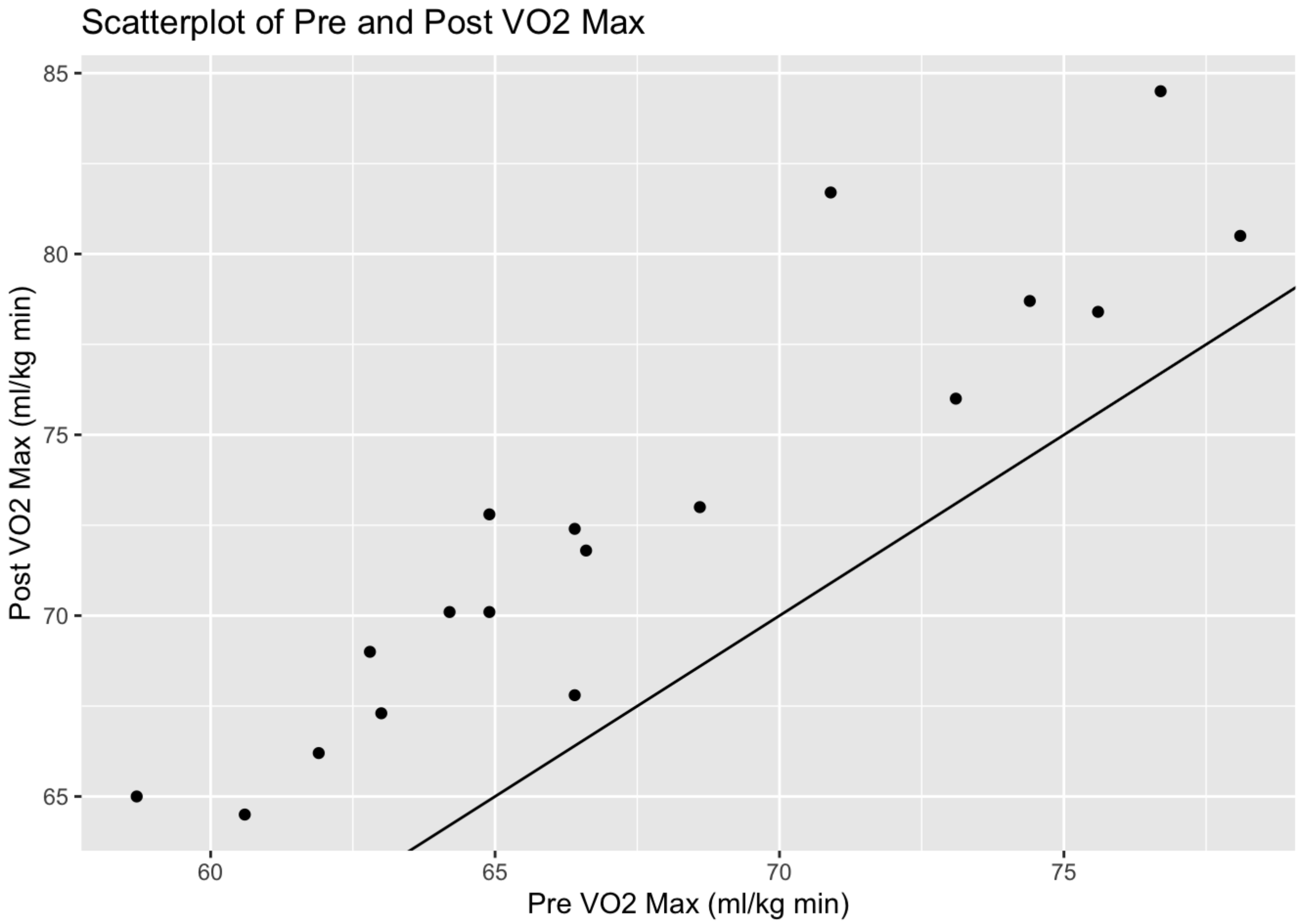
```
train.df %>% select(VO2.max_Pre,VO2.max_Post) %>%
            summarize(Pre_Mean=mean(VO2.max_Pre), Pre_SD= sd(VO2.max_Pre),
                      Post_Mean=mean(VO2.max_Post), Post_SD= sd(VO2.max_Post))
```

```
##   Pre_Mean   Pre_SD Post_Mean  Post_SD
## 1 67.65556 5.842564  72.76667 5.992839
```

Task: Interpret!

Scatterplot of Pre and Post with line of equality

```
train.df %>% ggplot(aes(x = VO2.max_Pre, y = VO2.max_Post)) +
        geom_point() +
  ggtitle("Scatterplot of Pre and Post VO2 Max") +
  ylab("Post VO2 Max (ml/kg min)") +
  xlab("Pre VO2 Max (ml/kg min)") +
  geom_abline(slope=1, intercept=0)
```



Task: Interpret!

Calculate the Improvement in V_IFT

Calculate a new variable, “improvement”, and have a look at the data frame to see that it has been created. High values of VO2 max are good so Post-Pre is a better measure than Pre-Post to capture this - what about V_IFT?

```
train.df <- train.df %>% mutate(Improvement = VO2.max_Post-VO2.max_Pre) %>%
              glimpse()
```

```
## Rows: 18
## Columns: 6
## $ ID           <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17…
## $ VO2.max_Pre  <dbl> 66.4, 70.9, 64.9, 68.6, 76.7, 75.6, 78.1, 73.1, 74.4, 64.…
## $ VO2.max_Post <dbl> 67.8, 81.7, 70.1, 73.0, 84.5, 78.4, 80.5, 76.0, 78.7, 72.…
## $ VIFT_Pre     <dbl> 23.8, 28.3, 25.2, 26.9, 30.1, 29.9, 29.5, 30.2, 31.5, 22.…
## $ VIFT_Post    <dbl> 23.3, 33.3, 25.8, 30.2, 34.6, 32.7, 31.8, 30.6, 32.0, 27.…
## $ Improvement  <dbl> 1.4, 10.8, 5.2, 4.4, 7.8, 2.8, 2.4, 2.9, 4.3, 7.9, 4.3, 5…
```

Mean and Standard Deviation of Improvement in V_IFT

```
train.df %>% select(Improvement) %>%
            summarize(Imp_Mean=mean(Improvement), Imp_SD= sd(Improvement))
```

```
##   Imp_Mean   Imp_SD
## 1 5.111111 2.258289
```

Task: Interpret!

Boxplot of Improvement in V_IFT

```
train.df %>% ggplot(aes(x = "", y = Improvement)) +
        geom_boxplot() +
  ggtitle("Boxplot of Improvement in VO2 Max") +
  ylab("Improvement in VO2 Max (ml/kg min)") +
  xlab("") +
  coord_flip()
```

Task: Interpret!

95% Confidence Interval Using the t.test function

```
train.df %>% select(Improvement) %>% t.test()
```

```
##
##  One Sample t-test
##
## data:  .
## t = 9.6022, df = 17, p-value = 2.798e-08
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  3.988090 6.234132
## sample estimates:
## mean of x
##  5.111111
```

Task: Based on the output given answer the following questions:

- What is the mean improvement in V_IFT the population of interest? Interpret the relevant 95% Confidence Interval carefully.
- Use the relevant interval estimate and p-value to decide whether there is sufficient evidence in the sample provided to claim that there is any improvement on average in V_IFT in the population of interest.
- What are the assumptions underlying the one sample t-test presented?
- Explain why or why not the assumptions seem justified based on the output provided.

95% Bootstrap CI for the mean

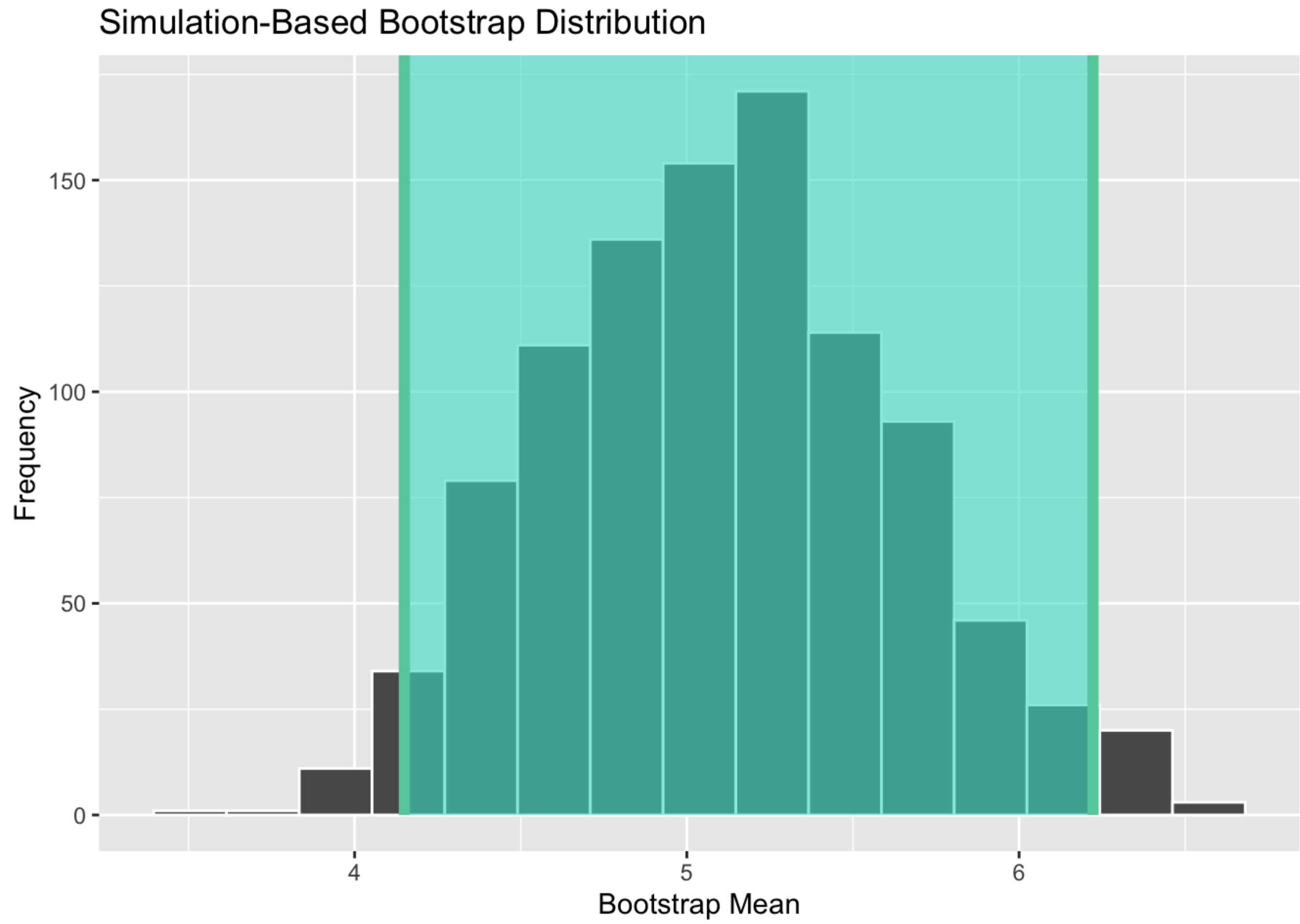
```
boot <- train.df %>%
  specify(response = Improvement) %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "mean")

percentile_ci <- get_ci(boot)
round(percentile_ci,2)
```

```
## # A tibble: 1 × 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1     4.15     6.22
```

Task: Interpret!

```
boot %>% visualize()+
  shade_confidence_interval(endpoints = percentile_ci) +
                  xlab("Bootstrap Mean") + ylab("Frequency")
```



Task: Interpret!

95% Bootstrap CI for the median improvement

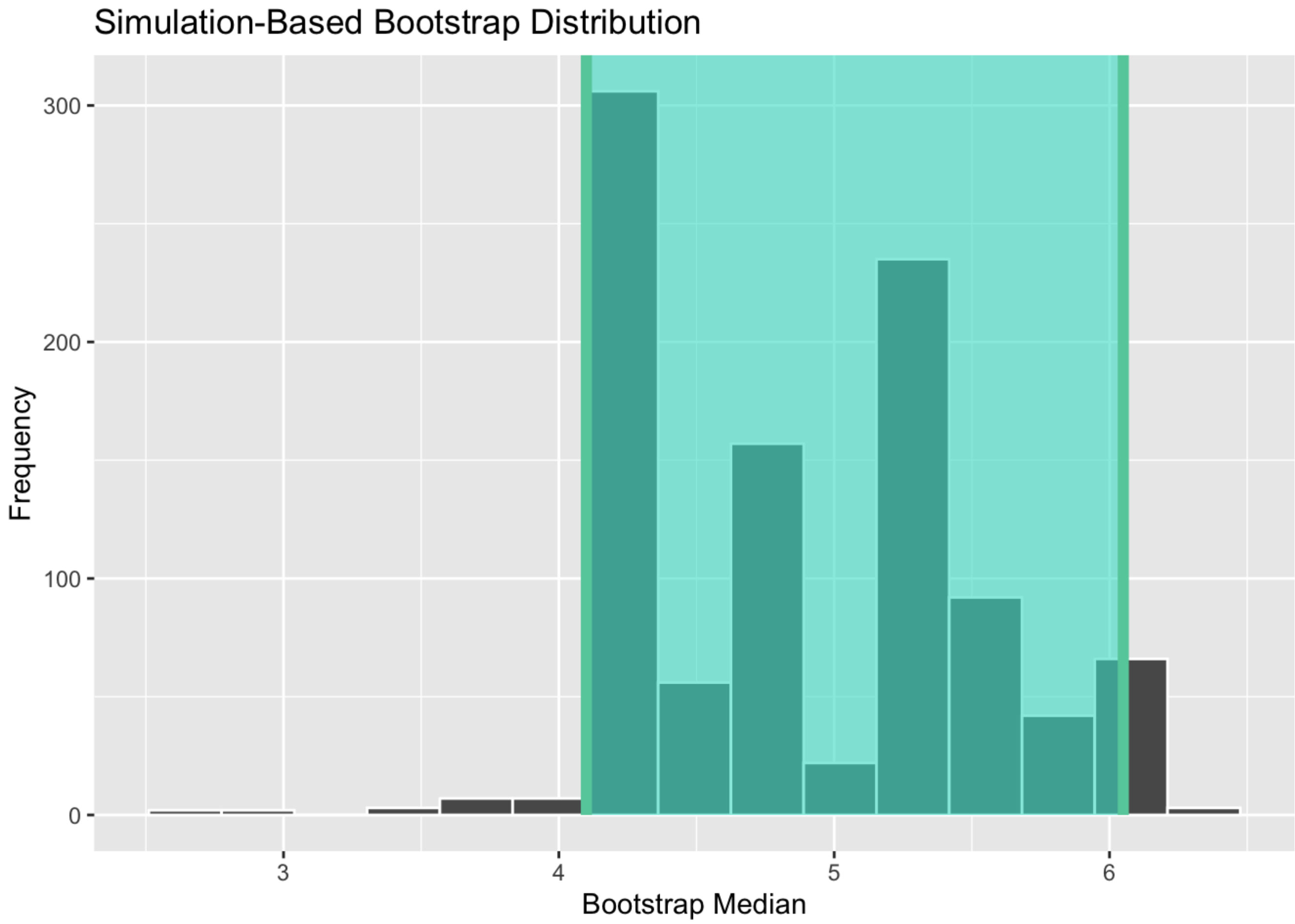
```
boot.median <- train.df %>%
  specify(response = Improvement) %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "median")

percentile_ci_median <- get_ci(boot.median)
round(percentile_ci_median,2)
```

```
## # A tibble: 1 × 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1      4.1     6.05
```

Task: Interpret!

```
boot.median %>% visualize()+
  shade_confidence_interval(endpoints = percentile_ci_median) +
                 xlab("Bootstrap Median") + ylab("Frequency")
```



Task: Interpret!

95% Tolerance Interval (Bonus Question)

Calculate a 95% tolerance interval covering 95% of V_IFT improvement values

```
normtol.int(train.df$Improvement, alpha = 0.05, P = 0.95)
```

```
##   alpha    P    x.bar 1-sided.lower 1-sided.upper
## 1  0.05 0.95 5.111111    -0.4283527      10.65057
```

Task: Interpret!

Overall Conclusion

Task: state your overall conclusion.