IS4300 Demonstration

Jason Holland

3/13/2021

# We check to see that we got the data by doing the following.  
cricketers[1:3,]

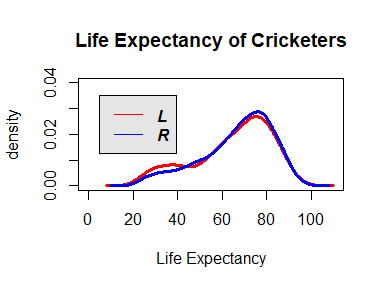
## # A tibble: 3 x 2  
## left right  
## <dbl> <dbl>  
## 1 82 82  
## 2 81 43  
## 3 77 47

# We can take this code out once we see the path works.

### Introduction

This demonstration involves a two-sample t-test and visualization of British Cricketers. We wish to analyze the life expectency across the factor **hand preference** to see if there is a significant difference between left and right handed cricketers. An abbreviated data set was created by the author and can be found at <https://github.com/jholland5/IS4300Project>. The full data set can be found at <https://vincentarelbundock.github.io/Rdatasets/datasets.html>

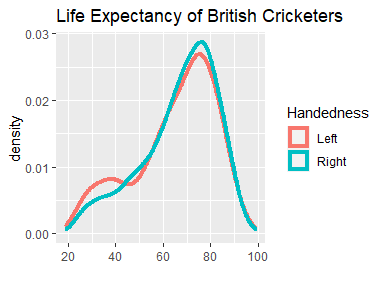
Before running the t-test, we visualize the distributions of life expectency for the two groups we are interested in. The following graph uses basic R graphing functions.



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

The plots are similar but there are some obvious deviations in life-expectency, especially around the ages 20 - 40.

We see if we can make a better graph using ggplot2. First we create a “tidy” data set for using ggplot2.



It is important to point out that there are a lot more right handers in the data set than left handers. Note that since the data is in tidy format now, we can use tidyverse functions to summarize the important variables. We demonstrate that next.

cricketdata %>% group\_by(Handedness) %>% summarize(number = n())

## # A tibble: 2 x 2  
## Handedness number  
## <fct> <int>  
## 1 Left 632  
## 2 Right 2755

We also summarize the life expectancy and we see that there is a fairly significant gap between the two averages.

cricketdata %>% group\_by(Handedness) %>% summarize(Mean = mean(LifeExpect))

## # A tibble: 2 x 2  
## Handedness Mean  
## <fct> <dbl>  
## 1 Left 65.0  
## 2 Right 66.7

### Two-Sample T-Test

It is easy to run a t-test in R. We use the following code.

t.test(left,right)

##   
## Welch Two Sample t-test  
##   
## data: left and right  
## t = -2.2503, df = 899.23, p-value = 0.02467  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -3.183791 -0.217410  
## sample estimates:  
## mean of x mean of y   
## 64.99051 66.69111

We can conclude that at the .05 level of significance, there is a statistically significant difference between the left and right handers life expectancy. We see that right handers are living longer by about 1.7 years.