IS4300 Final Project

Devin Bagley

5/6/2022

First let’s make sure the data has been imported correctly.

## # A tibble: 6 x 9  
## ...1 left year life dead acd kia inbed cause  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>  
## 1 1 right 1890 102 0 0 0 0 alive  
## 2 2 left 1892 100 0 0 0 0 alive  
## 3 3 right 1893 99 0 0 0 0 alive  
## 4 4 right 1894 98 0 0 0 0 alive  
## 5 5 right 1896 96 0 0 0 0 alive  
## 6 6 right 1896 96 0 0 0 0 alive

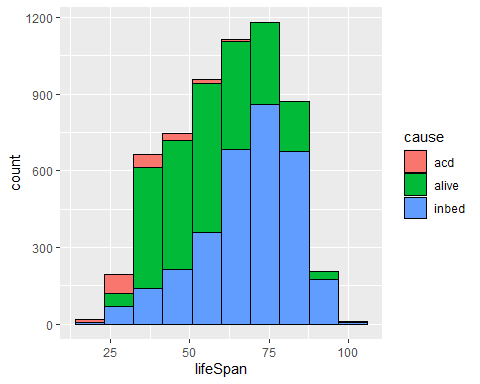
It worked.

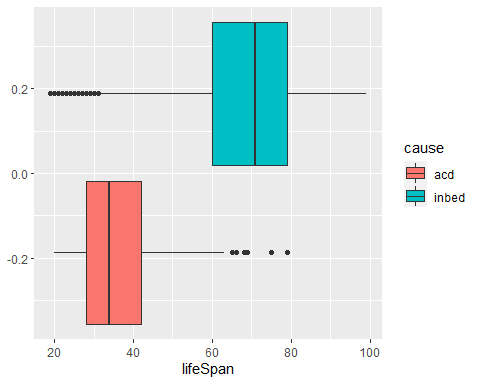
The variable names have been altered slightly from the excel sheet for continuity and clarity. Those changes are listed here:

x -> ID, left -> hand, year -> birthYear, life -> lifeSpan, acd -> accident, inbed -> inBed.

## Introduction

In this report we will look at a chi-squared test and charts for British Cricketers from the past. We will look at the difference in life span between people that died in an accident and the trends that lie within. This data set can be found at <https://github.com/Melee1114/IS3400Project>. The complete data set can be found at <https://vincentarelbundock.github.io/Rdatasets/datasets.html>.

First we’ll look at a chart that shows the causes of death based on the ages of the cricketers. Some are still alive so that is also portrayed.   From this we can see that most deaths caused by accidents take place before the age of 50. Cricketers dying in bed are more centralized around the age of 75.

This box chart shows the difference in the amount of deaths by accidents versus by in bed.  Based on this information, we can look at when a cricketer dies, the probability they die to an accident versus in bed.

## Chi-Squared Test

We can run the test in R by rolling a sample continuously and looking at both the expected probability and the actual probability. Here are the results:

## Sample10000  
## acd inbed   
## 559 9441

## acd inbed  
## Sample10000 559 9441  
## Expected 555 9445

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: Sample10000  
## X-squared = 0.0085555, df = 1, p-value = 0.9263

We conclude that our sample is close to the expected for randomness because of the low p-value.