MA388 Sabermetrics: SCORE Project

Module Proposal

CDTs Bridget Ge and Claire Tsay

1. Learning Goals
   A student who successfully completed this module should have the ability to:

a. Apply the six steps of the statistical investigation method to comparing two groups on a quantitative response.
b. Calculate the five-number summary (quartiles) and create histograms and boxplots to explore the data from two groups with a quantitative response variable.
c. Develop a null and alternative hypothesis for a research question for comparing two means.
d. Assess the statistical significance of the observed difference between two groups.
e. Apply the 3S Strategy to assess whether two sample means differ enough to conclude that there is a genuine difference in the population means or long-run means of a process.
f. Use the 2SD method to estimate a confidence interval for the difference in two means.
g. Determine the strength of evidence using the theory-based approach (two-sample t-test) for comparing two means.

2. Introduction
   The question that we are trying to answer using the data and our statistical exploration is "Is there an association between dominant foot and free kick accuracy for players in the European Soccer League?" We can use data from the teams to determine each player's preferred foot and their free-kick accuracy score (scale of 0 to 100).

3. Data
   The data for this project comes from the Kaggle Database titled "European Soccer Database: 25k+ matches, players & teams attributes for European Professional Football" uploaded by Hugo Mathien. The data for this dataset comes from multiple sources, including http://football-data.mx-api.enetscores.com/ The data for this dataset comes from multiple sources, including http://football-data.mx-api.enetscores.com/ (includes scores, lineup, team formation and events), http://www.football-data.co.uk/ (betting odds), and (player and team attributes).

The data includes a total of seven tables: Country, League, Match, Player, Player_Attributes, Team, and Team_Attributes, and these seven tables have a total of 199 columns. According to the description, the database contains data from 2008 to 2016 on more than 25,000 matches, 10,000 players their attributes, 11 European countries and their lead championship, team line ups, betting odds, and detailed match events (goal types, possession, corner, cross, fouls, cards, etc.)

Below is the code for loading the SQLite database and a preview of each of the seven tables in the database.

```
library(tidyverse)
library('RSQLite') # SQLite package for R
library(DBI) # R Database Interface.
library(knitr)
library(janitor)
```

```r
#Connect to Database
databaseConnection <- dbConnect(drv=RSQLite::SQLite(), dbname="database.sqlite")

#List Tables
tables <- dbListTables(databaseConnection)

#exclude sqlite_sequence (contains table information)
tables <- tables[tables != "sqlite_sequence"]

lDataFrames <- vector("list", length=length(tables))

#Create Dataframe for each table
for (i in seq(along=tables)) {
  lDataFrames[[i]] <- dbGetQuery(conn=databaseConnection, statement=paste("SELECT * FROM '", tables[[i]]
}

#label all of the dataframes
Country <- lDataFrames[[1]]
League <- lDataFrames[[2]]
Match <- lDataFrames[[3]]
Player <- lDataFrames[[4]]
Player_Attributes <- lDataFrames[[5]]
Team <- lDataFrames[[6]]
Team_Attributes <- lDataFrames[[7]]

dbGetQuery(databaseConnection, "SELECT* FROM league LIMIT 10")
```

```
##       id country_id                      name
## 1      1          1   Belgium Jupiler League
## 2   1729       1729   England Premier League
## 3   4769       4769           France Ligue 1
## 4   7809       7809    Germany 1. Bundesliga
## 5  10257      10257            Italy Serie A
## 6  13274      13274   Netherlands Eredivisie
## 7  15722      15722       Poland Ekstraklasa
## 8  17642      17642 Portugal Liga ZON Sagres
## 9  19694      19694  Scotland Premier League
## 10 21518      21518          Spain LIGA BBVA
```

```r
dbGetQuery(databaseConnection, "SELECT* FROM country LIMIT 10")
```

```
##       id        name
## 1      1     Belgium
## 2   1729     England
## 3   4769      France
## 4   7809     Germany
## 5  10257       Italy
## 6  13274 Netherlands
## 7  15722      Poland
## 8  17642    Portugal
## 9  19694    Scotland
## 10 21518       Spain
```

```r
dbGetQuery(databaseConnection, "SELECT* FROM match LIMIT 10")
```

```
##    id country_id league_id    season stage               date match_api_id
## 1   1          1         1 2008/2009     1 2008-08-17 00:00:00       492473
## 2   2          1         1 2008/2009     1 2008-08-16 00:00:00       492474
## 3   3          1         1 2008/2009     1 2008-08-16 00:00:00       492475
## 4   4          1         1 2008/2009     1 2008-08-17 00:00:00       492476
## 5   5          1         1 2008/2009     1 2008-08-16 00:00:00       492477
## 6   6          1         1 2008/2009     1 2008-09-24 00:00:00       492478
## 7   7          1         1 2008/2009     1 2008-08-16 00:00:00       492479
## 8   8          1         1 2008/2009     1 2008-08-16 00:00:00       492480
## 9   9          1         1 2008/2009     1 2008-08-16 00:00:00       492481
## 10 10          1         1 2008/2009    10 2008-11-01 00:00:00       492564
##    home_team_api_id away_team_api_id home_team_goal away_team_goal
## 1              9987             9993              1              1
## 2             10000             9994              0              0
## 3              9984             8635              0              3
## 4              9991             9998              5              0
## 5              7947             9985              1              3
## 6              8203             8342              1              1
## 7              9999             8571              2              2
## 8              4049             9996              1              2
## 9             10001             9986              1              0
## 10             8342             8571              4              1
##    home_player_X1 home_player_X2 home_player_X3 home_player_X4 home_player_X5
## 1              NA             NA             NA             NA             NA
## 2              NA             NA             NA             NA             NA
## 3              NA             NA             NA             NA             NA
## 4              NA             NA             NA             NA             NA
## 5              NA             NA             NA             NA             NA
## 6              NA             NA             NA             NA             NA
## 7              NA             NA             NA             NA             NA
## 8              NA             NA             NA             NA             NA
## 9              NA             NA             NA             NA             NA
## 10             NA             NA             NA             NA             NA
##    home_player_X6 home_player_X7 home_player_X8 home_player_X9 home_player_X10
## 1              NA             NA             NA             NA              NA
## 2              NA             NA             NA             NA              NA
## 3              NA             NA             NA             NA              NA
## 4              NA             NA             NA             NA              NA
## 5              NA             NA             NA             NA              NA
## 6              NA             NA             NA             NA              NA
## 7              NA             NA             NA             NA              NA
## 8              NA             NA             NA             NA              NA
## 9              NA             NA             NA             NA              NA
## 10             NA             NA             NA             NA              NA
##    home_player_X11 away_player_X1 away_player_X2 away_player_X3 away_player_X4
## 1               NA             NA             NA             NA             NA
## 2               NA             NA             NA             NA             NA
## 3               NA             NA             NA             NA             NA
## 4               NA             NA             NA             NA             NA
## 5               NA             NA             NA             NA             NA
## 6               NA             NA             NA             NA             NA
## 7               NA             NA             NA             NA             NA
## 8               NA             NA             NA             NA             NA
## 9               NA             NA             NA             NA             NA
```

```
## 10              NA             NA             NA             NA             NA
##    away_player_X5 away_player_X6 away_player_X7 away_player_X8 away_player_X9
## 1              NA             NA             NA             NA             NA
## 2              NA             NA             NA             NA             NA
## 3              NA             NA             NA             NA             NA
## 4              NA             NA             NA             NA             NA
## 5              NA             NA             NA             NA             NA
## 6              NA             NA             NA             NA             NA
## 7              NA             NA             NA             NA             NA
## 8              NA             NA             NA             NA             NA
## 9              NA             NA             NA             NA             NA
## 10             NA             NA             NA             NA             NA
##    away_player_X10 away_player_X11 home_player_Y1 home_player_Y2 home_player_Y3
## 1               NA              NA             NA             NA             NA
## 2               NA              NA             NA             NA             NA
## 3               NA              NA             NA             NA             NA
## 4               NA              NA             NA             NA             NA
## 5               NA              NA             NA             NA             NA
## 6               NA              NA             NA             NA             NA
## 7               NA              NA             NA             NA             NA
## 8               NA              NA             NA             NA             NA
## 9               NA              NA             NA             NA             NA
## 10              NA              NA             NA             NA             NA
##    home_player_Y4 home_player_Y5 home_player_Y6 home_player_Y7 home_player_Y8
## 1              NA             NA             NA             NA             NA
## 2              NA             NA             NA             NA             NA
## 3              NA             NA             NA             NA             NA
## 4              NA             NA             NA             NA             NA
## 5              NA             NA             NA             NA             NA
## 6              NA             NA             NA             NA             NA
## 7              NA             NA             NA             NA             NA
## 8              NA             NA             NA             NA             NA
## 9              NA             NA             NA             NA             NA
## 10             NA             NA             NA             NA             NA
##    home_player_Y9 home_player_Y10 home_player_Y11 away_player_Y1 away_player_Y2
## 1              NA              NA             NA             NA             NA
## 2              NA              NA             NA             NA             NA
## 3              NA              NA             NA             NA             NA
## 4              NA              NA             NA             NA             NA
## 5              NA              NA             NA             NA             NA
## 6              NA              NA             NA             NA             NA
## 7              NA              NA             NA             NA             NA
## 8              NA              NA             NA             NA             NA
## 9              NA              NA             NA             NA             NA
## 10             NA              NA             NA             NA             NA
##    away_player_Y3 away_player_Y4 away_player_Y5 away_player_Y6 away_player_Y7
## 1              NA             NA             NA             NA             NA
## 2              NA             NA             NA             NA             NA
## 3              NA             NA             NA             NA             NA
## 4              NA             NA             NA             NA             NA
## 5              NA             NA             NA             NA             NA
## 6              NA             NA             NA             NA             NA
## 7              NA             NA             NA             NA             NA
## 8              NA             NA             NA             NA             NA
```

```
## 9              NA             NA             NA             NA             NA
## 10             NA             NA             NA             NA             NA
##    away_player_Y8 away_player_Y9 away_player_Y10 away_player_Y11 home_player_1
## 1              NA             NA              NA              NA            NA
## 2              NA             NA              NA              NA            NA
## 3              NA             NA              NA              NA            NA
## 4              NA             NA              NA              NA            NA
## 5              NA             NA              NA              NA            NA
## 6              NA             NA              NA              NA            NA
## 7              NA             NA              NA              NA            NA
## 8              NA             NA              NA              NA            NA
## 9              NA             NA              NA              NA            NA
## 10             NA             NA              NA              NA            NA
##    home_player_2 home_player_3 home_player_4 home_player_5 home_player_6
## 1             NA            NA            NA            NA            NA
## 2             NA            NA            NA            NA            NA
## 3             NA            NA            NA            NA            NA
## 4             NA            NA            NA            NA            NA
## 5             NA            NA            NA            NA            NA
## 6             NA            NA            NA            NA            NA
## 7             NA            NA            NA            NA            NA
## 8             NA            NA            NA            NA            NA
## 9             NA            NA            NA            NA            NA
## 10            NA            NA            NA            NA            NA
##    home_player_7 home_player_8 home_player_9 home_player_10 home_player_11
## 1             NA            NA            NA             NA             NA
## 2             NA            NA            NA             NA             NA
## 3             NA            NA            NA             NA             NA
## 4             NA            NA            NA             NA             NA
## 5             NA            NA            NA             NA             NA
## 6             NA            NA            NA             NA             NA
## 7             NA            NA            NA             NA             NA
## 8             NA            NA            NA             NA             NA
## 9             NA            NA            NA             NA             NA
## 10            NA            NA            NA             NA             NA
##    away_player_1 away_player_2 away_player_3 away_player_4 away_player_5
## 1             NA            NA            NA            NA            NA
## 2             NA            NA            NA            NA            NA
## 3             NA            NA            NA            NA            NA
## 4             NA            NA            NA            NA            NA
## 5             NA            NA            NA            NA            NA
## 6             NA            NA            NA            NA            NA
## 7             NA            NA            NA            NA            NA
## 8             NA            NA            NA            NA            NA
## 9             NA            NA            NA            NA            NA
## 10            NA            NA            NA            NA            NA
##    away_player_6 away_player_7 away_player_8 away_player_9 away_player_10
## 1             NA            NA            NA            NA             NA
## 2             NA            NA            NA            NA             NA
## 3             NA            NA            NA            NA             NA
## 4             NA            NA            NA            NA             NA
## 5             NA            NA            NA            NA             NA
## 6             NA            NA            NA            NA             NA
## 7             NA            NA            NA            NA             NA
```

```
## 8             NA            NA            NA            NA            NA
## 9             NA            NA            NA            NA            NA
## 10            NA            NA            NA            NA            NA
##    away_player_11 goal shoton shotoff foulcommit card cross corner possession
## 1              NA <NA>   <NA>    <NA>       <NA> <NA>  <NA>   <NA>       <NA>
## 2              NA <NA>   <NA>    <NA>       <NA> <NA>  <NA>   <NA>       <NA>
## 3              NA <NA>   <NA>    <NA>       <NA> <NA>  <NA>   <NA>       <NA>
## 4              NA <NA>   <NA>    <NA>       <NA> <NA>  <NA>   <NA>       <NA>
## 5              NA <NA>   <NA>    <NA>       <NA> <NA>  <NA>   <NA>       <NA>
## 6              NA <NA>   <NA>    <NA>       <NA> <NA>  <NA>   <NA>       <NA>
## 7              NA <NA>   <NA>    <NA>       <NA> <NA>  <NA>   <NA>       <NA>
## 8              NA <NA>   <NA>    <NA>       <NA> <NA>  <NA>   <NA>       <NA>
## 9              NA <NA>   <NA>    <NA>       <NA> <NA>  <NA>   <NA>       <NA>
## 10             NA <NA>   <NA>    <NA>       <NA> <NA>  <NA>   <NA>       <NA>
##    B365H B365D B365A  BWH  BWD   BWA  IWH IWD IWA  LBH  LBD   LBA PSH PSD PSA
## 1   1.73  3.40  5.00 1.75 3.35  4.20 1.85 3.2 3.5 1.80 3.30  3.75  NA  NA  NA
## 2   1.95  3.20  3.60 1.80 3.30  3.95 1.90 3.2 3.5 1.90 3.20  3.50  NA  NA  NA
## 3   2.38  3.30  2.75 2.40 3.30  2.55 2.60 3.1 2.3 2.50 3.20  2.50  NA  NA  NA
## 4   1.44  3.75  7.50 1.40 4.00  6.80 1.40 3.9 6.0 1.44 3.60  6.50  NA  NA  NA
## 5   5.00  3.50  1.65 5.00 3.50  1.60 4.00 3.3 1.7 4.00 3.40  1.72  NA  NA  NA
## 6   4.75  3.40  1.67 4.85 3.40  1.65 3.70 3.2 1.8 5.00 3.25  1.62  NA  NA  NA
## 7   2.10  3.20  3.30 2.05 3.25  3.15 1.85 3.2 3.5 1.83 3.30  3.60  NA  NA  NA
## 8   3.20  3.40  2.20 2.55 3.30  2.40 2.40 3.2 2.4 2.50 3.20  2.50  NA  NA  NA
## 9   2.25  3.25  2.88 2.30 3.25  2.70 2.10 3.1 3.0 2.25 3.20  2.75  NA  NA  NA
## 10  1.30  5.25  9.50 1.25 5.00 10.00 1.30 4.2 8.0 1.25 4.50 10.00  NA  NA  NA
##     WHH  WHD  WHA  SJH  SJD   SJA  VCH  VCD  VCA  GBH  GBD   GBA  BSH  BSD  BSA
## 1  1.70 3.30 4.33 1.90 3.30  4.00 1.65 3.40 4.50 1.78 3.25  4.00 1.73 3.40 4.20
## 2  1.83 3.30 3.60 1.95 3.30  3.80 2.00 3.25 3.25 1.85 3.25  3.75 1.91 3.25 3.60
## 3  2.50 3.25 2.40 2.63 3.30  2.50 2.35 3.25 2.65 2.50 3.20  2.50 2.30 3.20 2.75
## 4  1.44 3.75 6.00 1.44 4.00  7.50 1.45 3.75 6.50 1.50 3.75  5.50 1.44 3.75 6.50
## 5  4.20 3.40 1.70 4.50 3.50  1.73 4.50 3.40 1.65 4.50 3.50  1.65 4.75 3.30 1.67
## 6  4.20 3.40 1.70 5.50 3.75  1.67 4.35 3.40 1.70 4.50 3.40  1.70   NA   NA   NA
## 7  1.83 3.30 3.60 1.91 3.40  3.60 2.10 3.25 3.00 1.85 3.25  3.75 2.10 3.25 3.10
## 8  2.70 3.25 2.25 2.60 3.40  2.40 2.80 3.25 2.25 2.80 3.20  2.25 2.88 3.25 2.20
## 9  2.20 3.25 2.75 2.20 3.30  3.10 2.25 3.25 2.80 2.20 3.30  2.80 2.25 3.20 2.80
## 10 1.35 4.20 7.00 1.27 5.00 10.00 1.30 4.35 8.50 1.25 5.00 10.00 1.29 4.50 9.00
```

```
dbGetQuery(databaseConnection, "SELECT* FROM team attributes LIMIT 10")
```

```
##    id team_api_id team_fifa_api_id    team_long_name team_short_name
## 1   1        9987              673          KRC Genk             GEN
## 2   2        9993              675      Beerschot AC             BAC
## 3   3       10000            15005  SV Zulte-Waregem             ZUL
## 4   4        9994             2007  Sporting Lokeren             LOK
## 5   5        9984             1750 KSV Cercle Brugge             CEB
## 6   6        8635              229    RSC Anderlecht             AND
## 7   7        9991              674          KAA Gent             GEN
## 8   8        9998             1747         RAEC Mons             MON
## 9   9        7947               NA     FCV Dender EH             DEN
## 10 10        9985              232 Standard de Liège             STL
```

```
dbGetQuery(databaseConnection, "SELECT* FROM player attributes LIMIT 10")
```

```
##    id player_api_id        player_name player_fifa_api_id            birthday
## 1   1        505942 Aaron Appindangoye             218353 1992-02-29 00:00:00
```

```
## 2   2        155782    Aaron Cresswell             189615 1989-12-15 00:00:00
## 3   3        162549        Aaron Doran             186170 1991-05-13 00:00:00
## 4   4         30572      Aaron Galindo             140161 1982-05-08 00:00:00
## 5   5         23780       Aaron Hughes              17725 1979-11-08 00:00:00
## 6   6         27316         Aaron Hunt             158138 1986-09-04 00:00:00
## 7   7        564793         Aaron Kuhl             221280 1996-01-30 00:00:00
## 8   8         30895       Aaron Lennon             152747 1987-04-16 00:00:00
## 9   9        528212       Aaron Lennox             206592 1993-02-19 00:00:00
## 10 10        101042      Aaron Meijers             188621 1987-10-28 00:00:00
##    height weight
## 1  182.88    187
## 2  170.18    146
## 3  170.18    163
## 4  182.88    198
## 5  182.88    154
## 6  182.88    161
## 7  172.72    146
## 8  165.10    139
## 9  190.50    181
## 10 175.26    170
```

```
dbGetQuery(databaseConnection, "SELECT* FROM team LIMIT 10")
```

```
##    id team_api_id team_fifa_api_id    team_long_name team_short_name
## 1   1        9987              673          KRC Genk             GEN
## 2   2        9993              675      Beerschot AC             BAC
## 3   3       10000            15005  SV Zulte-Waregem             ZUL
## 4   4        9994             2007  Sporting Lokeren             LOK
## 5   5        9984             1750 KSV Cercle Brugge             CEB
## 6   6        8635              229    RSC Anderlecht             AND
## 7   7        9991              674          KAA Gent             GEN
## 8   8        9998             1747         RAEC Mons             MON
## 9   9        7947               NA     FCV Dender EH             DEN
## 10 10        9985              232 Standard de Liège             STL
```

```
dbGetQuery(databaseConnection, "SELECT* FROM player LIMIT 10")
```

```
##    id player_api_id        player_name player_fifa_api_id            birthday
## 1   1        505942 Aaron Appindangoye             218353 1992-02-29 00:00:00
## 2   2        155782    Aaron Cresswell             189615 1989-12-15 00:00:00
## 3   3        162549        Aaron Doran             186170 1991-05-13 00:00:00
## 4   4         30572      Aaron Galindo             140161 1982-05-08 00:00:00
## 5   5         23780       Aaron Hughes              17725 1979-11-08 00:00:00
## 6   6         27316         Aaron Hunt             158138 1986-09-04 00:00:00
## 7   7        564793         Aaron Kuhl             221280 1996-01-30 00:00:00
## 8   8         30895       Aaron Lennon             152747 1987-04-16 00:00:00
## 9   9        528212       Aaron Lennox             206592 1993-02-19 00:00:00
## 10 10        101042      Aaron Meijers             188621 1987-10-28 00:00:00
##    height weight
## 1  182.88    187
## 2  170.18    146
## 3  170.18    163
## 4  182.88    198
## 5  182.88    154
## 6  182.88    161
```

```
## 7  172.72    146
## 8  165.10    139
## 9  190.50    181
## 10 175.26    170
```

```
#Used code from https://stackoverflow.com/questions/9802680/how-to-import-from-sqlite-database to creat
```

4. Methods/Instructional Content
   The two instructional content we picked were Introduction to Statistical Investigations, 2nd Edition and Intermediate Statistical Investigations, 1st Edition.
   Scholarly reference 1: Introduction to Statistical Investigations (Chapter 8: Comparing more than two proportions)
   This chapter provides theoretical knowledge of the key components of our module as the chapter includes all basic information. We used this reference to refresh our knowledge of the process of comparing multiple proportions. It also includes information on generalization and causation.
   Scholarly reference 2: Intermediate Statistical Investigations (Section 6.1 Comparing Proportions)
   This scholarly reference provides a detailed explanation of the different methods to compare proportions, including the two-sample t-testt. This source was helpful to the process of developing the model as it provides multiple examples of various cases of statistical investigations involving categorical datasets. By reading through the examples, we were able to develop the module by following the general question/exercise format as the examples, as it provides a very well-developed flow to guide readers through a problem.

5. Exercises/Activities
   The following are main topics related to the module. We will assign data exploration projects to practice the following skills to familiarize the students with the topic and ensure their success in this module.

- Ask a research question: Is there an association between preferred foot and free kick accuracy score in the European Soccer League?
  Null Hypothesis: There is no association between preferred foot and free kick accuracy score in the European Soccer League
  Alternative Hypothesis: There is an association between preferred foot and free kick accuracy score in the European Soccer League.

- Design study and collect data: Done earlier in Data section

- Explore the Data

```
#Data frame with some interesting attributes
select_player_attributes <- Player_Attributes %>%
  drop_na() %>%
  select(preferred_foot, ball_control, free_kick_accuracy, overall_rating)

#Table to show interesting attributes
select_player_attributes %>%
  group_by(preferred_foot) %>%
  summarize(mean_BC = mean(ball_control), mean_FCA = mean(free_kick_accuracy), mean_OR = mean(overall_ra
  kable(col.names = c("Preferred Foot", "Mean Ball Control Score", "Mean Free Kick Accuracy", "Mean Rati
```

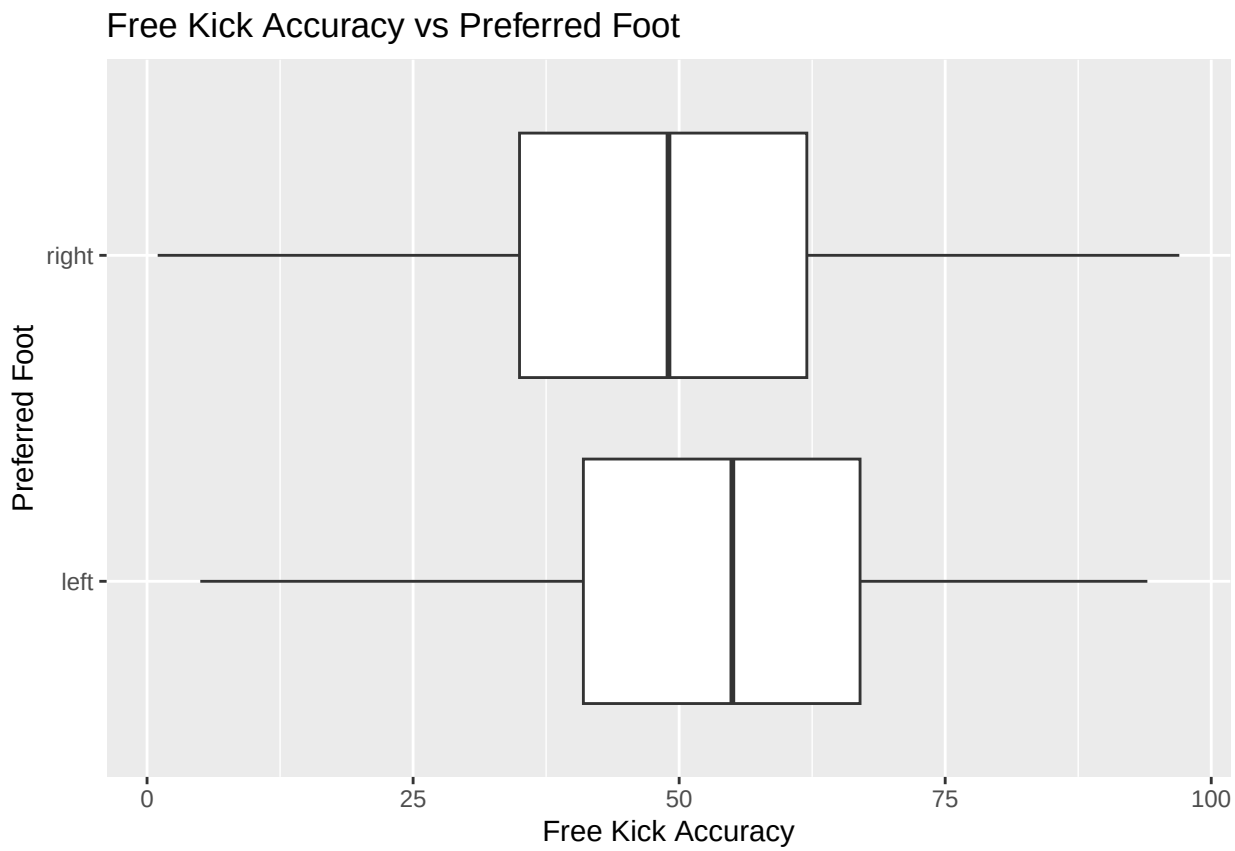
Preferred Foot	Mean Ball Control Score	Mean Free Kick Accuracy	Mean Rating	Number of Observations
left	65.44723	53.29136	68.65282	44107
right	62.80853	48.13070	68.62965	136247

Preferred Foot	Mean Ball Control Score	Mean Free Kick Accuracy	Mean Rating	Number of Observations

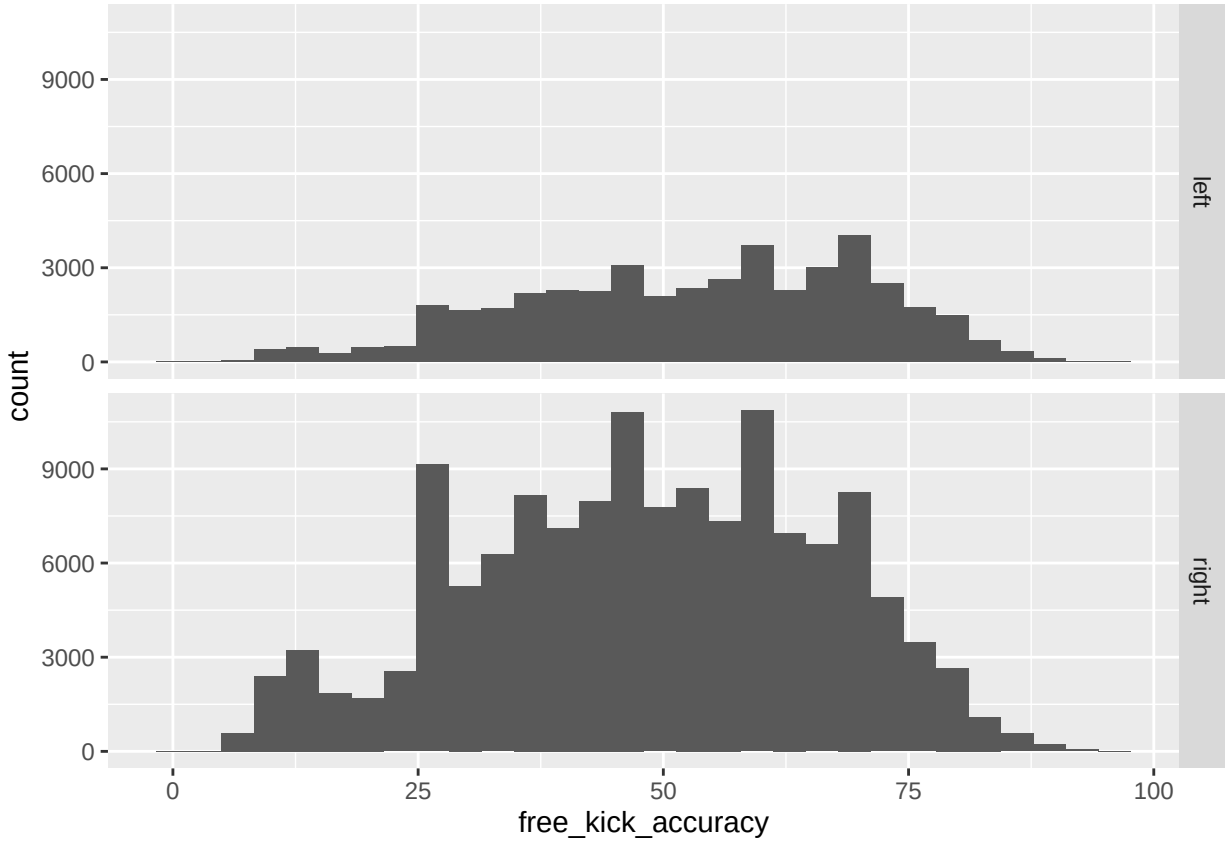
```
#Five number summary of each of the two populations
select_player_attributes %>%
  group_by(preferred_foot) %>%
  summarize(Minimum = min(free_kick_accuracy),
            LowerQuartile = quantile(prob =.25, free_kick_accuracy),
            Median = median(free_kick_accuracy),
            UpperQuartile = quantile(prob=.75, free_kick_accuracy),
            Maximum = max(free_kick_accuracy))
```

```
## # A tibble: 2 x 6
##   preferred_foot Minimum LowerQuartile Median UpperQuartile Maximum
##   <chr>            <int>         <dbl>  <int>         <dbl>   <int>
## 1 left                 5            41     55            67      94
## 2 right                1            35     49            62      97
```

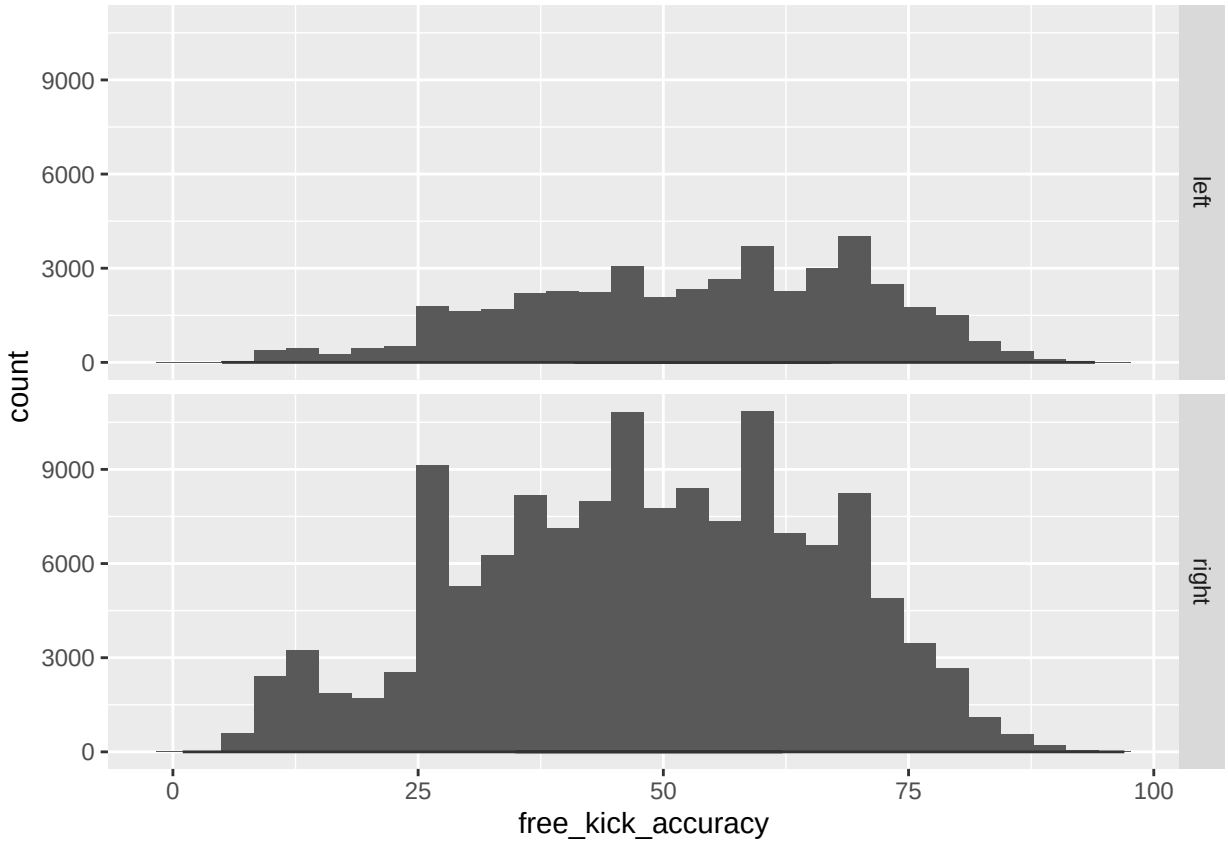
```
#Boxplot to illustrate the five-number summary
select_player_attributes %>%
  ggplot(aes(x = free_kick_accuracy,
             y = preferred_foot)) +
  geom_boxplot()+
  labs(y = "Preferred Foot",
       x = "Free Kick Accuracy",
       title = "Free Kick Accuracy vs Preferred Foot")
```



```
#Histogram to visualize data
select_player_attributes %>%
  ggplot(aes(x=free_kick_accuracy)) +
  geom_histogram() +
  facet_grid(preferred_foot~.)
```




```
#combination of boxplot and histogram
select_player_attributes %>%
  ggplot(aes(x=free_kick_accuracy)) +
  geom_histogram() +
  geom_boxplot() +
  facet_grid(preferred_foot ~.)
```



- Conduct Two Sample T-Test

```
#Look at Data
select_player_attributes %>%
  group_by(preferred_foot) %>%
  summarise(xbar = mean(free_kick_accuracy),
            s = sd(free_kick_accuracy),
            n = n())
```

```
## # A tibble: 2 x 4
##   preferred_foot  xbar     s      n
##   <chr>          <dbl> <dbl>  <int>
## 1 left            53.3  17.3  44107
## 2 right           48.1  17.8 136247
```

```
#Calculate Standardized Statistics
xbar_left = 53.291
xbar_right = 48.131
s_left = 17.325
s_right = 17.796
n_left = 44107
n_right = 136247
sd = sqrt(s_left^2/n_left+s_right^2/n_right)
null = 0
statistic = xbar_left-xbar_right
t = (statistic-null)/sd

#Calculate P-Value with n-2 degrees of freedom, two-tailed test
```

```r
n = n_left+n_right
pvalue = 2*pt(t,n-2, lower.tail = FALSE)
pvalue
```

```
## [1] 0
```

```r
#calculate Confidence interval at 99% confidence
multiplier = qt(.995,n-2)
se = sd
CI = c(statistic - multiplier*se, statistic + multiplier*se)
CI
```

```
## [1] 4.91388 5.40612
```

5. Wrap-Up/Conclusions:

Conclusions: Through an exploration of the data and the usage of a two-sample t-test to analyze the data, we were able to reject the null hypothesis that there is no difference in free kick accuracy score between right-footed and left-footed players and show evidence for the alternative that there is a difference in average free-kick accuracy between these two groups.
The p-value of 0 is less than our significance level of 0.01, and it shows that it is extremely unlikely that the observed difference between these two groups was due to chance alone. The confidence interval calculated shows that we can be 99% confident that the actual difference in free kick accuracy score between right-footed and left-footed players is between the values of 4.91388 and 5.40612, with left-footed kickers performing better.

Wrap-Up:
Recap the lessons learned both in terms of statistical techniques and in terms of the sports research question. Provide the reader at least one other sports application (could be the same sport) for this particular skill and at least one idea for a future skill to learn that builds on what you've presented.
In this lesson, students learned to conduct a statistical investigation by following the six-step statistical investigation method. The module targeted to answer the research question of whether there exists a correlation between dominant foot and free kick accuracy in the European Soccer League by comparing two population proportions using a two-sample t-test.