

# Class Test 2561551

Jinda(LousiY) Zhang

## Introduction

In the sport of rugby, there are players playing different positions, and it is of interest to see each position's trait such as speed, strength and accuracy. 89 backward position and forward position participates in the experiments and they each had their height in centimeter measured. The data including the position of each player, is analyzed in this report.

This report presents numerical and graphical summaries of the height of the player positions and fits a linear model to estimate the difference, on average, between the height of forward and back positions.

## Exploratory Data Analysis

Table 1: Summary statistics on height by position of 89 players.

position	n	Mean	St.Dev	Min	Q1	Median	Q3	Max
Back	33	182.7	4.4	173	180	183	185	193
Forward	56	181.2	3.6	173	180	181	183	190

Table 1 shows that there were more forward than back in the experiment(56 compared to 33) and that the summaries of the height of the back were greater than the corresponding summaries of the height of back. For example the mean height of the back were 182.7 cm compared to 181.2 cm for the mean height of forward. We also find that the variability in the back's height, as measured by the standard deviation of 4.4 cm, was larger than the standard deviation of 3.6 cm for forward's height. These differences also be found in the following boxplot that summarise the distribution of the height of back and forward cat.

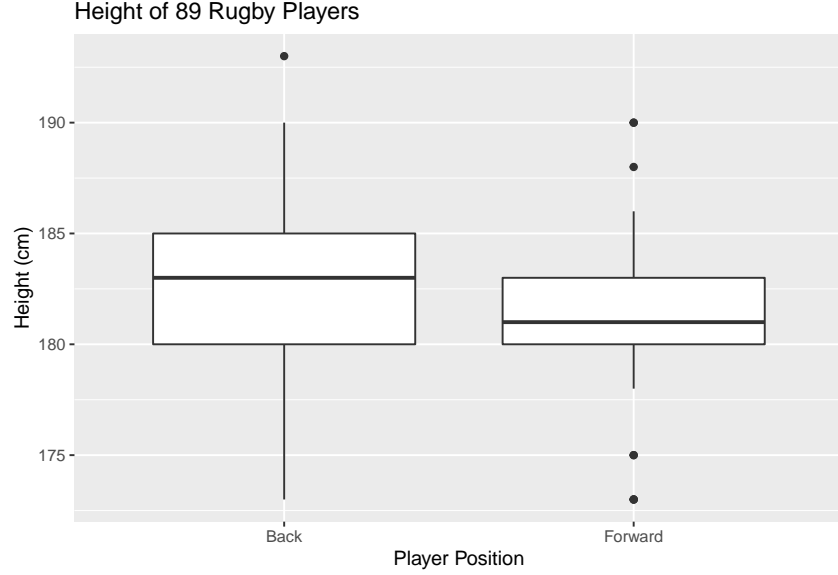


Figure 1: Height by position.

The boxplot shows that the back having larger height compared to the forward height and the height of the back are distributed more widely. There are also potentially 5 outliers(1 back and 4 forward) which have unusually height.

## Formal Data Analysis

To analyze the height of rugby player formally, we start by fitting the following model to the data.

$$\widehat{\text{Height}} = \hat{\alpha} + \hat{\beta}_{\text{Forward}} + \mathbb{I}_{\text{Forward}}(x)$$

where

- the intercept  $\hat{\alpha}$  is the mean height for the baseline category of Back;
- $\hat{\beta}_{\text{Forward}}$  is the difference in the mean height of a Forward relative to the baseline category Back; and
- $\mathbb{I}_{\text{Forward}}(x)$  is an indicator function s.t.

$$\mathbb{I}_{\text{Forward}}(x) = \begin{cases} 1 & \text{if Position of } x\text{th observation is Forward,} \\ 0 & \text{Otherwise.} \end{cases}$$

while the model is fitted to the data, the following estimate of  $\alpha$ (intercept) and  $\beta_{\text{Forward}}$ (positionForward) are returned:

Table 2: Estimates of the parameters from the fitted linear regression model.

term	estimate
intercept	182.697
positionForward	-1.447

As a result, the model estimates the average height of back position in 182.697cm and that the Forward's height are on average 1.447cm lower than the Back's height.

## Conclusions

In summary, we estimate that the forward position have height which weight 1.447cm less than the back position on average. Also, we estimate the average height of back position in 182.697cm and the average height of forward position is 181.2cm.

In addition to this, the centers of distributions of forward and backward position' height are different. We also find that the spread of the back height is greater than the spread of the forward's height. This may be probmatic if we want to use standard linear model to further analyze this data. As a result, we may want to use the models that allow for difference in variance within different group.



## FURTHER TASK

### Part a

```
setwd("/Users/kurisuuu/Downloads/Class\ Test\ Files-20210610")
rugby_full <- read_csv("rugby_full.csv")
```

```
rugby_full <- rugby_full %>%
  mutate(games_per_year = caps/years_playing)
```

### Part b

Yes, because each variable forms a column. Each observation forms a row. Each type of observational unit forms a table.

### Part c

```
ggplot(rugby_full, mapping=aes(x=caps, y=years_playing, group=position, color=position, shape=position))+
  geom_point()+
  facet_wrap(~team)+
  labs(x="Number of international appearances", y="number of years playing", title="number of international appearances")
```

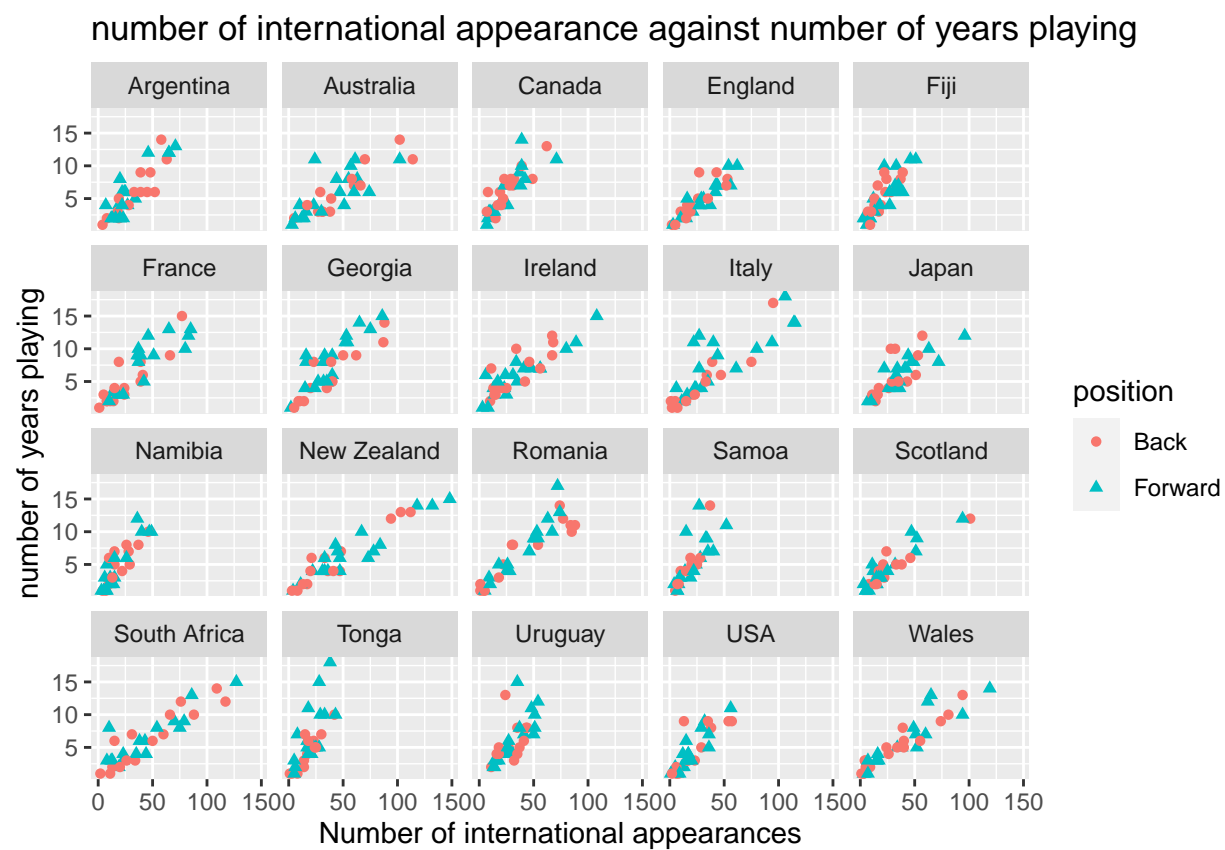


Figure 2: number of international appearance against number of years playing by position

## Part d

```
rugby_bri<- rugby_full %>%  
  filter(team == "England"|team=="Ireland"|team=="Scotland"|team=="Wales")  
  
ggplot(rugby_bri,mapping=aes(x=height_cm,y=weight_kg,group=position,color=position,shape=position))+  
  geom_point()+  
  labs(x="Height (cm)",y="Weight (kg)",title="Height and weight of players in all British Isles teams")  
  geom_smooth(method = "lm", se = FALSE, aes(color=position, linetype=position))
```

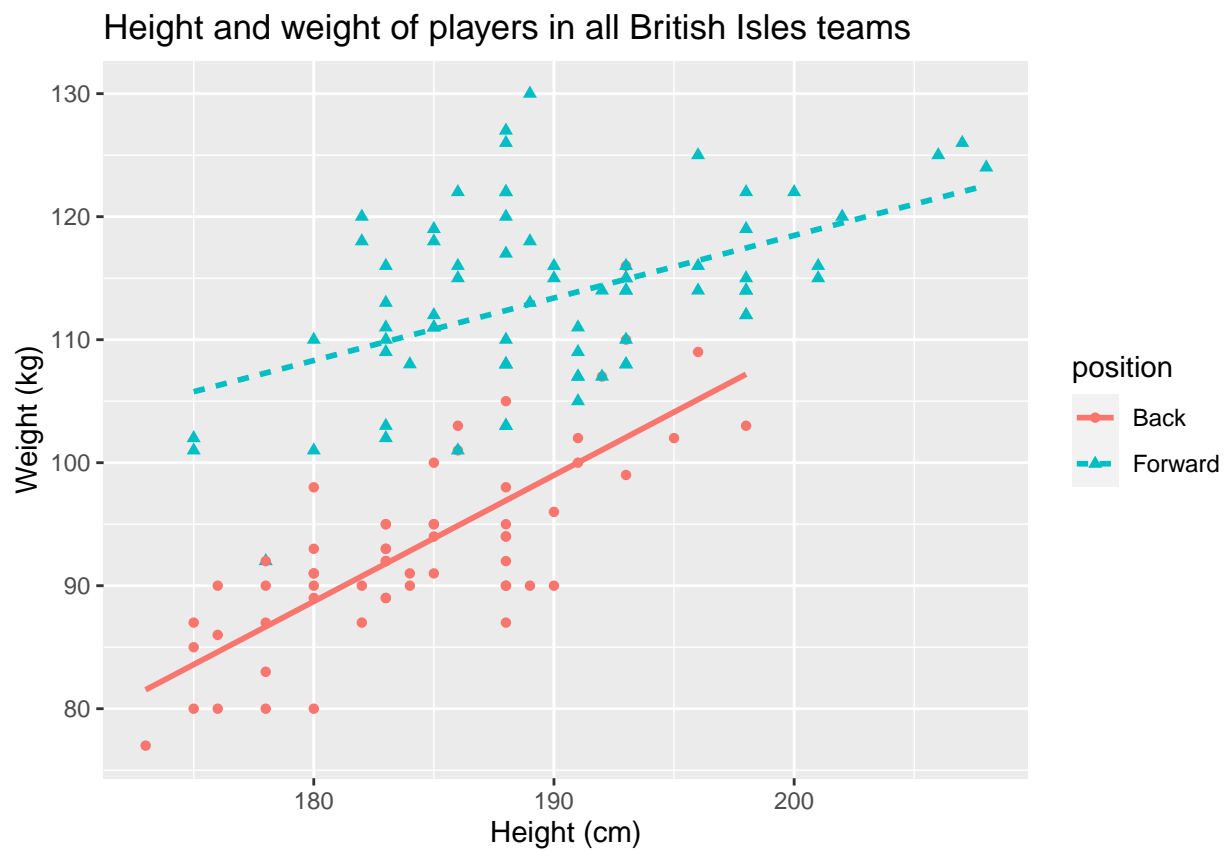


Figure 3: Height and weight of players in all British Isles teams by position