Analysis of the heart weights of cats

Student Number: 2516212c

# Introduction

In the sport of rugby there are many different positions a player could play but, in general, players are classified into one of two playing positions, namely ??forwards?? and ??backs??. From the data collected on rugby players from the 2015 Men??s Rugby World Cup, we want to figure out the relationship between position and heights. 89 rugby players were used in the research and they each had their own position and heights(in cm). This data is analysed in this report.

# Exploratory Data Analysis

Summary statistics of the heights of the rugby players are presented in the following table for each kind of position separately.

Summary statistics on height(in cm) by position of 89 rugby players.

position

n

Mean

St.Dev

Min

Q1

Median

Q3

Max

Back

33

183

4.4

173

180

183

185

193

Forward

56

181

3.6

173

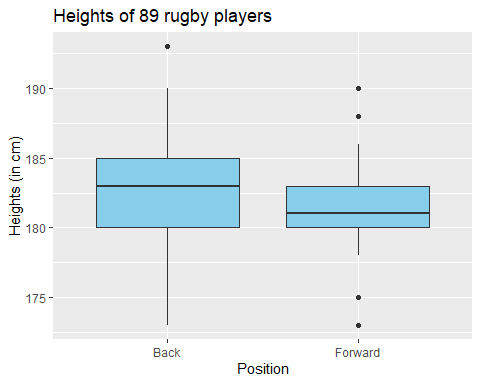
180

181

183

190

This table shows that there were almost twice as many rugby players in the sample (56 compared to 33) and that the summarises of the heights of the rugby players which position is ‘Back’ seem has no remarkable difference with these rugby players who played ‘Forward’. But from the median, the summarises of ‘Back’ was greater than ‘Forward’ although the difference was not large. For example the maximum of ‘Back’ was taller than the maximum of ‘Forward’ 3 cm and from the Standard deviation, we could see that the spread of ‘Back’ seems widely than the spread of ‘Forward’. These differences can be seen directly in the following boxplots which summarise the distribution of the heights of each kind of players.

 The boxplot shows that the rugby players’ heights of ‘Back’ is larger than ‘Forward’, in general, compared to the players’ heights of ‘Forward’ and that the players’ heights of ‘Back’ were more widely distributed. There are also potentially five outliers and most of them belong to ‘Forward’ which have special heights(quite not same with the mean), as shown by these points far away from the “whiskers” of the boxplots.

# Formal Data Analysis

To begin to analyse the heights of players data formally, we fit the following linear model to the data.

Estimates of the parameters from the fitted linear regression model.

term

estimate

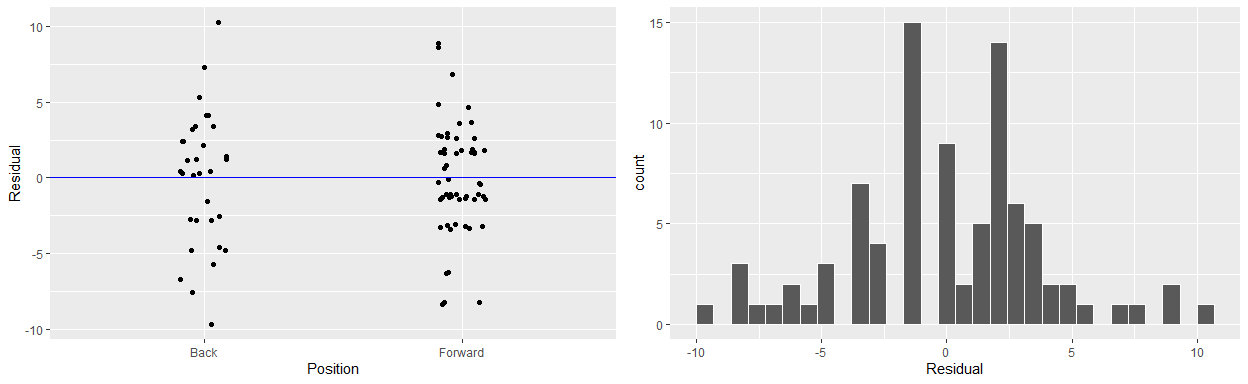
intercept

182.697

positionForward

-1.447

Hence the model estimates the average height of players in ‘Back’ is 182.697 cms (which agrees with the sample mean reported and that the ‘Forward’ players’ heights are, on average, 1.447 smaller than the ‘Back’ players’ heights.).

Before we can proceed to use the fitted model(for example to perform statistical inference) we must check the assumptions of the model. These are best considered in light of the residual plots in Figure 2.  The scatterplots show an approximately even spread of the residuals above and below the zero line for each position, therefore the assumption that the residuals have mean zero seems valid and the assumption that constant variance within the two positions seems available too in terms of the first scatterplot. The histogram shows the distribution of residuals are nomally distributed errors in the model, which the exception of a potential outliers.

Where is the expected value of the heart weight of the th cat in the sample; the intercept is the mean heart weight for the baseline category of females; is the difference in the mean heart weight of a males relatively to the baseline category females; is an indicator function such that

When this model is fitted to the data, the following estimates of (intercept) and are returned:

# Conclusions

In summary, we have estimated that, on averagem the male cats have hearts which weigh 2.121 grams more than the female cats’ hearts. In particular, we estimate the average heart weight of female cats is 9.202 grams and the average heart weight of male cats is 11.3 grams.

In addition to the centers of the distributions of male and female cats’ heart weights being different, we have also observed that the spread of the male heart weights is greater than the spread of the female cats’ heart weights. This may pose a problem if the standard linear model was used to further analyse this data, and therefore it is recommended that models which allow for differences in the variances with in different groups be used.

# FURTHER TASK