Report Title

Student Number

Introduction

#Introduction to the data being anaysed and to the question of interest.Marks deducted for copying the data description as given.

Experiments were conducted as part of research into "Digitalis", a heart medicine similar to toxins found in plants commonly known as foxglove. 144 domestic male and female adult cats were used in the experiments and they each had their heart weight in grams (Hwt) measured. This data, including the sex (Sex) of each cat, is analyzed in this report.

In particular, this report presents numerical and graphical summaries of the heart weights of the cats and fits a linear model to estimate the difference, on average, between the heart weights of male and female cats.

Exploratory Data Analysis

Summary statistics on heart weithgt by sex with appropriate comments.One mark removed if the output is simply copyed from R.

Table 1: Summary statistics on heart weight by sex of 144 adult cats.

Sex	n	Mean	St.Dev	Min	Q1	Median	Q3	Max
F	47	9.2	1.4	6.3	8.35	9.1	10.1	13.0
M	97	11.3	2.5	6.5	9.40	11.4	12.8	20.5

Comment:this table shows that there were apporximately twice as many male cats in the sample(97 compared to 47) and that the summaries of the heart weights of the male cats were consistently greater than the correponding summaries of the heart weight of female cats. For example the mean heart weight of the male cats were 11.3 grams compared to 9.2 grams for the mean heart weight of cats. We also note that the variability in the males hearts' weight, as measured by the standard deviation of 2.5 grams, was nearly twice as much as the standard deviation of 1.4 grams for the female hearts' weigths. These differences can be eaaily seen the in the following boxplot which summarise the distribution of the heart weight of male and female cats.

#Boxplot of heart weight by sex. One mark removed if the plot is not appropriately labelled, and axis labels not adjusted accordingly.

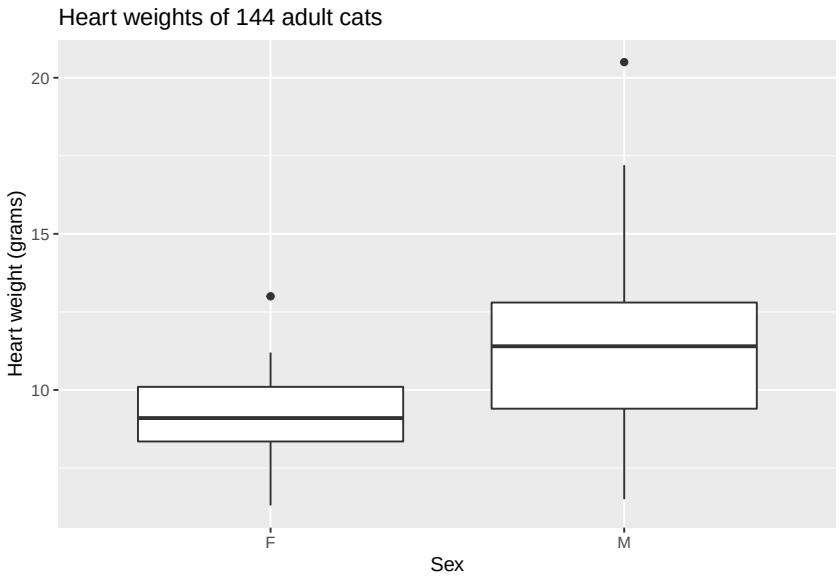

Figure 1: Heart weight by Sex.

Comment: the boxplot shows that the male cats having heavier hearts, in general, compared to the female cats' hearts and the weights of the male hearts were more widely distributed. There are also potentially two outliers(one male and one female) which have unusually heavy hearts, as shown by the two points beyond the whiskers of the boxplot.

Formal Data Analysis

Comment: To begin to analyse the cat heart weights data formally, we fit the following linear model to the data.

$$\widehat{\text{Hwt}} = \widehat{\alpha} + \widehat{\beta}_{\text{Male}} + \cdot\mathbb{I}_{\text{Male}}(x)$$

where

- the intercept $\widehat{\alpha}$ is the mean heart weight for the baseline category of Females;
- $\widehat{\beta}_{\text{Male}}$ is the difference in the mean heart weight of a Males relative to the baseline category Females; and
- $\mathbb{I}_{\text{Male}}(x)$ is an indicator function such that
-

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

Comment: when this model is fitted to the data, the following estimate of α(intercept) and β_{Male}(SexM) are returned:

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

term	estimate
intercept	9.202
SexM	2.121

Comment: Hence the model estimates the average heart weight of female cats in 9.202 grams(which agree with the sample mean reported in table 1) and that the male cats' heart weight are, on average, 2.121 grams heavier than the female cats' heart weights.

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.

NB: THE DIAGNOSTICS IN THE REMAINDER OF THIS ANALYSIS SECTION ARE NOT REQUIRED FOR THE CLASS TEST

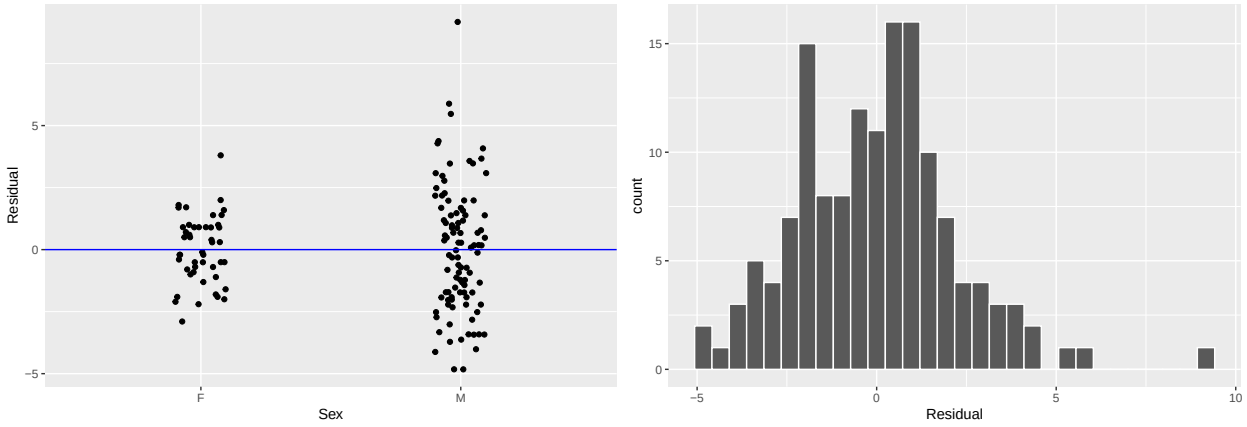

Figure 2: Scatterplots of the residuals by Sex (left) and a histogram of the residuals (right).

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

In addition to the centers of 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 difference in variance within different group is used.

FURTHER TASK

```
setwd("/Users/kurisuuu/Documents/glasgow_stats_2021/semester\ 2/das/Practice\ Class\ Test\ Files-202106
Glasgow_Ed_SIMD2020 <- read.csv("Glasgow_Edinburgh_SIMD2020.csv")
```

The data is not in tidy format since the measurement of interest is rank which there are 8 types spread over 8 columns.
In tidy format the rank measurements should be in a single column, with a seperate column indicating the type of rank.

To convert the data into a tidy format, use

```
Glasgow_Ed_SIMD2020_tidy2 <- gather(data = Glasgow_Ed_SIMD2020,
key = Type_of_Rank,
value = Rank,
-(Data_Zone:Working_Age_population))
Glasgow_Ed_SIMD2020_tidy2$Type_of_Rank <-
str_replace(Glasgow_Ed_SIMD2020_tidy2$Type_of_Rank, "_Rank", "")
```

```
Gla_Ed_SIMD2020 <- Glasgow_Ed_SIMD2020_tidy2 %>%
filter(Type_of_Rank == "SIMD") %>% #2 MARKS
mutate(Perc_Working = 100 *Working_Age_population/Total_population)

ggplot(Gla_Ed_SIMD2020)+
geom_point(mapping=aes(x=Perc_Working,y=Rank,group=Council_area,color=Council_area))+ #3 MARKS
labs(x="Employment Rate of Working Age Population",y="SIMD2020 Rank")
```


SIMD2020 Rank
6000
4000
2000
0
50
60
70
80
90
Employment Rate of Working Age Population
Council_area
City of Edinburgh
Glasgow City