Prepared by Richard Rennehan

NSCC Institute of Technology

DBAS3075 Introduction to Statistical Learning

Predicting Heart Weights of cats FROM Gender and Body weight

March 1, 2018

Contents

[Introduction 2](#_Toc507059239)

[Data Source 2](#_Toc507059240)

[Findings and Results 3](#_Toc507059241)

[Conclusion 5](#_Toc507059242)

[References 7](#_Toc507059243)

# Introduction

A cat’s heart weight can be predicted using its body weight and gender. The following report will display how a cat’s body weight and gender affect the weight of its heart.

# Data Used

The data used for this report comes from the following website:

https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/MASS/cats.csv

It consists of data for 47 female cats and 97 male cats. The three columns used for this report are Sex, Bwt (body weight), and Hwt (heart weight).

# Findings and Results

Figure 1: Heart weight vs body weight for all cats

First, it must be mentioned that body weights range from 2.00kg-3.00kg for females and 2.00kg-3.90kg for males. It is inappropriate to apply the results from this report to cats outside these body weight ranges.

Figure 1, Figure 2, and Figure 3 show scatterplots comparing a cat’s heart weight in grams to its body weight in kilograms.

Figure 1 shows there is a positive linear relationship between body weight and heart weight. As the cat’s body weight increases, its heart weight increases at a similar rate. Figure 2 and Figure 3 show this remains true for both males and females.

Figure 2: Heart weight vs body weight for male cats

Figure 4 displays the frequency of body weights among all cats. The histogram is right skewed. The reason is because all the cats used in this study had an above average body weight of at least 2.0kg, so the left side would have higher frequencies.

Figure 3: Heart weight vs body weight for female cats

There are 4 conditions that should be met before performing linear regression. The first is the straight enough condition. Figures 1, 2, and 3 overall demonstrate that a straight line can be reasonably drawn between the points. The second is the no categorical variable condition. Body weight and heart weight are both quantitative, so this condition is met. The third condition is what we call, “does the plot thicken?” If a line was drawn between the points, the distance the points are above and below the line should remain consistent. Finally, the outlier condition states that any obvious outlier points can negatively affect the linear regression model.

Figure 4: Frequency of body weights across all cats

Running a linear model to predict heart weight results in the following equation:

Hwt = -0.4149 -0.0821(Sex) + 4.0758(Bwt)

Where Hwt represents heart weight in grams, Bwt represents body weight in kg, and Sex equals 1 if male or 0 if female. The value of R2 is 0.6468, meaning around 65% of the variability in body weight can be explained by varying the heart weight.

For two cats of the same body weight but opposite genders, the male cat is predicted to have the lighter heart.

Say, for example, we had a male cat weighing 2.5kg. We can use the formula to predict his heart weight as shown:

Hwt = -0.4149 – 0.0821(1) + 4.0758(2.5)

Hwt = -0.4149 – 0.0821 + 10.1895

Hwt = 9.6925 grams

How about a female cat of body weight 2.2kg?

Hwt = -0.4149 – 0.0821(0) + 4.0758(2.2)

Hwt = -0.4149 + 8.96676

Hwt = 8.5519 grams

Keep in mind that a cat weighing 5.5kg, for example, should not be considered for this equation. It is outside the range of cat weights used for this study. There could be a small chance that the linear module starts to lose consistency outside this range.

# Conclusion

We can conclude that there is a positive linear relationship between cat body weights vs heart weights. As the cat’s body weight goes up, so does its heart rate. For two cats of the same body weight but opposite genders, the male is predicted to have a lighter heart.

# References

Data Set: https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/MASS/cats.csv

R Code:

cat\_data = read.csv("https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/MASS/cats.csv", header=TRUE, sep=",")
cat\_data = cat\_data[-1]

males = subset(cat\_data, Sex=="M")
females = subset(cat\_data, Sex=="F")

plot(cat\_data$Bwt,cat\_data$Hwt,main="Cat Heart Weight vs Body Weight",xlab="Body Weight (kg)",ylab="Heart Weight (grams)")

plot(males$Bwt,males$Hwt,main="Male Cat Heart Weight vs Body Weight",xlab="Body Weight (kg)",ylab="Heart Weight (grams)")

plot(females$Bwt,females$Hwt,main="Female Cat Heart Weight vs Body Weight",xlab="Body Weight (kg)",ylab="Heart Weight (grams)")

hist(cat\_data$Bwt,main="Histogram of Cat Body Weights", xlab="Cat Body Weight (kg)")

model1=lm(Hwt~., cat\_data)

summary(model1)