Midterm 1 Lab Portion Solutions - R Studio

Dwight Wynne

October 4, 2018

Problem 1

The cats.csv file on Titanium contains values of the following variables measured on 144 cats:

- Sex (M - Male, F - Female)
- Body_Weight: weight of the cat, in kilograms
- Heart_Weight: weight of the cat's heart, in grams

Part (A)

Find the mean and median of the variable Body_Weight, rounded to two decimal places. Interpret each value.

```
mean(cats$Body_Weight)
```

```
## [1] 2.723611
```

```
median(cats$Body_Weight)
```

```
## [1] 2.7
```

```
summary(cats)
```

```
##      Sex             Body_Weight     Heart_Weight
##  Length:144         Min.   :2.000   Min.   : 6.30
##  Class :character   1st Qu.:2.300   1st Qu.: 8.95
##  Mode  :character   Median :2.700   Median :10.10
##                     Mean   :2.724   Mean   :10.63
##                     3rd Qu.:3.025   3rd Qu.:12.12
##                     Max.   :3.900   Max.   :20.50
```

- 0.5 pts: For code and output similar to that above.
- 0.5 pts: The average body weight of the 144 cats is 2.72 kg.
- 0.5 pts: The average cat in the dataset weighs 2.70 kg.

Points awarded for equivalent, sensible interpretations of mean and median.

Part (B)

Compute the interquartile range of the variable Body_Weight, rounded to two decimal places.

```
quantile(cats$Body_Weight, 0.75) - quantile(cats$Body_Weight, 0.25)
```

```
##   75%
## 0.725
```

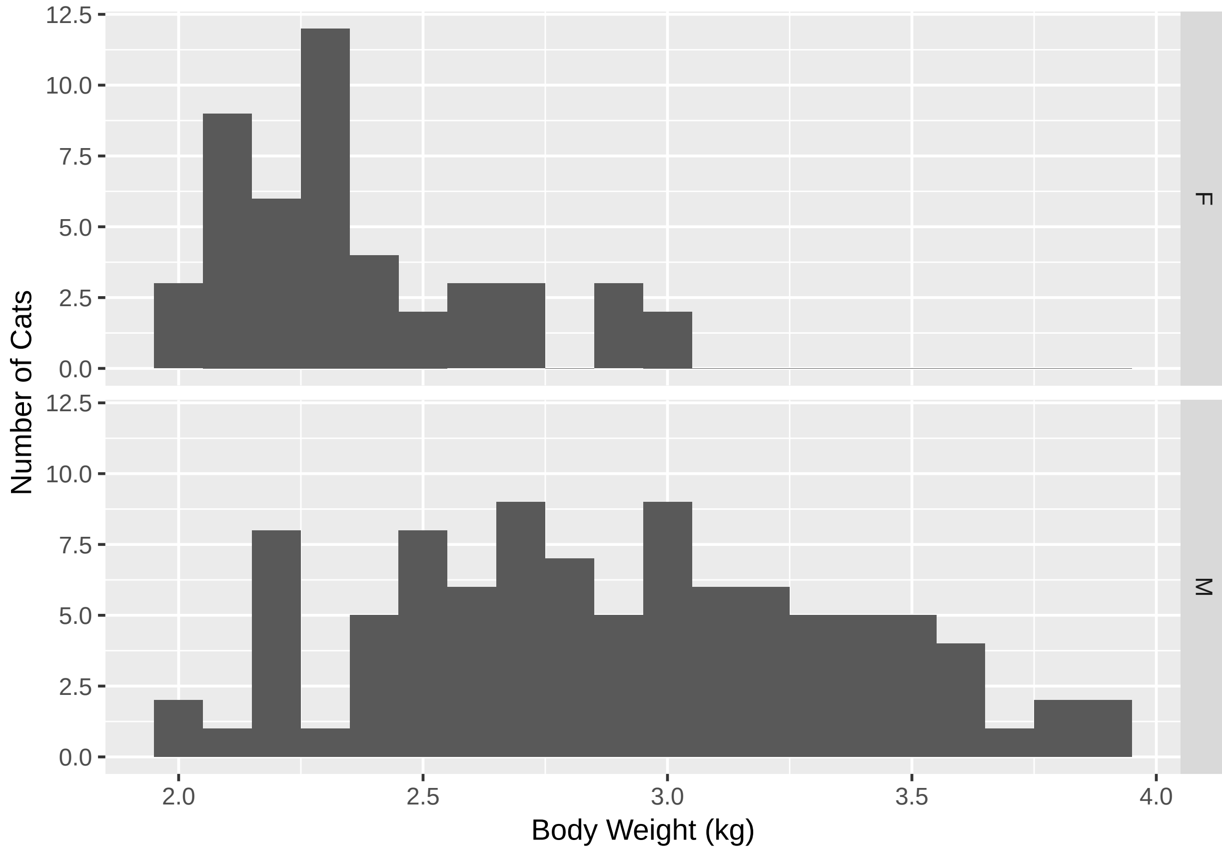
Or we can look at the summary and pick the 1st Qu. and 3rd Qu. and subtract them.

- 0.5 pts: The interquartile range is 0.73 kg.

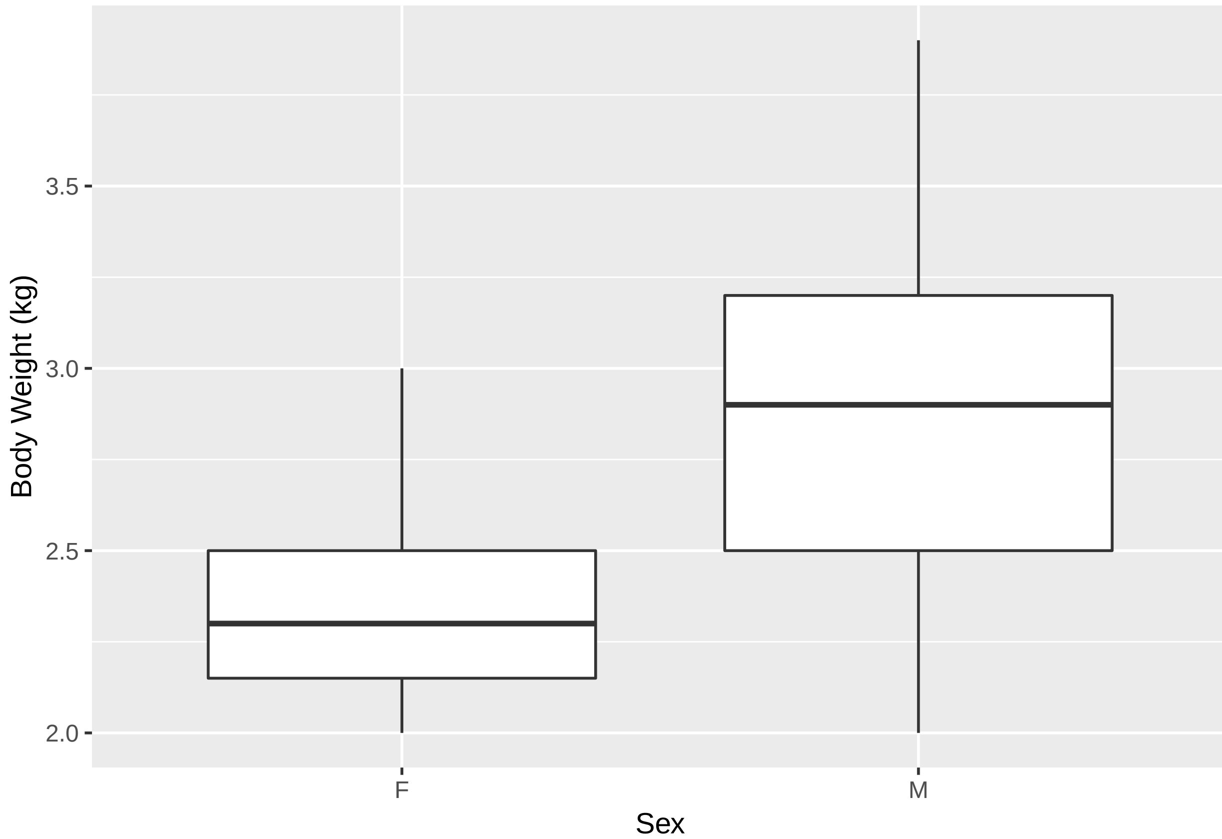
Part (C)

Create a set of graphs that shows how the distribution of body weights differs in Male vs. Female cats. Make sure your axis labels are well-defined!

```
library (ggplot2)
cats_hist <- ggplot(cats, mapping = aes(x = Body_Weight)) +
  geom_histogram(center = 2.5, binwidth = 0.1)
cats_hist_labeled <- cats_hist +
  labs(x = "Body Weight (kg)", y = "Number of Cats") +
  facet_grid(Sex~.)  # histograms one on top of the other
print(cats_hist_labeled)
```




```
library (ggplot2)
cats_boxplot <- ggplot(cats, mapping = aes(x = Sex, y = Body_Weight)) + geom_boxplot()
cats_boxplot_labeled <- cats_boxplot +
  labs(y = "Body Weight (kg)")
print(cats_boxplot_labeled)
```



- 1 pt for either the set of histograms or the set of boxplots above (or similar)
- 1 pt for appropriate x-axis and y-axis labels, such as those shown above

Part (D)

Briefly explain what your figure in part (C) indicates about how Body_Weight is affected by Sex.

The graphs clearly show that male cats weigh more than female cats, on average, and that the variability in male cat weights is greater.

- 1 pt for any reasonable comparison of body weight between male and female cats

Problem 2

Suppose that the heart weight of all female cats in a population is approximately $N(9, 1.5)$ and the heart weight of all male cats in a population is approximately $N(12, 2.5)$. Round all probabilities below to three decimal places (tenth of a percent).

Part (A)

What is the probability that a randomly selected female cat has a heart weight above 11 g?

```
pnorm(11, mean = 9, sd = 1.5, lower.tail = FALSE)
```

```
## [1] 0.09121122
```

The probability that a randomly selected female cat has a heart weight above 11 g is 0.091.

- 0.5 pts for the correct distribution
- 1 pt for using `pnorm` correctly

Part (B)

What is the probability that a randomly selected male cat has a heart weight between 11.5 and 12.5 g?

```
pnorm(12.5, mean = 12, sd = 2.5) - pnorm(11.5, mean = 12, sd = 2.5)
```

```
## [1] 0.1585194
```

The probability that a randomly selected male cat has a heart weight between 11.5 and 12.5 g is 0.159.

- 0.5 pts for the correct distribution
- 1 pt for using `pnorm` correctly
- 0.5 pts for subtracting to get the correct answer

Part (C)

What is the probability that a simple random sample of 47 female cats has a sample mean heart weight below 8.8 g?

```
srs_47cats <- 1.5/sqrt(47)
pnorm(8.8, mean = 9, sd = srs_47cats)
```

```
## [1] 0.1803355
```

The probability that a simple random sample of 47 female cats has a sample mean heart weight below 8.8 g is 0.180 (I'll accept anything between 0.180 and 0.182, depending on how you round your standard deviation).

- 1 pt: The sampling distribution of the sample mean is $N(9, 1.5/\sqrt{47})$ or approximately $N(9, 0.2188)$.
- 1 pt for using `pnorm` correctly

Part (D)

Suppose we obtain a simple random sample of 5 female cats and, independently, a simple random sample of 5 male cats. What is the probability that the sample of female cats has a higher mean heart weight?

```
mean_diff <- 12 - 9  # do male - female
var_male_cats <- 2.5^2/5
var_female_cats <- 1.5^2/5
sd_diff <- sqrt(var_male_cats + var_female_cats)
pnorm(0, mean = mean_diff, sd = sd_diff)  # F - M < 0
```

```
## [1] 0.01069878
```

```
mean_diff <- 9 - 12  # do female - male
var_female_cats <- 1.5^2/5
var_male_cats <- 2.5^2/5
sd_diff <- sqrt(var_female_cats + var_male_cats)
pnorm(0, mean = mean_diff, sd = sd_diff, lower.tail = FALSE)  # F - M > 0
```

```
## [1] 0.01069878
```

The probability that the sample of female cats has a higher mean heart weight is 0.011.

- 1.5 pts: The sample mean for Female is approximately $N(9, 0.6708)$ and the sample mean for Male is approximately $N(12, 1.118)$
- 0.5 pts: The difference of sample means is distributed approximately $N(3, 1.3038)$ if you use Male - Female, or $N(-3, 1.3038)$ if you use Female - Male.
- 1 pt for using `pnorm` correctly

Problem 3

In Problem 4 of the Lecture Exam, you (hopefully) modified the HQ Trivia rules to allow people to answer all 12 questions, even if they get a question wrong. Your young nephew has gotten a hold of the phone and is randomly touching the screen, so that he has a 1/3 chance of getting any individual question correct. Round all probabilities below to three decimal places (tenth of a percent).

Part (A)

Under these new rules (12 questions, 1/3 chance of getting each question correct), what is the probability that your nephew gets no questions correct?

```
n <- 12
p <- 1/3
dbinom(0, size = n, p = p)
```

```
## [1] 0.007707347
```

The probability that he gets no questions correct is 0.008.

- 0.5 pts for the correct distribution
- 1 pts for using `dbinom` with first argument 0

Part (B)

Under the new rules, what is the probability that your nephew gets 25% or fewer of the questions correct?

The probability that he gets 25% or fewer of the questions correct is the probability that he gets $(0.25)(12) = 3$ or fewer questions correct.

```
pbinom(3, size = n, p = p)
```

```
## [1] 0.3930747
```

The probability that he gets 25% or fewer of the questions correct is 0.393.

- 0.5 pts for the correct distribution
- 1.5 pts for using `pbinom` with first argument 3