Lab #3 – Probability and Data **CSC 3220** **Fall 2021**

Name(s): Caleb Lich, Jacob Blovins, Nathan Hynek, Michael Steets

Simple Probability *{40 points}*

Suppose that a woman in her forties goes for a mammogram and receives bad news: a "positive" mammogram. However, since not every positive result is real, *what is the probability that she actually has breast cancer*?

1. Create a matrix of probabilities using the following R code: *{15 points}*

```
> probs = matrix(c(11, 99, 3, 887), 2, 2, byrow = TRUE,
                 dimnames = list(c("M-True", "M-False"),
                  c("C-True", "C-False")))
```

   a) Show the matrix as a result of the above command:

	C-True	C-False
M-True	11	99
M-False	3	887

The upper left value represents what is called the *true positives* (TP) where the person was diagnosed with cancer (i.e., the mammogram was positive) and actually had cancer. The upper right is called the *false positives* (FP) where the patient was diagnosed with cancer but in fact did NOT have cancer. The lower left is called the *false negatives* (FN) where the person was diagnosed as NOT having cancer (i.e., the mammogram was negative) and actually DOES have cancer. The lower right is called *true negatives* (TN) where the person was diagnosed as NOT having cancer and actually did not have cancer. Ideally, one would want to maximize the diagonal (i.e., TP and TN).

   b) What is the chance (probability) that a person has breast cancer and received a negative mammogram?

The probability is 3/1000 or 0.3% chance

c) What is the chance that a person does not have cancer, but received a positive mammogram?

The probability is 99/1000 9.9% chance

2. What is the R code for producing a new matrix with marginal totals included – i.e., *summations* of rows (M) and columns (C) added across the right and bottom (respectively)? *{10 points}*

```
> probs = matrix(c(11, 99, (probs[1] + probs[3]), 3, 887, (probs
[2] + probs[4])),nrow=2, ncol=3, byrow = TRUE, dimnames = list(c
("M-TRUE", "M-False"),c("C-TRUE", "C-FALSE", "Totals")))
> probs
        C-TRUE C-FALSE Totals
M-TRUE      11      99    110
M-False      3     887    890
> |
```

3. Change the false positives to 14 and true negatives to 972 (i.e., move 85 results from FP to TN). Now, what is the chance that a person has cancer, conditional on a positive mammogram? *{15 points}*

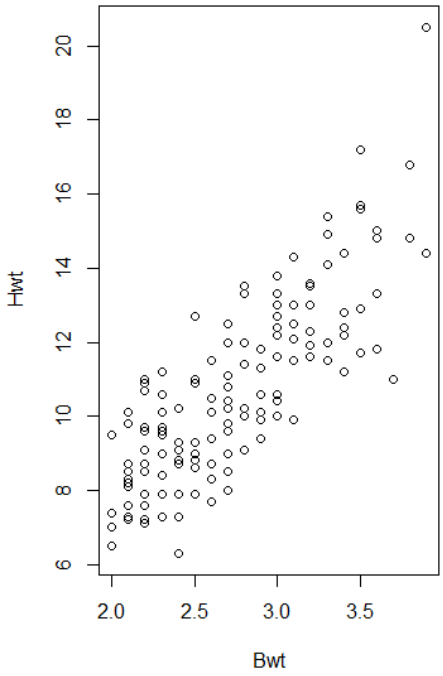
Probability is 11/1000 or 1.1% probability

Understanding the Data *{60 points}*

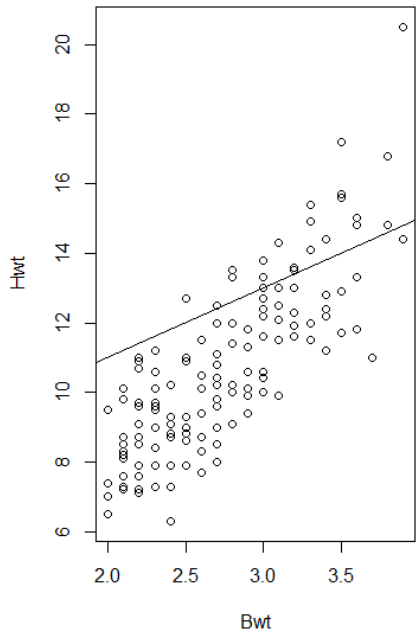
1. Load the data set *cats* from the library *MASS*. (Might want to check and make sure it loaded correctly, and see what this data set is.)

```
> library(MASS)
> install.packages("MASS")
Error in install.packages : Updating loaded packages
> data(cats)
> header(cats)
Error in header(cats) : could not find function "header"
> head(cats)
  Sex Bwt Hwt
1   F 2.0 7.0
2   F 2.0 7.4
3   F 2.0 9.5
4   F 2.1 7.2
5   F 2.1 7.3
6   F 2.1 7.6
>
```

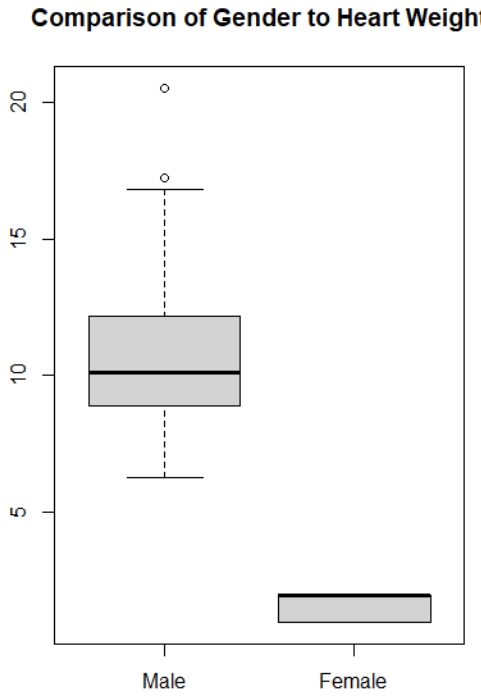
2. plot() the heart-weight against the body-weight, putting the body-weight on the horizontal axis. *{15 points}*



3. Obtain the best straight line through the points using the `abline()` command. *{15 points}*



4. So far, the analysis has not looked at all at the gender of the cat.
    a. Create a box plot that shows the heart weight based upon gender. *{15 points}*
    b. What do you observe? *{15 points}*



We observed that male was bigger than female, and that the average heart weight for male is varied and averages at 10. For female, their hearts are usually a lot less heavy.

(Submit your answers into the corresponding iLearn dropbox <u>as a single PDF</u> – be sure to include all of your names at the top of the file)