

# MLearn Week 1 Homework

Strongly recommended due date: October 12, before class starts

Official due date: October 16

## Probability

*For problems like these, please show your work (i.e., how you got to your answer). Acceptable ways to do this are: use paper and pencil and scan it, typeset it on a computer (bonus points for LaTeX), do it in R or Python and submit a script or notebook. In the latter case, please use very clear comments or variable names.*

Your friend has 19 ordinary six-sided dice (with sides numbered 1,2,3,4,5,6), and one weird die which has sides 0, 0, 1, 1, 2, 3 (two sides have a “0,” two have a “1,” one has a “2,” and one has a “3”).

(a) If your friend rolls one of the ordinary dice and the one weird die, what is the probability that the sum of the values shown will be 7 or greater. Symbolically, if  $S$  represents the sum of the values, find  $P(S \geq 7)$ .

(b) If your friend rolls one of the ordinary dice and the one weird die, what is the probability that the two dice will show the same value?

(c) Your friend puts all 20 dice into an opaque bag and then randomly pulls one out. What is the probability that the selected die is the weird one? Symbolically, if  $W$  represents the event that the weird die is selected, what is  $P(W)$ ?

(d) Your friend puts all 20 dice into an opaque bag and then randomly pulls one out and rolls it. The die shows the value 0. With this information, what is the probability that the weird die was selected?

(e) Your friend puts all 20 dice into an opaque bag and then randomly pulls one out and rolls it. The die shows the value 6. With this information, what is the probability that the weird die was selected?

(f) Your friend puts all 20 dice into an opaque bag and then randomly pulls one out and rolls it. The die shows the value 3. With this information, what is the probability that the weird die was selected?

(g) Your friend puts all 20 dice into an opaque bag and then randomly pulls one out and rolls it. The die shows the value 1. With this information, what is the probability that the weird die was selected?

(h) Your friend puts all 20 dice into an opaque bag and then randomly pulls *two* out. What is the probability that one of them is the weird one?

(i) Your friend puts all 20 dice into an opaque bag and then randomly pulls *two* out and rolls the pair of dice 10 times. Of those 10 trials, the two dice had the same value exactly once. With this information, what is the probability that the weird die was one of the two selected?

## Working with Data

*For this and other problems that involve working with actual data, submit an R or Python script or notebook. Please use clear variable names and/or comments so that a grader knows exactly what you're trying to doing. The script should be something the instructor or TA could run on their own computer to produce the answers. You may assume the relevant CSV file is in the script's working directory.*

Chapter 2 Exercise 8.

You can download College.csv from <http://www-bcf.usc.edu/~gareth/ISL/data.html>

You do not need to use the same commands that the book uses. (For example, instead of using `fix()`, I just display the value of the data frame.) Using either the built-in R plotting or `ggplot2` is fine. (If you're using Python, any mainstream plotting library is fine. Personally I disrecommend `matplotlib`, but you may like it if you're used to Matlab.)

## Baseline Models

We will again use the College.csv dataset for this problem.

Build a baseline model (a.k.a. constant regression) for acceptance rate (number of applications accepted divided by number of applications received). Use the mean acceptance rate, as we did in class. Evaluate the model (compute RMSE). Use a (randomly selected) 80% of the data for training and 20% for testing. Plot the residuals (y axis) against the outcome (x axis).

What is the median acceptance rate? Build a baseline model using the median instead of the mean. How does that compare (according to RMSE) to the mean model? (Be sure to use the same training and testing sets.)



