Unit 3: lab

Exercise One

There are 1000 3 digit numbers (I'm allowing leading zeroes in this count). 20 of these numbers have a "42" in them, where "42" is a substring.

1a) How many 5 digit numbers have a "42" in them?

1b) How many N digit numbers have a "42" in them? A closed form answer is desired.

1c) write a python or R code that creates a vector of N-digit integers and then converts them to strings, then finally counts how many strings in the vector contain '42' as a substring. Does the result agree with your analytical prediction?

Note: it's a little bit easier to code this in python than R, at least for me.

Exercise Two

MASS package contains data about 93 cars on sale in the USA in 1993. They're stored in Cars93 object and include 27 features for each car, some of which are categorical. So let's load the MASS package and look at the type of vehicles included in

2a) Select the types column. We have 6 types of cars. Perform the table function on the 'Type' column to tell how many of each type we have.

2b) use prop.table convert this table it into fractions

2c) use table and prop.table on the data on the 'Origin' variable

2d) create a contingency table: use table again, but with two arguments now. First will become row variable and second will become column variable, the two arguments will be the Type variable and the Origin variable data. Call prop.table to get the cell proportions **2e)** obtain the

row proportions and column proportions of this contingency table using the margin argument to the prop.table function call. (See lecture notes unit 3 for an example of how to do it on kobe dataframe)