Exercises for Week 2 – IST772

Instructions: Pascal's Triangle was named after 17th century French mathematician Blaise Pascal, but the structure appears to have emerged independently from India, China, and Persia as early as the second century BCE. Each element is the number of different ways of choosing k elements from a set with n elements. Think of n as the row number (starting with 0) *and* the number of coins we are tossing. Then k is the position within a row (starting with zero) and is also the number of heads. The third row below would read like this: Tossing two coins, there's one way to get zero heads, two ways to get one head, and one way to get two heads.

```
    1
  1   1
1   2   1
```

1. Create a Pascal's Triangle with this R code:

   lapply(0:7, function(i) choose(i, 0:i))

   If you were running trials for the **final row** *empirically* with actual coin tosses, how many coins would you need to throw to conduct each trial? Write a comment.

2. If all is well, the sum of the entries on your final row should be 128 (by the way, that is also two raised to the seventh power). If this was the number of heads across a total of 128 trials, *for what number of trials* would we have observed three heads? Convert the lowest layer of your triangle to probabilities using this R command: **choose(7, 0:7)/128**

3. Use rbinom() to create a list of 20 trials, where each trial consists of seven events (e.g., coin tosses). Set prob=0.50 to use a fair coin. Did your list include a zero or a seven? Why or why not?

4. Use the following line of code to produce the theoretical probabilities for a binomial distribution with seven events:

   dbinom(x=0:7, size=7, prob=0.5)

   Why are these theoretical probabilities slightly different from your results for question #3?

   For fun try: http://www.randomservices.org/random/apps/BinomialCoinExperiment.html

Exercises for Week 2 – IST772

Instructions: The table below shows accident reports from three different factories over the past month. Four types of accidents are represented. Each cell contains a count of the number of accidents of the given type at the particular factory:

Accidents	*Factory 1*	*Factory 2*	*Factory 3*	
Vehicle	0	6	4	
Spill	6	0	6	
Equipment	6	4	5	
Injury	4	9	0	

5. Add marginal totals to the table above for cross checking with your R results.

6. Recreate the matrix in R using the following code:

```
accMatrix <- matrix(data=c(0,6,4,6,0,6,6,4,5,4,9,0),
                    nrow=4,byrow=T,
                    dimnames=list(c("Vehicle","Spill","Equipment","Injury"),
                                  c("Factory 1", "Factory 2", "Factory 3")))
```

7. Compute a copy of accMatrix that contains proportions instead of counts.

8. Calculate marginal totals for accMatrix. Two helpful functions that can be called are rowSums( ) and colSums( ).

9. OSHA is auditing the factory that has **the worst accident record**. Add a comment in your code indicating which factory that is. For that factory, display the raw proportions of each type of accident, using the [ ] subsetting technique. For example, you could show the first column of accMatrix with this command: accMatrix[,1]

10. Putting your focus *only on accidents at that factory*, what's the probability of vehicle accidents at that factory? Write one line of R code that displays the result and include a comment describing what you see.

11. The insurance company for these factories wants to understand **the most prevalent type of accident** across all factories. Putting your focus *solely on the most prevalent kind of accident*, what's the probability of that kind of accident at each factory? Write a line of R code that displays the results and include a comment describing what you see. Which factory accounts for the most accidents in this category?