Stat 100 Assignment 5

1. Provide the graph of data, verbal description of the result with reference to Zipf's law. (3 marks)

2. (3 marks)
CPI
There are millions of different products, it's impossible to get the whole population information to construct an index of typical consumer prices without sampling products. Allowing all the consuming families from the population the same chance of being chosen, the chance of getting too many families that spend more or too many families that spend less is smaller. Random sampling is a simple method for ensuring a representative selection. Moreover, using random sampling can reduce the cost and time of data collection and data analysis.

Census
In a population census, the government attempts to collect information from every individual in the population. However, it is too time consuming to collect information from every individual and too costly to recruit and train enumerators. Asking only 5% of the population in detail in random is much cheaper. Random sampling can also reduce the time of data analysis. The analyzed result may be outdated since it takes a long time to analyze the whole population information.

Jury
Random sampling in the context of the jury article is proven to be useful because it can be used to minimize the percentage of people chosen to be juries that have the tendency to be bias against the defendant. By random digit dialing of the last 4 digits, the people being picked have different social views and demographic backgrounds. It generalizes their opinions to the overall population of potential jurors. Then, the defendant's lawyer can find out the percentage of age group and gender type that would be most favorable to the defendant.

3.   (3 marks: 1 mark for the mean, 1 mark for the SD and 1 mark for describing other characteristic of the distribution, and 1 bonus mark if the student gives the exact probabilities of the distribution)

Let x be the # of prizes you are going to win
X is a binomial distribution with $n = 10$ and $p = 100/1000 = 0.1$

$X =$	$P(X=x)$
0	0.348678
1	0.387420
2	0.193710
3	0.057396
4	0.011160
5	0.001488
6	0.000138
7	0.000009
8	0.000000
9	0.000000
10	0.000000

Expected # of prizes you are going to win
$= 0 \times 0.348678 + 1 \times 0.387420 + 2 \times 0.193710 + 3 \times 0.057396 + \ldots$
$= 1$

Therefore, your expected return will be 10 x $1 = $10

SD of 0 - 1 population $= \text{root}(0.1 \times 0.9) = 0.3$

SD of proportion (= SD of sample mean) $= .3/\text{root}(10)$

SD of winnings $= \$10*10*(.3/\text{root}(10)) = \$9.49$

Therefore, the distribution of the prizes that your $10 worth of tickets will get you has a mean of $10 and standard deviation $9.49.