Voting Preference Lab - Leila Erbay

**Voting Preference**

For the 2012 presidential elections, in the state of Massachusetts 1,921,290 people voted for the Democratic Party, while 1,188,314 voted for the Republican Party.

Within this problem we will ignore any other votes.

Suppose you ask 20 randomly selected people in MA who they voted for in the 2012 elections.

a) Write down a possible event for the outcome of this survey:

person 1 - D

person 2..20 - R

{D,R,R,R,R,R,R,R,R,R,R,R,R,R,R,R,R,R,R,R,R}

b) How many possible outcomes are in the sample space?

2^20

c) Now, we want to define a random variable to simplify our future analysis. We simply define “Ob” the number of people who voted for Obama in our survey.

*Ob <- PEOPLE WHO VOTED FOR OBAMA*

In general a random variable is a function that takes in events from the sample space and outputs a number. In this case we can input {R,R,D,D,D,R…etc} and count the number of D’s (people who voted Democrat) and the random variable then assigns this number to Ob, say Ob=13, out of the 20 people you have asked.

What possible values can Ob take?

from the sample of 20 people, the possible values Ob can take is

1 to 20

**PAUSE HERE, CHECK ANSWERS WITH CLASS, BRING FORWARD ANY QUESTIONS**

d) The number of different values that Ob can take is clearly less than your answer to b). Explain why this is the case.

Ob can take up to 20 values

This answer is smaller because the sample size we are looking at is much smaller.

e) As we saw in the previous lab activities, each event has a probability associated with it. What is your estimate of the probability that a specific person voted for the Democratic Party?

Probability that a person voted for D =

P(D) = 1,921,290 / (1,921,290 + 1,188,314) = 0.618

**PAUSE HERE, CHECK ANSWERS WITH CLASS, BRING FORWARD ANY QUESTIONS**

f) Because events have probabilities associated with them, we can calculate the probability of a random variable having a specific value. We start with a simple case:

Write down all the events that would lead to Ob=20, calculate their probabilities, and then sum them together to obtain:

P(Ob=20)

Ob = Binomial Variable

p = 0.618

P(Ob = 20) =( 20 C 20 ) (0..618^20) (0.618^0)

= 0.618^20

g) Random variables have a distribution associated with them. If we know the distribution, we can easily calculate the probability that the random variable takes on any particular number.

Ob is a binomially distributed random variable, which we write as follows:

Ob ~ B(n,p)

Where n=20

What is p?

p = 0.618

**PAUSE HERE, CHECK ANSWERS WITH CLASS, BRING FORWARD ANY QUESTIONS**

h) A single vote, also called a Bernoulli random variable (RV), is equal to 1 with probability p and 0 with probability (1-p). Recall p from part g).

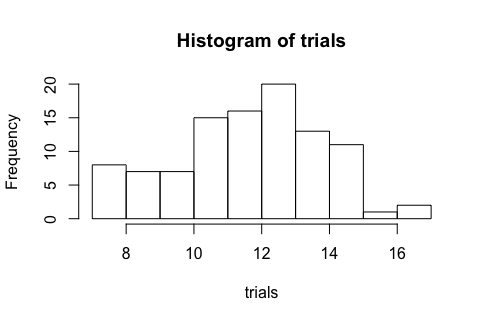
Now we will create a histogram of the data based on a simulation in R.

Use the rbinom() function as shown below **using your answer for g) in place of prob.**

rbinom(n, size, prob) # "n" is the number of observations, "size" is the number of trials and "prob" is the probability of success for each trial

Generate twenty votes, then compute the sum of those votes using sum(). This your Binomial RV.

Run 100 trials of this using a for loop. Save the sum of each trial (your random variable) to a vector. Create a histogram of the random variable values from your simulation.



i) What is the Expected Value of Ob? What is the average of all the random samples you obtained? Why aren’t they the same? What if you run 1000 trials? Try it.

E[Ob] = n\*p = 20\*0.618 = 12.36

E[Sample\_Ob] = 12.09

They are not the same because each trial is independent thus the average of 100 samples could be drastically different than E[Ob]

E[Ob\_1000] = 12.451

**PAUSE HERE, CHECK ANSWERS WITH CLASS, BRING FORWARD ANY QUESTIONS**

j) What is the variance of Ob? What is the variance in your sample? Why aren’t they the same?

V[Ob] = n\*p\*q = 20\*0.618 \* 0.382

V[Ob\_sample] = 4.931212

k) In the 2012 presidential elections, 41.4% of people in Texas voted for Obama.

We bring 15 Texans into a room and poll them about their voting history.

We define a new random variable “Ot” as the number of people in that room that voted for Obama.

What is the distribution of Ot?

Ot = # that voted for Obama

Ot - Binomial distribution

~Bin(15, .414)

**PAUSE HERE, CHECK ANSWERS WITH CLASS, BRING FORWARD ANY QUESTIONS**

l) For some weird reason, a pollster defines a new random variable “W” as W= 3\*Ob+2\*Ot +10

Go back to your R and simulate many rounds of Ot so you can obtain a random sample of W.

m) What is the expected value of W? What is the variance of W?

E[W] from sample: 59.28

V[W] = V[3Ob + 2Ot +10]

= V[3\* Ob] + E[2\*Ot] +V[10]

= 9 \*V[Ob] + 4V[Ot] + 0

= 9(20\*.618\*.382) + 4(15\*.414\*.586)

= 57.049

V[W] from sample: 61. 957

n) Compute the average and variance of your random sample. How do they compare to the answers above?

All ANSWERS ABOVE ^^^ (sample and random variable)

Note- If the variance looks different maybe you need more samples.

E[W] = E[3\*Ob + 2\*Ot + 10]

= E[3\*Ob] + E[2\*Ot] + E[10]

= 3E[Ob] + 2E[Ot] + 10

= 3\*(20\*0.618) + 2(15\*.414) + 10

= 59.5

V[W] = V[3Ob + 2Ot +10]

= V[3\* Ob] + E[2\*Ot] +V[10]

= 9 \*V[Ob] + 4V[Ot] + 0

= 9(20\*.618\*.382) + 4(15\*.414\*.586)

= 57.049

Make sure you properly understand the distinction between events, random variables, and the distribution of random variables. These distinctions are important to understand because we build on these ideas in future lectures.

**PAUSE HERE, CHECK ANSWERS WITH CLASS, BRING FORWARD ANY QUESTIONS**