PS 6800: Problem Set 1

Professor: Jane Esberg Teaching Assistant: Nicolás Idrobo

Due: Tuesday September 6, 2022 (8am)

1. Collaboration is allowed but you must write up your own answers. Do not copy someone else's code or answers (see the collaboration policy in the syllabus)
2. Submit a compiled PDF (in either LaTeX or RMarkdown) to Canvas by Tuesday at 8am. Use the format (lastname)_pset1.pdf.
3. For full credit show your work, not just the final answer. This means you should submit your code for any questions involving `R` as well.
4. Finally, when doing simulations or draws from a random distribution in `R`, please set your seed immediately using `set.seed(6800)`.

Problem 1

Our colleague Marc Meredith published a paper in 2020 estimating the prevalence of double voting in U.S. presidential elections. The core insight is that two *John Smiths* with the same birthday could very well be two different voters, because John Smith is a common name. Two people named Exa Dark Siderael (Elon Musk's daughter's name) with the same birthday are probably the same person voting twice.

Let's say we are given access to the voter files of American men. American men in this universe are all named *Hugo* or *Gus* (short for Augustus). We cannot see their birth year, but we know the date of their birth (e.g. July 23). We know the following about the distribution of the names *Hugo* and *Gus*:

- All months are equally likely (eg, men are equally likely to be born in February as December)
- Within any given month, every day is equally likely (eg, men are equally likely to be born on March 9 as March 26)
- *Hugo* and *Gus* are equally popular names over the course of a given year (50% of men are Hugos, and 50% are Guses).
- However, parents are more likely to name their child Augustus in August. 15% of *Guses* are born in August.

a) What is the overall probability of being born in August?

b) What is P(Name=*Gus*|Man born in August)? That is, conditional on being born in August, what is the probability that a man is named Gus?

c) What is P(Name=*Gus*|Man born in July)?

d) Imagine that there are 600 men born in 1982. Given the conditions above, what is the probability that at least two baby *Guses* are born on August 1, 1982?

e) What is the probability that at least two baby *Hugos* are born on August 1, 1982?

f) Given your answers to (d) and (e), what would you say to a pundit who pointed to two *Guses* with the same August 1, 1982 birthday as evidence of voter fraud in the form of double voting? What would you say to a pundit who pointed to two *Hugos* with the same August 1, 1982 birthday as evidence of voter fraud in the form of double voting?

Problem 2

Does social environment affect the political development of preadults? There has been some work [1] that shows that socializing agents do transmit certain political attitudes and values. Specifically, the authors argue that there exists a transmission of political values from parents to children, as manifested in their views during late adolescence. We will use this idea to solve the following questions.

Take that a parent can share (or not) a certain political attitude – here, we'll say xenophobia. If the parent is xenophobic, the children will share this same view with probability 3/5. On the other hand, if the parent does not have this political view (is not xenophobic), the children will not have it either. Take a parent, who has probability 1/3 of being xenophobic, and has 2 children.

a) What is the sample space for this experiment?

b) What is the probability that the younger child is xenophobic if their parent is xenophobic?

c) What is the probability that neither child are xenophobic?

d) Is whether the elder child has the political attitude independent of whether the younger child shares it?

e) Conditional on the parent holding the attitude, is the elder child's viewpoint independent of the younger child's?

f) We find an elder child who is not xenophobic. How does this change your beliefs about whether the parent is xenophobic? Show numerically.

g) We also find out that that the younger child does not share the value How does this change your evaluation of whether the parent is xenophobic? Show numerically.

Problem 3

Now let's verify whether our results hold doing simulations. In `R`, create three empty vectors for `parent`, `kid1` and `kid2`. Simulate 1,000 "families" where the probabilities of each individual holding the policy preference correspond to the probabilities given in Problem 2 above. After generating these 3 vectors of length 1,000, bind them together and assemble a `families` dataset using `data.frame(parent = vector1, child1 = vector2, child2 = vector3)`.

Some hints on functions that will be useful:

- `set.seed(6800)`: set your seed to 6800 - this makes sure that our randomization is the same
- `sample()`: randomly sample a population
- `c()`: create vectors

[1] Jennings, M. K., and Niemi, R. G. (1968). "The transmission of political values from parent to child", *American Political Science Review* 62(1), 169-184.

- `data.frame()`: create a data.frame
- `if()`: if statement (if a statement is true, do the things that follow in the {})
- `==,!=`: check if statement "the object on left is equal to object on right" is true; check if statement "the object on the left is not equal to the object on the right" is true.
- `data$variable1[data$variable2==X]`: returns the values of one variable matching a certain condition, data$variable2==X
- `sum()`: add up the numeric items in the parentheses
- `subset()`: subset a data.frame according to some rules
- `which()`: which elements where the statement in () is true

Based on your simulation:

a) What is the proportion of families in which no child holds the political attitude?

b) Among families with an elder child who is not xenophobic, what proportion of parents are xenophobic?

c) Among families where no child is xenophobic, what is the proportion of families where the parent is also not xenophobic?

d) Is whether the elder child xenophobic independent of whether the younger child is xenophobic?

e) Conditional on the parent's attitude, is the elder child independent of the younger child?

f) How well do these simulation results match with the analytic results in Problem 2 above?