American University, Department of Government

Ph.D. Methodology Qualifying Exam, Summer 2019, 4 Hours, answer all 8 questions.

1. **OUTPUT ANALYSIS.** Consider the table below.

   (a) Which of the coefficients meet the standard threshold for statistical significance? $(p < .05)$

   (b) Write out the estimated model representing these results in algebraic form.

   (c) Using the coefficients below, describe the relationship between Support for the women's movement and Support for Clinton *among men*.

   (d) Using the coefficients below, describe the relationship between Support for the women's movement and Support for Clinton *among women*.

   (e) Compare your results in 1c and 1d, and describe this pattern of results in lay terms.

Predictor variable	Support for Clinton
Support for women's movement	.75 (0.05)
Female	15.21 (4.19)
Support for women's movement x Female	-.13 (0.06)
Intercept	1.56 (3.04)
n	1466

Note: OLS regression coefficients with standard errors in parentheses. The outcome variable is the respondent's thermometer score for Hillary Clinton, ranging from 0 to 100. Support for women's movement also ranges from 0 to 100. Female is a dummy variable, where 0 = male and 1 = female.

2. **POISSON PROCESSES.** Suppose you have a Poisson process with rate parameter $\lambda = 5$.

   (a) What is the probability of getting exactly 7 events?

   (b) What is the probability of getting exactly 3 events?

   (c) These values are the same distance from the expected value of the Poisson distribution, so why are they different?

3. **STATISTICAL THEORY.**

   (a) Explain the role of the Law of Large Numbers in frequentist statistics.

   (b) Explain how the Central Limit Theorem relates to survey design and total error in surveys.

4. **NONLINEAR TESTING.** Fully describe the Wald test, the likelihood ratio test, and the Lagrange multiplier test. Explain how they are different, how they are the same, and the advantages of each one. Illustrate with a graph.

5. **LINEAR ALGEBRA.** Clogg, Petkova, and Haritou (1995) give detailed guidance for deciding between different linear regression models using the same data. In this work they define the matrices $\mathbf{X}$, which is $n \times (p+1)$ rank $p+1$, and $\mathbf{Z}$, which is $n \times (q+1)$ rank $q+1$, with $p < q$. They calculate the matrix $A = \left[\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right]^{-1}$. Find the dimension and rank of A.

6. R **CODE.** Describe what the following R code does. Additionally, comment on how algorithms of this type can be useful in research.

```
fun <- function(sampleSize){
    doors <- 1:3
    correct <- 0
    iterations <- 0
    while(iterations <= sampleSize){
        prizeDoor <- sample(doors, 1)
        guess <- sample(doors, 1)
        if(guess != prizeDoor){
            removed <- c(guess, prizeDoor)
        } else {
            removed <- c(guess, sample(doors[doors != guess], 1))
        }
        finalGuess <- removed[removed != guess]
        if(finalGuess == prizeDoor){
            correct <- correct + 1
        }
        iterations <- iterations + 1
    }
    res <- correct/sampleSize
    return(res)
}
fun(10000)
```

7. **PROBABILITY.**

(a) A weighted coin comes up Heads 40% of the time. What is the expected number of flips until the coin comes up Heads once?

(b) An urn contains 10 red balls, 10 blue balls, and 20 green balls. Five balls are randomly chosen from the urn without replacement. What is the probability that the sample contains at least one ball of each color?

(c) Suppose you suspect you may have a termite infestation in your home. An exterminator can either damage the walls or administer a test to determine whether you have termites, though the test costs $250. If a home has termites, the test correctly identifies this 90% of the time. The exterminator tells you that that 1% of all homes have termites, and that the test comes back positive 8% of the time. What is the probability that you have termites, if the test comes back positive?

8. **CAUSAL INFERENCE.** The following results refer to the New Haven voter mobilization experiment, in which a random subset of the subject pool was assigned to be canvassed, but only some of those assigned to be canvassed were actually canvassed. The outcome is voter turnout. (8 points each)

Voter turnout by experimental group, New Haven voter mobilization experiment.

	Treatment Group	Control Group
Turnout rate among those contacted by canvassers	54.43 (395)	
Turnout rate among those not contacted by canvassers	36.48 (1,050)	37.54 (5,645)
Overall turnout rate	41.38 (1,445)	37.54 (5,645)

Note: Entries are percent voting, with number of observations in parentheses.
Sample restricted to households containing a single registered voter.

(a) Define a "Complier."

(b) Estimate the proportion of Compliers in the subject pool.

(c) Show (with algebra) that under the assumptions of non-interference and excludability, the CACE (Complier-Average Causal Effect) is identified in this application.

(d) Are non-interference and excludability plausible in this example?

(e) Estimate (by hand) the CACE. Provide a substantive interpretation of your estimate.