Quantitative Empirical Methods Exam

Yale Department of Political Science, August 2023

You have 24 hours to complete the exam. The exam consists of three parts.

Back up your assertions with mathematics where appropriate and show your work. Good answers will provide a direct answer that illustrates an understanding of the question, and calculations or statistical arguments to validate the answer. Where applicable, exceptional answers will include all of these *as well as* proofs that are technically complete, including formally articulating sufficient assumptions and regularity conditions. Questions will not be weighted equally. A holistic score will be assigned to the exam. Therefore, it is important to demonstrate your understanding of the material to the best of your ability.

Part 1 (Theoretical section) consists of short questions that can be answered with pen and paper. You are allowed to consult textbooks and other reference material, but the questions are written so that well-prepared students should be able to answer them without such references.

Part 2 (Essay section) contains a recent, well-regarded empirical article. We will ask you to offer an evaluation of its methodological approach and presentation of results. In particular, we will advise you to pay particular attention to the identification conditions (either explicit or implicit), the associated estimation strategy, and possible threats to inference. Your response may be anywhere from 500 to 1,000 words.

Part 3 (Computer assisted section) will involve using statistical software to answer one longer exercise with several associated questions. A complete answer to Part 3 will include code and output, as well as your written answers.

For the whole exam, you are permitted access to any and all written materials, as well as unrestricted use of your own computer with access to the internet. The only restriction is that you may **not** interact with any person, online or otherwise.

Please turn in your answers as an email to colleen.amaro@yale.edu.

1 Theoretical section

1. Provide brief definitions of the following terms:

    (a) unbiasedness

    (b) confidence interval

    (c) p-value

    (d) statistical power

2. In a recent paper Muff, Nilsen, O'Hara, and Nater (2021)[1] propose to implement the recommendation "to regard p-values as what they are, namely, continuous measures of statistical evidence." By this they mean that a p-value less than 0.001 is very strong evidence, a p-value less than 0.01 is strong evidence, a p-value less than 0.05 is moderate evidence, a p-value less than 0.1 is weak evidence, and a p-value greater than 0.1 is little or no evidence. All else equal, is a p-value of 0.001 in fact stronger evidence against the null than a p-value of 0.04? Explain. Why is it important to specify "all else equal"? Be specific.

3. Suppose a scholar performs n independent hypothesis tests. Answer the following about multiple hypothesis testing.

    (a) Assume the null hypothesis is true in all n cases, so we have n independent p-values all distributed uniformly on the unit interval $[0, 1]$. In terms of n, what is the probability that a test with a Type I error rate of 0.03 will reject at least one of the null hypotheses?

    (b) Now suppose that 2 percent of the n alternative hypotheses tested are actually true. We run a test with a Type I error rate of 0.03, but under an optimistic scenario, has a Type II error rate of 0. *Among the tests that turn out with a p-value of* $p < 0.01$, what proportion are false positives?

4. The Canadian National Breast Cancer Study was a randomized controlled experiment on mammography (x-ray screening for breast cancer). The study found that screening was beneficial, especially for older patients. (Benefits were measured by comparing death rates in the treatment and control groups.) A critique argued that the randomization was not done properly: instead of following instructions, nurses were more likely to assign high risk patients to the treatment group. Would this bias the study? If so, would the bias make the treatment of screening look more or less effective than it really is? Explain your answer.

[1] You don't need to read the paper, but if you are interested see `https://www.sciencedirect.com/science/article/pii/S0169534721002846`.

5. A die is rolled 8 times. When a die is rolled, each of the six faces is equally likely to come up. Find the chances of:

   - Getting 8 sixes.
   - Every roll showing 5 or less.

6. A researcher conducts a survey to determine the distribution of income in a certain city. They draw a random sample of 1,000 households and conduct a door-to-door survey. After several visits, the interviewers are only able to reach 727 of the sample households. Rather than face this non-response rate, the researcher draws a new batch of households and uses the first 273 completed interviews in the second batch, bringing the total sample to the desired 1,000. The researcher calculates an average income of $42,000 in these 1,000 households and concludes in a report that "average household income in the city is around $42,000." Is this conclusion likely to be too low, too high, or about right? Explain.

7. National data show that on average, college students spend 6.8 hours per week going to parties. President Salovey does not believe that this figure applies at Yale. Yale has around 6,600 undergraduate students. President Salovey takes a random sample of 100 students and interviews all of them. On average, they report 5.9 hours a week going to parties, and the standard deviation is 8 hours. Is the difference between 5.9 and 6.8 meaningful? Explain.

8. Sometimes experimental researchers analyzing a survey experiment exclude subjects from their analysis because the subjects: (1) failed a pre-treatment attention check; (2) failed a post-treatment attention check; (3) took the survey too quickly (e.g., survey completion time was 3 standard deviations faster than the mean).

   (a) Suppose that the researcher is interested in the average treatment effect among subjects who are not dropped from the analysis. Which of these strategies (if any) are guaranteed to be unbiased?

   (b) Suppose now that the researcher is interested in the average treatment effect for all subjects who begin the survey experiment? Which of these strategies (if any) are guaranteed to be unbiased? (Assume there is no missingness; everyone who begins the survey experiment completes the survey experiment and answers every question in the survey.)

2 Essay section

Read the article attached to your exam.

1. First, start by summarizing the author's main estimating equation in notation, using Y as the outcome, D as the treatment, i to index municipalities and t to index time. Be sure to define, both statistically and substantively, any new notation you introduce too.

2. Next, offer a critical evaluation of its methodological approach and presentation of results. Note: "critical" does not imply that you must only criticize – it is recommended that you give credit to the authors if and when their arguments are convincing and/or novel with respect to standard practice.

Your response should be between 500 to **1000** words.

We advise you to pay particular attention to the identification conditions (either explicit or implicit), the associated estimation strategy, and possible threats to inference. Justify each of your claims and, where applicable, suggest ways in which this line of research might be improved. We do not expect you to have special expertise in the topic area, but we do expect you to bring to bear your general analytical skills as a political scientist.

Article:

> Albertus, Michael. 2023. "The Political Price of Authoritarian Control: Evidence from Francoist Land Settlements in Spain." *Journal of Politics*. `https://www.journals.uchicago.edu/doi/full/10.1086/723991`

3 Computer assisted section

`survey.csv` is a survey of $n = 823$ residents in China.[2] We are interested in inferring the proportion of the population that discusses politics. The survey includes the question, denoted with the variable `discuss`,

> "Do you discuss topics concerning government policies?"
>
> 1. Almost Never
> 2. Rarely
> 3. Often
> 4. Almost Every Day

A data dictionary defining the other variables is provided at the end of this question.

Define Y_i as a binary variable that is 1 if the respondent discusses politics "often" or "almost every day," and 0 otherwise. We also define R_i as a binary variable of whether or not the respondent answered the question. If $R_i = 0$, then the value of the `discuss` question is `NA`. If $R_i = 1$, we assume the respondent answers truthfully.

In the raw data, 36.3% of respondents for whom $R_i = 1$ have $Y_i = 1$. However, your coauthor is concerned that politically active respondents might refrain from responding to this question due to their fear of state surveillance. Another coauthor suggests the pattern might be the opposite: politically inactive respondents are more likely to refrain from answering this question.

In this exercise, we will ask you to evaluate this claim. Submit a **R script** and **written answers**, possibly with tables, with your solution. The R script must be anonymous (it does not include file paths with your name on it) and you should not use custom packages tailored to missing data to answer these questions.

1. Using a logit regression with all the covariates available in the data (except for `id` and the outcome), estimate a propensity score model for R_i. To evaluate the performance of your propensity score, hold out the first 100 rows of the dataset, estimate the model

[2] This dataset comes from Jennifer Pan, Zijie Shao, and Yiqing Xu (2021), "How government-controlled media shifts policy attitudes through framing", *Political Science and Research Methods*, `https://doi.org/10.1017/psrm.2021.35`. However, knowing what this article is about is not necessary nor especially helpful for you to answer this question.

on the remaining rows (The rows of the survey have already been randomly shuffled). Then, report the false negatives and false positives of your model in the held out dataset with a well-labeled 2 by 2 table, using a cutoff rule of declaring anything with a propensity score above 0.9 to be reported. That is, what proportion of cases you declare as missing are actually missing? And what proportion of cases you declare as reported are actually reported?

2. Report an updated estimate of the population average Y_i, $E[Y]$ using the propensity score model from Part 1 (but no other additional variables). In 3-4 sentences, *describe what your adjustment does*, explain *under what circumstances* is your estimator consistent for $E[Y]$, and summarize what your estimates suggest about the nature of non-response.

3. Estimate a linear regression model estimating the conditional expectation function (CEF) of Y_i only among those with $R_i = 1$. Then, construct a doubly-robust estimator that makes use of both your propensity score in Part 1 and this OLS model. In 3-4 sentences, *describe what your adjustment does*, explain *under what circumstances* is your estimator consistent for $E[Y]$, and summarize what your estimates suggest about the nature of non-response.

Data Appendix The following table lists the variables in the dataset.

Variable	Definition
`id`	Anonymous ID
`age`	Age
`college`	College or above (binary)
`minority`	Ethnic minority (binary)
`unmarried`	Not married (binary)
`english`	English fluency, likert scale from 1 (not at all) to 5 (fluent)
`income`	Income, low to high
`religious`	Religiosity (binary)
`ccp`	Communist Party member (binary)
`vpn`	Having used a VPN (binary)
`discuss`	Discuss government policies (1, 2, 3, 4; see above)