-- No calculators, phones, or other devices permitted. --

First Name and Last Name _____

This practice exam focuses on material new for Exam 3, but material eligible for Exam 3 includes material from prior exams and from online notes for Chapters 12 and earlier. This practice exam has a few open-ended items that could be converted to multiple-choice items for the actual exam and/or appear on open-ended Exam 4.

The most missed items on Exam 2 were 41, 31, 10, 12, 38, 32, 19, 25, 40, and 34.

[Directions for the actual exam, not this practice exam] Please write your first and last name on this page. Using a No. 2 pencil, on the opscan, write and bubble in the letters for your last name and first initial, write in "POL" for "DEPT.", write in "138" for "COURSE", and then bubble in your responses. No need for University ID or other information.

POL 138-003 Quantitative Reasoning in Political Science
KEY Practice Exam 3 · Fall 2025

1. Let's discuss two commonly used measures in political science.

   GDP per capita is often used as a measure of the on-average economic activity of a country, in which, compared to a lower GDP per capita, a higher GDP per capita suggests that the country has more economic activity per person on average.

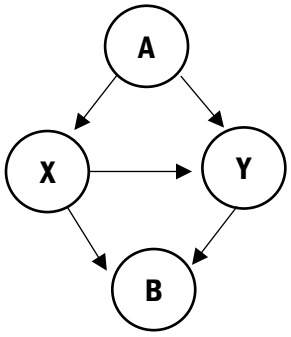
   The Gini coefficient is a measure of inequality that ranges from zero for perfect equality to 1 for perfect inequality. So, for example, regarding wealth, a Gini coefficient of 0 for a country would indicate that each person in the country has the same amount of wealth, and a Gini coefficient of 1 for a country would indicate that one person in the country has all of the wealth.

   Research has found that, on average over time across countries, the Gini coefficient has a negative correlation with GDP per capita ($p<0.05$). Is this sufficient evidence at the conventional level in political science to conclude that, at least on average, higher levels of inequality cause lower GDP per capita?
   A. Yes
   B. No

2. In political science, for peer review of papers that report a statistical analysis, is it typical for the peer reviewers to check the data to see whether the statistical analysis has been correctly conducted?
   A. Yes
   B. No

3. Study A has a sample size of 500 participants, and Study B has a sample size of 200 participants. Study A has an estimated effect size of 4, and Study B has an estimated effect size of 7. Which study should receive more weight in a meta-analysis?
   A. Study A, because Study A has a larger sample size
   B. Study B, because Study B has a larger estimated effect size

4. The arrows in the image below represent causal direction. For example, the image indicates that A influences X and indicates that X influences B. Suppose that we want to conduct a regression to estimate the full extent to which X influences Y. Which of the following variables would be better to add as a control variable in that regression?
   A. A
   B. B



5. Which of the following is the "blind" element of single-blind peer review of a paper?
   A. Authors are not told the names of the peer reviewers
   B. Peer reviewers are not told the names of the paper's authors

6. In a non-experimental analysis, omitting a control variable that should be included ___.
   A. cannot bias an estimate of an effect
   B. can bias an estimate of an effect only to be lower than it truly is
   C. can bias an estimate of an effect only to be higher than it truly is
   D. can bias an estimate of an effect to be lower than or higher than it truly is

7. In a non-experimental analysis, including a control variable that should be excluded ___.
   A. cannot bias an estimate of an effect
   B. can bias an estimate of an effect only to be lower than it truly is
   C. can bias an estimate of an effect only to be higher than it truly is
   D. can bias an estimate of an effect to be lower than or higher than it truly is

8. Franco et al. 2014 reported on a large set of social science experiments that were funded by the federal government. Franco et al. 2014 found that studies that did not provide sufficient evidence of an effect were ___ likely to be published than studies that provided strong evidence of an effect.
   A. much less
   B. just as
   C. much more

9. Suppose that we conduct a randomized experiment to estimate the effect of a treatment, but the experiment does not detect sufficient evidence that the treatment has any effect. One potential reason for this null result is heterogeneous effects, which refers to ___.
   A. socially desirable effects
   B. effects that differ between subpopulations
   C. effects biased by heteroskedasticity
   D. effects that regress toward the mean

10. Validity refers to the extent to which a measuring tool ___.
   A. produces consistent results
   B. produces statistically significant results
   C. measures what the tool is supposed to measure

11. Which type of validity concerns the ability to make correct claims about the sample?
   A. internal validity
   B. external validity

12. Which type of validity concerns the ability of a research result to generalize to the population?
   A. internal validity
   B. external validity

13. Of the following, which term best describes the process in which researchers publicly post ahead of time a plan for the research that they will conduct?
   A. Premonition
   B. Replication
   C. Pre-registration
   D. Post-registration

14. Of the following, which best helps reduce researcher flexibility in analyzing data, to prevent the researcher from misleadingly reporting results?
   A. Premonition
   B. Replication
   C. Pre-registration
   D. Post-registration

15. Suppose that a particular college offers majors only in physics and sociology, and that each student must select exactly one of these majors. At this college, compared to male physics majors, female physics majors have been more likely to receive a scholarship. Moreover, compared to male sociology majors, female sociology majors have been more likely to receive a scholarship. Can we therefore conclude that, across the college as a whole, compared to male students, female students have been more likely to receive a scholarship?
   A. Yes
   B. No

16. IRBs are organizations designed to protect human subjects in research. Can an IRB, in some circumstances, permit research involving human participants in which the participants do not provide informed consent before participating in the research?
    A. Yes
    B. No

17. Below are data for two studies:

Study	Estimated effect size	Sample size
A	6	200
B	3	500

    Which of the following is a correct formula for calculating the mean estimated effect size, weighted by sample size?
    A. 6×200 + 3×500
    B. 6×(200÷3) + 3×(500÷6)
    C. 6×(200÷500) + 3×(500÷200)
    D. 6×(200÷700) + 3×(500÷700)

18. The U.S. Constitution does not protect obscene material. For determining whether particular material is obscene, "[material] will not be deemed obscene unless it is patently offensive according to contemporary community standards, i.e., in the judgment of the average person in the community" (709 F.2d 132 2d Cir. 1983, internal quotations omitted). In the 2001 decision in *People* v. *Family Video Movie Club*, the Illinois Fifth District Appellate Court described a survey that was conducted by the defense as evidence of a community standard:

> ...the defense offered a survey conducted in its Belleville store in which patrons who commented on the obscenity citation were asked whether they would like to sign a petition in support or in opposition to defendant's offering of adult videos. Approximately 97 patrons signed the petition in support of defendant, and three patrons signed the petition in opposition.

    Of the following options, which best describes why results from this survey were not good evidence that the community standard supported the sale of the adult videos in question?
    A. The sample of the survey was not plausibly representative of the population.
    B. Support in the survey for the offering of adult videos was not 100 percent.
    C. The patrons signed a petition instead of filling out a survey.

19. In a recent year, about 70% of math majors at Harvard were male and about 30% were female. To assess the extent to which these percentages were fair, which of the following pieces of information would be more useful to know?
    A. the percentage female among high school students who score above average on the SAT-Math test
    B. the percentage female among high school students who earn a perfect score on the SAT-Math test

20. Suppose that researchers randomly assign students to seats in a classroom for a school year, to test the hypothesis that, compared to sitting in the back of the class, sitting in the front of the class will cause students to perform better on the 100-point exam at the end of the school year. The researchers do not detect evidence at $p<0.05$ to reject the null hypothesis of no effect, but which of 95% confidence interval below would better indicate an informative null for the estimated effect of sitting in the front of the class compared to sitting in the back of the class?
    A. [-2, +1]
    B. [-40, +40]

21. Discrimination in which unknown information for an individual is estimated based on known or perceived data for the individual's group is referred to as ___.
    A. statistical discrimination
    B. taste-based discrimination

22. Researchers conduct a randomized experiment. The researchers test the null hypothesis that the treatment has no effect among men, and the p-value for this test is $p=0.03$. The researchers test the null hypothesis that the treatment has no effect among women, and the p-value for this test is $p=0.20$. Does the combination of these results provide sufficient evidence to conclude at the conventional level in political science that the treatment was more effective among men participants than among women participants?
    A. Yes: the p-values of $p=0.03$ and $p=0.20$ provide sufficient evidence that the treatment worked among men participants but did not provide sufficient evidence that the treatment worked among women participants.
    B. No: p-values do not directly indicate anything about effect sizes, so we cannot conclude based on these p-values that the effect size was larger for men participants than for women participants.

23. Suppose that Researcher A and Researcher B each conduct an independent randomized experiment to collect data to test whether a treatment has an effect. Researcher A and Researcher B then each use their data to test the null hypothesis that the treatment has no effect. In reality, the treatment has no effect. What is the expected chance that at least one of these two experiments produces a p-value of 0.05 or less?
    A. 0%
    B. above 0% but less than 5%
    C. 5%
    D. above 5% but less than 10%
    E. 10%
    F. above 10%

24. Suppose that, in a particular country, admission to the best university is based solely on a high school student's score on a multiple-choice test that contains questions about mathematics, language, science, and social studies. Data from past years indicate that, among students who qualify for the university and attend the university, high-scoring students from schools that have low test scores do worse at the university, on average, compared to high-scoring students from schools that have high test scores. Which one of the following options could best help explain this pattern?
    A. overcontrolling
    B. reverse causality
    C. an ecological fallacy
    D. Simpson's paradox
    E. regression toward the mean

25. The phrase "informative null" describes the result of an analysis that did not provide sufficient evidence to reject the null hypothesis but did provide sufficient evidence to conclude that, if the null hypothesis is not true, then the deviation from the null hypothesis is at most small. For example, for testing the null hypothesis that a coin is fair, an informative null would be a situation in which the analysis does not permit us to claim that the coin is biased but the analysis does permit us to claim that, if the coin is biased, then the bias in the coin is likely at most small.

    Of the things in the list below, which would be most useful for assessing whether a null result is an informative null?
    A. a p-value
    B. a point estimate
    C. a 95% confidence interval

26. Suppose that Researcher A reviews applications from 2,000 female applicants and 2,000 male applicants that were submitted for an entry-level position with an insurance company. One hundred female applicants were hired (5%), and 140 male applicants were hired (7%), and this difference of 2% was statistically significant (p=0.014). Researcher B instead submits the same application to each of 800 insurance companies, but the applicant's name is randomly assigned to be a stereotypical female name or a stereotypical male name. Across the companies, callbacks for an interview were offered for 20 of the 400 applications that had a stereotypical female name (5%) and for 37 of the 400 applications that had a stereotypical male name (9%), and this difference of 4% was statistically significant (p=0.039). Who had the better research design for inferring gender bias in the hiring process?
    A. Researcher A, because Researcher A had a lower p-value
    B. Researcher A, because Researcher A had a larger sample size
    C. Researcher B, because Researcher B used a randomized experiment
    D. Researcher B, because Researcher B had a higher estimated difference in callback rates

-- No calculators, phones, or other devices permitted. --

27. Suppose that all students at Faber College take a multiple-choice political science knowledge test. This is a valid test of political science knowledge. On this test, the mean percentage correct was 80% among political science majors and was 50% among art majors. Amy is a political science major at Faber College, and Bob is an art major at Faber College. Amy and Bob each scored 90% correct on the political science knowledge test. Which of the following is more likely?
    A. Amy likely truly has more political science knowledge than Bob has.
    B. Bob likely truly has more political science knowledge than Amy has.
    C. Neither of the above

28. Suppose that, for a randomized experiment, the estimated effect size of the treatment is 6, with a p-value of $p<0.05$ for a test of the null hypothesis that the treatment effect is zero. But the research design is biased in a way that causes the estimated effect size to be higher than the true effect size. Which, if any, of the following can we therefore conclude at the conventional level in political science, about the true treatment effect for this experiment?
    A. The treatment effect truly exists.
    B. The treatment effect truly does not exist.
    C. Neither of the above

29. Suppose that we want to estimate how much political ads influence vote choice in U.S. presidential elections. We could conduct a meta-analysis in which we combine the results from all studies that have ever been conducted on this research question, to produce an estimate of the average effect of how much political ads influence vote choice in U.S. presidential elections. Explain the benefit of the average effect size estimate in our meta-analysis being weighted by the sample size of the studies.

    We want to give more emphasis to the larger sample studies, because, compared to smaller sample studies, larger sample studies provide more evidence about an effect (larger sample studies have more participants, more countries, etc).

[Items 30 and 31] Suppose that a researcher is interested in assessing whether there is gender bias in salaries of university instructors. The researcher has the 800 students in a large university psychology course participate in an experiment. Each student is given a course syllabus and asked how much the instructor should be paid, on a scale that ranges from a low of $40,000 per year to a high of $100,000 per year. The experimental manipulation is that 400 of the students are given a syllabus in which the instructor's name is "Jen Wilson", which is intended to suggest that the instructor is a woman, and 400 of the students are given a syllabus in which the instructor's name is "Joe Wilson", which is intended to suggest that the instructor is a man. Everything else about the syllabus is the same.

30. Does this study have a high amount of internal validity?

A. Yes, because the study has a large sample.

B. Yes, because the study is a randomized experiment that manipulated only the perceived gender of the instructor.

C. No, because college students are plausibly not representative of persons who decide how much university instructors are paid.

D. No, because the study did not have enough control variables to eliminate all plausible alternate explanations.

31. Does this study have a high amount of external validity?

A. Yes, because the study has a large sample.

B. Yes, because the study is a randomized experiment that manipulated only the perceived gender of the instructor.

C. No, because college students are plausibly not representative of persons who decide how much university instructors are paid.

D. No, because the study did not have enough control variables to eliminate all plausible alternate explanations.

-- No calculators, phones, or other devices permitted. --

32. On 25 May 2020, George Floyd died in police custody. Police officer Derek Chauvin, who had been videoed kneeling on Floyd's neck, was later convicted of second-degree murder and other charges regarding Floyd's death. The day after Floyd's death, protests against police misconduct began and continued across the United States. Later data analysis indicated that the number of Floyd protests in a state positively correlated at $p<0.05$ with the number of police officers in the state who were later shot in the line of duty. One potential cause for this association is that the Floyd protests lowered people's opinion about police and therefore people were more likely to commit violence against police. But propose another explanation for why the number of Floyd protests in a state positively correlated with the number of police officers in the state who were later shot in the line of duty.

Multiple acceptable responses, such as the population size of a state. For example, all else equal, compared to small population states such as North Dakota, large population states such as California should be expected to have had more Floyd protests and more police shot in the line of duty, merely because these states have more people to protest and more police on the streets. Therefore, the association between Floyd protests and police shot could both be due to the size of the state population.

33. Suppose that a POL 138 instructor is interested in measuring how much their students learn during the Spring semester. The POL 138 class has 40 students at the start of the semester, and no student drops or adds the course. The instructor gives each of the 40 students in attendance at the first class meeting in January a pretest, and then, in April, at the last class meeting before the final exam, the instructor gives each of the 30 students in attendance at that class meeting a posttest; the other ten students were absent from that class meeting. The instructor estimates how much the 40 students in the class learned on average during the Spring semester, by taking the mean posttest score among the 30 students who took both the pretest and the posttest and then subtracting the mean pretest score among these 30 students. Select which of the following you expect, considering selection bias in taking of the posttest:

    A. Compared to the true average amount of learning among the 40 students in the POL 138 course, the instructor's estimate of student learning based on the scores from the 30 students who took the pretest and posttest will be biased lower.

    B. Compared to the true average amount of learning among the 40 students in the POL 138 course, the instructor's estimate of student learning based on the scores from the 30 students who took the pretest and posttest will be biased higher.

    C. Compared to the true average amount of learning among the 40 students in the POL 138 course, the instructor's estimate of student learning based on the scores from the 30 students who took the pretest and posttest will not be biased.

34. Explain the reason for the option that you selected in the prior item.

    It's plausible that the students who attend class regularly are more conscientious than students who don't attend class regularly, and it's plausible that these more conscientious students know the course material better, for example, maybe because these more conscientious students are more likely to study or maybe because the students who regularly attend class are more likely to learn something in class. Therefore, if the students who attend and thus take both the pretest and posttest know more about the course material, compared to students who did not attend to take the posttest, then the estimate from the 30 students who took both the pretest and posttest should be expected to be higher than the true average amount of learning among the 40 students in the POL 138 course.

    Students can get credit for this item and the prior item by selecting a different response and providing a plausible explanation. For example, maybe students who know the course material well decide that they don't need to attend class, so that the 10 students who don't attend class to take the posttest are actually the students who know the material the best, so that the estimate from the 30 students who took both the pretest and posttest should be expected to be lower than the true average amount of learning among the 40 students in the POL 138 course.

-- No calculators, phones, or other devices permitted. --

35. Below is adapted from the abstract for Zhang et al. 2021:

> Based on a convenience sample, 56 letters of recommendation (LORs) that program directors wrote for underrepresented (Black, Latinx, women) and non-underrepresented applicants to a cardiology fellowship were analyzed using directed qualitative content analysis. In the LORs, underrepresented applicants were more often described using doubt raising language and concerns about career trajectory. Results thus indicated that that program directors used language in a biased way to describe underrepresented applicants in the LORs.

Which of the following best indicates the validities of this study, for understanding whether letters of recommendation from program directors unfairly use biased language when describing underrepresented applicants?

A. low internal validity
B. high internal validity

36. Explain your response to the item above.

The research design did not address alternate explanations, such as recommenders having legitimate unbiased differences in perceptions about underrepresented applicants and non-underrepresented applicants.