POL 138 Quantitative Reasoning in Political Science
KEY Practice Exam 3

Completing this practice exam is not worth points toward your POL 138 grade, but it might be useful to take the practice exam and check your responses against the key. Content on Exam 3 should closely match content on Practice Exam 3, but with potentially different ways of measuring understanding of concepts, such as if short answer items are converted to multiple-choice items or vice versa, or if understanding of a concept is measured a different way.

9 Imperfect indicators of quality

9.1 Peer review

1. 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

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

9.2 Pre-registration

3. Franco et al. 2014 found that null results were ___ likely to be published than strong results were to be published.
   A. much less
   B. just as
   C. much more

4. 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

9.3 Meta-analysis

5. Study A has an estimated effect of 10 units and a sample size of 100 participants. Study B has an estimated effect of 2 units and a sample size of 200 participants. Which study should receive more weight in a meta-analysis?
   A. Study A, because Study A has a larger estimated effect size

B. Study B, because Study B has a larger sample size

6. Below are data for two studies:

Study	Estimated effect size	Sample size
A	2	300
B	5	200

Which of the following is a correct formula for calculating the mean estimated effect size, weighted by sample size?
A. 2×(300) + 5×(200)
B. 2×(300/200) + 5×(200/300)
C. 2×(300/500) + 5×(200/500)
D. 300×(2/7) + 200×(5/7)

7. Calculate the average estimated effect size across the two studies below, weighted by sample size, and rounded to one decimal place. Show all of your work.

Study	Estimated effect size	Sample size
A	70	600
B	20	100

Total sample size is 700, so:
Weighted mean = 70*(600/700) + 20*(100/700) = 62.9

8. Suppose that two studies tested the same hypothesis. Study 1 had an effect size of 20 and a sample size of 600, and Study 2 had an effect size of 40 and a sample size of 200. Calculate, to one decimal place, the overall mean effect size across these two studies, weighted by sample size. Show all of your work.

Total sample size = 600 + 200 = 800

Weighted mean = 20*(600/800) + 40*(200/800) = 25.0

9. Explain why a meta-analysis might be better than a single well-done study as a source for information about a research question.

Meta-analyses collect data from multiple studies, so these meta-analyses should have larger sample sizes than an individual study in the meta-analysis and thus have more information. Moreover, meta-analyses collect data from different studies, so any particular idiosyncrasy from a study should hopefully even out or be overpowered when combined with other studies.

10. A meta-analysis can use all studies available on a research question to produce an average estimate from these studies. Explain the benefit of the meta-analysis average being weighted by sample size.

Compared to studies that have a smaller sample, studies that have a larger sample have more data and typically provide more information about the research question, so these larger sample studies should receive more emphasis in the meta-analysis than the smaller sample studies receive.

11. Explain why a meta-analysis might be better than a single well-done study as a source for information about a research question.

Meta-analyses collect data from multiple studies, so these meta-analyses should have larger sample sizes than an individual study in the meta-analysis and thus have more information. Moreover, meta-analyses collect data from different studies, so any particular idiosyncrasy from a study should hopefully even out or be overpowered when combined with other studies.

10 Threats to inference 1

10.1 Selection bias

12. Suppose that, for a POL 138 course, the instructor gives a pretest at the first class meeting and then gives the same test as a posttest at the last class meeting before the final exam. Each of the 100 students registered for the course attended the first class meeting and took the pretest, but the instructor did not require attendance, and only 60 of the 100 students attended the final class meeting before the final exam and took the posttest. For the 60 students who took both the pretest and the posttest, results indicated that, on average, these students increased their pretest score by 20% from the pretest to the posttest. But, for the final exam taken by all 100 students, the increase in score was only 12% on average. Of the following, which is the most plausible reason that the pretest/posttest comparison overestimated the amount of learning in the class, compared to the estimated learning from the final exam?
A. Kelley's paradox
B. Regression toward the mean
C. Selection bias
D. Simpson's paradox

13. The SAT test is optional in Pennsylvania, and only 48% of Pennsylvania high school students in the class of 2023 took the SAT. The mean SAT score among these students was 1078. Suppose that Pennsylvania had instead required that all high school students in the class of 2023 take the SAT. Indicate whether you expect the mean SAT score among all Pennsylvania students in the class of 2023 to have been 1078, lower than 1078, or higher than 1078, if all Pennsylvania students took the SAT. Then explain why.

One possible response that would get credit: It seems plausible that students who took the SAT are, on average, better students who would be expected to score higher on the SAT, compared to students who did not take the SAT. If so, then the expectation is that the mean SAT score among all Pennsylvania students in the class of 2023 is lower than 1078.

10.2 Per capita

[Items 14 and 15] Suppose that the island country of Madripoor has 200,000 native residents and 50,000 immigrants. In 2023, 300 native Madripoor residents and 100 immigrants to Madripoor were a victim of a hate crime.

14. There were more hate crimes against ___.
    A. immigrants to Madripoor
    B. native Madripoor residents

    300 compared to 100

15. The per capita rate of being a victim of a violent crime was higher among ___.
    A. immigrants to Madripoor
    B. natives of Madripoor

    per capita calculations: 300/200,000 compared to 100/50,000, which is equivalent to 0.0015 compared to 0.0020

---

16. Suppose that, at the start of 2020, Freedonia had 1,000 Asian residents and 3,000 Black residents. During 2020, covid caused the death of 100 of these Asian residents and 200 of these Black residents. Which group had a higher per capita rate of death from covid-19 in 2020?
    A. Asian residents of Freedonia
    B. Black residents of Freedonia

    per capita calculations:

Asian	**100/1,000 = 0.10**
Black	**200/3,000 = 0.07**

10.3 Influential outliers

[No exam items]

10.4 Using a less relevant measure

17. The Bone Student Scholar award is the most prestigious undergraduate student award at Illinois State University and is given to students who have at least four semesters at ISU

and who have a cumulative GPA of at least 3.7. For the 2021-2022 school year, 14 students were named a Bone Student Scholar. Nine of these 14 students were female (64%). To assess the extent to which this percentage is fair, which of the following pieces of information would be more useful to know?

A. the percentage of female students among all ISU students
B. the percentage of female students among all ISU students who have at least four semesters at ISU and who have a cumulative GPA of at least 3.7

To assess the extent to which this percentage is fair, we can get a comparison group that reflects the selection that would occur if the selection were fair. In this case, we would rather have an academically excellent comparison group, because a fair selection of Bone Student Scholars would be academically excellent.

18. Suppose that each public high school in Illinois selects their three best math students to represent the school in a competition answering math questions. For each student, we have data on the student's score on the most recent state standardized math test. EXPLAIN which one of the following would be most useful for predicting which school will win the competition:

- the 1st percentile score for each school on the state standardized math tests
- the mean score for each school on the state standardized math tests
- the median score for each school on the state standardized math tests
- the 99th percentile score for each school on the state standardized math tests

The 99th percentile score is the score under which 99 percent of scores fall; students at the 99th percentile are the highest scoring students, among the options available in the item. So, the higher this 99th percentile score is for a school, the better the best students at the school have performed on the standardized math tests. For this item, we are interested in the performance of the best students at each school, so, of the options, that 99th percentile score for each school would provide the best sense of how the best students in one school compare to the best students in the other schools.

19. Suppose that the political science department at Faber College administers a political science knowledge test to all of its political science majors. The department issues a "political science expert" award only to the ten students who have the ten highest scores on the political science knowledge test. For this test, the mean score among the 500 female political science majors was 50, and the mean score among the 500 male political science majors was 40, but the department gave 7 of the 10 "political science expert" awards to male political science majors. Explain how male political science majors could fairly receive the majority of the "political science expert" awards if female political science majors had a higher mean score on the test than male political science majors did.

The "political science expert" awards were given to students at the high end of the distribution of the test, so it is possible that the scores among male political science majors had a higher standard deviation than the scores among female political science majors.

10.5 Measurement error

20. Suppose that, over the past ten years, the reported number of burglaries has substantially decreased in Freedonia City. One possible explanation for this is that prevalence of burglary has decreased. But what else might this be due to?

Another potential explanation is that merely the reporting of the burglaries has decreased. Maybe, for instance, it has become well known that police have not recently investigated or solved many burglaries, so that a lot of people don't bother to report burglaries. In that case, the reported number of burglaries could decrease even if the number of burglaries has not decreased.

10.6 Restriction of range

21. Suppose that researchers test whether political science majors know more about politics on average than business majors know. Researchers claim that, in their data, political science majors do not know more about politics on average than business majors know. One explanation for this finding is that, in reality, political science majors do not know more about politics on average than business majors know. But explain how this finding might merely have been due to restriction of range in the researchers' research design.

The test that researchers used to measure knowledge about politics might have been so easy or so difficult that the test was not able to detect a true difference of knowledge about politics, on average, between political science majors and business majors.

10.7 Confounders

22. Suppose that an analysis indicated that, in head-to-head matchups between major party candidates in the United States, the candidate that raises more money wins in 90% of elections. Suppose also that the p-value is $p<0.05$ for a test of the null hypothesis that the candidate that raised more money was just as likely to win election as the candidate that raised less money. Explain whether this would be sufficient evidence at the conventional level of political science to conclude that raising more money than the other candidate causes a candidate to be more likely to win an election.

No, because there are reasonable alternate explanations that must first be addressed. For example, maybe candidates who are winning in the polls are more likely to receive campaign donations because they are winning in the polls. In that case, the chance of winning causes the donations and not vice versa.

10.8 Miscontrolling

23. Suppose we test whether variation in X causes variation in Y. Which of the following would be worse to add as a predictor to that regression?
A. a variable A that influences X and influences Y
B. a variable B that is influenced by X and is influenced by Y

The ideal for control variables is to help get the cases that we are comparing to be equal at the point of the difference in treatment, so that we can estimate the effect of the treatment. Therefore, we want to control for differences that have occurred before the difference in treatment, and we do not want to control for differences that occur because of the difference in treatment.

24. Omission of a relevant control variable in a non-experimental analysis ___.
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

25. Inclusion of an irrelevant control variable in a non-experimental analysis can ___.
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

10.9 Reverse causality and reciprocal causality

[No exam items]

11 Threats to Inference 2

11.1 Misinterpreting p>0.05

26. Sometimes a research study does not permit an inference that the effect of a treatment differs from zero; these inconclusive studies can be referred to as producing a null result. An "informative null" can be used to refer to a study that had a null result but nonetheless provided information that the treatment effect is at most small. Of the measures below, which measure is most useful for assessing whether a result from a study can be appropriately referred to as an informative null?
A. a p-value
B. a standard deviation
C. a 95% confidence interval

27. Suppose that researchers in Latveria are interested in whether married men have a different income on average than never married men have. For a representative sample of Latverian men, the researchers compare the mean income of married Latverian men to the mean income of never married Latverian men. The p-value is p=0.25 for a test of the

null hypothesis that these means equal each other. Which of the 95% confidence intervals below for the difference in mean incomes would be a more informative null?

A. [-$200, +$300]
B. [-$40,000, +$40,000]

The 95% confidence interval of [-$200, +$300] is more precise ($500 wide), compared to the 95% confidence interval [-$40,000, +$40,000], which is $80,000 wide. The 95% confidence interval of [-$200, +$300] better permits us to conclude that, if the true difference is not zero, then the true difference is at least relatively small.

11.2 Misinterpreting differences in statistical significance

28. Amy flips her coin a certain number of times, and Bob flips his coin a different number of times. Amy and Bob then each test the null hypothesis that their coin is fair. Amy's p-value is p=0.04, and Bob's p-value is p=0.50. Is this sufficient evidence to conclude at the conventional level in political science that Amy's coin is less fair than Bob's coin?

A. Yes
B. No

A common bad inference is to infer a difference between estimates merely because the p-value for one estimate falls below a p-value threshold (and is thus "statistically significant") and the p-value for the other p-value does not fall below that threshold (and is thus not "statistically significant"). This is a bad inference because, to infer something about the difference between estimates, we should have a p-value about that difference in estimates. p-values about each inference aren't useful for making an inference about the difference between the estimates.

29. Suppose that researchers test the effect of a treatment. Results provide evidence at p=0.03 that the treatment worked among men participants, but the p-value is p=0.20 for the test of the effect among women participants. Is this sufficient evidence to support the conclusion 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.

11.3 Multiple testing

30. Suppose that Researcher A and Researcher B independently test whether a treatment has an effect. Researcher A conducts a randomized experiment with a sample size of 800, to assess whether the treatment has an effect. Researcher B conducts a different randomized experiment with a sample size of 1600, to assess whether the 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 probability that at least one of these experiments produced 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 50%
E. 50%
F. above 50% but less than 95%
G. 95%
H. above 95% but less than 100%
I. 100%

If the null hypothesis is true, then there is a 5% chance that the p-value from a test of the null hypothesis is $p<0.05$. If we conduct more than one test of the null hypothesis, then that chance increases by some number less than 5%.

11.4 Regression toward the mean

31. Suppose that data from the Freedonia Department of Transportation indicate that the number of traffic crash fatalities from 2016 through 2022 had been 978, 990, 935, 910, 996, and 1134. After the jump from 996 fatalities to 1134 fatalities, Freedonia state troopers began a program to more frequently ticket speeding vehicles. The next year, the number of traffic crash fatalities fell from 1134 to 960. Using these data to conclude that the new ticketing policy caused the drop from 1134 to 960 best reflects a lack of consideration of which of the following?

A. an ecological fallacy
B. Kelley's paradox
C. regression toward the mean
D. Simpson's paradox

11.5 Ecological fallacy

[No exam items]

11.6 Simpson's paradox

32. Suppose that, in group A, men are on average older than women are. Suppose that, in group B, men are on average older than women are. If group A and group B are combined into group C, then, in group C, ___.

A. men will be older on average than women are
B. men will not necessarily be older on average than women are

11.7 Heterogenous effects

33. Suppose that we conduct a randomized experiment to estimate the effect of a treatment, but we do not detect sufficient evidence that the treatment differs from zero. 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 that regress toward the mean

11.8 Participant effects

[No exam items]

11.9 Lack of external validity

34. 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

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

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

37. Suppose that a researcher tests for racial bias. The researcher conducts a randomized experiment in which a large sample of students from the local college are randomly given a story about a Black man convicted of a DUI or a White man convicted of a DUI. Everything else in the story is the same, except for the race of the man convicted of a DUI. The researcher analyzes the responses to see whether the mean sentence length recommended for the Black man convicted of a DUI differs from the mean sentence length recommended for the White man convicted of a DUI. Does this study have a high amount of internal validity?
A. Yes, because the study is a randomized experiment that manipulated only the man's race.
B. Yes, because the study has a large sample.
C. No, because college students are plausibly not representative of persons in the legal system who will decide in real life the sentence length for a DUI.
D. No, because the study did not have enough control variables to eliminate all plausible alternate explanations.

38. Suppose that a researcher tests for racial bias. The researcher conducts a randomized experiment in which a large sample of students from the local college are randomly given a story about a Black man convicted of a DUI or a White man convicted of a DUI.

Everything else in the story is the same, except for the race of the man convicted of a DUI. The researcher analyzes the responses to see whether the mean sentence length recommended for the Black man convicted of a DUI differs from the mean sentence length recommended for the White man convicted of a DUI. Does this study have a high amount of external validity?

A. Yes, because the study is a randomized experiment that manipulated only the man's race.
B. Yes, because the study has a large sample.
C. No, because college students are plausibly not representative of persons in the legal system who will decide in real life the sentence length for a DUI.
D. No, because the study did not have enough control variables to eliminate all plausible alternate explanations.

11.10 Researcher bias or researcher error

39. Suppose that results from a large sample random representative survey of U.S. residents indicates that the percentage of Republicans who agree with the statement that "Covid-19 is less dangerous than the flu" is higher than the percentage of Democrats who agree with that statement, with a p-value of $p<0.05$ for a test of the null hypothesis that these percentages equal each other. Presuming that it is not true that Covid-19 is less dangerous than the flu, are these data sufficient to conclude that, at least at the time of the survey, the percentage of Republicans in the United States who were misinformed about the severity of Covid-19 is higher than the percentage of Democrats in the United States who were misinformed about the severity of covid-19?

A. Yes
B. No

People can be misinformed about the severity of covid-19 in multiple ways, such as underestimating its severity and overestimating its severity. The statement that "Covid-19 is less dangerous than the flu" can measure only underestimating the severity of covid-19 and thus cannot be used to conclude anything about which group is more misinformed about the severity of covid-19.

12 Ethical issues

12.1 IRBs

40. IRBs are organizations designed to protect human subjects in scientific research. Can an IRB, in some circumstances, permit research involving human participants in which the human participants do not provide informed consent?

A. Yes
B. No

12.2 Statistical discrimination

41. 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

12.3 Kelley's paradox

42. Suppose that, for a multiple-choice political science knowledge test, students from State A have a mean score of 40 and students from State B have a mean score of 60. Scores in both states follow a normal distribution, and the standard deviation of scores in both states is 10. Because of Kelley's paradox, a randomly selected student from State A who has a score of 80 will be expected to be more likely to have ___ true level of political science knowledge than a randomly selected student from State B who has a score of 80.
A. a lower
B. the same
C. a higher