

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

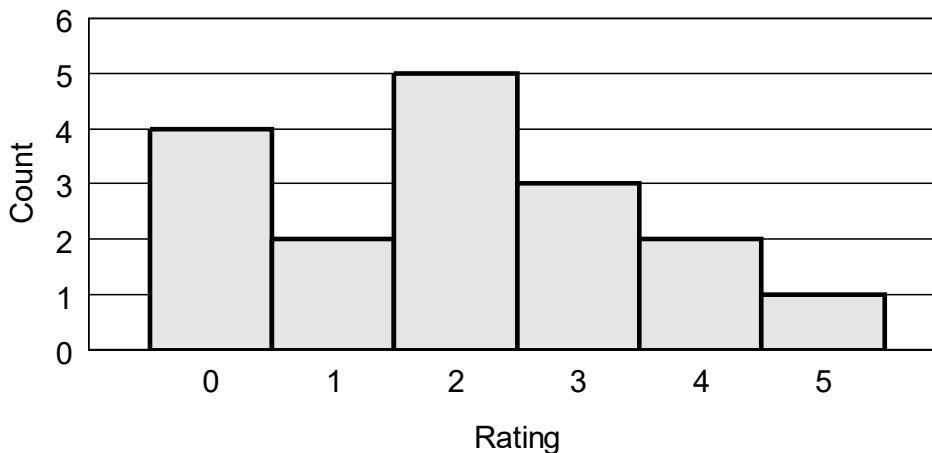
First Name and Last Name \_\_\_\_\_

**This practice exam focuses on material for the Final Exam, but material eligible for the Final Exam includes material from prior exams and from online notes for any chapters. This practice exam has a few open-ended items that could be converted to multiple-choice items.**

*[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 Practice Final Exam · Fall 2025**

1. Research focusing on numbers is \_\_\_\_.  
A. qualitative research  
B. quantitative research
2. Which of these is closest to what an inference is?  
A. a logical fallacy  
B. a conclusion  
C. the reason for a prediction
3. The histogram below indicates how a person rated 17 movies on a scale from 0 to 5. In this histogram, which is true?  
A. There is 1 rating of 2.  
B. There is 1 rating of 5.



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

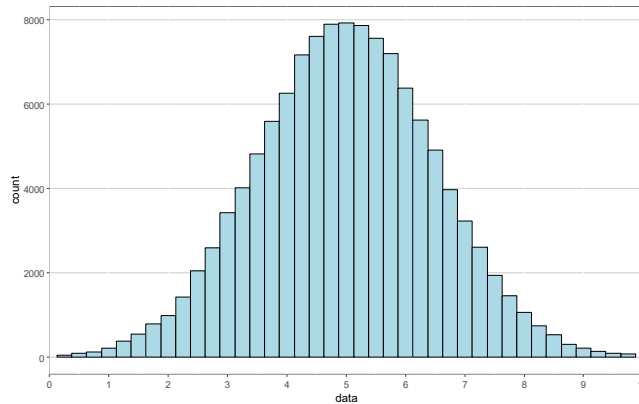
4. Adding an outlier to a set of data would be expected to have more influence on the \_\_\_ of the data.
  - A. mean
  - B. median
  
5. Removing an outlier from a set of data would be expected to have more influence on the \_\_\_ of the data.
  - A. mean
  - B. median
  
6. Standard deviation is a measure of \_\_\_\_\_.
  - A. validity
  - B. variation
  - C. reliability
  - D. correctness
  - E. central tendency
  
7. Which set of numbers {0, 2} or {12, 14} has a larger standard deviation?
  - A. {0, 2}
  - B. {12, 14}
  - C. The standard deviations for the two sets would be the same.
  
8. Which set of numbers {0, 3} or {12, 14} has a larger standard deviation?
  - A. {0, 3}
  - B. {12, 14}
  - C. The standard deviations for the two sets would be the same.
  
9. Suppose that a sample has 10 Democrats and 15 Republicans. What formula could be used to correctly calculate the proportion of the sample that is Democrat?
  - A.  $10 \div 15$
  - B.  $15 \div 10$
  - C.  $10 \div (10 + 15)$
  - D.  $(10 - 15) \div (10 + 15)$
  - E.  $(15 - 10) \div (10 + 15)$
  
10. Suppose that, in 2023, 40% of students at a college are women, but that, in 2024, 50% of students at the college are women. That change can be correctly expressed as an increase of \_\_\_\_\_.
  - A. 25 percent
  - B. 25 percentage points

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

11. Suppose that, in 2023, 40% of students at a college are women, but that, in 2024, 50% of students at the college are women. That change can be correctly expressed as an increase of \_\_\_\_\_.
- A. 10 percent
  - B. 10 percentage points
12. Which score below indicates a higher degree of political knowledge for a political knowledge test?
- A. scoring at the 15th percentile on the test
  - B. scoring at the 95th percentile on the test
13. The LSAT is a test that applicants to law school typically take. An LSAT score of 170 is at the 96th percentile. What does this indicate?
- A. An LSAT score of 170 is below 96 percent of LSAT scores
  - B. An LSAT score of 170 is above 96 percent of LSAT scores
  - C. A LSAT with a score of 170 was an LSAT that had 96 percent of items correct.
  - D. A LSAT with a score of 170 was an LSAT that had 96 percent of items incorrect.
  - E. None of the above
14. Suppose that a course has two exams: Exam 1 is worth 30% of the overall grade for the course, and Exam 2 is worth 70% of the overall grade for the course. If a student scored an 80% on Exam 1 and a 60% on Exam 2, which of the following could be used to correctly calculate that student's overall percentage for the course?
- A.  $(80 + 60) \div 2$
  - B.  $(0.30 \times 70) - (0.80 \times 60)$
  - C.  $(0.30 \times 80) + (0.70 \times 60)$
  - D.  $(0.30 \times 80) \times (0.70 \times 60)$
  - E.  $(0.30 \times 80) \div (0.70 \times 60)$
15. Political scientists weight survey data for which one of the following reasons?
- A. Because the sample is too small
  - B. Because the population is much larger than the sample
  - C. Because the sample characteristics do not match the population characteristics
16. Suppose that a test has a mean of 100 and a standard deviation of 10. Scores on the test follow a normal distribution. About 95% of scores should fall within which two scores?
- A. 0 and 100
  - B. 60 and 140
  - C. 70 and 130
  - D. 80 and 120
  - E. 90 and 110

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

17. The image below is an example of a \_\_\_\_.



- A. uniform distribution
- B. normal distribution

18. The sample for a research study is better described as which of the following?

- A. the things that are studied
- B. the things that the research study is interested in

19. The population for a research study is better described as which of the following?

- A. the things that are studied
- B. the things that the research study is interested in

20. Suppose that scores on a national test follow a normal distribution and have a mean of 100 and a standard deviation of 10. If Student A raises her score from 90 to 100, and Student B raises her score from 110 to 120, which of the following statements is true?

- A. Student A had the higher percentile increase on the test.
- B. Student B had the higher percentile increase on the test.
- C. Student A had the same percentile increase on the test as Student B had.

21. Suppose that scores on a national test follow a normal distribution and have a mean of 100 and a standard deviation of 10. If Student A raises her score from 90 to 100, and Student B raises her score from 100 to 110, which of the following statements is true?

- A. Student A had the higher percentile increase on the test.
- B. Student B had the higher percentile increase on the test.
- C. Student A had the same percentile increase on the test as Student B had.

22. Which best indicates what the null hypothesis is?

- A. The hypothesis that is true
- B. The hypothesis being tested
- C. The hypothesis that the effect is not zero
- D. The hypothesis that is most supported by the evidence

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

23. Of the following, which best describes what a p-value measures?
- A. the precision of an estimate
  - B. the size of an association controlling for other factors
  - C. the strength of evidence against the null hypothesis
24. Of the p-values below, which p-value is the strongest evidence against the null hypothesis?
- A. 0.03
  - B. 0.05
  - C. 0.91
  - D. 1.00
25. If we flipped a coin 18 times and got 10 heads and 8 tails, what would be the p-value for a test of the null hypothesis that the coin is fair?
- A. 0
  - B. 1
  - C. something between 0 and 1
26. If we flipped a coin 18 times and got 0 heads and 18 tails, what would be the p-value for a test of the null hypothesis that the coin is fair?
- A. 0
  - B. 1
  - C. something between 0 and 1
27. If we flipped a coin 18 times and got 9 heads and 9 tails, what would be the p-value for a test of the null hypothesis that the coin is fair?
- A. 0
  - B. 1
  - C. something between 0 and 1
28. What is the conventional p-value threshold in political science?
- A. 0.01
  - B. 0.05
  - C. 0.50
  - D. 0.95
  - E. 0.99
  - F. 1
29. For a test of the null hypothesis that there is no association, "statistically significant" evidence for the association refers to \_\_\_\_.
- A. sufficient evidence that the association exists
  - B. sufficient evidence that the association is large

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

30. If the p-value is  $p=0.00001$  for a single statistical test of a null hypothesis that there is no association, do we have enough evidence to claim that there is statistically significant evidence for the association?
- A. Yes
  - B. No
31. If the p-value is  $p=0.00001$  for a single statistical test of a null hypothesis that there is no association, do we have enough evidence to claim that the association is very large?
- A. Yes
  - B. No
32. If the p-value for a test of a null hypothesis is  $p=0.99$ , then we should do which of the following?
- A. accept the null hypothesis and accept the alternative hypothesis
  - B. reject the null hypothesis and reject the alternative hypothesis
  - C. accept the null hypothesis and reject the alternative hypothesis
  - D. reject the null hypothesis and accept the alternative hypothesis
  - E. none of the above
33. If the p-value for a test of a null hypothesis is  $p=0.01$ , then we should do which of the following?
- A. accept the null hypothesis and accept the alternative hypothesis
  - B. reject the null hypothesis and reject the alternative hypothesis
  - C. accept the null hypothesis and reject the alternative hypothesis
  - D. reject the null hypothesis and accept the alternative hypothesis
  - E. none of the above
34. A researcher tested the null hypothesis that an association is zero. The p-value for this test is  $p=0.03$ . Based on this p-value, which of the following should the researcher do, using the conventional level in political science?
- A. conclude that the association is zero
  - B. conclude that the association is not zero
  - C. neither of the above
35. A researcher tested the null hypothesis that an association is zero. The p-value for this test is  $p=0.70$ . Based on this p-value, which of the following should the researcher do, using the conventional level in political science?
- D. conclude that the association is zero
  - E. conclude that the association is not zero
  - F. neither of the above
36. A false positive occurs when we falsely believe that something is present, when in reality that thing is not present. Which one of the following p-value thresholds would be expected to produce a lower chance of a false positive?
- A.  $p=0.01$
  - B.  $p=0.10$

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

[Items 37 and 38] Suppose that the Faber College student body has this distribution of political party memberships: 3,000 Democrats, 1,000 Independents, 2,000 Republicans, and 10 Libertarians. Faber College randomly selects 5% of students to receive a scholarship. But, due to random error, not every political party has the same percentage of scholarship winners.

37. Which political party below is most likely to have had the lowest percentage of its membership win a scholarship?

- A. Democrats
- B. Independents
- C. Republicans
- D. Libertarians

38. Which political party below is most likely to have had the highest percentage of its membership win a scholarship?

- A. Democrats
- B. Independents
- C. Republicans
- D. Libertarians

---

39. Randomly assigning participants to groups helps a randomized experiment identify causes by \_\_\_\_.

- A. eliminating demand effects as much as possible
- B. helping as much as possible to avoid regression toward the mean
- C. getting the sample to be as representative of the population as possible without weighting
- D. getting the groups to be as similar to each other as possible before the difference in treatment

40. Suppose that, in a randomized experiment, the mean response from participants in the control group differs from the mean response from participants in the treatment group. One reason for this is that participants in the control group were treated differently than participants in the treatment group were treated. The other possible reason why the mean response from participants in the control group differed from the mean response from participants in the treatment group is \_\_\_\_.

- A. a ceiling effect
- B. Simpson's paradox
- C. random assignment error
- D. regression toward the mean

41. How do control variables help improve causal inference in a correlational study?

- A. reduce bias in measurements
- B. randomly assign participants
- C. remove sample bias
- D. help address alternate explanations

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

42. Which of the following best indicates what a placebo is?
- A. a treatment that has an effect
  - B. a treatment that has no effect
  - C. a treatment that has a positive effect
  - D. a treatment that has a negative effect
43. Suppose that researchers are interested in the extent, if any, to which having an internship at ISU affects the salary of a student six months after graduation. Six months after graduation, Researcher A compares the mean salary among ISU graduates who had an internship to the mean salary among ISU graduates who did not have an internship. Six months after graduation, Researcher B instead compares the mean salary among ISU graduates who had at least a 3.0 GPA and who had an internship to the mean salary among ISU graduates who had at least a 3.0 GPA and who did not have an internship. An advantage of Researcher B's research design over Researcher A's research design is that \_\_\_\_.
- A. Researcher B will address an alternate explanation
  - B. Researcher B will avoid bias due to regression toward the mean
  - C. Researcher B will avoid Simpson's paradox
  - D. Researcher B will have a smaller sample size
44. The table below indicates that, among a hypothetical set of men academics and women academics, 80 percent of men who applied for tenure received tenure, but only 40 percent of women who applied for tenure received tenure.

<b>Men academics</b>		<b>Women academics</b>	
<b>Tenure</b>	<b>Grants</b>	<b>Tenure</b>	<b>Grants</b>
No	0	No	0
Yes	1	No	0
Yes	1	No	0
Yes	2	Yes	1
Yes	2	Yes	2

Considering only statistical control for the number of grants, which inference below is most supported?

- A. The data suggest that women were less likely than men to receive tenure, controlling for number of grants.
- B. The data suggest that women were as likely as men to receive tenure, controlling for number of grants.
- C. The data suggest that women were more likely than men to receive tenure, controlling for number of grants.

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

45. Suppose that researchers survey legislative interns in the Freedonia Congress after the interns have served their legislator for two years. Results indicate that interns for female legislators have lower levels of modern sexism, on average, compared to interns for male legislators. The p-value is  $p < 0.05$  for a test of the null hypothesis that the mean level of modern sexism among interns for female legislators equals the mean level of modern sexism among interns for male legislators. Would this be sufficient evidence at the conventional level in political science to conclude that interning for a female legislator caused lower levels of modern sexism, at least among this set of interns and at least on average?
- A. Yes  
B. No
46. Suppose that the Oceania Congress randomly assigns interns to legislators. Results from a survey after the interns have served their legislators for two years indicate that interns randomly assigned to female legislators have lower levels of modern sexism, on average, compared to interns randomly assigned to male legislators. The p-value is  $p < 0.05$  for a test of the null hypothesis that the mean level of modern sexism among interns randomly assigned to female legislators equals the mean level of modern sexism among interns randomly assigned to male legislators. Would this be sufficient evidence at the conventional level in political science to conclude that interning for a female legislator caused lower levels of modern sexism, at least among this set of interns and at least on average?
- A. Yes  
B. No
47. On May 25, 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. Suppose that we want to assess the extent to which this incident affected the U.S. public's attitudes about police. Of the following, which would be a better research design?
- A. Compare the U.S. public's attitudes about police in the week before May 25, 2020, to the U.S. public's attitudes about police in the week after May 25, 2020.  
B. Compare the U.S. public's attitudes about police in the year before May 25, 2020, to the U.S. public's attitudes about police in the year after May 25, 2020.
48. Suppose that a researcher has the data below, from two participants (ID #1 and ID #2) from a survey in April and a survey in June of the same year. For each participant and for both months, the dataset has an indication of the participant's political party (D or R) and an indication of whether the participant supports or opposes affirmative action:

ID	April	June
1	D + Oppose	R + Oppose
2	R + Support	D + Support

- Based on these data only, which of the following is more supported?
- A. political party influences attitudes about affirmative action  
B. attitudes about affirmative action influences political party

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

49. Suppose that a researcher is interested in whether a Study Abroad experience affects students' political tolerance. The researchers collected data on June 1 (before the Study Abroad experience) and on August 1 (after the Study Abroad experience) from students who went on a Study Abroad experience and from similar students who did not go on a Study Abroad experience. Suppose that the researcher's data are in the table below, in which political tolerance is measured from 0 for extremely intolerant to 100 for extremely tolerant.

Group	Mean political tolerance	
	June 1	August 1
Study Abroad students	80	90
Comparison students	71	80

If the researcher used a difference-in-differences design that compared Study Abroad students to the comparison students, the researcher's estimate of the effect of Study Abroad on political tolerance would be that Study Abroad made students in the sample on average \_\_\_\_\_.

- A. about 1 unit more tolerant
  - B. about 3 units more tolerant
  - C. about 9 units more tolerant
  - D. about 10 units more tolerant
50. Suppose that, on January 1, 2024, Freedonia enacted the Unemployment Reduction Act. Researchers are interested in assessing the extent to which the Unemployment Reduction Act caused a change in Freedonia's unemployment rate. Atlantica and Oceania are countries immediately next to Freedonia and are similar to Freedonia in every way, except that Atlantica and Oceania did not enact any legislation to reduce unemployment.

	Unemployment Rate			
	2021	2022	2023	2024
Freedonia	9%	8%	9%	2%
Atlantica	10%	9%	10%	2%
Oceania	9%	8%	9%	2%

Considering a difference-in-differences method, what do the data in the table above suggest about the decrease in unemployment in Freedonia between 2023 and 2024?

- A. The data are consistent with the Unemployment Reduction Act being the reason for the decrease in unemployment in Freedonia between 2023 and 2024.
- B. The data are not consistent with the Unemployment Reduction Act being the reason for the decrease in unemployment in Freedonia between 2023 and 2024.

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

51. Suppose that each resident in Freedonia were offered an opportunity to take a flu vaccine on January 1, 2024. Some Freedonians took the flu vaccine, and some did not. Freedonians who took the flu vaccine must pay \$5 for the vaccine, but Freedonians who are over age 60 and/or who have an income under the poverty line got the flu vaccine for free. Suppose that researchers have data on all Freedonians, which indicates that Freedonians who took the flu vaccine were more likely to die in 2024, compared to Freedonians who did not take the flu vaccine. The p-value is  $p < 0.001$  for a test of the null hypothesis that the death rate among Freedonians who took the flu vaccine equals the death rate among Freedonians who did not take the flu vaccine.

Is this sufficient evidence at the conventional level in political science to conclude that taking the flu vaccine increased the probability that a Freedonian died in 2024?

- A. Yes, because Freedonians who took the flu vaccine were more likely to die, compared to Freedonians who did not take the flu vaccine, and the p-value for this difference is less than  $p = 0.05$ .
- B. No, because Freedonians who took the flu vaccine might have differed from Freedonians who did not take the flu vaccine in relevant characteristics other than taking the flu vaccine, such as the age of the persons who took the vaccine.
52. Membership in the Pi Sigma Alpha national political science honors society requires, among other things, at least a 3.0 GPA. Suppose that we are interested in estimating the effect, if any, that membership in Pi Sigma Alpha has had on the future income of college graduates. Researcher A plans to compare future income among any college graduates who were in Pi Sigma Alpha to future income among any college graduates who were not in Pi Sigma Alpha. Researcher B instead plans to compare future income among any college graduates who were in Pi Sigma Alpha and had a 3.0 GPA to future income among any college graduates who were not in Pi Sigma Alpha and had a 3.0 GPA. Which researcher has the better research design for estimating the effect, if any, that membership in Pi Sigma Alpha has had on the future income of college graduates?
- A. Researcher A, because Researcher A will have a much larger sample size and thus have a more precise estimate of the effect, if any, that membership in Pi Sigma Alpha has had on future income.
- B. Researcher B, because Researcher B's comparison will better isolate the effect of membership in Pi Sigma Alpha, because both of Researcher B's groups will be students of similar academic ability.

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

53. Recent data indicate that about 12% of U.S. residents receive "food stamp" benefits from the Supplemental Nutrition Assistance Program. Suppose that we conducted a large random survey of U.S. residents that asked participants to respond to the item below:

What percentage of U.S. residents receive food stamp benefits?

- Less than 15%
- 15% to 24%
- 25% to 50%
- More than 50%

Suppose that Republicans were less likely than Democrats to select the correct response of "Less than 15%", with a p-value of  $p < 0.05$  for the test of the null hypothesis that the percentage correct among Republicans equals the percentage correct among Democrats. Would that be sufficient evidence at the conventional level in political science to conclude that, at least at the time of the survey and at least on average, Republicans were more likely than Democrats were to be misinformed about the percentage of U.S. residents who receive food stamp benefits?

- A. Yes, because Republicans were less likely to select the correct response about the percentage of U.S. residents who receive food stamp benefits.
- B. No, because the survey item can identify only the misinformation of overestimating the percentage of U.S. residents who receive food stamp benefits, and Democrats might be more likely than Republicans to underestimate the percentage of U.S. residents who receive food stamp benefits.

[Items 54 and 55] A researcher randomly selects 2,000 people from the population of U.S. residents and then randomly assigns 1,000 of these people to a group that receives Treatment A and randomly assigns the other 1,000 people to a group that receives Treatment B.

54. The random assignment to groups \_\_\_\_.
- A. better permits the researcher to make an inference about U.S. residents
  - B. better permits the researcher to make an inference about whether Treatment A has a different effect than Treatment B has among participants in the sample
55. The random selection from the population \_\_\_\_.
- A. better permits the researcher to make an inference about U.S. residents
  - B. better permits the researcher to make an inference about whether Treatment A has a different effect than Treatment B has among participants in the sample

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

56. Suppose that, in a randomized experiment, the mean response from participants in the control group differs from the mean response from participants in the treatment group. One reason for this is that participants in the control group were treated differently than participants in the treatment group were treated. The other possible reason why the mean response from participants in the control group differed from the mean response from participants in the treatment group is \_\_\_\_.
- A. a ceiling effect
  - B. Simpson's paradox
  - C. random assignment error
  - D. regression toward the mean

[Items 57 and 58] Each table below has data for three persons in three months, indicating each person's LSAT score on the first day of that month and whether each person took an LSAT prep course that month. In both tables, there is a positive association between LSAT score and taking an LSAT prep course.

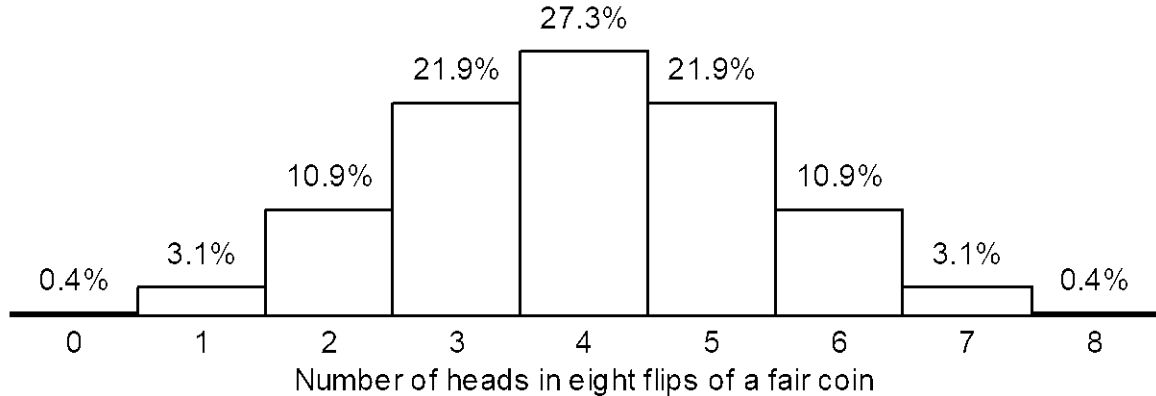
Table A			
Person	Month	LSAT score	Prep Course?
1	1	150	No
1	2	150	No
1	3	150	No
2	1	150	No
2	2	155	Yes
2	3	160	Yes
3	1	140	No
3	2	150	Yes
3	3	160	Yes

Table B			
Person	Month	LSAT score	Prep Course?
1	1	120	No
1	2	120	No
1	3	120	No
2	1	170	No
2	2	170	Yes
2	3	170	Yes
3	1	160	No
3	2	160	Yes
3	3	160	Yes

57. For which table do the data more strongly suggest that taking an LSAT prep course raises a test-takers LSAT score?
- A. Table A
  - B. Table B
58. For which table do the data more strongly suggest that high-scoring test-takers are more likely to take an LSAT prep course?
- A. Table A
  - B. Table B

---

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



59. The plot above is the distribution of the number of heads expected from eight flips of a fair coin. Which one of the following calculations will indicate the p-value for a test of the null hypothesis that a coin is fair, for a coin that lands on heads 2 times in 8 flips? Note that the numbers in the options below are decimal percentages, so that, for example, 3.1% is written as 0.031.
- A. 0.109
  - B.  $0.109 + 0.109$
  - C.  $0.109 + 0.219 + 0.273 + 0.219 + 0.109$
  - D.  $0.004 + 0.031 + 0.109 + 0.109 + 0.031 + 0.004$
60. GDP per capita rose in Fredonia 3 percent per year from 2020 to 2023. At the start of 2024, Fredonia enacted the GDP Growth Act, and, in 2024, GDP per capita rose 5 percent. For estimating how the GDP Growth Act affected GDP per capita growth in Fredonia, which of the following would provide the better comparison for a difference-in-differences design, based on only the information indicated below?
- A. Atlantica, in which GDP per capita rose by 5 percent per year from 2020 to 2023
  - B. Oceania, in which GDP per capita rose by 3 percent per year from 2020 to 2023
  - C. Pacifica, in which GDP per capita rose by 5 percent per year in 2024
  - D. Arctica, in which GDP per capita rose by 3 percent per year in 2024
61. Study A has an estimated effect size of 4, and Study B has an estimated effect size of 7. Study A has a sample size of 500 participants, and Study B has a sample size of 200 participants. Which study should receive more weight in a meta-analysis?
- A. Study A, because Study A has a smaller estimated effect size
  - B. Study A, because Study A has a larger sample size
  - C. Study B, because Study B has a larger estimated effect size
  - D. Study B, because Study B has a smaller sample size
62. 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

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

63. 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
64. 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, compared to studies that provided strong evidence of an effect, studies that did not provide sufficient evidence of an effect were \_\_\_\_.
- A. much less likely to be published
  - B. just as likely to be published
  - C. much more likely to be published
65. 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
66. 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
67. 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
68. 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

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

69. 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
70. Which type of validity concerns the ability to make correct claims about the sample?
- A. internal validity
  - B. external validity
71. Which type of validity concerns the ability to generalize to the population in the real world?
- A. internal validity
  - B. external validity
72. The following passage from the Allington et al. 2010 article in Reading Psychology, entitled "Addressing Summer Reading Setback Among Economically Disadvantaged Elementary Students", describes an experiment:

...852 students from 17 high-poverty schools were randomly selected to receive a supply of self-selected trade books on the final day of school over a 3-year period, and 478 randomly selected students from these same schools received no books and served as the control group. No further effort was provided in this intervention study. Outcomes on the state reading assessment indicated a statistically significant effect ( $p = .015$ ) for providing access to books for summer reading...

Which one of the following, if any, is a valid criticism of this study?

- A. The result is likely biased because the treatment group was larger than the control group.
  - B. The p-value is not small enough to eliminate random assignment error as a plausible alternate explanation at the conventional level in political science.
  - C. The study did not include control variables to address alternate explanations.
  - D. None of the above.
73. Suppose that we conduct a randomized experiment to estimate the effect of a treatment, but we do 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 biased by heteroskedasticity
  - C. effects that regress toward the mean
  - D. effects that differ between subpopulations

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

[Items 74 and 75] The image and output below are from a linear regression on data from the ANES 2020 Time Series Study that used a participant's age group (in a variable called AGEGRP, measured in three levels, of "Age 18 to 39", "Age 40 to 65", and "Age 66 to 80", with the first category as the omitted category) to predict the participant's rating about the Federal Bureau of Investigation (in a variable called FTFBI) on a 0-to-100 scale.

```
reg FTFBI i.AGEGRP
```

	FTFBI	Coefficient	p-value	[95% conf. interval]	
	AGEGRP				
	Age 40 to 65	4.64	0.000	3.35	5.94
	Age 66 to 80	6.02	0.000	4.57	7.47
constant/intercept		59.80	0.000	58.79	60.80

74. What does the 6.02 coefficient estimate for "Age 66 to 80" indicate?

- A. The mean rating about the FBI is predicted to be 6.02 among participants who are in the "Age 66 to 80" category.
- B. The mean rating about the FBI is predicted to be 6.02 higher among participants who are in the "Age 66 to 80" category, compared to participants at the average age.
- C. The mean rating about the FBI is predicted to be 6.02 higher among participants who are in the "Age 66 to 80" category, compared to participants who are in the "Age 18 to 39" category.
- D. None of the above

75. What does the 59.80 coefficient estimate for the constant/intercept indicate?

- A. The mean rating about the FBI is predicted to be 59.80 among the average participant.
- B. The mean rating about the FBI is predicted to be 59.80 among participants who are in the "Age 18 to 39" category.
- C. The mean rating about the FBI is predicted to be 59.80 among participants who are zero years old.
- D. None of the above

---

76. X positively associated with Y among the 200 members of Group A. X positively associated with Y among the 200 members of Group B. If members of Group A are combined with members of Group B into a 400-member Group C, then among members of Group C \_\_\_\_.

- A. X will definitely positively associate with Y
- B. X will definitely NOT positively associate with Y
- C. X might positively associate with Y, and X might not positively associate with Y

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

[Items 77 through 79] The output below is from a linear regression on data from the ANES 2020 Time Series Study that used a participant's age (in a variable called AGE, measured from age 18 to age 80) to predict the participant's rating about the Federal Bureau of Investigation (in a variable called FTFBI) on a 0-to-100 scale.

```
. reg FTFBI AGE
```

FTFBI	Coefficient	p-value	[95% conf. interval]	
AGE	0.14	0.000	0.11	0.17
_cons	56.15	0.000	54.41	57.88

77. What does the 56.15 intercept coefficient indicate?

- A. The predicted rating about the FBI among respondents who have an age of zero.
- B. The predicted rating about the FBI among respondents who have an age of 18.
- C. The predicted rating about the FBI among respondents who have an average age.
- D. The predicted rating about the FBI among respondents who have an age of 80.

78. What does the 0.14 coefficient for AGE indicate?

- A. AGE increases by 0.14 for each one-unit increase in FTFBI.
- B. The predicted rating about the FBI increases by 0.14 for each one-unit increase in AGE.
- C. The predicted rating about the FBI is 0.14 higher for older respondents than for younger respondents.

79. Which of the following is a correct linear regression equation for the output in the image, using X and Y?

- A.  $Y = 56.15 + 0.14$
- B.  $Y = 56.15X + 0.14$
- C.  $Y = 56.15 + 0.14X$
- D.  $Y = 56.15X + 0.14X$

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

[For items 80 through 82, see the regression output below and the description of the regression analysis on the prior page.]

```
. reg FTFBI AGE
```

FTFBI	Coefficient	p-value	[95% conf. interval]	
AGE	0.14	0.000	0.11	0.17
_cons	56.15	0.000	54.41	57.88

80. Did the analysis provide sufficient evidence to conclude, at the conventional level in political science, that a respondent's age associates with ratings about the FBI, at least on average among participants in the sample?

- A. Yes
- B. No

81. Did the analysis provide sufficient evidence to conclude, at the conventional level in political science, that going getting older caused a respondent to have a higher rating about the FBI, at least on average among participants in the sample?

- A. Yes
- B. No

82. The 95% confidence interval for AGE is [0.11, 0.17]. Which of the following is more likely to be the 99% confidence interval for AGE?

- A. [0.10, 0.18]
- B. [0.12, 0.16]

---

83. Of the following, which is a better reason to weight the average effect size estimate from a meta-analysis of studies by the sample size of the studies?

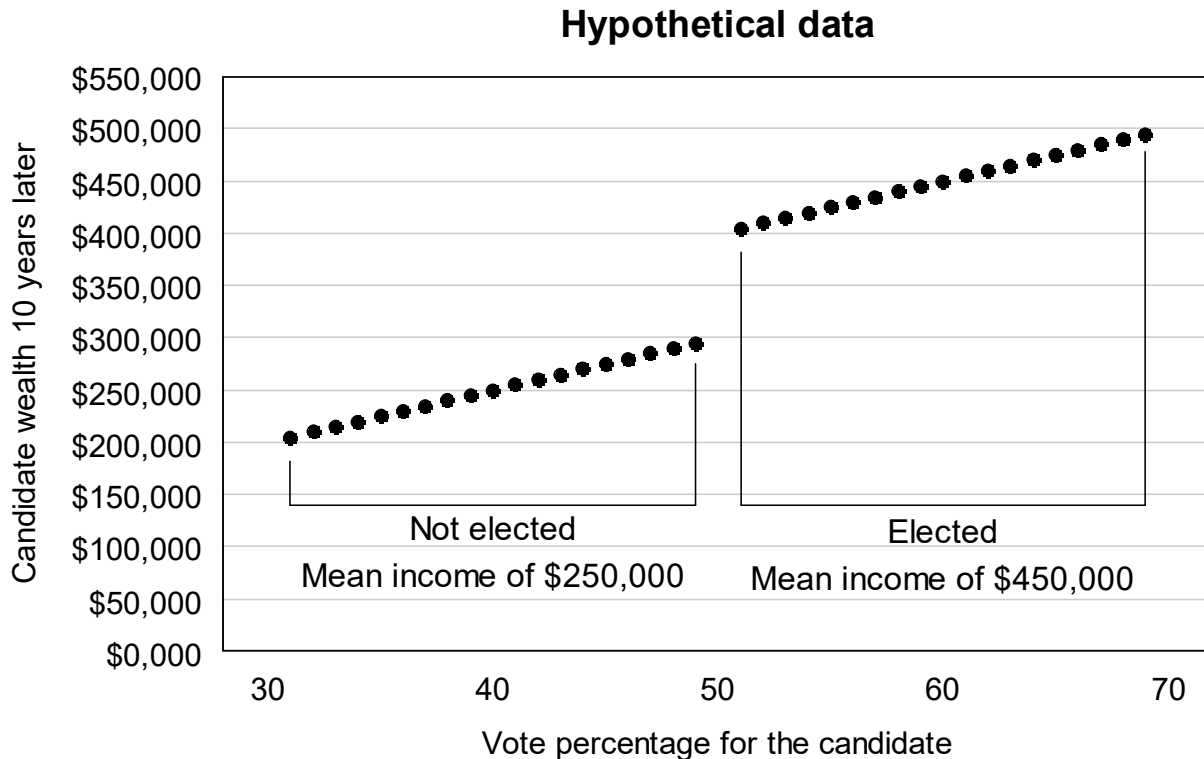
- A. Because this will increase the estimated effect size of the studies
- B. Because, compared to smaller sample studies, larger sample studies provide more evidence about an effect

84. 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?

- A. Yes
- B. No

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

85. Suppose that the data in the plot below are for political candidates in an election in a fictional country Freedonia. The x-axis indicates the percentage of the vote that a candidate received in an election, and the y-axis indicates the candidates' later-in-life wealth. In the plot, candidates who received more than 50 percent of the vote were elected into political office, and candidates who received less than 50 percent of the vote were not elected into political office.



Suppose that we use a discontinuity method to estimate how much being elected into office affects a candidate's wealth ten years later, by comparing how much the wealth among candidates who were just below the threshold for being elected differs from the wealth among candidates who were just above the threshold for being elected. Which of the following best indicates that estimate?

- A. Being elected did not increase candidate wealth, on average.
- B. Being elected increased candidate wealth by about \$100,000, on average.
- C. Being elected increased candidate wealth by about \$200,000, on average.
- D. Being elected increased candidate wealth by about \$400,000, on average.
- E. Being elected increased candidate wealth by about \$450,000, on average.
- F. Being elected increased candidate wealth by about \$500,000, on average.

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

[Items 86 through 88] The table below indicates that, among a hypothetical set of workers at a company, 50% of women were promoted and 50% of men were promoted.

Women				Men			
Promoted	Sales	Client Rating	Experience	Promoted	Sales	Client Rating	Experience
No	\$400	★★	0 years	No	\$400	★★	0 years
No	\$400	★★★★	5 years	No	\$400	★★	0 years
No	\$400	★★★★	5 years	No	\$400	★★	0 years
Yes	\$900	★★★★	5 years	Yes	\$900	★★	0 years
Yes	\$900	★★★★	5 years	Yes	\$900	★★★★	5 years
Yes	\$900	★★★★	5 years	Yes	\$900	★★★★	5 years

86. Controlling only for sales, the company's promotion practices seem \_\_\_\_.

- A. biased for women
- B. biased against women
- C. to not have a bias for or against women

87. Controlling only for client rating, the company's promotion practices seem \_\_\_\_.

- A. biased for women
- B. biased against women
- C. to not have a bias for or against women

88. Controlling only for experience, the company's promotion practices seem \_\_\_\_.

- A. biased for women
- B. biased against women
- C. to not have a bias for or against women

---

89. Between 2000 and 2015, women authored about 20 percent of articles published in the top three political science journals published in the United States. If we wanted to assess the extent to which this percentage was affected by gender bias among the editors of these journals who make the decisions about what to publish, of the following, which would be the best benchmark percentage?

- A. the percentage female among U.S. adults between 2000 and 2015
- B. the percentage female among U.S. political scientists between 2000 and 2015
- C. the percentage female among U.S. political scientists at top universities between 2000 and 2015

90. All else equal, which 95% confidence interval below indicates the more informative null?

- A. [-2,3]
- B. [-5, 5]
- C. [-50, 50]

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

91. Researchers are interested in whether a pill that contains Chemical X causes people to lose weight. Researcher A has a random sample of 400 persons. Researcher A directs each person to take the Chemical X pill once per day for six months. The mean weight loss among these 400 persons over the six months was 7 pounds, with a p-value of  $p=0.001$  for a test of the null hypothesis that the mean weight loss was zero pounds.

Researcher B has a random sample of 400 persons. Researcher B randomly assigned 200 of the persons to take the Chemical X pill once per day for six months, and randomly assigned the other 200 persons to take a placebo pill that has no effect once per day for six months. The mean weight loss among persons who took the Chemical X pill was 7 pounds, and the mean weight loss among persons who took the placebo pill was 3 pounds, and the p-value was  $p=0.010$  for a test of the null hypothesis that the mean weight loss among the persons who took the Chemical X pill equaled the mean weight loss among the persons who took the placebo pill.

Who had the better research design for inferring whether a pill with Chemical X causes weight loss?

- A. Researcher A, because Researcher A had a lower p-value
  - B. Researcher A, because Researcher A had a larger number of participants take the Chemical X pill
  - C. Researcher B, because Researcher B had a higher p-value
  - D. Researcher B, because Researcher B used a randomized experiment
92. Suppose that the final exam in POL 138 has 100 multiple-choice items, each with four response options, and suppose that the instructor randomly ordered the items, so that there is no association between the order of the items and how difficult the items are. This POL 138 section has 110 students who took the final exam. Suppose that, among the 25 *highest-scoring* students on the even-numbered items on the final exam, the mean percentage correct is 90%. Which of the following should be expected for the mean percent correct among these students on the odd-numbered items on the final exam?
- A. that mean percentage correct should also be 90%
  - B. that mean percentage correct should be lower than 90%
  - C. that mean percentage correct should be higher than 90%
93. Suppose that the final exam in POL 138 has 100 multiple-choice items, each with four response options, and suppose that the instructor randomly ordered the items, so that there is no association between the order of the items and how difficult the items are. This POL 138 section has 110 students who took the final exam. Suppose that, among the 25 *lowest-scoring* students on the even-numbered items on the final exam, the mean percentage correct is 35%. Which of the following should be expected for the mean percent correct among these students on the odd-numbered items on the final exam?
- A. that mean percentage correct should also be 35%
  - B. that mean percentage correct should be lower than 35%
  - C. that mean percentage correct should be higher than 35%

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

94. Suppose that Researcher A conducts a well-designed large-sample randomized experiment to test the null hypothesis that a treatment has no effect. In reality, the treatment has no effect. What is the expected probability that the p-value for the test of the null hypothesis is  $p=0.05$  or less?
- A. 0%
  - B. above 0% but less than 5%
  - C. 5%
  - D. above 5%
  - E. Cannot be determined without more information
95. Suppose that Researcher A conducts a well-designed large-sample randomized experiment to test the null hypothesis that a treatment has no effect. In reality, the treatment has an effect. What is the expected probability that the p-value for the test of the null hypothesis is  $p=0.05$  or less?
- A. 0%
  - B. above 0% but less than 5%
  - C. 5%
  - D. above 5%
  - E. Cannot be determined without more information
96. Suppose that Researcher A conducts two independent well-designed large-sample randomized experiments to test the null hypothesis that a treatment has no effect. In reality, the treatment has no effect. What is the expected chance that the p-value for at least one of the two tests of the null hypothesis is  $p=0.05$  or less?
- A. 0%
  - B. above 0% but less than 5%
  - C. 5%
  - D. above 5%
  - E. Cannot be determined without more information

97. Below are data for two studies:

Study	Sample size	Estimated effect size
A	100	4
B	400	5

Which of the following is a correct formula for calculating the mean estimated effect size, weighted by sample size?

- A.  $(4 \times 100) + (5 \times 400)$
- B.  $4 \times (100 \div 400) + 5 \times (400 \div 100)$
- C.  $4 \times (100 \div 500) + 5 \times (400 \div 500)$
- D.  $4 \div (100 \div 500) + 5 \div (400 \div 500)$

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

98. 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 lower 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
99. 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
100. Suppose that a randomized experiment tested the null hypothesis that a treatment has no effect. The experiment detected sufficient evidence to reject the null hypothesis for female participants ( $p = 0.01$ ) but did not detect sufficient evidence to reject the null hypothesis for male participants ( $p = 0.15$ ). Does the combination of these results provide sufficient evidence to conclude at the conventional level in political science that the treatment had a larger effect among female participants than among male participants?
- A. Yes
  - B. No